UNIVERSIDAD DE GRANADA

Departamento de Ciencias de la Computación e Inteligencia Artificial

Programa de Doctorado en Tecnologías de la Información y de la Comunicación

Nuevos Métodos basados en Soft Computing para Calibración de Modelos Basados en Agentes: Aplicaciones en Marketing y Ciencias Políticas

Memoria que presenta

Ignacio Moya Señas

Para optar al grado de Doctor

Julio 2021

Directores

Manuel Chica Serrano
Oscar Cordón García

Editor: Universidad de Granada. Tesis Doctorales
Autor: Ignacio Moya Señas
ISBN: 978-84-1117-058-1
URI: http://hdl.handle.net/10481/71116

El doctorando / The *doctoral candidate* **Ignacio Moya** y los directores de la tesis / and the thesis supervisor/s: **Manuel Chica** y **Oscar Cordón**

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

/

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date:

Granada, 20 de julio de 2021

Director/es de la Tesis / *Thesis supervisor/s;*

CHICA SERRANO MANUEL - 77341576M
Digitally signed by CHICA SERRANO MANUEL - 77341576M
Date: 2021.07.15 11:57:27 +01'00'

CORDON GARCIA OSCAR - 45281118Y
Firmado digitalmente por CORDON GARCIA OSCAR - 45281118Y
Fecha: 2021.07.15 13:42:56 +02'00'

Firma / Signed

Doctorando / *Doctoral candidate*:

MOYA SEÑAS IGNACIO - 11857158Z
Firmado digitalmente por MOYA SEÑAS IGNACIO - 11857158Z
Fecha: 2021.07.14 18:16:43 +02'00'

Firma / Signed

Caminante, son tus huellas
el camino, y nada más;
caminante, no hay camino:
se hace camino al andar.
Al andar se hace camino,
y al volver la vista atrás
se ve la senda que nunca
se ha de volver a pisar.
Caminante, no hay camino,
sino estelas en la mar.

– Antonio Machado, *Proverbios y cantares*

Agradecimientos

"I do not care who knows the truth now, tomorrow, or in ten thousand years. Loyalty is its own reward."
– Lion El'Jonhson, *Age of Darkness*

Esta tesis no habría sido posible sin el trabajo, el apoyo y el compromiso de mis directores. Tanto Manu como Oscar se han mostrado siempre cercanos y accesibles para cualquier cosa que les necesitara, tanto en lo personal como en lo profesional, y por ello les estaré siempre agradecido. También les agradezco el haber contado conmigo para el extinto proyecto NEWSOCO, proyecto que financió mi contrato y me dio cuatro años de tranquilidad económica mientras llevaba a cabo esta tesis. Así mismo me gustaría agradecer a Sergio la oportunidad que me dio al incorporarme a su equipo en el ECSC en Mieres, hace ya más de siete años, que no fue si no el principio de esta historia. Por lo mismo quiero agradecer a Jose su confianza durante los inicios de ZIO, que desde luego fueron entretenidos.

Tampoco creo que hubiera podido llegar aquí sin el apoyo de mi familia, que me proporcionó los medios para poder estudiar y me apoyó cuando les planteé la idea de cambiar Madrid por Asturias primero y por Granada después. Siempre tuvieron fe en que alcanzaría mis objetivos y no les puedo estar más agradecido por su cariño y su confianza. Gracias también a mi Ximenita, que es a mi vida lo que esa taza de café al levantarse de la cama: se empieza a tomar porque te gusta y sin darte cuenta ya no puedes vivir sin ella.

Me gustaría agradecerles su paciencia a todos los compañeros de fatigas que me han acompañado por los distintos despachos de Madrid, Mieres y el CITIC-UGR. Aunque la pandemia me haya terminado por transformar en alguien que prefiere trabajar en remoto, siempre he tenido un ambiente de trabajo fantástico en todos los despachos por los que he pasado y creo que he podido hacer muy buenos amigos. Prefiero no mencionar nombres porque iba a ser inevitable que se me olvidara alguno y me parece un feo innecesario, pero espero que el lector me permita hacer una excepción con Juanfrita. Gracias amigo por estar ahí desde el primer día que llegué a Granada, tanto en el trabajo como fuera de él. Sobretodo gracias por el día en que me presentaste a Ximena y mi vida no volvió a ser la misma.

También quiero agradecer a todos mis amigos, los que fueron compañeros de la universidad y los que no, por mostrar interés por mis progresos durante estos años y ayudarme a hacerlos más llevaderos. Por último, permíteme agradecerte a ti, lector, tu tiempo al leer estas líneas. Teniendo en cuenta que el grueso del material de la tesis está publicado en los artículos, dudo que este documento te pueda aportar mucho más que una vista de conjunto. Aun así, gracias y ten en cuenta que cada cita que me llega me hace mucha ilusión. Al fin y al cabo, *el negocio de esto está en publicar.*

Contents

Resumen

1 Introducción al problema

El modelado basado en agentes es una reconocida metodología de modelado que tiene como objetivo diseñar y simular modelos que utilizan entidades autónomas llamadas agentes [Eps06, JO06]. El comportamiento de estos agentes artificiales sigue reglas sencillas que definen sus acciones y reglas sociales que definen su interacción tanto con otros agentes como con el entorno. La capacidad de los modelos basados en agentes (ABM, por sus siglas en inglés) para recrear dinámicas emergentes y sistemas complejos a través de la agregación de dichas reglas les ha permitido ganar relevancia en los últimos años [Wal18]. En concreto, sus características les convierten en herramientas muy útiles para analizar fenómenos sociales, dado que permiten reproducir procesos de difusión y mecanismos *word-of-mouth* (WOM) usando redes sociales artificiales [CR17]. De este modo, la simulación de ABMs es usada con frecuencia para construir sistemas de soporte a la decisión que funcionan definiendo escenarios hipotéticos y *what-if*. Además, el enfoque *bottom-up* de los ABMs permite estudiar estos fenómenos tanto desde una perspectiva micro (a nivel de agente) como desde una perspectiva macro (a nivel de sistema).

Sin embargo, la construcción y utilización de estos modelos entraña mucha dificultad porque pueden contar con gran número de parámetros. Además, muchos de estos parámetros son susceptibles de ser estimados debido a la falta de datos que justifiquen un valor específico. El proceso de ajuste de parámetros del modelo para que se replique el comportamiento deseado se denomina calibración del modelo. Este proceso se puede llevar a cabo de forma automática [CR17] empleando un proceso computacional intensivo que ajusta los parámetros del modelo con un método de optimización. Este método de optimización considera una medida de error para comparar la salida del modelo con los datos históricos del fenómeno que se quiere reproducir. Dado que típicamente los parámetros de un ABM tienen interacciones no lineales complejas entre ellos, resultan apropiados los métodos de optimización como las metaheurísticas [Tal09], capaces de explorar todo el espacio de búsqueda. En este sentido, se han aplicado con éxito metaheurísticas muy asentadas como los algoritmos genéticos [BFM97]. Dado que, en principio, la calidad de la configuración final depende de la capacidad de la metaheurística elegida para explorar el espacio de búsqueda de los parámetros del modelo, la aplicación de metaheurísticas más avanzadas podría mejorar la calidad de las configuraciones calibradas.

En cualquier caso, después de aplicar el método de optimización, el modelador necesita revisar y validar los valores de los parámetros obtenidos, dado que un buen valor de ajuste a los datos históricos no garantiza la validez del modelo [Sar05]. Además, muchos modelos se diseñan considerando dos o más indicadores de rendimiento (KPIs, por sus siglas en inglés) referidos al comportamiento que se pretende simular. La calibración de estos modelos puede entenderse como un proceso de toma de decisiones multicriterio similar a un problema de optimización multiobjetivo [CLVV$^+$07], donde no existe una única configuración que satisfaga todos objetivos

sino que el usuario puede elegir una configuración a partir de un conjunto de soluciones igualmente preferibles (esto es, un conjunto de Pareto). Podemos observar que en este escenario la validación del modelo es incluso más compleja dado que hay varios conjuntos de configuraciones de parámetros potencialmente seleccionables por el modelador.

Por otra parte, el uso efectivo de la simulación de estos modelos requiere de un gran nivel de transparencia por parte del ABM subyacente. Los usuarios y los diseñadores de un ABM deben entender cómo el modelo recrea el comportamiento que se está estudiando en los distintos escenarios *what-if* propuestos. Esto puede conseguirse desde una perspectiva *white-box*, donde tanto los diseñadores como los usuarios de los modelos cuenten con herramientas que les permiten aumentar la explicabilidad del modelo. Este enfoque encaja con los objetivos del área emergente de *explainable artificial intelligence* [BDRD$^+$20, SWM17], que anima a los investigadores a abrir los modelos *black-box* con el fin de hacerlos transparentes y que su funcionamiento sea más fácil de explicar. Este enfoque aumenta la credibilidad de las soluciones obtenidas mediante estas técnicas, dado que aumentar la transparencia de las configuraciones calibradas debería incrementar la confianza en su comportamiento y su rendimiento. En este aspecto podemos destacar el uso de métodos y técnicas de visualización para calibración de modelos ABM, dado que son herramientas útiles para aumentar la comprensión del diseñador sobre el modelo y sus parámetros, lo que facilita la validación del modelo.

Como el éxito de estos modelos depende en parte de lo bien que ajusten su comportamiento a los datos históricos disponibles, los métodos utilizados para llevar a cabo calibración automática deben ser lo más efectivos posible durante la exploración del espacio de búsqueda. En este caso, es necesario estudiar la mejora en los resultados de calibración por parte de metaheurísticas novedosas, puesto que tanto en los modelos mono-objetivo como los modelos multiobjetivo se suelen aplicar metaheurísticas bien establecidas que podrían ser superadas por otras más recientes. Además, es necesario estudiar cómo el aumento de la dimensión del problema afecta al rendimiento de estas metaheurísticas, puesto que los modelos ABM suelen tener gran número de parámetros.

Por otra parte, es necesario mejorar los métodos de validación de modelos calibrados con métodos automáticos, especialmente en el caso de los modelos con múltiples KPIs. Esto podemos conseguirlo introduciendo un entorno de trabajo integral que combine dos elementos: algoritmos de optimización evolutiva multiobjetivo (EMO, por sus siglas en inglés) para calibrar los parámetros de modelos ABM automáticamente a partir de datos históricos, y un método avanzado de visualización que mejore la comprensión del proceso de calibración y de sus resultados [TCP$^+$18]. Este enfoque integral podría mostrar cómo la visualización es una herramienta indispensable para la calibración automática, mejorando la comprensión del modelo calibrado asistiendo al modelador durante el proceso de validación.

La calibración de ABMs tiene como objetivo la obtención de modelos que permitan reproducir y analizar distintos fenómenos sociales con la mayor fiabilidad posible. Específicamente, podemos identificar dos nichos de aplicabilidad de simulaciones sociales en el análisis de escenarios de ciencias políticas y en la definición de estrategias de marketing. En cuanto a las ciencias políticas, la simulación de modelos ABM puede permitir analizar el comportamiento del votante utilizando las distintas teorías de voto que se manejan actualmente [Mui10]. En cuanto a las estrategias de marketing, los modelos ABM son especialmente relevantes para evaluar escenarios *what-if*, donde la simulación tiene como objetivo evaluar la respuesta del mercado a distintas alternativas de marketing [CR17].

2 Desarrollo realizado

El trabajo desarrollado durante esta tesis se divide en cuatro bloques principales que tratan los distintos problemas identificados con el objetivo de mejorar los procesos de calibración y validación

de modelos ABM con cualquier número de KPIs y su aplicación a dos dominios concretos, las ciencias políticas y el marketing. Las siguientes subsecciones atienden a los bloques identificados. En primer lugar, se define el diseño e implementación de un sistema de simulación que permita instanciar distintos modelos ABM con aplicación a distintos escenarios de ciencias políticas y marketing, tanto mono-objetivo como multiobjetivo. Después, se considera el diseño, desarrollo y validación de un conjunto de algoritmos de calibración automática basados en metaheurísticas novedosas para aplicarlos a modelos mono-objetivo (es decir, con un único KPI). Posteriormente se atiende la extensión y validación de los métodos de calibración anteriores, con el fin de incorporar metaheurísticas multiobjetivo e incorporarlos a modelos ABM con múltiples KPIs. Finalmente, se considera el diseño de un entorno de trabajo integral para la validación de modelos multiobjetivo que combine los algoritmos de calibración multiobjetivo diseñados con un método de visualización avanzado.

2.1 ABMs para escenarios de ciencias políticas y marketing

El trabajo de esta tesis comenzó por el diseño e implementación de distintos ABMs para problemas de ciencias políticas y marketing que luego sirvieran de base sobre los que luego utilizar las técnicas de calibración y validación desarrolladas posteriormente. En cuanto a los escenarios de ciencias políticas, se diseñaron dos ABMs distintos que describen teorías distintas dentro de las ciencias políticas: la teoría del marco (*framing*) [CD07] y la teoría espacial del voto [EH84, EH94]. La teoría del marco se fundamenta en el efecto *framing*, que explica el proceso psicológico que permite a las personas desarrollar una conceptualización *ad hoc* de un evento, llegando incluso a reajustar su opinión. Esta teoría encaja con los eventos ocurridos en las elecciones generales de 2004, celebradas tres días después de los atentados de Atocha el 11 de marzo (11-M). Tras los atentados se generó una cantidad de información por parte del gobierno, políticos y medios de información, forzando a los candidatos a posicionarse con respecto a los atentados, una posición que los votantes incorporarían a su decisión de voto [Hol96]. Este evento tuvo un marco comunicativo que se estructuró en torno a dos *framings*: la autoría de los atentados por parte de ETA o Al'Qaeda. La primera era la posición defendida por el gobierno del Partido Popular (PP), mientras que el principal partido de la oposición, el Partido Socialista Obrero Español (PSOE), y otras fuerzas políticas defendían la segunda [Olm05]. Finalmente, el Centro de Investigaciones Sociológicas (CIS) estimó que el 11% de los votantes cambiaron su voto tras los atentados [Cen04].

El contexto del 11-M también puede ser estudiado desde el punto de vista de la teoría espacial del voto y de la teoría de gestión del terror. La teoría de gestión del terror [LSG+04, WA08] es una línea de investigación donde múltiples estudios sugieren que un conflicto político violento puede tener un efecto significativo en la opinión pública o incluso decidir unas elecciones [Bal07, Ran18, RHM13, RMA07]. En este sentido, posiciones defendidas por los partidos claramente implementaban una estrategia política con el objetivo de influenciar la decisión de voto. Además, los medios de comunicación fueron el canal de transmisión usado por los partidos para difundir esta estrategia debido a la gran demanda de noticias tras los atentados por parte de la opinión pública [PS15]. Por tanto, el ABM desarrollado nos permitió estudiar la influencia de las distintas estrategias de los partidos en la distancia ideológica de los votantes. La teoría espacial del voto propone que los partidos políticos y los votantes se posicionan en un medio unidimensional que refleja sus posiciones respecto a una cuestión política [Dow57, EH84, EH94, GG20, Gro85, Ked05, Ked09, RM89]. Por tanto, se puede definir la distancia ideológica como la distancia entre los partidos y los votantes dentro de este medio, asumiendo que los votantes eligen al partido más próximo. En respuesta al posicionamiento de los votantes, los partidos siguen una lógica utilitaria que les lleva a posicionarse donde pueda minimizar su distancia con cada uno de los votantes. Finalmente, usando este modelo podemos diseñar y validar distintos escenarios que modifican el mensaje de los medios de comunicación, produciendo distintos efectos. Estos escenarios son *the*

rally around the flag [Mue73], *the opinion leadership* [Cho11] y *the priming of public opinion and media coverage* [BS89].

En cuanto a los problemas de marketing, se optó por un ABM orientado al modelado del reconocimiento de marca y del número de conversaciones. Estos indicadores, también referidos como *brand awareness* y WOM *volume* en inglés, juegan un papel importante en la expansión de mercados y por ello han conseguido bastante relevancia a la hora de entender las dinámicas de problemas de marketing actuales [LMP13, MS00]. Una vez diseñado el ABM, se utilizaron datos reales de un escenario de banca en España para implementar e instanciar el modelo. Este primer modelo se utilizó después para componer un banco de pruebas de hasta 15 instancias con un número de parámetros incremental, llegando hasta las 175 variables de decisión.

2.2 Framework integral para calibración y validación de ABMs

Durante el desarrollo de la tesis se diseñó un entorno de trabajo para calibración y validación de ABMs y otros modelos de simulación por eventos discretos que consideren más de un KPI. Este entorno de trabajo sigue un enfoque integral que combina dos elementos: un algoritmo EMO y un método de visualización avanzado basado en técnicas de análisis de redes sociales. Por un lado, el algoritmo EMO permite calibrar automáticamente los parámetros de los ABMs usando datos históricos. Por otro, el método de visualización avanzado permite mejorar la comprensión del proceso de calibración y sus resultados, permitiendo además obtener conocimiento adicional sobre el modelo calibrado [Oli03, Sar05].

El algoritmo EMO es un componente principal del entorno de trabajo. Los algoritmos EMO son metaheurísticas poblacionales que pueden obtener un conjunto de configuraciones distintas para los parámetros calibrados en una única ejecución. Siguiendo este diseño del problema de calibración, cada solución del problema se codifica con distintas variables de decisión que representan cada configuración del ABM, siendo estas variables de decisión valores reales o enteros a conveniencia. Las configuraciones obtenidas por los algoritmo EMO tienen distintos valores en el espacio objetivo y componen una aproximación al conjunto Pareto-optimal. En principio, cualquier algoritmo EMO puede llevar a cabo este proceso y debería seleccionarse el más adecuado teniendo en cuenta las características del modelo a calibrar. Por ejemplo, si el problema contempla menos de cuatro KPI cualquiera de los algoritmos EMO más asentados como NSGA-II [DPAM02], SPEA2 [ZLT01] o MOEA/D [LZ09] pueden ser elegidos. En caso de tener cuatro o más, el problema de calibración debería tratarse como *many-objective* y el algoritmo EMO elegido debería rendir apropiadamente en este entorno. Algunos ejemplos de algoritmos EMO para optimización con cuatro o más objetivos pueden ser NSGA-III [DJ14], HypE [BZ11], GrEA [YLLZ13] o KnEA [ZTJ15].

En cuanto al método de visualización, se eligió *moGrams* [TCP+18] para representar los resultados de calibración. Esta metodología combina la visualización de soluciones no-dominadas tanto en el espacio de diseño como en el espacio objetivo. Un *moGram* es un grafo ponderado donde cada nodo representa una solución de la aproximación del conjunto de Pareto y cada arista representa una relación de similitud entre dos de estas soluciones en el espacio de variables. A través de *moGrams*, el usuario puede mejorar su comprensión del problema de calibración al poder identificar grupos de soluciones, detectar las más flexibles (aquellas que pueden ser intercambiadas por otras con un cambio mínimo en sus variables de decisión) y validar convenientemente la selección de parámetros a calibrar mediante las relaciones entre soluciones.

2.3 Algoritmos de calibración automática basados en metaheurísticas para ABM mono-objetivo

Como ya se ha señalado, la calidad de los resultados de calibración automática está supeditada al rendimiento del método de optimización seleccionado. En particular en el caso de los ABMs,

estos modelos suelen considerar muchos parámetros con interacciones complejas entre sí, por lo que los métodos de búsqueda no lineal son preferibles. Debido a esta circunstancia, distintas metaheurísticas han obtenido buenos resultados [CH06, CC15, CR17, Fab13, HS15, MSEH14]. Sin embargo, los trabajos anteriores tienden a utilizar diseños simples de metaheurísticas clásicas y no han considerado la aplicación de algoritmos más avanzados. Durante el desarrollo de esta tesis se ha explorado este espacio mediante la aplicación del proceso de calibración usando novedosas metaheurísticas bio-inspiradas: *coral reefs optimization* (CRO) [SSDSLT$^+$14] y *coral reefs optimization with substrate layers* (CRO-SL) [SSCGMH16], demostrando un gran rendimiento. Estas metaheurísticas emulan la formación y reproducción de arrecifes de corales, dando lugar a un algoritmo de optimización con un potente equilibrio entre exploración y explotación del espacio de búsqueda. CRO-SL es una versión mejorada de CRO que incluye un esquema de co-evolución en una única población que ya ha demostrado buenos resultados tratando distintos problemas de optimización complejos [BCD$^+$18, CGMMGGSS19, GHGHSM$^+$20, GHSMCM$^+$20, SSGHCG$^+$19].

Partiendo de estos resultados, hemos analizado el rendimiento de las metaheurísticas basadas en corales en el problema de calibración de ABMs. Este análisis incluye una comparación exhaustiva contra un conjunto diverso de metaheurísticas que incluye métodos bien asentados como *differential evolution* (DE) [SP97] y métodos más avanzados como *success-history based adaptive differential evolution with linear population size reduction* (L-SHADE) [TF14] y *restart CMA-ES with increasing population size* (IPOP-CMA-ES) [AH05]. Por último, se incluyeron modificaciones meméticas de todos los algoritmos considerados para estudiar la mejora producida por su hibridación con una búsqueda local de tipo *hill climbing* [RNI95].

El estudio compara el rendimiento de las metaheurísticas al calibrar 12 escenarios de ABMs para marketing con dimensionalidad creciente desde 24 hasta 129 parámetros. Estas instancias también se generaron a partir de un mercado real durante el desarrollo de la tesis (véase sección 2.1). Después de aplicar distintos tests estadísticos a los resultados, nuestro análisis concluyó que tanto CRO-SL como la versión memética de CRO tienen un rendimiento significativamente mejor para este problema. L-SHADE e IPOP-CMA-ES obtienen resultados cercanos a los de los algoritmos basados en corales para la mayoría de escenarios pero son superados en el *ranking* medio. En cuanto a la mejora de las versiones meméticas de los algoritmos, solo la mitad de los algoritmos mejoran su rendimiento mediante su hibridación con la búsqueda local elegida. En concreto, la versión memética de CRO-SL reduce su rendimiento según se incrementa la dimensionalidad de los escenarios, por lo que se puede observar que la búsqueda local afecta negativamente al buen equilibrio de CRO-SL entre exploración y explotación.

2.4 Algoritmos de calibración automática basados en metaheurísticas para ABMs con múltiples KPIs

De manera similar a los ABMs con un único KPI, la calidad de los resultados de calibración de ABMs con más de un KPI está muy influenciada por el método de optimización elegido. En concreto, hay gran cantidad de algoritmos EMO en la literatura especializada y seleccionar el mejor algoritmo para llevar a cabo el proceso de calibración automática puede ser una tarea muy exigente. Durante el desarrollo de esta tesis doctoral atendimos este asunto llevando a cabo un análisis exhaustivo del rendimiento de un grupo relevante de algoritmos EMO al calibrar un conjunto de instancias de un ABM con dos KPIs. Estas instancias fueron también generadas durante el desarrollo de la tesis (véase sección 2.1).

Los algoritmos EMO seleccionados pertenecen a las principales categorías en este tipo de algoritmos: basados en dominancia, indicadores y descomposición. Estos métodos son *non-dominated sorting genetic algorithm II* (NSGA-II) [DPAM02], *improved strength Pareto evolutionary algorithm* (SPEA2) [ZLT01], *general indicator-based evolutionary algorithm* (IBEA) [ZK04], *S metric selection multiobjective optimization algorithm* (SMS-EMOA) [BNE07] y

multiobjective evolutionary algorithm based on decomposition (MOEA/D) [LZ09, ZL07]. También se incluyeron dos algoritmos EMO recientes que han mostrado resultados competitivos resolviendo problemas reales [FCC18, MT17, ZVNGNAM17]. Estos son *many-objective metaheuristic based on the R2 indicator II* (MOMBI2) [HGC15] y *global weighting achievement scalarizing function genetic algorithm* (GWASF-GA) [SRL17]. Finalmente, se incluyó un método de optimización clásico en el estudio para que sirviera de *baseline* a los EMO analizados, el método *Nelder-Mead's simplex* [NM65]. Dado que este método está pensado para optimización mono-objetivo, fue necesario adaptarlo usando el método *adaptive* $\epsilon-$*constraint*, lo que constituye un recurso habitual [Eic08, Mav09].

Para analizar el rendimiento de los distintos algoritmos EMO se utilizaron principalmente indicadores de rendimiento multiobjetivo y *attainment surfaces* [ZTL$^+$03]. Además, se utilizó el test estadístico de Wilcoxon para analizar la significación estadística de los resultados de los indicadores [GMLH08]. Finalmente, se analizaron en profundidad los inconvenientes observados en el rendimiento de los distintos algoritmos EMO y se analizó la influencia de las propiedades de la instancias del banco de pruebas en el comportamiento de los algoritmos. Específicamente, las propiedades identificadas con mayor influencia en el rendimiento de los algoritmos fueron la dimensionalidad del problema, la forma de la región factible del espacio de búsqueda y la forma del frente de Pareto [ISMN17, MMNI18].

3 Conclusiones y Trabajos Futuros

Se puede afirmar que las técnicas de simulación como los ABMs son uno de los enfoques principales para el análisis de sistemas complejos, dado que permiten recrear en entornos computacionales los fenómenos que se pretenden estudiar. Sin embargo, estas técnicas deben utilizarse con rigor para asegurar su efectividad y no llegar a conclusiones erróneas por el uso de modelos no contrastados. La calibración y validación de ABMs son herramientas necesarias pero su uso no es trivial, dado que están condicionados por los datos y herramientas disponibles. La principal motivación de esta tesis doctoral es mejorar la calibración y validación de modelos ABM con cualquier número de KPIs mediante el diseño de técnicas novedosas que permitan superar estas dificultades. En este aspecto, diferentes técnicas de *soft computing* e inteligencia artificial como las metaheurísticas han demostrado su eficacia abordando problemas caracterizados por un alto nivel de complejidad.

Esta tesis presenta una serie de contribuciones con marcado carácter multidisciplinar, puesto que son transversales al *soft computing*, las ciencias políticas y el marketing. La primera contribución la conforman el conjunto de ABMs diseñados e implementados durante la tesis. Dentro de este conjunto se encuentran los dos ABMs diseñados para analizar las elecciones del 14-M, que hasta donde llega nuestro conocimiento son los primeros ABMs desarrollados para este escenario político. El primero de ellos es un modelo basado en la teoría del *framing* donde la resiliencia de un conjunto de votantes se expone a la presión mediática con dos sesgos diferenciados respecto de la autoría de los atentados del 11-M. Estos votantes, modelados como agentes en el modelo, ejercen su derecho al voto al final de la simulación, llegando a cambiar su intención de voto dependiendo de su resiliencia. El segundo modelo es un ABM basado en la teoría espacial del voto donde los efectos del gran impacto del atentado se estudian respecto a su influencia en la distancia ideológica entre los votantes y los partidos. Esta distancia ideológica se modela siguiendo el modelo de proximidad de Downs [Dow57] y permite analizar como las estrategias comunicativas de los partidos pudieron haber influenciado la distancia ideológica de los votantes durante las elecciones del 14-M.

Junto con los modelos diseñados para ciencias políticas, se diseñó e implementó un ABM para marketing usando datos reales de uno de los proyectos de investigación del grupo. Este modelo incluye dos KPIs relevantes: el reconocimiento y el número de conversaciones realizadas por los consumidores sobre una marca. Este modelo presenta a un conjunto de agentes que se ven expuestos

a la influencia de otros agentes cercanos y a influencias externas, que representan la publicidad que hacen las marcas en los medios de comunicación. Este modelo se empleó principalmente para componer un *benchmark* sintético con múltiples instancias de dimensionalidad incremental. De este modo cada instancia del modelo incluye una serie de características propias, dado que tiene valores históricos distintos para sus KPIs y además incluye un número distinto de medios de comunicación, lo que involucra también un número distinto de variables de decisión para el problema de calibración.

La segunda contribución de la tesis fue la propuesta de un entorno de trabajo integral para calibración de ABMs basado en algoritmos EMO y una técnica de visualización avanzada basada en el análisis de redes sociales. Este enfoque integral aporta una serie de ventajas a la calibración de ABMs con más de un KPI que van más allá de conseguir modelos calibrados de calidad. En concreto, hemos mostrado estas ventajas aplicando el entorno de trabajo desarrollado a varios de los modelos para marketing desarrollados durante la tesis. Para la primera etapa de este enfoque integral, utilizamos *NSGA-II* para la obtención de la aproximación del frente de Pareto del problema. Después usamos esta aproximación para seleccionar y analizar tres soluciones relevantes: las dos con mejor ajuste para cada KPI y la solución con mejor equilibrio entre los dos KPI. A partir de este análisis pudimos observar cómo el ajuste del reconocimiento de marca era más difícil de conseguir que el ajuste del número de conversaciones. Posteriormente, la utilidad de este enfoque se puso de manifiesto en la siguiente fase del entorno de trabajo al validar soluciones relevantes, evaluando propiedades como su flexibilidad respecto a otras soluciones de su aproximación del conjunto de Pareto. Por ejemplo, se pudo observar como las soluciones con mejor equilibrio entre los dos KPIs tenían un buen nivel de flexibilidad pero carecían de vecinos interesantes en el espacio de decisión. Además, se pudo seleccionar y validar soluciones alternativas que habrían sido difíciles de identificar en frentes de Pareto muy densos.

La tercera contribución de la tesis la conforman dos estudios sobre el rendimiento de distintas metaheurísticas en la calibración de ABMs para marketing. Para estos dos estudios se utilizaron las instancias del banco de pruebas desarrollado durante la tesis, lo que permitió comparar los algoritmos en un entorno con distintas propiedades. El primero de ellos introduce un nuevo enfoque a la calibración de ABMs con metaheurísticas basadas en corales usando codificación entera. En este aspecto, tanto CRO como CRO-SL demostraron un rendimiento competitivo a través de la emulación del ciclo de vida de los corales, lo que les aporta un buen equilibrio entre exploración e intensificación. El rendimiento de estas metaheurísticas fue comparado de manera exhaustiva contra otros algoritmos evolutivos relevantes (L-SHADE y IPOP-CMA-ES) usando tests estadísticos. Además, se estudió la posible mejora de rendimiento de todos los métodos analizados mediante su hibridación con una búsqueda local. El estudio concluyó que tanto CRO-SL como la versión memética de CRO-SL tuvieron un rendimiento significativamente mejor que el de los otros métodos.

El segundo estudio evaluó el rendimiento de distintos algoritmos EMO para la calibración de ABMs con más de un KPI. En concreto, se compararon siete algoritmos de las principales familias de algoritmos EMO y un método de optimización matemática clásico. El rendimiento de estos algoritmos se evaluó usando los habituales indicadores unarios y binarios junto con el test de Wilcoxon para evaluar la significación estadística de los resultados. Además, se utilizaron *attainment surfaces* para apoyar visualmente los resultados del estudio. A lo largo de este análisis se pudo ver cómo MOEA/D obtuvo un rendimiento excepcional, mostrando un rendimiento significativamente mejor que el del resto de algoritmos en la mayoría de las instancias. Esto sugiere que una estrategia basada en descomposición es claramente la mejor para este problema de calibración. Sin embargo, también se pudo observar cómo el rendimiento de MOEA/D se redujo al calibrar determinadas instancias del banco de pruebas (consecuencia habitual del *No Free Lunch theorem* [WM97]), lo que condujo a analizar el impacto de distintas propiedades del problema en el comportamiento de los algoritmos EMO. Las propiedades consideradas fueron la forma de la

región factible del espacio de búsqueda y la forma del frente de Pareto del problema, junto con la dimensionalidad de las instancias. Por otro lado, este estudio mostró como SMS-EMOA es el algoritmo EMO más robusto para este problema, siendo capaz de mantener un buen rendimiento en las instancias que mostraban características problemáticas.

Por tanto, podemos concluir que los resultados del trabajo realizado cumplen con los objetivos propuestos para esta tesis doctoral. Adicionalmente, nos gustaría resaltar que esta tesis ha conseguido tener una satisfactoria producción científica, dado que sus resultados se han reflejado en cinco publicaciones científicas en revistas internacionales indexadas en el JCR.

En cuanto a los trabajos futuros, se plantea el uso de técnicas de modelado más avanzadas para poder mejorar los ABMs desarrollados durante la tesis. Por ejemplo, el uso de lógica difusa y *computing with words* [GCCCH20] se está utilizando con éxito en modelos de marketing y podríamos utilizar un enfoque similar para mejorar los ABMs para escenarios de ciencias políticas. Este enfoque mejoraría los procesos de difusión de información linguística relacionada con la distancia ideológica o con la resistencia de los votantes al efecto *framing* en la red social de votantes artificiales. Siguiendo esta línea, el uso de *fuzzy cognitive maps* [Pap13] podría explorarse como una opción válida para mejorar el comportamiento de los votantes artificiales en estos escenarios.

En este sentido, los ABMs para marketing desarrollados podrían extenderse para considerar otros KPIs como por ejemplo las ventas. Esto haría que los métodos de calibración propuestos necesitaran ser extendidos para considerar objetivos adicionales, lo que definiría un escenario diferente donde sería necesario el uso de algoritmos EMO para optimización *many-objective*. A su vez, ABMs más complejos y con más KPIs podrían ser más costosos de calibrar, necesitando mayores tiempos de ejecución. Esto podría requerir el uso de funciones subrogadas por parte de los algoritmos de calibración, dado que una única configuración de un modelo necesita ejecutar múltiples simulaciones de Monte-Carlo [LRS18].

Otra línea que consideramos interesante como trabajo futuro es evaluar la mejora que podría incluir el uso de *qualitative pattern features* para minimizar la pérdida de información producida por la función de *fitness* durante el proceso de calibración [YB11, YB15]. Dado que esta funciones se centran principalmente en la distancia entre series de puntos como resultado de la naturaleza del ABM, la agregación de estos valores puede hacer que se pierda la forma del histórico en el proceso. Esto se podría resolver de varias formas, aunque futuras investigaciones decidirán qué opción produce los mejores resultados. Por ejemplo, las funciones de *fitness* utilizadas podrían modificarse para tener en cuenta esta información. Otra opción podría ser incluir objetivos adicionales en problema de calibración que atendieran esta restricción. A su vez, debido a la alta cardinalidad de las aproximaciones de los conjuntos de Pareto obtenidos durante nuestros experimentos, pensamos que también podría ser interesante extender el enfoque de calibración propuesto y evaluar el impacto de incorporar las preferencias del usuario durante el proceso de calibración [TMKM09].

Abstract

Model simulation is an established approach for analyzing complex systems but these computational models need to be carefully calibrated and validated before they can be useful. This task can be challenging in many cases as the lack of information prevents the users for precisely estimating the parameters of the model and the number of parameters to estimate can be too high. The agent-based model methodology is a well-known model simulation approach that relies in the behaviour of artificial agents, which are autonomous entities that act following simple rules and interacting with other agents and their environment. Due to the mentioned issues, the calibration of agent-based models is usually carried out using automated calibration methods, since they can estimate those parameters which cannot be set because of the lack of information. However, the use of automatic calibration does not release the user from carefully reviewing and validating the resulting parameter set, as a good fitting between the models' output and the calibration data is not a guarantee of a valid configuration.

This doctoral dissertation addresses these issues in several ways. First, it proposes multiple agent-based models for defining political scenarios and marketing strategies that serve as the foundation of the dissertation, since they are calibrated and validated with the techniques and methods proposed by this thesis. In addition, these agent-based models are useful for obtaining insights relevant to their application domain, which is two-fold. On the one hand, two models tackle the Spanish national elections on the 14th of March of 2004, that were severely influenced by the terrorist attacks that happened three days before. On the other hand, we introduce a benchmark containing several instances of a model for defining marketing strategies that considers the awareness of the brands and their word-of-mouth volume.

This doctoral dissertation reviews and compares several relevant metaheuristics to design the best performing method for agent-based model automatic calibration. The use of a well performing optimization method is important for automatic calibration as its success depends on the method's ability for exploring the complex and ill-defined parameter search space. Thus, this dissertation conducts an exhaustive experimentation comparing well-established and recent evolutionary algorithms and including their hybridization with local search procedures. The computational study analyzes the calibration accuracy of the metaheuristics using an integer coding scheme over the developed benchmark of instances of the agent-based model for marketing, which considers an increasing number of decision variables. This study reported the outstanding performance of the memetic coral reefs optimization algorithm after performing multiple statistical tests to the results.

The choice of the best performing metaheuristic is specially relevant for models calibrated over more than a single output, as its calibration requires handling different criteria jointly. This fact increases the problem complexity and can be achieved by using automated calibration and evolutionary multiobjective optimization methods as they can find a set of representative Pareto solutions under these restrictions and in a single run. However, selecting the best algorithm for performing automated calibration can be overwhelming. Therefore, this dissertation proposes

an exhaustive analysis of the performance of several evolutionary multiobjective optimization algorithms when calibrating different instances from the agent-based model for marketing. This analysis evaluates the performance of the compared algorithms by using multiobjective performance indicators and attainment surfaces, including a statistical test for studying the significance of the indicator values, and benchmarking their performance with respect to a classical mathematical optimization method. The results of this experimentation reflect that those algorithms based on decomposition perform significantly better than the remaining methods in most instances. Moreover, we also identify how different properties of the problem instances (i.e., the shape of the feasible search space region, the shape of the Pareto front, and the increased dimensionality) erode the behavior of the algorithms to different degrees.

Finally, this dissertation introduces a multicriteria integral framework to assist the modeler in the calibration and validation processes of agent-based models that combines evolutionary multiobjective optimization with network-based visualization. Up to our knowledge is the first integral approach to model calibration in the specialized literature. This approach combines the outstanding performance of evolutionary multiobjective optimization algorithms with an advanced visualization method to better understand the decision space and the set of solutions from the obtained Pareto set approximation. This proposal is tested by using two instances of the agent-based model for marketing. The final analysis of the calibrated solutions shows how the proposed framework eases the analysis of Pareto sets with high cardinality and helps with the identification of flexible solutions (i.e., those having close values in the design space), thus helping the designer in the agent-based model validation.

Part I. Report

1 Statement

This section introduces the challenges addressed by this doctoral dissertation. It starts by presenting a brief description of the features and characteristics of the faced problems. Then, a justification for the work developed is provided. Finally, we enumerate the goals and objectives of this dissertation and show its structure.

1.1 Introduction

The agent-based model (ABM) methodology is a well-known approach to modeling complex phenomena which relies on autonomous entities referred as agents [Bon02, Eps06, JJ03, MN05]. The behavior of these artificial agents follows simple rules that define their actions and social rules that define their interactions with other agents. The aggregation of these simple rules allows ABMs to recreate and simulate emergent behaviors and complex systems, which has increased the relevance of ABM in recent years [FF09, Wal18]. Specifically, the features and characteristics of ABM systems have turned them into useful tools for analyzing social phenomena, since they can replicate processes as information diffusion and word-of-mouth (WOM) [CR17, LMP13, Rog10] by using artificial social networks [BA99, WS98]. Thus, ABMs are frequently used for building decision support systems that allow the users to define hypothetical and *what-if* scenarios [CR17, JO06]. In addition, the bottom-up approach of ABMs enables their analysis from either a micro (on the agent level) or a macro perspective (on the system level).

However, the design and definition of these models can be difficult as they typically involve a high number of parameters. In addition, assigning specific values based on the available data can be troublesome for many of these parameters. The process of setting the ABM's parameter values so the model can properly replicate the actual system dynamics is known as the model calibration [CBK$^+$17, CCD$^+$16, Oli03]. Although this process can be carried out manually, this procedure has several disadvantages. On the one hand, the model needs to be simulated multiple times to check if the values of each parameter change its behavior, which can be too expensive in terms of time. On the other hand, ABM parameters tend to show complex relations and thus, cannot be adjusted separately. This implies that finding a parameter value that obtains the best fitting of the model when adjusted in isolation does not necessarily mean that this value obtains the best fitting overall (i.e., when other parameters are also modified). Last but not least, ABMs usually consider a high number of parameters, making the manual calibration process extremely difficult.

Due to these issues, model calibration is usually carried out in an automatic fashion [CR17, Oli03, Sar05]. This computationally intensive approach tunes the parameters of the model by considering three main components: an optimization method, a loss function, and the historical

data that that the model is intended to replicate. During automatic calibration, the optimization method iteratively compares the output of the model with the historical data using the selected loss function and adjusting the model parameters accordingly. Thus, in order to perform the automatic calibration of ABMs, it is important to select an optimization method that can handle the issues identified regarding ABM calibration. Metaheuristics [Tal09] are an example of such non-linear optimization methods that can search across a large span of the model parameter space [CBK+17, Mil98, SR14]. In this regard, several well-known metaheuristics like genetic algorithms [BFM97] have been used for ABM automatic calibration [CH06, CC15, CR17, Fab13, HS15, MSEH14]. Nevertheless, the selection of the metaheuristic for carrying out the calibration process heavily determines the quality of the resulting model parameter configuration as the model's accuracy relies on the ability of the method for exploring the parameter search space.

In any case, a good fitting between the output of the model and the historical data is not enough for ensuring the validity and correctness of the model [Sar05]. Therefore, the modeler needs to carefully review and validate the resulting model configuration. This issue gets even more difficult when the model is designed considering two or more key performance indicators (KPIs) regarding the behavior of the model, since these type of models will not usually have a single set of parameters that can satisfy all the objectives. Therefore, the calibration of these models needs to be addressed as a multicriteria decision process similar to a multiobjective optimization problem, where the user needs to choose a configuration from a Pareto set approximation. In this situation, the validation of the calibrated configurations presents more difficulties as several configuration sets could be potentially chosen by the modeler and multiple valid configurations may satisfy at least one KPI.

Moreover, the effective use and simulation of ABMs for representing a complex system is strongly influenced by the transparency of the underlying ABM. ABM modelers and stakeholders are required to understand how the model recreates a given behavior because ABM simulation is often used for developing *what-if* scenarios and forecast hypothetical scenarios [JO06, VKM+16]. This remarks the importance of using multiple visualization methods during the calibration of ABMs as they can play an important role for increasing the understanding of the modeler on the calibrated model and its parameter settings [CBK+17]. Thus, visually showing the underlying relationships between an input configuration and its corresponding model output becomes a critical component of the validation process.

1.2 Justification

The goal behind the calibration of ABMs is to allow them to reproduce and analyze different complex and social phenomena. As we have previously stated, the success on the automatic calibration of ABMs is severely influenced by a good fitting to the available historical data. Therefore, the methods employed for running automatic calibration should be as effective as possible exploring the search space of the problem [CBK+17]. Since new optimization methods and metaheuristics are being developed every day, we believe that the current results of well-established metaheuristics can be improved by some of the more recent ones like *coral reefs optimization* or newer versions of *differential evolution*. Additionally, this is also an issue of ABMs considering more than one KPI, so more recent methods from the evolutionary multiobjective optimization (EMO) area should be tested in these kind of problems [CLVV+07]. The improvement obtained by more recent metaheuristics should be analyzed, testing their robustness when dealing with the specific characteristics of ABMs such as their high dimensionality.

Moreover, there is a need to improve the available validation methods of ABMs calibrated using automatic approaches, specially when they consider multiple KPIs. This could be achieved by introducing an integral framework that comprises two elements: an EMO algorithm for calibrating the parameters of the ABM automatically using historical data and an advanced visualization

method that improves the understanding of the calibration process and its results [FT20a, TF15a, TCP+18, WEF13]. This integral approach could show how visualization is an indispensable tool for automatic calibration, as it assists the modelers during the validation process.

We can identify two niches of applicability in the analysis of scenarios in political sciences and the support of the definition of marketing strategies. With respect to political sciences, the simulation of ABMs can be used for analyzing voting behavior using different voting theories [Lav05, Mui10]. We can highlight the specific scenario of the elections on 14 March 2004 (14-M), since it is an important scenario which was not previously analyzed using ABMs. In this regard, political science shows that terrorism in general, and more particularly terrorist attacks occurred close to elections, has an impact on the vote and campaign strategy of political parties [BK07, BK08, Fis05, Ran18]. On the 11th of March 2004 (11-M), three days before the Spanish national elections on the 14th of March (14-M), a group of terrorists exploded various bombs on trains circulating to Atocha train station in Madrid. 193 people died and about 2000 were wounded. The attacks changed the electoral process: in the morning of the 11-M, the campaign was suspended. On the 12th of March, there were demonstration marches against terrorism in the main Spanish cities. And on the 13th of March, there was a demonstration in front of the headquarters of the People's Party (PP), the Spanish right-wing party who was in the government. Finally, voting surveys failed and the results of the 14-M elections revealed an unexpected change of government.

After the attacks took place, a large quantity of information was generated by government, politicians, and mass media. That huge amount of information pushed the 11-M candidates to position themselves in relation to this event. Voters incorporated this political position about the attacks into their voting decision processes [Hol96]. The communicative framework of this event had two political competing dimensions regarding the two terrorist groups which could be responsible for the attacks: ETA and Al'Qaeda. The first position was defended by PP's government, while the second was supported by the left-wing Spanish Socialist Workers' Party (PSOE), the main opposition party, and other opponents [Olm05]. Two main reasons influenced public opinion on PP and PSOE's positions: the evaluation of the management of the government against ETA terrorism and the active participation of Spain in the invasion of Iraq, in March 2003. The majority of the Spanish population positively evaluated the action of the PP government against ETA. Therefore, President Aznar's bureau declared ETA responsible for the 11-M attacks as an election strategy [LM06]. On the contrary, the Spanish government's decision to participate in the invasion of Iraq was against the majority of the public opinion and political parties. Given the socio-economical and political importance of these facts in the recent Spanish history, the analysis of the framing effect generated after the 11-M attacks and how it influenced the decision of those who would vote for PP, PSOE, or abstain after the attacks is a relevant contribution. These two parties and abstention were the three electoral options with the highest support in the 14-M elections.

In addition, once could analyze how the management of the crisis affected to the ideological distance between the voters and the main parties. In this regard, research on the electoral behavior in the 14-M elections agrees on the fact that the terrorist attack influenced the public opinion [Bal07, LM06, Mic05, Mon11, ML09, Olm05, RMA07, TR04]. They conclude that the attack contributed to some voters reorienting their voting intention from different scientific paradigms. However, none of these studies analyzes electoral behavior from the perspective of the spatial theory of voting, which assumes that voters and political parties are located on a bipolar continuum that reflects their positions on a political issue [Dow57, EH84, EH94, GG20, Gro85, Ked05, Ked09, RM89]. According to this theory, voters will choose the political party that is closer in this one-dimensional space. In response, the political parties are driven by a utilitarian logic that encourages them to position themselves in a position that minimizes the distance with each and every one of the voters. Therefore, the spatial theory of voting justifies the relevant role of the mass media in the

electoral process [CS08, DM11, Str04], since they are the channel used by the political parties for transmitting their messages and attracting the attention of voters, who demand information [Str15].

Regarding the design of marketing strategies, ABMs are specially relevant for assessing *what-if* scenarios where the simulation's goal is to evaluate the response of the market to different marketing alternatives [CR17]. These alternatives usually consider the role of certain KPIs that are key to market adoption, retention, and even new product development. In this regard we can highlight the importance of WOM and brand awareness, which have been identified as very relevant during market expansion [LMP13, MS00]. On the one hand, WOM is a powerful KPI for marketeers, as it can lead to new purchases without requiring additional investment by the brand. In addition, WOM has been proven useful for accelerating the adoption of customers that would eventually purchase the product anyway [LMP13].

On the other hand, the recall level of brand awareness is very relevant for building brand equity [HSV12]. This equity represents the value of the brand for the consumer, as it simplifies the decision-making of the purchase act by transferring responsibility to a specific manufacturer or distributor. As a consequence, unrecognized brands are likely to be discarded in certain scenarios. In contrast, if a brand is the first one to be recalled by the consumers (i.e., the top of mind), that status can be exploited by the brand, since it is considered a relevant indicator of costumer preferences on favor or against a global brand [Shi10].

1.3 Objectives

The main objective of this PhD dissertation is the development of new methods that improve the automatic calibration and the validation of ABMs with any number of KPIs. This objective implies the development of new methods for calibration and validation of ABMs. To do so, we aim to design real-world ABMs in political and marketing sciences which will be considered as comprehensible benchmarks for the automatic calibration methods. Therefore, we split up the general objective into four sub-objectives in order to reach the proposed overall goal:

- **Building ABMs for analyzing political scenarios and marketing strategies**. The designed engine should be compatible with the definition of models with any number of KPIs and must be validated in real-world scenarios. In addition, the developed ABMs will be used to define a calibration benchmark for ABMs that could be used for assessing and validating the rest of the methods proposed in this doctoral dissertation. Therefore, these model instances should consider the issues typical of ABM calibration, such as an increased dimensionality and correlated parameters.

- **Designing automatic single-objective calibration methods for ABMs based on advanced metaheuristics**. In this stage of the doctoral dissertation, the developed calibration methods will target single-objective ABMs. In this regard, both the ABMs considering a single KPI or those where their outputs can be aggregated as a single value can be considered. The methods developed for this sub-objective will be tested and validated using the ABM instances previously gathered by the first stage of the thesis.

- **Designing automatic multi-objective calibration methods for ABMs based on advanced metaheuristics**. EMO algorithms are considered the best approach to multiobjective optimization [Deb01] as they can obtain Pareto-optimal solutions in a reasonable time and can perform successfully without requiring specific properties of the optimized function. As a consequence, this line of research will be focused on providing the best design and identifying the best performing EMO for the tackled problem. The analyzed EMO algorithms will be tested and validated using the benchmarks with multiple KPIs generated when tackling with the first sub-objective.

- **Proposing an integral framework for multiobjective calibration and validation of ABMs**. The designed framework will combine the calibration algorithms designed for the former sub-objective (i.e., the EMO algorithms targeting ABM calibration with multiple KPIs) with an advanced visualization method that enhances the understanding of the calibration process and its results. This tool would ease the calibration and validation process for the modeler.

1.4 Structure

This doctoral dissertation splits into two clearly differentiated parts. The first one tackles the statement of the problem of calibrating ABMs, providing a preliminary description of both the ABM methodology and the metaheuristic algorithms used for running automatic calibration. It also considers the presentation, analysis, and discussions of the work done during the thesis development. Then, the second one collects the main scientific publications containing the main achievements resulting from this doctoral dissertation.

In detail, Part I is organized as follows. First, the current statement introduced the problem addressed by the current dissertation, provided its justification, and reviewed the proposed objectives in detail. In the next section, Section 2, we will provide a brief introduction to the ABM methodology and review the related work on ABMs for political and marketing scenarios. Section 3 describes the main metaheuristics considered during the development of this thesis. Later on, Section 4 addresses the problem of model calibration and validation and reviews the related work on automatic calibration. Section 5 summarizes and discusses the main results of this doctoral dissertation. Finally, Section 6 states some final remarks and draws some future research lines.

Part II details the work developed to accomplish the objectives stated before by assembling the following five scientific publications:

- I. Moya, M. Chica, J. L. Sáez-Lozano, O. Cordón. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections, Knowledge-Based Systems, vol. 123, pp. 200-216, 2017. DOI: 10.1016/j.knosys.2017.02.015. Impact Factor (JCR 2017): 4.396. Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Order: 14/132. Q1.

- I. Moya, M. Chica, O. Cordón. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization, Decision Support Systems, vol. 124, 113111, 2019. DOI: 10.1016/j.dss.2019.113111. Impact Factor (JCR 2019): 4.721. Category: OPERATIONAL RESEARCH & MANAGEMENT SCIENCE. Order: 8/83. D1. Q1.

- I. Moya, E. Bermejo, M. Chica, O. Cordón. Coral reefs optimization algorithms for agent-based model calibration, Engineering Applications of Artificial Intelligence, vol. 100, 104170, 2021. DOI: 10.1016/j.engappai.2021.104170. Impact Factor (JCR 2020): 6.212. Category: ENGINEERING, MULTIDISCIPLINARY. Order: 7/91. D1. Q1.

- I. Moya, M. Chica, O. Cordón. Evolutionary multiobjective optimization for automatic agent-based model calibration: A comparative study, IEEE Access, vol. 9, pp. 55284-55299, 2021. DOI: 10.1109/ACCESS.2021.3070071. Impact Factor (JCR 2020): 3.367. Category: ENGINEERING, ELECTRICAL & ELECTRONIC - SCIE. Order: 94/273. Q2.

- I. Moya, M. Chica, J. L. Sáez-Lozano, O. Cordón. Simulating the influence of terror management strategies on the voter ideological distance using agent-based modeling, Telematics and Informatics, vol. 63, pp. 101656, 2021. DOI: 10.1016/j.tele.2021.101656.

Impact factor (JCR 2020): 6.182. Category: INFORMATION SCIENCE & LIBRARY SCIENCE - SSCI. Order: 11/86. Q1.

2 Agent-based modeling for political and marketing scenarios

This section provides a general introduction to the ABM methodology, focusing on its main components. After this brief introduction, we review several applications of ABM to marketing and political scenarios in the literature. Finally, we describe our designs for two ABMs for political scenarios and an ABM for marketing strategies.

The ABM methodology [Bon02, Eps06, MN05, WR15] relies on a population of autonomous entities called agents that behave according to simple rules. The behavior of these agents can be defined in an heterogeneous manner, as different agents may follow different rules. Moreover, an agent could change its behavior multiple times during a single simulation, as its rationality is bounded by the amount of information that is accessible to it. As a consequence, we can observe that ABM supports a level of heterogeneity that can be hard to model using a top-down approach [WR15].

ABMs allow the interaction of individual agents with other agents of the population and their environment. These interactions can include a spatial structure so the agents only interact with the agents in its vicinity. Therefore, ABMs typically consider an artificial social network [BA99, WS98] that enables the interactions within the neighboring agents. This allows ABMs to realistically replicate the dynamics of epidemic and information diffusion processes [NBW06, PSV01] or word-of-mouth (WOM) mechanisms [CR17]. Additionally, the agents of the population can be exposed to external influences representing interactions with their environment or their exposure to external actions. The aggregation of the behavior of the different agents and their interactions permit the representation of complex and emerging dynamics.

ABM simulation also allows modelers and practitioners to define *what-if* and hypothetical scenarios [JO06, PBP17, PYS17, TR18]. These scenarios are driven by *explainability* purposes, as they allow the modelers and practitioners to obtain insights from the ABM that could be difficult to obtain using other methods. This has increased the visibility of ABM simulation in the last few years [FF09, Wal18]. *What-if* scenarios are designed by taking a calibrated and validated ABM and changing some of its properties to understand changes in behavior and explain future insights about the described phenomena. In practice, this usually involves changing the behavior of some of the agents in the population or modifying the external influences of the agents. Finally, it is important to remark that ABMs must be designed following certain guidelines for ensuring their rigor [RR11].

2.1 Related work on ABMs for political and marketing scenarios

2.1.1 ABMs for political scenarios

ABM techniques have been extensively applied in the field of political sciences for dealing with political scenarios and party competition [FS05, KMP92, KMP98, Lav05]. These approaches consider both parties and electors as moving entities that make decisions continuously. That is, electors react to politicians behavior and politicians reconsider their strategy regarding electors decisions. In [Mui10], authors extended this approach by including mass media influence. This new role for mass media is focused on campaign organization, where political parties use media for enhancing their image regarding their voters. The latter studies show how mass media can be useful to add realism to the model and better explain the political scenario.

In [GACK+07], mass media influence is studied by distinguishing two possible behaviors: global and local. The essential differences between these behaviors are focused on how they behave regarding time and space. Mass media effect and the role of mass media during campaigns regarding voters mobilizations is also analyzed in [FLPEC05]. This approach is interesting because it analyzes the abstention factor, instead of focusing on individual voting preferences. Authors of [MS06]

considered the polarization effect of mass media in opinion dynamics to examine how mass media affects individuals. In this model, mass media messages are propagated via social interactions, showing dynamic changes over different scenarios where strongly polarized messages influence the agent population.

In [Liu07], the authors examined how mass media influence the opinion formation through opinion leaders (i.e., influentials) [LBG65]. Using an ABM simulation, the authors highlight the importance of the communication networks used by opinion leaders to influence the public. Another approach to the influence cascade and its relationship with social networks can be found in [WD07]. Some contributions like [SKM13] incorporate topology restrictions for structuring social influence into voter communities, creating substructures where agents with similar attributes are grouped. These substructures are applied for modeling agents' social communications and influencing their political decision making.

ABM has also been used for studying the propagation of political perceptions. In [Zak14], the propagation of the agent's knowledge is modeled using a space based approach. In this case, the propagated agents' knowledge depends on the satisfaction of agents with the current political situation. This approach simplifies the topology problem by assigning different radius to each agent. Thus, some agents will share their perceptions with more agents than others. A similar approach is followed in [SBM11], where an ABM simulates the effect of social influence regarding voting preferences.

There are several contributions exploring the joint use of ABM and *spatial theory of voting*, a branch of political sciences considering that voters choose the party that is politically closest. [Clo08] explores the impact of uncertainty into an ABM using a Dowsonian approach [Dow57]. It concludes that typical models with complete information do not converge to the median distance but the simulations using certain level of uncertainty actually do. [PM08] introduces an ABM using a Dowsonian spatial model with multiple parties and multiple dimensions. They find that the number of competing parties and the likelihood to abstention increases the average distance between the parties and the center of the ideological scale.

[Lav05] explores spatial models of political competition where parties follow adaptive rules. Since voters are modeled to vote for the closest party, parties behave as stickers (who never change their position), aggregators (which try to converge on the average preferences of the voters), hunters (who move greedily for maximizing the number of supporters), or predators (that moves towards the largest party). This work is later extended by [LS07], where the authors study the survival of political parties by considering a dynamic environment where new parties can appear and the existing ones can be extinguished. The results of this model suggest that vote-seeking parties tend to make voters miserable, since their priority is to get new supporters instead of focusing on their current supporters [Lav11].

Other ABMs analyzing political competition consider the *spatial theory of voting*. [AM08] uses different adaptive agents for studying the Concordet criterion, which suggests that if a party candidate should win her/his rivals in pair-wise competitions, she/he should win the elections and result elected. By using the party behavior introduced by [Lav05], the authors conclude that the behavior and decision rules followed by the agents have a great impact in the Concordet efficiency. [WS15] extended the work of Laver by including oligarch agents. These oligarchs represent private industry leaders or lobbyists that influence the elections by using campaign donations. Using this new agent's role, they study the effect of the oligarchs on the legislative action, showing that oligarchs reduce their donations when parties are less likely to be influenced or when voters are opposed to the oligarchs' interests.

Finally, [LS18] introduces new decision rules and a coalition formation procedure into Laver's model. Their work concludes that aggregator parties outperform the other roles by being able to

satisfy both voter preferences and government policies. In addition, they find that a bigger set of parties improves the party system and government representation.

2.1.2 ABMs for marketing strategies

ABMs have been broadly applied to different problems in marketing. For instance, [JJ03] introduced the well-known Consumat model, an ABM for analyzing market dynamics where a set of artificial consumers choose between a set of products for purchase. This decision is made by weighting the personal preferences of the agent with the social utility of the product, which increases as more neighbors of the agent choose that product. The model by Jannsen and Jager have been employed for different applications such as the combination of viral marketing and influence maximization [RCC20]. In addition, ABMs have also been employed for validating online viral marketing strategies [SI16].

In [LMP13], an ABM is employed for evaluating the impact of WOM campaigns in marketing adoption policies. Their ABM focuses on the acceleration of adoption, the expansion of the market, and the increment of consumer equity under the competition of multiple brands. Marketing adoption policies are also analyzed using ABMs by [TRJ13]. In this case, the authors study the improvement on future predictions regarding diffusion dynamics by incorporating information regarding the structure of the underlying social network, since future products would involve subsets of the same network. Additionally, adoption and retention strategies are further analyzed by [SACM21] for the particular case of start-up companies. Among other insights, they point out the relevance of improving acquisition and retention in a proportional way for achieving an optimal growth.

[Gar05] analyzes new product growth and development by addressing the problem of resource allocation for an innovation firm, which requires to split resources between the exploration of new technologies and the further exploitation of the existing ones. This is done by proposing an ABM that splits its population between manufacturers and consumers. Then, some consumer agents are set to be early-adopters and the manufacturers are set to weight their investment between innovation and exploitation. [DJBJ10] introduces an ABM for innovation diffusion that analyzes the growth of new and future products. Their model is designed by combining the individual preferences of the consumers and the social influence of their neighbors. This combination is the result of weighting the individual utility for adoption and the social influence representing the number of neighbors that already adopted the product.

[HL13] introduces an ABM for targeting revenue leaders during the introduction of new products, as they should accelerate the adoption process. This revenue leaders would be consumers that are able to generate high profitability on their own with a social value higher than the average. In addition, [HLQS18] studies multiple targeting strategies that either focus on influentials, susceptibles, or unsusceptibles. Their experiments reveal that most approaches perform acceptably under different budget restrictions. On the one hand, targeting as many susceptibles as possible would be optimal on a limited budget. On the other hand, the optimal approach on large budgets would be to target unsusceptibles with free products. Finally, interested readers may find ABMs developed for modeling consumers decision-making by including fuzzy linguistic perceptions in [GCCCH20].

2.2 Building real-world ABMs to test automatic calibration

This section introduces the ABMs that will be subjects of calibration by the methods proposed in the following sections. After briefly introducing the components shared by these models, we review each of the ABMs in detail in the remaining of the section. Sections 2.2.2 and 2.2.3 consider the two models for political sciences that analyze the 14-M elections from two different approaches:

the framing theory and the spatial theory of voting. Finally, Sections 2.2.4 describes the ABM reproducing a banking scenario for marketing strategies in Spain.

2.2.1 Shared components

Although each of the proposed ABMs consider their own specific dynamics, the developed models consider a similar approach to the modeling of social dynamics and the definition of external influences. Thus, all the models consider that their agent population is connected by an artificial social network [BA99, WS98] where the agents participate in a WOM procedure and they are exposed to the influence of global touchpoints (e.g., television, radio, or online media, among others) [GACK+07].

The considered social network is modeled using an artificial scale-free network [BA99] because the most real networks match with this network model [BA99, NBW06, PSV01]. In these kinds of networks, the degree distribution follows a power law. This means that few nodes have a significantly large number of connections (hubs of the social network) and most nodes have a very low number of connections. Our scale-free network is generated using the Barabasi-Albert preferential attachment algorithm [BA99]. The preferential attachment algorithm has a main parameter m which regulates the network growth rate and its final density. The generation process starts with a small clique (a completely connected network) with m_0 nodes. At each generation step, a new node is added and connected to m different existing nodes. When a new node is included, the probability of choosing an existing node is proportional to its degree (preferential attachment). After t iterations, the Barabasi-Albert algorithm results in a social network with $m \cdot t$ edges. Finally, the average degree of the social network is $\langle k \rangle = 2 \cdot m$.

The external influences like brand advertising or mass media messages are modeled as global mass media [GACK+07], since they have the same probability of influencing any agent regardless of its connections in the social network. The external influences are parameterized to define the differences between the touchpoints. Each model considers a set of C mass media channels. Each mass media channel can have different values for its parameters even if they belong to the same media type. They can influence any number of agents at random depending on the channel potential for reaching the population. In addition, the total amount of reached agents depends on other factors such as the investment of each brand in the case of the marketing model and the audience of the media in the case of the politics model. However, the maximum population percentage that can be reached by a mass media channel is bounded by the nature of the channel itself, since some media are able to reach more people than others. For example, the maximum population percentage that can be reached by a campaign scheduled in the radio is bounded by the maximum population percentage that listens to the radio. We model this behavior with a reach parameter ($r_c \in [0, 1], \forall c \in C$), which defines the maximum percentage of agents that a channel c is able to hit during a single step.

In addition to other effects that are specific to each model, the effect of the external influences (i.e., either brand advertising or the media information during the 14-M elections) can produce a viral buzz effect in the reached agent. This buzz effect increases the number of conversations about the announced brand, modifying the talking probability (p_i^b) of the reached agents (this talking probability is further depicted with the specifics of each model in their respective sections). We model this effect through a variable called buzz increment (τ_c) for each channel $c \in C$. This increment of the agents talking probability is computed as a percentage increment over the initial talking probability ($p_i^b(0)$) of the agent. However, if the generated buzz is not reinforced, its effect could decay over time as previous interactions are forgotten. We model this effect with a variable called buzz decay ($d\tau_c$). The action of buzz decay reduces the previous increment of talking probability (σ_c) applied to the agent through channel c. The update process for the talking

probability value of agent i for brand b due to both buzz increment and decay effects of channel c is shown in Equation I.1.

$$\begin{aligned} p_i^b(t+1) &= p_i^b(t) - \sigma_{i\,c}^b(t) \cdot d\tau_c + p_i^b(0) \cdot \tau_c, \\ \text{where } \sigma_{i\,c}^b(t) &= \sum_{i=1}^{t} (p_i^b(t) - p_i^b(0) \cdot \tau_c). \end{aligned} \tag{I.1}$$

2.2.2 ABMs for political scenarios: the Framing theory

This ABM analyzes the 14-M Spanish elections in 2004 using the Framing theory paradigm [CD07]. Our proposed model simulates the 72 hours from the attacks to the election time: from March 11th at 8:00 AM to March 14th at 8:00 AM. The time-step of the ABM simulation is an hour, as it correctly fits with the mass media schedule. After the 72 steps of the simulation, every artificial voter or agent (which represents a group of real voters) votes and the simulation outputs the elections' results. During the simulation period, agents receive information from mass media and spread their political perceptions through their social network in a WOM process.

The initial perceptions of the agents of the simulation come from pre-electoral data of the Spanish government [Cen04], since this source resulted accurate predicting the results of the previous national elections. Therefore, the simulation starts with no framing effect over the voters. In order to model the voters (agents' population), we have divided agents into three segments: PSOE voters, PP voters, and abstainers (these three voting options cover more than the 84% of the election results). This segmentation is done to better fit the pre-election survey data [Cen04]. Using this segmentation, agents' parameters are defined at the segment level, so agents from different segments behave differently. This design decision makes the ABM simulation more realistic and heterogeneous as well as facilitates the definition of the model's parameters.

Figure 1 shows an illustrative diagram with the structure of the model and the flow of a mass media message. We use the size of the pre-election survey [Cen04], i.e., 24,109 respondents, as the ABM population size. This way we ensure enough granularity for the population to have a number of agents that represents the political conditions and available data from polls and National studies. Our target real population is the sum of PSOE and PP voters, plus abstainers, which represents 29,238,662 people. Thus, the ABM maps one virtual agent to a real voter in a 1:1,212.77 ratio.

The agent population will be influenced by the two framing effects generated after the 11-M attacks: ETA is responsible for the attacks and Al'Qaeda is behind the attacks. Each agent A manages the framing effect by the use of a state variable, called *resilience* and encoded by μ_A, a real-valued variable defined in interval $[0, 10)$. The resilience of the agent is the current vote positioning of the artificial voter with respect to the three voting possibilities. This value will change during the simulation depending on its resilience with respect to the received amount of external influence by mass media and other artificial voters. By using this framing effect variable μ we can obtain the voting alternative of each agent by applying Equation I.2. $v_A(\mu_A(t))$ represents the voting option of agent A at time-step t with $t = \{0, ..., 72\}$.

$$v_A(\mu_A(t)) = \begin{cases} 0, & \text{if } \mu_A(t) \in [0, 3.3)\,, \\ 1, & \text{if } \mu_A(t) \in [3.3, 6.7)\,, \\ 2, & \text{if } \mu_A(t) \in [6.7, 10)\,. \end{cases} \tag{I.2}$$

This function returns 0 if the the agent votes for PSOE, 1 if the agent abstains, and 2 if the agent votes for PP. Voters can change their vote v from $t-1$ to t, since they can be influenced by different sources (i.e., other agents and multiple mass media channels). The final voting option for agent A will be the result of $v_A(\mu_A(72))$. At the beginning of the simulation, the μ variable is

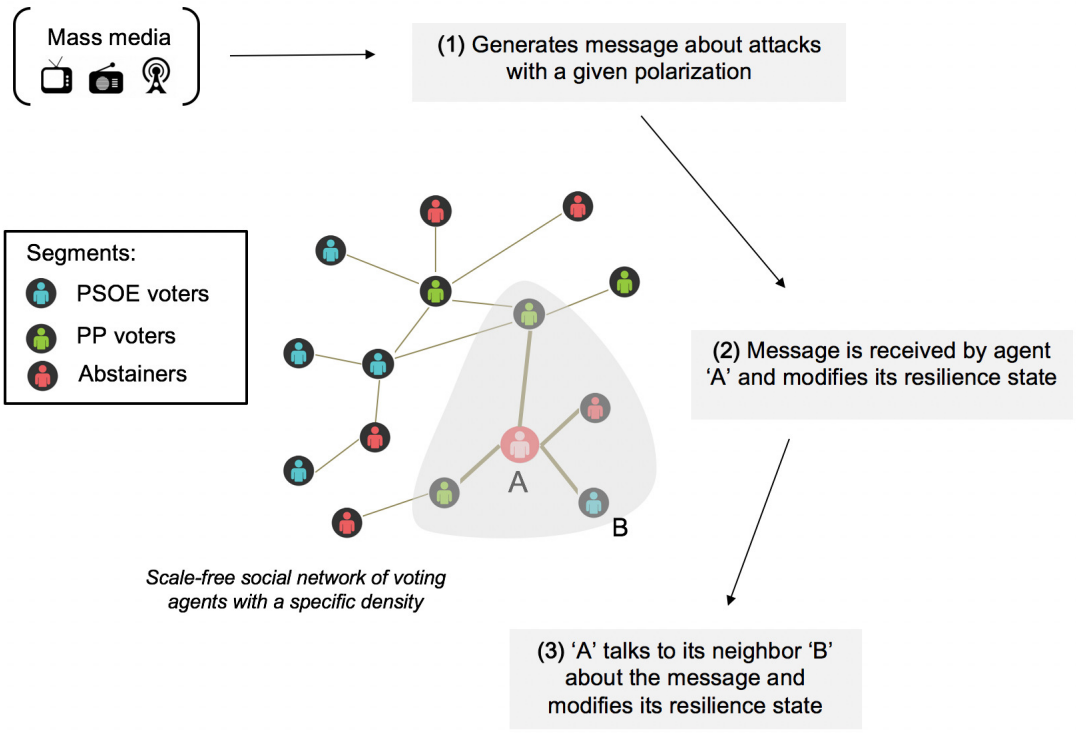


Figure 1: General structure of the model and example of mass media message sent to the voting agents, which react to it and spread their perceptions to their neighbors in the social network.

randomly initialized for each agent in each of the three segments using a uniform distribution in the segment-specific interval. These intervals are: [0, 3.3) for agents at the PSOE voters segment, [3.3, 6.7) for abstainers, and [6.7, 10) for agents at the PP voters segment.

If agent A resilience value (μ_A) moved to other option's interval, it will vote for that option, modifying its behavior as shown in Table I.1. For example, if an agent gets its resilience to a value between 3.3 and 6.7, it will abstain, even if it belongs to PP or PSOE voters segments. Thus, changes affecting resilience (μ) can generate four effects on the vote: reinforcement, conversion, activation [Ber54, LBG65], and deactivation [MiC08] (see Table I.1). Reinforced voters are those who voted the same electoral option at both steps $t = 0$ and $t = 72$. Converted voters are those who reoriented their vote, choosing another option at $t = 72$. Activated voters are those who did not aim to vote initially, but chose to vote at $t = 72$. Deactivated voters are those who aimed to vote at $t = 0$, but finally did not vote at $t = 72$.

The agents are able to spread their perceptions during the simulation using the artificial social network. We model this WOM interaction as a contagion process which allows the spreading of the information through the nodes of the social network depending on the influence and connections of these nodes (i.e., agents or artificial voters) [NBW06, ZZWZ13]. Every agent A has a talking probability ($p_A(t) \in [0, 1]$) to spread its perceptions about the current framing (i.e., its μ value) at each step of the simulation. When the probability check passes, the agent will talk with all of its neighbors of the social network. We will model this interaction using a variable called influence change (Δ), which modifies the strength of the agent's influences to its neighbors. This interaction is modeled in a directed way, meaning that the talking agent influences its neighbors and not in the opposite way. This change on the resilience value is controlled by Equation I.3, where $\mu_B(t)$ refers to resilience value of the listening agent B when talking with agent A. Δ_A refers to the influence change value of agent A.

Before attacks $V(0)$	**Election day** $V(72)$	**Framing effect on a voter**
PP	PP	Reinforcing
PSOE	PSOE	Reinforcing
Abstention	Abstention	Reinforcing
PP	PSOE	Converting
PSOE	PP	Converting
Abstention	PP	Activating
Abstention	PSOE	Activating
PP	Abstention	Deactivating
PSOE	Abstention	Deactivating

Table I.1: Framing effects in 72 hours for the elections March 14th, 2004.

$$\mu_B(t+1) = \mu_B(t) + |\mu_B(t) - \mu_A(t)|\,\Delta_A. \tag{I.3}$$

We also include in our model a variable called influence decay ($d\Delta$) which modulates how previous influence is forgotten over time. This decay effect is applied at the beginning of each step for every agent and reduces accumulated influence. Previous accumulated influence ($\delta_A(t)$) is computed following Equation I.4 and represents the sum of previous changes to μ performed by WOM from the initial step 0 to current step t.

$$\delta_A(t) = \sum_{i=1}^{i=t} \left(\mu_{A_i} - \mu_{A_{i-1}}\right). \tag{I.4}$$

The resilience value change for agent A due to decay is defined in Equation I.5, where $d\Delta_A$ represents the decay rate which modifies the accumulated influence. Let us finally remind that, to make the ABM more heterogeneous, each segment of the model can have different values for the talking probability ($p(t)$), influence (Δ), and influence decay ($d\Delta$).

$$\mu_A(t+1) = \mu_A(t) - (\delta_A(t) d\Delta_A). \tag{I.5}$$

Registered media audience from 11-M to 14-M represents the external influences for the agents [Aso04, LG04]. The selection of the messages forming the communicative framework that originated after the attack has been performed considering three criteria. First, the communicative diversification, because we analyze the messages broadcast by television, radio and newspapers, instead of focusing on a single type of mass media. Second, we select mass media channels that are broadcast at a national scale. Finally, we select messages that respect the plurality of information. This selection was developed including any message that contained information regarding the attacks, from regular news sessions to special bulletins and statements from political figures. Following those criteria, we design a complete communicative framework that covers the main information broadcast between the 11th and the 14th of March, 2004. Specifically, we include the main mass media channels broadcasting in Spain during this period: El Pais (press), El Mundo (press), ABC (press), Cadena Ser (radio), TVE (television), Antena 3 (television), and Telecinco (television). We should note that the Internet did not have enough influence in 2004.

These channels will also spread different messages at any step t, depending on which terrorist group is considered as responsible by its broadcast information (ETA or Al Qaeda). Each transmitted message has a polarization value, modeled as $m_C(t) \in [-2, 2]$. $m_C(t)$ represents the information bias of the message (ETA *versus* Al Qaeda) at time-step t broadcast by specific channel

C. In order to set the values of these polarized messages, a team of three media experts reviewed and studied press, radio, and television information during the three days period. Because of the subjectivity of the process, they individually scored them and later, agreed about the polarization value of all the media messages. $m_C(t)$ is set to -2 if the information of the message strongly points to Al Qaeda and to 2 if it strongly points to ETA. Both -1 and 1 refer to a weak authority and 0 refers to not assigning attacks' authority to any specific group. As our simulation steps by hour, we use the average polarization for a given time slot when two or more messages appear within it for a specific media. In a similar way, if a single message was assigned to different values during the experts' classification process, its final value was agreed and calculated using the averaged value.

Additionally to the parameters depicted in Section 2.2.1, the mass media channel considers a set of parameters for regulating their influence in the agents. When a mass media channel impacts an agent A, its message influences the resilience value of the agent (μ_A). We define the influence change parameter (Δ') to modulate this effect. This behavior is similar to the one defined for the social interaction between agents. This way, resilience change is performed using the received message and the influence change value for the media channel. As the same message could be received multiple times by the same agent, its maximum influence is limited to the overall influence value (Δ^{max}). Additionally, we represent the previous influence accumulated by the channel (δ') analogously to WOM. The resilience value change of agent A after the influence of channel C is formulated by Equation I.6, where $m_C(t)$ refers to the transmitted message and Δ^{max} refers to the maximum influence value. In addition, agents may forget what they just watched or read as the novelty of the message expires [WH07, YL10]. We include a parameter for measuring how media influences can be forgotten by the agents. This effect is modeled as influence decay ($d\Delta'$) which reduces previous influence, similarly to the one defined for social interaction. Equation I.7 defines the decay update for agent A due to the influence of channel C.

$$\mu_A(t+1) = \mu_A(t) + (\Delta_C^{max} - \delta'_C)\Delta'_C m_C(t). \tag{I.6}$$

$$\mu_A(t+1) = \mu_A(t) - (\delta'_C d\Delta'_C). \tag{I.7}$$

2.2.3 ABMs for political scenarios: the spatial theory of voting

The designed ABM is used for testing the following hypothesis: the 11-M attacks and its management by the Spanish government and the opposition influenced the ideological distance between the voters and the political parties. In this influence the mass media played a critical role as they connect the information produced by political parties with the voters. Our approach is to analyze the changes on the ideological distance itself, which makes it the main variable of the model, not to use the ideological distance as a method for explaining the behavior of the voters neither the outcome of the elections. The following model computes the ideological distance following the proximity model by [Dow57], since the available data does not allow us to reproduce other models based on intensity.

Similarly to the previous model, the current ABM considers a terminating simulation of 72 steps that represents the 72 hours between the attacks and the elections (from March 11 at 8:00 AM to March 14 at 8:00 AM) using a time-step of an hour. The model simulates the behavior of N agents representing artificial voters and their reaction to the information received from C mass media channels along with the diffusion of this information through the social network due to a WOM process. The information supplied by the mass media channels contains a polarized message from different political leaders from the P main parties (i.e., PP, PSOE, and IU), which concentrated 85.26% of the total votes. Due to the strong positioning of the parties regarding

the authority of the attacks, these messages influence how the agents position the parties in the ideological space.

This message polarization modifies the perception of the agents with respect to the ideological distances between them and the main parties. We model this effect by using the state variable *distance*, encoded as Λ_i^k, a real-valued variable, with $i \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, P\}$. The value of the latter variable changes during the simulation depending on the amount of external influences supplied by the mass media and the other neighboring agents. The distance variable is initialized by using the locations of the ideological space of voters and political parties [Cen04]. These data are introduced as v_i and p_i^k using integer values defined in interval $[1, 10]$, where v_i is the ideological location of the voter i and p_i^k is the ideological position that i assigns to the party $k \in \{1, \ldots, P\}$, as P is the number of parties participating in the election. Therefore, we can compute the initial ideological distance as $\Lambda_i^k = \left|v_i - p_i^k\right|$, and thus is defined in interval $[0, 9]$. In addition, the state variable $\Phi_i = \{0, 1, \ldots, P\}$ represents the political dispositions of each agent i, which identifies it as voter of a particular party (i.e., the agent voted for IU, PP, PSOE, or the remaining parties in the previous elections) or as a non voter (i.e., abstainers). Thus, Φ_i takes the value of the party when the agent is a voter and 0 if it is a non-voter or abstainer.

Agents can share their perceptions regarding their ideological distances with the main parties. These interactions between the agents can be modeled again as a contagion process [LK14, NBW06] as they spread their distance values through the social network. At every step of the simulation, each agent i considers a talking probability ($p_i(t) \in [0, 1]$) of spreading its distance values for the different parties (collected in the Λ_i^k values). Therefore, the agent spreads its values with all of its neighbors each time the probability check is passed. The variable *influence* (Δ) models the influence of an agent with its neighbors. When an agent shares its perceptions, it does it in a directed-only way (i.e., from the active agent to its neighbors). The update of the distance value due to social interactions is defined by Equation I.8, where $\Lambda_j^k(t)$ refers to the distance value of the neighbor agent j of party k when the active agent i shares its perceptions. This equation regulates the final influence using two additional values, since it would not be realistic that agents with very distant ideological positions were to influence each other. First, $\Theta(x)$ represents the Heaviside step activation function with $x = |\Lambda_j^k(t) - \Lambda_i^k(t)| - \psi$. $\Theta(x)$ returns 1 when $x \geq 0$ and 0 otherwise. Therefore, it disables the WOM influence when the distance difference is greater than a given threshold ψ. Second, voters with different political dispositions (i.e., voters of different political parties or non-voters) are less likely to influence each other. This is resembled by the parameter $\phi \in (0, 1]$, which regulates the influence when $\Phi_i \neq \Phi_j$, otherwise it is equal to 1.

$$\Lambda_j^k(t+1) = \Lambda_j^k(t) + \left|\Lambda_j^k(t) - \Lambda_i^k(t)\right| \Theta(x) \phi_{\Phi_i \neq \Phi_j} \Delta. \tag{I.8}$$

The proposed model considers an additional parameter referred as influence decay ($d\Delta$) that regulates how social influence erodes over time if it is not reinforced with further stimulus. Therefore, every agent reduces its accumulated social influence at the beginning of each simulation step due to this decay effect. The accumulation of social influence ($\delta_i^k(t)$) due to WOM interactions follows Equation I.9, which represents the accumulated changes to Λ from the start of the simulation to the current step t. Finally, the distance value change experienced by agent i with respect to party k due to the decay effect is defined in Equation I.10.

$$\delta_i^k(t) = \sum_{s=1}^{s=t} \left(\Lambda_i^k(s) - \Lambda_i^k(s-1)\right). \tag{I.9}$$

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) - \left(\delta_i^k(t) d\Delta\right). \tag{I.10}$$

The registered media audience between 11-M and 14-M are the external influences to the agents during the simulation [Aso04, LG04], which can influence any agent during any simulation

step. Again, the selected media are written press, radio, and television channels, since the Internet did not have enough influence by that time. In addition, we model cell phone messages and similar communications using WOM, as it is well known they had a strong activity and voting influence during the studied period [Olm05]. The information supplied by the mass media channels considers any message containing information about the attacks, regardless if it appeared in regular news sessions, were included as special bulletins, or were taken from statements of political figures. This information was selected following three main criteria: diversification (we consider multiple types of mass media channels), scale (we include mass media channels operating nationally), and plurality (messages were included avoiding discrimination of sources). The selected channels are those considered by the previous model (see the previous section).

In addition, and because there are two competing frames aligned with the considered parties, the transmitted polarization modifies the distance of parties differently. If the resulting polarization of a message is biased towards ETA (i.e., polarization > 0) then the agent's distance with the PP party is reduced and the distance with the other parties is increased. In contrast, if the message is biased towards Al Qaeda, the perceived distance with the PP party is increased and the distance with the other parties is reduced. Finally, the polarization values were scored by different experts due to the subjectivity of this task and agreeing the final values using the average.

Besides the transmitted message, mass media channels are modeled with respect to their influence, their reach, and their buzz (the modeling of the latter is addressed by Section 2.2.1, along with the other shared components). Influence parameter (Δ'_c) modulates the influence achieved by a mass media channel after impacting a given agent. This influence works similarly to the produced by WOM interactions, but in this case the distance change is calculated by using the specific change value of the channel c and the polarization of that channel during that time-slot. Since an agent could receive the same message multiple times, the maximum influence is bounded by an overall influence value (Δ_c^{max}). In addition, the influence previously accumulated by the channel ($\delta_c'^k$) is treated similarly to the one accumulated by WOM. An agent i experiences a change of its distance values for party k by the influence of a given channel c following Equation I.11, where $m_c^k(t)$ refers to the resulting polarization for party k in time-step t and Δ_c^{max} represents the maximum amount of influence that can be supplied by c. However, as the simulation progresses and new messages are produced by mass media channels, the agents tend to forget previous messages [WH07, YL10]. We model this effect using the influence decay ($d\Delta_c$) parameter, that regulates the rate at which the agents are forgetting previous influences, analogously to WOM. The distance value update of agent i for party k due to the effect of decay of channel c is defined by Equation I.12.

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) + (\Delta_c^{max} - \delta_c'^k(t))\Delta_c m_c^k(t). \tag{I.11}$$

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) - \delta_c'^k(t) d\Delta_c. \tag{I.12}$$

2.2.4 An ABM for defining marketing strategies

The ABM for marketing strategies considers a terminating simulation with T weeks of a virtual market that comprises a set of brands B. Using a time-step of a week, the model simulates the behavior of N agents and their reaction to social influence through a social network in a WOM process and external influences (advertisement) through a set of C mass media channels. The model has two main outputs or KPIs: brand awareness and WOM volume (i.e., the number of WOM interactions among the consumers). We select these KPIs because they have an important role in market expansion [LMP13, MS00]. A general scheme of this model is presented in Figure 2.

Brand awareness represents the customer's recognition of a certain brand, which is a key indicator in certain markets [HSV12]. If an agent is not aware of the brand, it will neither be

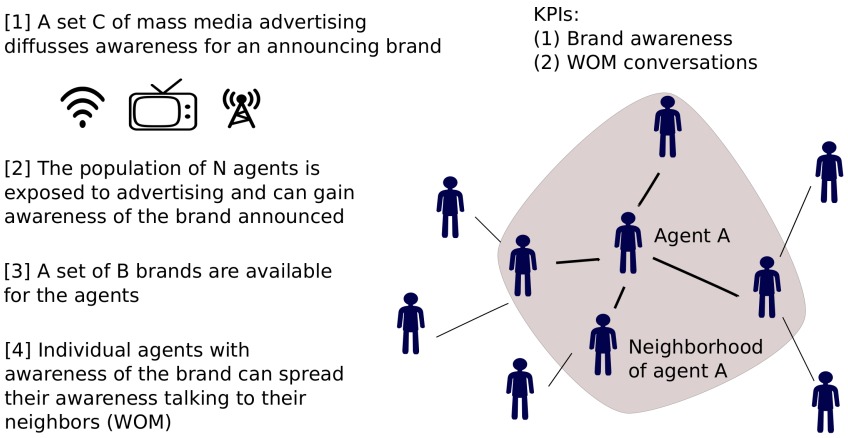


Figure 2: General scheme and structure of the ABM with an example of a brand advertised using mass media. The agents exposed to advertising can gain awareness of the brand announced and talk about it to their neighbors.

able of talking about it or purchasing it. The awareness values of the agents are modeled using a state variable called $a_i^b \in \{0,1\}$. If $a_i^b(t) = 1$, then the agent i is aware of brand b at time step t. Otherwise, the agent does not have awareness of brand b. This state variable is initialized using an initial awareness parameter set for each brand ($a^b(0) \in [0,1]$) which is the global awareness of the population and fulfills $a^b(0) = \frac{1}{N}\sum_{i=1}^{N} a_i^b(0)$. Therefore, the initial awareness parameter for each brand specifies the percentage of agents that have awareness of that brand at the beginning of the simulation. This process is carried out during the initialization of the agents, where they activate their awareness of each brand b with probability $a^b(0)$.

The awareness values of the agents do not remain static but change during the simulation: agents may loose or gain awareness of any brand at each step of the simulation. On the one hand, agents may gain awareness of a brand due to advertising or due to interacting with other agents through a WOM diffusion process. On the other hand, if the awareness of a brand is not reinforced, it may be lost over time because of a deactivation process [WH07, YL10].

We model these losing/gaining effects with additional parameters. The parameter regulating the rate at which awareness is lost over time is called awareness deactivation probability ($d \in [0,1]$). This parameter is modeled as follows. At the beginning of each step t, the agent i checks all the brands the agent is aware of ($a_i^b(t) = 1, \forall b \in B$). Each of these awareness values will be deactivated with a probability d by setting $a_i^b(t) = 0$. If the deactivation takes effect, the agent could still re-gain awareness due to the WOM diffusion and/or the mass media channels during the same simulation step, but it will not check for deactivation until the next simulation step.

In addition, each agent stores the number of conversations produced during its diffusion process in order to compute the WOM volume for each brand ($\omega_i^b(t)$). This way, every time an agent starts a diffusion process and talks with its neighborhood, the variable $\omega_i^b(t)$ will be updated by increasing it with the number of agents' neighbors (i.e., conversations). Finally, it will update the global $\omega^b(t)$ variable for the respective brand and time step.

The agents of the model can spread their awareness values during the simulation through the artificial social network. We model this social interaction as a contagion process which allows information diffusion through the nodes of the social network depending on their connectivity [NBW06, STSH10, ZKZ11, ZKZ19]. Every agent i has a talking probability ($p(t)_i^b \in [0,1]$) to spread the brands it is aware of at time step t (i.e., for every brand b where $a_i^b = 1$). This probability p_i^b specifies when agent i talks with all of its neighbors in the artificial social network, having the chance of transferring its awareness (i.e., a contagion process). We model this contagion

effect using a parameter called WOM awareness impact ($\alpha^{WOM} \in [0, 1]$), which represents the probability for an agent in the neighborhood to be aware of a brand after having a conversation about it.

```
begin
    reach_step = 0;
    total_hits = χ_c^b(t) · 0.01 · N;
    reach_increment = 1 / N ;
    i = 0;
    while i < total_hits do
        select agent randomly;
        if selected agent was already hit then
            impact agent;
            i++;
        else if reach_step + reach_increment ≤ r_c then
            impact agent;
            i++;
            reach_step += reach_increment;
    end
end
```

Algorithm 1: Pseudo-code of the advertising scheduling of the model for a given brand, time step and channel.

The advertising campaigns of the mass media channels are modeled using gross rating points (GRPs). In advertising, a GRP is a measure of the magnitude of the impressions scheduled for a mass media channel [FBPR10]. Specifically, we use the convention that one GRP means reaching 1% of the target population. The investment units in GRPs for channel c by brand b and time step t is modeled by the variable $\chi_c^b(t)$. Each channel has different costs for GRP and the brands need to carefully choose their investment since increasing the population awareness using mass media channels implies a monetary cost. Using both the supplied GRPs for a given brand and the reach values for a mass media channel, we are able to model brand advertising. Algorithm 1 shows the scheduling algorithm for modeling impacts of the media channels over the population.

Name	Description
N	Number of agents
$\|B\|$	Number of brands
$\|C\|$	Number of mass media channels
T	Number of time-steps
$a^b(0)$	Initial awareness for brand b
d	Awareness deactivation probability
m	Parameter for social network BA algorithm
$p_i^b(0)$	Initial talking probability, shared by all brands
α^{WOM}	Awareness impact for social interactions
χ_c^b	GRP units invested by brand b in channel c
r_c	Reach for channel c
α_c	Awareness impact for channel c
τ_c	Buzz increment for channel c
$d\tau_c$	Buzz decay for channel c

Table I.2: List of parameters of our proposed marketing model.

In addition to the parameters depicted in Section 2.2.1, each mass media channel has an awareness impact parameter ($\alpha_c \in [0, 1], \forall c \in C$) that defines the probability of the agent becoming

aware of the brand after one impact. If the agent is not aware of the brand at a given time step t ($a_i^b(t) = 0$), this probability α_c will activate the awareness of the agent for brand b. Finally, a summary of the complete set of model parameters is listed in Table I.2.

3 Metaheuristics for ABM calibration

This section reviews the main metaheuristics addressed during this doctoral dissertation. As previously introduced, metaheuristics are a family of approximate non-linear optimization techniques that provide *acceptable* solutions in a reasonable time for solving hard and complex problems in science and engineering [Tal09]. The remaining contents split between those metaheuristics designed for single-objective optimization problems (reviewed in Section 3.1) and those designed for multiobjetive ones (reviewed in Section 3.2).

3.1 Single-objective algorithms

This section introduces the design of relevant metaheuristics for calibrating ABMs by considering a single function to optimize. In addition, a memetic design based on their hybridization with a local search procedure is described in detail.

3.1.1 Steady-State Genetic Algorithm

The first considered metaheuristic is a steady-state genetic algorithm (SSGA) [BFM97, Sys89]. Genetic algorithms perform the search process by creating a population of individual solutions and evolving them through a determined number of generations. At each generation, the algorithm creates new individuals by applying crossover and mutation operators over the population, which are applied with a probability p_c and p_m, respectively. The crossover operator selects two parent individuals and generates two offspring combining the information of both parents. The mutation operator modifies the information of a solution by slightly changing its values. In our steady state approach, only two offspring are generated in each generation and they compete with the worst two individuals of the population to survive. The population inherently keeps the best solutions found so far (elitism). Our selected design for SSGA includes a BLX-α crossover operator [ES93] and an uniform mutation mechanism. In addition, the parent individuals for the crossover operator are selected using a k tournament selection mechanism [Tal09].

3.1.2 Differential Evolution

Differential evolution (DE) [SP97] is a metaheuristic considered in several automatic calibration approaches [LKGG$^+$19, ZC15, CLG14] and has been shown to provide a better performance than the basic genetic algorithms used in other contributions when calibrating ABMs [HS15]. DE generates new solutions by combining the existing individuals with a *donor vector* created following the next equation: $x_i(G+1) = x_{r1}(G) + F(x_{r2}(G) - x_{r3}(G))$, where x_i is the generated donor vector, F is a mutation rate, and $r1$, $r2$, and $r3$ are different solutions at generation G. For each generation, a donor vector x_i is generated for every individual i and its values are combined with those of the original individual by means of an uniform crossover according to a crossover rate (CR) parameter. For every gene of the newly created individual, the algorithm takes the value of the donor vector with probability CR, otherwise it takes the original value of individual i. If the fitness value of the resulting individual is better or equal than that of individual i, it will replace i in the population. Additionally, the selected design considers the variant DE/best/1/bin [SP97]. In this variant, "best" refers to the fact that the best individual in the population is always selected as the individual $x_{r1}(G)$ in the previous operator to obtain the mutated offspring $x_i(G+1)$. Meanwhile, the values "1" and "bin" respectively stand for the use of a single vector difference in that operator, as shown in the former equation, and of the uniform crossover operator, as also described.

3.1.3 Success-history based adaptive differential evolution with linear population size reduction

Success-history based adaptive differential evolution with linear population size reduction (L-SHADE) [TF14] is an extension of SHADE, a history-based variant of DE [TF13]. SHADE and other adaptive variants of DE self-adapt the values of CR and F during the optimization process, which are the main control parameters. In the case of SHADE, the *successful* values of CR and F are stored into a historical memory. A parameter value is *successful* if the solution generated using it improves the previous individual. In addition, L-SHADE extends SHADE by including a mechanism for reducing the population after each generation. This population reduction is computed linearly with respect of the number of fitness evaluations, which improves the convergence of the algorithm as the search process advances.

3.1.4 Restart CMA-ES with increasing population size

Restart CMA-ES with increasing population size (IPOP-CMA-ES) [AH05] is an evolution strategy that relies on a covariance matrix for sampling new search points. At each iteration of the optimization process, λ new solutions are independently sampled according to a multi-variate normal distribution. Then, the best μ solutions (or search points, as referred by CMA-ES literature) are weighted and summed for obtaining a new mean value for the distribution. CMA-ES employs the fitness information of previous iterations (called the evolution path) in combination with the latest solutions sampled for updating the covariance matrix. IPOP-CMA-ES extends the typical CMA-ES design by including a restart strategy that increases the current population size by a given factor each time the algorithm restarts because a stopping criteria is met. These stopping criteria typically involve a stagnation of the search process [AH05].

3.1.5 Coral Reefs Optimization

CRO [SSDSLT$^+$14, SS17, SSMBV17] is an evolutionary metaheuristic based on the processes occurring in a coral reef. A coral reef is represented by a two dimensional grid able to allocate a colony of corals (i.e., a population with variable size), where a coral represents a solution to the optimization problem. The CRO algorithm initializes some positions of the grid with random solutions, leaving the rest available spots empty. Algorithm 2 illustrates the flowchart diagram of the CRO algorithm, where the different mechanisms available in nature are recreated by applying known evolutionary process in four sequential stages:

1. **External reproduction.** This stage consists of the random selection of a fraction (F_b) of the existing solutions and the generation of a pool of offspring solutions. These new solutions are generated by applying a crossover operator or any other exploration strategy with two existing solutions as parents. Once a solution has been selected as parent, it is not chosen for reproduction purposes during an iteration.

2. **Internal reproduction.** This reproduction is modeled as a random mutation mechanism that takes place on the remaining fraction of solutions ($1-F_b$). A percentage P_i of the solution is mutated. The subset of mutated solutions is added to the pool of children solutions.

3. **Replacement.** Once the children solutions are formed either through external or internal reproduction, they will try to set in the grid. Each solution in the pool will randomly try to set in a position of the grid and settle if the location is free. Otherwise, the new solution will replace the existing one only if its fitness is better. Each child in the pool can attempt to occupy a position at each iteration only a determined number of tries η.

4. **Elimination.** At the end of each reproduction iteration, a small number of solutions in the grid are discarded, thus releasing space in the grid for the next generation. The elimination operator is applied with a very small probability (P_d) to a fraction (F_d) of the grid size with worse fitness.

Require: Valid values for the parameters controlling the CRO algorithm
Ensure: A single feasible individual with optimal value of its *fitness*
1: Initialize the algorithm
2: **for** each iteration of the simulation **do**
3: Update values of influential variables: reproduction probabilities, etc.
4: Reproduction processes
5: Replacement of new solutions in the grid
6: Elimination process
7: Evaluate the new population in the grid
8: **end for**
9: Return the best individual (final solution) from the grid

Algorithm 2: Pseudo-code for the CRO algorithm

3.1.6 CRO with substrate layers

There are other relevant interactions in real reef ecosystems that can be incorporated into the CRO approach for improving its performance in optimization and search problems. For example, different recent studies have shown that successful recruitment in coral reefs (i.e., successful settlement and subsequent survival of larvae) strongly depends on the type of substrate on which they fall after the reproduction process [Ver05]. This specific characteristic of coral reefs was first included in the CRO algorithm in [SSMBV17] in order to solve different instances of the Model Type Selection Problem for energy applications, resulting in CRO-SL (CRO with substrate layers). In [SSCGMH16], CRO-SL is enhanced with several substrate layers providing a different search procedure.

The inclusion of substrate layers in CRO can be carried out in a straightforward way: the artificial reef considered in the CRO is redefined in such a way that each cell of the square grid Ψ is now associated with a different exploration layer that indicates which structure it belongs to (search operator in this case). Each solution in the grid is then processed in a different way (with a different search operator) depending on the region (specific layer) in which it falls after the reproduction process. Figure 3 illustrates the flowchart diagram of the CRO-SL algorithm. Any exploration operator can be integrated in the algorithm as a substrate layer as long as it follows the determined coding scheme. Figure 4 shows an example grid with a few substrates, which are detailed below.

- **Uniform mutation**. The uniform random mutation is a traditional mutation operator that replaces the value of a gene for a given individual by a random value. This value is generated using an uniform distribution from an interval defined by lower and upper bounds.

- **Random walk mutation**. This operator modifies the values of a given individual by including neighborhood information into the mutation process. In the context of the present problem, the neighbors of a solution are accessed increasing or decreasing its values. This way, this operator modifies the value of each parameter of the individual solution with probability p_m by randomly moving to a neighbor. After moving to a random neighbor, the operator will keep moving to a random neighbor until the probability check fails.

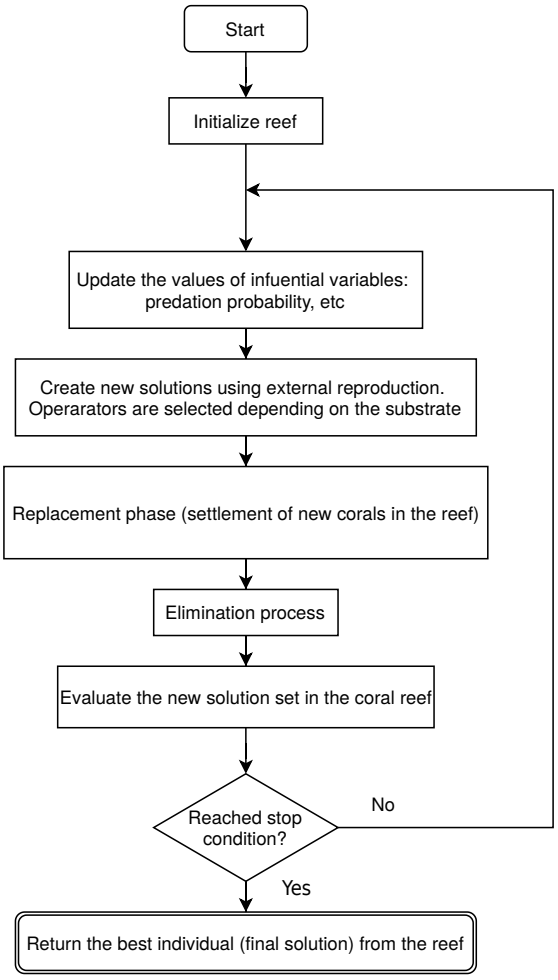


Figure 3: Flowchart of the CRO-SL algorithm, which modifies the reproduction stages of the original CRO by selecting the operator mechanisms depending on the substrate layers.

- **Simulated binary crossover** (SBX). SBX [DA94] performs the crossover operation emulating the behavior of a single-point crossover from binary-encoding. This operator works as follows: given two parents $P_1 = (p_{11}, ..., p_{1n})$ and $P_2 = (p_{21}, ..., p_{2n})$, SBX generates two springs $O_1 = (o_{11}, ..., o_{1n})$ and $O_2 = (o_{21}, ..., o_{2n})$ as $o_{1i} = \bar{X} - \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$ and $o_{2i} = \bar{X} + \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$, where $\bar{X} = \frac{1}{2}(p_{1i} + p_{2i})$. $\bar{\beta}$ is a random number fetched from a probability distribution which is employed as a spread factor.

- **BLX-α crossover**. BLX-α [ES93] takes new values from the interval defined by $[c_{\min} - I\alpha, c_{\max} + I\alpha], c_{\max} = \max(v_i^1, v_i^2)$, $c_{\min} = \min(v_i^1, v_i^2)$ and $I = c_{\max} - c_{\min}$. v_i^1, v_i^2 represent the decoded values from the chromosome of the parents. In this regard, α defines the level of exploration for the operator. For example, if α is set to 0, BLX-α works as a flat crossover.

Nevertheless, it is noted that the recent surge of novel bio-inspired algorithms has been subject of controversy [Sör15] due to the lack of scientific rigor behind some of these algorithms. CRO-SL is not one of those waste-of-time "novel" algorithms, as it can be justified beyond the novelty of the metaphor and only focusing on its actual design and performance [BCD+18]. The novelty of CRO-SL resides in an excellent exploration-exploitation trade-off and robustness as a consequence of the combination of multiple search patterns in its scheme. Thus, the CRO-SL general approach aims for a grid-based competitive co-evolution [FM08] in just one population where

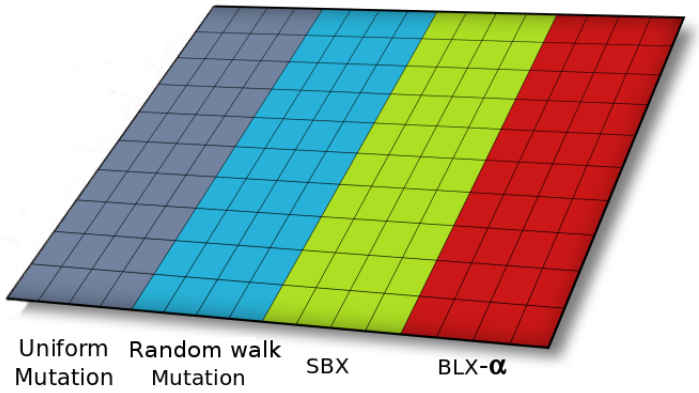


Figure 4: Example of the exploration layers in the CRO-SL algorithm when considering uniform mutation, random walk mutation, SBX, and BLX-α .

each substrate layer represents a different search process (different models, operators, parameters, constraints, repairing functions, etc.). Therefore, CRO-SL is a powerful multi-method ensemble where multiple search strategies coexist in the same population and can be classified as a low-level competitive single population approach [DOM+19, WMS19]. Such design provides CRO-SL with an exceptional versatility, being able to adapt different approaches to tackle a wide variety of problems while considering a single, general algorithmic scheme.

Finally, CRO-SL usually converges quickly to high quality solutions even in multimodal search spaces, being suitable for computationally expensive optimization problems both satisfying quality and computation time constraints. However, its performance varies significantly depending on the CRO's parameters and the different substrates included in the simulated reef. This idea of CRO with substrate layers as a competitive co-evolution algorithm was successfully tested in different applications and problems such as micro-grid design [SSCGMP+16], vibration cancellation in buildings [SSCGM+17], 3D medical image registration [BCD+18], as well as used for the evaluation of novel non-linear search procedures [SS16].

3.1.7 Memetic algorithms

Memetic algorithms [Mos89] combine the exploration capability of global search methods with the intensification of local search procedures, improving the quality of the candidate solutions found during the regular execution of the algorithm. The most important design issue is to define a trade-off between the local search procedure and the global search method [IYM03]. Some common approaches to memetic algorithms apply this local search refinement to every solution obtained during its run but this strategy can be time-consuming and it has proven to not always lead to the best performing memetic design [KS00]. Therefore, the most popular approach performs local search refinement selectively with a probability p_{LS} [KS00, LHKM04]. Thus, all the solutions in the population are candidates for receiving local search refinement since each solution runs a probability check at the end of a generation.

3.2 Evolutionary multiobjective optimization algorithms

In multiobjective optimization problems, the quality of a candidate solution is evaluated regarding multiple conflicting criteria instead of considering a single error measure. Thus, the optimization algorithm aims to either maximize or minimize $F(x) = f_1(x), ..., f_m(x)$, where m represents the number of objectives and x is the set of decision variables for the optimization problem (i.e., decision space).

The quality of the provided decision variables is evaluated using typical semantics from multiobjective optimization such as the Pareto *dominance* concept [CLVV+07]. Given two feasible solutions u and v from the decision space with $u \neq v$, u *dominates* v if $u_i \leq v_i, \forall i : 1 \leq i \leq m$ and $\exists j : 1 \leq j \leq m : u_j < v_j$, i.e., if u is equal or better than v for every objective and strictly better for at least one objective. However, these inequalities should be reversed for any objective that is being maximized (to *dominate* means to be better). Using the dominance concept, the global Pareto-optimal solutions are those vectors u such that there is no feasible vector v that dominates u. A set of u solutions where there is no v that dominates any of the other solutions is called a Pareto-optimal set. In addition, the representation of the solutions in the Pareto set as points from the objective space is called a Pareto-optimal front [CLVV+07].

Evolutionary multiobjective optimization (EMO) algorithms can be considered the best approach to multiobjective optimization, as they obtain a set of representative Pareto-optimal solutions (i.e. a Pareto-optimal set approximation) in a single run. In addition, EMO algorithms obtain Pareto-optimal solutions in a reasonable time and can perform successfully without requiring specific properties of the optimized functions [Deb01]. There are three categories of EMO algorithms depending on their behavior and how they carry out the optimization process: based on Pareto dominance, indicators, and decomposition. A set of the most extended EMO algorithms are reviewed as follows. In addition, we provide an overview of the most common methods for assessing the performance of the EMO algorithms.

3.2.1 EMO algorithms based on Pareto dominance

Pareto dominance-based EMO algorithms assign the quality of the solutions (thus guiding the selection mechanism) according to their dominance over other solutions in the population. The most extended Pareto dominance-based algorithms are NSGA-II and SPEA2.

- **NSGA-II** [DPAM02] can be identified as the most popular and well-known EMO algorithm. NSGA-II's approach relies on non-dominated sorting, which allows it to combine elitism with good levels of diversity in a single population while being computationally fast, specially for problems with two or three objectives. NSGA-II produces an offspring set Q_t at each generation using the solutions of the previous set P_t. Then, both sets are merged into the temporary set R_t where previous and newly generated solutions are ranked according to its non-dominance level. The non-dominance level of a solution corresponds with the number of solutions that dominate it. The next set P_{t+1} is generated by selecting the solutions with the best ranking, which are the solutions not dominated by other solutions in the previous set. This process is iterated for the next ranks until a population size $|P|$ is reached. This strategy guides the algorithm to non-dominated regions while a set of non-dominated solutions are maintained in the population. The first solution set that does not fit P_{t+1} is filtered using a crowding mechanism for boosting the diversity of the new population.

- **SPEA2** [ZLT01] is a well-known EMO algorithm that computes the fitness of its individuals calculating a "strength" value that represents how many solutions each of them dominates. Then, the fitness value for each solution is computed by summing the "strength" values of the solutions that dominate it. SPEA2 considers a separate population, named the "archive" ($\overline{P}_t$), designed to store non-dominated solutions. At each step, non-dominated solutions in P_t and $\overline{P}_t$ are copied to $\overline{P}_{t+1}$. If $\overline{P}_{t+1}$ exceeds the size of P, then its solutions are filtered using a truncation operator inspired in the *k-th* nearest neighbor method that selects the solutions with the minimum distance. If there are not enough solutions for filling $\overline{P}_{t+1}$ then the dominated solutions with the minimum fitness are included until $|\overline{P}_{t+1}| = P$. Then a mating pool is set using binary tournament on $\overline{P}_{t+1}$. Finally P_{t+1} is the result of applying the crossover and mutation operators to the mating pool.

3.2.2 EMO algorithms based on indicators

Indicator-based EMO algorithms assign the fitness of the solutions using indicator values (see Section 3.2.4). Some relevant indicator-based EMO algorithms are IBEA, SMS-EMOA, and MOMBI2.

- **IBEA** [ZK04] is a classic EMO algorithm that qualifies solutions regarding their relative contribution to a given performance indicator with respect to the rest of solutions of the population. Therefore, IBEA computes the loss of quality of removing a solution from the population using dominance preserving binary indicators. In order to carry out this task, some suitable indicators would be the additive I_ϵ or the I_{HD} indicator, that is based on the concept of hypervolume [ZT99]. Using these concepts, IBEA's fitness evaluation for solution x using a binary indicator I and a scaling parameter κ is computed as $F(x) = \sum_{y \in P \setminus x} -exp[-I(y,x)/\kappa]$. Finally, IBEA performs elitism and only the worst solutions of the population are removed, althought this implies that the fitness of the remaining solutions need to be updated each time a solution is removed from the population.

- **SMS-EMOA** [BNE07] introduces the maximization of the dominated hypervolume into the search process for approximating the true Pareto front. SMS-EMOA borrows NSGA-II's non-dominated sorting mechanism for merging the current population P_t with the offspring population Q_t into P_{t+1}. However, SMS-EMOA considers a replacement strategy that targets the solutions from the worst front with the lesser contribution to the hypervolume of their respective front. This process maximizes the quality of the population regarding their hypervolume [ZT99]. In addition, as the repeated calculation of hypervolume values is computationally expensive, SMS-EMOA follows a steady-state scheme for easing the replacement mechanism, thus allowing an easy parallelization of the fitness evaluation. Unlike other EMO algorithms like SPEA2, SMS-EMOA does not consider a separate archive for storing non-dominated solutions. Instead, it maintains a population of constant size that includes dominated and non-dominated solutions, as NSGA-II does. SMS-EMOA also preserves the extreme solutions (i.e., the ones with best fitness for one objective and worst fitness for the other) into the population for biobjective problems such as our ABM calibration problem instead of requiring a reference point for computing hypervolume. For problems with more objectives, a reference point is calculated dynamically at each generation.

- **MOMBI2** [HGC15] relies in the R2 quality indicator for ranking the solutions, a Pareto compliant indicator with a reduced computational cost. This quality indicator uses a utility function for mapping each objective into a single value. A common MOMBI2 configuration employs the achievement scalarizing function (also used by GWASF-GA) since it allows the algorithm to obtain weekly Pareto optimal solutions, although there are several candidate utility functions for the algorithm. In addition, instead of updating the nadir point at each generation, MOMBI2 updates this reference point taking into account its historic values during previous generations. This update takes two parameters α and ϵ as the threshold and the tolerance threshold, respectively. These historic values are used for estimating how far current solutions are from the true Pareto front: a high variance suggests that the solutions are far from it and a low variance suggests that the solutions are close. The solution ranking using R2 proceeds as follows: first, the solutions with the best rank (i.e., those that optimize the weight vectors) are selected, removed from P_t and introduced into P_{t+1}. Then, this process goes on ranking solutions until every solution has been ranked and $|P|$ solutions are selected. In case two solutions provide the same utility value, the solution with the lowest Euclidean distance is selected.

3.2.3 EMO algorithms based on decomposition

Decomposition-based EMO algorithms transform a given multiobjective problem into several subproblems. Some relevant decomposition-based EMO algorithms are MOEA/D and GWASF-GA.

- **MOEA/D** [LZ09, ZL07] is an evolutionary algorithm that has received great attention in the evolutionary computation literature in the last few years. It employs decomposition techniques for reducing the multiobjective problem into as many subproblems as individuals ($|P|$). Then, MOEA/D solves every subproblem jointly by evolving its solution population (P_t), which contains the best solution found for each subproblem. The optimization of each subproblem is performed by only using information from its neighboring subproblems. Although MOEA/D is compatible with any decomposition approach that transforms the Pareto approximation problem into several scalar optimization problems, we choose the Tchebycheff approach in this dissertation, as recommended by the authors [LZ09, ZL07]. In addition, MOEA/D uses an external population for storing the non dominated solutions found during the execution of the algorithm, similarly to SPEA2. Finally, we select a DE operator as the crossover strategy instead of the SBX employed in the other algorithms, also following authors' recommendation [LZ09]. This operator generates each offspring $C = (c_1, ..., c_n)$ as $c_i = P_1(i) + F \cdot (P_2(i) - P_3(i))$ with probability CR and $c_i = P_1(i)$ with probability $1 - CR$, where P_1, P_2, and P_3 are the donor individuals and CR and F are the control parameters.

- **GWASF-GA** [SRL17] is a recent aggregation-based evolutionary algorithm. GWASF-GA approximates the true Pareto front transforming the original problem into a set of scalar subproblems that are minimized using the achievement scalarizing function, based on the Tchebychev distance. This scalarizing function uses two reference points: the nadir point and the utopian point. The former is a point containing the worst objective values of the solutions of the entire Pareto-optimal set. The latter is a point that is chosen for dominating the ideal point and that will not be obtainable for any solution. During each algorithm iteration, every solution in the population is classified into different fronts by computing their achievement scalarizing function values using the two mentioned reference points and a set of weight vectors. Each of these fronts contains the solutions with the lowest scalarizing function value for the weight vectors in the set. The set of weight vectors is predefined for ensuring that its inverse is well distributed, ensuring that the algorithm maintains diversity. Then, the fronts with the lowest function values are introduced into the next population until $|P|$ solutions are selected.

3.2.4 Analysis of EMO algorithms' performance

There are two different kinds of multiobjective performance indicators, unary and binary. Unary performance indicators evaluate a single Pareto front approximation individually. We have selected hyper-volume ratio (HVR) [CLVV$^+$07] as our unary performance indicator. HVR measures the distribution and convergence of a given Pareto front approximation. It is defined as $HVR = \frac{HV(P)}{HV(P*)}$, with $HV(P)$ being the volume of the given Pareto front approximation and $HV(P*)$ the volume of the true Pareto front. However, in the ABM calibration problem tackled by the current dissertation, we do not know the true Pareto front for any of the model instances, so we use a pseudo-optimal Pareto front for computing the HVR values instead. The pseudo-optimal Pareto front is an approximation obtained by merging all the Pareto front approximations generated by every algorithm for that instance in every independent execution and removing the dominated solutions.

A binary performance indicator compares two given Pareto front approximations generated for the same problem. Our selected binary performance indicator is the multiplicative I_ϵ measure [ZTL+03]. The calculation of $I_\epsilon(P,Q)$ for Pareto front approximations P and Q is shown in the following equation: $I_\epsilon(P,Q) = \inf_{\epsilon\in\mathbb{R}}\{\forall z^2 \in B\ \exists z^1 \in A : z^1 \succeq_\epsilon z^2\}$. The value computed by $I_\epsilon(P,Q)$ represents the minimum factor required to multiply every element in P in order to weekly dominate Q. That is, the minimum ϵ so P ϵ-dominates Q. As our calibration problem constitutes a minimization problem, if $I_\epsilon(P,Q) < I_\epsilon(Q,P)$ then we can assume that P is better than Q.

In addition, it is necessary to develop a statistical test and study the significance of the I_ϵ values to avoid that isolated results could bias the analysis. This test can be performed following the methodology described in [SV08, CÁCD13a]: let N be the number of repetitions of two algorithms A and B; then let A_i and B_j be two arbitrary resulting Pareto front approximations with $1 \leq i \leq N$ and $1 \leq j \leq N$; finally let $p_{A_i}(B_j)$ be 1 if A_i dominates B_j based on the computed I_ϵ value (i.e., $I_\epsilon(A_i,B_j) \leq 1$ and $I_\epsilon(B_j,A_i) > 1$) and 0 otherwise. Using $p_{A_i}(B_j)$, we can define P by the equation $P_{A_i}(B) = 1/N \cdot \sum_{j=1}^{N} p_{A_i}(B_j)$ as the percentage or resulting Pareto front approximations obtained by algorithm B that are dominated by A_i. The resulting ϵ dominance percentage values can be easily visualized using boxplots.

Finally, the vector $P_A(B) = P_{A_1}(B), P_{A_2}(B), \ldots, P_{A_N}(B))$ can be considered as a random variable representing the percentage of times that algorithm A outperforms algorithm B, since it is the proportion of resulting Pareto set approximations of algorithm A dominating the Pareto set approximations delivered by algorithm B. Therefore, if the expectation of $P_A(B)$ is greater than the expectation of $P_B(A)$ we can claim that A is better than B because it is more likely that the resulting Pareto set approximations of A dominate those obtained by B. In this stage a statistical test or an post-hoc procedure can be applied on the generated percentage values. For example, the Wilcoxon ranksum test can be chosen (null hypothesis $E(P_A(B)) = E(P_B(A))$, alternate hypothesis $E(P_A(B)) > E(P_B(A))$), since it has proven to be useful when analyzing the performance of evolutionary algorithms [GMLH08].

3.2.5 Visualization assisting multiobjective optimization

The use of visualization techniques is one of the main resources assisting users and decision makers when dealing with multiobjective-problems [TF15b]. They complement the information obtained by other tools such as performance metrics, since they enable the identification of insights that cannot be easily otherwise [GNL19]. We can see how many of the visualization methods used in EMO are mainly generic data visualization techniques commonly employed in data analysis, specially if the analyzed multiobjective problem considers two or three objectives [FT18]. However, problems with a higher number of objectives require of specific visualization methods developed in the EMO field that can deal with the special properties of the solutions forming a Pareto set approximation (for instance, the dominance relations) [FT18, TD18, TD20].

Filipič and Tušar [FT18] proposed a taxonomy of methods and techniques for visualizing Pareto front approximations. A general scheme of their taxonomy is displayed by Figure 5. In order to show the differences between some of the following visualization methods, Figure 6 considers a benchmark with three typical front shapes: spherical, linear, and knee-shaped. This taxonomy mainly splits the visualization methods between those employed for visualizing a single Pareto approximation set and those used for visualizing repeated approximation sets. Then, the methods visualizing a single Pareto set approximation are also split between those intended for visualizing solutions independently from each other and those intended for showing certain properties of the set. The former category is ultimately split between those methods showing the original objective values of the represented solutions and those methods transforming the solutions values in any way.

In the category of methods showing the original objective values we can find several simple and intuitive methods such as the scatter plot matrices [FT18], the parallel coordinates [Ins85],

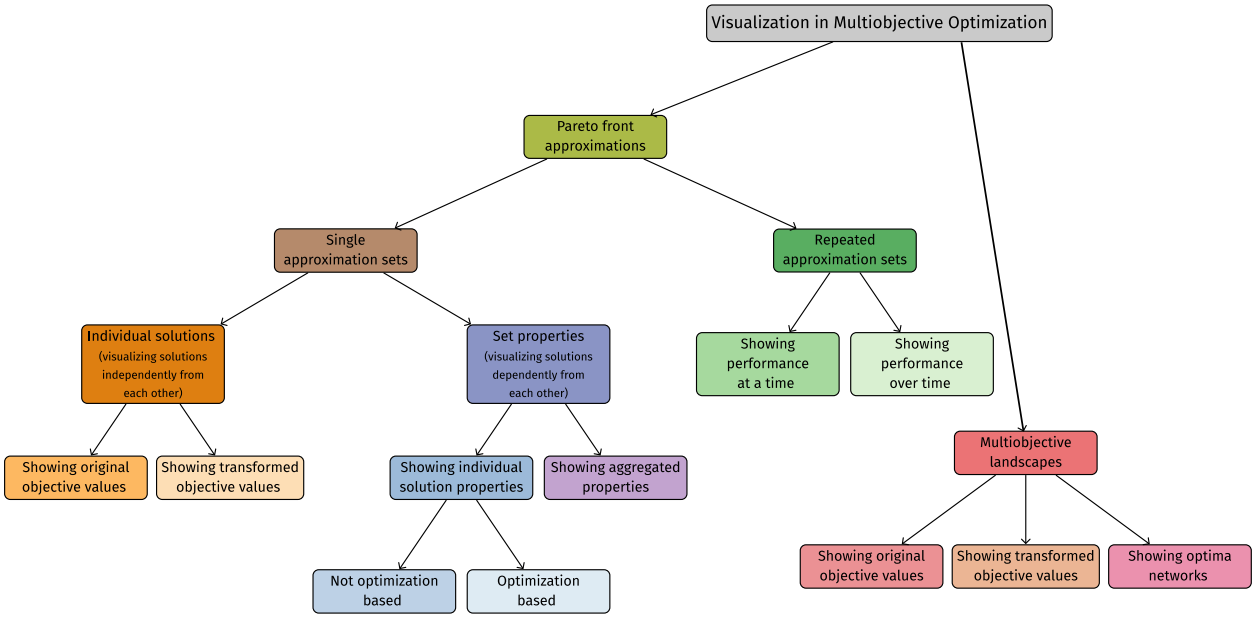


Figure 5: Taxonomy of methods and techniques for visualization in multiobjective optimization. The left branch focuses on the methods for visualizing Pareto front approximations. The small branch in the right includes the visualization of multiobjective landscapes, which applies to the visualization of solution sets in the decision space. Diagram extracted from [FT20b].

or the use of heat maps [PMN07]. In addition, more elaborated methods like interactive decision maps [LKB04] and the Nasseh method [LKB04] fit this category. The methods transforming the solutions values include radial coordinate visualization [HGM+97], level diagrams[BHSM08], polar plots [HY16], PalletteViz [TD20], and prosection plots [TF15b]. An example of the application of radial coordinate visualization is displayed by Figure 7a.

The visualization methods intended for identifying set properties are further split into different categories as well. First, the methods showing individual solution properties can be distinguished from the ones showing aggregated properties. Then, the former are divided into the ones that require an optimization procedure for generating the resulting visualization and the ones that do not. Several methods depending on an underlying optimization procedure can be highlighted, such as Sammon mapping [VB07], Isomap [KY12], and distance-based and dominance-based mappings [FE13]. Some examples of the visualization methods not involving optimization procedures are Pareto shells [WEF10], hyper-space diagonal counting [ABL+], and treemaps [Wal15]. Finally, self-organizing maps [OS03], trend mining [BN19], and moGrams [TCP+18] can be noted as visualization methods showing aggregated properties. Figure 7 shows an example of the use of Sammon mapping (7c) and hyper-space diagonal counting (7b).

Additionally, the methods considering the visualization of repeated approximation sets are split into two categories depending on how they measure the optimization progress. This differentiation is made depending on whether the progress of the optimizer is shown at a time (i.e., the performance improvement of an iteration with respect of the others) or if the progress is shown over time, as the optimization advances. The methods in the former category rely on the empirical attainment function [GdFFH01], which computes the frequency for different points on the objective of being dominated by the points in the different sets. Few methods following this approach would be line plots [FF96], heat maps [LIPS10], or direct volume rendering [TF]. The methods showing performance over time contain consider the average runtime attainment function[BAHT17], which represents the average number of evaluations to obtain each of the non-dominated solutions. The

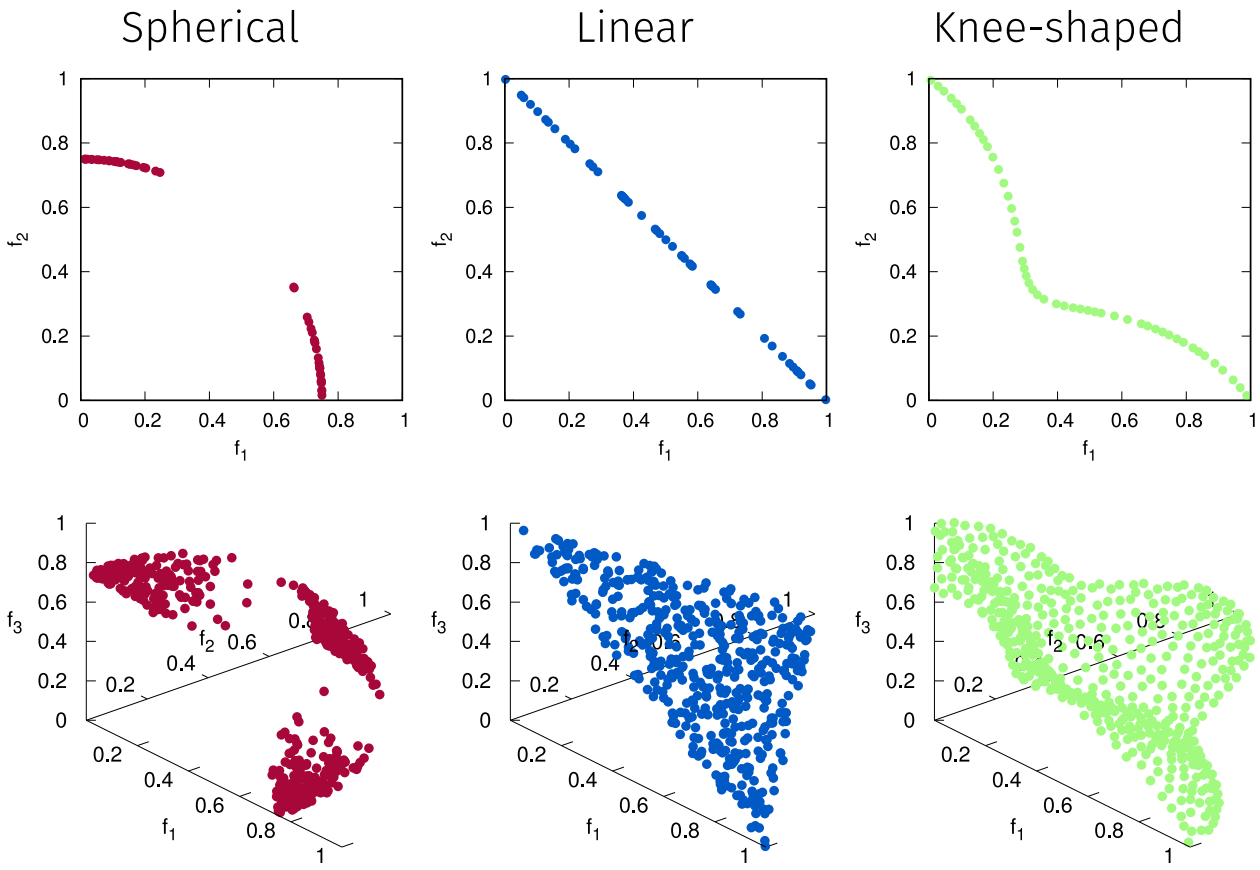
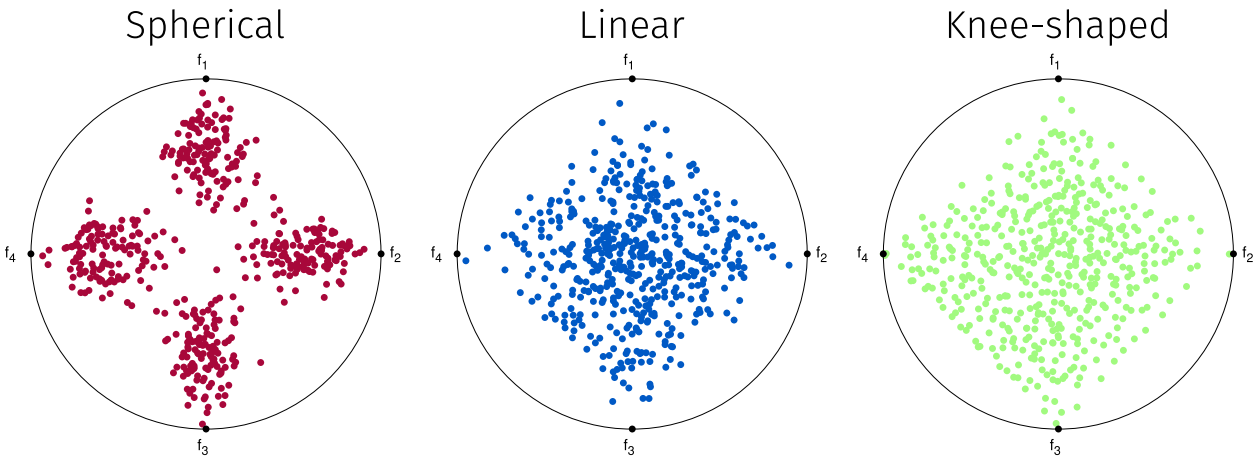


Figure 6: Visualization benchmark considering three typical front shapes: spherical, linear, and knee-shaped. Image extracted from [FT20b].

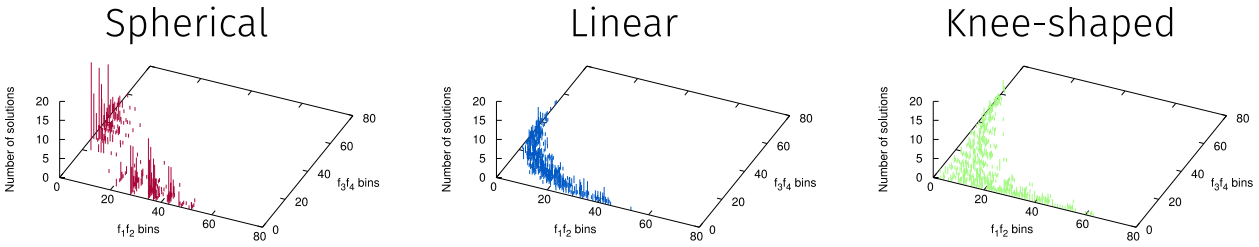
grid-based sampling [BAHT17] and circular indicator plot [WC20] are two methods fitting this category.

This taxonomy was latter extended for including the visualization of multiobjective landscapes [FT20b], which applies to the visualization of solution sets in the decision space. This last category is split into the visualization methods showing the original objective values, those showing a transformation of the objective values, and those showing optima networks. Two samples of visualization methods showing the original objective values would be the level sets (consisting on curves connecting points with the same objective value) and the line walks [BTAH19, VNKT19]. Few examples of visualization methods transforming the objective values would be dominance rank [Fon95], local dominance [FCAM19], and cumulative gradient [KG17]. Finally, the methods showing optima networks include the Pareto local optimal solutions network [LDV+18a], the Pareto local optima network [LDV+18b], and the dominance-neutral optima network [LDV+18b].

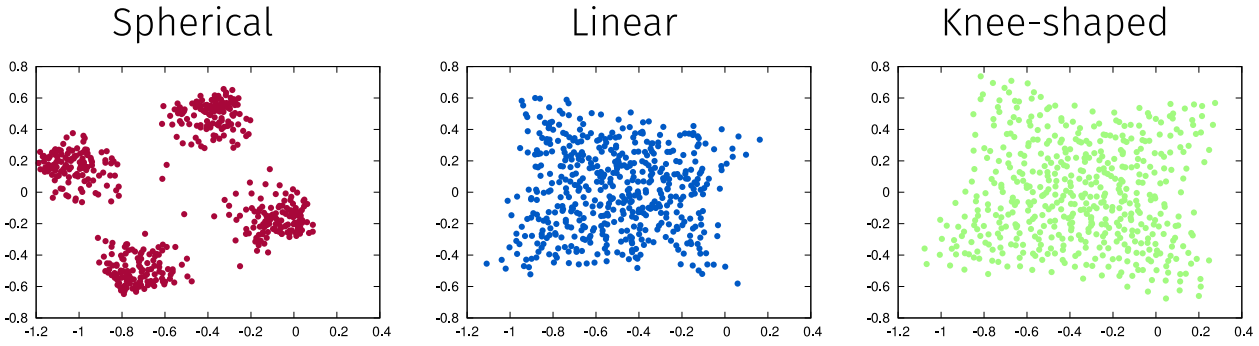
We can see how most of the contributions available in the literature focus on the visualization of the non-dominated solutions in the objective space [TF15a, WEF13]. In contrast, only a few contributions tackle the visualization of the solutions in the design space, which is the most interesting for discovering knowledge about the parameter values, and even less proposals derive joint visualizations for both the objective and design spaces [TCP+18]. From those few existing approaches of such kind, we belive that the *moGrams* methodology [TCP+18] is the best suited for its application in the problem tackled by this dissertation, as it mutually analyzes and visualizes the solutions obtained by EMO algorithms in the decision and objective spaces. More details regarding *moGrams* are supplied in Section 4.3.



(a) **Radial coordinate visualization**



(b) **Hyper-space diagonal counting**



(c) **Sammon mapping**

Figure 7: Some examples of the highlighted visualization techniques that belong to different categories in the taxonomy. Radial coordinate visualization shows individual solutions showing transformed objective values. Hyper-space diagonal counting displays individual solution properties without considering optimization. Sammon mapping visualizes individual solution properties considering optimization. Images extracted from [FT20b].

4 Model calibration and validation

Section 4.1 formally introduces the problem of model calibration. Then, Section 4.2 reviews relevant work on model calibration considering either single or multiple key performance indicators. Section 4.3 addresses the important role of visualization for model calibration. Afterwards, Section 4.4 introduces our experiments on the proposed ABMs for analyzing the 14-M elections. Finally, Section 4.5 reviews the different experiments developed using the proposed ABM for marketing strategies.

4.1 Problem definition

Model simulation is useful for representing and analyzing complex systems, but computational models require to be calibrated and validated before they can be used. This can be problematic because validating a model is not straightforward, specially if the modeling methodology involves the definition and setting of many parameters. In addition, creating and configuring a model for a specific problem from scratch can be difficult for designers and decision makers. If some of the model parameter values cannot be specified using the available information and knowledge, the modeler needs to manually estimate them for properly simulating the underlying system dynamics. The process of adjusting the values is known as model calibration and it is a crucial step during the validation of the model [CBK$^+$17, CCD$^+$16, Oli03]. This calibration step can be performed manually by the modeler, which results in a process similar to a global sensitivity analysis [tBvVL16], where the modeler repeatedly simulates the model and tunes its parameters based on the observed output. However, this approach is impracticable for many realistic models, which are characterized for considering many parameters.

This issue is dealt with using automatic calibration, a data-rich and computationally intensive process that attempts to discover the best parameter values for the model. This is done by comparing real-world data to the model output and tuning a selected set of model parameter values so that the output can properly match the data [Oli03, Sar05]. Figure 8 shows the main components for running an automatic calibration process. Automatic calibration requires a set of historical data, an error measure, and an optimization method for searching the parameter values space in a systematic way by minimizing the error measure. However, after the application of the calibration process, the resulting parameter values need to be carefully reviewed and validated, since a good fitting of the historical data does not ensure the validity of the model. Additionally, typical parameters of computational models exhibit non-linear interactions and usually the best approach is to consider a non-linear optimization algorithm such as a metaheuristic [Tal09]. Metaheuristics can search across a large span of the model parameter values space [CBK$^+$17, Mil98, SR14].

In an automatic calibration process, the values of the model parameters are adjusted to match the model outputs with the data-driven reality of the modeled scenario. Thus, each parameter configuration $X = (x_1, ..., x_n)$ can be defined as a vector of n decision variables. The modeler should carefully select the model parameters that will be estimated by automatic calibration since the difficulty of validating the calibrated configurations (i.e., the search space dimension) increases with the number of calibrated parameters. On the one hand, the modeler should consider those parameters being the most uncertain and the hardest to define by her/him according to the available information. On the other hand, sensitive parameters should also be considered for calibration since small changes in their values can significantly affect the model response and the global output. The parameters to be calibrated are usually coded as either integer or real values, but the optimization method should be able to properly handle the required coding scheme. Additionally, the selection of the optimization method for carrying out the calibration process heavily determines the quality of the resulting parameter configuration as the model accuracy relies on the ability of the method for exploring the parameter search space.

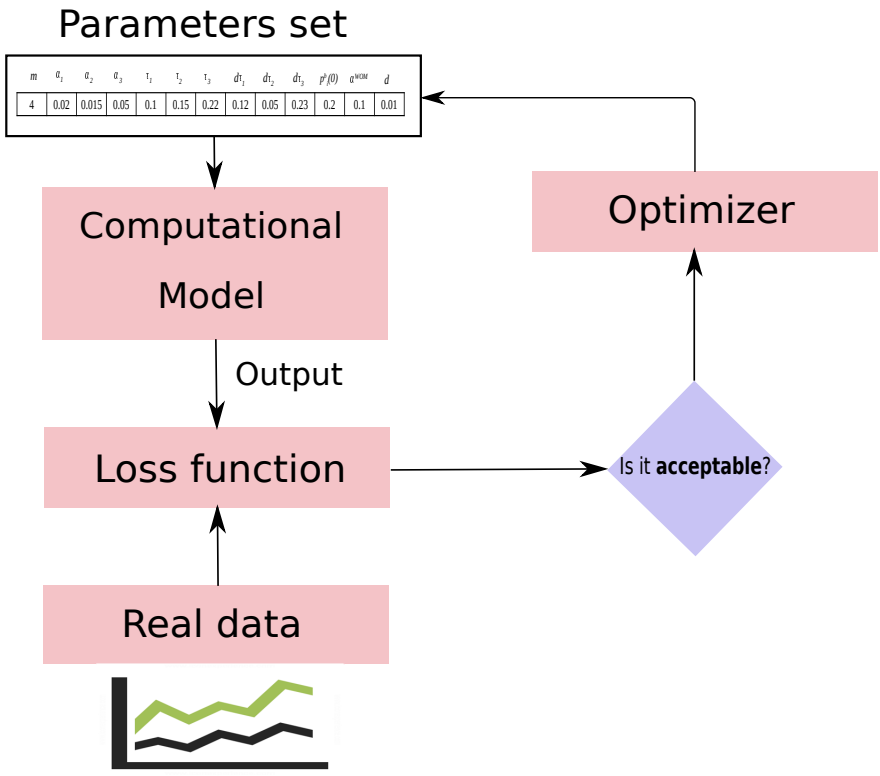


Figure 8: General scheme and main components for the automatic calibration of a computational model. The model is run considering a specific set of parameter values so its corresponding output can be compared with real data using a loss function. The result of this comparison is later used by an optimization method for generating a new set of parameter values. This loop is iterated until an *acceptable* configuration is found.

The fitting over the training historical data is computed using a deviation measure ϵ (i.e., the loss function) that guides the optimization method by calculating the error between the provided ground-truth data and the model output for a given objective. This distance can be computed using any of the standard deviation measures (RMSE, MAPE, or MARE [HK06], for instance). The modeler should choose the most appropriate for each output susceptible of being calibrated. Thus, the goal of the optimization method is to minimize $F(X)$ as defined by Equation I.13, where $\widetilde{o}$ represents the target ground-truth values for the calibrated output and $o(X)$ represents the simulated values of the model using the parameter configuration X. Note that the error measure ϵ is independent of the calibrated output and different deviation measures can be used when calibrating different outputs.

$$\min F(X) = \epsilon(o(X), \widetilde{o}). \tag{I.13}$$

4.2 Related work on model calibration

4.2.1 Calibration of a single key performance indicator

Automatic calibration techniques have been applied for calibrating the parameters of non-linear computational models from different areas. A few examples are market modeling [CR17, NMA+10], crowd modeling [ZC15], traffic simulation [KR03, NM12], and growth modeling [CLG14]. Some approaches consider the use of exact methods like simplex-based [KR03] or gradient-based [TKG14]. However, these approaches are employed for calibrating a relatively low number of parameters (i.e., no more than 20 calibration parameters) and its application to models involving more

than 100 parameters seems computationally prohibitive. Another approach is the cross entropy method [NM12], a stochastic optimization algorithm which results effective dealing with calibration problems with multiple local optima but it is also unclear how this approach can perform when dealing with high dimensional parameter value spaces.

In contrast, the use of metaheuristics is more convenient for the calibration problem when having higher dimensionality. There are several contributions addressing the application of metaheuristics for model calibration and parameter estimation. For instance, the interested readers can find genetic algorithms [DYX+09], evolution strategies [MD13, CLG14], and several versions of differential evolution [LKGG+19, ZC15, CLG14]. In terms of the ABM calibration, the use of metaheuristics is clearly the most extended approach. Thus, there are examples of metaheuristics for calibration that tackle models designed for different areas, such as social and biological sciences [CH06, CC15, CR17, Fab13, HS15, MSEH14]. However, one can see from these examples that different versions of genetic algorithms [BFM97] are usually the *de facto* approach, which could be improved by employing recent and more powerful evolutionary metaheuristics. Unfortunately, the state of the art in ABM calibration does not consider a reference benchmark for comparing these methods. Moreover, none of these previous efforts considers an exhaustive comparison of several metaheuristics for ABM calibration, as it is done in this dissertation.

Additionally, efficiently calibrating ABM is troublesome since the model needs to be simulated for evaluating the quality of a given set of parameter values, thus leading to a simheuristic approach [CJB+20]. Besides, the proper evaluation of an ABM requires multiple runs of the model for each parameter configuration by means a Monte-Carlo simulation, thus significantly increasing the computational burden. It can be noted the use of surrogate models for reducing the computational cost of estimating these values [vdH19, LRS18]. Specifically, machine learning algorithms have been employed for training a fast surrogate model that responds similarly to the changes of parameter values from the original model. Nevertheless, it is unknown how this approach could perform with ABMs having a high number of parameters, but authors of the latter publications claimed it can work with more than 30 parameters.

Finally, the use of surrogate fitness functions for tackling expensive optimization has also been addressed in evolutionary optimization [Bha13, BABR17]. In this case, the authors propose to replace the original fitness function by an approximate (faster) function. This surrogate function is designed relaxing the conditions of the original fitness function. For example, a fitness function for an ABM could be relaxed by either reducing the simulation steps of the model or reducing the number of agents. However, this approach can be problematic from the calibration point of view because the use of an appropriate surrogate function is problem-dependent and its validation is not straightforward.

4.2.2 Calibration of multiple key performance indicators

The calibration of computational models considering multiple criteria is less frequent but there are some examples of the use of EMO algorithms for multicriteria calibration of computational models [ALCJ15, GZL+14, KO13, Liu09, LS13, MD13, ZST16]. Many of them are focused in the calibration of hydrological models, such as the soil and water assessment tool [BN07, CW07, ZSL10, ZSVL08], the rainfall–runoff models [Liu09], empirical hydrological models for streamflow forecasting [GZL+14], and an integrated water system model [ZST16]. The thorough review of these contributions reveals that their usual approach relies on employing the NSGA-II for running the calibration process, probably because it is the most popular EMO algorithm. Apart from NSGA-II, we can also find some studies using SPEA2 [GZL+14, KO13, ZSL10, ZSVL08].

The application of EMO for multicriteria calibration of ABMs is not frequent although there are few examples tackling this issue [FNR+16, NMM06, RART16]. Farhadi et al. [FNR+16] present a framework for sustainable groundwater management including a Nash bargaining model, which is

implemented as an ABM incorporating cooperative and non-cooperative agents that consume the water of the modeled scenario at different ratios. The parameters of the model are calibrated using NSGA-II and considering three objectives and a single calibration scenario. Narzisi et al. [NMM06] deal with the calibration of an ABM for emergency response planning using NSGA-II. Their model is calibrated for minimizing the percentage of fatalities and the average waiting time of the population before receiving attention at the hospitals. This calibration process is applied to a single scenario considering ten real-coded model parameters with several restrictions. Finally, Read et al. [RART16] introduce the calibration of artificial murine multiple sclerosis simulation, an established immunological ABM for computational biology. They use NSGA-II in the calibration of 16 integer and real parameters with respect to four objectives. The authors consider a single scenario for the model and run three independent calibration executions for its analysis.

Therefore, NSGA-II is the recurrent EMO algorithm for multicriteria calibration of ABMs and there is not any comparative study on different EMO algorithms for this problem. In addition, it can also be recognized that the methodology followed by these contributions generally limit their experimentation to a single run of the EMO algorithm, which can be explained by the high computational cost of simulating multiple times a single model configuration for every evaluation. However, this approach is not taking into account that EMO algorithms are stochastic, thus requiring multiple runs using different seeds. Analyzing the results of a single calibration algorithm execution reduces the amount of information provided by the calibration process because valuable model configurations may be skipped in the initial run, specially if the EMO algorithm is not properly tuned.

4.3 Visualization for model calibration

The effective use of an ABM for representing a complex system heavily relies on the transparency of the underlying model, as ABM modelers and stakeholders require to understand how the model recreates a given behavior. This can be achieved from a white-box perspective [BDRD$^+$20, SWM17], where both modelers and stakeholders can make use of visualization tools for increasing the *explainability* of the model. Improving the understanding of artificial intelligence-based models is one of the goals of the emerging area of explainable artificial intelligence [BDRD$^+$20, SWM17]. Explainable artificial intelligence also empowers the solutions delivered by white-box models, since boosting the transparency of the delivered solutions should increase the trust in the behavior and performance of these solutions.

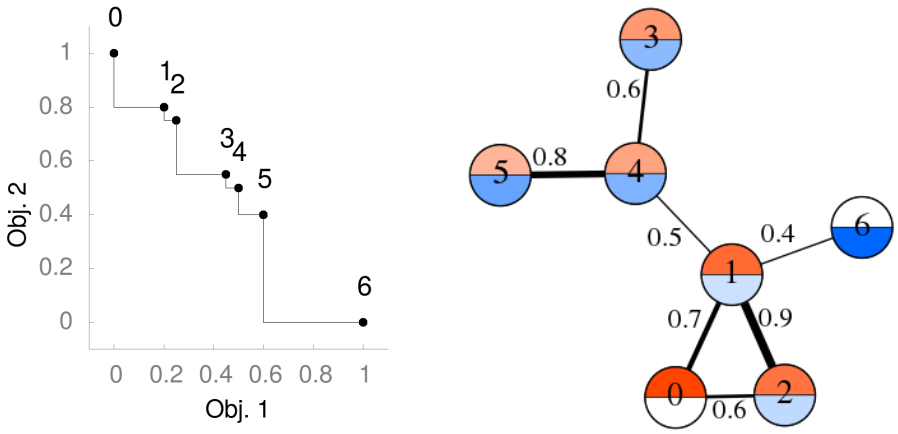
This fact highlights the role of visualization methods for model calibration since they are powerful tools that increase the understanding of the modeler on the calibrated model and its parameter settings [CBK$^+$17]. The use of visualization increases the transparency of the quality indicators (mostly focused on the fitting of the model to real data) for the validation of the calibrated model [BCG$^+$13, CBK$^+$17, LFLZ$^+$15]. Thus, visually representing the underlying relationships between an input configuration and its corresponding model output becomes a critical component of the validation process.

When the model considers two or more conflicting outputs, multiobjective visualization methods are specifically required for the validation of the model. As previously introduced, the *moGrams* methodology is one of the few existing approaches that mutually analyzes and visualizes the solutions obtained by EMO algorithms in the decision and objective spaces. moGrams [TCP$^+$18] follows a social network analysis-based visualization approach that represents the non-dominated solutions in a Pareto set approximation as nodes in a weighted network where each edge stands for a relationship between the connected solutions in the design space. The weights of the edges are computed using a similarity metric specifically defined for each problem by the designer. In order to improve the readability of the network, moGrams employs the Pathfinder network pruning algorithm [SDD89] for reducing the edges of the network leaving only the most

salient ones from a global viewpoint. In addition, each node is divided into sectors of the same size, each of them associated to a different objective, which are colored differently with its opacity proportional to the quality of the solution for the respective objective. For example, if a problem considers four objectives, the node is divided into four sectors with different colors. This way, the modeler has access to the whole information of both the objective and the design spaces at the same time. When considered in our problem (i.e., the calibration and validation of a model) it will provide additional information regarding the parameter settings of the different calibrated model configurations, highlighting similarities between them.

Figure 9 shows an example of the moGram that was generated for the Pareto set approximation of a problem with two objectives. The Pareto front approximation obtained for the generated moGram network is also shown since the joint visualization of both elements is suggested for better understanding the relationships in the design space. In this network example, we can identify two separate clusters connected by the bridge edge between Solution 4 and Solution 1, with the latter being the most connected solution. From the neighbors of Solution 1, we can observe that Solution 2 has the highest similarity relationship. This means that the parameter configuration of Solution 2 is highly similar to the one of Solution 1.

However, the high similarity between Solutions 1 and 2 could lead us to think that both solutions are close in the objective space. In this regard, moGrams provides the user with other relationships that are not intuitive, such as the relationship between Solution 1 and Solution 6. This relationship reveals that the closer configuration to Solution 6 is the one defined by Solution 1, which is located at the other end of the Pareto front approximation. Thus, the decision maker can detect parameters that drastically change the behavior of the solutions through this relationship. In addition, the topology of the moGram network allows the modeler to identify Solution 1 as the most flexible solution, which can be swapped with other solutions (such as Solution 2) with minimum changes on its decision variables.



(a) Pareto front approximation. (b) moGram representation.

Figure 9: Generated moGram network example for a given Pareto set approximation corresponding to a problem with two objectives.

4.4 Experiments: calibration of ABMs for the 14-M elections

This section reviews our experiments on the proposed ABMs for analyzing the 14-M elections. First, Section 4.4.1 depicts the data feeding our models. After that, Section 4.4.2 addresses the parameters of the ABMs that are tuned during their calibration. Then, Section 4.4.3 reviews our experimentation using the ABM tacking the elections from the point of view of the Framing

theory. Finally, Section 4.4.4 analyzes the impact of terror management strategies on the ideological distance between the voters and the political parties during the same period.

4.4.1 Data description

We use the data obtained from the following sources to set the values for some parameters and to restrict the search for the unknown parameter values:

- The final election results to match the model output have been take from [Obs04].
- The 2553 and 2555 study of the Spanish Centre for Sociological Research [Cen04] were used with a two-fold purpose:
  - An estimation of the voting intention before the 11-M attacks, with a sample size of 24,109 interviews [Cen04]. From that interview, we focus on the question regarding who they were willing to vote on the next national elections to be held in March 14th. We use these data to model initial resilience values for the agent population at step $t = 0$.
  - In addition, the ideological positions for the voter agents (v_i) and the ideological position where they allocate the considered parties (p_i^k) are also taken from this study. In the case of this question, the study considers a sample size of 1,500 interviews. In addition, we use the 2555 study (the post-elections survey) for calibrating the model. However, in this case we consider the average distance of the interviewed individuals as the target average distance values for the agents.
- Three different sources have been used for setting the polarization message value of $m_C(t)$ for the whole simulation:
  - We took the television information from the informational volume 19-20 from *Quaderns del Consell de l'Audiovisual de Catalunya* [Con04].
  - The audio from the radio was gathered from Cadena Ser, since it had the highest audience rate at that moment and received special attention during this period.
  - The information about the chosen newspapers (ABC, El Mundo, and El País) was directly collected from them and can be accessed from the collected on-line database [1].
- There has been a thorough analysis of the broadcast informational content from 07:30 PM (March 11) to 12:00 AM (March 14): television (Antena 3, Tele 5, and Spanish National Television), radio (Cadena SER), and newspapers (ABC, El Mundo, and El País). As said, the most important media were selected for the study according to *Encuesta General de Medios* (General Media Survey, in English) [Aso04]. The share of the three analyzed television channels exceeded a 75% [LG04]. For radio, we included Cadena Ser, a radio channel which had the most important role for this political event [Olm05]. In addition, the three selected newspapers were the most read during this period of time prior to the elections.

4.4.2 Calibrated parameters for the ABMs for political scenarios

Both ABMs developed for the analysis of the 14-M elections consider a similar set of parameters to be calibrated during automatic calibration. The selected parameters are those related with the control of WOM and mass media diffusion, which are both the most uncertain and the hardest to

[1] This database can be accessed from `http://ugr.mynews.es/hu/`

estimate with the existing information. For each defined parameter, we also use a parameter range to set its possible values during optimization. In the case of the model based on the framing theory, a set of 44 parameters is selected. A brief description of those parameters is as follows:

- WOM diffusion parameters. For each defined segment S, we will calibrate its initial talking probability ($p_S(t)$), influence change (Δ_S), and influence decay ($d\Delta_S$), i.e., 9 parameters.
- Mass media parameters. For each defined mass media channel C, we calibrate its maximum influence (Δ_C^{max}), influence change (Δ'_C), influence decay ($d\Delta'_C$), buzz increment (τ_C), and buzz decay ($d\tau_C$); i.e., 35 parameters.

In the case of the model based on the spatial theory of voting, the automatic calibration procedure only adjusts the 35 parameters of the model that are related to the mass media influence and buzz parameters. This is due to the lack of population segmentation in this model as all the artificial voters are included in the same segment. In addition, the social parameters can be manually set for this model since we can assume a baseline where the attacks do not happen and the distance values of the agents remains stable overtime.

4.4.3 ABM based on the theory of Framing

This section is a summary of the main aspects of the experimentation regarding the ABM based on the theory of Framing. Interested readers are referred to the manuscript in Section 1 in Part II of this dissertation. This experiment considers a memetic algorithm composed of a SSGA and a local search procedure as the automated calibration algorithm (both described in sections 3.1.1 and 3.1.7, respectively). The fitness function measures the distance between the election results and the simulated output by using a symmetric mean absolute percentage error (SMAPE) (see Equation I.14). This function is selected because it increases the sensitivity for miss-voting agents. In this equation, A_t represents actual election results and F_t represent the simulated election results. The mapping ratio was 1:1,212.77. The algorithm considers a population of 100 integer-coded individuals and 10,000 evaluations as stopping criteria. In addition, the algorithmic setup considers 3-tournament selection, a BLX-α crossover ($\alpha = 0.51$, $p_c = 1$), and a uniform random mutation mechanism ($p_m = 0.1$).

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{\frac{(|A_t|+|F_t|)}{2}}. \tag{I.14}$$

4.4.3.1 Calibration results

We show the calibration results using historical voting data as well as different validation scenarios. These validation scenarios are built by removing some components of the ABM to observe its behavior with respect to the historical trends. First, we calibrate the ABM with all the designed components. Due to the lack of empirical data about the social network of voters before the elections, we choose to set the parameter m of the Barabasi-Albert algorithm to 2 in order to generate the scale-free social network. In our case, as we have a population of 24,109 agents (i.e., a social network with 24,109 nodes), this results in a network density of $1.7 \cdot 10^{-5}$ and an average degree $\langle k \rangle = 4$.

The calibration method needs to simulate 30 Monte-Carlo runs for each model configuration. In addition to the complete calibrated model, three validation scenarios are presented: one without mass media, another without WOM diffusion, and the last one with neither mass media nor WOM diffusion. These additional scenarios are variations of the complete model where certain

modules are disabled. By setting these scenarios, the designed model can be validated as a whole, facing its global behavior with respect to removing any of its main modules. The scenario without mass media, called "No Media", disables media effect on the agents. Thus, mass media channels will neither influence agents nor increase buzz activity. In this scenario, only WOM diffusion is performed by agents through their social network. The scenario without WOM diffusion, referred as "No Diffusion", does not include agent diffusion through the social network. In this scenario, only mass media channels influence the agent population but there is not any buzz effect generated from its impact. The last scenario, called "No Influence", does not include neither media effect on the agents nor diffusion through the social network. In this scenario, the agent population is not exposed to any kind of influence, thus the election results are directly those predicted by pre-election opinion surveys.

The comparison between the whole model and the three additional validation scenarios is shown in Table I.3. These results are obtained averaging the results of 30 Monte-Carlo runs of the ABM simulations. Percentage values represent SMAPE accuracy on the final election results, scaling both simulated and real number of votes to the top third. This computation facilitates the understanding of the fitting results in a 0% to 100% scale.

In this Table I.3, votes are displayed by party in the top block of the table and computed error is shown in the bottom block. Fitting results show good accuracy values for the complete model, displaying an accuracy value higher than 99%. This implies that the model is correctly simulating elections turnout. Observed errors also suggest that only WOM or mass media information in isolation are not enough to match final votes, and there is a need to use both components in the model. In fact, the latter two scenarios have a higher number of PP voters than the final results. That suggests that, when used in isolation, defined dynamics are not modeling voting turnout reality after the attacks in an accurate way.

	Real Data	**Models**			
Party	**Election Results**	**Complete**	**No Media**	**No Diffusion**	**No Influence**
Votes					
PSOE	11,026,163	11,020,144	10,329,618	10,259,170	9,941,145
PP	9,763,144	9,766,804	10,577,439	10,948,887	11,403,078
Abstention	8,449,355	8,451,711	8,331,602	8,030,602	7,894,438
% Total votes					
PSOE	37.71%	37.69%	35.33%	35.09%	34%
PP	33.39%	33.4%	36.18%	37.45%	39%
Abstention	28.9%	28.91%	28.5%	27.47%	27%
Global fitting		**99.13%**	**84.49%**	**77.06%**	**68.39%**

Table I.3: Fitting values of the calibrated model and three additional model variation scenarios.

4.4.3.2 Analysis of the model's outputs

This section is devoted to evaluate the model behavior to ensure its validity. In the first place, the evolution of the averaged resilience μ of the population is displayed in Figure 10. In order to compute these values, we average resilience for all the agents of each segment at each time-step. This evolution is stepped by hours, starting on March 11th at 08:00 AM and finishing on March 14th at 08:00 AM.

This chart presents stronger changes when news are on television. It corresponded to the prime time for news in Spain. Additionally, the first simulation steps show more intense changes

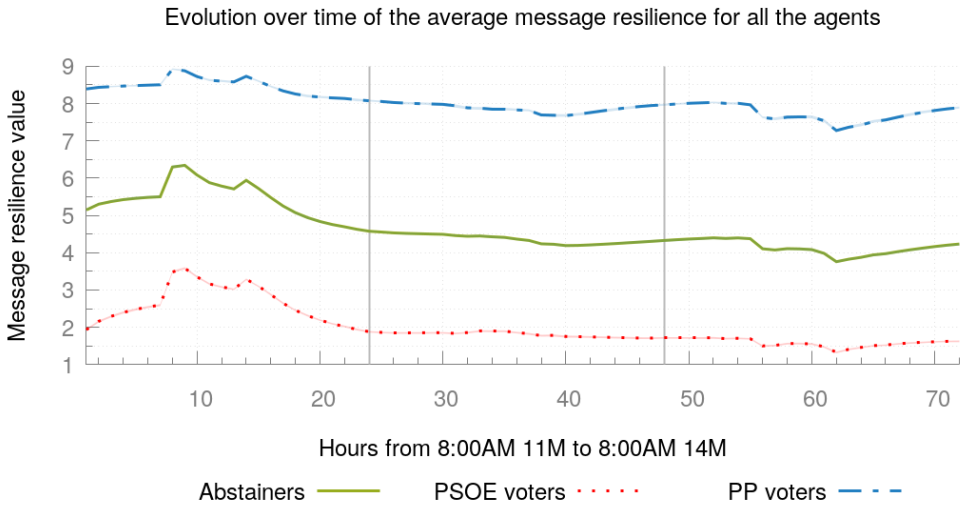


Figure 10: Averaged message resilience (μ) over time for all the agents.

in the perceptions of the agents than the subsequent ones. The main responsible for this behavior is the polarization of the message transmitted by mass media. The content of this message, that was uniform at the start of the simulation, turns mixed at the end, resulting in a smoother curve. Because mass media exposure is not biased by segment, its impact over resilience evolution is similar (averaged resilience curves present a similar shape in the three segments). This evolution is slightly softer for the PP voters segment at the beginning of the simulation. This effect can also be observed at the end of the simulation, but for the opposite direction. PSOE voters change their perception smoother than before.

Figure 11 shows the evolution of votes by day, plotting a track of the voters intention every 24 hours. Votes track is done at 08:00 AM every day. Tracks for the first day, i.e., 11-M, do not have deviation as they correspond to the initial poll data. On the second day, when the message content is still confusing, both PP and PSOE voters reduce their number while abstainers have a significant increase. The reduction is much larger for PSOE voters as ETA's authority is still likely. From that point, the messages are more focused and clear and both PP and abstainers reduce while PSOE voters significantly increase. Overall, box-plots of Figure 11 shows how PP decreases its votes in favor of PSOE and abstention along the simulation. This behavior is consistent with the surveys closer to the elections whose results suggested that the gap between PSOE and PP was reduced as the elections approached [LM06].

Finally, the WOM behavior is mainly validated by using the number of conversations and the evolution of the messages' polarization. Figure 12 shows the percentage of conversations by step (also called WOM volume). In these values we can see that the highest buzz is achieved at prime time, just like in the resilience evolution chat. As happened in the deviation chart, blurred areas represent Monte-Carlo variations. The peaks shown in the number of conversations are heavily related with the buzz increment generated by the mass media. Thus, the increment in the number of conversations is consistent, as the highest audience level is achieved during these time slots, enabling the media to generate its biggest buzz. Figure 13 shows the polarization of the conversations during the simulation. A positive polarization value (above 0) means that conversations are increasing the averaged μ value of the agents (moving it towards 10). Otherwise, a negative polarization suggests that the averaged μ value of the agents decreases (moving it towards 0). Net polarization is heavily influenced by the information transmitted by mass media during the simulation. As a result, this

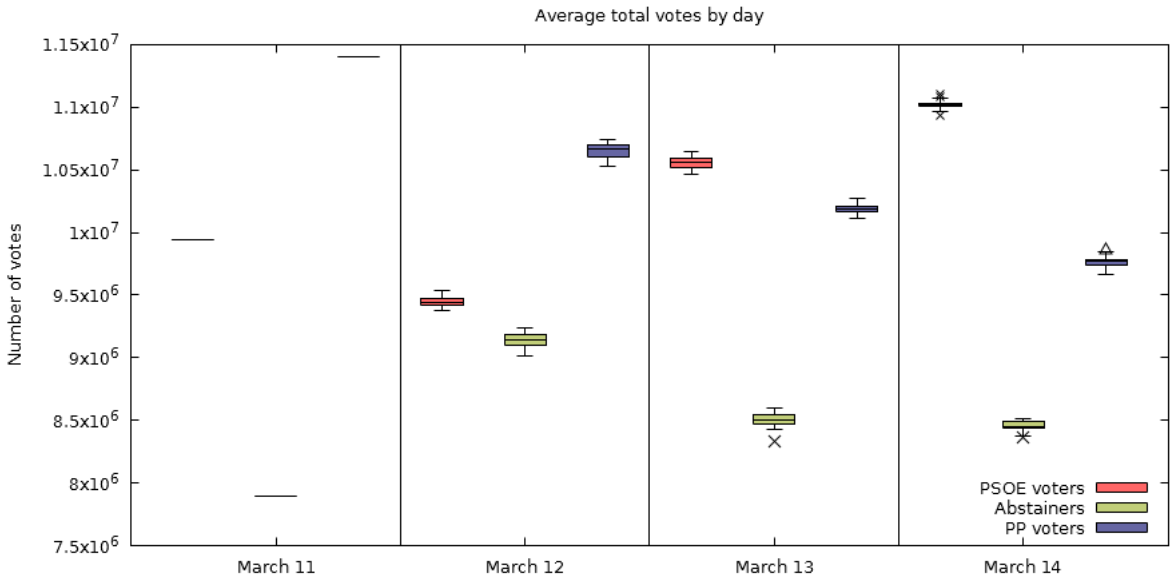


Figure 11: Average votes by political option for every day between the attacks and the elections.

chart looks similar to the one displaying resilience evolution. The trend followed by polarization reaches its minimum value towards the end of the simulation, when the media information contains more messages blaming Al Qaeda. At the end of the simulation the message polarization is negative and therefore, there are more conversations regarding Al Qaeda's involvement in the attacks than conversations regarding ETA's authority. Because Al Qaeda's framing is defended by the PSOE party, this situation increases PSOE votes. This evolution was also observed in Figure 10 where we showed the message resilience over time for all the agents of the population.

4.4.3.3 Deployment of what-if political scenarios

This section analyzes different what-if scenarios using the previously validated model. The study focuses on two scenarios: Section 4.4.3.3.1 analyzes WOM influence in the voter segments and Section 4.4.3.3.2 analyzes changes on mass media messages.

4.4.3.3.1 WOM influence in the voters segments

This scenario is focused on the information spread through the social network. We perform a sensitivity analysis on the parameters which control the diffusion mechanisms. Those parameters are the voter talking probability ($p(t)$) and the parameter to generate the scale-free social network (m, see Section 2.2.1) which affects the social network density and the average degree of the agents, thus affecting the volume of conversations. Both of them are increased when setting higher values for m and consequently the speed of the diffusion process is higher. The sensitivity analysis is performed following the one-factor-at-a-time methodology [tBvVL16] which modifies each parameter in an isolated way by keeping the rest of the parameters fixed to its original value.

In the resulting charts of Figure 14 we show the model response to changes to the talking probability and network connectivity variations as connected points even if they correspond to discrete values obtained from simulation runs for each specific parameter configurations. We consider six different values for the talking probability, from 0.01 to 0.06, with a fixed increment step of 0.01. Another six values for m are studied, from 1 to 6, which correspond to network average

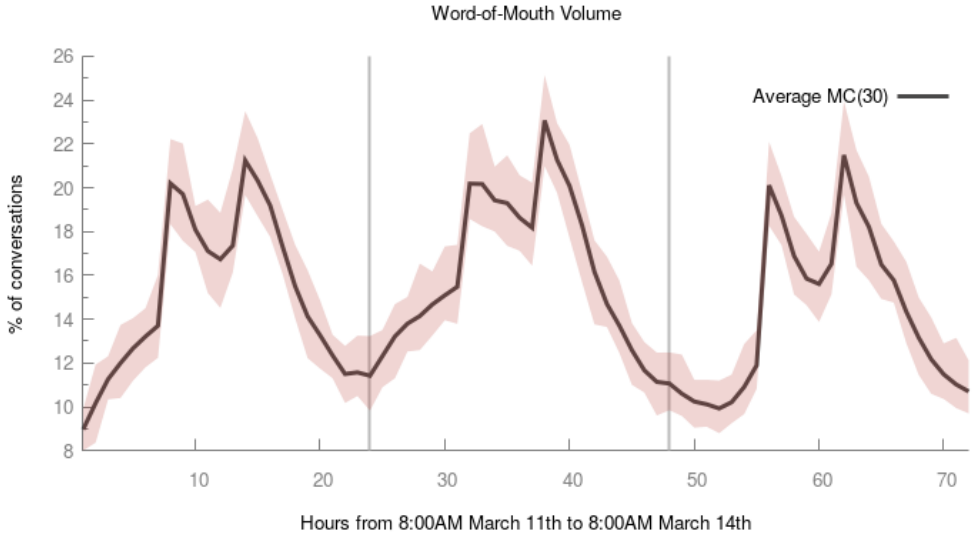


Figure 12: Percentage increase of the conversations made by agents within their social network.

degrees $\langle k \rangle$ of $2, 4, 6, 8, 10$, and 12. This experimentation is performed at both segment and global level. First, we will modify the values of m and the talking probability for each segment (chart lines formed by circles, crosses, and pluses). Later, we will do the same by modifying the parameter values for all the segments at the same time (chart lines formed by boxes).

This analysis shows that the model behavior and its historical fitting are sensible to changes on the WOM parameters. The three existing segments obtain different voting results when their parameters are modified. In the case of PSOE voters, this segment seems to have an important participation in the diffusion process because the number of votes for PSOE party changes significantly when modifying their own diffusion parameters, reaching its maximum and minimum values. In contrast, although the number of PSOE votes changes when altering the diffusion parameters of the other segments, the effect is softer than when modifying the values for the PSOE segment. With respect to abstainers, modifying its parameter values has a relatively small effect on the other segments that could suggest a secondary role in the diffusion process. Finally, in the case of PP voters, these results show a fall in the number of PP votes for most scenarios when diffusion increases. This fact suggests that PP segment cannot influence the other segments even when PP's diffusion parameters have a high value. This could also suggest that message polarization penalizes extreme values when having a highly connected network.

4.4.3.3.2 Content changes on the mass media messages

These scenarios are focused on the polarization of the message transmitted by mass media channels. Using the original polarization as a reference, we perform a sensitivity analysis over the message content transmitted by each mass media channel. Instead of modifying a single parameter, a group of parameters are changed for each scenario [tBvVL16]. For each mass media category (press, radio, and television), its message polarization is modified towards either ETA or Al Qaeda. These polarization variations are applied to all the mass media channels and to every information transmitted by those mass media channels contained in each category.

The interested reader can find the complete results of this study in the published manuscript at Section 1 in Part II of this dissertation. However, in order to keep the analysis brief, we focus on the results in Figure 15. This chart shows the evolution of abstentions with respect to the modified

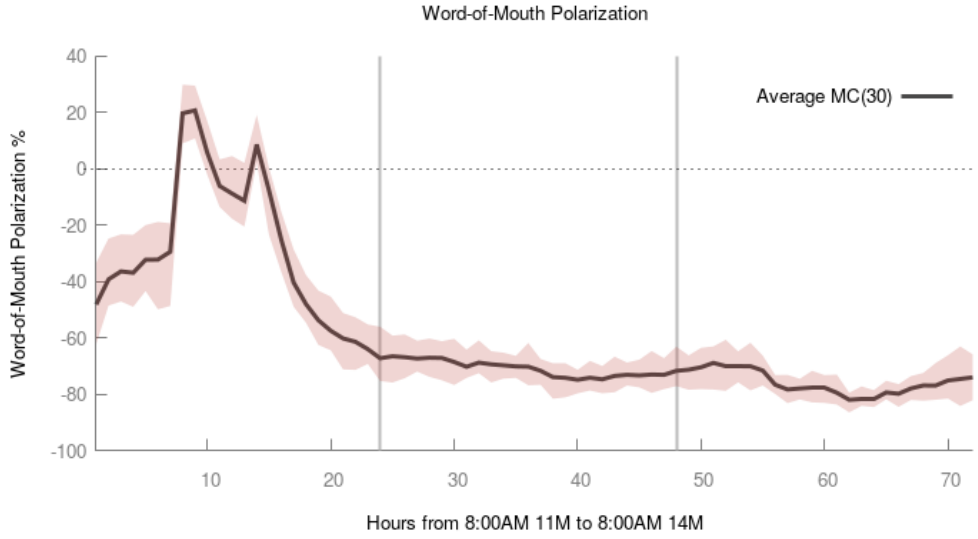


Figure 13: Net variation of the polarization of the message transmitted by the WOM process.

polarization. This way, results obtained with the original message are placed at 0 in the x-axis. This polarization is gradually increased from 0 to 1 and from 0 to -1. Both extremes represent full message content towards Al Qaeda (-1) or ETA (1).

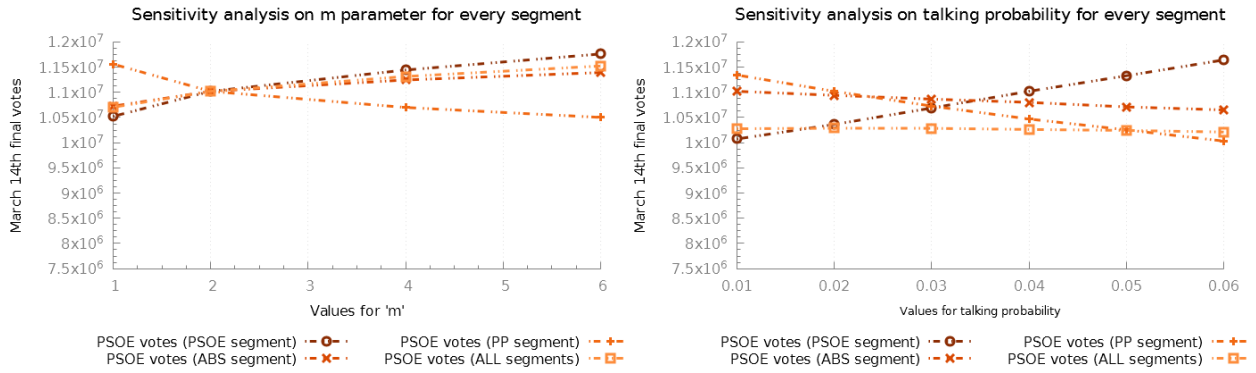
Resulting abstentions are displayed by mass media categories: press, radio, television, and all of them together. These results (and those for the other political options) clearly highlight television as the most influencing media channel. Additionally, the joint effect of all the mass media channels seems interesting because its maximum result surpasses individual categories.

This chart is also interesting because its non-symmetric shape. If message polarization is strongly moved towards Al Qaeda (-1 variation), the simulation results saturate quickly. There is almost no change in the number of abstentions when message is pushed beyond -0.5. On the opposite, when message is pushed towards ETA (+1 variation), the model results saturate around 0.9. In the case of television, its results saturate before the rest of the media categories.

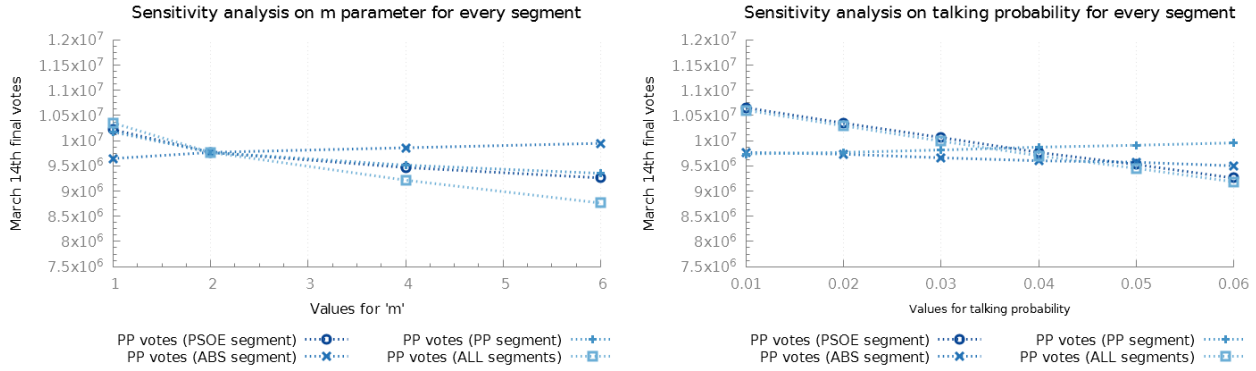
4.4.4 ABM based on the spatial theory of voting

This section is a summary of the main aspects of the experimentation regarding the ABM based on the spatial theory of voting. Interested readers are referred to the manuscript in Section 2, Part II of the dissertation. As introduced in Section 4.4.1, the ideological positions for the voter agents (v_i) and the ideological position where they allocate the considered parties (p_i^k) are taken for the 2553 study of the CIS [Cen04]. It is important to clarify that these surveys do not contain enough information for modeling the ideological distance using the approaches based on intensity, since the post-election survey does not ask how important (intense) the attacks were for the voters' decision. The polarized values ($m_c^k(t)$) of the multiple messages supplied by the different mass media channels are defined using the same information as in the previous model. The considered data sources are depicted in Section 4.4.1.

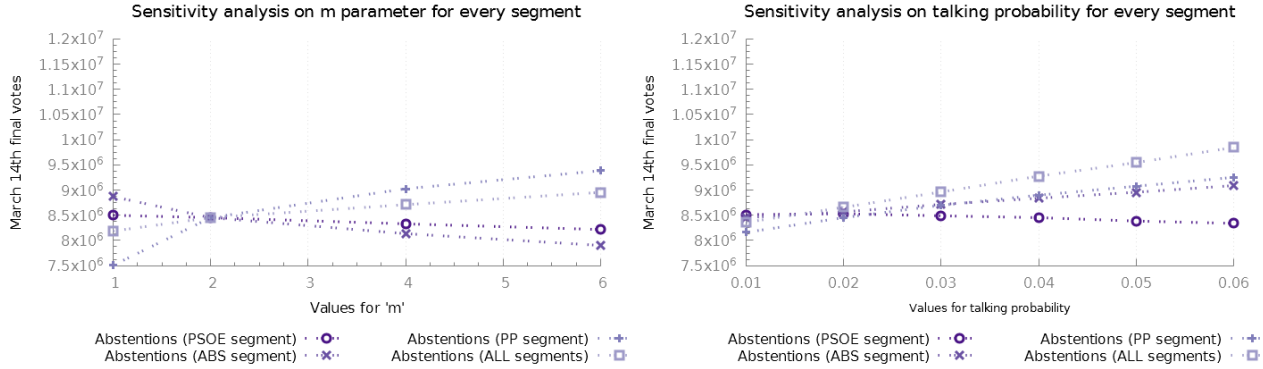
Using these data, the simulation setup considers a set of $N = 21,280$ agents. This population size is set by extending the number of interviews of the pre-elections survey and instancing multiple agents for each of the interviewed voters. In addition, those parameters regulating the WOM interactions are set to $\psi = 3$ and $\phi = 0.5$, respectively. On the one hand, a value of $\psi = 3$ disables WOM interactions when the difference in the distance values of the interacting agents is greater than 3 (i.e., a third of the maximum range of the variable). On the other hand, a value of



(a) PSOE votes



(b) PP votes



(c) Abstentions

Figure 14: Sensitivity analysis for the talking probability and scale-free generation parameter m on each of the voters' segments.

$\phi = 0.5$ halves the resulting influence of two agents that represent voters from different political dispositions, offering a good balance by allowing WOM influences between agents from different groups while penalizing these exchanges. Finally, we run the automatic calibration procedure using a memetic algorithm that comprises a SSGA and local search refinement in a similar way to the previous model. With respect to its configuration, the algorithm considers 3-tournament selection, uniform random mutation ($p_m = 0.1$), and a BLX-α crossover ($p_c = 1$ and $\alpha = 0.51$).

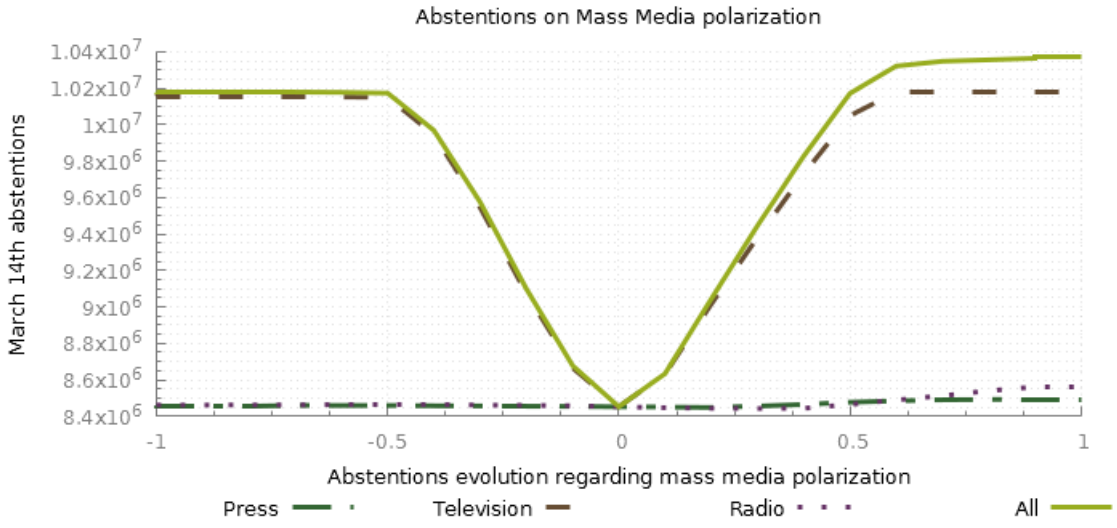


Figure 15: Abstentions resulting from variations on the polarization of mass media messages.

4.4.4.1 Calibration results

As first introduced in Section 4.4.2, the selected automatic calibration procedure adjusts 35 parameters of the model. For each of the selected mass media channels, we calibrate five of its parameters: its maximum influence (Δ_c^{max}), influence change (Δ_c), influence decay ($d\Delta_c$), buzz increase (τ_c), and buzz decay ($d\tau_c$). The fitting function computes the average of the absolute deviation error for each party. This absolute deviation is computed as $|o_j - s_j|$, with o_j being the observed value and s_j being the simulated value. This fitting is computed using the values of the post-electoral survey [Cen04], where the participators were asked to position both themselves and the parties in an ideological scale between 1 and 10. For each participant, the ideological distance is computed as the absolute difference between their position on the ideological continuum and the position of each political party. Therefore, the average values of all the voter distances are used for evaluating the fitness of a given model configuration by comparing them with the simulated distance values at the end of the simulation.

Table I.4 shows the fitting results for the calibrated model. In these values, we can observe that the calibrated model obtains an excellent fitting for IU and PP parties, since their absolute deviation error is equal to or lower than 0.05 for both parties. If this deviation is translated to a percentage error ($100 \cdot |o_j - s_j|/o_j$) with respect to the actual values from the post-electoral survey (2.43 and 3.42, respectively) the deviation is still lower than a 3%. However, the average distance for the PSOE party results harder to fit since the absolute deviation is 0.18. As the target value is the lowest of all parties (1.57), the relative error (around a 12%) is greater than the other parties but we can argue that the calibrated result is acceptable. Additionally, we note that there is not any other model configuration improving the fitting of PSOE distances without reducing the fitting of the other two parties.

4.4.4.2 Model's output and analysis

The main output of the model is the evolution of the distance values for each party during the different steps of the simulation. Figure 16 shows the average of the distance values for the agents of the model by each party at each step of the simulation. These values are computed using

Party	Pre	Post	Simulated	Deviation	Percentage
IU	2.55	2.43	2.39	0.03	1.23
PP	3.22	3.42	3.43	0.005	0.14
PSOE	1.92	1.57	1.77	0.19	12.1

Table I.4: Distance values from the election surveys and simulation results. Additionally, the simulation deviation error is shown both as an absolute value and as a percentage value.

the average of the 30 Monte-Carlo runs with a resulting negligible deviation, hence it is not shown in the charts. The overtime values show that the average distance increases during the beginning of the simulation for the IU and PSOE parties and reaches its maximum value at step 16 (i.e., at 00:00 AM March 12) . After this peak, a moment where a significant change in the information broadcast by the mass media arose as a consequence of the findings by the Spanish police, the values decrease until stabilizing during the final steps of the simulation. We can see how the average values have a similar behavior over time for both parties (i.e., IU and PSOE). In contrast, we can see that the average statistics show an opposite behavior for the PP party as they reduce their values during the first steps of the simulation and increase them in the subsequent steps.

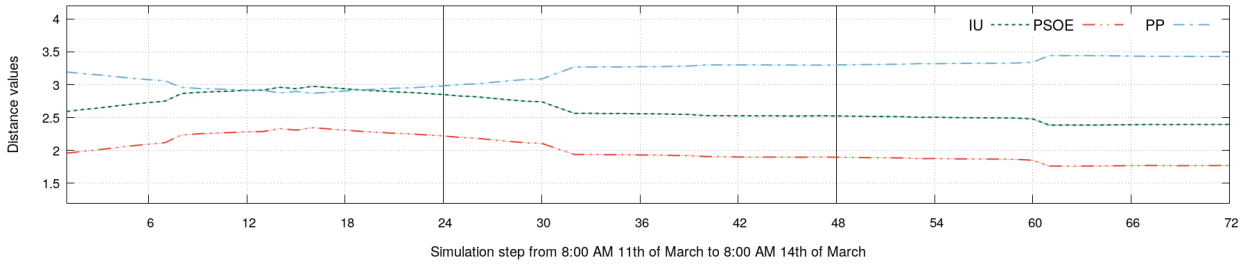


Figure 16: Average ideological distance values of the agents for each party. The displayed values represent the average of the 30 Monte-Carlo repetitions at each step of the simulation. The Monte-Carlo repetitions show a negligible deviation and, hence, it is not shown in the charts.

Additionally, the social behavior of the agents can be evaluated regarding the number of WOM interactions and the effect those interactions. Figure 17 shows the average percentage increase on the number of WOM interactions during the simulation, with the blurred areas representing the Monte-Carlo variability. These values show two peaks for each day of the simulation corresponding with the news in the afternoon and in the evening. Hence the news on prime time had the biggest share for the televisions and caused a high buzz for the following steps.

The effect of the agent social interactions can be approached as the *sentiment* of WOM. It reflects the impact of these interactions on the agent distance values showing if there is a majority of conversations increasing or decreasing the distance values. This sentiment value indicates the trend of the social interactions and hence an unfavorable sentiment means that there are more conversations increasing the distance values than decreasing them. Therefore, a sentiment value of 5 means that there are 5% more unfavorable WOM interactions than favorable ones (i.e., those decreasing the distance values). Figure 18 shows the average sentiment of WOM interactions for each party. We can identify how the peaks in the sentiment values match with the timing of the news, as observed in the WOM volume. In addition, we can see how the sentiment trend shifts during the simulation. At the beginning of the simulation, the sentiment for both IU and PSOE parties is unfavorable while the sentiment of PP is favorable. Then, these trends shift as the mass media polarization changes. This can be also observed in the overall aggregated sentiment behavior.

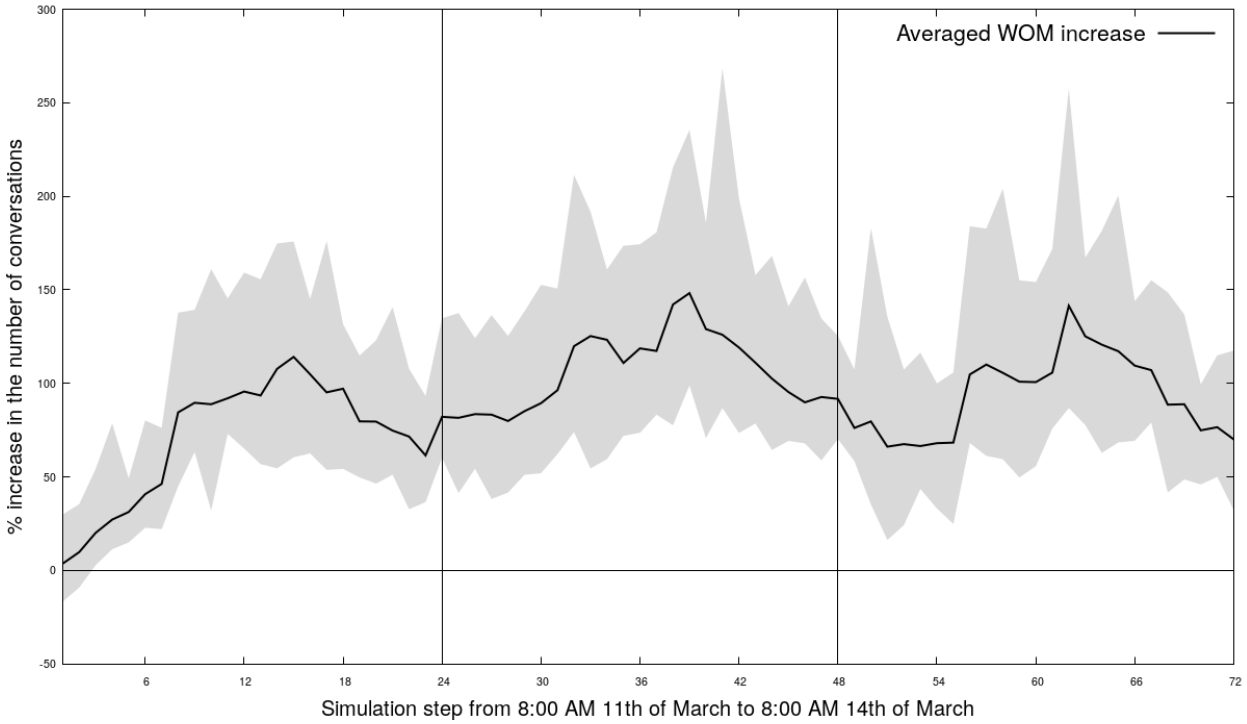


Figure 17: Percentage increase in the number of WOM interactions by step, shown as the average of the multiple Monte-Carlo repetitions. The variability in these repetitions is shown using the blurred area between the maximum and minimum values obtained for each step.

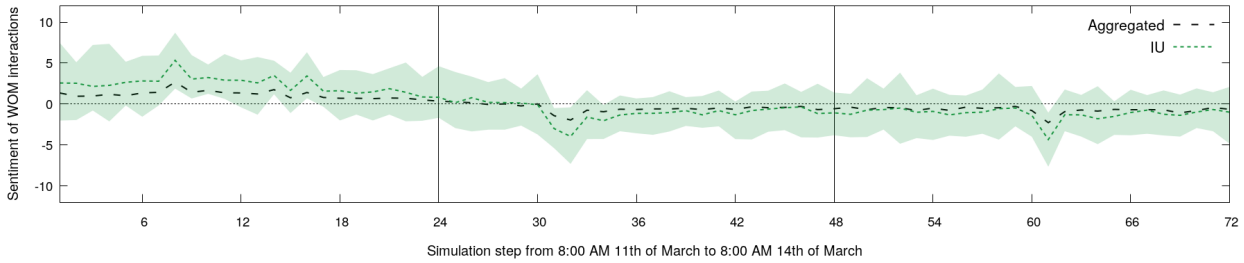
4.4.4.3 Analysis of what-if political scenarios

Once the ABM is calibrated and validated, it can be used to analyze political scenarios. In particular, we believe that the political scenarios from the *theory of terror management* are a good fit to our model due to the circumstances of the 14-M elections. The interested reader is referred to Part II: Section 2 of this dissertation, where the main approaches explaining the mechanisms that operate in the relationship between terrorism and public opinion are addressed in detail. In addition, the current analysis of the what-if scenarios is summarized in order to keep Part I brief. The approaches addressed by these experiments are *the rally around the flag*, *the opinion leadership*, and *the priming of public opinion and media coverage*, which are tackled by the current experiment.

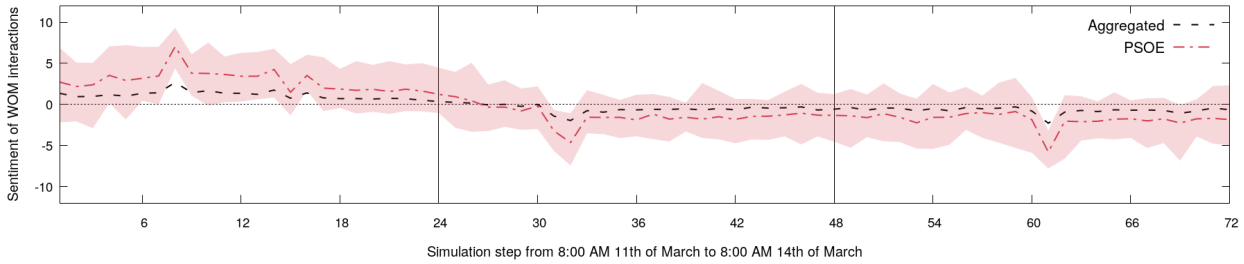
We can see the impact of these scenarios on the resulting ideological distance values of the artificial voters by analyzing the values obtained at the end of the simulation. Hence, Figure 19 displays the absolute variation of the resulting average distance values with respect to the baseline calibrated model for the multiple Monte-Carlo simulation runs using boxplots.

The rally around the flag scenario is simulated by modifying the polarization of the mass media channels to have only those messages that support the government version towards the authority of the attacks. Therefore, the messages that either blame Al Qaeda or which accuse the government of lying are disabled and do not take effect during the simulation. In view of the results in Figure 19, we can recognize that this scenario has the highest impact on the parties since the average distance values for each of them show a deviation of around 0.5.

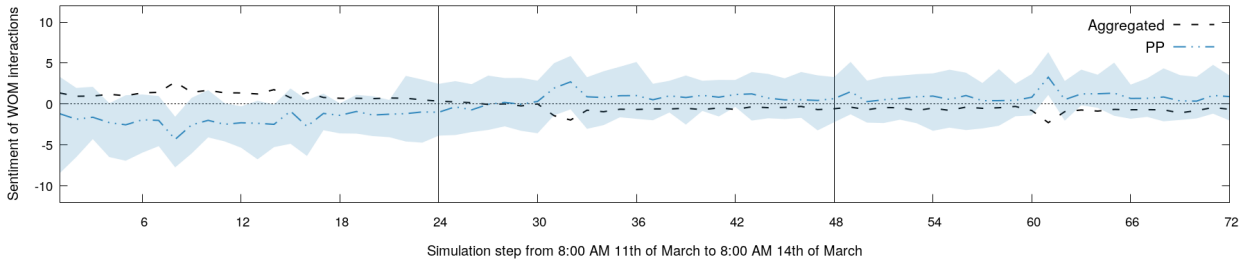
We can simulate *the opinion leadership* scenario by reducing the messages of the mass media channels to those statements of opinion leaders that claim that the government lies regarding the authority of the attacks. This is done by disabling the messages that mention the authority of the attacks by either ETA or Al Qaeda as well as the messages supporting the government. Since



(a) IU



(b) PSOE



(c) PP

Figure 18: Average sentiment of WOM interactions for each party. Blurred areas represent the maximum and minimum values between the multiple Monte-Carlo repetitions. In addition, average overall sentiment resulting from aggregating the parties' values is also included. A dotted line at 0 represents a neutral sentiment, separating the favorable and unfavorable sentiment areas.

the remaining messages do not compose a sample with enough size for reproducing *the opinion leadership* effect, we propose to analyze the hypothetical scenario by inserting additional messages of opposing leaders where they criticize the government. This allows us to analyze *the opinion leadership* scenario according to the assumptions of the *theory of terror management*. Thus, we include several messages from this category in the TV channels during their afternoon and evening news. By including these new messages during the TV channels prime time their effects should be observed clearly, since these are the mass media with the highest audience.

As seen in the variation of the distance values with respect to the baseline simulation, this scenario produces a similar effect in all the parties, reducing the distance of IU and PSOE and increasing the distance of the PP party by the same amount. This shows how a systematic appearance of the political leaders criticizing the government during the 2004 Spanish elections would have had a strong impact in the ideological distance of the voters and highlights the role of the leaders of opposing parties.

The simulation of *the priming of public opinion and media coverage* scenario is designed by focusing the broadcast information in the messages pointing out Al Qaeda's authority of the

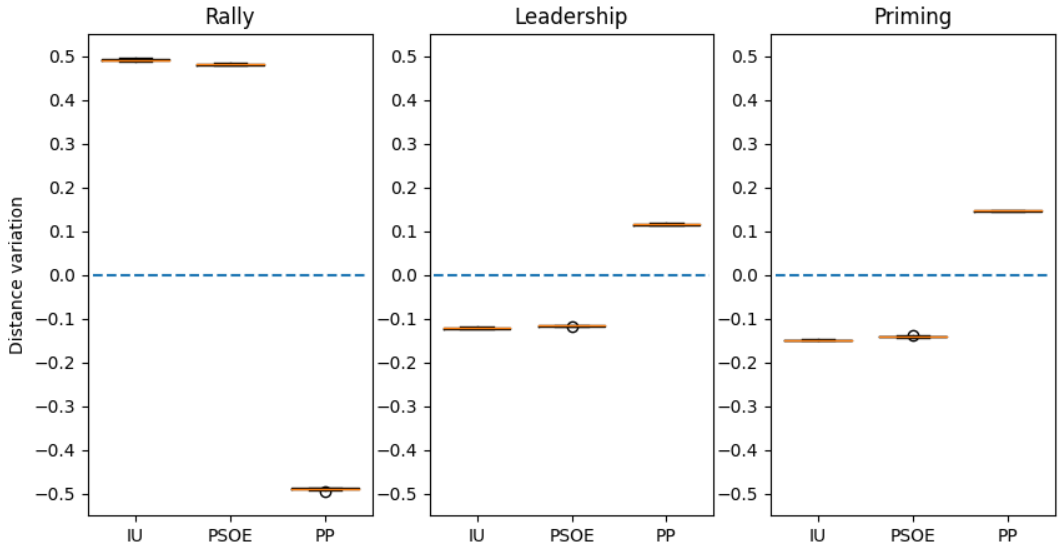


Figure 19: Boxplots showing the distance variation at the end of the simulation for each defined what-if political scenario with respect to the baseline calibrated model. The reduced size of the boxes indicates the robustness of the results across the 30 Monte-Carlo runs.

attacks. Thus, the messages supporting that ETA is responsible for the attacks and those either blaming or support the Government are disabled. This results in a deviation of around a 0.15 for every party in their distance values (see Figure 19).

4.5 Experiments: calibration of ABMs for marketing strategies

This section reviews the different experiments developed using the proposed ABM for marketing strategies. The purpose of these experiments is two-fold. On the one hand, it focuses on evaluating the performance of different metaheuristics for automatic calibration. On the other hand, we proposed how the calibration and validation of ABMs can be improved by using an advanced visualization method. Thus, the rest of the section goes as follows. Section 4.5.1 reviews the different instances of the ABM for marketing developed for assessing the calibration methods. Then, Section 4.5.2 briefly introduces the parameters of the ABM that act as decision variables during the automatic calibration procedures and the fitting functions used by the metaheuristics. Section 4.5.3 addresses the comparative of the metaheuristics considered for single KPI calibration. Section 4.5.4 evaluates the performance of the different EMO algorithms selected for the joint automatic calibration of two KPIs. Finally, Section 4.5.5 introduces our integral approach to ABM calibration using EMO algorithms and network-based visualization.

4.5.1 A benchmark for ABM calibration

We consider a benchmark with 15 instances of the model corresponding to different market configurations with an increasing number of mass media channels. These instances are the result of synthetically generating 14 additional instances from an initial real-world, baseline instance, referred as P1(25). Notice that, this original instance corresponds to a market with seven channels, thus resulting in 25 parameters to be calibrated, the number enclosed in brackets (Section 4.5.2 includes details regarding the relationship between the number of channels and the number of parameters). The additional instances are generated applying variations on the initial baseline

instance. Each model variation incorporates new mass media channels that are generated from the existing ones by perturbing their investment values. The new instances also include modifications on the target data for the fitting of both outputs (i.e., WOM values and awareness values). Each new instance increases the dimensionality of the previous one as the parameters of the new channels are added as new decision variables, enabling a deeper analysis of the algorithm performance.

On the one hand, the perturbations on the existing mass media channels C consist of multiplying the investment of each brand at each time step by a given factor. We consider reductions in the original investment by 15%, 30%, 45%, and 60%. In addition, we increase the original investment by 100%, 200%, 300%, and 400%. The decision of whether increasing or decreasing a brand investment is made at random and remains constant for each step. On the other hand, the modifications on the target historical values for both objectives involve directly adding or subtracting a different value for each brand to each of its time steps. In order to avoid unrealistic values we truncate the resulting awareness values to a maximum of 100% and a minimum of 1%. Each addition/subtraction on the awareness target values will be by 2%, 5%, 8%, or 10%. In the case of WOM volume each addition/subtraction will be by 1,000, 2,000, 4,000, or 6,000 conversations, with a minimum value of 0. Similarly to mass media investment, the decision of whether increasing or decreasing the target values is made at random and remains constant for each step.

The new generated instances are labeled according to their dimensionality: P2(40), P3(46), P4(55), P5(61), P6(70), P7(76), P8(85), P9(91), P10(100), P11(115), P12(130), P13(145), P14(160), and P15(175). The parameter configuration of baseline P1 considers $|I| = 1000$ agents, $|B| = 8$ brands, $|C| = 7$ media channels, and $T = 52$ time steps (a year of simulation with each time step representing a single week). Awareness is initialized to $a(0) = (0.7, 0.75, 0.58, 0.25, 0.08, 0.42, 0.39, 0.34)$ and mass media channels consider $r_c = (0.92, 0.57, 0.54, 0.03, 0.43, 0.38, 0.69)$. The generated instances share this baseline configuration, including the reach parameters values r_c of the new channels, that take the value of the channel used for its generation. For example, if a new channel 9 is generated from the original channel 5, the reach value of the former channel is set to the value of the latter (i.e., $r_9 = r_5$).

4.5.2 Parameters selected for calibration and fitting functions

For this model, we select the parameters that either modify the agent awareness values or their number of conversations for the automated calibration process since they are the most uncertain and the hardest to estimate. These parameters regulate the awareness and talking probability gained by mass media and social interactions with the addition of the awareness deactivation probability (d) and the social network generation parameter (m). Notice that, the density of the agent social network is a parameter that is always difficult to identify. This way, each of the selected model calibration parameters corresponds with one decision variable in the coding scheme of the metaheuristic algorithm. The final set of parameters to be calibrated for each model instance is determined by the size of the modeled scenario: three parameters for each mass media channel plus four fixed social parameters. Briefly, those parameters are the following:

- Social network parameters. The following parameters are calibrated: the initial talking probability ($p_i^b(0)$), social awareness impact (α^{WOM}), awareness deactivation probability (d), and social network generation parameter (m).
- Mass media parameters. For each defined mass media channel $c \in C$, we calibrate its awareness impact (α_c), buzz increment (τ_c), and buzz decay ($d\tau_c$).

Therefore, the number of calibrated parameters is $3 \cdot |C| + 4$. Figure 20 shows the coding scheme for a model instance using three mass media channels, that is, composed of 13 genes.

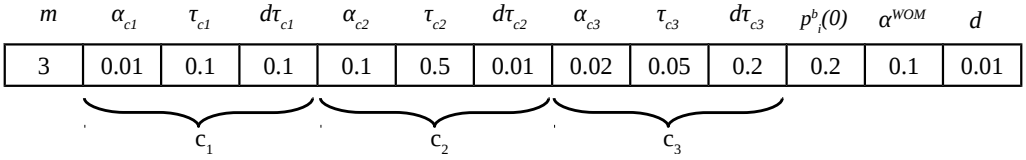


Figure 20: Example coding scheme for a model instance with three mass media channels. The first gene (m parameter) is an integer value bounded to $\{2, ..., 8\}$. The rest of the genes in the chromosome represent real-coded parameters that are defined in $[0, 1]$.

With respect to the fitting functions for this ABM, Equations I.15 and I.16 define the objective fitting functions for the two KPIs, f_1 (awareness deviation error) and f_2 (WOM volume deviation error), respectively. These functions compute the deviation error between the provided series of target data and the model outputs for each objective using the standard mean absolute percentage error (MAPE) function, where $\widetilde{a}$ and $\widetilde{\omega}$ represent the ground-truth target values of awareness and WOM volume of the whole population, respectively. The simulated values are generated using Monte-Carlo simulations by computing the average of those independent runs. Due to the time-consuming task of simulating multiple times for every parameter configuration, we limit the number of Monte-Carlo runs to 15 (instead of 30 as in the political ABMs), which anyway is a reliable number of runs for the experimentation.

$$f_1 = \frac{100}{T \cdot |B|} \sum_{b=1}^{|B|} \sum_{t=1}^{T} \left| \frac{a^b(t) - \widetilde{a}^b(t)}{\widetilde{a}^b(t)} \right|, \tag{I.15}$$

$$f_2 = \frac{100}{T \cdot |B|} \sum_{b=1}^{|B|} \sum_{t=1}^{T} \left| \frac{\omega^b(t) - \widetilde{\omega}^b(t)}{\widetilde{\omega}^b(t)} \right|. \tag{I.16}$$

4.5.3 Comparative study of metaheuristics for single-objective ABM calibration

Two baseline methods and five evolutionary algorithms from those depicted at Section 3.1 are considered for benchmarking their performance when calibrating our ABM for marketing. The first baseline method follows a random search approach (RND). This procedure simply creates valid solutions by randomly generating values for each of the decision variables. This way, random solutions are generated until the stopping criteria is met and the method returns the best solution found. The second baseline method is a local search procedure implementing a Hill Climbing (HC) strategy, similar to the one employed by the memetic algorithms. However, instead of starting for a given solution from the population, the baseline HC starts from a random individual generated using a uniform distribution. The HC search continues until the stopping criteria is met, which involves a much higher number of steps than that considered for the memetic variants. The considered metaheuristics were DE, L-SHADE, IPOP-CMA-ES, CRO, CRO-SL, and the corresponding memetic variants for the algorithms.

4.5.3.1 Experimental setup

Each metaheuristic runs 20 independent times using different random seeds. Every algorithm execution considers 10,000 evaluations as stopping criteria. During each evaluation of a candidate solution, the objective function f is computed using the output of each independent run with the provided historical data considering the parameter $\beta \in [0, 1]$: $f = \beta \cdot f_1 + (1 - \beta) \cdot f_2$. The considered value for the weighted combination is $\beta = 0.5$ (i.e. the same weight is considered for both KPIs).

With respect to the coding scheme, we have chosen an integer-coded scheme. This approach is selected because representing the values of real parameters following an integer-coded scheme allows the modeler to set the desired granularity for the parameter values [CBK+17]. Thus, since the parameters of this calibration problem are real values limited to $[0, 1]$, the integer-coded approach transforms the parameter values using $\min_i = 0$, $\max_i = 1$, and $\epsilon_i = 0.001$, for every parameter i, resulting in 1,001 possible values. The parameter setup for the metaheuristics was specified via a preliminary experimentation and the final values are as follows:

- HC movements increase or decrease the integer value of the genes[2] by one unit. Each HC run takes 200 iterations and applies 50 movements during each iteration.
- DE considers a population size of 100 individuals. The parameter value for the crossover rate is set to CR= 0.9 and the value of the mutation rate is set to F= 0.5.
- L-SHADE uses an initial population size of 100 individuals and a external archive size of 200. During the reproduction phase, the p value for the current-to-pbest/1/bin strategy is set to $p = 0.1$. Finally, the size of the historical memory is set to the dimensionality of the problem.
- IPOP-CMA-ES considers a population size of $\lambda = 15$ with $\mu = 6$ and an increasing factor of 2. In addition, the learning rates are set to $c_\sigma = 0.568$, $c_c = 0.6962$, and $c_{\text{cov}} = 0.4897$. Since the calibration problem transforms the search space from the interval $[0, 1]$ into $\{0, \cdots, 1000\}$, a relatively high $\sigma^{(0)}$ value of $\sigma^{(0)} = 56.747$ is selected. Finally, the dampening for the step size update is set to $d_\sigma = 4.2939$.
- CRO uses a reef size of 50 individuals. Regarding the BLX-α crossover, the probability is set to $p_c = 0.2$ using $\alpha = 0.25$ and tournament selection of size 3. The mutation probability of the random mutation is set to $p_m = 0.1$. The rest of CRO parameter values are $k = 3$, $\rho_0 = 0.6$, $F_a = 0.05$, $F_b = 0.9$, $F_d = 0.08$, and $P_d = 0.15$.
- CRO-SL defines a reef populated by 50 individuals. Its CRO-based values are set to $k = 3$, $\rho_0 = 0.6$, $F_a = 0.05$, $F_b = 0.9$, $F_d = 0.05$, and $P_d = 0.15$. The substrate layers of CRO-SL integrate uniform random mutation, random walk mutation, SBX, and BLX-α, and its parameter values are $p_m = 0.2$, $p_c = 1$, and $\alpha = 0.51$.
- Regarding the memetic variants, the refinement probability is set to $p_{\text{LS}} = 0.0625$ since it is the recommended value for similar problems [LHKM04]. Each time an individual is refined using the local search procedure it will use 50 evaluations. The rest of the parameters of the memetic algorithms are shared with its corresponding global search algorithm (i.e., MCRO has the same setup as CRO).

4.5.3.2 Calibration results

This section contains a brief summary of the full analysis of the different metaheuristics performance, depicted in detail in the published manuscript at Part II: Section 3. The analysis splits into two different stages: a preliminary one only considering the baseline and the evolutionary algorithms and an advanced one which includes the designed memetic approaches. Table I.5 shows the average ranking of the metaheuristics for the preliminary stage. This table also considers several post-hoc procedures for highlighting significant differences in the metaheuristics performance. Specifically, Friedman's non-parametric test [Fri40], Bonferroni-Dunn's test [Dun61], and Holm's test [Hol79] are used. With respect to the Friedman's test, the result of applying the test is

[2] Although local optimizers do not refer to the optimization variables in the solution as genes, we have decided to keep that nomenclature to make a clear difference with the model parameters we are calibrating.

	Rank	**Bonferroni** p	**Holm** p
CRO-SL	1.5	–	–
CRO	2.25	1	0.52
IPOP-CMA-ES	2.42	1	0.52
L-SHADE	4.33	0.003	0.001
DE	4.5	0.001	0.001
RND	6.08	$1.4 \cdot 10^{-7}$	$1.2 \cdot 10^{-7}$
HC	6.92	$2.6 \cdot 10^{-10}$	$2.6 \cdot 10^{-10}$

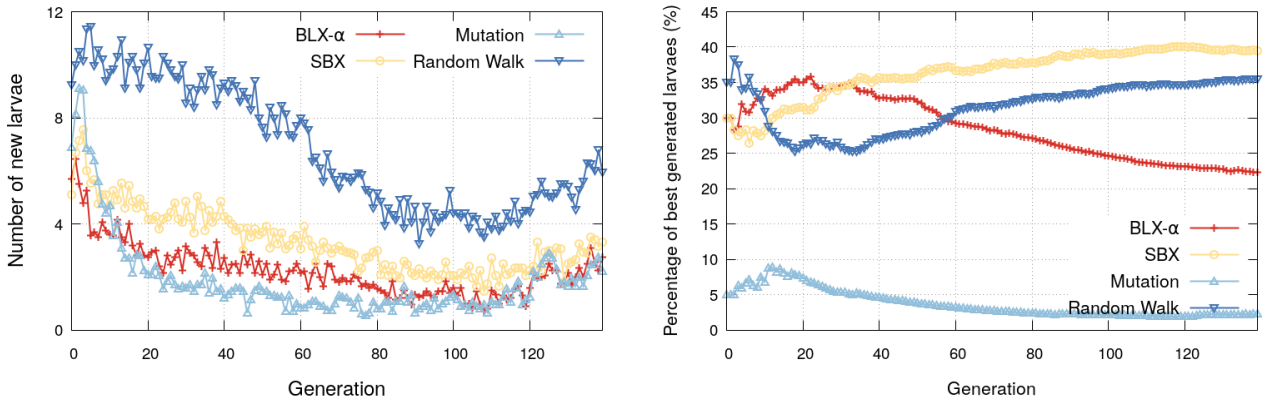
Table I.5: Average ranking of the metaheuristics and their resulting $p-$values for Bonferroni's and Holm's test using CRO-SL as the control method.

$\chi_F^2 = 64.35$ and its corresponding $p-$value is $5.8 \cdot 10^{-12}$. As the $p-$value is lower than the desired level of significance ($\alpha = 0.01$), the test concludes that there are significant differences in the algorithms performance.

These results show how CRO-SL outperforms the other algorithms as it ranks first achieving the lowest mean rank (1.5) and finds the configuration with lowest fitness value for almost every model instance. After CRO-SL, CRO and IPOP-CMA-ES rank second and third with close ranking values (2.25 and 2.42, respectively). L-SHADE and DE rank forth and fifth but perform significantly worse than the control method CRO-SL, since the $p-$values from both Bronferroni-Dunn's and Holm's tests (0.003 and 0.001 in the case of L-SHADE, 0.001 and 0.001 in the case of DE) are lower than the considered significance level ($\alpha = 0.05$). The baseline methods HC and RND also perform significantly worse than the control method and rank last for every model instance. In addition, the performance of CRO-SL can be analyzed in terms of the behavior of the different search procedures when solving a single problem instance. Figure 21 shows the number of settling larvae (i.e., new solutions in the population) obtained by each substrate at each generation and the percentage of times each substrate produces the best solution of the generation for P12 instance.

These plots show that SBX and random walk are the best performing substrates since they end up generating 40% and 35% of the best quality solutions, respectively. Random walk also stands out as the substrate generating the higher number of solutions which replace other solutions in the reef due to their quality. It can also be observed that every substrate is productive during the first generations when there is space in the reef available for new solutions. A higher number of solutions during the first generations are produced by the mutation substrates, in a period when exploration of the search space is crucial. BLX-α decays after obtaining some of the best solutions during the first generations and, later on, SBX stands out by delivering good quality solutions. Thus, both exploration and exploitation are maintained by the different operators and balanced until the convergence of the algorithm.

Table I.6 presents the corresponding average ranking after including the memetic variants together with the results of the original metaheuristics. In addition, this stage omits the less performing methods from the earlier stage, since they were already outperformed. Table I.6 includes the resulting $p-$values of Bonferroni-Dunn's and Holm's tests, this time using MCRO as the control method. In this case, the Friedman's test results in $\chi_F^2 = 62.94$ and a $p-$value of $3.8 \cdot 10^{-11}$, which once again concludes that there are significant differences between the algorithms performance. The results of the memetic variants reveal that not every hybridized algorithm improves the performance of its original counterpart since MCRO and MIPOP improve the original algorithms but ML-SHADE and MCRO-SL do not. MCRO ranks first and achieves the lowest average ranking (1.5), improving CRO in every instance and ranking first in eight instances. MIPOP obtains a small



(a) Larvae generated by each substrate

(b) Percentage of best generated larvae (%)

Figure 21: (a) Number of new larvae settling in the reef per substrate at each generation. (b) Percentage of best larvae obtained by a substrate at each generation. These values are computed as the average of the multiple Monte-Carlo runs of CRO-SL for the P12 instance.

	Rank	**Bonferroni** p	**Holm** p
MCRO	1.5	–	–
CRO-SL	2.5	1	0.287
MIPOP	4	0.054	0.015
CRO	4.25	0.024	0.013
IPOP	4.25	0.024	0.013
MCRO-SL	4.58	0.007	0.005
L-SHADE	6.92	$5.9 \cdot 10^{-8}$	$5 \cdot 10^{-8}$
ML-SHADE	8	$3.3 \cdot 10^{-11}$	$3.3 \cdot 10^{-11}$

Table I.6: Average ranking of the metaheuristics including the memetic algorithms. The resulting $p-$values corresponds to Bonferroni's and Holm's tests using MCRO as the control method.

improvement compared to the original IPOP-CMA-ES performance for most instances but ranks third with an average ranking of 4, only ranking first in P3.

CRO-SL is ranked second in average (2.5) and ranks first for the remaining three model instances. In addition, CRO-SL still finds the best solution for most model instances. In contrast, MCRO-SL ranks sixth (4.58) and is significantly outperformed by MCRO. Since CRO-SL shows a good balance between exploration and exploitation (note that it obtains the most of the best solutions), the addition of the local search component can negatively modify this balance by reducing its diversity and causing the algorithm to converge prematurely. With the inclusion of the memetic variants, CRO obtains the same rank as IPOP-CMA-ES and both algorithms are tied with an average rank of 4.25 in the forth/fifth position. However, these algorithms are significantly outperformed by MCRO. L-SHADE and its memetic counterpart ML-SHADE are the worst ranked methods.

The impact of hybridization on the performance of these algorithms can be visually corroborated by displaying their fitness values for the instance with the highest dimensionality, P12. Figure 22 shows the multiple runs of the algorithms for this instance using points: the X axis refers to the fitness values and the Y axis marks the different algorithm runs. This plot shows how MCRO obtains the highest improvement with respect to its non-memetic counterpart. MIPOP

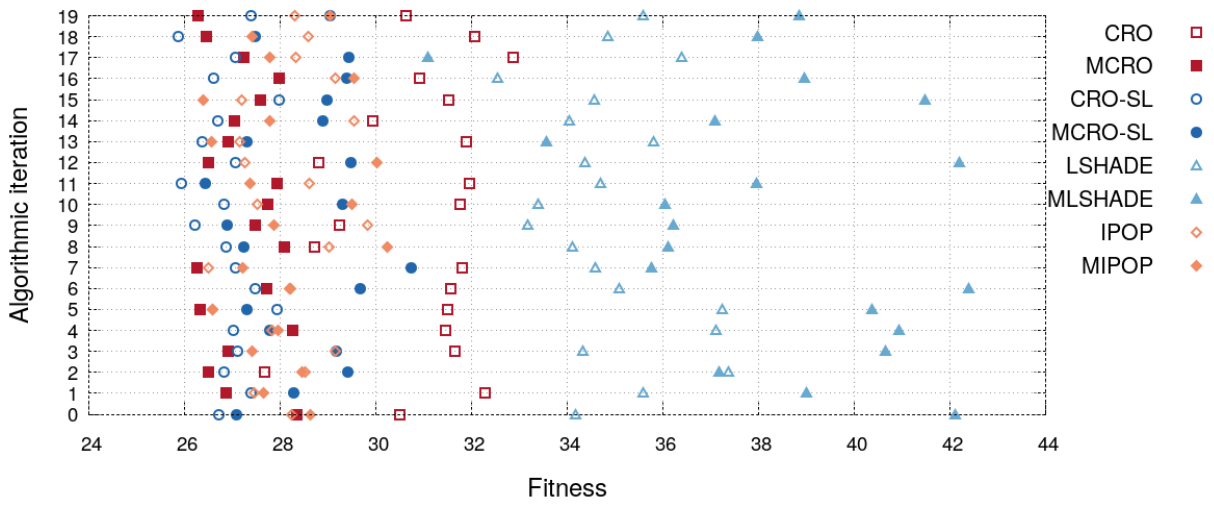


Figure 22: Fitness results of the different runs of the eight best performing algorithms for the problem instance P12. The values in the Y axis refer to the different runs and the values in the X axis shows the fitness error for the problem instance.

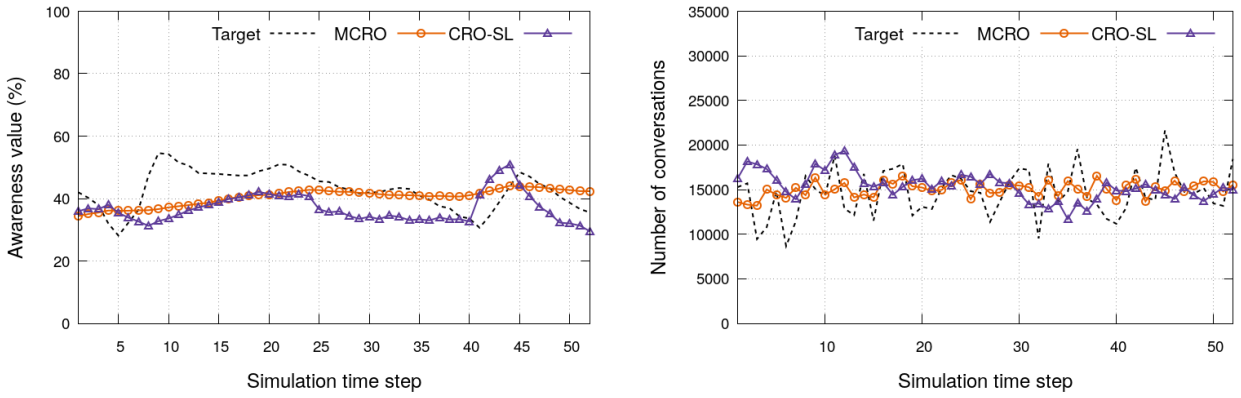
slightly improves IPOP in many of the runs but with a reduced margin and not in every run. In addition, the loss of performance of MCRO-SL and MLSHADE is also noticeable.

However, there are still some limitations to our approach to automatic calibration that need to be addressed. One of these limitations can be identified by visually analyzing the output fitting of the calibrated solutions. Figure 23 shows the fitting for both awareness and WOM volume of the best calibrated solutions of MCRO and CRO-SL (i.e., those with the lowest fitness) for P12. Only the KPI values of brands 6 and 8 are shown for boosting the clarity of the figure. The represented values are the average of the multiple Monte-Carlo simulations. These values show that even when having solutions with good fitness values, the calibrated solutions are only able to capture the trend of the target values for both brands. This effect is a consequence of the objective fitting function f, which relies in the numeric deviation error computed by the loss measure but ignores the pattern characteristics of the historical values. In addition, we also acknowledge that these synthetic instances could be harder to calibrate than others based on real data as they were generated in a random fashion.

4.5.4 Comparative study of EMO algorithms for two-objective ABM calibration

The present experiment considers the EMO algorithms collected at Section 3.2 when calibrating the instances of the benchmark using both KPIs. These algorithms, which are NSGA-II, SPEA2, IBEA, SMS-EMOA, MOMBI2, MOEA/D and GWASF-GA, share the following characteristics across the experiment. In addition, we have included a classical mathematical optimization method in our experiments, the Nelder-Mead simplex method [NM65]. This classical method allows us to benchmark the performance obtained by the different EMO algorithms when compared with traditional approaches. Each candidate solution has n decision variables corresponding to the model parameters being calibrated, which can either be integer-coded or real-coded values. In contrast with the previous study depicted by Section 4.5.3, in this particular case we choose to focus on a real-coded scheme. This is motivated by the multicriteria nature of the problem, as we aim to obtain the most accurate approximation possible to the Pareto set for every model instance.

The considered algorithms include polynomial mutation [Deb01] as their mutation strategy. It modifies the values of a solution variable with a probability $p_m \in [0, 1]$ using a polynomial



(a) Awareness output for brand 8

(b) WOM volume output for brand 6

Figure 23: Overtime awareness fitting for brand 8 (a) and overtime number of conversations (WOM volume) fitting for brand 6 (b) for P12. These values are obtained using the average overtime values of the multiple Monte-Carlo simulations. The lines with circles represent the best solution obtained by MCRO and the lines with purple triangles represent the best solution obtained by CRO-SL. The dashed lines represent the target historical values.

distribution. This mutation strategy uses a distribution index parameter that regulates the strength of the mutation. Unless stated otherwise in their description at Section 3.2, the proposed algorithms use simulated binary crossover (SBX) [Deb01] with a crossover probability $p_c \in [0, 1]$ as their crossover strategy. SBX emulates the operation of a single-point crossover from binary-encoding when performing crossover into real-coding decision variables. SBX operates as follows: given two parents $P_1 = (p_{11}, ..., p_{1n})$ and $P_2 = (p_{21}, ..., p_{2n})$, SBX generates two offsprings $C_1 = (c_{11}, ..., c_{1n})$ and $C_2 = (c_{21}, ..., c_{2n})$ as $c_{1i} = \bar{X} - \bar{\beta}/2 \cdot (p_{2i} - p_{1i})$ and $c_{2i} = \bar{X} + \bar{\beta}/2 \cdot (p_{2i} - p_{1i})$, where $\bar{X} = 1/2 \cdot (p_{1i} + p_{2i})$. $\bar{\beta}$ is a random value fetched from a random distribution initialized by setting a distribution index that acts as the spread factor of the operation.

4.5.4.1 Experimental setup

Each EMO algorithm is run 30 times using different seeds to account for the probabilistic nature of the calibration algorithms considered. Every algorithm considers a population of 100 individuals ($P = 100$) and evolves for 100 generations with a stopping criteria of 10,000 evaluations. The distribution index of the mutation operator is set to 10 and the mutation probability value is set to $p_m = 1/n$ where n is the number of parameters being calibrated for the model instance (i.e., decision variables). The SBX crossover operator considers a crossover probability of $p_c = 1.0$ and sets its distribution index value to 5. In addition, the EMO algorithms designed to use a set of weights, such as MOEA/D, MOMBI2, and GWASF-GA, initialize their values by generating a uniform set of 100 vectors, a usual setup when dealing with two objectives and only 100 individuals. In addition, MOMBI2 is set to $\epsilon = 0.001$ and $\alpha = 0.5$. Finally, MOEA/D uses a neighborhood size of 20 and its DE operator considers $CR = 0.5$ and $F = 0.5$.

Finally, in order to adapt the Nelder-Mead algorithm to our multiobjective problem, we employ the adaptive ϵ-constraint method [Eic08, Mav09], which allows single-objective optimization methods to deal with multiple objectives. The Nelder-Mead's approach also involves starting from different solutions for obtaining a Pareto set approximation. Therefore, each run generates 50 random solutions that are optimized until reaching 200 evaluations. This setup slightly modifies the one employed by the EMO algorithms because the Nelder-Mead simplex method requires a model evaluation for every single modification in the decision variables.

4.5.4.2 Analysis of the EMO algorithms performance

This section briefly reviews the performance of the selected EMO algorithms for this experiment, which is a summary of the full experimentation on the published manuscript, addressed in Section 4 in Part II of this dissertation. The performance of EMO algorithms is evaluated using extended unary and binary multiobjective performance indicators (introduced at Section 3.2.4). Table I.7 shows the computed values of HVR for the resulting Pareto front approximations of each algorithm for every model instance. These values are presented using the average of the individual HVR values computed for the individual Pareto front approximations resulting in each of the 30 algorithm executions. The average HVR values show that MOEA/D consistently achieves better values than the other algorithms for most model instances, obtaining the best average HVR in all but four instances. MOEA/D also obtains values close to the best ones for these four instances. For example, in P9 and P12 the best HVR values are obtained by MOMBI2 (0.863 and 0.824), closely followed by MOEA/D (0.857 and 0.812). In addition, MOEA/D obtains the best average value across the 15 instances with the second lowest standard deviation. These results also highlight the poor performance of the Nelder-Mead simplex method when calibrating our problem instances. It obtains the lowest HVR value for every problem instance with a very significant difference with respect to the EMO approaches (an average value of 0.321 while the worst performing EMO algorithm is over 0.83).

	HVR																
EMO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	**Avg.**	σ
MOEA/D	**0.946**	0.956	**0.917**	**0.885**	**0.906**	**0.936**	**0.912**	**0.93**	0.857	0.892	**0.902**	0.812	**0.794**	**0.895**	**0.884**	**0.895**	0.045
SPEA2	0.904	0.969	0.817	0.851	0.859	0.914	0.878	0.803	0.757	0.901	0.825	0.748	0.688	0.794	0.803	0.834	0.073
SMS-EMOA	0.901	0.969	0.795	0.857	0.85	0.93	0.87	0.879	0.836	0.908	0.863	0.802	0.779	0.868	0.843	0.863	0.051
IBEA	0.859	0.95	0.758	0.832	0.778	0.919	0.851	0.876	0.851	0.878	0.863	0.809	0.737	0.84	0.845	0.843	0.056
NSGA-II	0.902	**0.972**	0.821	0.856	0.865	0.926	0.888	0.821	0.757	**0.909**	0.839	0.764	0.746	0.822	0.822	0.847	0.065
GWASFGA	0.881	0.959	0.769	0.816	0.808	0.899	0.825	0.853	0.827	0.863	0.861	0.742	0.69	0.855	0.841	0.833	0.065
MOMBI2	0.873	0.921	0.799	0.846	0.831	0.894	0.838	0.857	**0.863**	0.843	0.878	**0.824**	0.758	0.839	0.857	0.848	**0.039**
Nelder-Mead	0.161	0.169	0.074	0.177	0.17	0.284	0.334	0.353	0.608	0.549	0.383	0.501	0.288	0.348	0.411	0.321	0.155

Table I.7: Average HVR values for every algorithm and model instance. The best value for each model instance is shown in bold font. Additionally, the average HVR values across the multiple instances is shown along with the standard deviation (σ).

The values for the multiplicative I_ϵ indicator were computed as the average of each possible $I_\epsilon(P, Q)$, with P and Q being any pair of Pareto front approximations of different algorithms, resulting from any of the 30 independent executions (i.e., a pair-wise comparison of every run). The interested reader is referred to the tables in the published manuscript, but these results can be summarized by just stating that MOEA/D outperforms the remaining algorithms for most model instances. Moreover, it obtains a lower average I_ϵ value for every comparison with the other algorithms, with the exception of P2, where GWASF-GA obtains a better indicator value. In order to statistically check these results, we applied a statistical test and study the significance of the I_ϵ values to avoid that isolated results could bias our former analysis. We perform this test following the methodology proposed by [SV08, CÁCD13a], described at Section 3.2.4. The interested reader is referred again to Part II: Section 4, where boxplots show the resulting ϵ dominance percentage values, which contains the computed $P_A(B)$ for every pair of EMO algorithms. Briefly, these charts show how MOEA/D generally obtains bigger dominance percentages than the remaining algorithms, since their boxes and whiskers cover a considerable percentage of the interval, implying a big dominance probability.

Table I.8 shows the significance for the resulting $p-$values of the Wilcoxon ranksum test (null hypothesis $E(P_A(B)) = E(P_B(A))$, alternate hypothesis $E(P_A(B)) > E(P_B(A))$) considering a significance level of 0.05. These results are again consistent with the previous indicator values,

as MOEA/D shows an outstanding and robust behavior, being able to perform significantly better than the remaining algorithms in most instances. Hence, MOEA/D is the best performing decomposition-based algorithm for our problem, since it almost always outperforms GWASF-GA. Regarding the performance of the remaining methods, we can see how the Pareto dominance-based EMO variants (NSGA-II and SPEA2) outperform most of the algorithms for the first seven instances. However, if we compare these two algorithms, we can observe that SPEA2 does not significantly outperform NSGA-II in any instance, suggesting that NSGA-II is the best algorithm from this family when dealing with the ABM calibration problem. SMS-EMOA would be the best performing indicator-based EMO algorithm but we can find some instances like P12 where it is outperformed by MOMBI2.

Nevertheless, it is clear that the behavior of MOEA/D is eroded when dealing with specific instances like P6 or some of the bigger instances like P9, P12, and P15. Although MOEA/D obtains the best HVR values for some of these instances, the statistical tests revealed that it is dominated by other algorithms. In the case of P6, SMS-EMOA and NSGA-II are the best performing methods and significantly outperform MOEA/D. As we pointed out, MOMBI2 arises as the best performing one for the P12 model instance. The p-values for the P9 instance, where SMS-EMOA and IBEA perform significantly better than the remainder, with SMS-EMOA finally outperforming IBEA. SMS-EMOA shows better convergence for this instance, which explains the better dominance $P_B(A)$ values corroborated by the statistical test.

4.5.4.3 Influence of the instance properties on the algorithm performance

In view of the results obtained by both the unary and binary indicators, we can observe how specific properties of the problem instances are affecting the performance of the evaluated EMO algorithms. These properties are the shape of the feasible region, the shape of the Pareto front, and the dimensionality of the problem instance. Some studies [ISMN17, MMNI18] have pointed out a relationship between the performance of decomposition-based EMO algorithms (such as MOEA/D, the best performing algorithm in our study) and the shapes of both the feasible search space region and the Pareto front. Figure 24 shows an approximation to the search space landscape of the problem instances using scatter-plots. The shape of the feasible region is approximated by sampling 100,000 random configurations for each problem instance. In the plots in Figure 24, we can observe that the search space extent is considerably bigger for P2, P6, P7, and P10 when compared with the rest of the problem instances. Therefore, the shape of the feasible region for these instances can explain their difficulty, specially for the performance of the EMO algorithms that employ reference points [MMNI18].

In addition, Figure 25 displays the shape of the Pareto fronts. We approximate the shape of the global Pareto fronts by using the aggregated Pareto front approximations obtained for every EMO, which contain the overall non-dominated solutions found for each problem instance. In the plot in Figure 25, we can observe that P2, P6, and P10 have a long tail shape compared with the rest of instances. These long tail shapes can explain a performance reduction for the algorithms using weight vectors because these shapes are non-symmetric and mismatch a distribution of uniformly generated weight vectors [ISMN17]. Hence, the problem instances with these properties may require a customized set of weight vectors for improving its accuracy.

We can also observe how the specific properties of instances P2, P6, P7, and P10 produce different effects on the behavior of the EMO algorithms, namely:

- In the case of P2, it was observed how most EMO algorithms obtain HVR values over 0.95 and compete similarly, since MOMBI2 is the only EMO that results dominated by the rest of EMO algorithms. Thus, the long tail shape of the P2 instance is not sufficient for eroding

MOEA/D	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
SPEA2	-	-	+	-	+	-	+	+	-	-	-	+	+	+	+
SMS-EMOA	-	-	+	+	+	-	+	-	-	-	-	-	+	-	-
IBEA	+	-	+	+	+	-	+	+	-	-	+	-	+	+	+
NSGA-II	-	-	+	-	+	-	+	+	+	-	+	+	+	+	+
GWASF-GA	+	-	+	+	+	+	+	+	-	-	+	+	+	+	+
MOMBI2	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+

SPEA2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SMS-EMOA	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
IBEA	+	-	+	+	+	-	+	-	-	+	-	-	-	-	-
NSGA-II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GWASF-GA	+	-	+	+	+	+	+	-	+	-	-	+	-	-	-
MOMBI2	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-

SMS-EMOA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+
SPEA2	-	-	-	-	-	+	-	-	+	+	+	+	+	+	+
IBEA	+	-	+	+	+	+	+	+	-	+	+	-	+	+	+
NSGA-II	-	-	-	-	-	+	-	+	+	-	+	+	-	+	+
GWASF-GA	+	-	+	+	+	+	+	+	+	-	+	+	-	+	+
MOMBI2	+	+	-	-	-	+	+	+	+	-	-	-	-	+	+

IBEA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
SPEA2	-	-	-	-	-	-	-	-	+	-	-	+	+	+	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NSGA-II	-	-	-	-	-	-	-	-	+	-	-	+	-	+	+
GWASF-GA	-	-	-	+	-	-	+	-	+	-	-	+	-	-	-
MOMBI2	-	+	-	-	-	-	+	+	-	-	-	-	-	+	-

NSGA-II	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
SPEA2	-	-	+	+	-	+	+	-	-	+	+	+	+	+	+
SMS-EMOA	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-
IBEA	+	-	+	+	+	+	+	+	-	+	-	-	+	-	-
GWASF-GA	+	-	+	+	+	+	+	+	+	-	-	+	-	-	-
MOMBI2	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-

GWASF-GA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SPEA2	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IBEA	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
NSGA-II	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
MOMBI2	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-

MOMBI2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
SPEA2	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
IBEA	-	-	+	+	+	-	-	-	-	-	+	+	+	-	+
NSGA-II	-	-	-	-	-	-	-	-	+	-	-	+	+	-	+
GWASF-GA	-	-	+	+	+	-	+	-	+	-	+	+	+	-	+

Table I.8: Significance of the results with respect of the computed $p-$values for every EMO algorithm against the other methods for every model instance. Cells marked as + represent statistical significance.

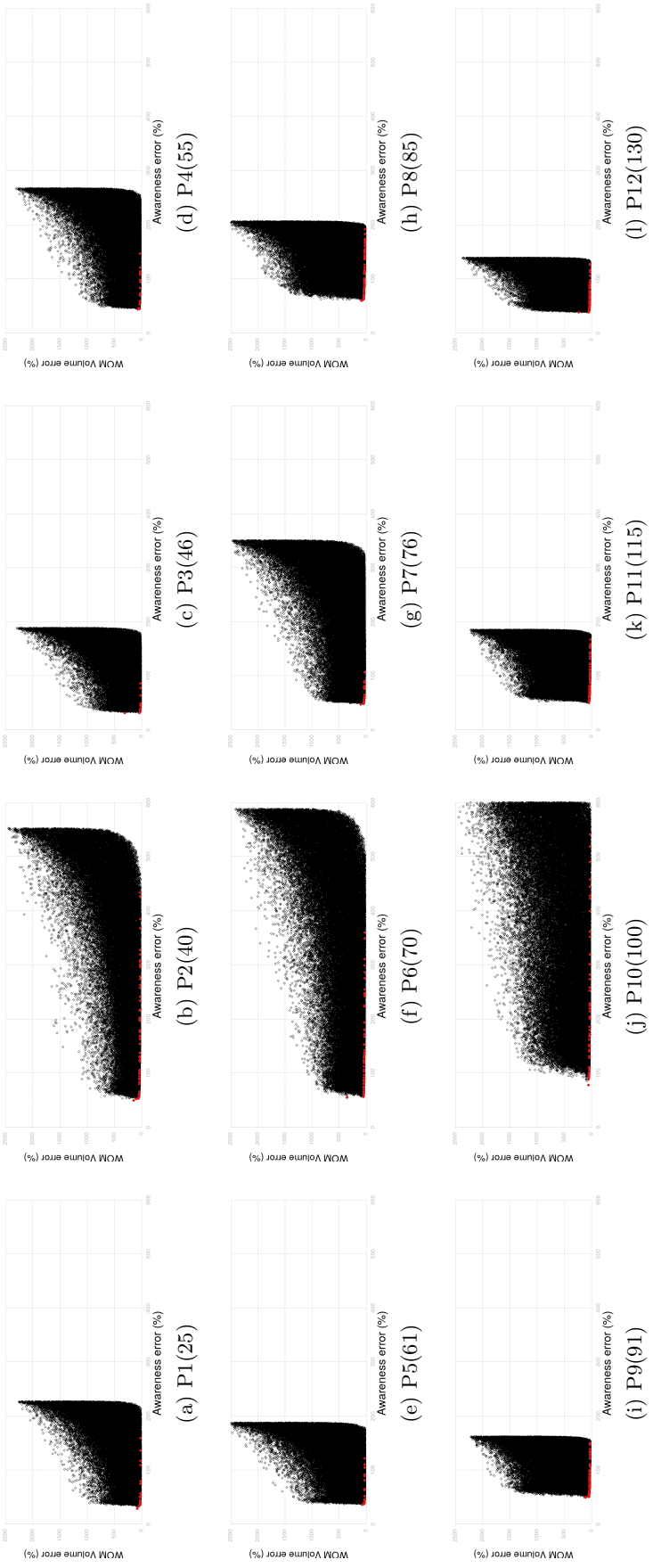


Figure 24: Sampled solutions for the different instances using 100,000 random configurations. The non-dominated solutions are coloured in red. The axis of the charts have been fixed for comparing the different shapes of the feasible region for each problem instance.

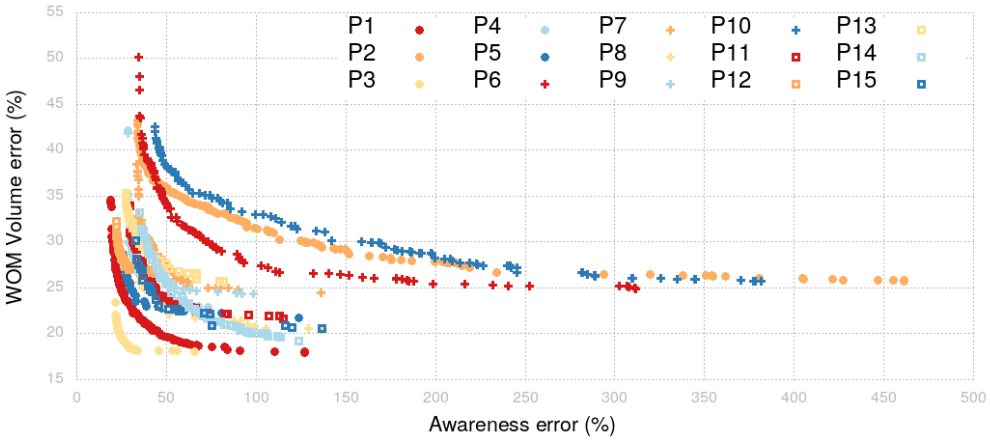


Figure 25: Approximation to the shape of the Pareto fronts using the aggregated Pareto front approximations for each problem instance. Each front contains the overall non-dominated solutions obtained for each problem instance.

the performance of the selected EMO algorithms but this could be explained by the reduced number of variables considered by this instance (only 40).

- With respect to the P6 instance, we have seen that MOEA/D obtains the best HVR values, closely followed by SMS-EMOA and NSGA-II. Despite that, the results of the statistical test pointed out that MOEA/D is dominated by SMS-EMOA and NSGA-II. A deeper analysis of the Pareto sets obtained by MOEA/D's in its individual runs reveals that for some of these runs MOEA/D performed poorly. This lack of consistency solving P6 explains why it is dominated by SMS-EMOA and NSGA-II although it obtains a better average HVR value across the 30 runs.

- It can be observed that the shape of the feasible region for P7 is not so long-tailed as other instances but it is still remarkable. However, the results for this instance are comparable with those obtained for other regular instances, as MOEA/D is clearly the best performing algorithm (obtains the best HVR values with some margin and significantly outperforms the remaining EMOs). This suggests that the shape of the feasible region for P7 is not wide enough for eroding the behavior of MOEA/D.

- In the case of P10, NSGA-II obtains the best HVR value, closely followed by SMS-EMOA and SPEA2. However, the results of the statistical tests for P10 showed that no EMO algorithm is able to significantly dominate more than two of the remaining algorithms. Similarly to the P2 instance, it could be argued that most EMO algorithms are performing similarly but in this case the HVR values are sensibly lower for P10 than for P2. Because MOEA/D is the best performing algorithm for most of the instances in this study, we can argue that its behavior is more influenced by P10 properties than NSGA-II and SMS-EMOA.

4.5.5 An integral approach to ABMs calibration and validation

In this section we briefly describe our proposal of an integral multicriteria framework for model calibration using EMO algorithms and network-based visualization, which is depicted in detail in the published manuscript at Section 5 in Part II of this dissertation. A diagram illustrating the components of our framework is shown in Figure 26. Our experimentation considers two different model instances of the ABM for marketing strategies. The first instance is the initial *baseline* instance P1(25), described in Section 4.5.1. In addition, we consider an instance with increased

dimensionality, P4(55), which was synthetically generated. P4 includes 10 additional mass media channels and increases the dimensionality of the baseline real-world instance by adding 30 new decision variables, enabling a more complete analysis of the algorithm performance.

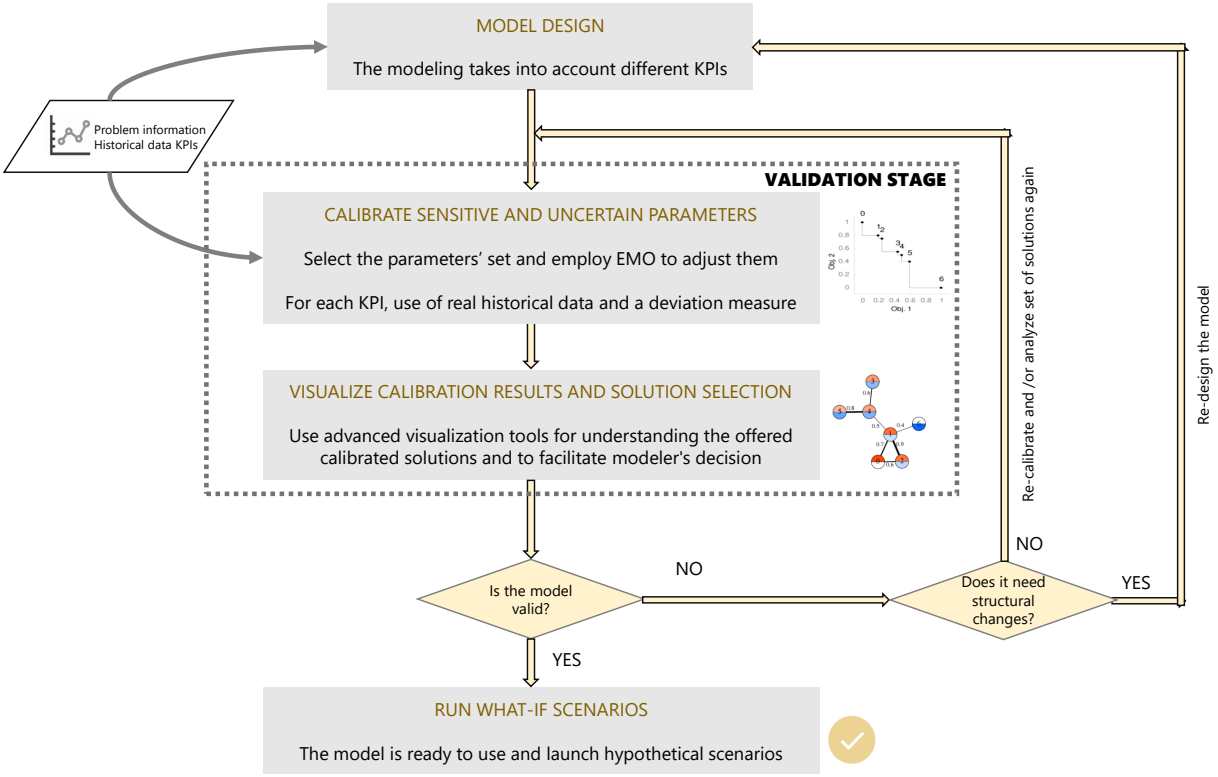


Figure 26: Diagram illustrating the components and the flow of our multicriteria integral framework for model calibration.

NSGA-II is the selected EMO algorithm for these experiments. Its setup considers SBX, which is applied with probability $p_c = 1.0$, and a polynomial mutation operator applied with probability $p_m = 1/n$, where n is the number of decision variables (i.e., parameters of the model to be calibrated). The mutation operator uses a distribution index value of 10 and a different mutation probability value depending on the number of parameters of each model instance. These experiments consider 20 runs of NSGA-II for each scenario, using different seeds for each run. The NSGA-II has a population of 100 individuals ($|P| = 100$) and evolves during 100 generations using 10,000 evaluations as stopping criteria. In addition, the first 39 steps of the historical data (i.e., 75% of total) are used for training, leaving the remaining values as hold-out.

4.5.5.1 First stage: results of the EMO algorithm

The results of the EMO algorithm are analyzed by visualizing the solutions of the generated Pareto set approximations with respect to the two conflicting objectives for P1 and P4 (see Figure 27 with the two Pareto front approximations). We have selected three of the most representative solutions for the two instances: a) the solution with lowest awareness error, b) the solution with lowest WOM volume error, and c) the solution with the best trade-off for both objectives. In order to select the best trade-off solution we use the procedure followed in [CACD13b]. We generate 1,000 random weights $w \in [0, 1]$ and compute the average value

of the aggregation function of both objectives f_1 and f_2. Since the values of f_1 are much bigger than those of f_2, we apply a normalization factor δ in order to scale them $\delta = \frac{1}{|S|}\sum_{i=1}^{|S|}\frac{f_2(s_i)}{f_1(s_i)}$, where S is the set of solutions in the Pareto set approximation. We formulate this process as $\overline{F}(s_i) = \frac{1}{1000}\sum_{j=1}^{1000}\delta \cdot w_j \cdot f_1(s_i) + (1 - w_j) \cdot f_2(s_i)$. The selected solutions are highlighted in their respective Pareto front approximations in Figure 27.

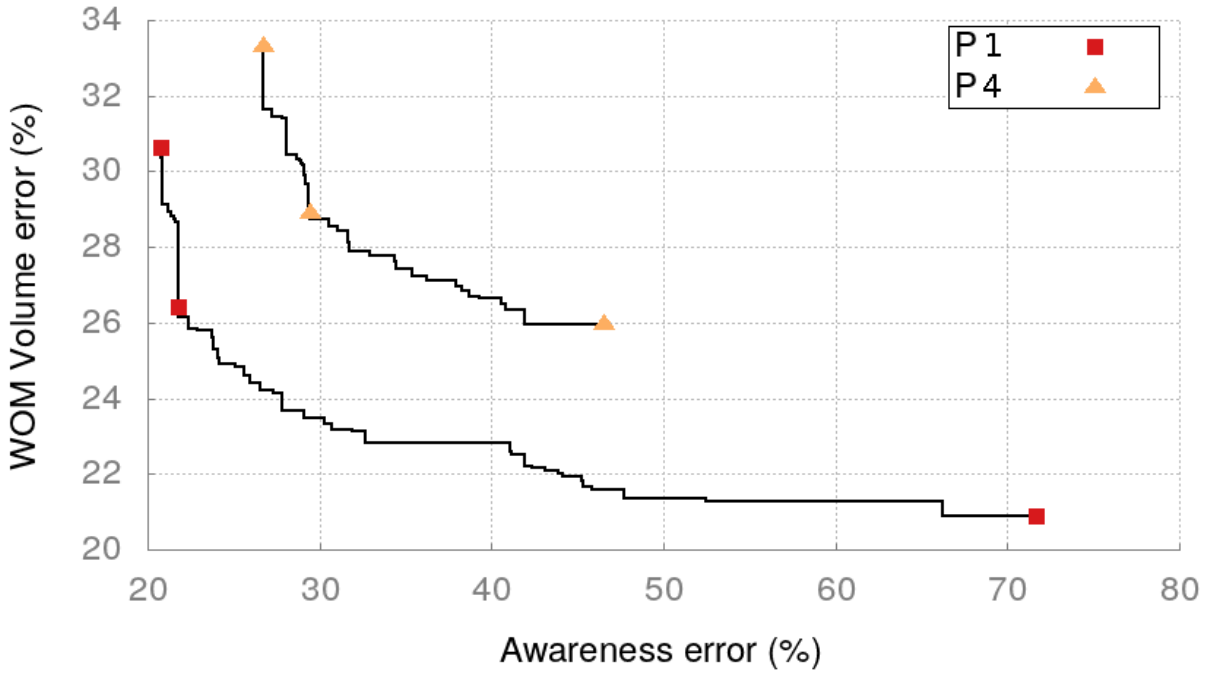
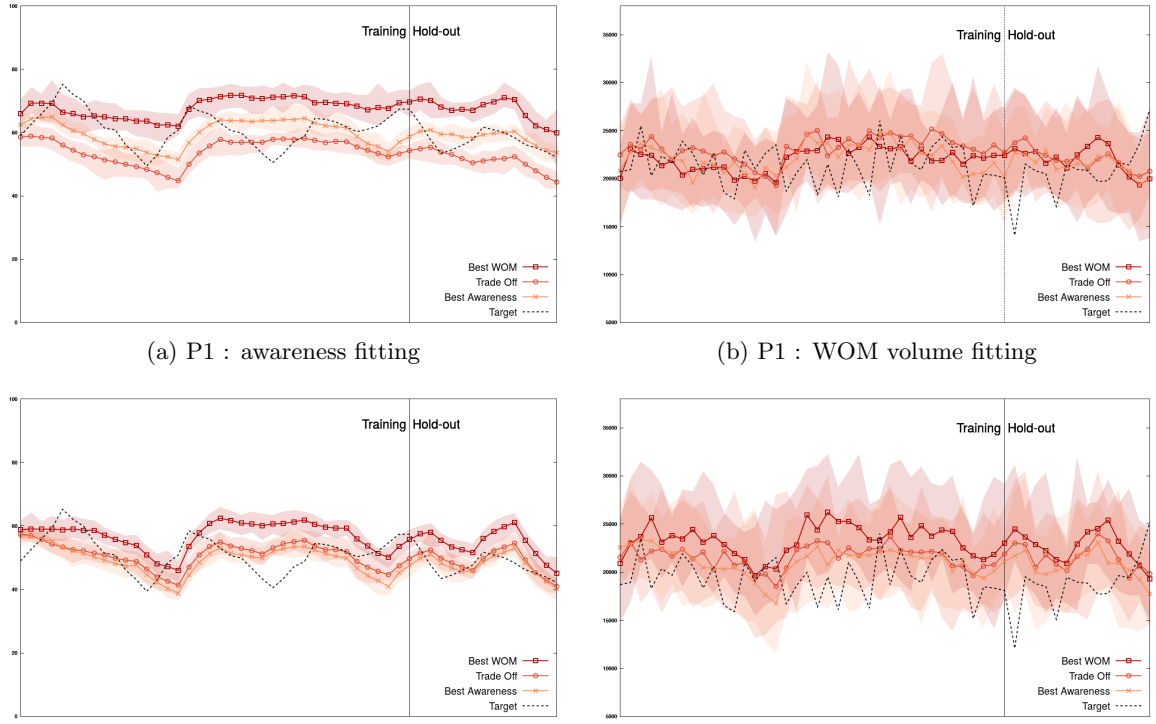


Figure 27: Selected solutions for the two model instances (P1 and P4): lowest awareness error, lowest WOM volume error, and best trade-off. Solutions from P1 are represented using squares and solutions from P4 are displayed using triangles.

We visualize the outputs of the model using the calibrated configurations setting the focus on some specific brands in order to carry out an understandable analysis of the behavior of the selected solutions. The KPI evolution along the simulation steps for brands 3 and 6 is shown in Figures 28 and 29 respectively. These brands were chosen because their behavior is a good resemblance of the rest of the brands for both objectives. These charts show the model output for both the training and the hold-out sets, where we can note than the latter obtain similar fitting than the former and therefore the model can generalize this behavior. Both figures show that adjusting the behavior the dynamics of the awareness evolution over time is harder than the WOM volume dynamics. In contrast, as already identified in the controlled experiment, WOM volume dynamics are more sensible to Monte-Carlo variability for both model instances (as seen in the blurred areas in Figures 28b, 28d, 29b, and 29d).

This is specially relevant for the P1 instance, as shown in Figures 28a and 29a, since the best calibrated solutions only capture the trend of the target values. In the case of the P4 instance, the awareness output of the solutions is wavier but the resulting values are far from the target data and the final awareness error for this model instance ends up being greater (as seen in Figure 27). However, this could be a consequence of the synthetically generated target values, that could be too difficult to match. We extract similar conclusions for the WOM volume objective. Although trade-off and best awareness solutions achieve reasonable WOM volume outputs for the P1 instance (Figures 28b and 29b), final WOM volume errors are higher for the P4 instance even when considering the fittest solutions (as shown in Figures 28d and 29d).



(a) P1 : awareness fitting

(b) P1 : WOM volume fitting

(c) P4 : awareness fitting

(d) P4 : WOM volume fitting

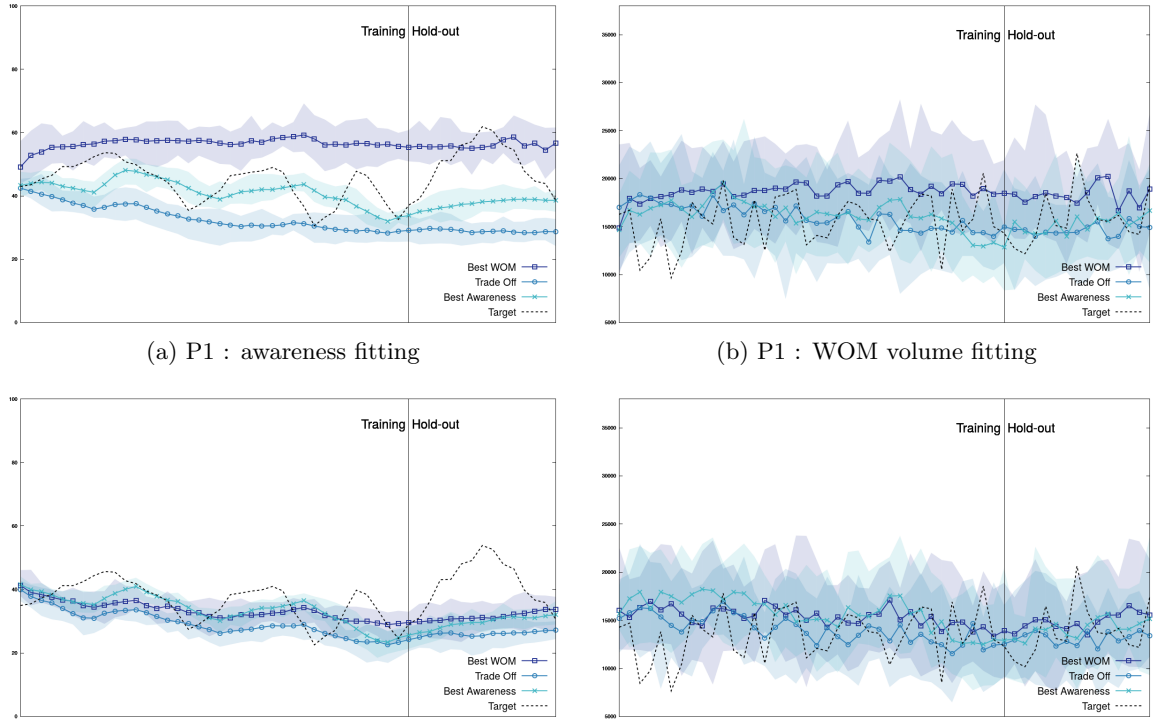
Figure 28: Awareness output and WOM volume over time for P1 and P4 regarding brand 3. In these charts, the central line represents the average of the Monte-Carlo simulation and the blurred areas represent the minimum and maximum values obtained for all the Monte-Carlo 15 independent runs. In addition, the dashed lines represent target values. Best WOM and best awareness (lowest error) solutions are represented with pointed lines containing squares and crosses respectively. Trade-off solutions are represented using lines with circles.

4.5.5.2 Second stage: visual and qualitative multicriteria analysis using moGrams

We continue the analysis of the calibrated model instances using moGrams, which composes the second stage of our framework. As mentioned before, moGrams is a visualization methodology that combines the visualization of both the design and the objective spaces that aids the decision maker enhancing her understanding of the problem [TCP+18]. Our approach is similar to the one followed during the behavior analysis: we apply moGrams to two Pareto set approximations obtained by NSGA-II for both P1 and P4 model instances. In order to perform moGrams generation, we need to define a similarity metric for our calibration problem. Our similarity metric $Sim(X_i, X_j) \in [0, 1]$ compares two solutions (i.e., set of calibrated parameters) X_i and X_j using the normalized Euclidean distance, since our calibration problem considers many independent decision variables. The similarity metric is defined in Equation I.17.

$$Sim(X_i, X_j) = 1 - \sqrt{\frac{(x_{i1} - x_{j1})^2 + ... + (x_{in} - x_{jn})^2}{n}}. \tag{I.17}$$

The generated moGram for the P1 model instance is shown in Figure 30 and its associated Pareto front approximation is displayed in Figure 31. We can see that, given the relatively high cardinality of the Pareto front approximation (46 solutions), the decision making process for this model instance seems too complex to deal with if a visualization method is not considered. Following the moGrams methodology, each node in the generated network is associated to an individual



(a) P1 : awareness fitting

(b) P1 : WOM volume fitting

(c) P4 : awareness fitting

(d) P4 : WOM volume fitting

Figure 29: Awareness output and WOM volume over time for P1 and P4 regarding brand 6. In addition, the dashed lines represent target values. Best WOM and best awareness (lowest errors) solutions are represented with pointed lines containing squares and crosses respectively. Trade-off solutions are represented using lines with circles.

solution (i.e., model parameter setting) from the Pareto set approximation. We draw each node as a pie where the upper pie segment represents the awareness error objective using degradation between orange and white while the lower pie segment represents the WOM volume error objective using degradation between blue and white. For both objectives, a more intense color means a better value with a white color being the worst possible value. In addition, we have included indexes for the solutions in both the network and the Pareto front approximation for making their relation clearer (see Figures 30 and 31). We provide several observations from the moGrams visualization:

- Regarding the structure of the network, we can identify multiple clusters of solutions (i.e., groups of solutions) in the design space. Two of these subsets of solutions, located in the left side of the network, are connected to the general network by Solution 31, which bridges with another subset through Solution 28. In addition, Solution 15 can be identified as another hub that connects to another subset of solutions located in the right side of the network.

- From those clusters we can identify Solution 31 and Solution 28 as the most connected ones, since they are the only solutions with degree 4. Due to their connectivity and the additional information provided by moGrams, these solutions could be interesting configurations for the modeler. In terms of similarity, both solutions have values of 0.9, which suggests they have good flexibility and can be swapped by other solutions with minimum parameter changes.

- The moGrams visualization methodology assists us in validating the best trade-off solution (Solution 9, located in the right side of the map), which could be a suitable model configuration due to its good balance for both objectives. This solution is connected with

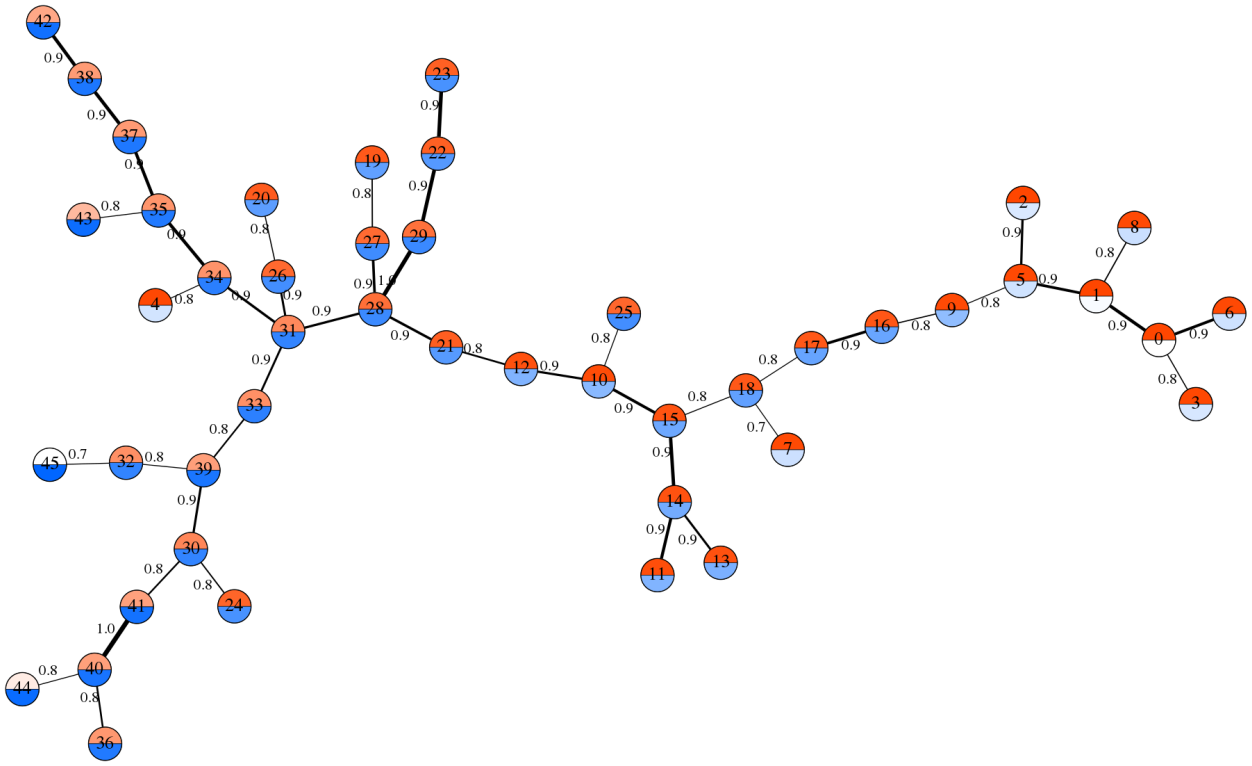


Figure 30: moGrams network representing the non-dominated calibration solutions for P1 model instance.

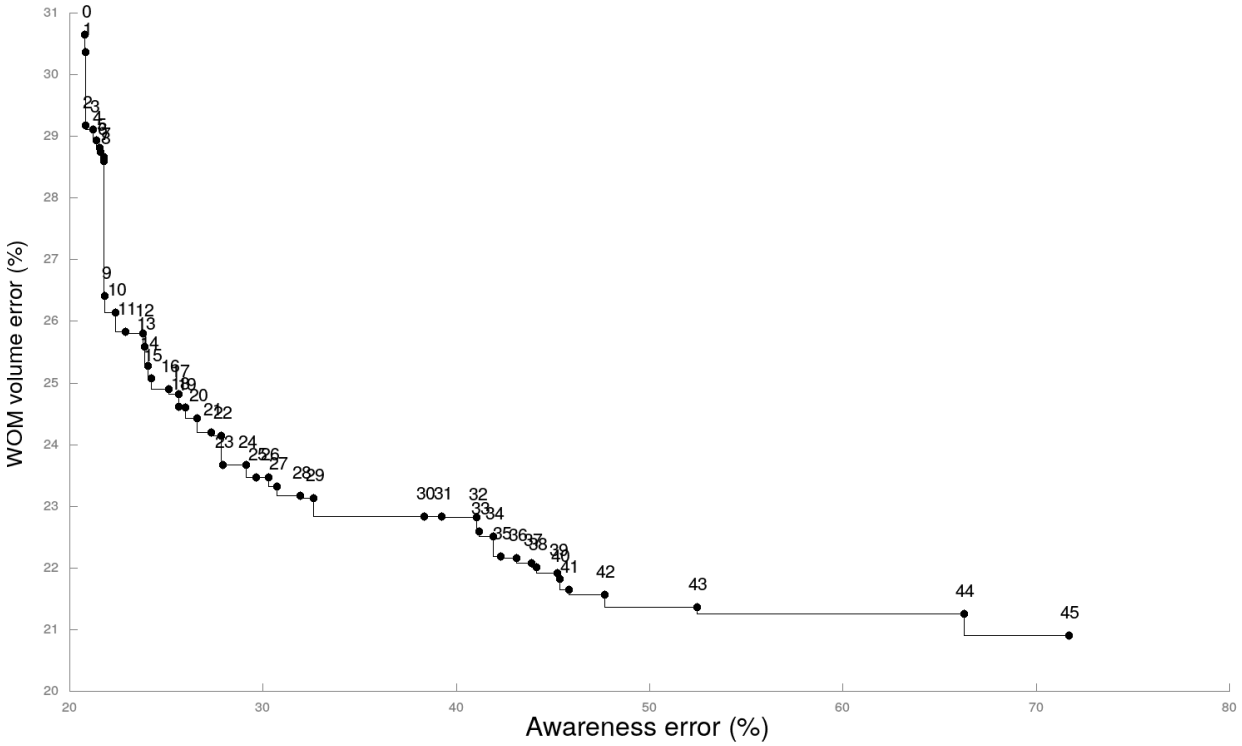


Figure 31: Pareto front approximation for P1 model instance associated to the moGram of Figure 30.

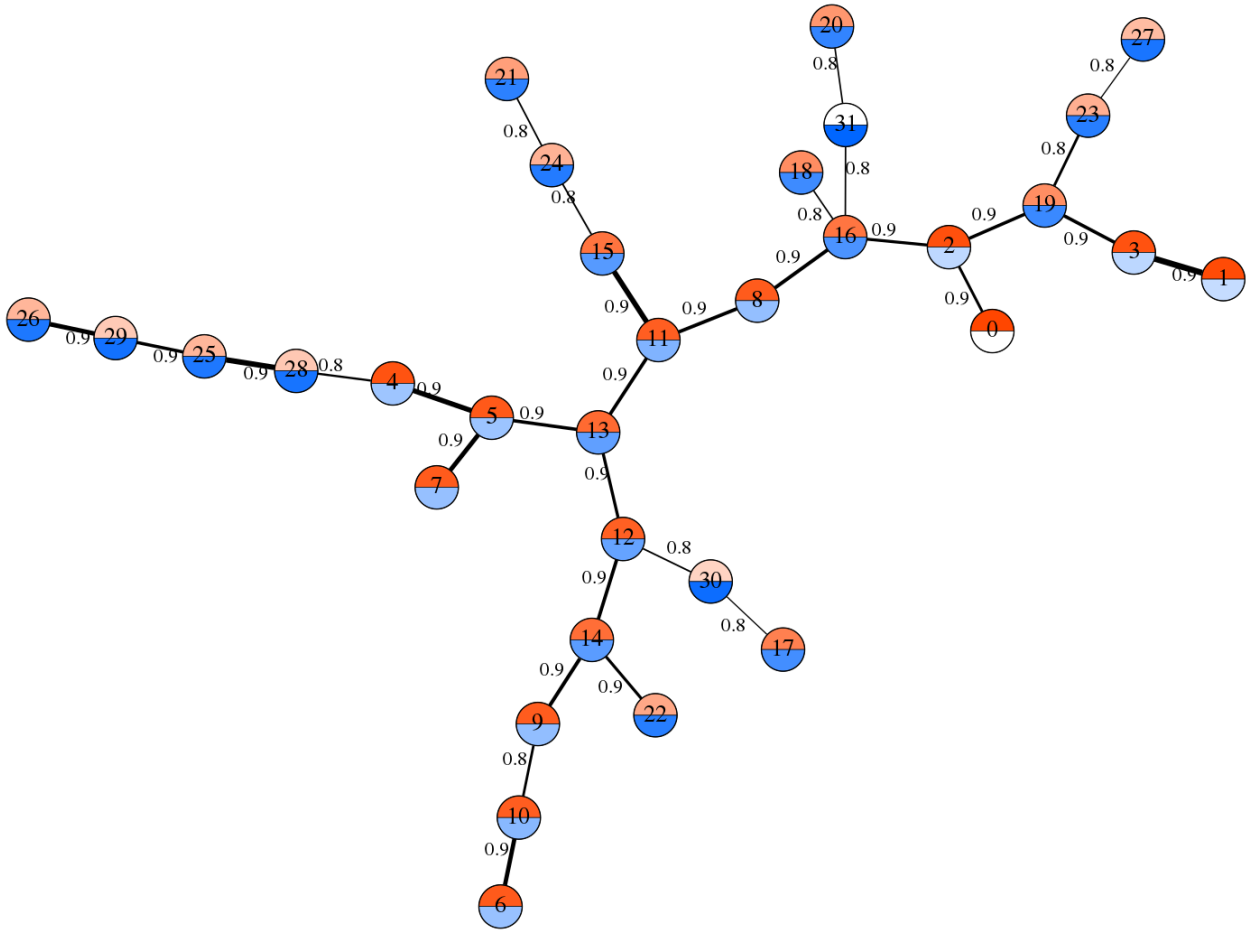


Figure 32: moGrams network representing the non-dominated calibration solutions for P4 model instance.

Solutions 5 and 16 with similarity values of 0.8. However, these solutions are located in the same region of the Pareto front approximation, reducing the interest of swapping Solution 9 with any of its neighbors.

- With respect to the best solutions for each objective (Solutions 0 and 45), both of them are located in opposite regions of the network. Solution 0 has three neighbors with similarity values beyond 0.8, meanwhile Solution 45 has a single connection to Solution 32 with a low similarity value (0.7). However, the neighborhood of Solution 0 may not be really interesting, since all the solutions are close in the Pareto front approximation.

Figure 32 shows the generated moGram for the P4 model instance while its associated Pareto front approximation is displayed in Figure 33. Similarly to the previous moGram, its associated Pareto front approximation has a relatively high cardinality (32 solutions). Again, we can provide the following observations and interesting insights for the modeler from a validation point of view:

- Due to the star topology of the network, three main subsets of solutions in the design space arise, which grow as tree-shaped subnetworks from Solution 13. We can identify Solution 16 as the most connected, since it is the only with four neighbors. This solution could be interesting for the modeler since its connections include solutions from opposite regions of the Pareto front approximations (i.e., Solution 31, the solution with the lowest WOM error and Solution 2, which is close to the solution with the lowest awareness error).

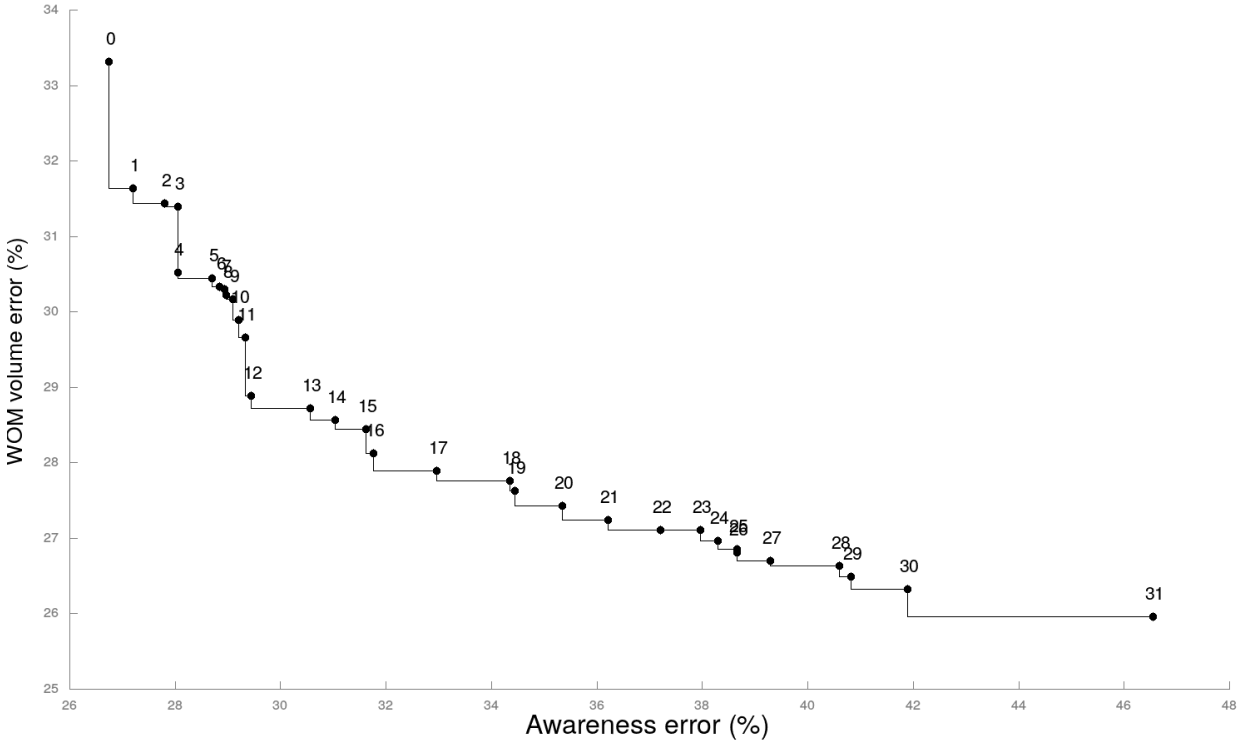


Figure 33: Pareto front approximation for P4 model instance associated to the moGram of Figure 32.

- In addition, several other solutions have three connections. From those, we could highlight Solution 2 and Solution 12 since their connections are more diverse and include solutions from other regions of the Pareto front approximation. Solution 12 can also be identified as the trade-off solution, which could be a plus for the modeler. The most interesting neighbor of Solution 12 is Solution 30 with a similarity value of 0.8.

- Finally, the best solutions for each objective are located in the same subset of solutions. As already pointed out, Solution 31 is connected to the hub defined by Solution 16. However, Solution 0 is isolated with a single connection to Solution 2, which belongs to the same region of the Pareto front approximation. Additionally, we can observe that the proximity in the network for both solutions could be relevant for the modeler since it suggests that modifying the value of some sensible parameters can drastically change the behavior of the model.

5 Summary and discussion of the results

This section addresses the main results obtained during the development of the dissertation, which are discussed in their corresponding subsections. These are the following: two ABMs for analyzing and understanding the 14-M national elections in 2004 in Spain, the development of automatic calibration methods for ABMs using novel metaheuristics based on coral reefs, the analysis of EMO algorithms for automatic calibration of ABMs with multiple KPIs, and a multicriteria integral framework for assisting the modeler during ABM calibration.

5.1 ABMs for analyzing the 14-M 2004 elections in Spain

Both the proposed models have proven useful for analyzing the 14-M national elections in 2004 in Spain. The results of the experiments using the ABM based on the framing theory suggested that the framing effect generated during this period could actually influence the election results by both mobilizing abstainers and deactivating voters from PP party. Although this fact was already pointed out by several other contributions [Bal07, LM06, Mic05, Mon11, ML09, Olm05, RMA07, TR04], this is the first time it is demonstrated using a computational model and the developed what-if scenarios allowed us to obtain additional insights.

First, these scenarios showed how the social network have a key role when exposing an agent population to highly polarized messages. This can be observed in these experiments as a significant swap of votes arises when modifying the social network generation parameter m and the agent talking probability, which regulate the density of the social network and the dynamics of the WOM process. In addition, it was observed that this swap of votes is not linear since some political options are more sensitive to modifying the diffusion rate of the network. Another interesting insight is the aggregated behavior of the mass media channels since they seem to achieve stronger effects than the addition of those channels applied individually. Although the television can be clearly distinguished as the most influencing channel, radio and press increase the aggregated media effect in a noticeable way. This is a relevant insight for the 14-M elections since it corroborates that not only television channels had an important role in the diffusion of the 11-M events but the presence of other media channels were decisive for the elections turnout.

On the other hand, the ABM based on the spatial theory of voting enabled us to obtain insights regarding how the management of the politicians' management of the 11-M attacks could have influenced the ideological distance between the voters and the PP, PSOE, and IU parties. The results of the experiments using this model showed how the combined effect produced by all the media channels supporting the same communicative framework produced a significant and stable impact on the distance values of the voting agents. In addition, the observed impact on the distance values of the voting agents suggests that certain approaches to terror management could prolong over time and produce a long-term effect on the ideological distance. These findings provide sound evidence to suggest that the effects of the politicians management after a shock like the 11-M attacks can produce a change in the ideological distance in the short term.

5.2 Automatic calibration of ABMs using novel bio-inspired metaheuristics

The results of our experiments on automatic calibration of ABMs using a single KPI concluded that both MCRO and CRO-SL performed significantly better than the remaining methods considered. They consistently achieved competitive results across every instance of the benchmark, showing a robust behavior even for those with high dimensionality. L-SHADE and IPOP-CMA-ES also obtained competitive results but they ended up outranked by the coral reefs-based metaheuristics. In addition, the memetic variants of the metaheuristics showed very different behavior depending on the original metaheuristic. As a matter of fact, only half of the memetic algorithms improved

their performance when compared with their corresponding non-memetic counterparts. The case of MCRO-SL was specially interesting because it showed good performance for P1, the smallest instance, but its performance was reduced as the dimensionality of the instances increased. This suggests that the local search procedure had a negative impact on the original trade-off of CRO-SL between exploration and exploitation, which offered better results. Finally, we could observe in the experiments of Section 4.5.3 that our approach still has some limitations despite the good performance of the selected bio-inspired methods when calibrating a single KPI.

5.3 EMO algorithms for automatic calibration of ABMs with multiple KPIs

The results of our experiment on automatic calibration of ABMs using with multiple KPIs concluded that MOEA/D shows outstanding and robust behavior for our problem, being able to perform significantly better than the other EMO algorithms in most instances. This suggests that the decomposition-based strategy proposed by MOEA/D is clearly the best performing for the search space of the analyzed problem. However, we could identify how the performance of MOEA/D was reduced when dealing with specific instances. A deeper analysis of the shape of the search space feasible region and the shape of the Pareto front of these instances revealed that some of them have certain characteristics that can affect the performance of decomposition-based algorithms. For example, long-tailed Pareto fronts are known for reducing the performance of decomposition-based EMO algorithms because these non-symmetric shapes mismatch a generic distribution of weight vectors [ISMN17, MMNI18].

In addition to these characteristics, the calibration results on the high-dimensional instances have shown that its dimensionality erodes the performance of most of the compared EMO algorithms. Although this could be expected, it allows us to identify SMS-EMOA as the most robust EMO algorithm for our problem, as it shows a competitive behavior regardless of dimensionality of the instance and the shapes of its feasible region and Pareto front. The increased dimensionality has the biggest impact on the behavior of SPEA2 and NSGA-II, since they obtain good results and event outperform most of the remaining EMO algorithms for instances having less than 90 decision variables. However, their performance decays for the biggest instances and IBEA and GWASF-GA are able to outperform them.

5.4 A multicriteria integral framework for ABM calibration

The results of our experiments with the proposed multicriteria integral framework showed how this approach can help the modeler in the decision making of the best model parameter setting and thus in the ABM validation. The proposed framework reaches this goal by enhancing both the quality of the calibrated ABM and the understanding of the calibrated parameter configurations. This was validated by two separated experiments. First, a controlled experiment proved that our approach identifies the optimal model configurations for each objective, successfully searching across the parameter search space and providing high-quality calibrated models. Then, a second experiment analyzed the resulting Pareto front approximations by selecting three solutions: the solution with best awareness error (f_1), the solution with best WOM volume error (f_2), and the solution with best trade-off for both objectives. This analysis showed how awareness dynamics were more difficult to adjust than the WOM volume for the calibrated instances, specially for the instance with highest dimensionality.

In addition, the second stage of the proposed framework analyzed the design space of the calibrated ABM instances by visualizing individual Pareto front approximations using moGrams. This demonstrated the usefulness of our approach for validating relevant solutions and assessing their flexibility (i.e., the solution with best trade-off for both objectives) from the Pareto front approximation. By using moGrams we could identify how solutions with the best trade-off had

good flexibility but they did not have interesting neighboring solutions in the decision space. In addition, it allowed the identification of other solutions that could be relevant for the modeler. Due to the dimensionality of the analyzed Pareto front approximations, we can observe that identifying these solutions without considering an advanced visualization methodology such as moGrams can be difficult for modelers and stakeholders.

6 Final remarks and future work

The present dissertation tried to elaborate on the usefulness of ABM simulations and the importance of calibrating and validating these kinds of models. Simulating ABMs and other computational models is useful for representing and analyzing complex systems but validating a model is not straightforward, specially if the modeling technique involves the definition and setting of many parameters. Thus, the main motivation of this dissertation is to improve the tools for calibrating and validating ABMs. In order to accomplish this goal, we presented several contributions across this thesis.

First, we developed several real-world ABMs to be subject of the calibration and validation techniques developed during the current dissertation. These models included two ABMs for analyzing the 14-M 2004 national elections in Spain and an ABM for defining marketing strategies. The first ABM focusing on the 14-M elections studied the framing effect during the 2004 Spanish elections following the 11-M attacks. The designed and implemented ABM simulation replicated elector behavior into artificial voting agents connected by a social network and influenced by the most significant Spanish mass media messages. This model recreated the environmental conditions from 11-M to 14-M from the mass media information point of view. The model was calibrated and validated, achieving a model fitting of 99.13% and employing different validation cases of the model components in the ABM simulation. Using this model we defined several experiments that suggested an actual influence of the framing effect on the election results that mobilized abstainers and deactivated voters from the PP party, in addition to other insights reviewed in Section 5.

The second ABM studied the same political event by focusing on the effects of the politicians management of the 11-M attacks on the ideological distance of voters with the three main parties in the 14-M elections: PP, PSOE, and IU. Multiple model outputs were analyzed for its validation: the average distance for each party (i.e., the main output for our study), the number of WOM interactions in the social network, and the sentiment of the latter interactions. This model allowed us to analyze the effect of the management of the 11-M attacks by the politicians in the ideological distance between the voters and the PP, PSOE, and IU parties. This analysis included three closely related political scenarios from theory of terror management, which enabled us to obtain several insights depicted in Section 5.

In addition, the proposed ABM for marketing strategies was employed for generating multiple instances with increased dimensionality in order to provide a benchmark for comparing different metaheuristic-based methods for automatic calibration. This ABM considered global awareness and WOM volume as its main outputs. Starting from an initial model instance representing a virtual market built with real data, we have synthetically generated 14 additional instances by changing the market characteristics to achieve a progressive dimensionality increase. Using this set of benchmarks we conducted two exhaustive studies of metaheuristics for automatic ABM calibration, which are another contribution of this dissertation.

The first study considered a comparison of novel bio-inspired metaheuristics for automatic ABM calibration considering a single KPI. These experiments allowed us to confirm that coral reefs-based metaheuristics are very successful when calibrating ABMs with an integer-coded scheme. Specifically, both MCRO and CRO-SL performed significantly better than a significant number of competing metaheuristics since they consistently achieved competitive results across every model instance, including those with high dimensionality. The second study assessed the performance of seven EMO algorithms from the three main families for calibrating ABMs with multiple KPIs. This study also considered a classical optimization method such as the Nelder-Mead's simplex method, which clearly showed how classical methods struggle when dealing with multiple objectives. The results of our study concluded that MOEA/D was the best performing EMO algorithm for our problem after considering multiple performance indicators, visualization tools, and statistical tests. Nevertheless, even if MOEA/D performed significantly better than the other EMO algorithms in

most instances, we could observe how certain features of a few specific problem instances reduced its performance. A thorough analysis of these features revealed that the shape of the search space feasible region and the shape of the Pareto fronts of these instances affected the performance of decomposition-based algorithms [ISMN17, MMNI18].

Finally, we can remark the proposal of a multicriteria integral framework for ABM calibration as another contribution of this dissertation. This framework combines the use of EMO algorithms to find the best set of configuration parameters and an advanced visualization method to help the modeler in the decision making of the best model parameter setting. We showed how this framework improves the process of calibration and validation by using ABM instances from our benchmark of instances for marketing. The first experiment was conducted on a controlled environment where the EMO algorithm successfully identified the optimal model configurations for each objective, proving that this approach is able to obtain high-quality calibrated models. A second experiment considered the analysis of three relevant solutions from a Pareto front approximation, which visually showed the difficulty of adjusting awareness dynamics even with reasonable fitting solutions. Then, we analyzed the design space of the calibrated instances by using moGrams on individual Pareto set approximations. This analysis allowed us to identify and validate alternative solutions that would have been ignored without the use of an advanced visualization method due to the high cardinality of the Pareto set approximation.

As a summary, we can conclude that the results obtained by the different experiments developed through this dissertation are sufficient for accomplishing its considered goals. We would also like to remark its success in terms of scientific production, as the outcome of our research produced five publications in top JCR-indexed journals among other related works.

The latter contributions for ABM calibration and validation leave room for several future lines of research. For example, future work will address alternative strategies for modeling the behavior of the agents in the ABMs developed during this dissertation. These more advanced mechanics could involve using fuzzy logic [Zad65] for modeling agent decision making with linguistic information and computing with words [CCD$^+$16, GCCCH20]. In addition, fuzzy cognitive maps [Pap13] could also be used for modeling the behavior of the agents. This could be applied both to the ABMs for political and marketing scenarios. On the one hand, fuzzy cognitive maps could improve the modeling of the voter decision making. On the other hand, an extended marketing model considering additional KPIs such as sales would require an advanced modeling technique for guiding the agent behavior.

In addition, future work will be focused on evaluating the possible improvement of including qualitative pattern features in the fitness function of the automatic calibration methods. Those patterns could be useful for minimizing the already identified loss of information produced by the point-based error fitness functions [YB11, YB15]. Since fitness functions like those employed in our study mainly focus on the distance between series of points, the aggregation of these values can potentially loose the shape of the series in the process. This issue can be solved in multiple ways. For example, the current fitness functions could be modified or additional objectives related to each of the model outputs could be incorporated. Further research should clarify which alternative produces the best results.

Apart from the use of qualitative patterns, other ABM consumer models may require the calibration of additional KPIs, such as the calibration of sales. Calibrating more outputs could be approached by including them as additional objectives, which defines a new scenario where the use of many-objective EMO algorithms will be required. Besides we believe that surrogate fitness functions would be useful for future studies due to the high computational costs of simulating multiple times for every evaluation of a single model configuration [LRS18]. Additionally, we also consider interesting to extend our calibration approach to evaluate the impact of including the

modeler preferences during the calibration process due to the high cardinality of the Pareto set approximations obtained during our experiments [TMKM09].

Part II. Publications

This chapter introduces the scientific contributions published during the development of this dissertation. As a whole, they tackle the work carried out to achieve the stated objectives in this thesis. Each paper is presented in a different section.

1 An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections

- I. Moya, M. Chica, J. L. Sáez-Lozano, O. Cordón. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections, Knowledge-Based Systems, vol. 123, pp. 200-216, 2017. DOI: 10.1016/j.knosys.2017.02.015.
  - State: Published.
  - Impact Factor (JCR 2017): 4.396.
  - Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Order: 14/132. Q1.

An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections

Ignacio Moya[a,*], Manuel Chica[b], José L. Sáez-Lozano[c], Óscar Cordón[a]

[a] *DECSAI and CITIC-UGR, University of Granada, 18071 Granada, Spain*
[b] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia*
[c] *Departament of International and Spanish Economics and GIADE-UGR, University of Granada, 18071 Granada, Spain*

Abstract

Government, politicians, and mass media generated a large quantity of information after the bombing attacks in Madrid on the 11th of March 2004. This information had two competing dimensions on the terrorist group responsible for the attacks: ETA and Al'Qaeda. The framing theory could explain how this information influenced the Spanish national elections on the 14th of March, three days after the attacks. We propose to analyze this political scenario using agent-based modeling to recreate the environment and framing effect of the three days prior to the elections. Using our model we define several experiments where we observe how media communications influence agent voters after calibrating the model with real data. These experiments are what-if scenarios where we analyze alternatives for mass media communication messages and word-of-mouth behaviors. Our results suggest that the framing effect affected the election results by influencing voters. These results also outline the aggregated impact of mass media channels and the different role of each party segment of voters during this period.

Keywords— Social Simulation, Agent-Based Modeling, Voting, Framing Effect, Terrorist Attack 11-M

1. Introduction

On the 11th of March 2004 (11-M), three days before the Spanish national elections on the 14th of March (14-M), a group of terrorists exploded various bombs on trains circulating to Atocha train station in Madrid. 193 people died and about 2000 were wounded. The attacks changed the electoral process: in the morning of the 11-M, the campaign was suspended; on the 12th of March, there were demonstration marches against terrorism in the main Spanish cities; and on the 13th of March, there was a demonstration in front of the headquarters of the People's Party (PP), the Spanish right-wing party who was in the government. Finally, voting surveys failed and the results of the 14-M elections revealed an unexpected change of government.

After the attacks took place, a large quantity of information was generated by government, politicians, and mass media. That huge amount of information pushed the 11-M candidates to

*Corresponding author
Email addresses: imoya@correo.ugr.es (Ignacio Moya), manuel.chica.serrano@gmail.com (Manuel Chica), josaez@ugr.es (José L. Sáez-Lozano), ocordon@decsai.ugr.es (Óscar Cordón)

position themselves in relation to this event. Voters incorporated this political position about the attacks into their voting decision processes [33]. The communicative framework of this event had two political competing dimensions regarding the two terrorist groups which could be responsible for the attacks: ETA and Al'Qaeda. The first position was defended by PP's government, while the second was supported by the left-wing Spanish Socialist Workers' Party (PSOE), the main opposition party, and other opponents [55]. Two main reasons influenced public opinion on PP and PSOE's positions: the evaluation of the management of the government against ETA terrorism and the active participation of Spain in the invasion of Iraq, in March 2003. The majority of the Spanish population positively evaluated the action of the PP government in the fight against ETA. Therefore, President Aznar's bureau declared ETA responsible for the 11-M attacks as an election strategy [38]. On the contrary, the Spanish government's decision to participate in the invasion of Iraq was against the majority of public opinion and political parties.

The presence of terrorism has influenced pre-electoral environments in the past. In the US, the hostage crisis in Iran embassy was few weeks before the presidential elections of November 1980 [65]. In the Netherlands, the mayor of the city of Rotterdam was assassinated nine days before the local elections of 2002 [57]. However, none of the previous cases was comparable to the dimension of the 11-M attacks. As a consequence, the analysis of the turnout of the 14-M elections cannot be performed by comparing it with previous observations because there is not any similar electoral incident.

The post-election studies of the 14-M elections showed that the 11-M attacks influenced the decision of many voters, a thesis that has been corroborated by existing research studies on the 11-M and its impact on the elections [3, 11, 38, 39, 59, 70]. The authors supporting this thesis tend to interpret the elections turnout as a punishment to the ruling party for their mismanagement of the attack, along with their foreign policies. However, no previous study was devoted to explain the framing effect that was generated right after the attack by recreating the main communicative framework. Chong and Druckman [16] defined the framing effect as the psychological process that allows people to develop an *ad hoc* conceptualization of an issue or event, and to readjust their opinion. For instance, an important study showed that 11% of voters changed their minds and decided to go to vote after the attack [10]. This percentage rises to 15.5% in the survey conducted by the Regional Political Observatory [53] but it decreases to 6% in the opinion poll by TNS / Demoscopia [38].

Given the socio-economical and political importance of these facts in the recent Spanish history, the main goal of the current contribution is to analyze the framing effect generated after the 11-M attacks and how it influenced the decision of those who would vote for PP, PSOE, or abstain after the attacks. These two parties and abstention were the three electoral options with the highest support in the 14-M elections. PP and PSOE obtained 81% of votes cast, and 24.83% of the voters abstained, meaning that these three options covered the 84% of the overall voting population. We belief that focusing on the major parties benefits this study since the rest of the votes cast are either received by local parties or achieved an insignificant parliamentary representation. This is because the Spanish electoral system which follows the d'Hondt system [24]. In fact, PP and PSOE obtained 312 of the 350 members of the parliament (i.e., an 89%) with the 81% of total votes.

Our analysis involves studying the influence of mass media treatment of the attack's responsibility into voters and how this influence was spread by individual voters. We propose to model this political scenario using an agent-based model (ABM) methodology [7, 23, 45]. ABM has been broadly applied for social simulation [27, 37, 46, 61] and for modeling political scenarios [40, 43, 51].

The ABM methodology relies on a population of autonomous entities called agents which behave according to simple rules and by interacting with other agents. The aggregation of these simple rules and interactions allow the representation of complex and emerging dynamics as well as defining what-if scenarios and forecasting hypothetical scenarios [34].

By using this ABM framework we simulate the 72 hours next to the attacks and study how this period of time affects the Spanish population when voting for the 14-M elections. The simulated population is segmented using real pre-electoral data to replicate the main political options: PP, PSOE, and abstention. Our ABM simulation framework also reproduces mass media information from real tracking data and the word-of-mouth (WOM) mechanisms [14, 42, 60] by using artificial social networks [4, 72]. Specifically, WOM is modeled by spreading voters' perceptions [5, 20] through a scale-free network [4]. We include mass media information by gathering and modeling the main broadcast media involved in the event (i.e., television, radio, and press) for this period within the simulation.

Using real pre-electoral data as our input, we validate our designed model to fit its behaviour to the actual 14-M election results, calibrating some of its parameters using the election's turnout as the target data. Although the values of some model parameters are directly set using reliable real data sources, some other parameters are more difficult to estimate. These parameters involve WOM volume and media influences, among others, and must be estimated through automatic calibration. The automatic calibration process tunes these parameters and it is a crucial phase in model validation [14, 54, 58, 62]. More specifically, we have implemented our calibration process using metaheuristics [69]. The selected metaheuristic is a memetic algorithm [50] based on a genetic algorithm [2, 28] and a local search procedure which adjusts the main WOM and media parameters to replicate the reality.

We define several experiments where we observe how media communications influenced voters through their corresponding agents for the ABM-based calibrated simulation model. These experiments are what-if scenarios where we analyze alternatives for mass media communication messages and WOM behaviors. Alternatives for mass media involve different communication strategies, such as altering media messages to favor one of the identified framings. In the case of WOM, these alternative behaviors involve modifying how segmented voters react to WOM. Additionally, the proposed set of what-if scenarios is used for studying the impact of both media treatment and WOM in the 14-M election results. This study is carried out by monitoring the elections turnout for the different scenarios.

The structure of this paper is as follows. Section 2 discusses related work and motivation for analyzing the framing effect in our study. Then we introduce the description of the model and its structure in Section 3. Section 4 presents the model validation with real data. In Section 5, we run the what-if scenarios where we study how the designed model behaves under different communication strategies. Finally, in Section 6, conclusions and final remarks are discussed.

2. Related work and motivation

2.1. ABM for simulating political scenarios and mass-media influence

ABM techniques have been extensively applied in the field of political sciences for dealing with political party competition [35, 36, 40]. These approaches consider both parties and electors as moving entities that make decisions continuously. That is, electors react to politicians behavior and politicians reconsider their strategy regarding electors decisions. In [51], authors extended this

approach by including mass media influence. This new role for mass media is focused on campaign organization, where political parties use media for enhancing their image regarding their voters. The latter studies show how mass media can be useful to add realism to the model and better explain the political scenario. We will explain how we incorporate mass media to our model in Section 3.4.

In [29], mass media influence is studied by distinguishing two possible behaviors: global and local. The essential differences between these behaviors are focused on how they behave regarding time and space. Mass media effect and the role of mass media during campaigns regarding voters mobilizations is also analyzed in [26]. This approach is interesting because it analyses the abstention factor, instead of focusing on individual voting preferences. Authors of [47] considered the polarization effect of mass media in opinion dynamics to examine how mass media affects individuals. In this model, mass media messages are propagated via social interactions, showing dynamic changes over different scenarios where strongly polarized messages influence the agent population.

In [43], the authors examined how mass media influence the opinion formation through opinion leaders (i.e., influentials) [41]. Using an ABM simulation, the authors highlight the importance of the communication networks used by opinion leaders to influence the public. Another approach to the cascade of influence and its relationship with social networks can be found in [71]. Some contributions like [68] incorporate topology restrictions for structuring social influence into voter communities, creating substructures where agents with similar attributes are grouped. The substructures created this way are applied for modeling agents' social communications and influencing their political decision making.

ABM has also been used for studying the propagation of political perceptions. In [76], the propagation of the agent's knowledge is modeled using a space based approach. In this case, the propagated agents' knowledge depends on the satisfaction of agents with the current political situation. This approach simplifies the topology problem by assigning different radius to each agent, thus some agents will share their perceptions further, reaching more agents. A similar approach is followed in [63], where an ABM models the effect of social influence regarding voting preferences.

2.2. Using the framing effect for explaining the voting process in the 14-M elections

Due to the special nature of the 14-M elections, several publications were dedicated to study this phenomenon from the political perspective. The main topic of study is to find out if the attacks influenced the election process [3, 11, 38, 70]. Most authors seem to agree on interpreting that the elections turnout involved a punishment to the ruling party for its mismanagement of the attack, along with its foreign policies. The study carried out in [38] showed the influence of the attacks on the voting population when compared to a similar scenario where no attacks had taken place. In this case, authors used counter-factual simulations to analyze the influence of both the information treatment of the attacks and the foreign policies of the government [25], concluding that the management of the attacks by the government could have influenced the voters. However, their experiments did not reproduce the communicative framework and they neither simulated scenarios with alternative information treatments as done in the current contribution.

In our case, we aim at studying this communication environment and information treatment by using the framing theory. The framing theory [15, 16, 22] focuses on how communication can emphasize some features of the transmitted message to influence how this message is perceived. Because of the wide range where it can be applied, studies analyzing the framing effect appear in many social disciplines [8, 19, 48]. In the framing effect, two types of subjects take part: i) the speakers, who invoke the communication, and ii) the public opinion, which could modify their

political attitude after receiving the informational content. In the 14-M elections, the external event of the campaign was the 11-M attacks; speakers were political elites, mass media, and social activists; and the public opinion were the voters. The psychological process of the framing effect on the 14-M elections assumes that voters received the messages of the communication framework that was generated after the attack [66], but they mainly paid attention on those who helped them decide their vote [32].

The framing theory has been previously applied to the 11-M attacks for studying the integration of the Islamic community living in Spain [21]. The framing effect analyzed in that study is the one generated from the months following the attacks and how it influenced Islamic segregation in Spain. However, there is no previous study about the framing effect generated by the attacks on the 14-M elections.

Most of the publications considering framing theory used other approaches but ABM. Among them, we can distinguish those focused on political party competition over time and its relationship with public opinion [19, 30, 48, 64], and those interested in framing competition [8, 15]. Therefore, the current manuscript presents a methodological novelty for modeling the framing effect in political scenarios.

3. Model description

This section describes the main ABM design's issues. Section 3.1 presents the general structure of our model. Section 3.2 describes the mechanics and behavior of agent's voting. Section 3.3 introduces the artificial social network and its features, and Section 3.4 shows how mass media channels are modeled. Finally, Section 3.5 addresses the calibration process of the model with respect to real data.

3.1. ABM general structure

Our proposed model simulates the 72 hours from the attacks to the election time: from March 11th at 08:00 am to March 14th at 08:00 am. The time-step of the ABM simulation is an hour, as it correctly fits with the mass media schedule. After the 72 steps of the simulation, every artificial voter or agent (which represents a group of real voters) votes and the simulation outputs the elections' results. During the simulation period, agents receive information from mass media and spread their political perceptions through their social network in a WOM process.

The initial perceptions of the agents of the simulation come from pre-electoral data of the Spanish government [10], since this source resulted accurate predicting the previous national elections results (this data is further described in Section 4.1). Therefore, the simulation starts with no framing effect over the voters. In order to model the voters (agents' population), we have divided agents into three segments: PSOE voters, PP voters, and abstainers. This segmentation is done to better fit the pre-election survey data [10]. Using this segmentation, agents' parameters are defined at the segment level, so agents from different segments behave differently. This design decision makes the ABM simulation more realistic and heterogeneous as well as facilitates the definition of the model's parameters.

Figure 1 shows an illustrative diagram with the structure of the model and the flow of a mass media message. The next sub-sections will describe all the parameters and processes of the model. We use the size of the pre-election survey [10], i.e., 24,109 respondents, as the ABM population size. This way we ensure enough granularity for the population to have a number of agents that represents the political conditions and available data from polls and National studies. Our target

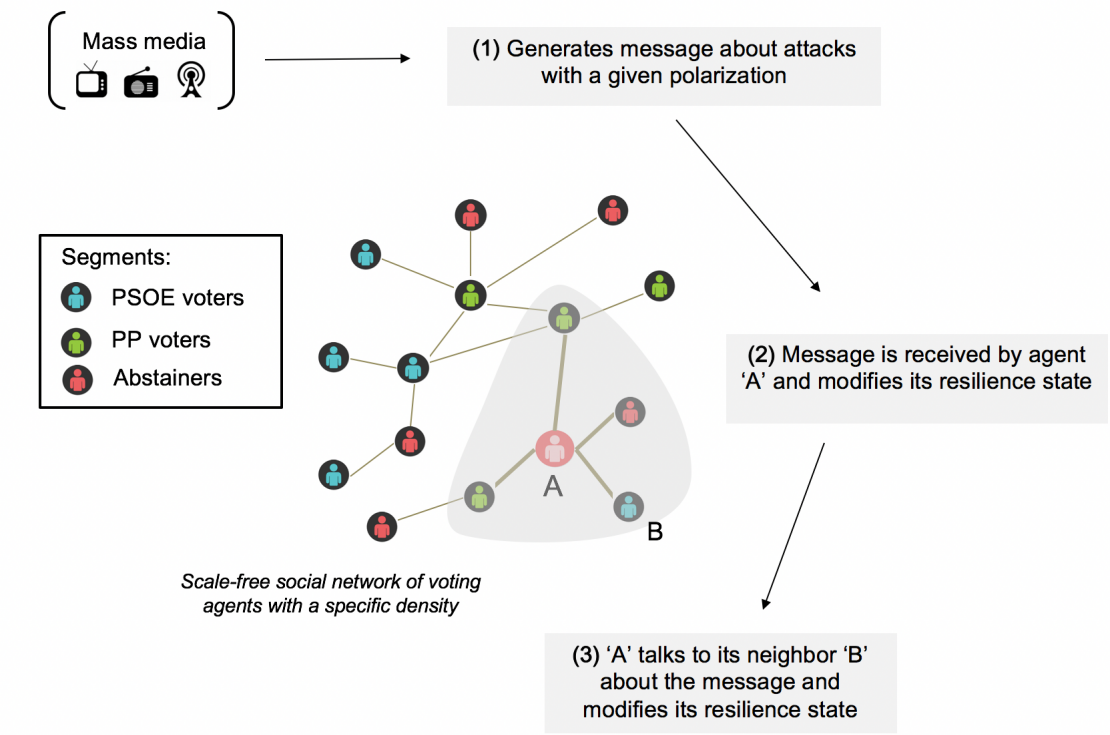


Figure 1: General structure of the model and example of mass media message sent to the voting agents, which react to it and spread their perceptions to their neighbors in the social network. See more details about parameters and mass media messages from Section 3.2 to 3.4.

real population is the sum of PSOE and PP voters, plus abstainers, which represents 29,238,662 people. Thus, the ABM maps one agent/voter with a 1:1,212.77 ratio.

3.2. Agents' state and update rule

Agent population will be influenced by the two framing effects generated after the 11-M attacks: ETA is responsible for the attacks and Al'Qaeda is behind the attacks. Each agent A manages the framing effect by the use of a state variable, called *resilience* and encoded by μ_A, a real-valued variable defined in interval $[0, 10)$. The resilience of the agent is the current vote positioning of the artificial voter with respect to the three voting possibilities. This value will change during the simulation depending to its resilience with respect to the received amount of external influence by mass media and other artificial voters. By using this framing effect variable μ we can obtain the voting alternative of each agent by applying Equation 1. $v_A(\mu_A(t))$ represents the voting option of agent A at time-step t with $t = \{0, ..., 72\}$.

$$v_A(\mu_A(t)) = \begin{cases} 0, & \text{if } \mu_A(t) \in [0, 3.3)\,, \\ 1, & \text{if } \mu_A(t) \in [3.3, 6.7)\,, \\ 2, & \text{if } \mu_A(t) \in [6.7, 10)\,. \end{cases} \tag{1}$$

This function returns 0 if the the agent votes for PSOE, 1 if the agent abstains, and 2 if the agent votes for PP. Voters can change their vote v from $t-1$ to t, since they can be influenced

Before attacks $V(0)$	**Election day** $V(72)$	**Framing effect on a voter**
PP	PP	Reinforcing
PSOE	PSOE	Reinforcing
Abstention	Abstention	Reinforcing
PP	PSOE	Converting
PSOE	PP	Converting
Abstention	PP	Activating
Abstention	PSOE	Activating
PP	Abstention	Deactivating
PSOE	Abstention	Deactivating

Table 1: Framing effects in 72 hours for the elections March 14th, 2014.

by different sources (i.e., other agents and multiple mass media channels). The final voting option for agent A will be the result of $v_A(\mu_A(72))$. At the beginning of the simulation, the μ variable is randomly initialized for each agent in each of the three segments using a uniform distribution in the segment-specific interval. These intervals are: [0, 3.3) for agents at the PSOE voters segment, [3.3, 6.7) for abstainers, and [6.7, 10) for agents at the PP voters segment.

If agent A resilience value (μ_A) moved to other option's interval, it will vote for that option, modifying its behavior as shown in Table 1. For example, if an agent gets its resilience to a value between 3.3 and 6.7, it will abstain, even if it belongs to PP or PSOE voters segments. Thus, changes affecting resilience (μ) can generate four effects on the vote: reinforcement, conversion, activation [6, 41], and deactivation [17] (see Table 1). Reinforced voters are those who voted the same electoral option at both steps $t = 0$ and $t = 72$. Converted voters are those who reoriented their vote, choosing another option at $t = 72$. Activated voters are those who did not want to vote initially, but chose to vote at $t = 72$. Deactivated voters are those who wanted to vote at $t = 0$, but finally did not vote at $t = 72$.

3.3. Social network of agents and their word-of-mouth interactions

Our agents comprise an artificial social network [4, 72]. We choose to model this social network using an artificial scale-free network [4] because of the lack of information about the real social network in 2004 and the fact that most real networks match with this network model [4, 52, 56].

Degree distributions in scale-free networks follow a power law distribution [4]. This kind of topology has been pointed out as a realistic way of modeling real networks, where few nodes have a significantly large number of connections (hubs of the social network) and most nodes have few connections. Scale-free networks can be generated iteratively using the Barabasi-Albert preferential attachment algorithm. The Barabasi-Albert algorithm has a main parameter m which regulates the network's growth rate and its final density [4]. Network generation starts with a small clique (a completely connected graph) with m_0 nodes. At each successive iteration, a new node is added and connected to m different existing nodes. When a new node is randomly connected to an existing node, the probability of choosing an existing node is proportional to the degree. After t iterations, the Barabasi-Albert algorithm results in a social network with mt edges. The average degree of the social network is then $\langle k \rangle = 2m$.

The agents are able to spread its perceptions during the simulation using the artificial social network. We model this WOM interaction as a contagion process which allows the spreading of the information through the nodes of the social network depending on the influence and connections of these nodes (i.e., agents or artificial voters) [52, 77]. Every agent A has a talking probability ($p_A(t) \in [0, 1]$) to spread its perceptions about the current framing (i.e., its μ value) at each step of the simulation. When the probability check passes, the agent will talk with all of its neighbors of the social network. We will model this interaction using a variable called influence change (Δ), which modifies the strength of the agent' influences to its neighbors. This interaction is modeled in a directed way, meaning that the talking agent influences its neighbors and not in the opposite way. This change of the resilience value controlled by Equation 2, where $\mu_B(t)$ refers to resilience value of the listening agent B when talking with agent A. Δ_A refers to the influence change value of agent A.

$$\mu_B(t+1) = \mu_B(t) + |\mu_B(t) - \mu_A(t)|\, \Delta_A. \tag{2}$$

We also include in our model a variable called influence decay ($d\Delta$) which modulates how previous influence is forgotten over time. This decay effect is applied at the beginning of each step for every agent and reduces accumulated influence. Previous accumulated influence ($\delta_A(t)$) is computed following Equation 3 and represents the sum of previous changes to μ performed by WOM from the initial step 0 to current step t.

$$\delta_A(t) = \sum_{i=1}^{i=t} \left(\mu_{A_i} - \mu_{A_{i-1}}\right). \tag{3}$$

The resilience value change for agent A due to decay is defined in Equation 4, where $d\Delta_A$ represents the decay rate which modifies the accumulated influence. Let us finally remind that, in order to make the ABM more heterogeneous, each segment of the model can have different values for the talking probability ($p(t)$), influence (Δ), and influence decay ($d\Delta$).

$$\mu_A(t+1) = \mu_A(t) - (\delta_A(t) d\Delta_A). \tag{4}$$

3.4. Modeling the mass media

Registered media audience from 11-M to 14-M is modeled as an external influence for the agents of the ABM simulation [1, 44]. These external influences work as global mass media [29], with the same probability of influencing any agent regardless of the social network and segment. Moreover, these external influences are parameterized to define differences between press, radio, and television[1]. Mass media channels can influence any number of agents at random depending on the channel audience at each step.

The selection of the messages forming the communicative framework that originated after the attack has been performed considering three criteria. First, the communicative diversification, because we analyze the messages broadcast by television, radio and newspapers, instead of focusing on a single type of mass media. Second, we select mass media channels that broadcast at a national scale. Finally, we select messages that respect the plurality of information. This selection was

[1]In 2004, the Internet influence was not strong enough to be considered in our model. Communications via phone messages are modeled using WOM.

Channel name	Type
El Pais	Press
El Mundo	Press
ABC	Press
Cadena Ser	Radio
TVE	Television
Antena 3	Television
Telecinco	Television

Table 2: Selected mass media channels

developed including any message that contained information regarding the attacks, from regular news sessions to special bulletins and statements from political figures.

Following those criteria, we design a complete communicative framework that covers the main information broadcast between the 11th and the 14th of March, 2004. Specifically, we include the main mass media channels broadcasting in Spain during this period: El Pais (press), El Mundo (press), ABC (press), Cadena Ser (radio), TVE (television), Antena 3 (television), and Telecinco (television). A summary of the selected mass media channels is depicted at Table 2.

Each mass media channel can have different values for its parameters even if they belong to the same media type. For instance, the existing television channels (i.e., TVE, Antena 3, and Telecinco) have different parameter values. These channels will also spread different messages at any step t, depending on which terrorist group is considered as responsible by its broadcast information (ETA or Al Qaeda). Each transmitted message has a polarization value, modeled as $m_C(t) \in [-2, 2]$. $m_C(t)$ represents the information bias of the message (ETA *versus* Al Qaeda) at t broadcast by specific channel C. In order to set the values of these polarized messages, a team of three media experts reviewed and studied press, radio, and television information during the three days period. Because of the subjectivity of the process, they individually scored them and later, agreed about the polarization value of all the media messages.

$m_C(t)$ is set to -2 if the information of the message strongly points to Al Qaeda and 2, if it strongly points to ETA. Both -1 and 1 refer to a weak authority and 0 refers to not assigning attacks' authority to any specific group. As our simulation steps by hour, we use the average polarization for a given time slot when two or more messages appear within it for a specific media. In a similar way, if a single message was assigned to different values during the experts' classification process, its final value was agreed and calculated using the averaged value.

Additionally, mass media is modeled by the following parameters:

- Reach. This parameter models the maximum amount of people each channel is able to hit. It this sense, some media are able to reach more people than others. Moreover, there is a difference between the amount of people that can be influenced for a given time slot or during the whole simulation. Thus, we use a reach parameter r_{short} for the percentage of agents that may be influenced within an hour (time-step). Another parameter, called r_{long}, is used for the maximum percentage of agents that can be influenced during the simulation. Data for setting the values of the latter two parameters were taken from Zenith study from 2013 [2],

[2] http://blogginzenith.zenithmedia.es/estudio-zenith-los-medios-en-espana-y-portugal-un-terreno-cambiante/

that considers the percentage of the population consuming media content both in the short and in the long term. Because of the huge attention that the media received during this period, we have matched the r_{short} value with the maximum coverage that a media channel may achieve within a single time slot. With respect to the r_{long} parameter, we set it to the overall maximum coverage. Additionally, these values are displayed along with the evolution of mass media in Spain between 2006 and 2013. Since the selected media channels show steady values in their evolution for the first years of the study, we can approximate the reach parameters in our period of study (March 2004) to the values shown in 2006.

- Influence. When a mass media channel impacts an agent A, its message influences the resilience value of the agent (μ_A). We define the influence change parameter (Δ') to modulate this effect. This behavior is similar to the one defined for the social interaction between agents. This way, resilience change is performed using the received message and the influence change value for the media channel. As the same message could be received multiple times by the same agent, its maximum influence is limited to the overall influence value (Δ^{max}). Additionally, we represent the previous influence accumulated by the channel (δ') analogously to WOM. The resilience value change of agent A after the influence of channel C is formulated by Equation 5, where $m_C(t)$ refers to the transmitted message and Δ^{max} refers to the maximum influence value.

$$\mu_A(t+1) = \mu_A(t) + (\Delta_C^{max} - \delta'_C)\Delta'_C m_C(t). \tag{5}$$

  In addition, agents may forget what they just watched or read as the novelty of the message expires [73, 74]. We include a parameter for measuring how media influences can be forgotten by the agents. This effect is modeled as influence decay ($d\Delta'$) which reduces previous influence, similarly to the one defined for social interaction. Equation 6 defines the decay update for agent A due to the influence of channel C.

$$\mu_A(t+1) = \mu_A(t) - (\delta'_C d\Delta'_C). \tag{6}$$

- Buzz. Information during those events can get a critical media impact and may generate a viral buzz effect. We model this effect through a variable called buzz increment (τ_C) for a channel C. This increment is applied to the agents' talking probability as a percentage increment to the initial talking probability ($p(0)$) of the agent. In contrast, as information is getting older, its buzz effect decreases over time. In a similar way to media influence, we model this effect with a variable called buzz decay ($d\tau_C$). Buzz decay decreases the talking probability depending on the previous amount of talking probability that has been previously incremented to the agent (σ). The update of the talking probability of agent A due to both buzz increment and decay effects of channel C, defined by Equations 7 and 8 respectively.

$$p_A(t+1) = \begin{cases} p_A(t) + (p_A(0)\tau_C), & \text{if } (p_A(t) + (p_A(0)\tau_C)) \leq 1, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

$$p_A(t+1) = p_A(t) - (\sigma_A d\tau_C). \tag{8}$$

By using these parameters, we can model how media spread their messages to the entire population of agents during the whole period of time. As previously commented, mass media information during these three days period suffered from strong polarization, moving from one position to the opposite one. At the beginning, media information strongly pointed to ETA's authority, but later declared Al Qaeda's authority. Thus, the content of the message transmitted via a certain mass media channel C will change during the simulation and we model it by scheduling the different message polarization values for each channel ($m_C(t)$).

3.5. Calibration

Automated calibration is a data-rich and computationally intensive process that uses an error measure to compare real-world data to model-data and then, tunes the parameters of the model to match the data [54, 62]. Automated calibration attempts to discover the best parameters of the model that fit the model output to the data. Therefore, automated calibration requires an error measure and an optimization method for modifying the parameters in a systematic way in order to minimize the error measure. After calibration is finished, the resulting parameter values need to be carefully reviewed and validated.

With regards to the optimization method, since the parameters in computational models exhibit non-linear interactions, the best option is to use a non-linear optimization algorithm that can search across a large span of the model parameters space [12, 49, 67]. Metaheuristics are a family of approximate non-linear optimization techniques that provide *acceptable* solutions in a reasonable time for solving hard and complex problems in science and engineering [69].

The optimization process will asses the quality of the model by running the computational model and comparing its outputs to the elections data. By doing this, we adjust the parameters of the model to match the model's output with the 11-M reality. The selected parameters for automatic calibration are those related with the parameters that control WOM and mass media diffusion, which are both the most uncertain and the hardest to estimate with the existing information. For each defined parameter, we also use a parameter range to set its possible values during optimization. The set of 44 real-valued parameters to be calibrated are shown in Table 3. A brief description of those parameters is as follows:

- WOM diffusion parameters. For each defined segment S, we will calibrate its initial talking probability ($p_S(t)$), influence change (Δ_S), and influence decay ($d\Delta_S$), i.e., 9 parameters.
- Mass media parameters. For each defined mass media channel C, we calibrate its maximum influence (Δ_C^{max}), influence change (Δ'_C), influence decay ($d\Delta'_C$), buzz increment (τ_C), and buzz decay ($d\tau_C$); i.e., 35 parameters.

The selected automated calibration algorithm is a memetic algorithm [50] composed of a steady-state genetic algorithm [2, 28] and a local search procedure. The pseudo-code guiding the memetic algorithm is shown at Algorithm 1. The calibration algorithm initializes its population generating feasible solutions. Thus, every generated individual is a feasible configuration of models' parameters. The creation of the population of the memetic algorithm is randomly performed by selecting a value for each gene between a range of values. The fitness function designed for guiding the optimization algorithm measures the distance between the election results and the simulated output. Fitness values are computed using a symmetric mean absolute percentage error (SMAPE), defined in Equation 9, which facilitates to increase the sensitivity for miss-voting agents. In this

Parameters to be calibrated (44 parameters)				
	Mass media (35 in total)			
	El Pais	**El Mundo**	**ABC**	**Cadena Ser**
Buzz	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$
Influence	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$
	TVE	**Antena 3**	**Telecinco**	
Buzz	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$	
Influence	$(\delta_{max}, \Delta, d\Delta)$	$(\delta_{max}, \Delta, d\Delta)$	$(\delta_{max}, \Delta, d\Delta)$	
	WOM process (3 per segment, 9 in total)			
Initial talking prob.	$(p_{PSOE}(0), p_{ABST}(0), p_{PP}(0))$			
Influence change	$(\Delta_{PSOE}, \Delta_{ABST}, \Delta_{PP})$			
Influence decay	$(d\Delta_{PSOE}, d\Delta_{ABST}, d\Delta_{PP})$			

Table 3: List of parameters to be automatically calibrated by the memetic algorithm.

Algorithm 1: Pseudo-code for the memetic algorithm.

```
1  begin
2      Initialize population P with valid random solutions;
3      Evaluate each individual in P;
4      while number of evaluations < 10,000 do
5          A = Select individual using 3-tournament selection;
6          B = Select individual using 3-tournament selection;
7          randC = random(U(0,1));
8          if randC ≤ Crossover Probability then
9              Offspring = Breed A and B using BLX-α crossover;
10         else
11             randM = random(U(0,1));
12         if randM ≤ Mutation Probability then
13             Offspring' = Mutate Offspring;
14         else
15             Offspring' = Offspring;
16         Evaluate Offspring';
17         W = worst individual in population P;
18         if Offspring' is better than W then
19             Replace W by Offspring' in P;
20     Improve best individual found using local search;
```

equation, A_t represents actual election results and F_t represent the simulated election results. The sensitivity of the calibration is 0.0248% as the mapping ratio was 1:1,212.77. As previously exposed in Section 3.1, this ratio defines the relation between the number of agents and the size of the real population.

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{\frac{(|A_t|+|F_t|)}{2}}. \quad (9)$$

The algorithm follows a steady-state approach with a population of 100 real-coded chromosomes and 10,000 evaluations as stopping criteria. The algorithm also uses 3-tournament selection, a BLX-α crossover [31], and a uniform random mutation mechanism. The crossover operator generates two offspring by crossing two parents with a probability $p_c = 1$. It truncates the selected values over the gene set of feasible values after selecting it from interval $[c_{min} - I\alpha, c_{max} + I\alpha]$, $c_{max} = max(v_i^1, v_i^2)$, $c_{min} = min(v_i^1, v_i^2)$ and $I = c_{max} - c_{min}$. v_i^1, v_i^2 are the feasible decoded values from the genes of the parents. α defines the level of exploration for the operator. If α is set to 0, BLX-α is equivalent to the flat crossover. Regarding mutation operator, we choose to assign mutation probability $p_m = 0.1$ for each gene. When the probability check passes, a new value is generated for that gene using an uniform distribution and the specific range of values for that gene.

4. Validation of the model with elections' data

4.1. Data description

We use the data obtained from the following sources to set the values for some parameters and to restrict the search for the unknown parameters' values (i.e., using an automatic calibration method):

- The final election results to match the model's output have been take from Observatorio Político Autonómico [53].
- The voting intention before the 11-M attacks has been extracted from the 2555 study of the Spanish Centre for Sociological Research. Centro de Investigaciones Sociológicas [10] is a National survey that was executed between January 24 and February 15, 2004, with a sample size of 24,109 interviews. From that interview, we focus on the question regarding who they were willing to vote on the next national elections to be held in March 14th. We use these data to model initial resilience values for the agents' population at step $t = 0$. Surveys for previous elections suggest that this source is accurate enough to be used for this purpose. For example, in the national elections of 2000, the pre-election survey predicted that PP will win the elections with 47.8% of votes and in the end they won with 44.52%.
- Three different sources have been used for setting the polarization message value of $m_C(t)$ for the whole simulation:
  - We took the television information from the informational volume 19-20 from *Quaderns del Consell de l'Audiovisual de Catalunya* [18].
  - The audio from the radio was gathered from Cadena Ser, since it had the highest audience rate at that moment and received special attention during this period.

Parameters	Values obtained from available data			
	El Pais	**El Mundo**	**ABC**	**Cadena Ser**
$\{r_{short}, r_{long}\}$	$\{0.396, 0.584\}$	$\{0.396, 0.584\}$	$\{0.396, 0.584\}$	$\{0.543, 0.611\}$
	TVE	**Antena 3**	**Telecinco**	
$\{r_{short}, r_{long}\}$	$\{0.46, 0.995\}$	$\{0.46, 0.995\}$	$\{0.46, 0.995\}$	

Table 4: List of parameters set by the modelers using the available data (without including them in the automatic calibration method).

  - The information about the chosen newspapers (ABC, El Mundo, and El País) was directly collected from them and can be accessed from the collected on-line database [3].

- There has been a thorough analysis of the broadcast informational content from 07:30 pm (March the 11th) to 12:00 am (March the 14th): television (Antena 3, Tele 5, and Spanish National Television), radio (Cadena SER), and newspapers (ABC, El Mundo, and El País). As said, the most important media were selected for the study according to *Encuesta General de Medios* (general media survey, in English) [1]. The share of the three analyzed television channels exceeded a 75% [44]. For radio, we included Cadena Ser, a radio channel which had the most important role for this political event [55]. In addition, the three selected newspapers were the most read ones during this period of time prior to the elections.

4.2. Model calibration results

In order to test the model behavior we show the calibration results using historical voting data as well as different validation scenarios. These validation scenarios are built by removing some components of the ABM simulation model to observe its behavior with respect to the historical trends. First, we calibrate the ABM with all the designed components. Due to the lack of empirical data about the social network of voters before the elections, we choose to set the parameter m of the Barabasi-Albert algorithm to 2 in order to generate the scale-free social network. In our case, as we have a population of 24,109 agents (i.e., a social network with 24,109 nodes), this results in a network density of 0.00017 and an average degree $\langle k \rangle = 4$. The influence of different parameter m values is studied later at Section 5.1, where we evaluate the model behavior when generating networks with different densities using these m values to create different network configurations.

The calibration method needs to simulate, for each model configuration, 30 Monte-Carlo runs which takes 109.165s. Additionally, we studied the empirical computational complexity of the model simulation in Figure 2. This chart displays the evolution of the computational cost of the ABM model (Y axis) when increasing the number of agents (X axis). The chart also shows the evolution in comparison with standard Big O notation functions (i.e., O(*log* n), O(n), and O(n *log* n)). We can observe that the computational cost is lower than a linear cost O(n) with a low-medium number of agents but the cost drastically increases with larger populations. Therefore, the computational complexity of the model approximates an exponential shape function.

Table 4 shows the values of the parameters which were set without the use of an automatic calibration because of the existence of available data (mainly, mass media reach parameters).

[3] This database can be accessed from `http://ugr.mynews.es/hu/`

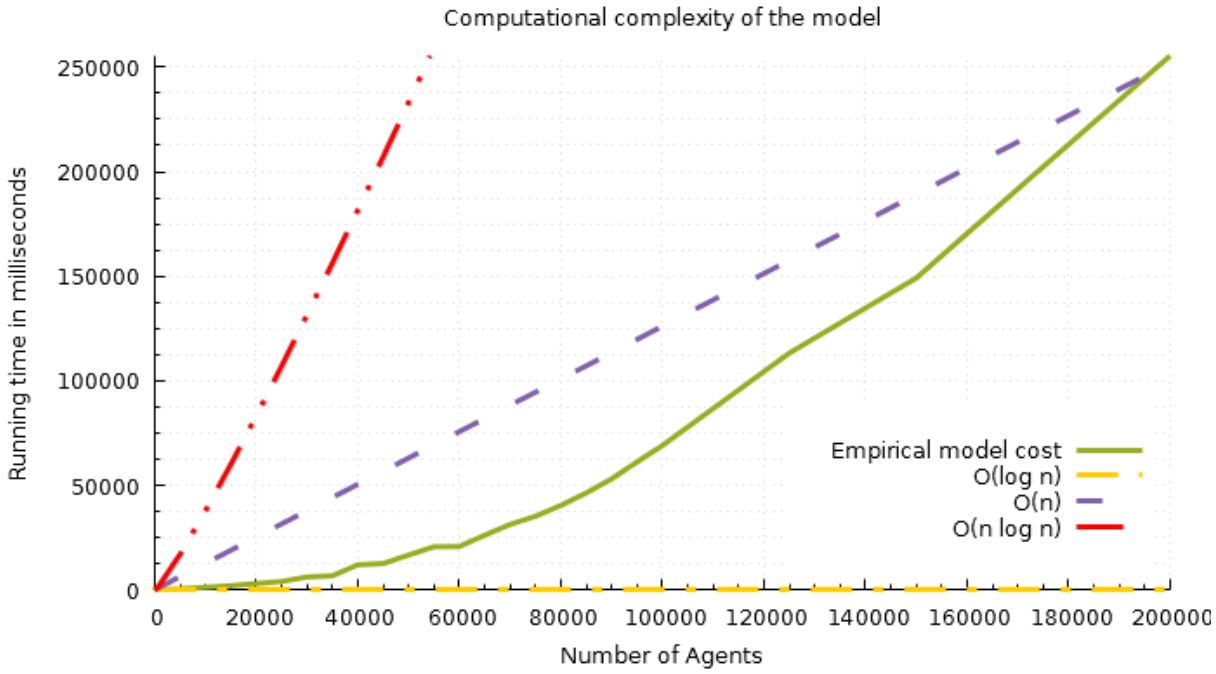


Figure 2: Analysis of the computational complexity of the model when increasing the number of agents in the population.

Parameters	**Calibrated values for the 44 parameters**			
	El Pais	**El Mundo**	**ABC**	**Cadena Ser**
Buzz $\{\tau, d\tau\}$	$\{1.4, 0.19\}$	$\{0.9, 0.11\}$	$\{1.4, 0.19\}$	$\{1.6, 0.13\}$
Influence $\{\Delta_{max}, \Delta, d\Delta\}$	$\{1.5, 0.79, 0.16\}$	$\{1.6, 0.78, 0.21\}$	$\{2.1, 0.81, 0.25\}$	$\{2.5, 0.94, 0.18\}$
	TVE	**Antena 3**	**Telecinco**	
Buzz $\{\tau, d\tau\}$	$\{1.8, 0.2\}$	$\{2.4, 0.18\}$	$\{0.7, 0.15\}$	
Influence $\{\delta_{max}, \Delta, d\Delta\}$	$\{4.2, 0.79, 0.18\}$	$\{3.6, 0.87, 0.14\}$	$\{3.8, 0.94, 0.15\}$	
	WOM process (PSOE, ABST., PP)			
Talking prob. $\{p(0)\}$	$\{0.04, 0.01, 0.02\}$			
Influence change $\{\Delta\}$	$\{0.19, 0.11, 0.06\}$			
Influence decay $\{d\Delta\}$	$\{0.27, 0.1, 0.38\}$			

Table 5: List of parameters automatically calibrated by the memetic algorithm.

Additionally, Table 5 shows the full list of 44 calibrated parameters using the memetic algorithm explained in Section 3.5 with their final values. After running the calibration method, we also analyzed the values for the parameters and evaluated them in order to have coherent setting for the model according to the available data and knowledge (i.e., first part of the validation process of the model).

In addition to the complete calibrated model, three validation scenarios are presented: one without mass media, another without WOM diffusion, and the last one with neither mass media nor WOM diffusion. These additional scenarios are variations of the complete model. This way we create new models where certain modules are disabled. By setting these scenarios, the designed model can be validated as a whole, facing its global behavior with respect to removing any of its main modules. This corresponds to hypothetical scenarios where some of the social interactions produced before the 14-M elections would have not been occurred. Thus, this process also enriches our social analysis as well as model validation. As in the case of the complete model, the pre-election opinion surveys are employed as the initial state of the simulation.

The scenario without mass media, called "No Media", disables media effect on the agents. Thus, mass media channels will neither influence agents neither increase buzz activity. In this scenario, only WOM diffusion is performed by agents through their social network. The scenario without WOM diffusion, referred as "No Diffusion", does not include agent diffusion through the social network. In this scenario, only mass media channels influence the agent population, but there is not any buzz effect generated from its impact. The last scenario, called "No Influence", does not include neither media effect on the agents nor diffusion through the social network. In this scenario, the agents' population is not exposed to any kind of influence, thus the elections results are directly those predicted by pre-election opinion surveys.

The comparison between the whole model and the three additional validation scenarios is shown in Table 6. These results are obtained averaging the results of 30 Monte-Carlo iterations of the ABM simulations. Percentage values represent SMAPE accuracy using final election results, scaling both simulated and real number of votes to the top third. This computation facilitates the understanding of the fitting results in a 0% to 100% scale.

In this Table 6, votes are displayed by party in the top block of the table and computed error is shown in the bottom block. Fitting results show good accuracy values for the complete model, displaying an accuracy value higher than 99%. This implies that the model is correctly simulating elections turnout. Observed errors also suggest that only WOM or mass media information in isolation are not enough to match final votes, and there is a need to use both modules in the model. In fact, the latter two scenarios have higher PP voters than the final results. That suggests that, when used in isolation, defined dynamics are not modeling voting turnout reality in an accurate way after the attacks. It also corroborates our initial assumption that both WOM and mass media information had significant influence on the 14-M election results.

	Real Data	Models			
Party	**Election Results**	**Complete**	**No Media**	**No Diffusion**	**No Influence**
Votes					
PSOE	11,026,163	11,020,144	10,329,618	10,259,170	9,941,145
PP	9,763,144	9,766,804	10,577,439	10,948,887	11,403,078
Abstention	8,449,355	8,451,711	8,331,602	8,030,602	7,894,438
% Total votes					
PSOE	37.71%	37.69%	35.33%	35.09%	34%
PP	33.39%	33.4%	36.18%	37.45%	39%
Abstention	28.9%	28.91%	28.5%	27.47%	27%
Global fitting		**99.13%**	**84.49%**	**77.06%**	**68.39%**

Table 6: Fitting values of the calibrated model and three additional model variation scenarios.

4.3. Analysis of the model's outputs

Once we have compared the model results with the real elections data, we will further evaluate the model behavior to ensure its validity. In the first place, the evolution of the averaged resilience μ of the population is displayed in Figure 3. Let us remind the reader resilience represents the amount of external influence needed to change its vote. For example, if an agent gets its resilience to a value between 3.3 and 6.7, it will abstain, even if it belongs to PP or PSOE voters segments. In order to compute these values, we average resilience for all the agents of each segment at each time-step. This evolution is stepped by hours, starting on March 11th at 08:00 am and finishing on March 14th at 08:00 am.

This chart presents stronger changes when news are on television. It corresponds to the prime time for news in Spain by that time. Additionally, the first simulation steps show more intense changes in the perceptions of the agents than the subsequent ones. The main responsible for this behavior is the polarization of the message transmitted by mass media. The content of this message, that was uniform at the start of the simulation, turns mixed at the end, resulting in a smoother curve. Because mass media exposure is not biased by segment, its impact over resilience evolution is similar (averaged resilience curves present a similar shape in the three segments). This evolution is slightly softer for the PP voters segment at the beginning of the simulation. This effect can also be observed at the end of the simulation, but for the opposite direction. PSOE voters change their perception smoother than before.

We also present the evolution of the resilience standard deviation in Figure 4 to show differences between segments. This chart shows Monte-Carlo variances as blurred areas. These curves are consistent with perception evolution, because deviation is increased when mass media exposure gets stronger. Deviation increases when agents are influenced by the framings. This can be observed between hours 0 and 20 for PSOE voters and abstainers, or from hour 30 for PP voters. In the former case, mass media channels are transmitting a message about the implication of ETA in the attacks. This increments resilience value, increasing PSOE and abstainers deviation, and reducing PP deviation. Because both PSOE voters and abstainers are more influenced than the PP voters during the first 20 hours of the simulation, its resilience deviation is similar (and higher) for that period. Eventually the messages from the framing blaming Al Qaeda outnumber the messages blaming ETA, which reduces PSOE voters and abstainers dispersion, and increases the dispersion of PP voters. This effect is more visible for abstainers and PP voters, whose deviation values get

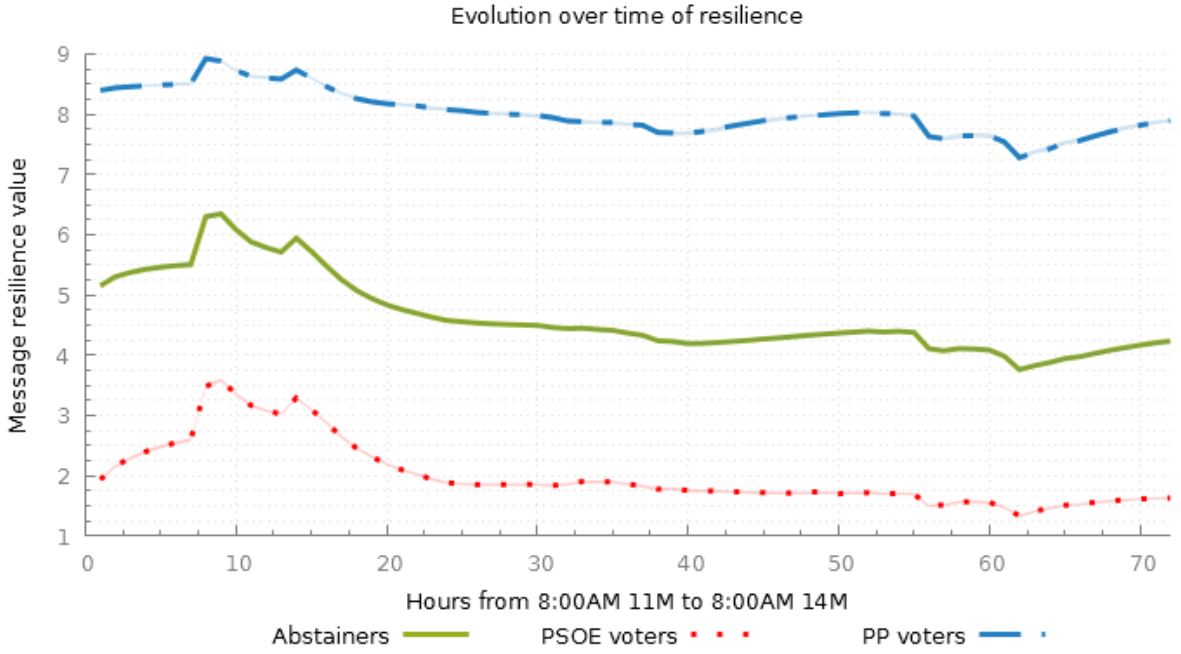


Figure 3: Averaged message resilience (μ) over time for all the agents.

a more similar shape as the simulation goes forward.

Figure 5 shows the evolution of votes by day, plotting a track of the voters intention every 24 hours. Votes track is done at 08:00 am every day. Tracks for the first day, i.e., 11-M, do not have deviation as they are collected when the simulation starts. On the second day, when the message content is still confusing, both PP and PSOE voters reduce their number while abstainers have a significant increase. The reduction is much larger for PSOE voters as ETA's authority is still likely. From that point, the messages are more focused and clear and both PP and abstainers reduce while PSOE voters significantly increase. Overall, box-plots of Figure 5 shows how PP decreases its votes in favor of PSOE and abstention along the simulation. This behavior is consistent with the surveys closer to the elections whose results suggested that the gap between PSOE and PP was reduced as the elections approached [38].

Finally, the WOM behavior is validated using two indicators: the number of conversations and the evolution of the messages' polarization. Figure 6 shows the percentage of conversations by step (also called WOM volume). In these values we can see that the highest buzz is achieved at prime time, just like in the resilience evolution chat. As happened in the deviation chart, blurred areas represent Monte-Carlo variations. The peaks shown in the number of conversations are heavily related with the buzz increment generated by the mass media. This way, the increment in the number of conversation is consistent, as the highest audience level is achieved during these time slots, enabling the media to generate its biggest buzz. Figure 7 shows the polarization of the conversations during the simulation. A positive polarization value (above 0) means that conversations are increasing the averaged μ value of the agents (moving it towards 10). Otherwise, a negative polarization suggests that the averaged μ value of the agents decreases (moving it towards 0). Net polarization is heavily influenced by the information transmitted by mass media during

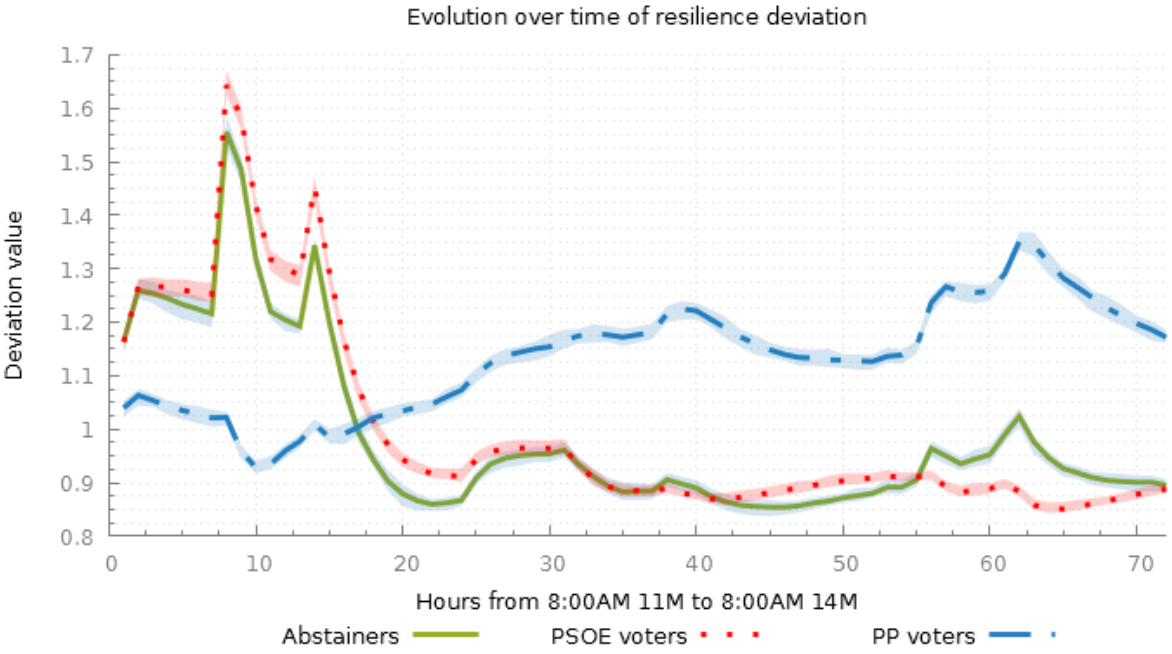


Figure 4: Averaged deviation of resilience (μ) for all the agents.

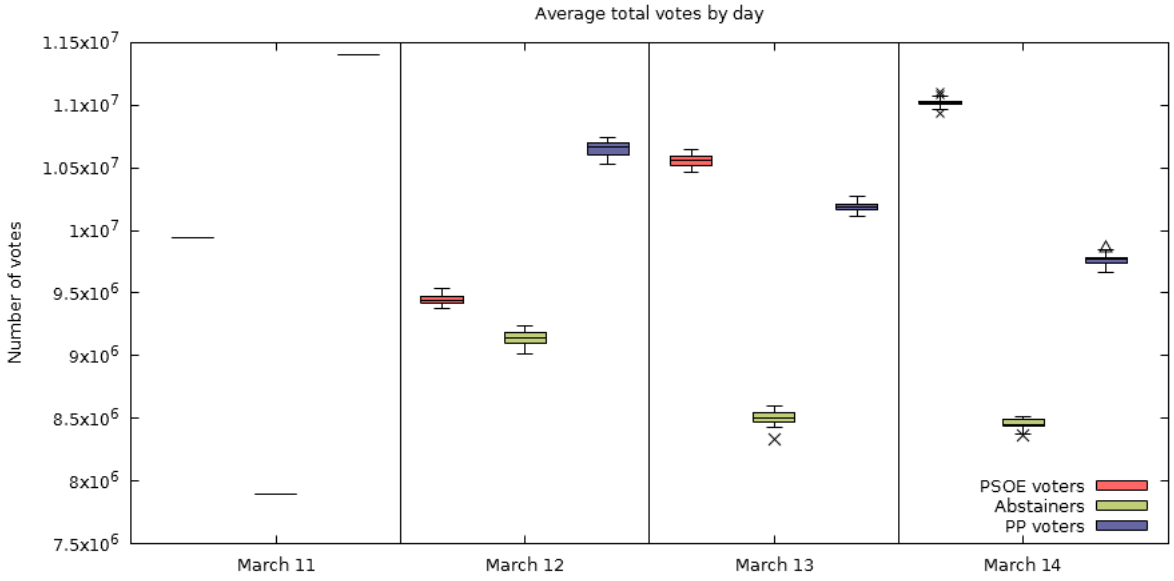


Figure 5: Average votes by political option for every day between the attacks and the elections.

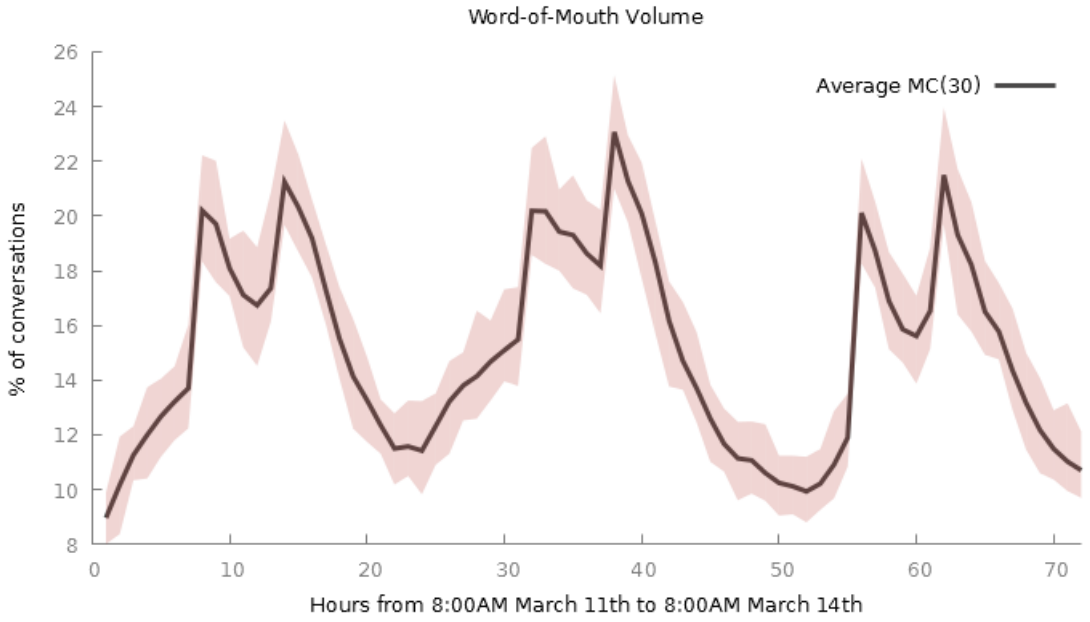


Figure 6: Percentage of conversations made by agents within their social network.

the simulation. As a result, this chart looks similar to the one displaying resilience evolution. The trend followed by polarization reaches its minimum value towards the end of the simulation, when the media information contains more messages blaming Al Qaeda. At the end of the simulation the message polarization is negative and therefore, there are more conversations regarding Al Qaeda's involvement in the attacks than conversations regarding ETA's authority. Because Al Qaeda's framing is defended by the PSOE party, this situation increases PSOE votes. This evolution was also observed in Figure 3 where we showed the message resilience over time for all the agents of the population.

5. Deployment of what-if political scenarios

We will analyze in this section different what-if scenarios using the previously validated model. Our study is mainly focused on two scenarios. First, Section 5.1 analyses WOM influence in the voters' segments. Then, changes on mass media messages are analyzed in Section 5.2.

5.1. WOM influence in the voters segments

This scenario is focused on the information spread through the social network. In order to study WOM influence, we perform a sensitivity analysis on the parameters which control the diffusion mechanisms. Those parameters are the talking probability ($p(t)$) and the parameter to generate the scale-free social network (m, see Section 3.3) which affects the social network density and the average degree of the agents. Both of them are increased when setting higher values for m and consequently, the speed of the diffusion process is higher. The sensitivity analysis is performed following the one-factor-at-a-time methodology [9] which modifies each parameter in an isolated way by keeping the rest of the parameters fixed to its original value.

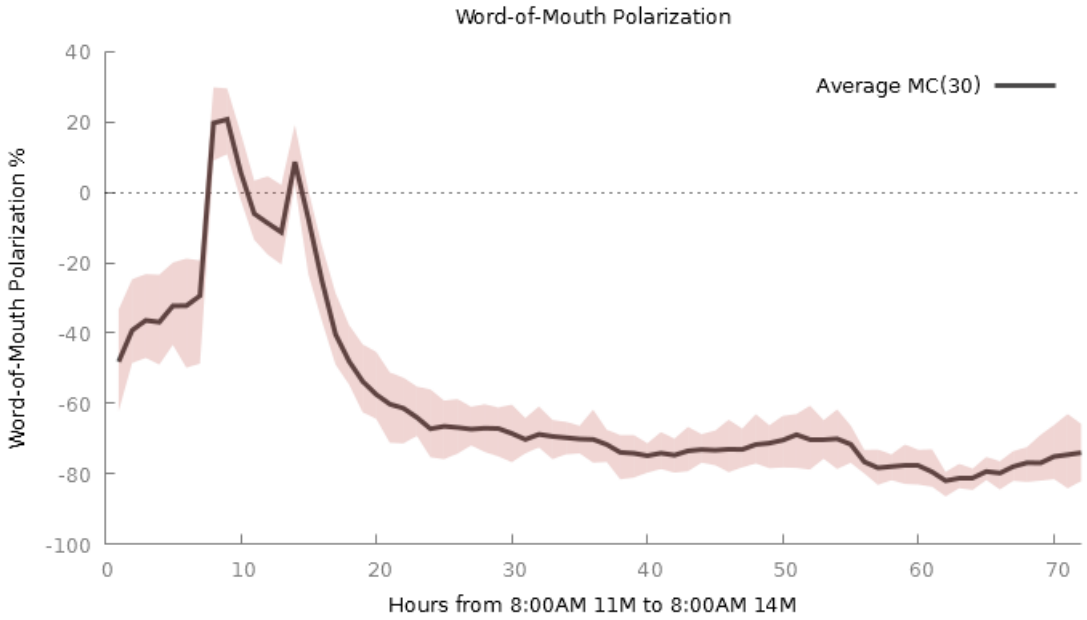
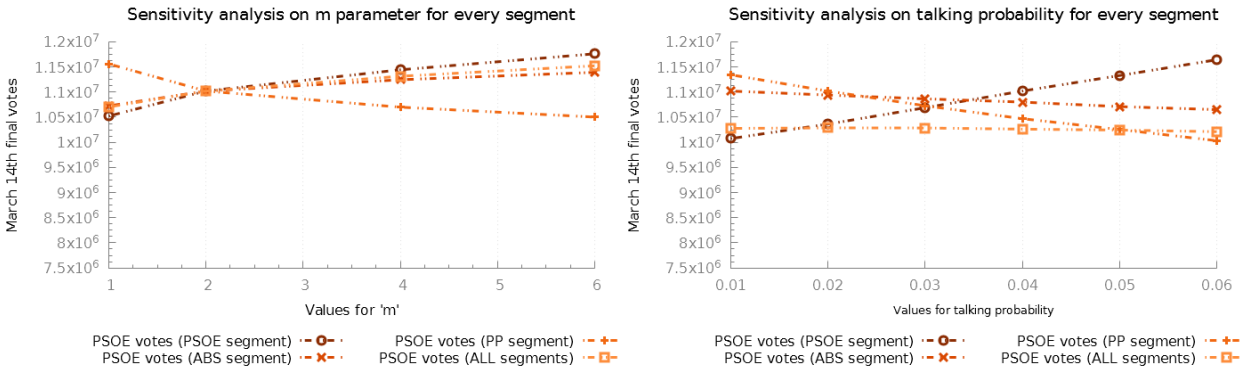


Figure 7: Net variation of the polarization of the message transmitted by the WOM process.

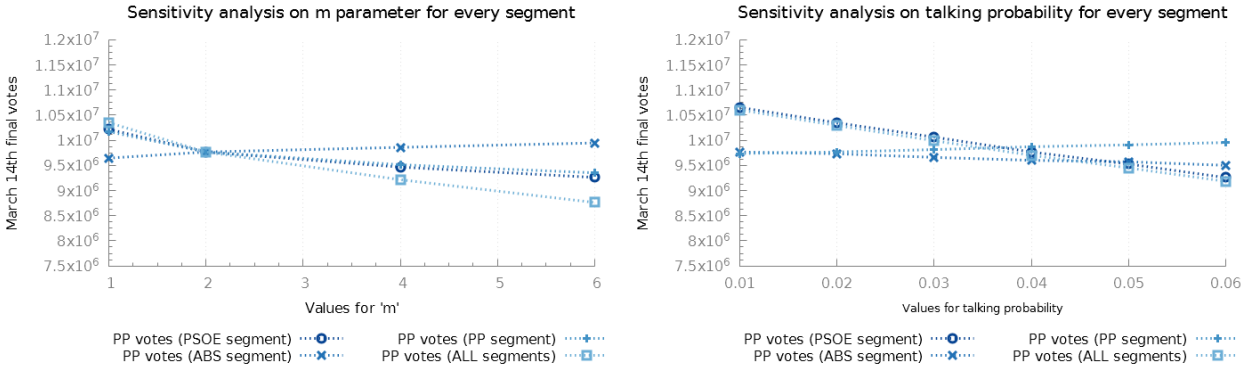
In the resulting charts of Figure 8 we show the model's response to changes to the talking probability and network connectivity variations as connected points even if they correspond to discrete values obtained from simulation runs for each specific parameters' configurations. We consider six different values for the talking probability, from 0.01 to 0.06, with a fixed increment step of 0.01. Another six values for m are studied, from 1 to 6, which correspond to network average degrees $\langle k \rangle$ of $2, 4, 6, 8, 10$, and 12 (we remind the reader that $\langle k \rangle = 2m$ in a scale-free network generated by the Barabasi-Albert algorithm). This experimentation is performed at both segment and global level. First, we will modify the values of m and the talking probability for each segment (chart lines formed by circles, crosses, and pluses). Later, we will do the same by modifying the parameters' values for all the segments at the same time (chart lines formed by boxes).

Regarding the variations over the PSOE voters segment, the results present strong changes on both PSOE and PP voters, keeping abstentions stable. It is also remarkable to notice that variations on talking probability produce linear variations for PSOE votes, keeping almost the same trend when increasing and decreasing. Variations over connectivity through parameter m behave slightly softer.

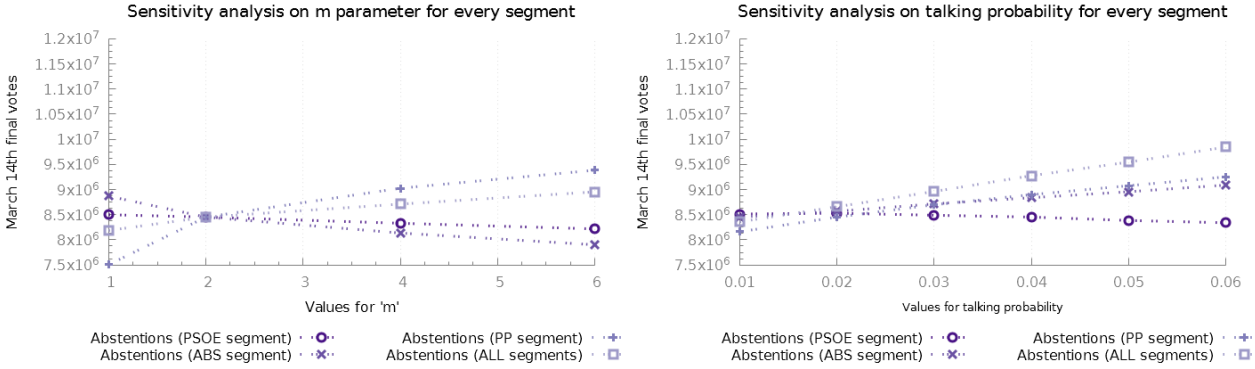
The variations over the PP voters segments show a different behavior from the other studied segments. When increasing/amplifying the values of the diffusion parameters, the three political options tend to obtain similar votes, reducing the difference between them. In addition, when decreasing the values of the diffusion parameters, the difference between the three political options rises. However, altering the PP voters segment parameters on either talking probability or m (i.e., network generation parameter) produce similar results. In the case of the talking probability parameter, a tilt of its value produces a transition between PSOE votes and abstainers, with a slower transition between abstainers and PP votes. In the case of the m parameter, increasing its value (and, as a consequence, increasing the speed of the diffusion process) only produces a growth in the number of abstentions, while PSOE and PP votes drop. For instance, when m is



(a) PSOE votes



(b) PP votes



(c) Abstentions

Figure 8: Sensitivity analysis for the talking probability and scale-free generation parameter m on each of the voters' segments.

greater than 5, PP party obtains a similar support to the abstention option. This behavior is also interesting because when the social network has a lower density (due to lower m values), both PP and PSOE obtain a higher number of votes. Instead of being favored by more connections and a faster diffusion, these political options are penalized in a more connected WOM scenario.

The results when altering the abstainers segment values show a more stable behavior for every

segment. However, we can notice that changes over talking probability slightly favor different diffusing messages than changes on m. On the one hand, increasing talking probability also increases abstentions, reducing votes for the other political alternatives in a similar way. On the other hand, increasing m reduces abstentions, specially increasing PSOE votes.

Figure 8 also shows the simulation results after changing talking probability and m for all the segments of the model at the same time. The study, previously applied to segments, is now performed to all the agents of the population at once. The results of this study are in line with those gathered when the variations were only applied to PSOE voters, but having some differences. Focusing on the PP response, its evolution follows a similar trend with respect to PSOE. Variations on the social network density (produced by altering the m parameter) produce a fall on PP votes while increasing abstentions and PSOE votes. While PSOE increment is similar to the one showed when altering PSOE values only, abstentions and PP variations are similar to the ones found when only altering PP values.

To sum up with this analysis, we can observe that the model's behavior and its reality fitting are sensible to changes on the WOM parameters. The three existing segments obtain different voting results when their parameters are modified. In the case of PSOE voters, this segment seems to have an important participation in the diffusion process because the number of votes for PSOE party changes more aggressively when modifying their own diffusion parameters, reaching its maximum and minimum values. In contrast, although its number of votes changes significantly when altering the diffusion parameters of the other segments, the effect is softer on them in comparison with the PSOE segment. With respect to abstainers, modifying its parameters has a relatively small effect on the other segments that could suggest a secondary role in the diffusion process. Finally, in the case of PP voters, these results show a fall in the number of PP votes for most scenarios when diffusion increases. This fact suggests that PP segment cannot influence the other segments even when PP's diffusion parameters have a high value. This could also suggest that message polarization penalizes extreme values when having a highly connected network.

5.2. Content changes on the mass media messages

These scenarios are focused on the polarization of the message transmitted by mass media channels. As previously explained, message polarization from March 11th to March 14th gradually changed from pointing ETA to pointing Al Qaeda as long as new insights were progressively known from the developed investigation. Using the original polarization as a reference, we perform a sensitivity analysis over the message content transmitted by each mass media channel.

Instead of modifying a single parameter, a group of parameters are changed for each scenario [9]. For each mass media category (press, radio, and television), its message polarization is modified towards ETA and towards Al Qaeda. These polarization variations are applied to all the mass media channels and applied to every information transmitted by those mass media channels contained in each category.

The results of this study are shown by political option in Figures 9, 10, and 11. For each option, the evolution of votes is displayed regarding the amount of modified polarization. This way, results obtained with the original message are placed at 0 in the x-axis. This polarization is gradually increased from 0 to 1 and from 0 to -1. Both extremes represent full message content towards Al Qaeda (-1) or ETA (1).

Figure 9 shows the results of the polarization variations for the number of abstentions. Resulting abstentions are displayed by mass media categories: press, radio, television, and all of them together. These results (and those for the other political options) clearly highlight television as

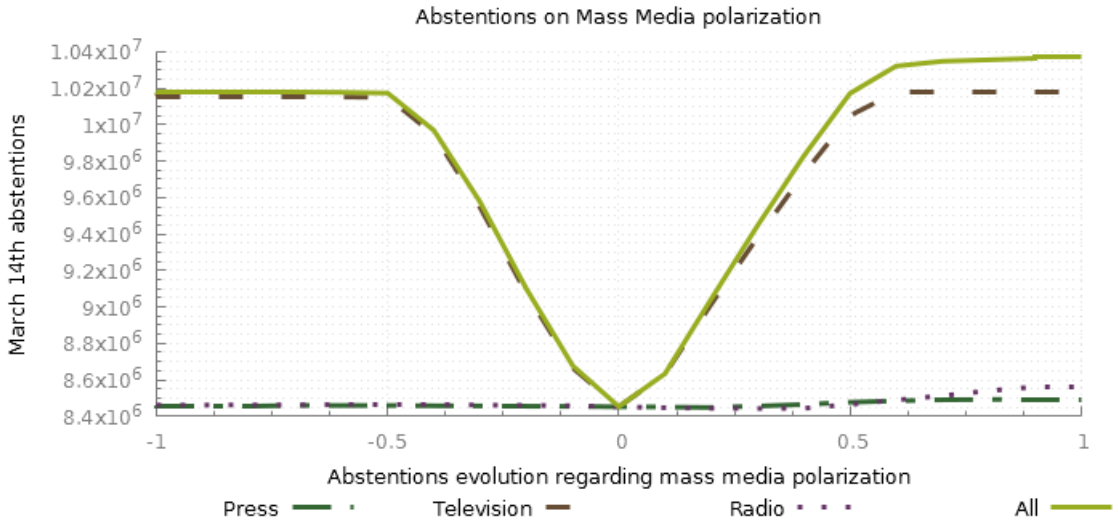


Figure 9: Abstentions resulting from variations on the polarization of mass media messages.

the most influencing media channel. Additionally, the joint effect of all the mass media channels seems interesting because its maximum result surpasses individual categories.

This chart is also interesting because its non-symmetric shape. If message polarization is strongly moved towards Al Qaeda (-1 variation), simulation results collapse quickly. There is almost no change in the number of abstentions when message is pushed beyond -0.5. On the opposite, when message is pushed towards ETA (+1 variation), the model's results saturate around 0.9. In the case of television, its results saturate before the rest of the media categories. These results suggest that the number of abstentions is more sensitive to stronger messages towards ETA.

Figure 10 shows the results of the polarization variations for the PP voting results. Again in this case, television polarization achieves the highest change. Moreover, this chart also shows a non-symmetric shape, where all categories saturate around -0.5. Most mass media channels have its maximum number of votes close to the maximum influence towards ETA (+1 variation).

We can also notice the amount of votes achieved by increasing the polarization towards ETA is relatively small when compared with the amount of votes lost when decreasing the polarization towards Al Qaeda (-1 variation). This effect may be caused by the original message of some channels like television channels which changed their message content during the three days period moving from one framing to the other. Results then show that the number of PP votes is more influenced by polarization towards Al Qaeda than polarization towards ETA.

Finally, Figure 11 shows the results of the polarization variations for PSOE. As happened with previous political options, television is the channel that causes the highest change. In addition, the polarization variations toward ETA involve significant fluctuations. These results also suggest a resilient behavior regarding polarization when pushed to Al Qaeda. In fact, the amount of increased votes is small compared to the amount of votes lost when polarization is pushed towards ETA. This drop is the highest one of all the political options.

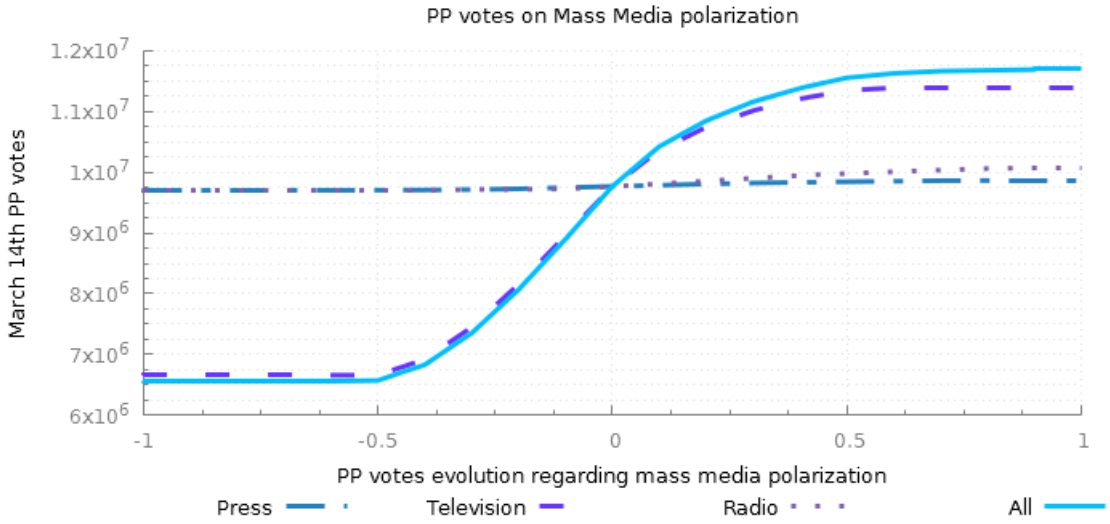


Figure 10: PP votes resulting from variations on the polarization of mass media messages.

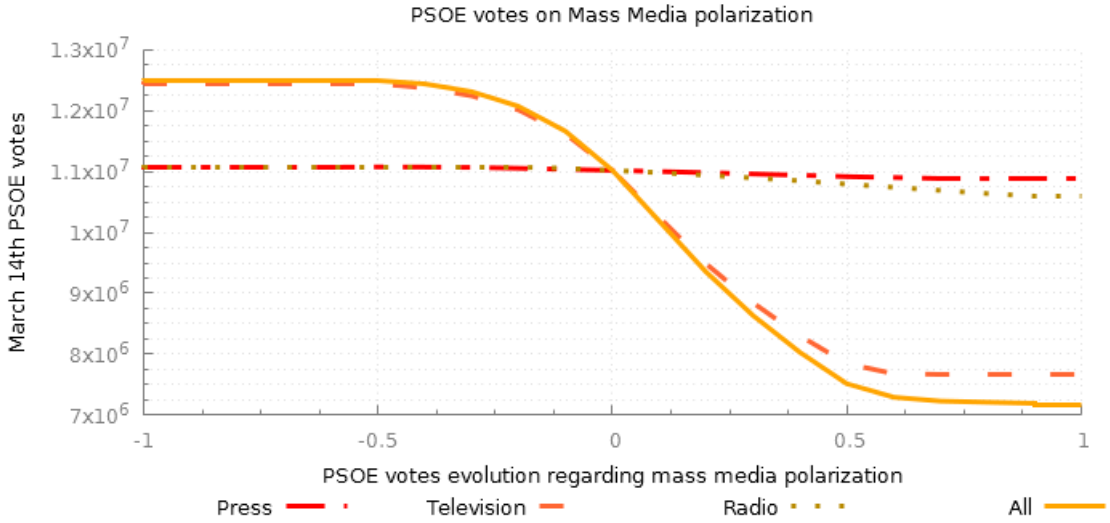


Figure 11: PSOE votes resulting from variations on the polarization of mass media messages.

6. Final remarks

In this paper we have presented a study on the framing effect during the 2004 Spanish elections following the 11-M attacks. We have designed and implemented an ABM simulation to replicate electors' behavior into artificial voting agents connected by a social network and influenced by the most significant Spanish mass media messages. Our model recreated the environmental conditions from 11-M to 14-M from the mass media information point of view. We calibrated and validated the model by achieving a model fitting of 99.13% and employing different validation cases of the modules in the ABM framework.

The results of the experiments suggest that the framing effect could actually influence the election results by both mobilizing abstainers and deactivating voters from PP party. Other important conclusions of the model's results in our what-if scenarios are:

- Diffusion mechanisms have an important role during this period because a significant swap of votes arises when modifying the density of the social network and the dynamics of the WOM process (i.e., defined by the social network generation parameter m and the agents' talking probability). This conclusion seems consistent with other works regarding diffusion of political beliefs [63, 68, 76]. Moreover, the experiments we present in this paper show that the social network have a key role when exposing an agent population to highly polarized messages.

- The swap of votes through diffusion does not seem to follow linear increments. Instead, some political options achieve votes when increasing its diffusion rates, but others maintain or barely increase its number of votes. PSOE is the most influenced political option by WOM diffusion.

- In the same way, the diffusion of polarized messages using mass media communications does not produce linear changes of votes. On the one hand, the number of votes for every party reaches its maximum rather quickly when the transmitted message is polarized towards Al Qaeda by not responding to strategies with a stronger message. On the other hand, the number of votes changes smoothly when polarizing the message towards ETA. This behavior suggests a higher sensitivity of the model when messages are polarized towards Al Qaeda.

- Aggregating mass media channels seems to achieve stronger effects than the addition of those channels applied individually. Even if television is clearly distinguished as the most influencing channel, radio and press increase the aggregated media effect remarkably. This also corroborates that not only television channels had an important role in the diffusion of the 11-M events but the presence of other media channels were decisive for the elections turnout.

Future work will be focused on studying alternative strategies for modeling more complex social network diffusion and voting behaviors for the agents of the political scenario. These more advanced mechanics could involve using fuzzy logic [75] for modeling voter's decision making with linguistic information as done in Chica et al. [13]. Also, one could combine framing effects with the individual political positioning of the voters.

Acknowledgments

This work is supported by Spanish Ministerio de Economía y Competitividad under the NEWSOCO project (ref. TIN2015-67661-P), including European Regional Development Funds (ERDF).

7. References

[1] Asociación para la Investigación de Medios de Comunicación (AIMC), 2004. AIMC February to November 2004. General Recap. EMG. `http://www.aimc.es/-Datos-EGM-Resumen-General-.html`. Accessed: 2016-03-28.
[2] Back, T., Fogel, D.B., Michalewicz, Z., 1997. Handbook of evolutionary computation. IOP Publishing Ltd., Bristol (UK).
[3] Bali, V.A., 2007. Terror and elections: Lessons from Spain. Electoral Studies 26, 669–687.
[4] Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
[5] Bass, F.M., 1969. A new product growth for model consumer durables. Management Science 15, 215–227.
[6] Berelson, B., 1954. Voting; a study of opinion formation in a presidential campaign. University of Chicago Press, Chicago.
[7] Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99, 7280–7287.
[8] Borah, P., 2011. Conceptual issues in framing theory: A systematic examination of a decade's literature. Journal of communication 61, 246–263.
[9] ten Broeke, G., van Voorn, G., Ligtenberg, A., 2016. Which sensitivity analysis method should I use for my agent-based model? Journal of Artificial Societies & Social Simulation 19, 5.
[10] Centro de Investigaciones Sociológicas, 2004. Estudio 2555. CIS Data bank.
[11] Chari, R., 2004. The 2004 Spanish election: terrorism as a catalyst for change? West European Politics 27, 954–963.
[12] Chica, M., Barranquero, J., Kajdanowicz, T., Cordón, O., Damas, S., 2017. Multimodal optimization: an effective framework for model calibration. Information Sciences 375, 79–97.
[13] Chica, M., Cordón, O., Damas, S., Iglesias, V., Mingot, J., 2016. Identimod: modeling and managing brand value using soft computing. Decision Support Systems 89, 41–55.
[14] Chica, M., Rand, W., 2016. Building agent-based decision support systems for word-of-mouth programs. a freemium application. Journal of Marketing Research (in press) .
[15] Chong, D., Druckman, J.N., 2007a. Framing public opinion in competitive democracies. American Political Science Review 101, 637–655.
[16] Chong, D., Druckman, J.N., 2007b. Framing theory. Annual Review of Political Science 10, 103–126.
[17] Martínez i Coma, F., 2008. ¿ Por qué importan las campañas electorales? (in Spanish). Centro de Investigaciones Sociológicas.
[18] Consell de l'Audiovisual de Catalunya, 2004. 11-14M: la construcció televisiva (in catalonian). Quaderns del CAC , 89–204.
[19] Druckman, J.N., Peterson, E., Slothuus, R., 2013. How elite partisan polarization affects public opinion formation. American Political Science Review 107, 57–79.
[20] Duzevik, D., Anev, A., Funes, P., Gaudiano, P., 2007. The effects of word-of-mouth: An agent-based simulation of interpersonal influence in social networks. Word of Mouth Research Symposium .
[21] Edling, C., Rydgren, J., Sandell, R., 2016. Terrorism, belief formation, and residential integration population dynamics in the aftermath of the 2004 Madrid terror bombings. American Behavioral Scientist , in press. DOI:0002764216643127.
[22] Entman, R.M., 1993. Framing: Toward clarification of a fractured paradigm. Journal of communication 43, 51–58.
[23] Epstein, J.M., 2006. Generative social science: Studies in agent-based computational modeling. Princeton University Press.
[24] Farrell, D.M., 2011. Electoral systems: a comparative introduction. Palgrave macmillan.
[25] Fearon, J.D., 1991. Counterfactuals and hypothesis testing in political science. World politics 43, 169–195.
[26] Fieldhouse, E., Lessard-Phillips, L., Edmonds, B., Crossley, N., 2005. The complexity of turnout: An agent-based simulation of turnout cascades. Electoral Studies 24.
[27] Gao, L., Durnota, B., Ding, Y., Dai, H., 2012. An agent-based simulation system for evaluating gridding urban management strategies. Knowledge-Based Systems 26, 174–184.

[28] Goldberg, D.E., Holland, J.H., 1988. Genetic algorithms and machine learning. Machine Learning 3, 95–99.
[29] González-Avella, J.C., Cosenza, M.G., Klemm, K., Eguíluz, V.M., San Miguel, M., 2007. Information feedback and mass media effects in cultural dynamics. Journal of Artificial Societies and Social Simulation 10, 9.
[30] Hänggli, R., Kriesi, H., 2010. Political framing strategies and their impact on media framing in a Swiss direct-democratic campaign. Political Communication 27, 141–157.
[31] Herrera, F., Lozano, M., Verdegay, J.L., 1998. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. Artificial Intelligence Review 12, 265–319.
[32] Higgins, E.T., 1996. Knowledge activation: Accessibility, applicability, and salience. Guilford Publications.
[33] Holbrook, T., 1996. Do campaigns matter? Contemporary American Politics. SAGE publications.
[34] Janssen, M.A., Ostrom, E., 2006. Empirically based, agent-based models. Ecology and Society 11, 37.
[35] Kollman, K., Miller, J.H., Page, S.E., 1992. Adaptive parties in spatial elections. American Political Science Review 86, 929–937.
[36] Kollman, K., Miller, J.H., Page, S.E., 1998. Political parties and electoral landscapes. British Journal of Political Science 28, 139–158.
[37] Kuhn, J.R., Courtney, J.F., Morris, B., Tatara, E.R., 2010. Agent-based analysis and simulation of the consumer airline market share for Frontier Airlines. Knowledge-Based Systems 23, 875–882.
[38] Lago, I., Montero, J.R., 2006. The 2004 election in Spain: Terrorism, accountability, and voting. Institut de Ciències Polítiques i Socials, Barcelona.
[39] Lago Peñas, I., Montero, J.R., Torcal, M., 2005. Del 11-M al 14-M: Los mecanismos del cambio electoral (in Spanish). Claves de la razón práctica 149, 36–45.
[40] Laver, M., 2005. Policy and the dynamics of political competition. American Political Science Review 99, 263–281.
[41] Lazarsfeld, P.F., Berelson, B., Gaudet, H., 1965. The people's choice; how the voter makes up his mind in a presidential campaign. Columbia Univ. Press, New York.
[42] Libai, B., Muller, E., Peres, R., 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. Journal of Marketing Research 50, 161–176.
[43] Liu, F.C., 2007. Constrained opinion leader influence in an electoral campaign season: Revisiting the two-step flow theory with multi-agent simulation. Advances in Complex Systems 10, 233–250.
[44] López García, G., 2004. El 11-M y el consumo de medios de comunicación (in Spanish). Sala de Prensa 71.
[45] Macal, C.M., North, M.J., 2005. Tutorial on agent-based modeling and simulation, in: Proceedings of the 37th conference on Winter simulation, ACM. pp. 2–15.
[46] Martínez-Miranda, J., Pavón, J., 2012. Modeling the influence of trust on work team performance. Simulation 88, 408–436.
[47] Mckeown, G., Sheehy, N., 2006. Mass media and polarisation processes in the bounded confidence model of opinion dynamics. Journal of Artificial Societies and Social Simulation 9, 11–42.
[48] Miceli, M.S., 2005. Morality politics vs. identity politics: Framing processes and competition among christian right and gay social movement organizations, in: Sociological Forum, Springer. pp. 589–612.
[49] Miller, J.H., 1998. Active nonlinear tests (ANTs) of complex simulation models. Management Science 44, 820–830.
[50] Moscato, P., 1989. On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Technical Report 826. Caltech Concurrent Computation Program. Pasadena, USA.
[51] Muis, J., 2010. Simulating political stability and change in the netherlands (1998-2002): an agent-based model of party competition with media effects empirically tested. Journal of Artificial Societies and Social Simulation 13, 4.
[52] Newman, M., Barabási, A.L., Watts, D.J., 2006. The structure and dynamics of networks. Princeton University Press.
[53] Observatorio Político Autonómico, 2004. Resultados encuesta postelectoral elecciones generales 14 de marzo de 2004 (in Spanish).
[54] Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research 151, 552–568.
[55] Olmeda, J.A., 2005. Fear or falsehood? Framing the 3/11 terrorist attacks in Madrid and electoral accountability. Boletín Elcano .
[56] Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. Physical Review Letters 86, 3200.
[57] Pennings, P., Keman, H., 2003. The Dutch parliamentary elections in 2002 and 2003: The rise and decline of the Fortuyn movement. Acta Politica 38, 51–68.

[58] Rand, W., Rust, R.T., 2011. Agent-based modeling in marketing: Guidelines for rigor. International Journal of Research in Marketing 28, 181–193.
[59] i Rigo, E.O., 2005. Aznar's political failure or punishment for supporting the Iraq war? Hypotheses about the causes of the 2004 Spanish election results. American Behavioral Scientist 49, 610–615.
[60] Rogers, E.M., 2010. Diffusion of innovations. Simon and Schuster.
[61] Roozmand, O., Ghasem-Aghaee, N., Hofstede, G.J., Nematbakhsh, M.A., Baraani, A., Verwaart, T., 2011. Agent-based modeling of consumer decision making process based on power distance and personality. Knowledge-Based Systems 24, 1075–1095.
[62] Sargent, R.G., 2005. Verification and validation of simulation models, in: Proceedings of the 37th conference on Winter simulation, pp. 130–143.
[63] Singh, V.K., Basak, S., Modanwal, N., 2011. Agent based modeling of individual voting preferences with social influence, in: Trends in Computer Science, Engineering and Information Technology. Springer, pp. 542–552.
[64] Slothuus, R., de Vreese, C.H., 2010. Political parties, motivated reasoning, and issue framing effects. The Journal of Politics 72, 630–645.
[65] Sorauf, F.F.J., Beck, P.A., 1988. Party politics in America. Scott Foresman.
[66] Stapel, D.A., Koomen, W., Zeelenberg, M., 1998. The impact of accuracy motivation on interpretation, comparison, and correction processes: Accuracy × knowledge accessibility effects. Journal of Personality and Social Psychology 74, 878–893.
[67] Stonedahl, F., Rand, W., 2014. When does simulated data match real data? Comparing model calibration functions using genetic algorithms, in: Advances in Computational Social Science. Springer, Japan. volume 11 of *Agent-Based Social Systems*, pp. 297–313.
[68] Sudo, Y., Kato, S., Mutoh, A., 2013. The impact of exchanging opinions in political decision-making on voting by using multi-agent simulation, in: PRIMA 2013: Principles and Practice of Multi-Agent Systems. Springer, pp. 340–354.
[69] Talbi, E.G., 2009. Metaheuristics: from design to implementation. John Wiley & Sons.
[70] Torcal, M., Rico, G., 2004. The 2004 Spanish General Election: In the Shadow of Al Qaeda. South European Society and Politics 9, 107–121.
[71] Watts, D.J., Dodds, P.S., 2007. Influentials, networks, and public opinion formation. Journal of Consumer Research 34, 441–458.
[72] Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.
[73] Wu, F., Huberman, B.A., 2007. Novelty and collective attention. Proceedings of the National Academy of Sciences 104, 17599–17601.
[74] Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 599–608.
[75] Zadeh, L.A., 1965. Fuzzy sets. Information and control 8, 338–353.
[76] Zakaria, N., 2014. Modeling political belief and its propagation, with Malaysia as a driving context. Open Journal of Political Science 4, 58–75.
[77] Zhang, X., Zhu, J., Wang, Q., Zhao, H., 2013. Identifying influential nodes in complex networks with community structure. Knowledge-Based Systems 42, 74 – 84.

2 Simulating the influence of terror management strategies on the voter ideological distance using agent-based modeling

- I. Moya, M. Chica, J. L. Sáez-Lozano, O. Cordón. Simulating the influence of terror management strategies on the voter ideological distance using agent-based modeling, Telematics and Informatics, vol. 63, 101656, 2021. DOI: 10.1016/j.tele.2021.101656.

  - State: Published.
  - Impact Factor (JCR 2020): 6.182.
  - Category: INFORMATION SCIENCE & LIBRARY SCIENCE - SSCI. Order: 11/86. Q1.

Simulating the influence of terror management strategies on the voter ideological distance using agent-based modeling

Ignacio Moya*,[a], Manuel Chica[a,b], José L. Sáez-Lozano[c], Óscar Cordón[a]

[a] *Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, 18071, Granada, Spain*
[b] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia*
[c] *Department of International and Spanish Economics and GIADE-UGR, University of Granada, 18071 Granada, Spain*

Abstract

This paper simulates the effect of the strategies implemented by politicians after the terrorists attacks in Madrid on 11 March 2004 on the ideological distance between voters and political parties. The attacks took place three days before the elections and changed the campaign's agenda, which centered around the issue of who was responsible for the attack: ETA or Al Qaeda. It also altered the agenda of the mass media, which focused its informative activity on broadcasting news related to this issue. We did an exhaustive selection process of all the news broadcast on television, radio, and newspapers that made reference to the authorship of the attack. Using these messages we developed an agent-based model for explaining how the political strategies implemented by political parties influenced the ideological distance. The proposed model is based on the ideological proximity model by Downs (1957). After calibrating and validating the model with real data, we simulated the effect of three political strategies from the *theory of terror management* on the ideological distance between voters and political parties: the rally around the flag, the opinion leadership, and the priming theory of public opinion and media coverage. The results show that these strategies have a significant and stable impact on the ideological distance. In particular, *the rally around the flag* can have a lasting effect, capable of changing the ideological distance in the short term after a terrorist attack.

Keywords— Spatial theory of voting, terrorist attack, terror management, agent-based modeling.

1. Introduction

Political science shows that terrorism in general, and more particularly terrorist attacks close to elections, have an impact on the vote and campaign strategy of political parties (Berrebi and Klor, 2007, 2008; Fishman, 2005; Randahl, 2018). The attack on 11 September 2001 in the United States is one of the best known and most studied cases. The literature shows that this terrorist attack influenced the perception that public opinion had of George W. Bush's management and

*Corresponding author
Email addresses: imoya@ugr.es (Ignacio Moya), manuelchica@ugr.es (Manuel Chica), josaez@ugr.es (José L. Sáez-Lozano), ocordon@decsai.ugr.es (Óscar Cordón)

the electoral support of the Republican Party (Randahl, 2018). The elections on 14 March 2004 (14-M) in Spain also constitute a relevant milestone in the history of electoral behavior and the management of terror by political parties due to the attacks on the 11th of March in Madrid (11-M). Bali (2007) identified two events, which turned this electoral process into something extraordinary. On the one hand, there was a terrorist attack that caused around 200 deaths and 2000 wounded. On the other hand, there was a great mobilization of public opinion during the three days that preceded the elections that influenced the electoral result. After the attacks the protagonists of the electoral campaign were the political parties, the media, and public opinion. The agenda focused on the issue of the authorship of the attack: ETA or Al Qaeda.

Both the People's Party (PP) (i.e., the right-wing party in the government) and the main opposition parties, the Spanish Socialist Workers' Party (PSOE) and the United Left (IU), implemented a political strategy aimed at influencing the vote decision. The PP claimed that the attacks were perpetrated by ETA, while the PSOE and IU blamed Al Qaeda. The strategies of the parties are framed within the *theory of terror management*, which is a line of research where several studies have been published suggesting that a violent political conflict can have a significant impact on the public opinion or even decide the election results (Bali, 2007; Randahl, 2018; Robbins et al., 2013; Rose et al., 2007). The mass media were the transmission channel used by political parties to broadcast their political strategy, as there was a high demand for news from the public after the attacks (Piolatto and Schuett, 2015). There are two major strands in the literature on the role of the media in democratic processes: studies that show that the media is a relevant actor because it offers information to voters to guide their voting decision, and the theories of the agenda setting, priming, and framing. The latter postulate that the mass media promote propaganda aimed at exploiting the cognitive errors committed by the voters (Strömberg, 2015). Specifically, previous studies from the framing theory (Moya et al., 2017) suggested that the mass media influenced the reorientation of the voting intention in the 14-M elections.

Research on the electoral behavior in the 14-M elections agree on the fact that the terrorist attack influenced the public opinion (Bali, 2007; Lago and Montero, 2006; Michavila, 2005; Montalvo, 2011; Montero and Lago, 2009; Moya et al., 2017; Olmeda, 2005; Rose et al., 2007; Torcal and Rico, 2004). They conclude that the attack contributed to some voters reorienting their voting intention from different scientific paradigms. However, none of these studies analyzes electoral behavior from the perspective of the spatial theory of voting, which assumes that voters and political parties are located on a bipolar continuum that reflects their positions on a political issue (Downs, 1957; Enelow and Hinich, 1984, 1994; González and Granic, 2020; Grofman, 1985; Kedar, 2005, 2009; Rabinowitz and Macdonald, 1989). According to this theory, voters will choose the political party that is closer in this one-dimensional space. In response, the political parties are driven by a utilitarian logic that encourages them to position themselves in a position that minimizes the distance with each and every one of the voters. Therefore, the spatial theory of voting justifies the relevant role of the mass media in the electoral process (Chan and Suen, 2008; Duggan and Martinelli, 2011; Strömberg, 2004), since they are the channel used by the political parties for transmitting their messages and attracting the attention of voters, who demand information (Strömberg, 2015).

We propose to study the influence of the political strategies implemented by political parties on the ideological distance. The goal is to analyze whether after the terrorist attack, the news related to the issue of the authorship of the 11-M attack changed the ideological distance between voters and political parties. We carry out our analysis by using agent-based modeling (ABM) (Bonabeau, 2002; de Holanda et al., 2008; de Marchi and Page, 2014; Epstein, 2006; Lee et al., 2020; Macal and

North, 2005; Wilensky and Rand, 2015). ABM is a powerful methodology commonly employed for analyzing complex and emergent problems that has been successfully applied to analyze political scenarios (Laver, 2005; Liu, 2007; Moya et al., 2017; Muis, 2010). ABM relies on a set of autonomous entities called agents that behave according to simple rules and interacting with other agents. The aggregation of both the agents' individual actions and their interactions allows the modelers to reproduce complex and dynamic behaviors which would be difficult to model using a top-down approach. Hence, we can study how the management of the crisis, spread by the mass media, influences the ideological distance of a set of artificial voter agents with respect to IU, PP, and PSOE, which were the main three parties in Spain in 2004. Mass media information during these 72 hours is reproduced by our ABM simulation considering real tracking data and by including the main broadcast media at that time: radio, television, and written press. In addition, our ABM considers an artificial social network (Barabási and Albert, 1999; Watts and Strogatz, 1998) for reproducing the word-of-mouth (WOM) (Chica and Rand, 2017; Libai et al., 2013) interactions of the voter agents.

Our model is calibrated and validated using the data from the pre-electoral and the post-electoral surveys (Spanish Centre for Sociological Research, 2004. Estudio 2555. CIS Data bank). Thus, those model parameters whose value cannot be adjusted manually with the available information are set so the distance values of the agents fit the data surveyed after the elections. This calibration process is carried out using automatic calibration (Chica and Rand, 2017; Moya et al., 2019; Rand and Rust, 2011) based on an optimization method and a deviation function with respect to the historical data. The unknown parameters are mainly those regulating the influence of mass media messages and WOM interactions in the agents. Specifically, our automatic calibration approach considers a memetic algorithm (Moscato, 1989) (i.e., a bio-inspired metaheuristic) that comprises a steady-state genetic algorithm and local search refinement (Back et al., 1997; Moya et al., 2017).

Once calibrated and validated, we analyze how the management of crises like the 11-M attacks could influence the ideological distances between voters and political parties. Therefore, using the calibrated model we define three what-if scenarios selected from the *theory of terror management* (Landau et al., 2004; Willer and Adams, 2008) where we can analyze the evolution of the distance values of the voters during the simulation. These scenarios modify the message of the mass media channels so they can reproduce different effects: *the rally around the flag* (Mueller, 1973), *the opinion leadership* (Chowanietz, 2011), and *the priming of public opinion and media coverage* (Brody and Shapiro, 1989). These scenarios are selected because they are the most relevant to the events occurring in the 2004 Spanish elections.

This contribution is structured as follows. Section 2 reviews the background on the spatial theory of voting, the theory of terror management, and the related ABMs developed according to the spatial theory of voting. Then, Section 3 introduces the specifics of our ABM for ideological distances. The calibration and validation of the model using real data is presented in Section 4. Section 5 develops the simulation of three political scenarios from the theory of terror management. Finally, Section 6 depicts our final remarks and main conclusions of the study.

2. Background

2.1. Theory of terror management and ideological distance

Originally, the *theory of terror management* focused on the analysis of the efficiency of government policies to combat terrorism since it conditioned the perception that citizens had on their

government's management (Willer and Adams, 2008). After the military operations in Iraq and Afghanistan, the terrorist attacks of 9/11 in the US, and the 11-M in Spain, several studies showed that the political division regarding terrorism has increased its influence on public opinion. Within the *theory of terror management*, a new line of research has emerged analyzing both the relevance of terror and terrorism in the perception of voters for those leaders or parties that manage the defense of the country (Landau et al., 2004).

There are three main approaches explaining the mechanisms that operate in the relationship between terrorism and public opinion: *the rally around the flag*, *the opinion leadership*, and *the priming of public opinion and media coverage*, and all of them will be tackled by the current contribution. First, Mueller defines three criteria for generating a rally effect: "*i) it is international, ii) involves the United States and particularly the president directly; and iii) specific, dramatic, and sharply focused*" (Mueller, 1973). We can see that the 11-M attacks fit these criteria, and therefore should have generated a rally around the flag effect during the 2004 Spanish elections. This effect indicates that the 11-M attacks put the government of the PP in the focus of attention of national public opinion. Therefore, these attacks should have generated an increase in the PP's voting intention because the government issued several messages guided at generating a patriotic reaction in order to induce the majority of the voters to vote for the PP as the government party was then the most confident to protect the country from a terrorist attack.

In contrast with *the rally around the flag*, several authors proposed *the opinion leadership* paradigm (Baker and Oneal, 2001; Hetherington and Nelson, 2003; Colaresi, 2007; Chowanietz, 2011). According to this paradigm, in front of an event such as the terrorist attacks on 11-M, the main leaders in the political opposition should engage in criticizing the government for its management of the crisis. A thorough analysis of the contributions supporting the opinion leadership shows that the nature and scope of the messages emitted by this elite (i.e., relevant leaders from the opposing parties) depend on the coverage provided by the media. In addition, it shows that not all the messages achieve similar relevance to public opinion. The persuasion of messages broadcast by the elite depends on their credibility, which is conditioned by the interaction between sender, receiver, message, and media (Edwards and Swenson, 1997; Colaresi, 2007; Groeling and Baum, 2008; Randahl, 2018).

Finally, the *priming theory of public opinion and media coverage* is based on the role of the media, which intensifies the news related to political and governmental action during an electoral campaign (Brody and Shapiro, 1989). In the context of the 11-M attacks, the media had a strong incentive to highlight news of interest because the elections were only three days later. This interest can be exploited by the main parties: if the news favored the interests of the ruling party, then its content is emphasized and amplified by the party leaders; if the news damaged the popularity of the government, the opposition leaders found it advantageous to spread them since it would favored their electoral interests (Bali, 2007). These incentives could be consistent with earlier studies on the 11-M attacks: Montalvo (2011) concluded that the attacks altered the election results as the conservative party would have won the elections in its absence. The priming of public opinion and media coverage theory can be differentiated from the former two approaches because the media can influence the perception of the public opinion with respect to how the Government managed the crisis. This differentiation involves highlighting certain news and ignoring others (Edwards et al., 1995; Iyengar, 1994; Krosnick and Brannon, 1993; Zaller, 1992). In addition, in the electoral campaign, the government is an additional actor sharing the leading role with the elite (Bali, 2007).

In view of this background, we can hypothesize that different strategies in terror management

can induce different responses on voters' ideological distance through the effect of mass media. Therefore, the role of mass media is essential for developing this relationship, as mass media has proven to have multiple effects on politics (Piolatto and Schuett, 2015; Strömberg, 2015). This ideological distance between the voters and the political parties can be modeled following different approaches. One of the earliest is the *proximity voting model* by Downs (1957), which considers the Euclidean squared distance between the political position of the voter and the candidate's position. It defines a continuous one-dimensional space where both voters and candidates are located considering their ideological positioning (i.e., left-right). Voters evaluate each party and compute the distance/proximity between each candidate and their personal preferences. The closer the ideological position of a party is to the ideological preferences of the voter, the higher the utility obtained by voting to that specific party. The decision rule of the proximity vote proposed by Downs (1957) is expressed by the utility function P_i^j, where v_i is the ideological positioning of the voter i and p_i^k is the ideological location of party k.

$$P_i^k = (v_i - p_i^k)^2. \tag{1}$$

Later on, Enelow and Hinich (1984, 1994) developed the *spatial theory of voting* based on the previous contributions by Downs (1957) and Black et al. (1958). They assumed that the political parties and the voters are placed in a continuum that reflects their positions before a political issue. Therefore, the latter model predicts that each voter votes to the closest party and the political parties try to locate themselves at that point which allows them to maximize the number of votes, driven by the logic of maximizing their electoral support. Rabinowitz and Macdonald (1989) specified a *distance-based voting model* that reflects the intensity with which voters and candidates want to change their ideological position. Thus, they proposed a model that reflects the intensity with which both voters and candidates hold their preferences for a certain direction of policy making. Finally, a recent development of space theory is Kedar's *compensational voting model* (Kedar, 2005, 2009), an extension of previous proximity and discounting models by including a counterfactual thought-experiment where outcome-oriented voters compare the utility of the current party system with respect to the obtained by a hypothetical one. In this hypothetical system, a given candidate is removed from the policy-making process, and therefore, the difference of these systems represents the utility of the candidate (Kedar, 2009).

2.2. Related work in ABM and spatial modeling

ABM is a popular approach for analyzing political scenarios (Fowler and Smirnov, 2005; Kollman et al., 1998; Moya et al., 2017; Muis, 2010; Sudo et al., 2013). Furthermore, there are several contributions exploring the joint use of ABM and *spatial theory of voting*. From those, there are some studies using a Dowsonian approach like the one followed in the current manuscript (Clough, 2008; Plümper and Martin, 2008). Clough (2008) explores the impact of uncertainty into an ABM, concluding that typical models with complete information do not converge to the median distance but the simulations using higher levels of uncertainty actually do. Plümper and Martin (2008) introduce an ABM using a Dowsonian spatial model with multiple parties and multiple dimensions. They find that the number of competing parties and the likelihood to abstention increases the average distance between the parties and the center of the ideological scale. However, we can see that these studies tackle issues that are not directly linked to our problem, as we here analyze the effects of information broadcast into the ideological distance between voters and political parties.

In addition, another line of research worth mentioning is the one using ABMs and spatial models for analyzing political competition. Laver (2005) explores spatial models of political competition where parties follow adaptive rules. Since voters are modeled to vote for the closest party, parties behave as stickers (who never change their position), aggregators (which aim to the mean preference of the existing voters), hunters (who move greedily for maximizing the number of supporters), or predators (that moves towards the largest party). This work is later extended by Laver and Schilperoord (2007), where the authors study the survival of political parties by considering a dynamic environment where new parties can appear and the existing ones can be extinguished. The results of this model suggest that vote-seeking parties tend to make voters miserable, since their priority is to get new supporters instead of focusing on their current supporters (Laver, 2011). Laver's approach to political competition has been extended by several contributions (Adams and Mayer, 2008; Lehrer and Schumacher, 2018; Wright and Sengupta, 2015). Finally, we can observe how the parties of our study would have been considered as stickers by Laver, since the three of them do not move their position regarding the authority of the attacks, which was the main issue during this period.

3. Description of the agent-based model for ideological distances

The designed ABM is used for testing our hypothesis: the 11-M attacks and its management by the Spanish government and the opposition influenced the ideological distance between the voters and the political parties. In this influence the mass media played a critical role as they connect the information produced by political parties with the voters. Our approach is to analyze the changes on the ideological distance itself, which makes it the main variable of the model, not to use the ideological distance as a method for explaining the behavior of the voters neither the outcome of the elections. Finally, our model computes the ideological distance following the proximity model by Downs (1957), since the available data does not allow us to reproduce other models based on intensity (this issue is further addressed in Section 4.1).

3.1. General structure and agent's state

Our model considers a terminating simulation of 72 steps, which represents the 72 hours between the attacks and the elections (from March 11 at 8:00 AM to March 14 at 8:00 AM) using a time-step of an hour. The model simulates the behavior of N agents representing artificial voters and their reaction to the information received from C mass media channels along with the diffusion of this information through the social network due to a WOM process. The information supplied by the mass media channels contains a polarized message from different political leaders from the P main parties (i.e., PP, PSOE, and IU), which concentrated 85.26% of the total votes. Due to the strong positioning of the parties regarding the authority of the attacks, these messages influence how the agents position the parties in the ideological space.

This message polarization modifies the perception of the agents with respect to the ideological distances between them and the main parties. We model this effect by using the state variable *distance*, encoded as Λ_i^k, a real-valued variable, with $i \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, P\}$. The value of the latter variable changes during the simulation depending on the amount of external influences supplied by the mass media and the other neighboring agents. The distance variable is initialized by using the locations of the ideological space of voters and political parties (CIS, 2004). These data are introduced as v_i and p_i^k using integer values defined in interval $[1, 10]$, where v_i is the ideological location of the voter i and p_i^k is the ideological position that i assigns to the

party $k \in \{1, \ldots, P\}$, as P is the number of parties participating in the election. Therefore, we can compute the initial ideological distance as $\Lambda_i^k = \left|v_i - p_i^k\right|$, and thus is defined in interval $[0, 9]$. In addition, the state variable $\Phi_i = \{0, 1, \ldots, P\}$ represents the political dispositions of each agent i, which identifies it as voter of a particular party (i.e., the agent voted for IU, PP, PSOE, or the remaining parties in the previous elections) or as a non voter (i.e., abstainers). Thus, Φ_i takes the value of the party when the agent is a voter and 0 if it is a non-voter or abstainer.

3.2. Social behavior of the agents and word-of-mouth

The N agents of the model are connected by an artificial social network (Barabási and Albert, 1999; Watts and Strogatz, 1998) modeled as a scale-free network (Barabási and Albert, 1999). We select the scale-free approach because many real-world networks match this topology (Barabási and Albert, 1999; Newman et al., 2006). The degree distribution of these networks is shaped as a power law, where most nodes have few connections but few nodes have multiple connections, that are referred as the hubs of the network. Barabasi-Albert's preferential attachment algorithm (Barabási and Albert, 1999) allows to generate scale-free networks relying on the parameter m. This parameter modulates the growth rate of the network and its final density (Barabási and Albert, 1999). Barabasi-Albert's algorithm starts with a fully connected graph with m_0 initial nodes. Then, the algorithm iterates by adding a new node to the network and connecting it to m existing nodes which are selected with a probability proportional to their degree. This procedure continues until the network reaches the desired size. The final average degree of the resulting network can be calculated as $\langle k \rangle = 2m$.

Agents can share their perceptions regarding their ideological distances with the main parties. These interactions between the agents can be modeled as a contagion process (Lee and Kim, 2014; Newman et al., 2006) as they spread their distance values through the social network. At every step of the simulation, each agent i considers a talking probability ($p_i(t) \in [0, 1]$) of spreading its distance values for the different parties (collected in the Λ_i^k values). Therefore, the agent spreads its values with all of its neighbors each time the probability check is passed. The variable *influence* (Δ) models the influence of an agent with its neighbors. When an agent shares its perceptions, it does it in a directed-only way (i.e., from the active agent to its neighbors). The update of the distance value due to social interactions is defined by Equation 2, where $\Lambda_j^k(t)$ refers to the distance value of the neighbor agent j of party k when the active agent i shares its perceptions. This equation regulates the final influence using two additional values, since it would not be realistic that agents with very distant ideological positions were to influence each other. First, $\Theta(x)$ represents the Heaviside step activation function with $x = |\Lambda_j^k(t) - \Lambda_i^k(t)| - \psi$. $\Theta(x)$ returns 1 when $x \geq 0$ and 0 otherwise. Therefore, it disables the WOM influence when the distance difference is greater than a given threshold ψ. Second, voters with different political dispositions (i.e., voters of different political parties or non-voters) are less likely to influence each other. This is resembled by the parameter $\phi \in (0, 1]$, which regulates the influence when $\Phi_i \neq \Phi_j$, otherwise it is equal to 1.

$$\Lambda_j^k(t+1) = \Lambda_j^k(t) + \left|\Lambda_j^k(t) - \Lambda_i^k(t)\right| \Theta(x) \phi_{\Phi_i \neq \Phi_j} \Delta. \tag{2}$$

Our model considers an additional parameter referred as influence decay ($d\Delta$) that regulates how social influence erodes over time if it is not reinforced with further stimulus. Therefore, every agent reduces its accumulated social influence at the beginning of each simulation step due to this decay effect. The accumulation of social influence ($\delta_i^k(t)$) due to WOM interactions follows

Equation 3, which represents the accumulated changes to Λ from the start of the simulation to the current step t. Finally, the distance value change experienced by agent i with respect to party k due to the decay effect is defined in Equation 4.

$$\delta_i^k(t) = \sum_{s=1}^{s=t} \left(\Lambda_i^k(s) - \Lambda_i^k(s-1) \right). \tag{3}$$

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) - \left(\delta_i^k(t) d\Delta \right). \tag{4}$$

3.3. External influences using mass media channels

The registered media audience between 11-M and 14-M are the external influences to the agents during the simulation (Asociación para la Investigación de Medios de Comunicación (AIMC), 2004; López García, 2004), which can influence any agent during any simulation step. These media, modeled as global mass media (González-Avella et al., 2007), are parameterized for resembling the differences between the multiple channels. These channels are able to reach any agent randomly depending on the audience of the channel for that step. The selected media are written press, radio, and television channels, since the Internet did not have enough influence in 2004. In addition, we model cell phone messages and similar communications using WOM, as it is well known they had a strong activity and voting influence during the studied period (Olmeda, 2005).

The information supplied by the mass media channels considers any message containing information about the attacks, whether it appeared in regular news sessions, were included as special bulletins, or were taken from statements of political figures. This information was selected following three main criteria: diversification (we consider multiple types of mass media channels), scale (we include mass media channels operating nationally), and plurality (messages were included avoiding discrimination of sources). The selected channels are Cadena Ser (radio), El Mundo (written press), El Pais (written press), ABC (written press), Antena 3 (television), Telecinco (television), and TVE (television).

During the simulated period, these selected channels where the most relevant operating in Spain. Thus, its combination considers the main messages produced between the attacks and the elections. The analyzed television channels had more than 75% of share and 45% of radio users listened to Cadena Ser (Asociación para la Investigación de Medios de Comunicación (AIMC), 2004), which is also known for having a relevant role during this political event (Olmeda, 2005). Finally, the selected written press was the most read at the time.

Mass media channels can spread different messages at any step t and can have different values for their parameters despite of belonging to the same media type. Each message transmitted by the media channels considers a polarization value modeled as $m_c(t) \in [-2, 2]$, that models the content of a message broadcast by specific channel c at time-step t (i.e., authority of ETA *versus* authority of Al Qaeda). $m_c(t)$ is set to -2 in case the message clearly informs that Al Qaeda as the author of the terrorist attack and 2 if it clearly informs of ETA's authority. The values -1 and 1 are assigned to the interventions of political leaders that either support or criticize the actions and transparency of the Spanish government during the crisis, instead of discussing specifically the authority of the attacks. Since most of the messages from this period are concerned with the authority of the attacks, the messages of the latter category represents around 5% of the total messages and mostly appear in the second half of the simulation. Finally, 0 refers to the message not being biased to any specific terrorist organization. Because the simulation runs hourly, for any

time-slot where two or more messages appear the resulting polarization value is computed as the average of these messages.

In addition, because there are two competing frames aligned with the considered parties, the transmitted polarization modifies the distance of parties differently. If the resulting polarization of a message is biased towards ETA (i.e., polarization > 0) then the agent's distance with the PP party is reduced and the distance with the other parties is increased. In contrast, if the message is biased towards Al Qaeda, the perceived distance with the PP party is increased and the distance with the other parties is reduced. Finally, the polarization values were scored by different experts due to the subjectivity of this task and agreeing the final values using the average.

Besides the transmitted message, mass media channels are modeled with respect to their reach, their influence, and their buzz. The reach parameter ($r_c \in [0,1], \forall c \in \{1,\ldots,C\}$) models the maximum percentage of the agent population that each channel can hit in a single step, since some channels can potentially reach to more people than others (Moya et al., 2019). The data for setting this parameter is taken from the Zenith's media track [1] that studies mass media consumption in Spain in 2013. Since it includes data since 2006, we can approximate the reach parameters for the 2004 Spanish elections from the mass media reach values of 2006.

The influence parameter (Δ'_c) modulates the influence achieved by a mass media channel after impacting a given agent. This influence works similarly to the produced by WOM interactions, but in this case the distance change is calculated by using the specific change value of the channel c and the polarization of that channel during that time-slot. Since an agent could receive the same message multiple times, the maximum influence is bounded by an overall influence value (Δ_c^{max}). In addition, the influence previously accumulated by the channel ($\delta_c'^{k}$) is treated similarly to one accumulated by WOM. An agent i experiences a change of its distance values for party k by the influence of a given channel c following Equation 5, where $m_c^k(t)$ refers to the resulting polarization for party k in time-step t and Δ_c^{max} represents the maximum amount of influence that can be supplied by c. However, as the simulation progresses and new messages are produced by mass media channels, the agents tend to forget previous messages (Moya et al., 2017; Wu and Huberman, 2007; Yang and Leskovec, 2010). We model this effect using the influence decay ($d\Delta_c$) parameter, that regulates the rate at which the agents are forgetting previous influences, analogously to WOM. The distance value update of agent i for party k due to the effect of decay of channel c is defined by Equation 6.

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) + (\Delta_c^{max} - \delta_c'^{k}(t))\Delta_c m_c^k(t). \tag{5}$$

$$\Lambda_i^k(t+1) = \Lambda_i^k(t) - (\delta_c'^{k}(t) d\Delta_c). \tag{6}$$

Additionally, the information supplied during this critical events is likely to trigger a viral buzz effect on the listening agents. Thus, we include a buzz increase parameter (τ_c) for each channel c that increases the agent's talking probability by a given percentage of its initial value ($p(0)$). However, this buzz effect decreases over time as newer information is spread by the media. The buzz decay parameter ($d\tau_c$) reduces the previously increased talking probability in the agents because of the effect of τ_c. Therefore, the talking probability values of an agent i are updated according to Equations 7 and 8, respectively.

[1] http://blogginzenith.zenithmedia.es/estudio-zenith-los-medios-en-espana-y-portugal-un-terreno-cambiante/

$$p_i(t+1) = \begin{cases} p_i(t) + (p_i(0)\tau_c), & \text{if } (p_i(t) + (p_i(0)\tau_c)) \leq 1, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

$$p_i(t+1) = p_i(t) - (\sigma_i d\tau_c). \tag{8}$$

4. Model calibration and validation

This section introduces the processes used for calibrating and validating the model along with the employed data. Section 4.1 describes the setup of the simulation and the data used for initializing and adjusting the model. Then, Section 4.2 introduces our automatic calibration approach and discusses its results. Finally, Section 4.3 reviews the different outputs of the calibrated model.

4.1. Data description and simulation setup

The ideological positions for the voter agents (v_i) and the ideological position where they allocate the considered parties (p_i^k) are taken for the 2553 study of the CIS (2004). This is the pre-elections survey developed between December 8 and 15 of 2003 that considers a sample size of 1,500 interviews. As detailed in the survey, the interviewed individuals were selected using a *multistaged, stratified, cluster sampling, with the selection of primary units sampling (municipalities) and the secondary units (neighborhood) proportionally random, and the last units (individuals) by random routes and quotas of sex and age.* In addition, we use the 2555 study (the post-elections survey) for calibrating the model[2]. However, in this case we consider the average distance of the interviewed individuals as the target average distance values for the agents. In addition, these surveys does not contain enough information for modeling the ideological distance using the approaches based on intensity, since the post-election survey does not ask how important (intense) the attacks were for the voters' decision.

The polarized values ($m_c^k(t)$) of the multiple messages supplied by the different mass media channels are defined using information from different sources. In the case of television we used the informational volume 19-20 from *Quaderns del Consell de l'Audiovisual de Catalunya* (Consell de l'Audiovisual de Catalunya, 2004). The messages from the radio were analyzed using the audio from Cadena Ser [3], since this is the only radio channel providing access to their audios for the analyzed period. The written press values (i.e., El País, El Mundo, and ABC) were extracted from the *MyNews* on-line database [4]. Finally, the assessment of the messages was taken from a previous study (Moya et al., 2017), where several experts already evaluated the polarization of the considered messages.

[2]Notice that, we are aware that using this source of data to initialize and calibrate our model could be slightly problematic. On the one hand, the ability of the pre-electoral data of the Spanish Center for Sociological Research to predict the election results in Spain has been deceptive in some elections. On the other hand, post-electoral surveys could tend to overestimate the voters for the winning party. Nevertheless, the final goal of our case study is not to predict the results on the elections but to elaborate on the voters' ideological positioning, which should be less sensitive to the latter problems. Besides, the surveys of the Spanish Centre for Sociological Research are the only source including all the required information to build and calibrate our model, and it has been obtained from a representative sample of the population in Spain.

[3]`http://www.cadenaser.com/static/especiales/2005/sonidos11_14/dia11.html`

[4]`http://ugr.mynews.es/hu/`

Using these data, the simulation setup considers a set of $N = 21.280$ agents. This population size is set by extending the number of interviews of the pre-elections survey and instancing multiple agents for each of the interviewed voters. In addition, the regulators of the WOM interactions are set to $\psi = 3$ and $\phi = 0.5$, respectively. On the one hand, a value of $\psi = 3$ disables WOM interactions when the difference in the distance values of the interacting agents is greater than 3 (i.e., a third of the variable's max range). On the other hand, a value of $\phi = 0.5$ halves the resulting influence of two agents that represents voters from different political dispositions, offering a good balance by allowing WOM influences between agents from different groups while penalizing these exchanges. Finally, the simulation considers 30 Monte-Carlo runs.

4.2. Results of the automatic calibration algorithm

We use automatic calibration for adjusting the parameters of the model. It is an automatic procedure that tunes a selection of the model's parameters using an optimization method and a deviation function which compares the model's simulated output with the target real data. The optimization method minimizes the deviation function iteratively by running independent model simulations for each candidate parameter configuration. After the adjustment the resulting model configuration requires to be reviewed in detail to check its validity, which is performed in the following sections.

The automatic calibration procedure adjusts 35 parameters of the model. These parameters are those regulating the diffusion of information from the C channels, since those are both the hardest to set manually using the available information. For each of the selected mass media channels, we calibrate five of its parameters: its maximum influence (Δ_c^{max}), influence change (Δ_c), influence decay ($d\Delta_c$), buzz increase (τ_c), and buzz decay ($d\tau_c$). These parameters are calibrated within the $(0, 1]$ interval, with the exception of the maximum influence parameters (Δ_c^{max}) which take a value in the $(1.5, 2.5]$ interval.

A memetic algorithm (Moscato, 1989) that comprises a steady-state genetic algorithm (Back et al., 1997) and local search refinement is selected as our optimization method. The algorithm is initialized considering a population of 100 feasible solutions that represents valid values for the selected models' parameters using an integer coding. These integer-coded values are the result of splitting the given real-coded intervals of valid parameter values with a size-step of 0.001, as done in Chica et al. (2017). The algorithm iterates until reaching a stopping criteria of 10,000 evaluations. Each evaluation involves running 30 individual Monte-Carlo simulations of the ABM in order to obtain the fitting of a model parameter configuration.

In addition, the algorithm considers 3-tournament selection, uniform random mutation, and a BLX-α crossover (Herrera et al., 1998). The mutation operator has a mutation probability $p_m = 0.1$ of modifying each decision variable. This operator resets the value of the mutated gene by generating a new random value from its specific interval using an uniform distribution. The crossover operator activates with probability $p_c = 1$ and generates two offspring solutions by crossing two existing parent solutions. These new solutions are generated by selecting new values from the interval $[c_{min} - I\alpha, c_{max} + I\alpha]$, with $c_{max} = max(v_i^1, v_i^2)$, $c_{min} = min(v_i^1, v_i^2)$ and $I = c_{max} - c_{min}$, and v_i^1, v_i^2 representing the decoded values of the parent solutions. Theses values are also truncated according to the set of feasible values for each decision variable.

Model fitting is computed using the values of the post-electoral survey (CIS, 2004), where the participators were asked to position both themselves and the parties in an ideological scale between 1 and 10. For each participant, the ideological distance is computed as the absolute difference between their position on the ideological continuum and the position of each political

party. Therefore, the average values of all the voters' distances are used for evaluating the fitness of a given model configuration by comparing them with the simulated distance values at the end of the simulation.

Table 1 shows the fitting results for the calibrated model. In these values, we can observe that the calibrated model obtains an excellent fitting for IU and PP parties, since their absolute deviation error (computed as $|o_j - s_j|$, with o_j being the observed value and s_j being the simulated value) is equal to or lower than 0.05 for both parties. If this deviation is translated to a percentage error ($100|o_j - s_j|/o_j$) with respect to the actual values from the post-electoral survey (2.43 and 3.42, respectively) the deviation is still lower than a 3%. However, the average distance for the PSOE party results harder to fit, since the absolute deviation is 0.18. As the target value is the lowest of all parties (1.57), the relative error (around a 12%) is greater than the other parties but we can argue that the calibrated result is acceptable. Additionally, we note that no other model configuration can improve the fitting of PSOE's distances without reducing the fitting of the other two parties.

Party	Pre	Post	Simulated	Deviation	Percentage
IU	2.55	2.43	2.39	0.03	1.23
PP	3.22	3.42	3.43	0.005	0.14
PSOE	1.92	1.57	1.77	0.19	12.1

Table 1: Distance values from the election surveys and simulation results. Additionally, the simulation deviation error is shown both as an absolute value and as a percentage value.

4.3. Model's output and analysis

The main output of the model is the evolution of the distance values for each party during the different steps of the simulation. Figure 1 shows the average of the distance values for the agents of the model by each party at each step of the simulation. These values are computed using the average of the 30 Monte-Carlo runs with a resulting negligible deviation, hence it is not shown in the charts. The overtime values show that the average distance increases during the beginning of the simulation for the IU and PSOE parties and reaches its maximum value at step 16. After this peak, a moment where a significant change in the information broadcast by the mass media arose as a consequence of the findings by the Spanish police, the values decrease until stabilizing during the final steps of the simulation. We can see how the average values have a similar behavior over time for both parties (i.e., IU and PSOE). In contrast, we can see that the average statistics show an opposite behavior for the PP party as they reduce their values during the first steps of the simulation and increase them in the subsequent steps.

Additionally, the social behavior of the agents can be evaluated regarding the number of WOM interactions and the effect those interactions. Figure 2 shows the average percentage increase on the number of WOM interactions during the simulation, with the blurred areas representing the Monte-Carlo variability. These values show two peaks for each day of the simulation corresponding with the news in the afternoon and in the evening. Hence the news on prime time had the biggest share for the televisions and caused a high buzz for the following steps.

The effect of the agents' social interactions can be approached as the *sentiment* of WOM. It reflects the impact of these interactions on the agent distance values showing if there is a majority of conversations increasing or decreasing the distance values. Sentiment value indicates the trend of

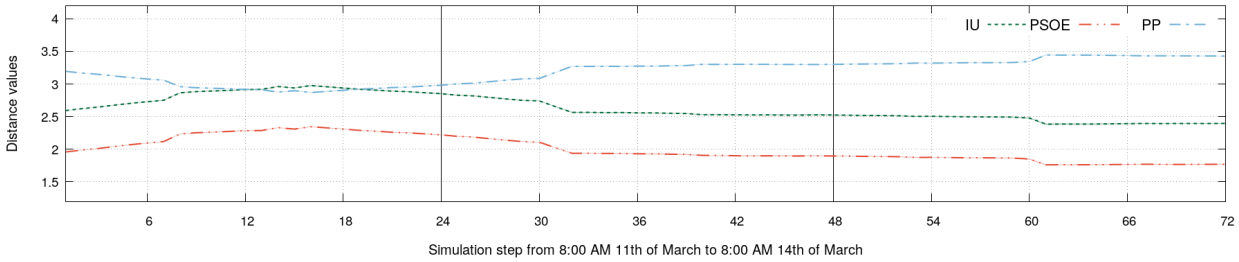


Figure 1: Average ideological distance values of the agents for each party. The displayed values represent the average of the 30 Monte-Carlo repetitions at each step of the simulation. The Monte-Carlo repetitions show a negligible deviation and, hence, it is not shown in the charts.

the social interactions and hence an unfavorable sentiment means that there are more conversations increasing the distance values than decreasing them. Therefore, a sentiment value of 5 means that there are 5% more unfavorable WOM interactions than favorable ones (i.e., those decreasing the distance values). Figure 3 shows the average sentiment of WOM interactions for each party. We can identify how the peaks in the sentiment values match with the timing of the news, as observed in the WOM volume. In addition, we can see how the sentiment trend shifts during the simulation. At the beginning of the simulation, the sentiment for both IU and PSOE parties is unfavorable while the sentiment of PP is favorable. Then, these trends shifts as the mass media polarization changes. This can be also observed in the overall aggregated sentiment behavior.

5. Analysis of what-if political scenarios

Using our calibrated and validated model we can simulate and analyze the political scenarios from the *theory of terror management* identified in Section 2.1. Thus, *the rally around the flag* is analyzed in Section 5.1, *the opinion leadership* is studied in Section 5.2, and *the priming of public opinion and media coverage* is tackled in Section 5.3. We can see the impact of these scenarios on the resulting ideological distance values of the artificial voters by analyzing the values obtained at the end of the simulation. Hence, Figure 10 displays the absolute variation of the resulting average distance values with respect to the baseline calibrated model for the multiple Monte-Carlo simulation runs using boxplots.

5.1. The rally around the flag

The rally around the flag scenario is simulated by modifying the polarization of the mass media channels to have only those messages that support the government's version towards the authority of the attacks. Therefore, the messages that either blame Al Qaeda or which accuse the government of lying are disabled and do not take effect during the simulation. Figure 4 shows the average ideological distance values for this scenario. In these values, we can observe that *the rally around the flag* has a significant impact on the ideological distance values for every party.

The distance values produced by this scenario can be clearly distinguished from the baseline scenario, specially after step 16 where the polarization towards Al Qaeda would have started to become stronger in the baseline scenario (see Figure 1). We can see that the combined effect produced by *the rally around the flag*, where all the media channels support the same communicative

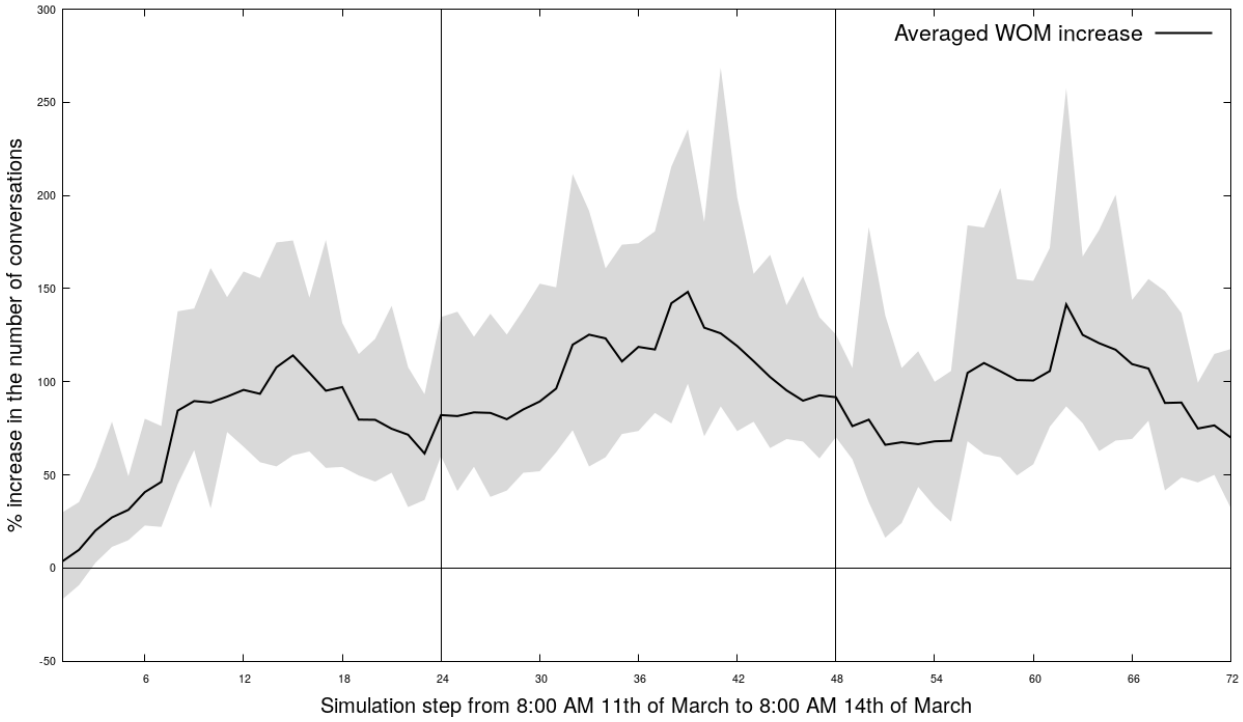
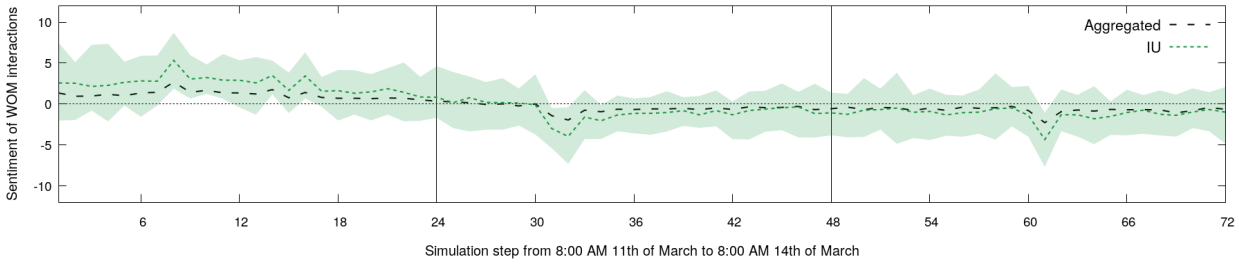
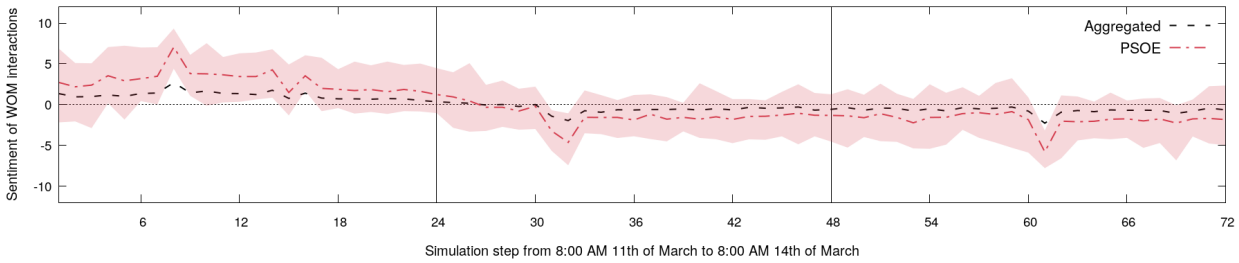


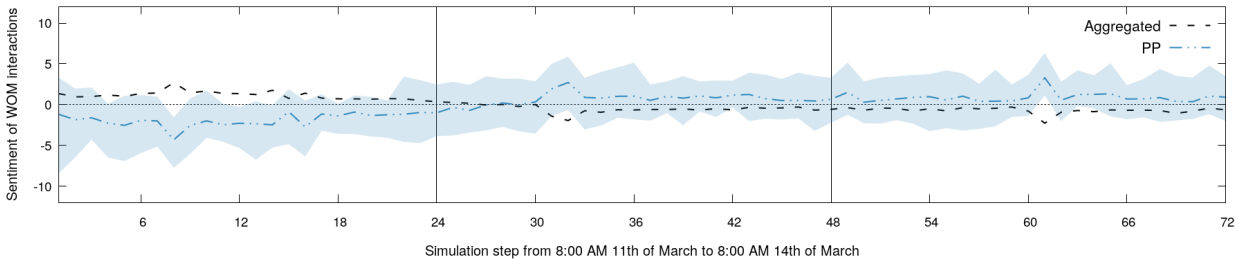
Figure 2: Percentage increase in the number of WOM interactions by step, shown as the average of the multiple Monte-Carlo repetitions. The variability in these repetitions is shown using the blurred area between the maximum and minimum values obtained for each step.



(a) IU



(b) PSOE



(c) PP

Figure 3: Average sentiment of WOM interactions for each party. Blurred areas represent the maximum and minimum values between the multiple Monte-Carlo repetitions. In addition, average overall sentiment resulting from aggregating the parties' values is also included. A dotted line at 0 represents a neutral sentiment, separating the favorable and unfavorable sentiment areas.

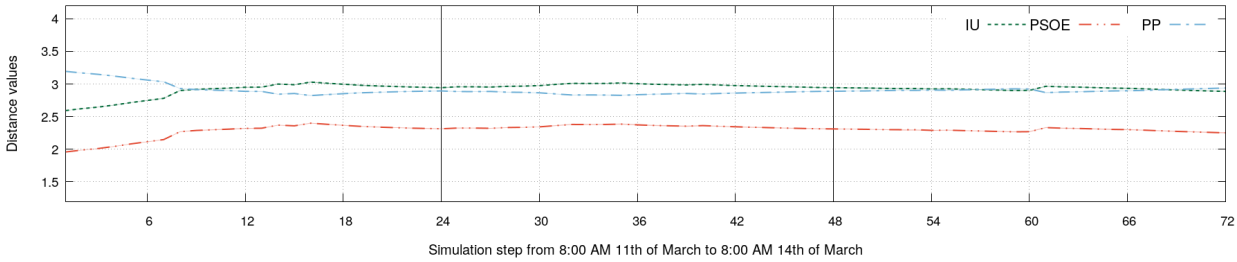


Figure 4: Average ideological distance values for *the rally around the flag scenario*. The displayed values represent the average of the 30 Monte-Carlo repetitions at each step of the simulation. The Monte-Carlo repetitions show a negligible deviation and, hence, it is not shown in the charts.

framework, produces stable values that can resist the accumulated decay effect during the simulation. In view of the results in Figure 10, we can recognize that this scenario has the highest impact on the parties since the average distance values for each party show a deviation of around 0.5.

Finally, Figure 5 shows the sentiment of the WOM interactions for this scenario. In these values we can observe that the sentiment trend for the PP party is mostly favorable during the whole simulation, which implies that the distance values of the PP party are constantly being reduced. In contrast, the sentiment values for IU and PSOE are mostly unfavorable or neutral, reflecting a constant increase on their distance values.

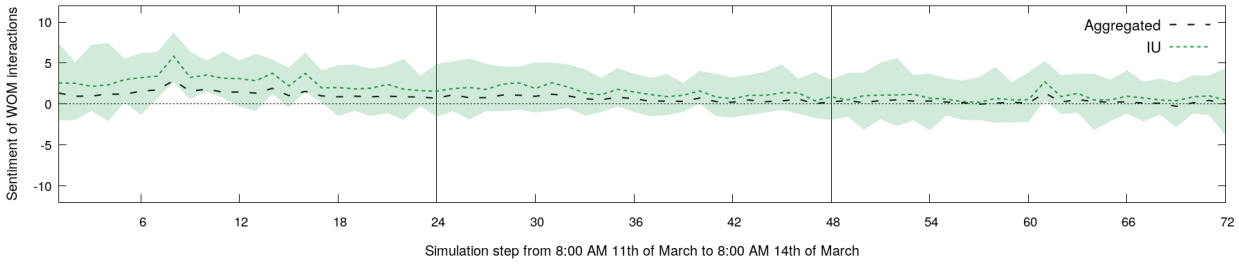
5.2. The opinion leadership

We can simulate *the opinion leadership* scenario by reducing the messages of the mass media channels to those statements of opinion leaders that claim that the government lies regarding the authority of the attacks. This is done by disabling the messages that mention the authority of the attacks by either ETA or Al Qaeda as well as the messages supporting the government. Since the remaining messages do not compose a sample with enough size for reproducing *the opinion leadership* effect, we propose to analyze the hypothetical scenario by inserting additional messages of opposing leaders where they criticize the government. This allows us to analyze *the opinion leadership* scenario according to the assumptions of the *theory of terror management*. Thus, we include several messages from this category in the TV channels during their afternoon and evening news. By including these new messages during the TV channels prime time their effects should be observed clearly, since these are the mass media with the highest audience.

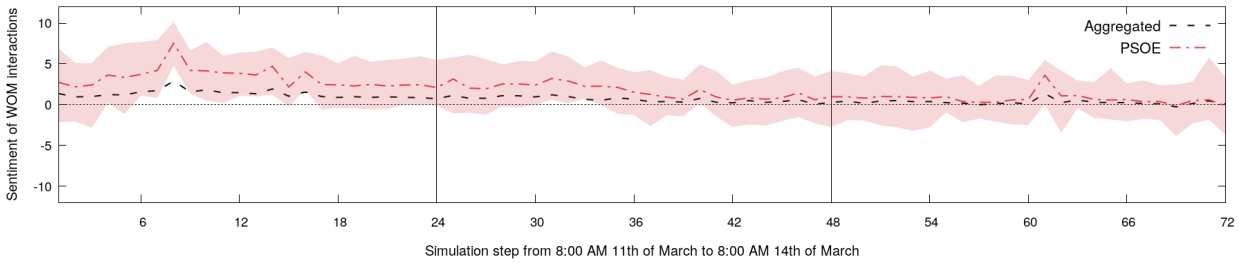
Figure 6 shows the average distance values for this scenario. From this figure we can observe that this scenario has noticeable effects on the average distance values for the analyzed parties in the selected time steps. We can see that the accumulated change produced by these new messages involves a stable variation on the average distance for every party, although some change is lost over time due to the separation of the news time steps.

The effect of these messages can be corroborated by the associated sentiment values, depicted in Figure 7. We can observe how the newly added messages produce individual spikes in the sentiment values during the corresponding news time step. This involves a strong favorable sentiment for IU and the PSOE party and an unfavorable sentiment for the PP party. However, we can also recognize how the sentiment variation for the PP is lower than the variation shown by the other parties.

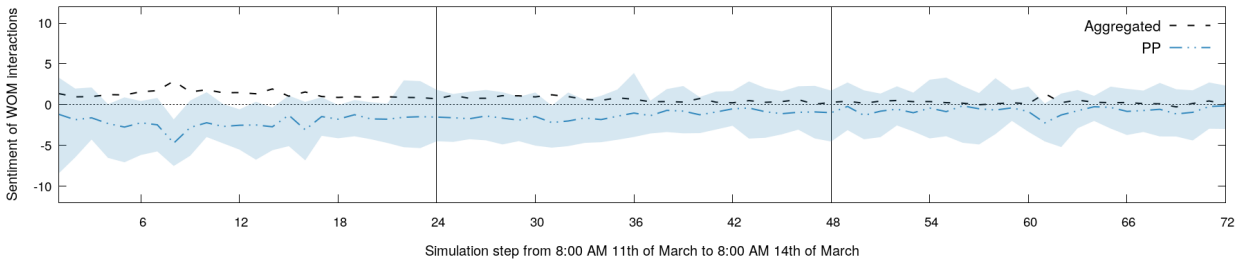
As seen in the variation of the distance values for this scenario with respect to the baseline



(a) IU



(b) PSOE



(c) PP

Figure 5: Average sentiment of WOM interactions for each party for *the rally around the flag scenario*. Blurred areas represent the maximum and minimum values between the multiple Monte-Carlo repetitions. In addition, the average overall sentiment resulting from aggregating the parties' values is also included. A dotted line at 0 represents neutral sentiment, separating the favorable and unfavorable sentiment areas.

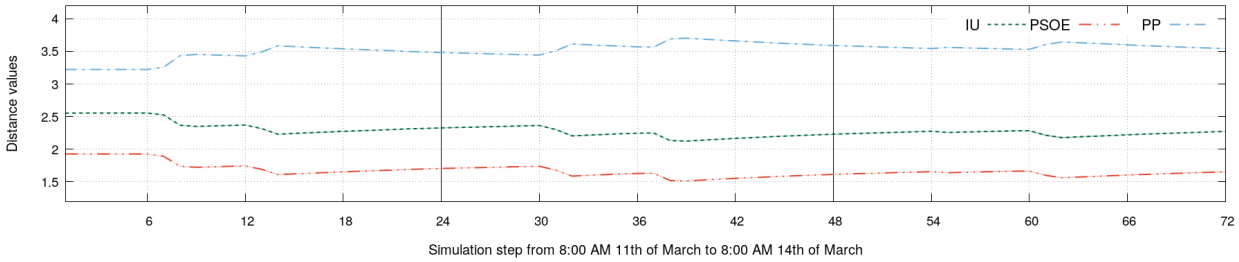


Figure 6: Average ideological distance values for *the opinion leadership* scenario. The displayed values represent the average of the 30 Monte-Carlo repetitions at each step of the simulation. The Monte-Carlo repetitions show a negligible deviation and, hence, it is not shown in the charts.

simulation, showed at Figure 10, this scenario produces a similar effect in all the parties, reducing the distance of IU and PSOE and increasing the distance of the PP party by the same amount. This shows how a systematic appearance of the political leaders criticizing the government during the 2004 Spanish elections would have a strong impact in the ideological distance of the voters and highlights the role of the leaders of opposing parties.

5.3. Priming of public opinion and media coverage

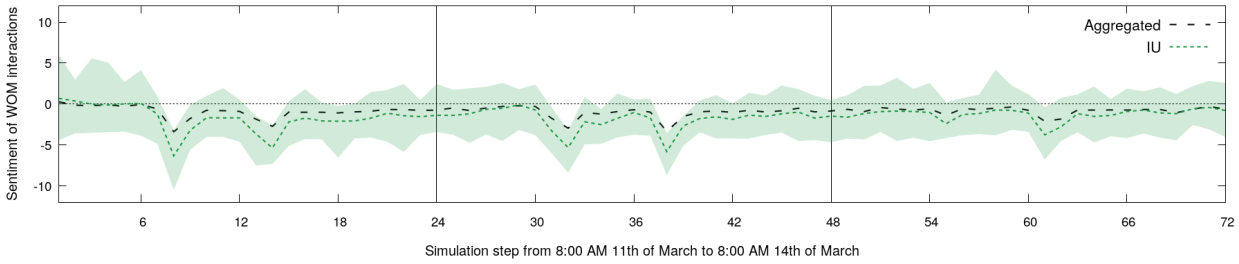
The simulation of *the priming of public opinion and media coverage* scenario is designed by focusing the broadcast information in the messages pointing out Al Qaeda's authority of the attacks. Thus, the messages supporting that ETA is responsible for the attacks and those either blaming or support the Government are disabled. The resulting distance overtime values for this scenario are displayed at Figure 8. These values resemble a significant impact on the average distance values of the voters with each party: in the cases of the PSOE and IU the average distance is reduced below the values of the baseline; in contrast, for the PP it increases its values beyond the baseline.

This can also be observed at Figure 10, which shows a deviation of around a 0.15 for every party in their distance values. Similarly to the results obtained by *the rally around the flag* scenario, the average distance values at the end of the simulation stay stable and seem to resist to the erosion caused by decay. This suggests that the effect of this scenario could have prolonged over time. In contrast with the results observed in *the opinion leadership* scenario, the overtime evolution of the average values is constant, instead of being pushed by the information supplied during a specific time-step. This also resembles how the information regarding Al Qaeda's authority of the attacks was constant during the three days period.

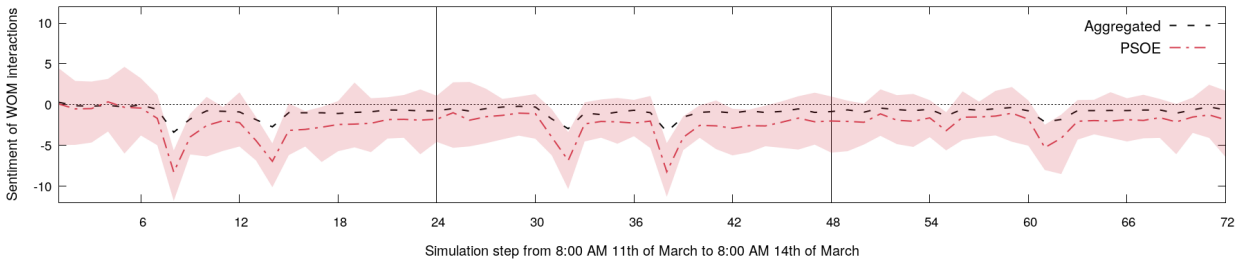
The sentiment values for this scenario (shown at Figure 9) can support the latter conclusion. The observed sentiment values show that the interactions regarding the PP party are mostly unfavorable for the whole simulation, which increase the average distance value. In the case of IU and PSOE, we can see how they maintain a favorable sentiment trend during the simulation, including a maximum favorable sentiment higher than -5.

6. Concluding remarks

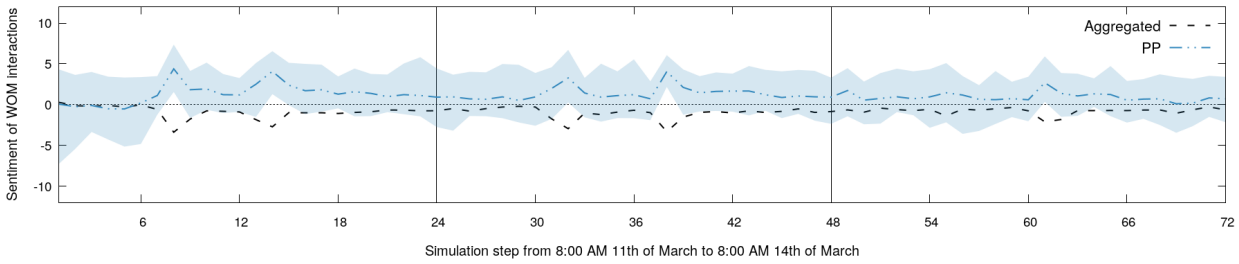
In this paper we have analyzed the effects of the politicians' management of the 11-M attacks on the ideological distance of voters with PP, PSOE, and IU in 2004 Spanish elections. We carried out our analysis by designing and implementing an ABM that simulates the three days between the



(a) IU



(b) PSOE



(c) PP

Figure 7: Average sentiment of WOM interactions for each party for *the opinion leadership* scenario. Blurred areas represent the maximum and minimum values between the multiple Monte-Carlo repetitions. In addition, the average overall sentiment resulting from aggregating the parties' values is also included. A dotted line at 0 represents a neutral sentiment, separating the favorable and unfavorable sentiment areas.

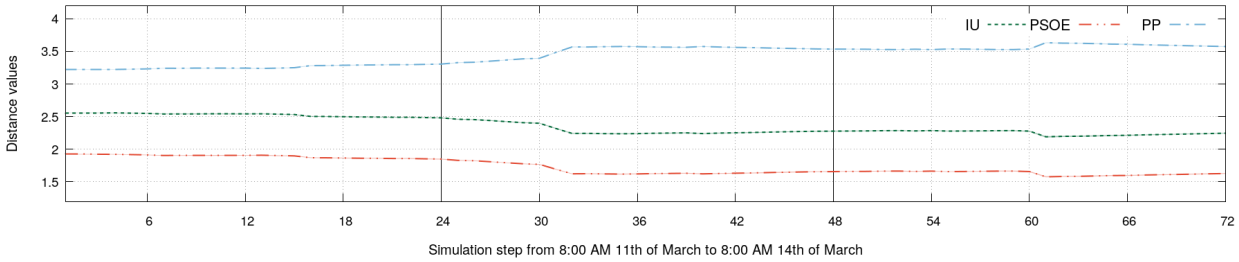


Figure 8: Average ideological distance values for *the priming of public opinion and media coverage* scenario. The displayed values represent the average of the 30 Monte-Carlo repetitions at each step of the simulation. The Monte-Carlo repetitions show a negligible deviation and, hence, it is not shown in the charts.

attacks and the elections and includes the information spread by different mass media channels. In our simulation, the artificial voter agents are exposed to both social interactions with other agents within an artificial social network and to the external influences of the main mass media channels in Spain in 2004. We calibrated our ABM with a memetic algorithm which comprises a steady state genetic algorithm with local search refinement.

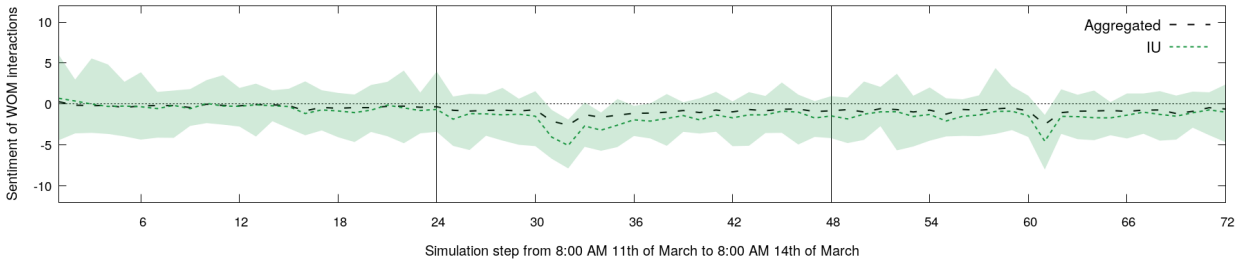
Multiple model outputs were analyzed for its validation: the average distance for each party (i.e., the main output for our study), the number of WOM interactions in the social network, and the sentiment of the latter interactions. These performance indicators showed how the politicians' management of the 11-M attacks could have influenced the ideological distance between the voters and the PP, PSOE, and IU parties. By using the calibrated and validated model, we were able to analyze three related political scenarios from theory of terror management: *the rally around the flag*, *the opinion leadership*, and *the priming of public opinion and media coverage*.

The simulated scenarios showed how the combined effect produced by all the media channels supporting the same communicative framework produced a significant and stable impact on the distance values of the voting agents. In these scenarios, we have identified that the distance values for both IU and PSOE behave in a similar way, which is consistent with them being in the opposition. On the contrary, the PP shows a different behavior for each scenario, which is consistent with it being the party in government. In addition, the observed impact on the distance values of the voting agents suggests that certain approaches to terror management could have a long-term effect on the ideological distance. Therefore, we have shown that the effects of the politicians' management after a shock like the 11-M attacks can produce a change in the ideological distance in the short term.

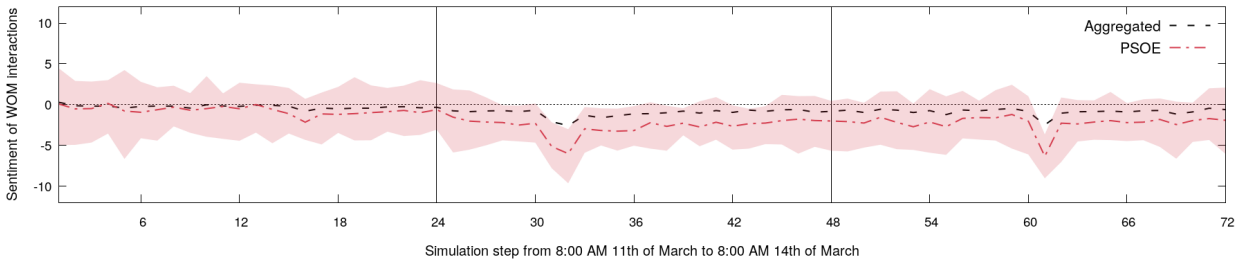
Future work will study employing more advanced techniques for modeling these complex political scenarios. For example, the use of fuzzy logic and computing with words (Giráldez-Cru et al., 2020) could improve the modeling and spread of linguistic information like the one describing ideological distance between the voters and the parties. In addition, fuzzy cognitive maps (Papageorgiou, 2013) could be used for modeling the behavior of the voters during these political scenarios.

Acknowledgments

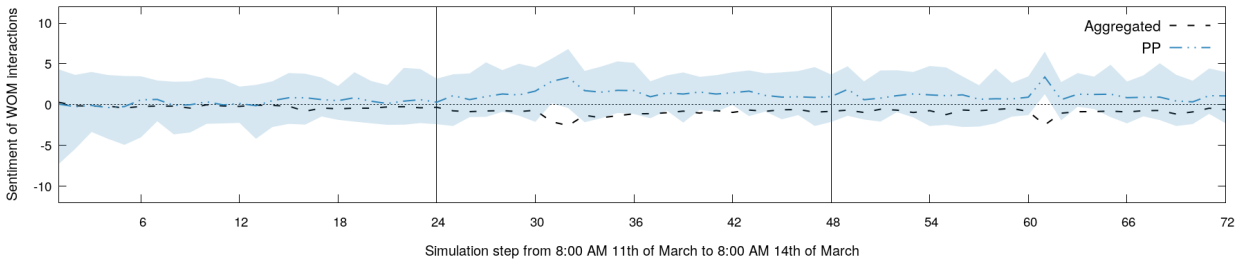
This work is supported by the Spanish Agencia Estatal de Investigación, the Andalusian Government, the University of Granada, and European Regional Development Funds (ERDF) under



(a) IU



(b) PSOE



(c) PP

Figure 9: Average sentiment of WOM interactions for each party for *the priming of public opinion and media coverage* scenario. Blurred areas represent the maximum and minimum values between the multiple Monte-Carlo repetitions. In addition, the average overall sentiment resulting from aggregating the parties' values is also included. A dotted line at 0 represents a neutral sentiment, separating the favorable and unfavorable sentiment areas.

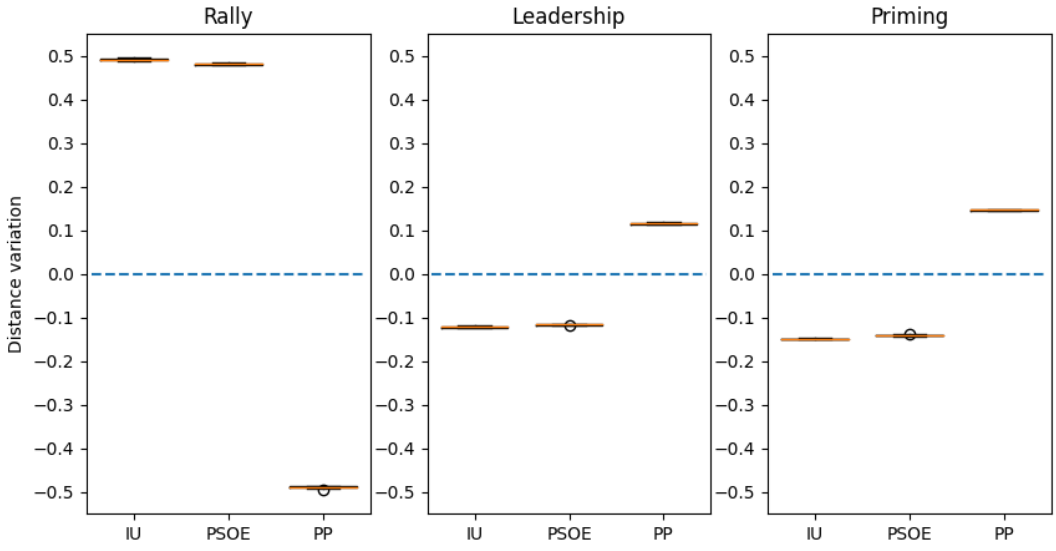


Figure 10: Boxplots showing the distance variation at the end of the simulation for each defined what-if political scenario with respect to the baseline calibrated model. The reduced size of the boxes indicates the robustness of the results across the 30 Monte-Carlo runs.

grants EXASOCO (PGC2018-101216-B-I00) and AIMAR (A-TIC-284-UGR18). We are also supported by Project I+D+i from the FEDER Operational Program, Andalusia (A-SEJ-154-UGR18). M. Chica is also supported through the Ramón y Cajal program (RYC-2016-19800).

References

Adams, J., Mayer, A.K., 2008. Condorcet efficiency with adaptive parties in a spatial model. Mathematical and Computer Modelling 48, 1298–1307.

Asociación para la Investigación de Medios de Comunicación (AIMC), 2004. AIMC February to November 2004. General Recap. EMG. `http://www.aimc.es/-Datos-EGM-Resumen-General-.html`. Accessed: 2016-03-28.

Back, T., Fogel, D.B., Michalewicz, Z., 1997. Handbook of evolutionary computation. IOP Publishing Ltd., Bristol (UK).

Baker, W.D., Oneal, J.R., 2001. Patriotism or opinion leadership?: The nature and origins of the "rally 'round the flag" effect. Journal of Conflict Resolution 45, 661–687.

Bali, V.A., 2007. Terror and elections: Lessons from Spain. Electoral Studies 26, 669–687.

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Berrebi, C., Klor, E.F., 2007. The Impact of Terrorism on Voters' Preferences. RAND Santa Monica, CA.

Berrebi, C., Klor, E.F., 2008. Are voters sensitive to terrorism? direct evidence from the Israeli electorate. The American Political Science Review 102, 279–301.

Black, D., Newing, R.A., McLean, I., McMillan, A., Monroe, B.L., 1958. The theory of committees and elections. Springer.

Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99, 7280–7287.

Brody, R.A., Shapiro, C.R., 1989. Policy failure and public support: The Iran-Contra affair and public assessment of President Reagan. Political Behavior 11, 353–369.

Chan, J., Suen, W., 2008. A Spatial Theory of News Consumption and Electoral Competition. The Review of Economic Studies 75, 699–728.

Chica, M., Barranquero, J., Kajdanowicz, T., Cordón, O., Damas, S., 2017. Multimodal optimization: an effective framework for model calibration. Information Sciences 375, 79–97.
Chica, M., Rand, W., 2017. Building agent-based decision support systems for word-of-mouth programs: A freemium application. Journal of Marketing Research 54, 752–767.
Chowanietz, C., 2011. Rallying around the flag or railing against the government? Political parties' reactions to terrorist acts. Party Politics 17, 673–698.
CIS, 2004. Spanish centre for sociological research, estudio 2555. CIS Data bank.
Clough, E., 2008. Still converging? A downsian party system without polls. Journal of Theoretical Politics 20, 461–476.
Colaresi, M., 2007. The benefit of the doubt: Testing an informational theory of the rally effect. International Organization 61, 99–143.
Consell de l'Audiovisual de Catalunya, 2004. 11-14M: la construcció televisiva (in Catalonian). Quaderns del CAC , 89–204.
de Holanda, G.M., Ávila, I.M., Martins, R.B., 2008. Mapping users' perspectives and outlining social impacts from digitalization of terrestrial tv in brazil. Telematics and Informatics 25, 19–35.
de Marchi, S., Page, S.E., 2014. Agent-Based Models. Annual Review of Political Science 17, 1–20.
Downs, A., 1957. An economic theory of political action in a democracy. Journal of Political Economy 65, 135–150.
Duggan, J., Martinelli, C., 2011. A Spatial Theory of Media Slant and Voter Choice. The Review of Economic Studies 78, 640–666.
Edwards, G.C., Mitchell, W., Welch, R., 1995. Explaining presidential approval: The significance of issue salience. American Journal of Political Science 39, 108–134.
Edwards, G.C., Swenson, T., 1997. Who rallies? the anatomy of a rally event. The Journal of Politics 59, 200–212.
Enelow, J.M., Hinich, M.J., 1984. The spatial theory of voting: An introduction. CUP Archive.
Enelow, J.M., Hinich, M.J., 1994. A test of the predictive dimensions model in spatial voting theory. Public Choice 78, 155–169.
Epstein, J.M., 2006. Generative social science: Studies in agent-based computational modeling. Princeton University Press.
Fishman, R., 2005. On the continuing relevance of the Weberian methodological perspective (with applications to the Spanish case of elections in the aftermath of terrorism). Working Paper of the Helen Kellogg Institute for International Studies , 1–28.
Fowler, J.H., Smirnov, O., 2005. Dynamic parties and social turnout: An agent-based model. American Journal of Sociology 110, 1070–1094.
Giráldez-Cru, J., Chica, M., Cordón, O., Herrera, F., 2020. Modeling agent-based consumers decision-making with 2-tuple fuzzy linguistic perceptions. International Journal of Intelligent Systems 35, 283–299.
González-Avella, J.C., Cosenza, M.G., Klemm, K., Eguíluz, V.M., San Miguel, M., 2007. Information feedback and mass media effects in cultural dynamics. Journal of Artificial Societies and Social Simulation 10, 9.
González, T.A., Granic, G.D., 2020. Spatial voting meets spatial policy positions: An experimental appraisal. American Political Science Review 114, 285—290.
Groeling, T., Baum, M.A., 2008. Crossing the water's edge: Elite rhetoric, media coverage, and the rally-round-the-flag phenomenon. The Journal of Politics 70, 1065–1085.
Grofman, B., 1985. The neglected role of the status quo in models of issue voting. The Journal of Politics 47, 230–237.
Herrera, F., Lozano, M., Verdegay, J.L., 1998. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. Artificial Intelligence Review 12, 265–319.
Hetherington, M.J., Nelson, M., 2003. Anatomy of a rally effect: George W. Bush and the war on terrorism. PS: Political Science & Politics 36, 37–42.
Iyengar, S., 1994. Is anyone responsible?: How television frames political issues. University of Chicago Press.
Kedar, O., 2005. When moderate voters prefer extreme parties: Policy balancingin parliamentary elections. American Political Science Review 99, 185–199.
Kedar, O., 2009. Voting for policy, not parties: How voters compensate for power sharing. Cambridge University Press.
Kollman, K., Miller, J.H., Page, S.E., 1998. Political parties and electoral landscapes. British Journal of Political Science 28, 139–158.
Krosnick, J.A., Brannon, L.A., 1993. The impact of the gulf war on the ingredients of presidential evaluations: Multidimensional effects of political involvement. American Political Science Review 87, 963–975.
Lago, I., Montero, J.R., 2006. The 2004 election in Spain: Terrorism, accountability, and voting. Institut de Ciències

Polítiques i Socials, Barcelona.
Landau, M.J., Solomon, S., Greenberg, J., Cohen, F., Pyszczynski, T., Arndt, J., Miller, C.H., Ogilvie, D.M., Cook, A., 2004. Deliver us from evil: The effects of mortality salience and reminders of 9/11 on support for President George W. Bush. Personality and Social Psychology Bulletin 30, 1136–1150.
Laver, M., 2005. Policy and the dynamics of political competition. American Political Science Review 99, 263–281.
Laver, M., 2011. Why vote-seeking parties may make voters miserable. Irish Political Studies 26, 489–500.
Laver, M., Schilperoord, M., 2007. Spatial models of political competition with endogenous political parties. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 1711–1721.
Lee, D., Kim, H., 2014. The effects of network neutrality on the diffusion of new internet application services. Telematics and Informatics 31, 386–396.
Lee, J., Kim, J., Kim, H., Hwang, J., 2020. Sustainability of ride-hailing services in china's mobility market: A simulation model of socio-technical system transition. Telematics and Informatics 53, 101435.
Lehrer, R., Schumacher, G., 2018. Governator vs. hunter and aggregator: A simulation of party competition with vote-seeking and office-seeking rules. PLOS ONE 13, 1–23.
Libai, B., Muller, E., Peres, R., 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. Journal of Marketing Research 50, 161–176.
Liu, F.C., 2007. Constrained opinion leader influence in an electoral campaign season: Revisiting the two-step flow theory with multi-agent simulation. Advances in Complex Systems 10, 233–250.
López García, G., 2004. El 11-M y el consumo de medios de comunicación (in Spanish). Sala de Prensa 71.
Macal, C.M., North, M.J., 2005. Tutorial on agent-based modeling and simulation, in: Proceedings of the 37th conference on Winter simulation, ACM. pp. 2–15.
Michavila, N., 2005. Guerra, terrorismo y elecciones: incidencia electoral de los atentados islamistas en madrid. Boletín Elcano , 34.
Montalvo, J.G., 2011. Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. The Review of Economics and Statistics 93, 1146–1154.
Montero, J.R., Lago, I., 2009. Voto, terrorismo y rendición de cuentas: las elecciones de 2004 en España (in Spanish). Il cittadino-elettore in Europa e America. Studi in onore di Giacomo Sani. Bologna: Il Mulino , 137–175.
Moscato, P., 1989. On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Technical Report 826. Caltech Concurrent Computation Program. Pasadena, USA.
Moya, I., Chica, M., Cordón, O., 2019. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization. Decision Support Systems 124, 113111.
Moya, I., Chica, M., Sáez-Lozano, J.L., Cordón, O., 2017. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 spanish elections. Knowledge-Based Systems 123, 200–216.
Mueller, J.E., 1973. War, presidents, and public opinion. John Wiley & Sons.
Muis, J., 2010. Simulating political stability and change in the netherlands (1998-2002): an agent-based model of party competition with media effects empirically tested. Journal of Artificial Societies and Social Simulation 13, 4.
Newman, M., Barabási, A.L., Watts, D.J., 2006. The structure and dynamics of networks. Princeton University Press.
Olmeda, J.A., 2005. Fear or falsehood? Framing the 3/11 terrorist attacks in Madrid and electoral accountability. Boletín Elcano .
Papageorgiou, E.I., 2013. Fuzzy cognitive maps for applied sciences and engineering: from fundamentals to extensions and learning algorithms. volume 54. Springer Science & Business Media.
Piolatto, A., Schuett, F., 2015. Media competition and electoral politics. Journal of Public Economics 130, 80–93.
Plümper, T., Martin, C.W., 2008. Multi-party competition: A computational model with abstention and memory. Electoral Studies 27, 424–441.
Rabinowitz, G., Macdonald, S.E., 1989. A directional theory of issue voting. American Political Science Review 83, 93–121.
Rand, W., Rust, R.T., 2011. Agent-based modeling in marketing: Guidelines for rigor. International Journal of Research in Marketing 28, 181–193.
Randahl, D., 2018. Terrorism and public opinion: The effects of terrorist attacks on the popularity of the president of the united states. Terrorism and Political Violence 30, 373–383.
Robbins, J., Hunter, L., Murray, G.R., 2013. Voters versus terrorists: Analyzing the effect of terrorist events on voter turnout. Journal of Peace Research 50, 495–508.
Rose, W., Murphy, R., Abrahms, M., 2007. Does terrorism ever work? The 2004 Madrid train bombings. International Security 32, 185–192.

Strömberg, D., 2004. Mass Media Competition, Political Competition, and Public Policy. The Review of Economic Studies 71, 265–284.
Strömberg, D., 2015. Media and politics. Annual Review of Economics 7, 173–205.
Sudo, Y., Kato, S., Mutoh, A., 2013. The impact of exchanging opinions in political decision-making on voting by using multi-agent simulation, in: PRIMA 2013: Principles and Practice of Multi-Agent Systems. Springer, pp. 340–354.
Torcal, M., Rico, G., 2004. The 2004 Spanish General Election: In the Shadow of Al Qaeda. South European Society and Politics 9, 107–121.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.
Wilensky, U., Rand, W., 2015. Introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo. MIT Press.
Willer, R., Adams, N., 2008. The threat of terrorism and support for the 2008 presidential candidates: Results of a national field experiment. Current Research in Social Psychology 14, 1–22.
Wright, M., Sengupta, P., 2015. Modeling oligarchs' campaign donations and ideological preferences with simulated agent-based spatial elections. Journal of Artificial Societies and Social Simulation 18, 3.
Wu, F., Huberman, B.A., 2007. Novelty and collective attention. Proceedings of the National Academy of Sciences 104, 17599–17601.
Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 599–608.
Zaller, J.R., 1992. The Nature and Origins of Mass Opinion. Cambridge Studies in Public Opinion and Political Psychology, Cambridge University Press.

3 Coral reefs optimization algorithms for agent-based model calibration

- I. Moya, E. Bermejo, M. Chica, O. Cordón. Coral reefs optimization algorithms for agent-based model calibration, Engineering Applications of Artificial Intelligence, vol. 100, 104170, 2021. DOI: 10.1016/j.engappai.2021.104170.

  - State: Published.
  - Impact Factor (JCR 2020): 6.212.
  - Category: ENGINEERING, MULTIDISCIPLINARY. Order: 7/91. Q1. D1.

Coral reefs optimization algorithms for agent-based model calibration

Ignacio Moya[*,a], Enrique Bermejo[a], Manuel Chica[a,b], Óscar Cordón[a]

[a] *Andalusian Research Institute DaSCI "Data Science and Computational Intelligence", University of Granada, 18071 Granada, Spain*

[b] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia*

Abstract

Calibrating agent-based models involves estimating multiple parameter values. This can be performed automatically using automatic calibration but its success depends on the optimization method's ability for exploring the parameter search space. This paper proposes to carry out this process using coral reefs optimization algorithms, a new branch of competitive bio-inspired metaheuristics that, beyond its novel metaphor, has shown its good behavior in other optimization problems. The performance of these metaheuristics for model calibration is evaluated by conducting an exhaustive experimentation against well-established and recent evolutionary algorithms, including their hybridization with local search procedures. The study analyzes the calibration accuracy of the metaheuristics using an integer coding scheme over a benchmark of 12 problem instances of an agent-based model with an increasing number of decision variables. The outstanding performance of the memetic coral reefs optimization is reported after performing statistical tests to the results.

Keywords— Evolutionary computation, metaheuristics, coral reefs optimization, model calibration, agent-based modeling.

1. Introduction

Agent-based modeling (ABM) (Epstein, 2006; Janssen and Ostrom, 2006; Wilensky and Rand, 2015) is a well-established methodology for designing and simulating computational models that relies on autonomous entities called agents. The behavior of these artificial agents is managed through the definition of simple rules and their interactions with other agents. The ability of ABM for recreating complex and emerging dynamics through the aggregation of the agents' rules and social interactions has made them receive increased attention in the last few years (Farmer and Foley, 2009; Waldrop, 2018; Coates et al., 2019). However, building agent-based models is difficult because the values of a large number of parameters must be set in order to design the model. In addition, the modeler is commonly forced to estimate many of those parameter values due to lack of appropriate data. The process of adjusting these parameter values to correctly replicate the desired dynamics is addressed as model calibration (Chica et al., 2017; Oliva, 2003).

Model calibration can be performed automatically using automatic calibration, a computationally intensive process that adjusts the model's parameters using a optimization method. This

[*]Corresponding author

Email addresses: imoya@ugr.es (Ignacio Moya), enrique.bermejo@decsai.ugr.es (Enrique Bermejo), manuelchica@ugr.es (Manuel Chica), ocordon@decsai.ugr.es (Óscar Cordón)

optimization method considers an error measure for comparing the model's output and the real data from the phenomena simulated by the model (Oliva, 2003; Sargent, 2005). Since the values of different model parameters are typically unrelated, it is desirable to use non-linear optimization methods such as metaheuristics (Talbi, 2009) that can search through the whole parameter space (Chica et al., 2017; Stonedahl and Rand, 2014). Nevertheless, the selection of the metaheuristic for carrying out the calibration process heavily determines the quality of the resulting model parameter configuration as the model's accuracy relies on the ability of the method for exploring the parameter search space.

The current manuscripts proposes to carry out the calibration process using novel and competitive bio-inspired metaheuristics based on coral reefs: coral reefs optimization (CRO) (Salcedo-Sanz et al., 2014) and coral reefs optimization with substrate layers (CRO-SL) (Salcedo-Sanz et al., 2016b). Coral reefs-based metaheuristics emulate the formation and reproduction of coral reefs, resulting in an optimization algorithm with a powerful trade-off between exploration and exploitation of the search space. CRO-SL is an enhanced version of CRO that also includes a cooperative co-evolution scheme and has shown outstanding performance tackling some complex optimization problems, such as (Bermejo et al., 2018; Camacho-Gómez et al., 2019; Garcia-Hernandez et al., 2020a,b; Salcedo-Sanz et al., 2019). Taking these good results as a base, the performance of the coral reefs-based metaheuristics should be extensively analyzed in the significantly complex problem of calibrating ABM. The authors acknowledge that the recent surge of novel bio-inspired algorithms has been subject of controversy due to the lack of scientific rigor behind some of these algorithms (Sörensen, 2015). However, CRO-SL is not one of those waste-of-time "novel" algorithms because it can be justified getting past the novel metaphor and focusing on its actual design and performance (see Bermejo et al. (2018)).

This study analyzes the behavior of the coral reefs-based metaheuristics using an integer coding scheme for estimating the ABM model's parameters. This approach is selected because representing the values of real parameters following an integer-coded scheme allows the modeler to set the desired granularity for the parameter values (Chica et al., 2017). This increases the control of the modeler over the calibration process and eases the management of the calibrated parameters, since handling long-tailed real values could be troublesome. Additionally, CRO and CRO-SL are extended by hybridizing them with local search procedures with the goal of analyzing their improvement with respect to the original algorithms. Thus, memetic variants (Moscato, 1989; Moscato et al., 2004) are designed and implemented for the metaheuristics.

The performance of the coral reefs-based metaheuristics is evaluated by conducting an exhaustive comparison of CRO and CRO-SL against a well-known metaheuristic that have been previously applied for calibrating agent-based models such as differential evolution (DE) (Storn and Price, 1997). DE is one of the most commonly used metaheuristics because it obtains good results despite of being easy to use and simple to implement. However, there are newer and more advanced metaheuristics in the field of evolutionary computation that can obtain good results calibrating ABM. For example, success-history based adaptive differential evolution with linear population size reduction (L-SHADE) (Tanabe and Fukunaga, 2014) and restart CMA-ES with increasing population size (IPOP-CMA-ES) (Auger and Hansen, 2005) are two advanced and highly competitive metaheuristics that have established as a reference for evolutionary computation, where variants of these algorithms are constantly being developed (Molina et al., 2017). Thus, the performance of CRO and CRO-SL is compared against L-SHADE and IPOP-CMA-ES in addition to DE. On the one hand, L-SHADE is a recent metaheuristic extending the original DE design which is becoming

the main DE variant due to its good performance across different benchmarks (Molina et al., 2017). On the other hand, IPOP-CMA-ES is a highly competitive evolutionary algorithm extending the original and very extended CMA-ES metaheuristic (Hansen, 1997) that has proven outstanding results solving complex problems (Biedrzycki, 2017; Biswas and Biswas, 2017; Marín, 2012; Molina et al., 2017).

In order to fairly benchmark the proposed coral reefs-based metaheuristics with the selected evolutionary methods, memetic variants for L-SHADE and IPOP-CMA-ES have also been implemented. The design of this study compares the performance of the metaheuristics when calibrating 12 scenarios. Each of these scenarios is defined by a different instance of an ABM for marketing, with every instance having different dimensionality (i.e., number of decision variables). Therefore, the mentioned calibration instances consider between 24 and 129 parameters to be calibrated, allowing the design of a test suite composed of several ABM calibration problem instances involving several large-dimension optimization problems where a significant number of parameters (beyond 100 parameters) are to be solved. The study also considers two baseline non-evolutionary methods: a hill climbing (HC) (Russell et al., 1995) and a random search procedure. Thus, the resulting battery of experiments is based on five global search metaheuristics, four memetic algorithms, and two baseline methods. Additionally, different statistical tests are performed for evaluating the significance of the results. Hence, the main contributions of the present study are:

- The evaluation of the performance of coral reefs-based metaheuristics when calibrating ABM scenarios using an integer coding scheme.
- The analysis of the improvement obtained by their hybridization with local search procedures.
- The design of an appropriate experimental setup for the study, which considers eleven calibration methods and composed of 12 ABM scenarios with up to 129 decision variables.

The paper is structured as follows. Section 2 reviews the background on model calibration and discusses the related work in the literature. Section 3 introduces the description of the ABM used in the experimentation. Section 4 details the coral reefs-based metaheuristics, the considered coding scheme, and the design of the memetic variants. Section 5 presents the problem instances, competing metaheuristics, and experimental setup. Finally, Section 6 reports the experimental analysis and Section 7 discusses the final remarks.

2. Background and related work

There is a need to carefully validate computational models before they can be used. In this regard, the calibration of the model is recognized as an important step during model validation (Chica et al., 2017; Oliva, 2003). The modeler can perform this process manually similarly to a global sensitivity analysis (ten Broeke et al., 2016), where the modeler repeatedly simulates the model and tunes its parameters based on the observed output. However, this approach is impracticable for many realistic models, which are characterized for considering many parameters. Instead, modelers tend to employ automatic calibration, which is an effective approach to model calibration (Chica et al., 2017; Oliva, 2003).

Automatic calibration techniques have been applied for calibrating the parameters of computational non-linear models from different areas. A few examples would be market modeling (Chica and Rand, 2017; North et al., 2010), crowd modeling (Zhong and Cai, 2015), traffic simulation (Kim

and Rilett, 2003; Ngoduy and Maher, 2012), or growth modeling (César Trejo Zúñiga et al., 2014). Some approaches consider the use of exact methods like simplex-based (Kim and Rilett, 2003) or gradient-based (Thiele et al., 2014) methods. However, these approaches are employed for calibrating a relatively low number of parameters (i.e., no more than 20 calibration parameters) and its application to models involving more than 100 parameters seems computationally prohibitive. Another approach is the cross entropy method (Ngoduy and Maher, 2012), a stochastic optimization algorithm which results effective dealing with calibration problems with multi-local optima, but it is also unclear how this approach can perform when dealing with more than 20 calibration parameters.

In contrast, the use of metaheuristics is more convenient for the calibration problem when having higher dimensionality. There are several contributions addressing the application of metaheuristics for model calibration and parameter estimation. For instance, the interested readers can find genetic algorithms (Dai et al., 2009), evolution strategies (Muraro and Dilão, 2013; César Trejo Zúñiga et al., 2014), and several versions of differential evolution (LaTorre et al., 2019; Zhong and Cai, 2015; César Trejo Zúñiga et al., 2014). In terms of the ABM calibration, the use of metaheuristics is clearly the most extended approach. Thus, there are examples of metaheuristics for calibrating ABM that tackle models designed for different areas, such as social and biological sciences (Calvez and Hutzler, 2006; Canessa and Chaigneau, 2015; Chica et al., 2017; Fabretti, 2013; Herrmann and Savin, 2015; Malleson et al., 2014; Moya et al., 2019, 2017). However, one can see from these examples that different versions of genetic algorithms (Back et al., 1997) are usually the default approach, which could be improved by employing recent and more powerful evolutionary metaheuristics. Unfortunately, the state of the art in ABM calibration does not consider a reference benchmark for comparing these methods. Moreover, none of these previous efforts consider an exhaustive comparison of several metaheuristics for ABM calibration, as it is done in this manuscript.

Additionally, efficiently calibrating ABM is troublesome since the model needs to be simulated in order to evaluate the quality of a given set of parameter values, thus leading to a simheuristic approach (Chica et al., 2020). It can be noted the use of surrogate models for reducing the computational cost of estimating these values (van der Hoog, 2018; Lamperti et al., 2018). Specifically, machine learning algorithms have been employed for training a fast surrogate model that responds similarly to the changes of parameter values of the original model. Nevertheless, it is unknown how this approach could perform with ABM having a high number of parameters, but authors of the latter publications claimed it can work with more than 30 parameters.

Finally, the use of surrogate fitness functions for tackling expensive optimization has also been addressed in evolutionary optimization (Bhattacharya, 2013; Branke et al., 2017). In this case, the authors propose to replace the original fitness function by an approximate (faster) function. This surrogate function is designed relaxing the conditions of the original fitness function. For example, a fitness function for an ABM could be relaxed by reducing the simulation steps of the model or reducing the number of agents. However, this approach can be problematic from the model calibration point of view because the use of an appropriate surrogate function is problem-dependent (i.e., specific for each ABM) and its validation is not straightforward.

3. Model description

This section introduces the composition of the ABM for marketing employed in the following experiments (Moya et al., 2019). First, the general structure of the model and the behavior of

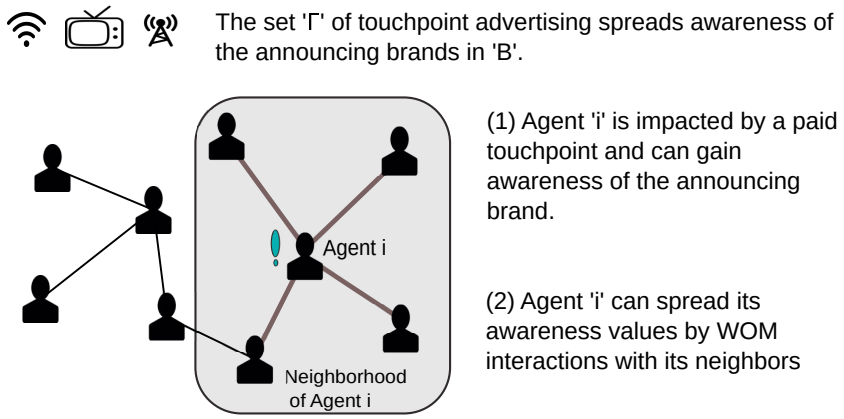


Figure 1: Summary and main elements of the ABM model. Agents exposed to the advertising of the paid touchpoints can gain awareness of the brand announced and talk about it to their neighbors.

the agents are presented in Section 3.1. Section 3.2 introduces the features of the artificial social network where the agents are embedded. Section 3.3 presents the modeling of external influences as paid touchpoints. Finally, the parameters selected for calibration and the fitness function for the problem are summarized in Section 3.4.

3.1. General structure and definition of agents' behavior

The ABM simulates a given number of weeks ($t_{\max}$) of a market composed of $|B|$ competing brands. It considers the behavior of z agents exposed to the advertising of a set of Γ paid touchpoints and the social interaction between the agents. The model uses a time-step of a week and involves the computation of two outputs or key performance indicators (KPIs): brand awareness and number of word-of-mouth (WOM) interactions among the consumers, that is referred as WOM volume. These KPIs are selected due to their importance with respect to market expansion (Libai et al., 2013; Macdonald and Sharp, 2000). A summary of the core elements of the model is presented in Figure 1.

Agents' awareness values are modeled using the state variable $a_i^b \in \{0, 1\}$. $a_i^b(t) = 1$ represents that the agent i is aware of brand b at time step t while $a_i^b(t) = 0$ stands for the agent not being aware of brand b. This variable is initialized using initial awareness ($a^b(0) \in [0, 1]$), a parameter that represents the fraction of agents that have awareness of the brands at the beginning of the simulation, which satisfies $a^b(0) = \frac{1}{z}\sum_{i=1}^{z} a_i^b(0)$.

The awareness values of the agents are dynamic because awareness may be lost or gained for any of the brands at each step of the simulation. For example, an agent may gain awareness of a brand due to advertising or due to the interaction with its neighbors from the social network. In contrast, brand awareness can be lost because of a deactivation process (Wu and Huberman, 2007; Yang and Leskovec, 2010) if it is not reinforced by new stimulus.

These losing/gaining effects are modeled by including additional parameters. The parameter setting the probability for an agent to deactivate the awareness of a brand is called awareness deactivation ($d \in [0, 1]$). This process is modeled by checking each brand b the agent i is aware of ($a_i^b(t) = 1$) at the start of each step t and deactivating it ($a_i^b(t) = 0$) with a probability d. The modeling processes of the awareness obtained due to the effect of WOM diffusion and paid touchpoints are further described in Sections 3.2 and 3.3, respectively.

Additionally, each agent saves its interactions with the social network for computing the WOM volume generated by each brand ($\omega_i^b(t)$). Thus, whenever an agent begins a diffusion process by talking with its neighborhood, $\omega_i^b(t)$ will be updated by increasing it with the total number of conversations, which corresponds to the number of agents' neighbors. Finally, global variable $\omega^b(t)$ is updated at every step of the simulation for all the brands.

3.2. Agents' social network and word-of-mouth interactions

The agent population is placed in an artificial social network (Barabási and Albert, 1999; Watts and Strogatz, 1998) modeled as a scale-free network since this topology matches those of many real networks (Barabási and Albert, 1999; Newman et al., 2006). In this network model the degree distribution follows a power-law (Barabási and Albert, 1999), where few nodes (referred as hubs) are highly connected but most nodes consider a reduced number of neighbors. The network is generated using the Barabasi-Albert preferential attachment algorithm (Barabási and Albert, 1999), which considers a parameter m for regulating the network's growth rate while maintaining a power-law degree distribution of $P(k) = k^{-3}$.

During the simulation, the agents can talk with their neighbors in the social network and spread the awareness of the brands they are aware of. This social interaction is modeled as a contagion process where the information flows through the network depending on the number of connections of the different nodes (Newman et al., 2006). Every agent i has a talking probability $p_i^b(t) \in [0, 1]$ to spread its awareness at time step t for every brand b where $a_i^b(t) = 1$. This talking event involves all the neighbors of agent i that have a chance of activate its awareness through a contagion process. This contagion effect is modeled using the WOM awareness impact parameter ($\alpha^{\text{WOM}} \in [0, 1]$), that models the probability for a neighbor agent j to gain awareness of a brand after interacting with agent i.

3.3. Modeling of paid touchpoints

The external influences of the agents are modeled as paid touchpoints (i.e., global mass media (González-Avella et al., 2007)) and mainly act as brand advertising. Using a similar modeling approach to that already applied for social interactions, different parameters are considered for representing the natural differences among the touchpoints (i.e., television, radio, and press). Paid touchpoints γ from Γ can influence any agent of the model at random. The maximum percentage of reached agents depends on the amount invested by the brand and the touchpoints' capability for reaching the agents, which is bounded by the properties of the touchpoint itself. For example, an ad included in the press will impact the population percentage that reads the press at best. These different properties are modeled by a reach parameter ($r_\gamma \in [0, 1], \forall \gamma \in \Gamma$), which bounds the maximum number of agents that a touchpoint γ can try to influence at a single step.

Touchpoint advertising is modeled according to an investment measure called gross rating points (GRPs). In advertising (Farris et al., 2010), a GRP measures the potential of the campaigns scheduled in mass media assuming that one GRP represent enough impressions for reaching 1% of the target population. The variable $\chi_\gamma^b(t)$ models the investment units in GRPs for touchpoint γ by brand b and time step t. Because increasing the awareness or the number of conversations of the target population using paid touchpoints implies a monetary cost, the brands need to define their marketing mix considering that each touchpoint has a different cost for the invested GRP units. Touchpoint advertising is scheduled by computing the number of actual impressions for each band and step using the GRP values and the size of the agent population. Then, these impressions

Market parameters	
z: number of agents running in the model	$\|B\|$: number of considered brands
$\|\Gamma\|$: number of paid touchpoints	$t_{\max}$: number of steps of the simulation
$a^b(0)$: Initial awareness for brand b	d: awareness deactivation probability
WOM parameters	
$p_i^b(0)$: initial talking probability, same value for each brand b	α^{WOM}: awareness impact for social interactions
	m: parameter for social network generator
Touchpoint parameters	
χ_γ^b: GRP units invested by brand b in touchpoint γ	r_γ: reach for paid touchpoint γ
	$d\tau_\gamma$: buzz decay for paid touchpoint γ
α_γ: awareness impact for paid touchpoint γ	τ_γ: buzz increment for paid touchpoint γ

Table 1: Summary of the parameters of the agent-based model for marketing scenarios.

are individually assigned at random between the agents respecting the reach restrictions for the touchpoint.

Similarly to the social interactions, each touchpoint is assigned an awareness impact parameter ($\alpha_\gamma \in [0,1], \forall \gamma \in \Gamma$) that models the probability of the agent to activate its awareness of the ad brand after a single touchpoint impact. In addition, the effect of the advertising transmitted by paid touchpoints can create a viral effect in the reached agent, as done in (Moya et al., 2017). The buzz effect created by the touchpoints produces an increment of the number of conversations regarding the announcing brand by increasing the talking probability (p_i^b) of the impacted agents. This effect is modeled using a buzz increment parameter (τ_γ) defined for each touchpoint $\gamma \in \Gamma$. The increment produced on the agents' talking probability is calculated as a percentage increment over the initial talking probability ($p_i^b(0)$) of the agent. Similarly to the awareness, buzz effect decays over time if it is not reinforced. This effect is modeled by Equation 1, where the action of the buzz decay parameter ($d\tau_\gamma$) erodes the increment of talking probability (σ_γ) previously supplied to the agent by touchpoint γ. Finally, a summary of the variables and parameters of the calibrated ABM is shown at Table 1.

$$p_i^b(t+1) = p_i^b(t) - \sigma_{i\,\gamma}^b(t) \cdot d\tau_\gamma + p_i^b(0) \cdot \tau_\gamma, \qquad (1)$$
$$\text{where } \sigma_{i\,\gamma}^b(t) = \sum_{s=1}^{t} (p_i^b(s) - p_i^b(0) \cdot \tau_\gamma).$$

3.4. Calibration parameters and fitness function

The parameters selected for automatic calibration are those that control either the dynamics of the agents' awareness values or their number of conversations since those are sensitive parameters that are hard to estimate by the modeler using the available data (Moya et al., 2019). The calibration process assigns each of the selected model parameters to a single decision variable limiting the range of possible parameter values to a real-coded interval, $[0, 1]$. Additionally, the final set of parameters that are selected for calibration is determined by the size of the model

instance (i.e., the number of paid touchpoints $|\Gamma|$): three parameters for each touchpoint plus the three social parameters. Briefly, for each defined paid touchpoint $\gamma \in \Gamma$, the calibration considers its buzz increment (τ_γ), buzz decay ($d\tau_\gamma$), and awareness impact (α_γ). In addition, the initial talking probability ($p_i^b(0)$), social awareness impact (α^{WOM}), and awareness deactivation (d) are also calibrated. Thus, the overall number of parameters being calibrated is computed as $(|\Gamma|+1)\cdot 3$.

Finally, Equations 2 and 3 define the selected fitting functions for the historical data of the awareness (f_1) and WOM volume (f_2). Both functions compute the deviation error comparing the simulated outputs values with the target data of the objectives. This deviation is computed using the standard mean absolute percentage error function, where $\widetilde{a}$ represents the target awareness values and $\widetilde{\omega}$ represent the target WOM volume values. The simulated values are the result of averaging multiple independent Monte-Carlo simulations in order to account for the random nature of ABMs. Using the latter fitting functions, the objective function f is defined as its weighted combination using parameter $\beta \in [0,1]$: $f = \beta \cdot f_1 + (1-\beta)\cdot f_2$.

$$f_1 = \frac{100}{t_{\max}\cdot |B|}\sum_{b=1}^{|B|}\sum_{t=1}^{t_{\max}}\left|\frac{a^b(t)-\widetilde{a}^b(t)}{\widetilde{a}^b(t)}\right|, \tag{2}$$

$$f_2 = \frac{100}{t_{\max}\cdot |B|}\sum_{b=1}^{|B|}\sum_{t=1}^{t_{\max}}\left|\frac{\omega^b(t)-\widetilde{\omega}^b(t)}{\widetilde{\omega}^b(t)}\right|. \tag{3}$$

4. Coral reefs-based metaheuristics for model calibration

This section introduces the design of CRO and CRO-SL metaheuristics for calibrating ABM. Then, the coding scheme considered for the problem and the memetic design due to their hybridization with a local search procedure are described in detail.

4.1. CRO and CRO-SL algorithms

CRO (Salcedo-Sanz et al., 2014; Salcedo-Sanz, 2017; Salcedo-Sanz et al., 2017b) is an advanced evolutionary metaheuristic based on the processes occurring in a coral reef. In the analogy, a coral reef is represented by a two dimensional grid able to allocate a colony of corals (i.e., a population with variable size), where a coral represents a solution to the optimization problem. The CRO algorithm initializes some positions of the grid with random solutions, leaving the rest available spots empty. The second phase simulates the processes of coral reproduction and reef formation which are recreated by applying known evolutionary process in four sequential stages:

1. **External reproduction.** This stage consists of the random selection of a fraction (F_b) of the existing solutions and the generation of a pool of child solutions. These new solutions are generated by applying a crossover operator or any other exploration strategy with two existing solutions as parents. Once a solution has been selected as parent, it is not chosen for reproduction purposes during an iteration.

2. **Internal reproduction.** This reproduction is modeled as a random mutation mechanism that takes place on the remaining fraction of solutions ($1-F_b$). A percentage P_i of the solution is mutated. The subset of mutated solutions is added to the pool of children solutions.

3. **Replacement.** Once the children solutions are formed either through external or internal reproduction, they will try to set in the grid. Each solution in the pool will randomly try to set in a position of the grid and settle if the location is free. Otherwise, the new solution will replace the existing one only if its fitness is better. Each child in the pool can attempt to occupy a position at each iteration only a determined number of tries η.

4. **Elimination.** At the end of each reproduction iteration, a small number of solutions in the grid are discarded, thus releasing space in the grid for next generation. The elimination operator is applied with a very small probability (P_d) to a fraction (F_d) of the grid size with worse fitness.

However, there are other relevant interactions in real reef ecosystems that can be incorporated into the CRO approach for improving its performance in optimization and search problems. For example, different recent studies have shown that successful recruitment in coral reefs (i.e., successful settlement and subsequent survival of larvae) strongly depends on the type of substrate on which they fall after the reproduction process (Vermeij, 2005). This specific characteristic of coral reefs was first included in the CRO algorithm in (Salcedo-Sanz et al., 2017b) in order to solve different instances of the Model Type Selection Problem for energy applications, resulting in CRO-SL (CRO with substrate layers). In (Salcedo-Sanz et al., 2016b), CRO-SL is enhanced with several substrate layers providing a different search procedure.

The inclusion of substrate layers in CRO can be carried out in a straightforward way: the artificial reef considered in the CRO is redefined in such a way that each cell of the square grid Ψ is now associated with a different exploration layer that indicates which structure it belongs to (search operator in this case). Each solution in the grid is then processed in a different way (with a different search operator) depending on the region (specific layer) in which it falls after the reproduction process. Figure 2 illustrates the flowchart diagram of the CRO-SL algorithm. Any exploration operator can be integrated in the algorithm as a substrate layer as long as it follows the determined coding scheme. In order to select a well-performing combination of substrate layers for the ABM calibration problem tackled in the current contribution, a preliminary test was developed considering different search operators known to perform well when tackling the addressed problem. The considered exploration operators are detailed below.

- **Uniform mutation**. The uniform random mutation is a traditional mutation operator that replaces the value of a gene for a given individual by a random value. This value is generated using an uniform distribution from an interval defined by lower and upper bounds.

- **Random walk mutation**. This operator modifies the values of a given individual by including neighborhood information into the mutation process. In the context of the present problem, the neighbors of a solution are accessed increasing or decreasing its values. This way, this operator modifies the value of each parameter of the individual solution with probability p_m by randomly moving to a neighbor. After moving to a random neighbor, the operator will keep moving to a random neighbor until the probability check fails.

- **Simulated binary crossover** (SBX). SBX (Deb and Agrawal, 1994) performs the crossover operation emulating the behavior of a single-point crossover from binary-encoding. This operator works as follows: given two parents $P_1 = (p_{11}, ..., p_{1n})$ and $P_2 = (p_{21}, ..., p_{2n})$, SBX generates two springs $O_1 = (o_{11}, ..., o_{1n})$ and $O_2 = (o_{21}, ..., o_{2n})$ as $o_{1i} = \bar{X} - \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$

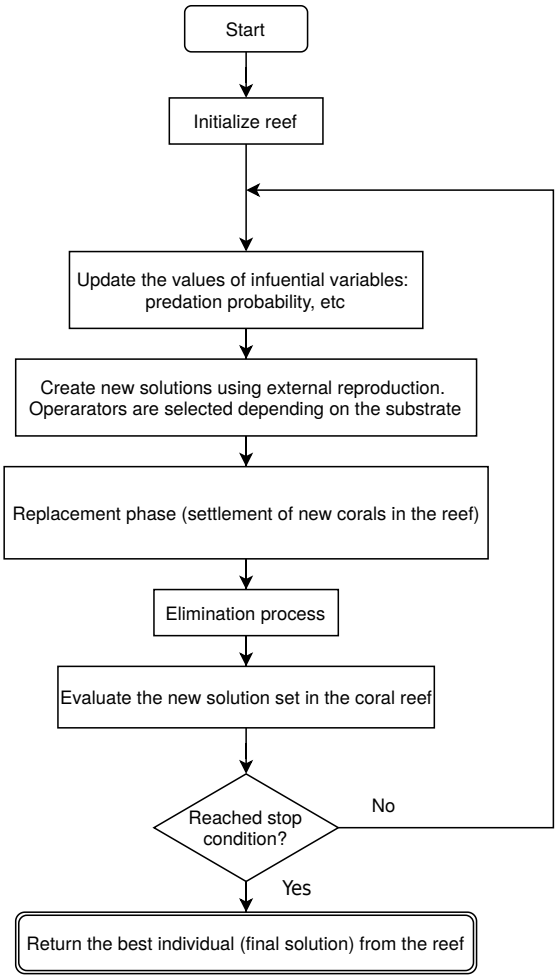


Figure 2: Flowchart of the CRO-SL algorithm, which modifies the reproduction stages of the original CRO by selecting the operator mechanisms depending on the substrate layers.

and $o_{2i} = \bar{X} + \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$, where $\bar{X} = \frac{1}{2}(p_{1i} + p_{2i})$. $\bar{\beta}$ is a random number fetched from a probability distribution which is employed as a spread factor.

- **BLX-α crossover**. BLX-α (Eshelman and Schaffer, 1993) takes new values from the interval defined by $[c_{\min} - I\alpha, c_{\max} + I\alpha], c_{\max} = \max(v_i^1, v_i^2)$, $c_{\min} = \min(v_i^1, v_i^2)$ and $I = c_{\max} - c_{\min}$. v_i^1, v_i^2 represent the decoded values from the chromosome of the parents. In this regard, α defines the level of exploration for the operator. For example, if α is set to 0, BLX-α works as a flat crossover.

Nevertheless, it is noted that the recent surge of novel bio-inspired algorithms has been subject of controversy (Sörensen, 2015) due to the lack of scientific rigor behind some of these algorithms. CRO-SL is not one of those waste-of-time "novel" algorithms, as it can be justified beyond the novelty of the metaphor and only focusing on its actual design and performance (Bermejo et al., 2018). The novelty of CRO-SL resides in an excellent exploration-exploitation trade-off and robustness as a consequence of the combination of multiple search patterns in its scheme. Thus, the CRO-SL

general approach aims for a grid-based competitive co-evolution (Floreano and Mattiussi, 2008) in just one population where each substrate layer represents a different search process (different models, operators, parameters, constraints, repairing functions, etc.). Therefore, CRO-SL is a powerful multi-method ensemble where multiple search strategies coexist in the same population and can be classified as a low-level competitive single population approach (Del Ser et al., 2019; Wu et al., 2019). Such design provides CRO-SL with an exceptional versatility, being able to adapt different approaches to tackle a wide variety of problems while considering a single, general algorithmic scheme.

Finally, CRO-SL usually converges quickly to high quality solutions even in multi-modal search spaces, being suitable for computationally expensive optimization problems both satisfying quality and computation time constraints. However, its performance varies significantly depending on the CRO's parameters and the different substrates included in the simulated reef. This idea of CRO with substrate layers as a competitive co-evolution algorithm was successfully tested in different applications and problems such as micro-grid design (Salcedo-Sanz et al., 2016a), vibration cancellation in buildings (Salcedo-Sanz et al., 2017a), 3D medical image registration (Bermejo et al., 2018), and in the evaluation of novel non-linear search procedures (Salcedo-Sanz, 2016).

4.2. Coding scheme

As previously described in Section 3.4, the presented design assigns each of the selected model parameters to a single decision variable. Thus, during the metaheuristics execution each candidate solution consists of an array of values that represent the values of the parameters being calibrated. This design considers an integer coding scheme which encodes the individual solutions as discrete values. This coding scheme was selected because it provides a good trade-off between precision and usability, since it does not require the users for precisely adjusting long-tailed real values.

This integer coding scheme samples a given continuous interval into several discrete values following Equation 4, where the number of possible values of parameter i is computed using a minimum parameter value ($\min_i$), a maximum parameter value ($\max_i$), and a step (ϵ_i) value that regulates the granularity of the search procedure. Figure 3 shows the representation of an individual solution using the integer coding scheme for an example model instance which considers three touchpoints using $\min = 0$, $\max = 1$, and $\epsilon = 0.001$ for each parameter in the example.

$$n_i = \left\lfloor \frac{\max_i - \min_i}{\epsilon_i} \right\rfloor + 1. \tag{4}$$

α_1	α_2	α_3	τ_1	τ_2	τ_3	$d\tau_1$	$d\tau_2$	$d\tau_3$	$p^b(0)$	α^{WOM}	d
20	15	50	100	150	220	120	50	230	200	100	10

Figure 3: Example representation of a model instance which considers three paid touchpoints using an integer coding scheme. The genes of the chromosome represent real parameters limited to $[0, 1]$.

4.3. Memetic algorithms

Memetic algorithms (Moscato, 1989) combine the exploration capability of global search methods with the intensification of local search procedures, improving the quality of the candidate

solutions found during the regular execution of the algorithm. The most important design issue is to define a trade-off between the local search procedure and the global search method (Ishibuchi et al., 2003). Some common approaches to memetic algorithms apply this local search refinement to every solution obtained during its run but this strategy can be time-consuming and it has proven to not always lead to the best performing memetic design (Krasnogor and Smith, 2000). Therefore, the selected approach performs local search refinement selectively with a probability p_{LS} (Krasnogor and Smith, 2000; Lozano et al., 2004). Thus, all the solutions in the population are candidates for receiving local search refinement since each solution runs a probability check at the end of a generation.

The designed memetic algorithms are the result of the hybridization of CRO and CRO-SL with a local search component. This local optimizer implements a hill climbing strategy. It refines the quality of the individuals by increasing or decreasing the encoded parameter values by a given quantity specified by a step parameter (ϵ). The result of applying this operation over an individual is called a neighbor generation. Thus, the set of all possible neighbors for a given solution is called its neighborhood. The local search component moves to the first neighbor that improves the current solution quality until no neighbor shows an improvement. This search is repeated for a given number of iterations since it is integrated with the global search of the coral reefs-based metaheuristics. The memetic algorithms proposed for CRO and CRO-SL are labeled MCRO and MCRO-SL, respectively.

5. Experimental details

This section reviews the design of the experimentation. First, the model instances considered for the experiments are presented in Section 5.1. Then, the selected metaheuristics for benchmarking the coral reefs-based methods are introduced in Section 5.2. Afterwards, Section 5.3 reviews the experimental setup and the parameter configuration of the different metaheuristics.

5.1. Calibration scenarios and datasets

The experimentation considers 12 different instances of the ABM model. These model instances were generated using an initial *baseline* instance, referred as P1(24), which models a real banking marketing scenario from one of the authors' research contracts. The rest of the instances are synthetically generated increasing the complexity of the initial instance. Each model variation introduces additional touchpoints that are built from the initial ones by modifying its investment values. In addition, each model includes a perturbation of the target historical values for both KPI, awareness and WOM. Each of the newly generated instances increases the dimensionality of the previous one, including new decision variables that enable a more complete comparison of the different metaheuristics.

The modifications on the existing touchpoints $\gamma \in \Gamma$ consist of either increasing or reducing the original investment of each brand for each of its steps, multiplying its value by a given factor. The following reduction factors for the original values are considered: 15%, 30%, 45%, and 60%. In addition, the considered increase factors are 100%, 200%, 300%, and 400%. These factors were inspired by the investment data used during the design of the baseline model where paid touchpoints can receive different investment in different years. The election of an increase or reduction factor is made at random and stays unchanged for each step. In contrast, the modifications in the target values of awareness and WOM volume are applied adding or subtracting a given quantity to each of its time steps. This way, each modification on the target awareness values adds or subtracts 2%,

5%, 8%, or 10% to the target brand values. The resulting awareness values are truncated between 1% and 100% for avoiding unrealistic target values. With respect to target WOM volume, each modification is of either 1,000, 2,000, 4,000, or 6,000 conversations, always keeping the conversation values above 0. Like in the case of touchpoint investment, the decision of increasing or reducing the target values is made at random and maintained at each step.

All problem instances are labeled using its number of decision variables: P1(24), P2(39), P3(45), P4(54), P5(60), P6(69), P7(75), P8(84), P9(90), P10(99), P11(114), and P12(129). The parameter values of the baseline instance for each parameter not modified during calibration are the following. The main parameters are $z = 1000$, $|B| = 8$, $|\Gamma| = 7$, and $t_{\max} = 52$. The awareness of the agent population is initialized as $a^b(0) = (0.71, 0.75, 0.58, 0.25, 0.08, 0.42, 0.39, 0.34)$ and the touchpoint reach parameters are set to $r_\gamma = (0.92, 0.57, 0.54, 0.035, 0.43, 0.38, 0.69)$. Finally, the network generation parameter in the Barabasi-Albert algorithm is set to $m = 4$. A summary with these simulation parameters is shown in Table 2.

General market settings	Initial awareness	Touchpoints' reach
$z = 1000$	$a^1(0) = 0.71$	$r_1 = 0.92$
$\|B\| = 8$	$a^2(0) = 0.75$	$r_2 = 0.57$
$\|\Gamma\| = 7$	$a^3(0) = 0.58$	$r_3 = 0.54$
$t_{\max} = 52$	$a^4(0) = 0.25$	$r_4 = 0.035$
	$a^5(0) = 0.08$	$r_5 = 0.43$
	$a^6(0) = 0.42$	$r_6 = 0.38$
	$a^7(0) = 0.39$	$r_7 = 0.69$
	$a^8(0) = 0.34$	

Table 2: Summary of the simulation parameters that are not modified during the calibration process.

The generated instances share this initial setup along with its corresponding reach parameter value r_γ for the new paid touchpoints, which take the value of the original touchpoint employed in its generation. This way, if a new touchpoint 12 is generated using the original touchpoint 3, they share the reach parameter value (i.e., $r_{12} = r_3$).

5.2. Metaheuristics considered for ABM automatic calibration task

Two baseline methods and three evolutionary algorithms were selected for benchmarking the coral reefs-based metaheuristics. The first baseline method follows a random search approach (RND). This procedure simply creates valid solutions by randomly generating values for each of the decision variables. This way, random solutions are generated until the stopping criteria is met and the method returns the best solution found. The second baseline method is a local search procedure implementing a HC strategy, similar to the one employed by the memetic algorithms. However, instead of starting for a given individual of the coral population, the baseline HC starts from a random individual generated using a uniform distribution. The HC search continues until the stopping criteria is met, which involves a much higher number of steps than that considered for the memetic variants.

The first evolutionary algorithm considered in the study is differential evolution (DE) (Storn and Price, 1997). This is the metaheuristic considered in several automatic calibration approaches (LaTorre et al., 2019; Zhong and Cai, 2015; César Trejo Zúñiga et al., 2014) and has been shown to provide a better performance than the basic genetic algorithms used in other contributions

when calibrating ABMs (Herrmann and Savin, 2015; Zhong and Cai, 2015). DE generates new solutions by combining the existing individuals with a *donor vector* created following the next equation: $x_i(G+1) = x_{r1}(G) + F(x_{r2}(G) - x_{r3}(G))$, where x_i is the generated donor vector and $r1$, $r2$, and $r3$ are different solutions at generation G. For each generation, a donor vector x_i is generated for every individual i and its values are combined with those of the original individual by means of an uniform crossover according to a CR parameter. For every gene of the newly created individual, the algorithm takes the value of the donor vector with probability CR, otherwise it takes the original value of individual i. If the fitness value of the resulting individual is better or equal than that of individual i, it will replace i in the population. Additionally, the selected design considers the variant DE/best/1/bin (Storn and Price, 1997). In this variant, "best" refers to the fact that the best individual in the population is always selected as the individual $x_{r1}(G)$ in the previous operator to obtain the mutated offspring $x_i(G+1)$. Meanwhile, the values "1" and "bin" respectively stand for the use of a single vector difference in that operator, as shown in the former equation, and of the uniform crossover operator, as also described.

The second evolutionary algorithm considered in the study is success-history based adaptive differential evolution with linear population size reduction (L-SHADE) (Tanabe and Fukunaga, 2014). L-SHADE is an extension of SHADE, a history-based variant of differential evolution (Tanabe and Fukunaga, 2013). SHADE and other adaptive variants of DE self-adapt the values of the crossover rate (CR) and the mutation rate (F) during the optimization process, which are the main control parameters. In the case of SHADE, the *successful* values of CR and F are stored into a historical memory. A parameter value is *successful* if the solution generated using it improves the previous individual. In addition, L-SHADE extends SHADE by including a mechanism for reducing the population after each generation. This population reduction is computed linearly with respect of the number of fitness evaluations, which improves the convergence of the algorithm as the search process advances.

Finally, the third evolutionary algorithm is a restart CMA-ES with increasing population size (IPOP-CMA-ES) (Auger and Hansen, 2005). CMA-ES is an evolution strategy that relies on a covariance matrix for sampling new search points. At each iteration of the optimization process, λ new solutions are independently sampled according to a multi-variate normal distribution. Then, the best μ solutions (or search points, as referred by CMA-ES literature) are weighted and summed for obtaining a new mean value for the distribution. CMA-ES employs the fitness information of previous iterations (called the evolution path) in combination with the latest solutions sampled for updating the covariance matrix. IPOP-CMA-ES extends the typical CMA-ES design by including a restart strategy that increases the current population size by a given factor each time the algorithm restarts because a stopping criteria is met. These stopping criteria typically involve a stagnation of the search process (Auger and Hansen, 2005).

In addition, memetic variants for L-SHADE and IPOP-CMA-ES were also included, which are referred as ML-SHADE and MIPOP (for short). These memetic algorithms include the same local optimizer employed for MCRO and MCRO-SL and the local search refinement is applied with a probability p_{LS} to any solution of the population. To sum up, 11 metaheuristics are compared for the automatic ABM calibration problem: the coral reefs-based metaheuristics, two baseline methods, three evolutionary algorithms, and four memetic variants.

5.3. Experimental setup

The selected metaheuristics were implemented in Java using the ECJ framework (Luke, 1998). Since this framework does not include the coral reefs-based algorithms, the CRO implementation is

integrated by modifying some ECJ components. Readers can access the code for the algorithms and the experimentation in GitHub: `https://github.com/nachomsenias/abmcalibration`. Each metaheuristic runs 20 independent times using different random seeds. Every algorithm execution considers 10,000 evaluations as stopping criteria. During each evaluation of a candidate solution, the objective function f is computed using the output of each independent run with the provided historical data considering $\beta = 0.5$ (i.e. the same weight is considered for both KPIs). Due to the highly time-consuming task of simulating multiple times for every parameter configuration, the final fitness of the individual is calculated as the average value of f for 15 Monte-Carlo simulations of the ABM using the parameters encoded in the individual.

With respect to the coding scheme, since the parameters of this calibration problem are real values limited to $[0, 1]$, the integer-coded approach transforms the parameter values using $\min_i = 0$, $\max_i = 1$, and $\epsilon_i = 0.001$, for every parameter i, resulting in 1,001 possible values. The parameter setup for the metaheuristics was specified via a preliminary experimentation and the final values are as follows:

- HC movements increase or decrease the integer value of the genes by one unit. Each HC run takes 200 iterations and applies 50 movements during each iteration.

- DE considers a population size of 100 individuals. The parameter value for the crossover rate is set to CR= 0.9 and the value of the mutation rate is set to F= 0.5.

- L-SHADE uses an initial population size of 100 individuals and a external archive size of 200. During the reproduction phase, the p value for the current-to-pbest/1/bin strategy is set to $p = 0.1$. Finally, the size of the historical memory is set to the dimensionality of the problem.

- IPOP-CMA-ES considers a population size of $\lambda = 15$ with $\mu = 6$ and an increasing factor of 2. In addition, the learning rates are set to $c_\sigma = 0.568$, $c_c = 0.6962$, and $c_{\text{cov}} = 0.4897$. Since the calibration problem transforms the search space from the interval $[0, 1]$ into $[0, 1000]$, a relatively high $\sigma^{(0)}$ value of $\sigma^{(0)} = 56.747$ is selected. Finally, the dampening for the step size update is set to $d_\sigma = 4.2939$.

- CRO uses a reef size of 50 individuals. Regarding the BLX-α crossover, the probability is set to $p_c = 0.2$ using $\alpha = 0.25$ and tournament selection of size 3. The mutation probability of the random mutation is set to $p_m = 0.1$. The rest of CRO parameter values are $k = 3$, $\rho_0 = 0.6$, $F_a = 0.05$, $F_b = 0.9$, $F_d = 0.08$, and $P_d = 0.15$.

- CRO-SL defines a reef populated by 50 individuals. Its CRO-based values are set to $k = 3$, $\rho_0 = 0.6$, $F_a = 0.05$, $F_b = 0.9$, $F_d = 0.05$, and $P_d = 0.15$. The substrate layers of CRO-SL integrate uniform random mutation, random walk mutation, SBX, and BLX-α, and its parameter values are $p_m = 0.2$, $p_c = 1$, and $\alpha = 0.51$.

- Regarding the memetic variants, the refinement probability is set to $p_{\text{LS}} = 0.0625$ since it is the recommended value for similar problems (Lozano et al., 2004). Each time an individual is refined using the local search procedure, it will use 50 evaluations using the local search procedure. The rest of the parameters of the memetic algorithms are shared with its corresponding global search algorithm (i.e., MCRO has the same setup as CRO).

6. Analysis of results

In order to conduct a clear and comprehensive analysis of the different metaheuristics' performance, the analysis of the calibration results considers different stages. After the preliminary analysis, the improvement of including memetic approaches to the original metaheuristics is studied. Finally, the best performing methods are selected and studied at the end of this section.

Table 3 presents the performance of the original metaheuristics (i.e., without considering the memetic component). This information is complemented with a statistical test considering the ranking of the metaheuristics and with several post-hoc procedures for highlighting significant differences in their performance. Specifically, Friedman's nonparametric test (Friedman, 1940), Bonferroni-Dunn's test (Dunn, 1961), and Holm's test (Holm, 1979) are used. The average ranking and the resulting $p-$values of Bonferroni-Dunn's test and Holm's test are shown in Table 4. With respect to the Friedman's test, the result of applying the test is $\chi_F^2 = 64.35$ and its corresponding $p-$value is $5.8 \cdot 10^{-12}$. As the $p-$value is lower than the desired level of significance ($\alpha = 0.01$), the test concludes that there are significant differences in the algorithms' performance.

It can be observed in these results that CRO-SL outperforms the other algorithms since it ranks first achieving the lowest mean rank (1.5) and finds the configuration with lowest fitness value for almost every model instance. After CRO-SL, CRO and IPOP-CMA-ES rank second and third with close ranking values (2.25 and 2.42, respectively). L-SHADE and DE rank forth and fifth but perform significantly worse than the control method CRO-SL, since the $p-$values from both Bronferroni-Dunn's and Holm's tests (0.003 and 0.001 in the case of L-SHADE, 0.001 and 0.001 in the case of DE) are lower than the considered significance level ($\alpha = 0.05$). Finally, the baseline methods HC and RND also perform significantly worse than the control method and rank last for every model instance.

The performance of CRO-SL can be analyzed in terms of the behavior of the different search procedures when solving a single problem instance. Thus, the P12(129) instance is selected as it is the instance with the highest dimensionality. Figure 4 shows the number of settling larvae (i.e., new solutions in the population) obtained by each substrate at each generation and the percentage of times each substrate produces the best solution of the generation. These plots show that SBX and random walk are the best performing substrates since they end up generating 40% and 35% of the best quality solutions, respectively. Random walk also stands out as the substrate generating the higher number of solutions which replace other solutions in the reef due to their quality. It can also be observed that every substrate is productive during the first generations when there is space in the reef available for new solutions. A higher number of solutions during the first generations are produced by the mutation substrates, in a period when exploration of the search space is crucial. BLX-α decays after obtaining some of the best solutions during the first generations and later on the operator SBX stands out by delivering good quality solutions. Thus, both exploration and exploitation are maintained by the different operators and balanced until the convergence of the algorithm.

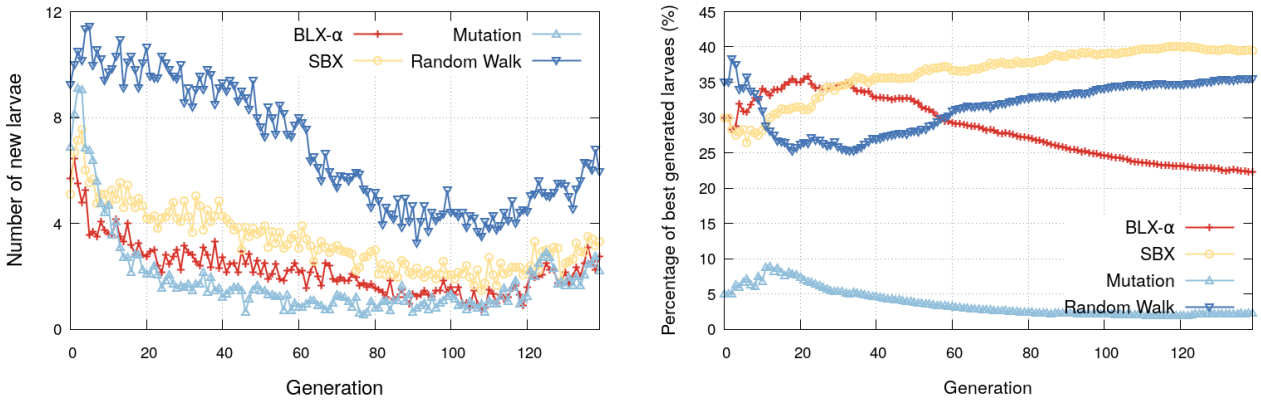
Table 5 shows the results of the memetic variants together with the results of the original metaheuristics so their values can be compared. At this stage of the analysis, the results of the baseline methods are not shown because they are significantly outperformed by the other algorithms. Similarly, the improvement of a DE-based memetic algorithm is not studied since it was the worst ranked evolutionary algorithm during the preliminary analysis. Besides, a more advanced DE variant, L-SHADE, is considered. Table 6 presents the corresponding average ranking and the resulting $p-$values of Bonferroni-Dunn's and Holm's tests using MCRO as the control method. In

		RND	HC	DE	LSHADE	IPOP	CRO	CRO-SL
P1(24)	Avg.	34.03	37.46	31.48	29.57	**25.79**	27.87	26.98
	Std. dev.	0.84	1.24	2.12	1.65	0.27	2.86	2
	Best	32.18	34.45	26.82	25.69	25.51	25.19	24.96
P2(39)	Avg.	53.26	51.9	42.87	42.74	43.02	40.76	**38.13**
	Std. dev.	1.63	5.84	0.57	1.02	0.55	1.19	0.24
	Best	49.8	45.05	42.11	40.86	42.35	38.36	37.85
P3(45)	Avg.	31.56	34.57	28.87	27.59	**23.99**	26.83	25.17
	Std. dev.	0.55	1.76	1.05	0.64	0.94	2.44	1.16
	Best	30.48	31.57	24.93	26.18	23.42	23.79	23.65
P4(54)	Avg.	42.93	46.04	37.59	37.48	**31.85**	31.95	32.55
	Std. dev.	0.66	1.46	2.17	1.51	0.42	1.87	3.34
	Best	41.55	42.5	33.68	34.41	31.41	30.86	29.19
P5(60)	Avg.	37.54	39.58	32.54	32.17	**26.66**	27.53	26.72
	Std. dev.	0.56	1.17	1.76	1.13	0.25	1.38	1.95
	Best	36.47	37.11	28.33	29.01	26.35	25.58	25.05
P6(69)	Avg.	55.39	59.58	47.54	43.52	42.41	40.82	**39.32**
	Std. dev.	1.01	1.48	0.82	0.88	0.93	1.51	0.92
	Best	53.47	57.1	45.35	42.32	41.22	39.37	38.34
P7(75)	Avg.	46.33	48.29	39.09	40.63	35.41	**35.09**	35.14
	Std. dev.	0.66	1.03	1.22	1.81	0.95	1.09	2.22
	Best	45.08	45.51	34.88	35.86	34.5	34.35	32.59
P8(84)	Avg.	55.16	58.1	37.63	43.41	32.82	32.59	**28.7**
	Std. dev.	0.94	1.89	0.22	2.5	1.61	2.1	1.87
	Best	53.02	53.79	37.02	38.76	31.66	29.36	25.91
P9(90)	Avg.	49.14	51.35	31.04	35.14	27.91	27.2	**26.54**
	Std. dev.	0.6	0.94	1.05	2.76	1.62	1.99	2.47
	Best	47.84	49.87	27.93	30.7	26.05	25.38	24.23
P10(99)	Avg.	72.37	80.35	54.72	52.79	48.05	46.02	**40.81**
	Std. dev.	1.64	3.9	0.93	1.06	1.75	4.05	1.18
	Best	69.54	71.84	51.33	50.28	46.09	41.36	39.01
P11(114)	Avg.	48.61	50.93	33.98	37.63	31.04	30.28	**26.47**
	Std. dev.	0.47	1.03	1.02	2.99	1.94	2.22	1.85
	Best	47.8	49	29.95	34.94	27.71	28.15	24.55
P12(129)	Avg.	41.96	43.81	34.07	34.88	28.24	30.94	**26.92**
	Std. dev.	0.56	0.79	1.08	1.25	0.65	1.38	0.57
	Best	40.55	42.51	31.96	32.55	27.46	27.68	25.87

Table 3: Fitness function values obtained by the metaheuristics for every ABM instance. The values are shown using average, standard deviation, and minimum result of the 20 different runs.

	Rank	Bonferroni p	Holm p
CRO-SL	1.5	–	–
CRO	2.25	1	0.52
IPOP-CMA-ES	2.42	1	0.52
L-SHADE	4.33	0.003	0.001
DE	4.5	0.001	0.001
RND	6.08	$1.4 \cdot 10^{-7}$	$1.2 \cdot 10^{-7}$
HC	6.92	$2.6 \cdot 10^{-10}$	$2.6 \cdot 10^{-10}$

Table 4: Average ranking of the metaheuristics and their resulting $p-$values for Bonferroni's and Holm's test using CRO-SL as the control method.



(a) Larvae generated by each substrate

(b) Percentage of best generated larvae (%)

Figure 4: (a) Number of new larvae settling in the reef per substrate at each generation. (b) Percentage of best larvae obtained by a substrate at each generation. These values are computed as the average of the multiple Monte-Carlo runs of CRO-SL for P12(129) instance.

this case, the Friedman's test results in $\chi^2_F = 62.94$ and a $p-$value of $3.8 \cdot 10^{-11}$. Once again the test concludes that there are significant differences between the algorithms' performance.

The results of the memetic variants reveal that not every hybridized algorithm improves the performance of its original counterpart since MCRO and MIPOP improve the original algorithms but ML-SHADE and MCRO-SL do not. MCRO ranks first and achieves the lowest average ranking (1.5), improving CRO in every instance and ranking first in eight instances. MIPOP obtains a small improvement compared to the original IPOP-CMA-ES performance for most instances but ranks third with an average ranking of 4, ranking first in P3.

CRO-SL is ranked second in average (2.5) and ranks first for the remaining three model instances. In addition, CRO-SL still finds the best solution for most model instances. In contrast, MCRO-SL ranks sixth (4.58) and is significantly outperformed by MCRO. MCRO-SL improves the original CRO-SL in the first instance but its performance decreases for the rest. In order to assess the performance loss of MCRO-SL, it was checked if the stopping criteria can restrict the memetic performance by rerunning the experimentation fixing the stopping criteria in 20,000 evaluations for CRO-SL and MCRO-SL. However, the results were similar to those originally delivered for 10,000 evaluations showing again that MCRO-SL cannot improve CRO-SL. Therefore, the performance reduction cannot be explained by the stopping criteria but for the role of the local search component. Since CRO-SL shows a good balance between exploration and exploitation (note that it obtains the most of the best solutions), the addition of the local search component can negatively modify this balance by reducing its diversity and causing the algorithm to converge prematurely.

With the inclusion of the memetic variants, CRO obtains the same rank as IPOP-CMA-ES and both algorithms are tied with an average rank of 4.25 in the forth/fifth position. However, these algorithms are significantly outperformed by MCRO since the $p-$values of both Bonferroni and Holm tests (0.024 and 0.013, respectively) are below the considered significance threshold. L-SHADE and its memetic counterpart ML-SHADE are the worst ranked algorithms. ML-SHADE does not obtain any performance increment from its hybridization with local search and ranks last for every instance.

		LSHADE	MLSHADE	IPOP	MIPOP	CRO	MCRO	CRO-SL	MCRO-SL
P1(24)	Avg.	29.57	32.53	25.79	25.98	27.87	**25.73**	26.98	26.75
	Std. dev.	1.65	1.65	0.27	0.68	2.86	1.34	2	2.19
	Best	25.69	28.37	25.51	25.53	25.19	25.07	24.96	25.13
P2(39)	Avg.	42.74	46.39	43.02	42.57	40.76	39.42	**38.13**	40.09
	Std. dev.	1.02	2.23	0.55	0.62	1.19	0.73	0.24	1.64
	Best	40.86	42.42	42.35	41.95	38.36	38.45	37.85	38.34
P3(45)	Avg.	27.59	30.33	23.99	**23.81**	26.83	26.1	25.17	27.23
	Std. dev.	0.64	1.04	0.94	0.39	2.44	3.21	1.16	1.04
	Best	26.18	27.05	23.42	23.47	23.79	22.42	23.65	24.91
P4(54)	Avg.	37.48	40.22	31.85	31.79	31.95	**30.62**	32.55	34.09
	Std. dev.	1.51	2.39	0.42	0.5	1.87	0.51	3.34	3.22
	Best	34.41	34.93	31.41	31.19	30.86	29.52	29.19	30.02
P5(60)	Avg.	32.17	34.73	26.66	27.09	27.53	**25.87**	26.72	27.54
	Std. dev.	1.13	2.23	0.25	1.08	1.38	0.74	1.95	2.33
	Best	29.01	29.4	26.35	25.97	25.58	25.06	25.05	25.24
P6(69)	Avg.	43.52	52.18	42.41	42.76	40.82	**39.2**	39.32	41.49
	Std. dev.	0.88	1.94	0.93	0.8	1.51	0.26	0.92	2.15
	Best	42.32	47.09	41.22	42.05	39.37	38.83	38.34	38.85
P7(75)	Avg.	40.63	44.82	35.41	35.71	35.09	**33.6**	35.14	36.24
	Std. dev.	1.81	2.34	0.95	0.59	1.09	0.47	2.22	2.45
	Best	35.86	39.54	34.5	35.02	34.35	32.82	32.59	33.29
P8(84)	Avg.	43.41	46.99	32.82	32.41	32.59	**27.61**	28.7	34.49
	Std. dev.	2.5	3.65	1.61	2.03	2.1	0.71	1.87	2.97
	Best	38.76	38.83	31.66	29.34	29.36	26.46	25.91	28.1
P9(90)	Avg.	35.14	42.01	27.91	27.69	27.2	**24.69**	26.54	27.03
	Std. dev.	2.76	3.56	1.62	0.76	1.99	0.38	2.47	2.72
	Best	30.7	32.9	26.05	26.81	25.38	24.24	24.23	24.3
P10(99)	Avg.	52.79	59.1	48.05	48.08	46.02	42.55	**40.81**	45.51
	Std. dev.	1.06	4.67	1.75	1.09	4.05	1.92	1.18	2.01
	Best	50.28	55.39	46.09	46.34	41.36	39.91	39.01	42.75
P11(114)	Avg.	37.63	39.42	31.04	30.59	30.28	**25.97**	26.47	30.08
	Std. dev.	2.99	5.31	1.94	1.71	2.22	0.89	1.85	3.55
	Best	34.94	34.51	27.71	28.53	28.15	24.83	24.55	25.45
P12(129)	Avg.	34.88	37.03	28.24	28.03	30.94	27.22	**26.92**	28.47
	Std. dev.	1.25	2.83	0.65	0.51	1.38	0.7	0.57	1.18
	Best	32.55	31.08	27.46	27.42	27.68	26.26	25.87	26.43

Table 5: Fitness function values obtained by the memetic algorithms for every model instance. The values are shown using average, standard deviation, and best result of the 20 different runs.

	Rank	Bonferroni p	Holm p
MCRO	1.5	–	–
CRO-SL	2.5	1	0.287
MIPOP	4	0.054	0.015
CRO	4.25	0.024	0.013
IPOP	4.25	0.024	0.013
MCRO-SL	4.58	0.007	0.005
L-SHADE	6.92	$5.9 \cdot 10^{-8}$	$5 \cdot 10^{-8}$
ML-SHADE	8	$3.3 \cdot 10^{-11}$	$3.3 \cdot 10^{-11}$

Table 6: Average ranking of the metaheuristics including the memetic algorithms. The resulting $p-$values corresponds to Bonferroni's and Holm's tests using MCRO as the control method.

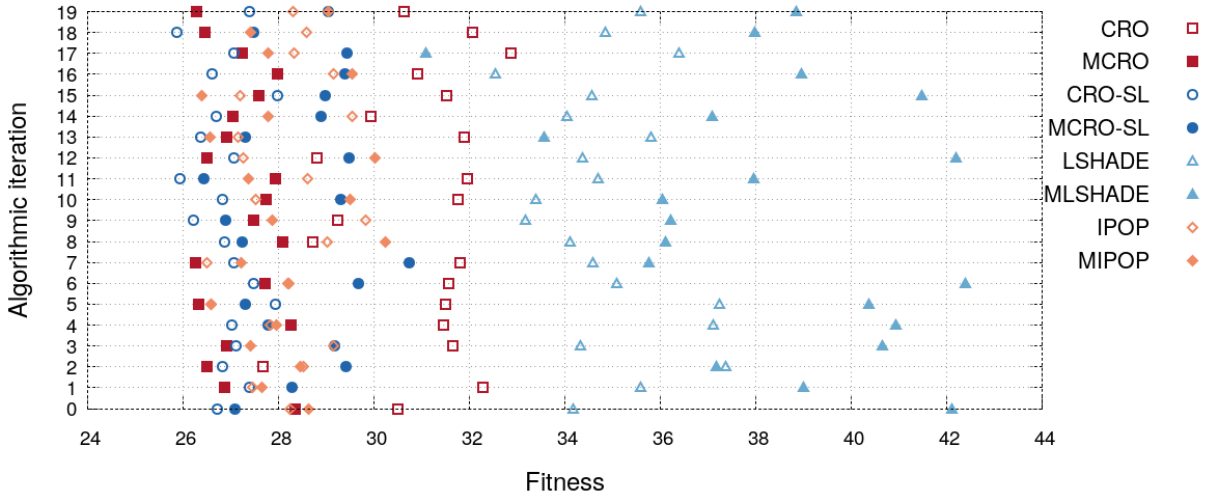


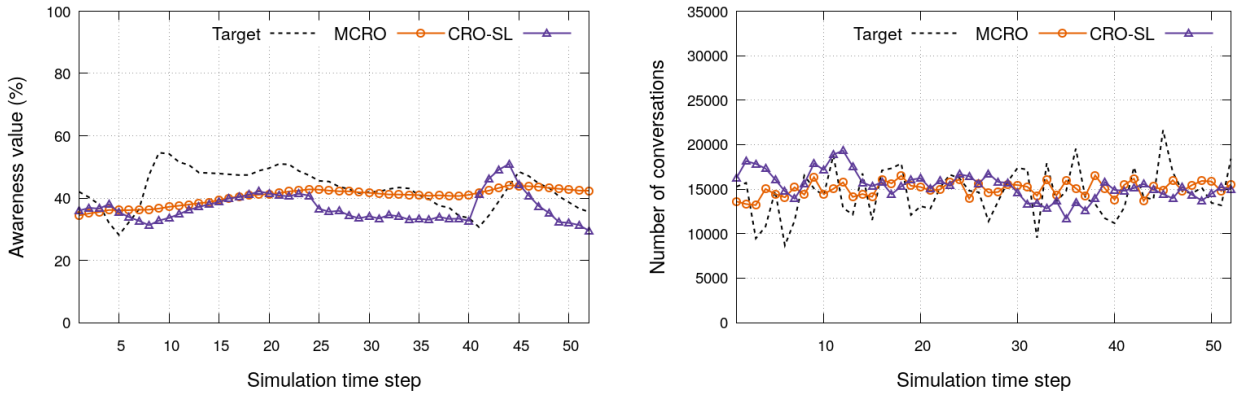
Figure 5: Fitness results of the different runs of the eight best performing algorithms for the problem instance P12(129). The values in the Y axis refer to the different runs and the values in the X axis shows the fitness error for the problem instance.

The impact of hybridization on the performance of these algorithms can be visually corroborated by displaying their fitness values for the instance with the highest dimensionality, P12(129). Figure 5 shows the multiple runs of the algorithms for this instance using points: the X axis refers to the fitness values and the Y axis marks the different iterations. This plot shows how MCRO obtains the highest improvement with respect to its non-memetic counterpart. MIPOP slightly improves IPOP in many of the runs but with a reduced margin and not in every run. In addition, the loss of performance of MCRO-SL and MLSHADE is also noticeable.

An additional statistical test is applied to the best ranking algorithms (MCRO, CRO-SL, and MIPOP) using pairwise comparisons. Hence, the analysis performs the Wilcoxon's ranksum test (null hypothesis $R_i = R_j$, alternate hypothesis $R_i < R_j$, where R_i and R_j represent the ranking of two different algorithms) (García et al., 2008). The test is applied to every pair of the best performing metaheuristics, considering a level of significance $\alpha = 0.05$. The resulting $p-$values are

MCRO	vs	CRO-SL	MIPOP
$p-$values	–	0.018	$4 \cdot 10^{-4}$
CRO-SL	vs	MCRO	MIPOP
$p-$values	–	0.984	0.016

Table 7: Resulting $p-$values corresponding to Wilcoxon ranksum test when comparing MCRO, CRO-SL, and MIPOP using pairwise comparisons.



(a) Awareness output for brand 8

(b) WOM volume output for brand 6

Figure 6: Overtime awareness fitting for brand 8 (a) and overtime number of conversations (WOM volume) fitting for brand 6 (b) for P12(129). These values are obtained using the average overtime values of the multiple Monte-Carlo simulations. The lines with circles represent the best solution obtained by MCRO and the lines with purple triangles represent the best solution obtained by CRO-SL. The dashed lines represent the target historical values.

shown in Table 7. The results of Wilcoxon's test conclude that MCRO performs significantly better than CRO-SL and MIPOP since the resulting $p-$values are lower than the defined significance level. In addition, CRO-SL outperforms MIPOP with statistical significance.

Finally, there are some limitations of the presented approach that need to be addressed. One of these limitations can be identified by visually analyzing the output fitting of the calibrated solutions. Figure 6 shows the fitting for both awareness and WOM volume of the best calibrated solutions of MCRO and CRO-SL (i.e., those with the lowest fitness) for P12(129). In order to improve clarity, the KPI values refer to only two brands (6 and 8) since their behavior is a good indicator of the remaining brands. The displayed values represent the average of the multiple Monte-Carlo runs. These values show that even when having solutions with good fitness values, the calibrated solutions are only able to capture the trend of the target values for both brands. This effect is a consequence of the objective fitting function f, which relies in the numeric deviation error computed by the loss measure but ignores the pattern characteristics of the historical values. In addition, these synthetic instances could be harder to calibrate than others based on real data as they were generated using randomness.

7. Final remarks and future work

In this paper, a new approach to automatic ABM calibration using coral reefs-based metaheuristics with an integer-coded scheme was introduced. These methods, referred as CRO and CRO-SL, exhibit competitive performance due to the emulation of the processes involving the life cycle of corals, which empower the resulting heuristic search algorithms with a strong trade-off between diversification and intensification. The performance of the coral reefs-based methods was assessed by developing an exhaustive comparison against relevant evolutionary algorithms and by checking the significance of the results with statistical tests. This analysis also considered several memetic algorithms resulting from the hybridization of the selected metaheuristics with a local search procedure. This study was carried out by applying the selected methods to the calibration of an ABM for marketing scenarios considering brand awareness and WOM volume.

The analysis concluded that both MCRO and CRO-SL performed significantly better than the other methods. Both metaheuristics consistently achieved competitive results across every model instance, including those with high dimensionality. L-SHADE and IPOP-CMA-ES achieve competitive results to the coral reefs-based metaheuristics for most model instances. However, they end up outranked in their average ranking, which may be related with the selected integer coding scheme that may have reduced their performance. With respect to the memetic algorithms, the action of the local search procedure has a different behavior depending on the original metaheuristic, since only half of the memetic algorithms improved their performance when compared with their corresponding non-memetic counterparts. MCRO-SL reduces its performance when the dimensionality of the instances increases, which suggests that the local search procedure negatively affects the good balance of CRO-SL between exploration and exploitation.

Future work will be focused on analyzing the problem of calibrating ABM from a multiobjective perspective since some ABM are calibrated considering multiple KPIs. This could also involve tackling the calibration of ABM whose search space have special properties (e.g., multimodal search space). In addition, future work can address the limitations of the current calibration approach (exposed at the end of Section 6) by including qualitative pattern features (Yücel and Barlas, 2011).

Acknowledgments

This work is supported by the Spanish Agencia Estatal de Investigación, the Andalusian Government, the University of Granada, and European Regional Development Funds (ERDF) under grants EXASOCO (PGC2018-101216-B-I00), SIMARK (P18-TP-4475), and AIMAR (A-TIC-284-UGR18). M. Chica is also supported through the Ramón y Cajal program (RYC-2016-19800). Authors thank the Centro de Servicios de Informática y Redes de Comunicaciones (CSIRC), University of Granada, for providing the computing resources.

References

Auger, A., Hansen, N., 2005. A restart CMA evolution strategy with increasing population size, in: 2005 IEEE Congress on Evolutionary Computation, pp. 1769–1776.

Back, T., Fogel, D.B., Michalewicz, Z., 1997. Handbook of evolutionary computation. IOP Publishing Ltd., Bristol (UK).

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Bermejo, E., Chica, M., Damas, S., Salcedo-Sanz, S., Cordón, O., 2018. Coral reef optimization with substrate layers for medical image registration. Swarm and Evolutionary Computation 42, 138–159.

Bhattacharya, M., 2013. Evolutionary approaches to expensive optimisation. International Journal of Advanced Research in Artificial Intelligence 2.

Biedrzycki, R., 2017. A version of IPOP-CMA-ES algorithm with midpoint for CEC 2017 single objective bound constrained problems, in: 2017 IEEE Congress on Evolutionary Computation (CEC), pp. 1489–1494.

Biswas, A., Biswas, B., 2017. Analyzing evolutionary optimization and community detection algorithms using regression line dominance. Information Sciences 396, 185–201.

Branke, J., Asafuddoula, M., Bhattacharjee, K.S., Ray, T., 2017. Efficient use of partially converged simulations in evolutionary optimization. IEEE Transactions on Evolutionary Computation 21, 52–64.

ten Broeke, G., van Voorn, G., Ligtenberg, A., 2016. Which sensitivity analysis method should I use for my agent-based model? Journal of Artificial Societies & Social Simulation 19, 5.

Calvez, B., Hutzler, G., 2006. Automatic tuning of agent-based models using genetic algorithms, in: Sichman, J.S., Antunes, L. (Eds.), Multi-Agent-Based Simulation VI, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 41–57.

Camacho-Gómez, C., Marsa-Maestre, I., Gimenez-Guzman, J.M., Salcedo-Sanz, S., 2019. A coral reefs optimization algorithm with substrate layer for robust Wi-Fi channel assignment. Soft Computing 23, 12621–12640.

Canessa, E., Chaigneau, S., 2015. Calibrating agent-based models using a genetic algorithm. Studies in Informatics and Control 24, 79–90.

Chica, M., Barranquero, J., Kajdanowicz, T., Cordón, O., Damas, S., 2017. Multimodal optimization: an effective framework for model calibration. Information Sciences 375, 79–97.

Chica, M., Juan, A., Bayliss, C., Cordon, O., Kelton, D., 2020. Why simheuristics? benefits, limitations, and best practices when combining metaheuristics with simulation. SORT 22, 1–35.

Chica, M., Rand, W., 2017. Building agent-based decision support systems for word-of-mouth programs: A freemium application. Journal of Marketing Research 54, 752–767.

Coates, G., Li, C., Ahilan, S., Wright, N., Alharbi, M., 2019. Agent-based modeling and simulation to assess flood preparedness and recovery of manufacturing small and medium-sized enterprises. Engineering Applications of Artificial Intelligence 78, 195–217.

César Trejo Zúñiga, E., López Cruz, I.L., García, A.R., 2014. Parameter estimation for crop growth model using evolutionary and bio-inspired algorithms. Applied Soft Computing 23, 474–482.

Dai, C., Yao, M., Xie, Z., Chen, C., Liu, J., 2009. Parameter optimization for growth model of greenhouse crop using genetic algorithms. Applied Soft Computing 9, 13–19.

Deb, K., Agrawal, R.B., 1994. Simulated Binary Crossover for Continuous Search Space. Technical Report. Departement of Mechanical Enginering, Indian Institute of Technology, Kanpur, India.

Del Ser, J., Osaba, E., Molina, D., Yang, X.S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P.N., Coello Coello, C.A., Herrera, F., 2019. Bio-inspired computation: Where we stand and what's next. Swarm and Evolutionary Computation 48, 220–250.

Dunn, O.J., 1961. Multiple comparisons among means. Journal of the American statistical association 56, 52–64.

Epstein, J.M., 2006. Generative social science: Studies in agent-based computational modeling. Princeton University Press.

Eshelman, L.J., Schaffer, J.D., 1993. Real-coded genetic algorithms and interval-schemata, in: WHITLEY, L.D. (Ed.), Foundations of Genetic Algorithms. Elsevier. volume 2 of *Foundations of Genetic Algorithms*, pp. 187–202.

Fabretti, A., 2013. On the problem of calibrating an agent based model for financial markets. Journal of Economic Interaction and Coordination 8, 277–293.

Farmer, J.D., Foley, D., 2009. The economy needs agent-based modelling. Nature 460, 685–686.

Farris, P.W., Bendle, N.T., Pfeifer, P.E., Reibstein, D.J., 2010. Marketing metrics: The definitive guide to measuring marketing performance. Wharton School Publishing. 2nd edition.

Floreano, D., Mattiussi, C., 2008. Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies. The MIT Press.

Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11, 86–92.

García, S., Molina, D., Lozano, M., Herrera, F., 2008. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. Journal of Heuristics 15, 617–644.

Garcia-Hernandez, L., Garcia-Hernandez, J., Salas-Morera, L., Carmona-Muñoz, C., Alghamdi, N., de Oliveira, J.V., Salcedo-Sanz, S., 2020a. Addressing unequal area facility layout problems with the coral reef optimization algorithm with substrate layers. Engineering Applications of Artificial Intelligence 93, 103697.

Garcia-Hernandez, L., Salas-Morera, L., Carmona-Muñoz, C., Garcia-Hernandez, J., Salcedo-Sanz, S., 2020b. A novel island model based on coral reefs optimization algorithm for solving the unequal area facility layout problem.

Engineering Applications of Artificial Intelligence 89, 103445.
González-Avella, J.C., Cosenza, M.G., Klemm, K., Eguíluz, V.M., San Miguel, M., 2007. Information feedback and mass media effects in cultural dynamics. Journal of Artificial Societies and Social Simulation 10, 9.
Hansen, N., 1997. Convergence properties of evolution strategies with the derandomized covariance matrix adaptation: the ($\mu/\mu_$l, λ)-CMA-ES, in: Proceedings of the Fifth European Congress on Intelligent Techniques and Soft Computing (EUFIT'97), Aachen, Germany, pp. 650–654.
Herrmann, J., Savin, I., 2015. Evolution of the electricity market in Germany: Identifying policy implications by an agent-based model. VfS Annual Conference 2015 (Muenster): Economic Development - Theory and Policy. Verein für Socialpolitik / German Economic Association.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70.
van der Hoog, S., 2018. Surrogate Modelling in (and of) Agent-Based Models: A Prospectus. Computational Economics (in press) .
Ishibuchi, H., Yoshida, T., Murata, T., 2003. Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling. IEEE Transactions on Evolutionary Computation 7, 204–223.
Janssen, M.A., Ostrom, E., 2006. Empirically based, agent-based models. Ecology and Society 11, 37.
Kim, K.O., Rilett, L.R., 2003. Simplex-based calibration of traffic microsimulation models with intelligent transportation systems data. Transportation Research Record 1855, 80–89.
Krasnogor, N., Smith, J., 2000. A memetic algorithm with self-adaptive local search: TSP as a case study, in: Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation, Morgan Kaufmann Publishers Inc.. pp. 987–994.
Lamperti, F., Roventini, A., Sani, A., 2018. Agent-based model calibration using machine learning surrogates. Journal of Economic Dynamics and Control 90, 366–389.
LaTorre, A., Kwong, M.T., García-Grajales, J.A., Shi, R., Jérusalem, A., Peña, J.M., 2019. Model calibration using a parallel differential evolution algorithm in computational neuroscience: simulation of stretch induced nerve deficit. Journal of Computational Science , 101053.
Libai, B., Muller, E., Peres, R., 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. Journal of Marketing Research 50, 161–176.
Lozano, M., Herrera, F., Krasnogor, N., Molina, D., 2004. Real-coded memetic algorithms with crossover hill-climbing. Evolutionary Computation 12, 273–302.
Luke, S., 1998. ECJ evolutionary computation library. Publicly available at http://cs.gmu.edu/∼eclab/projects/ecj/.
Macdonald, E.K., Sharp, B.M., 2000. Brand awareness effects on consumer decision making for a common, repeat purchase product:: A replication. Journal of Business Research 48, 5–15.
Malleson, N., See, L., Evans, A., Heppenstall, A., 2014. Optimising an agent-based model to explore the behaviour of simulated burglars, in: Theories and Simulations of Complex Social Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 179–204.
Marín, J., 2012. How landscape ruggedness influences the performance of real-coded algorithms: a comparative study. Soft Computing 16, 683–698.
Molina, D., Moreno-García, F., Herrera, F., 2017. Analysis among winners of different ieee cec competitions on real-parameters optimization: Is there always improvement?, in: 2017 IEEE Congress on Evolutionary Computation (CEC), pp. 805–812.
Moscato, P., 1989. On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Technical Report 826. Caltech Concurrent Computation Program. Pasadena, USA.
Moscato, P., Cotta, C., Mendes, A., 2004. Memetic algorithms, in: New optimization techniques in engineering. Springer, pp. 53–85.
Moya, I., Chica, M., Cordón, Ó., 2019. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization. Decision Support Systems 124, 113111.
Moya, I., Chica, M., Sáez-Lozano, J.L., Cordón, Ó., 2017. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections. Knowledge-Based Systems 123, 200–216.
Muraro, D., Dilão, R., 2013. A parallel multi-objective optimization algorithm for the calibration of mathematical models. Swarm and Evolutionary Computation 8, 13–25.
Newman, M., Barabási, A.L., Watts, D.J., 2006. The structure and dynamics of networks. Princeton University Press.
Ngoduy, D., Maher, M., 2012. Calibration of second order traffic models using continuous cross entropy method. Transportation Research Part C: Emerging Technologies 24, 102–121.
North, M.J., Macal, C.M., Aubin, J.S., Thimmapuram, P., Bragen, M., Hahn, J., Karr, J., Brigham, N., Lacy, M.E., Hampton, D., 2010. Multiscale agent-based consumer market modeling. Complexity 15, 37–47.

Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research 151, 552–568.

Russell, S., Norvig, P., Intelligence, A., 1995. A modern approach. Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs 25, 27.

Salcedo-Sanz, S., 2016. Modern meta-heuristics based on nonlinear physics processes: A review of models and design procedures. Physics Reports 655, 1–70.

Salcedo-Sanz, S., 2017. A review on the coral reefs optimization algorithm: new development lines and current applications. Progress in Artificial Intelligence 6, 1–15.

Salcedo-Sanz, S., Camacho-Gómez, C., Mallol-Poyato, R., Jiménez-Fernández, S., Del Ser, J., 2016a. A novel Coral Reefs Optimization algorithm with substrate layers for optimal battery scheduling optimization in micro-grids. Soft Computing 20, 4287–4300.

Salcedo-Sanz, S., Camacho-Gomez, C., Molina, D., Herrera, F., 2016b. A coral reefs optimization algorithm with substrate layers and local search for large scale global optimization. 2016 IEEE Congress on Evolutionary Computation (CEC) , 3574–3581.

Salcedo-Sanz, S., Camacho-Gómez, C., Magdaleno, A., Pereira, E., Lorenzana, A., 2017a. Structures vibration control via Tuned Mass Dampers using a co-evolution Coral Reefs Optimization algorithm. Journal of Sound and Vibration 393, 62–75.

Salcedo-Sanz, S., Del Ser, J., Landa-Torres, I., Gil-López, S., Portilla-Figueras, J.A., 2014. The Coral Reefs Optimization algorithm: A novel metaheuristic for efficiently solving optimization problems. The Scientific World Journal 2014, 1–15.

Salcedo-Sanz, S., García-Herrera, R., Camacho-Gómez, C., Alexandre, E., Carro-Calvo, L., Jaume-Santero, F., 2019. Near-optimal selection of representative measuring points for robust temperature field reconstruction with the cro-sl and analogue methods. Global and Planetary Change 178, 15–34.

Salcedo-Sanz, S., Muñoz-Bulnes, J., Vermeij, M.J., 2017b. New coral reefs-based approaches for the model type selection problem: a novel method to predict a nation's future energy demand. International Journal of Bio-inspired Computation 10, 145–158.

Sargent, R.G., 2005. Verification and validation of simulation models, in: Proceedings of the 37th conference on Winter simulation, pp. 130–143.

Sörensen, K., 2015. Metaheuristics—the metaphor exposed. International Transactions in Operational Research 22, 3–18.

Stonedahl, F., Rand, W., 2014. When does simulated data match real data? Comparing model calibration functions using genetic algorithms, in: Advances in Computational Social Science. Springer, Japan. volume 11 of *Agent-Based Social Systems*, pp. 297–313.

Storn, R., Price, K., 1997. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11, 341–359.

Talbi, E.G., 2009. Metaheuristics: from design to implementation. John Wiley & Sons.

Tanabe, R., Fukunaga, A., 2013. Success-history based parameter adaptation for differential evolution, in: 2013 IEEE Congress on Evolutionary Computation, pp. 71–78.

Tanabe, R., Fukunaga, A.S., 2014. Improving the search performance of shade using linear population size reduction, in: 2014 IEEE Congress on Evolutionary Computation (CEC), pp. 1658–1665.

Thiele, J.C., Kurth, W., Grimm, V., 2014. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using Netlogo and 'R'. Journal of Artificial Societies and Social Simulation 17, 11.

Vermeij, M., 2005. Substrate composition and adult distribution determine recruitment patterns in a caribbean brooding coral. Marine Ecology Progress Series 295, 123–133.

Waldrop, M.M., 2018. Free agents. Science 360, 144–147.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Wilensky, U., Rand, W., 2015. Introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo. MIT Press.

Wu, F., Huberman, B.A., 2007. Novelty and collective attention. Proceedings of the National Academy of Sciences 104, 17599–17601.

Wu, G., Mallipeddi, R., Suganthan, P.N., 2019. Ensemble strategies for population-based optimization algorithms – A survey. Swarm and Evolutionary Computation 44, 695–711.

Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 599–608.

Yücel, G., Barlas, Y., 2011. Automated parameter specification in dynamic feedback models based on behavior pattern features. System Dynamics Review 27, 195–215.

Zhong, J., Cai, W., 2015. Differential evolution with sensitivity analysis and the Powell's method for crowd model calibration. Journal of Computational Science 9, 26–32. Computational Science at the Gates of Nature.

4 Evolutionary multiobjective optimization for automatic agent-based model calibration: A comparative study

- I. Moya, M. Chica, O. Cordón. Evolutionary multiobjective optimization for automatic agent-based model calibration: A comparative study, IEEE Access, vol. 9, pp. 55284-55299, 2021. DOI: 10.1109/ACCESS.2021.3070071.

  - State: Published.
  - Impact Factor (JCR 2020): 3.367.
  - Category: ENGINEERING, ELECTRICAL & ELECTRONIC - SCIE. Order: 94/273. Q2.

Evolutionary multiobjective optimization for automatic agent-based model calibration: A comparative study

Ignacio Moya[*,a], Manuel Chica[a,b], Óscar Cordón[a]

[a] *Andalusian Research Institute DaSCI "Data Science and Computational Intelligence", University of Granada, 18071 Granada, Spain*

[b] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia*

Abstract

Complex problems can be analyzed by using model simulation but its use is not straight-forward since modelers must carefully calibrate and validate their models before using them. This is specially relevant for models considering multiple outputs as its calibration requires handling different criteria jointly. This can be achieved using automated calibration and evolutionary multiobjective optimization methods which are the state of the art in multiobjective optimization as they can find a set of representative Pareto solutions under these restrictions and in a single run. However, selecting the best algorithm for performing automated calibration can be overwhelming. Therefore, we propose to deal with this issue by conducting an exhaustive analysis of the performance of several evolutionary multiobjective optimization algorithms when calibrating several instances of an agent-based model for marketing with multiple outputs. We analyze the calibration results using multiobjective performance indicators and attainment surfaces, including a statistical test for studying the significance of the indicator values, and benchmarking their performance with respect to a classical mathematical method. The results of our experimentation reflect that those algorithms based on decomposition perform significantly better than the remaining methods in most instances. Besides, we also identify how different properties of the problem instances (i.e., the shape of the feasible region, the shape of the Pareto front, and the increased dimensionality) erode the behavior of the algorithms to different degrees.

Keywords— Model calibration, agent-based modeling, evolutionary multiobjective optimization.

1. Introduction

Model simulation is a common approach to the analysis of complex phenomena. It allows users and stakeholders to recreate the desired dynamics so these phenomena could be studied in a controlled environment, where different policies or strategies can be tested. The agent-based model (ABM) methodology [6, 15, 34] is a well-known model simulation technique that relies in the behaviour of artificial agents, which are autonomous entities that act following simple rules and interacting with other agents. The aggregation of the agent's behavior and the social interactions

[*]Corresponding author

Email addresses: imoya@ugr.es (Ignacio Moya), manuelchica@ugr.es (Manuel Chica), ocordon@decsai.ugr.es (Óscar Cordón)

allow the modeler to simulate complex emergent dynamics using a bottom-up approach [15]. This approach has proven useful in both the forecasting of hypothetical scenarios and the definition of what-if scenarios [27], which has increased the visibility of ABM in the latter years [18, 55].

However, the use of these models is not straight-forward as they require the modelers to deal with several issues. On one hand, ABMs must be designed following certain guidelines and methodologies for ensuring their rigor [48]. On the other hand, they require to be properly calibrated and validated before being used [9, 52]. The calibration of a model refers to the process of adjusting its parameters so it can correctly simulate the desired dynamics. This stage in model development can be carried out manually, since many parameters are usually set based on data, but it can be impracticable for models with high dimensionality.

This limitation can be overcome using automated calibration [46], a procedure that relies on two main components for adjusting the model's parameters: a given error measure and an optimization method. With this approach, the error measure is minimized using the optimization method by adjusting the model's parameters so the simulated output can match the provided real data [8, 53]. Nevertheless, the success of an automated calibration process depends on the capability of its optimization method for exploring the model's parameters search space. This is specially relevant if the model being calibrated considers multiple conflicting criteria since the optimization method requires to handle these criteria jointly [10].

Multiobjective optimization methods are specially tailored for working under these restrictions [10]. In particular, evolutionary multiobjective optimization (EMO) algorithms can be considered the best approach to multiobjective optimization, as they obtain a set of representative Pareto solutions in a single run. In addition, EMO algorithms obtain Pareto-optimal solutions in a reasonable time and can perform successfully without requiring specific properties of the optimized function [12] and have proven to be successful when dealing with dynamic multiobjective optimization problems which are common in real-world applications [50, 21, 22, 47, 5]. However, there is a large number of EMO algorithms available in the specialized literature and finding the best algorithm for conducting the automated calibration process can be overwhelming.

We propose to delve into this issue by conducting an exhaustive analysis of the performance of the most prominent and recent EMO algorithms when calibrating multiple instances of an ABM jointly optimizing different key performance indicators. The study considers well-known EMO algorithms from the main EMO categories: based on Pareto dominance, indicators, and decomposition. The selected methods include the non-dominated sorting genetic algorithm II (NSGA-II) [13], the improved strength Pareto evolutionary algorithm (SPEA2) [66], the general indicator-based evolutionary algorithm (IBEA) [65], the S metric selection multiobjective optimization algorithm (SMS-EMOA) [4], and the multiobjective evolutionary algorithm based on decomposition (MOEA/D) [30, 61]. Additionally, we also make use of two recent EMO developments that have shown competitive results when dealing with real problems [16, 38, 60]. Specifically, we incorporate to the study the many-objective metaheuristic based on the R2 indicator II (MOMBI2) [24] and the global weighting achievement scalarizing function genetic algorithm (GWASF-GA) [51]. Finally, we included a classical optimization method such as the Nelder-Mead's simplex method [44] as the baseline for the EMO algorithms, that we adapted to our multiobjective problem by using the adaptive $\epsilon-$constraint method, which is a common approach [14, 37]. We analyze the results of the selected algorithms using multiobjective performance indicators and attainment surfaces [68]. In addition, we perform a statistical test for studying the significance of the performance indicator values.

Our study considers a benchmark of 15 instances of an ABM for marketing, which is the selected computational model for our experiments. This ABM tackles marketing scenarios involving two conflicting outputs or key performance indicators: the global awareness of the consumers regarding the brands available in the market and the number of word-of-mouth consumer interactions for those brands. Both the instances and historical data for our calibration benchmark are taken from a real banking marketing scenario in Spain. Although there is previous work using EMO for multicriteria calibration of ABMs [17, 42, 49], none of these contributions considers a rigorous and exhaustive comparison of several EMO algorithms for calibrating multiple model instances and the subsequent analysis of the algorithms' performance according to the problem characteristics.

Hence, the main contributions of the current manuscript are:

- An exhaustive analysis of the performance of relevant EMO algorithms when calibrating multiple instances of an ABM for marketing considering different outputs.
- The design of an appropriate experimental setup for the study, which is based on a benchmark comprising 15 instances considering two key performance indicators and up to 175 decision variables.
- A comprehensive analysis of the results which considers both unary and binary performance indicators, attainment surfaces, statistical significance/tests, and two discussion sections addressing both the influence of the instances' properties on the algorithms' performance and the drawbacks of the methods.

The structure of this paper is as follows. Section 2 addresses the related work on the use of EMO algorithms for multicriteria calibration of computational models. Then, our approach to the multicriteria calibration of ABMs using EMO algorithms is depicted in Section 3, which reviews several concepts of EMO-based model calibration and the algorithms selected for our study. The ABM to be calibrated in our experiments is described in Section 4. The analysis of the results is thoroughly reviewed in Section 5. Finally, Section 6 discusses our final remarks and Section 7 reviews the practical implications and future directions of our work.

2. Related work

There are some examples of the use of EMO algorithms for multicriteria calibration of computational models [1, 23, 28, 32, 33, 41, 64]. Many of them are focused in the calibration of hydrological models, such as the soil and water assessment tool [3, 11, 62, 63], the rainfallrunoff models [32], empirical hydrological models for streamflow forecasting [23], and an integrated water system model [64]. The thorough review of these contributions reveals that their usual approach relies on employing the NSGA-II for running the calibration process, probably because it is the most popular EMO algorithm. Apart from NSGA-II, we can also find some studies using SPEA2 [23, 28, 62, 63].

The application of EMO for multicriteria calibration of ABMs is not frequent although there are few examples tackling this issue [17, 42, 49]. Farhadi et al. [17] present a framework for sustainable groundwater management including a Nash bargaining model, which is implemented as an ABM incorporating cooperative and non-cooperative agents that consume the water of the modeled scenario at different ratios. The parameters of the model are calibrated using NSGA-II and considering three objectives and a single calibration scenario. Narzisi et al. [42] deal with the calibration of an ABM for emergency response planning using NSGA-II. Their model is calibrated

for minimizing the percentage of fatalities and the average waiting time of the population before receiving attention at the hospitals. This calibration process is applied to a single scenario considering ten real-coded model parameters with several restrictions. Finally, Read et al. [49] introduce the calibration of artificial murine multiple sclerosis simulation, an established immunological ABM for computational biology. They use NSGA-II in the calibration of 16 integer and real parameters with respect to four objectives. The authors consider a single scenario for the model and run three independent calibration executions for its analysis.

Therefore, NSGA-II is the recurrent EMO algorithm for multicriteria calibration of ABMs and there is not any comparative study on different EMO algorithms for this problem. In addition, it can also be recognized that the methodology of these contributions generally limit their experimentation to a single run of the EMO algorithm, which can be explained by the high computational cost of simulating multiple times for every evaluation of a single model configuration. However, this approach is not taking into account that EMO algorithms are stochastic, thus requiring multiple runs using different seeds. Analyzing the results of a single algorithmic execution reduces the amount of information provided by the calibration process because valuable model configurations may be skipped in the initial run, specially if the EMO algorithm is not properly tuned.

3. Multicriteria calibration of ABMs using EMO algorithms

This section describes the key features of using EMO algorithms for multicriteria calibration of ABMs. First, Section 3.1 reviews some basics on multiobjective optimization and Section 3.2 presents the common design of the EMO algorithms. Then, Section 3.3 includes several subsections for introducing the selected EMO algorithms according to their category, depending on their operation mode.

3.1. Multiobjective optimization

The multicriteria calibration of ABMs can be approached as a multiobjective optimization problem since there is usually a need for calibrating the model according to different outputs. In these kinds of problems, the quality of a setting is evaluated regarding multiple conflicting criteria instead of considering a single error measure. Thus, the optimization algorithm aims to minimize $F(x) = f_1(x), ..., f_m(x)$, where m represents the number of objectives and x is the set of decision variables for the optimization problem (i.e., decision space). Each function f_i computes the quality of the parameter setting to a calibration objective using a deviation error measure ϵ_i. In a calibration problem, each objective is associated to one simulated output o_i, resulting in $f_i(x) = \epsilon_i(o_i(x), \widetilde{o}_i)$, with $\widetilde{o}_i$ being the historical target values for the i-th output. Any of the well-known deviation measures such as MAPE, RMSE, or MARE [25] can be chosen for computing this deviation error.

Multicriteria model configurations thus need to be analyzed using multiobjective semantics like the Pareto *dominance* concept [10]. Given two feasible configurations u and v from the decision space with $u \neq v$, u *dominates* v if $u_i \leq v_i, \forall i : 1 \leq i \leq m$ and $\exists j : 1 \leq j \leq m : u_j < v_j$, i.e., if u is equal or better than v for every objective and strictly better for at least one objective. However, these inequalities should be reversed for any objective that is being maximized (to *dominate* means to be better). Using the dominance concept, the global Pareto-optimal configurations are those vectors u such that there is no feasible vector v that dominates u. A set of u configurations where there is no v that dominates any of the other solutions is called a Pareto-optimal set. In addition,

the representation of the solutions in the Pareto set as points from the objective space is called a Pareto-optimal front [10].

3.2. Common design

Before describing the selected EMO algorithms, we introduce their common characteristics. Each candidate solution has n decision variables corresponding to the model parameters being calibrated, which can either be integer-coded or real-coded values. The considered algorithms include polynomial mutation [12] as their mutation strategy. It modifies the values of a solution's variables with a probability $p_m \in [0, 1]$ using a polynomial distribution. This mutation strategy uses a distribution index parameter that regulates the strength of the mutation. Unless stated otherwise, the proposed algorithms use simulated binary crossover (SBX) [12] with a crossover probability $p_c \in [0, 1]$ as their crossover strategy. SBX emulates the operation of a single-point crossover from binary-encoding when performing crossover into real-coding decision variables. SBX operates as follows: given two parents $P_1 = (p_{11}, ..., p_{1n})$ and $P_2 = (p_{21}, ..., p_{2n})$, SBX generates two springs $C_1 = (c_{11}, ..., c_{1n})$ and $C_2 = (c_{21}, ..., c_{2n})$ as $c_{1i} = \bar{X} - \bar{\beta}/2 \cdot (p_{2i} - p_{1i})$ and $c_{2i} = \bar{X} + \bar{\beta}/2 \cdot (p_{2i} - p_{1i})$, where $\bar{X} = 1/2 \cdot (p_{1i} + p_{2i})$. $\bar{\beta}$ is a random value fetched from a random distribution initialized by setting a distribution index that acts as the spread factor of the operation.

3.3. Considered EMO algorithms

3.3.1. Algorithms based on Pareto dominance

Pareto dominance-based algorithms assign the quality of the solutions (thus guiding the selection mechanism) according to their dominance of other solutions in the population. The selected Pareto dominance-based algorithms are NSGA-II and SPEA2.

- NSGA-II [13] can be identified as the most popular and well-known EMO algorithm. NSGA-II's approach relies on non-dominated sorting, which allows it to combine elitism with good levels of diversity in a single population while being computationally fast, specially for problems with two or three objectives. NSGA-II produces an offspring set Q_t at each generation using the solutions of the previous set P_t. Then, both sets are merged into the temporary set R_t where previous and newly generated solutions are ranked according to its non-dominance level. The non-dominance level of a solution corresponds with the number of solutions that dominate it. The next set P_{t+1} is generated by selecting the solutions with the best ranking, which are the solutions not dominated by other solutions in the previous set. This process is iterated for the next ranks until a population size $|P|$ is reached. This strategy guides the algorithm to non-dominated regions while a set of non-dominated solutions are maintained in the population. The first solution set that does not fit P_{t+1} is filtered using a crowding mechanism for boosting the diversity of the new population.

- SPEA2 [66] is a well-known EMO algorithm that computes the fitness of its individuals calculating a "strength" value that represents how many solutions it dominates. Then, the fitness value for each solution is computed by summing the "strength" values of the solutions that dominate it. SPEA2 considers a separate population, named the "archive" ($\overline{P}_t$), designed to store non-dominated solutions. At each step, non-dominated solutions in P_t and $\overline{P}_t$ are copied to $\overline{P}_{t+1}$. If $\overline{P}_{t+1}$ exceeds the size of P, then its solutions are filtered using a truncation operator inspired in the *k-th* nearest neighbor method that selects the solutions with the minimum distance. If there are not enough solutions for filling $\overline{P}_{t+1}$ then the dominated

solutions with the minimum fitness are included until $|\overline{P}_{t+1}| = P$. Then a mating pool is set using binary tournament on $\overline{P}_{t+1}$. Finally P_{t+1} is the result of applying the crossover and mutation operators to the mating pool.

3.3.2. Algorithms based on indicators

Indicator-based algorithms assign the fitness of the solutions using indicator values. The selected indicator-based EMO algorithms are IBEA, SMS-EMOA, and MOMBI2.

- IBEA [65] is a classic EMO algorithm that qualifies solutions regarding their relative contribution to a given performance indicator with respect to the rest of solutions of the population. Therefore, IBEA computes the loss of quality of removing a solution from the population using dominance preserving binary indicators. In order to carry out this task, some suitable indicators would be the additive I_ϵ or the I_{HD} indicator, that is based on the concept of hypervolume [67]. Using these concepts, IBEA's fitness evaluation for solution x using a binary indicator I and a scaling parameter κ is computed as $F(x) = \sum_{y \in P \setminus x} -exp[-I(y,x)/\kappa]$. Finally, IBEA performs elitism and only the worst solutions of the population are removed, althought this implies that the fitness of the remaining solutions need to be updated each time a solution is removed from the population.

- SMS-EMOA [4] introduces the maximization of the dominated hypervolume into the search process for approximating the true Pareto front. SMS-EMOA borrows NSGA-II's non-dominated sorting mechanism for merging the current population P_t with the offspring population Q_t into P_{t+1}. However, SMS-EMOA considers a replacement strategy that targets the solutions from the worst front with the lesser contribution to the hypervolume of their respective front. This process maximizes the quality of the population regarding their hypervolume [67]. In addition, as the repeated calculation of hypervolume values is computationally expensive, SMS-EMOA follows a steady-state scheme for easing the replacement mechanism and allowing an easy parallelization of the fitness evaluation. Unlike other EMO algorithms like SPEA2, SMS-EMOA does not consider a separate archive for storing non-dominated solutions. Instead, it maintains a population of constant size that includes dominated and non-dominated solutions (as NSGA-II does). SMS-EMOA also preserves the extreme solutions (i.e., the ones with best fitness for one objective and worst fitness for the other) into the population for biobjective problems such as our ABM calibration problem instead of requiring a reference point for computing hypervolume. For problems with more objectives, a reference point is calculated dynamically at each generation.

- MOMBI2 [24] relies in the R2 quality indicator for ranking the solutions, a Pareto compliant indicator with a reduced computational cost. This quality indicator uses a utility function for mapping each objective into a single value. A common MOMBI2 configuration employs the achievement scalarizing function (also used by GWASF-GA) since it allows the algorithm to obtain weekly Pareto optimal solutions, although there are several candidate utility functions for the algorithm. In addition, instead of updating the nadir point at each generation, MOMBI2 updates this reference point taking into account its historic values during previous generations. This update takes two parameters α and ϵ as the threshold and the tolerance threshold, respectively. These historic values are used for estimating how far current solutions are from the true Pareto front: high variance suggests that the solutions are far from it and a low variance suggests that the solutions are close. The solution ranking using R2 proceeds as

follows: first, the solutions with the best rank (i.e., those that optimize the weight vectors) are selected, removed from P_t and introduced into P_{t+1}; then, this process goes on ranking solutions until every solution has been ranked and $|P|$ solutions are selected. In case two solutions provide the same utility value, the solution with the lowest Euclidean distance is selected.

3.3.3. Algorithms based on decomposition

Decomposition-based algorithms transform a given multiobjective problem into several subproblems. The selected decomposition-based EMO algorithms are MOEA/D and GWASF-GA.

- MOEA/D [30, 61] is an evolutionary algorithm that has received great attention in the evolutionary computation literature in the last few years. It employs decomposition techniques for reducing the multiobjective problem into as many subproblems as individuals ($|P|$). Then, MOEA/D solves every subproblem jointly by evolving its solution population (P_t), which contains the best solution found for each subproblem. The optimization of each subproblem is performed by only using information from its neighboring subproblems. Although MOEA/D is compatible with any decomposition approach that transforms the Pareto approximation problem into several scalar optimization problems, we choose the Tchebycheff approach in this paper, as recommended by the authors [30, 61]. In addition, MOEA/D uses an external population for storing the non dominated solutions found during the execution of the algorithm, similarly to SPEA2. Finally, we select a differential evolution operator as the crossover strategy instead of the SBX employed in the other algorithms, also following authors' recommendation [30]. This operator generates each offspring $C = (c_1, ..., c_n)$ as $c_i = P_1(i) + F \cdot (P_2(i) - P_3(i))$ with probability CR and $c_i = P_1(i)$ with probability $1 - CR$, where P_1, P_2, and P_3 are the donor individuals and CR and F are the control parameters.

- GWASF-GA [51] is a recent aggregation-based evolutionary algorithm. GWASF-GA approximates the true Pareto front transforming the original problem into a set of scalar subproblems that are minimized using the achievement scalarizing function, based on the Tchebychev distance. This scalarizing function uses two reference points: the nadir point and the utopian point. The former is a point containing the worst objective values of the solutions of the entire Pareto-optimal set. The latter is a point that is chosen for dominating the ideal point and that will not be obtainable for any solution. During each algorithm iteration, every solution in the population is classified into different fronts by computing their achievement scalarizing function values using the two mentioned reference points and a set of weight vectors. Each of these fronts contains the solutions with the lowest scalarizing function value for the weight vectors in the set. The set of weight vectors is predefined for ensuring that its inverse is well distributed, ensuring that the algorithm maintains diversity. Then, the fronts with the lowest function values are introduced into the next population until $|P|$ solutions are selected.

4. ABM description

The current section summarizes the main characteristics of the selected ABM [39]. Section 4.1 presents the general structure of the model and the mechanics of the agents. Then, Section 4.2 summarizes the parameters selected for calibration.

4.1. ABM general structure

The model performs a terminating simulation during a given number of steps T of a market with a set of B competing brands, where the time-step represents a natural week. The model simulates a set of agents I and their behavior when exposed to social interactions and the advertisement of C mass media channels. It considers two outputs relevant to market expansion [31, 35]: the word-of-mouth interactions between consumers (referred as WOM volume) and the awareness of the brands. During the simulation, the I consumer agents are exposed to the information spread by mass media channels and the WOM process generated through their social network. These two processes are connected because the activation of brand awareness using advertising also increases WOM volume due to the buzz effect produced by the campaign. Additionally, WOM interactions spread the agents' brand awareness through the social network. Therefore, we can see that both outputs cannot be adjusted separately, since improving the fitting of one output decreases the fitting of the other. As a consequence, there is no configuration that jointly satisfies the fitting of both outputs. We present a general scheme of our marketing model in Figure 1.

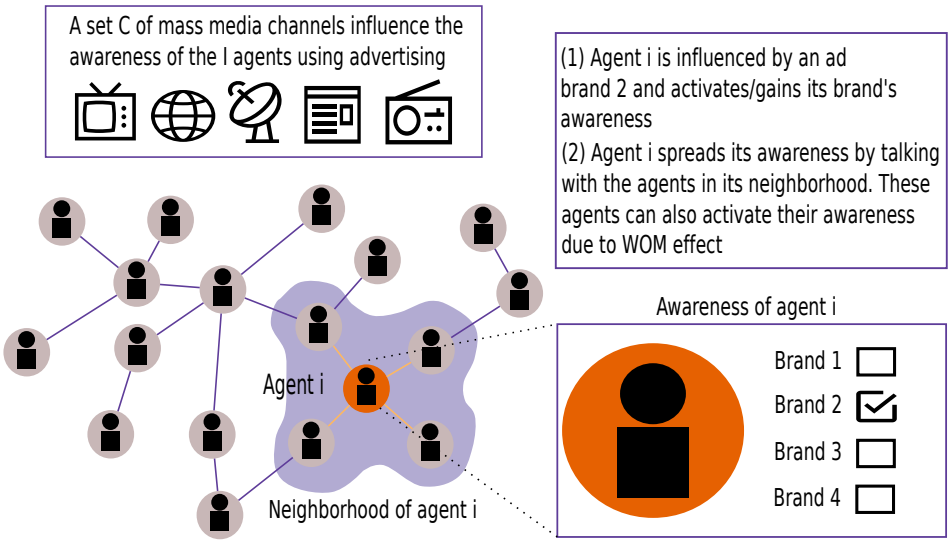


Figure 1: General scheme of the ABM showing an example of the effects of an ad to an agent. The agents reached by advertising can activate its awareness value for the ad brand and spread it to other agents in the social network.

We model agent's awareness with respect to the different brands ($b \in B$) using a binary state variable $a_i^b \in \{0, 1\}$. If $a_i^b(t) = 1$ at a given time step $t \in [1, T]$, the awareness of agent $i \in I$ is active for brand b (i.e., it is aware of the brand). In contrast, a value of 0 represents that the awareness is not active at this time step. This variable is initialized by using a parameter called initial awareness ($a^b(0) \in [0, 1]$) that represents the percentage of the agent population whose awareness is active at the start of the simulation for each brand ($a^b(0) = 1/|I| \cdot \sum_{i=1}^{|I|} a_i^b(0)$).

Agents' awareness values are dynamic because the awareness of any brand can be activated or deactivated at every simulation step. On the one hand, brand awareness can be activated by WOM interactions within the social network or through brand advertising. On the other hand, agents' brand awareness can be deactivated [56, 57] if it is not reinforced during the simulation. This activation/deactivation effect is modeled with additional parameters. First, the awareness deactivation parameter ($d \in [0, 1]$) models the probability of an agent deactivating its awareness

of a given brand. This parameter takes effect during the start of each step t, when every agent i checks each brand b that it is aware of ($a_i^b(t) = 1$). Then, each brand awareness value is deactivated with probability d by switching $a_i^b(t)$ to 0.

The agents in I compose an artificial scale-free social network [2] where their awareness values flow due to their social interactions. This effect is modeled as a contagion process [45] by the probability of talking of the agents ($p(t)_i^b \in [0, 1]$). Thus, each agent i can spread their brand awareness for every brand $b \in B$ where $a_i^b = 1$ at time step t. The WOM awareness impact parameter ($\alpha^{WOM} \in [0, 1]$) regulates the probability for each neighboring agent to activate its awareness of a brand after one of its neighbors talked about it. Additionally, the scale-free network is generated by Barabasi-Albert's preferential attachment algorithm [2], which considers a main parameter m that regulates the network's growth rate and its final density. Using m, the average degree of the social network can be computed as $\langle k \rangle = 2 \cdot m$. Finally, variable $\omega_i^b(t)$ stores the number of new conversations of the agent i about brand b in the WOM process.

The mass media channels in C influence agents randomly depending of the capability of the channel to reach high percentages of the population and the amount of investment employed. The reach parameter ($r_c \in [0, 1], \forall c \in C$) regulates the maximum percentage of the population that can be reached by channel c during a single step. We can schedule the resulting media impressions over the agent population using the given investment by assigning them at random between the agents without violating the reach constraints for the media channel [39, 40].

All the channels $c \in C$ consider an awareness impact parameter ($\alpha_c \in [0, 1]$) that regulates the likelihood for an impacted agent to activate its awareness of the announcing brand after a single media impact. In addition, the advertising scheduled in mass media may generate a viral buzz effect [40], increasing the talking probability (p_i^b) of the reached agent for the announcing brand. This effect is modeled by the buzz increment parameter (τ_c), which increases the current talking probability by a percentage of its initial value ($p_i^b(0)$). Nevertheless, the buzz created by advertising is dynamic and decays over time by buzz decay parameter ($d\tau_c$) if it is not reinforced.

4.2. Parameters selected for calibration and objective fitting functions

The parameters selected for automated calibration are those involved in modifying the agents' volume of conversations and their awareness values as those are the most uncertain and the hardest to set manually. From the parameters controlling the WOM behavior, we calibrate the m parameter for the network's generation, the initial talking probability ($p_i^b(0)$), the WOM's awareness impact (α^{WOM}), and the awareness deactivation probability (d). From the parameters which control the behavior of the mass media channels, we include the awareness impact (α_c), buzz increment (τ_c), and buzz decay ($d\tau_c$).

Each parameter is represented by a real-coded decision variable in $[0, 1]$, with the exception of m that is limited to $\{2, ..., 8\}$ because it requires integer values. The overall number of parameters to be calibrated depends on the number of channels $|C|$ in the market. Hence, the final dimensionality of each instance is $|C| \cdot 3 + 4$. Figure 2 shows an example of a chromosome encoding a candidate solution for a market instance with $|C| = 3$.

The objective fitting functions for our model are defined by Equations 1 and 2. These functions are a specification of the generic multicriteria calibration approach given in Section 3.1: f_1 computes the awareness deviation error and f_2 computes the error for the number of conversations. Both of these fitting functions are similar to a standard MAPE function. The series of target data are represented by $\widetilde{a}$ and $\widetilde{\omega}$, with the former being the target awareness values and the latter being the target WOM volume values. The simulated outputs are the result of running multiple Monte-Carlo

m	α_1	α_2	α_3	τ_1	τ_2	τ_3	$d\tau_1$	$d\tau_2$	$d\tau_3$	$p^b_i(0)$	α^{WOM}	d
4	0.02	0.015	0.05	0.1	0.15	0.22	0.12	0.05	0.23	0.2	0.1	0.01

Figure 2: Example chromosome for a model instance with $|C| = 3$ where the first decision variable matches the network's generation parameter m, an integer parameter bounded to $\{2, ..., 8\}$. The remaining decision variables match the real-coded parameters of the model, which are limited to $[0, 1]$.

simulations for each set of calibrated model's parameters encoded in a chromosome and averaging the values of these independent runs. As introduced in Section 4.1, both model's outputs are in conflict and should not be adjusted separately since improving the fitting of one output decreases the fitting of the other [39].

$$f_1 = \frac{100}{T \cdot |B|} \sum_{t=1}^{T} \sum_{b \in B} \left| \frac{a^b(t) - \widetilde{a}^b(t)}{\widetilde{a}^b(t)} \right| \tag{1}$$

$$f_2 = \frac{100}{T \cdot |B|} \sum_{t=1}^{T} \sum_{b \in B} \left| \frac{\omega^b(t) - \widetilde{\omega}^b(t)}{\widetilde{\omega}^b(t)} \right| \tag{2}$$

5. Experimentation

Section 5.1 explains the setup of our experiments, which includes the description of the different benchmark instances and algorithms' configuration. Section 5.2 discusses the calibration results using multiobjective performance indicators, attainment surfaces, and statistical tests. Finally, Section 5.3 reviews the drawbacks observed in the performance of the EMO algorithms and Section 5.4 analyzes the influence of the instances' properties on the behavior of the algorithms.

5.1. Setup

We consider a benchmark with 15 instances of the model corresponding to different market configurations with a variable number of mass media channels. These instances are the result of synthetically generating 14 additional instances from an initial real-world, baseline instance, referred as P1(25) [39]. Notice that, this original instance corresponds to a market with 7 channels, thus resulting in 25 parameters to be calibrated, the number enclosed in brackets. The additional instances are generated applying variations on the initial baseline instance. Each model variation incorporates new mass media channels that are generated from the existing ones by perturbing their investment values. The new instances also include modifications on the target data for the fitting of both outputs (i.e., WOM values and awareness values). Each new instance increases the dimensionality of the previous one as the parameters of the new channels are added as new decision variables, enabling a deeper analysis of the algorithms' performance.

On the one hand, the perturbations on the existing mass media channels C consist of multiplying the investment of each brand at each time step by a given factor. We consider reductions in the original investment by 15%, 30%, 45%, and 60%. In addition, we increase the original investment by 100%, 200%, 300%, and 400%. The decision of whether increasing or decreasing a brand investment is made at random and remains constant for each step.

On the other hand, the modifications on the target historical values for both objectives involve directly adding or subtracting a different value for each brand to each of its time steps. In order to avoid unrealistic values we truncate the resulting awareness values to a maximum of 100% and a minimum of 1%. Each addition/subtraction on the awareness target values will be by 2%, 5%, 8%, or 10%. In the case of WOM volume each addition/subtraction will be by 1,000, 2,000, 4,000, or 6,000 conversations, with a minimum value of 0. Similarly to mass media investment, the decision of whether increasing or decreasing the target values is made at random and remains constant for each step.

The new generated instances are labeled according to their dimensionality: P2(40), P3(46), P4(55), P5(61), P6(70), P7(76), P8(85), P9(91), P10(100), P11(115), P12(130), P13(145), P14(160), and P15(175). The parameter configuration of baseline P1(25) considers $|I| = 1000$, $|B| = 8$, $|C| = 7$, and $T = 52$. Awareness is initialized to $a(0) = (0.7, 0.75, 0.58, 0.25, 0.08, 0.42, 0.39, 0.34)$ and mass media channels consider $r = (0.92, 0.57, 0.54, 0.03, 0.43, 0.38, 0.69)$. The generated instances share this baseline configuration, including the reach parameters values r_c of the new channels, that take the value of the channel used for its generation. For example, if a new channel 9 is generated from the original channel 5, the reach value of the former channel is set to the value of the latter (i.e., $r_9 = r_5$).

Each EMO algorithm is run 30 times using different seeds for each run to account for the probabilistic nature of the calibration algorithms considered. Every algorithm considers a population of 100 individuals ($P = 100$) and evolves for 100 generations with a stopping criteria of 10,000 evaluations. Due to the highly time-consuming task of simulating multiple times for every parameter configuration, each evaluation of a candidate solution involves 15 Monte-Carlo runs. The distribution index of the mutation operator is set to 10 and the mutation probability value is set to $p_m = 1/n$ where n is the number of parameters being calibrated for the model instance (i.e., decision variables). The SBX crossover operator considers a crossover probability of $p_c = 1.0$ and sets its distribution index value to 5. In addition, the EMO algorithms designed to use a set of weights, such as MOEA/D, MOMBI2, and GWASF-GA, initialize their values by generating a uniform set of 100 vectors. That is the usual setup when dealing with two objectives and only 100 individuals. In addition, MOMBI2 is set to $\epsilon = 0.001$ and $\alpha = 0.5$. Finally, MOEA/D uses a neighborhood size of 20 and its differential evolution operator considers $CR = 0.5$ and $F = 0.5$. We have implemented all the EMO algorithms in Java using the jMetal framework [43]. Table 1 shows a brief summary of the EMO's parameters.

In addition, we have included a classical mathematical optimization method in our experiments, the Nelder-Mead simplex method [44]. This classical method allows us to benchmark the performance obtained by the different EMO algorithms when compared with traditional approaches. In order to adapt the Nelder-Mead algorithm to our multiobjective problem, we employ the adaptive ϵ-constraint method [14, 37], which allows single-objective optimization methods to deal with multiple objectives. The Nelder-Mead's approach also involves starting from different solutions for obtaining a Pareto set approximation. Therefore, each run generates 50 random solutions that are optimized until reaching 200 evaluations. This setup slightly modifies the one employed by the EMO algorithms because the Nelder-Mead simplex method requires a model evaluation for every single modification in the decision variables.

5.2. Analysis of the EMO algorithms' performance

We evaluate the performance of the selected EMO algorithms using widely-used unary and binary multiobjective performance indicators. First, the calibration results are analyzed using

Parameter	Value
Shared by all the EMO	
Population	100
Maximum evaluations	$10,000$
Mutation probability	$1/n$
Mutation distribution index	10
SBX p_c	1
SBX distribution index	5
MOEA/D	
Crossover rate	0.5
F	0.5
Neighborhood size	20
MOMBI2	
α	0.5
ϵ	0.001

Table 1: Parameters settings for the EMO algorithms.

a unary performance indicator and attainment surfaces in Section 5.2.1. Then, we continue the analysis by means of a binary performance indicator and a statistical test in Section 5.2.2.

5.2.1. Unary performance indicator and attainment surfaces

Unary performance indicators evaluate a single Pareto front approximation individually. We have selected hyper-volume ratio (HVR) [10] as our unary performance indicator. HVR measures the distribution and convergence of a given Pareto front approximation. It is defined as $HVR = HV(P)/HV(P*)$, with $HV(P)$ being the volume of the given Pareto front approximation and $HV(P*)$ the volume of the true Pareto front. However, we do not know the true Pareto front for any of the model instances, so we use a pseudo-optimal Pareto front for computing the HVR values instead. The pseudo-optimal Pareto front is an approximation obtained by merging all the Pareto front approximations generated by every algorithm for that instance in every independent execution and removing the dominated solutions.

HVR	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	Avg.	σ
MOEA/D	**0.946**	0.956	**0.917**	**0.885**	**0.906**	**0.936**	**0.912**	**0.93**	0.857	0.892	**0.902**	0.812	**0.794**	**0.895**	**0.884**	**0.895**	0.045
SPEA2	0.904	0.969	0.817	0.851	0.859	0.914	0.878	0.803	0.757	0.901	0.825	0.748	0.688	0.794	0.803	0.834	0.073
SMS-EMOA	0.901	0.969	0.795	0.857	0.85	0.93	0.87	0.879	0.836	0.908	0.863	0.802	0.779	0.868	0.843	0.863	0.051
IBEA	0.859	0.95	0.758	0.832	0.778	0.919	0.851	0.876	0.851	0.878	0.863	0.809	0.737	0.84	0.845	0.843	0.056
NSGA-II	0.902	**0.972**	0.821	0.856	0.865	0.926	0.888	0.821	0.757	**0.909**	0.839	0.764	0.746	0.822	0.822	0.847	0.065
GWASFGA	0.881	0.959	0.769	0.816	0.808	0.899	0.825	0.853	0.827	0.863	0.861	0.742	0.69	0.855	0.841	0.833	0.065
MOMBI2	0.873	0.921	0.799	0.846	0.831	0.894	0.838	0.857	**0.863**	0.843	0.878	**0.824**	0.758	0.839	0.857	0.848	0.039
Nelder-Mead	0.161	0.169	0.074	0.177	0.17	0.284	0.334	0.353	0.608	0.549	0.383	0.501	0.288	0.348	0.411	0.321	0.155

Table 2: Average HVR values for every algorithm and model instance. The best value for each model instance is shown in bold font. Additionally, the average HVR values across the multiple instances is shown along with the standard deviation (σ).

Table 2 shows the computed values of HVR for the resulting Pareto front approximations of each algorithm for every model instance. These values are presented using the average of the

individual HVR values computed for the individual Pareto front approximations resulting in each of the 30 algorithm executions. The average HVR values show that MOEA/D consistently achieves better values than the other algorithms for most model instances, obtaining the best average HVR in all but four instances. However, for these four instances MOEA/D obtains values close to the best ones. For example, in P9 and P12 the best HVR values are obtained by MOMBI2 (0.863 and 0.824), closely followed by MOEA/D (0.857 and 0.812). In addition, MOEA/D obtains the best average value across the 15 instances with the second lowest standard deviation. These results also highlight the poor performance of the Nelder-Mead simplex method when calibrating our problem instances. It obtains the lowest HVR value for every problem instance with a very significant difference with respect to the EMO approaches (an average value of 0.321 while the worst performing EMO algorithm is over 0.83).

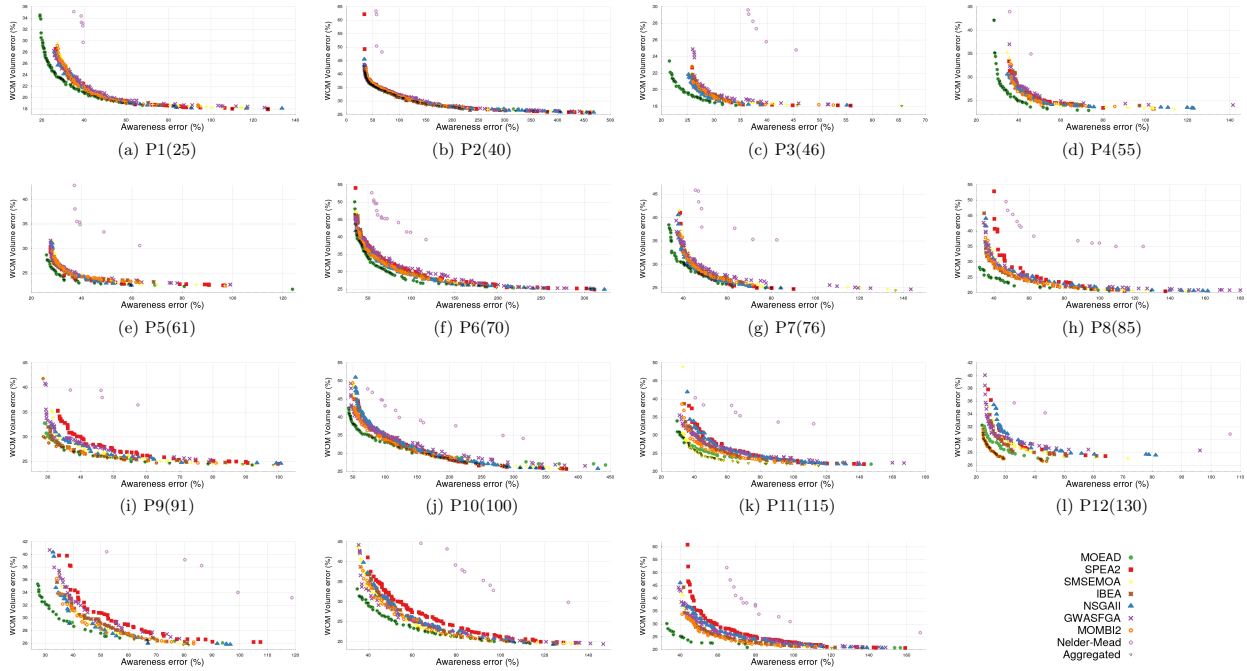
The results of HVR can be visually corroborated using attainment surfaces [19] for each model calibration instance, reported in Figure 3. These attainment surfaces exhibit that the surface obtained by MOEA/D (represented using green filled circles) outperforms the other algorithms for most instances. These surfaces are coherent with the HVR values and P12 is the only problem instance where MOEA/D is visually dominated by other algorithm (MOMBI2, represented by orange circles). Nevertheless, these surfaces also reflect how every EMO algorithm performs competitively for certain instances such as P2 where every attainment surface converges to the aggregated Pareto front approximation. By using these surfaces we can also observe the visual difference between Nelder-Mead and the EMO algorithms, where Nelder-Mead's surface is outperformed by every EMO algorithm.

5.2.2. Binary performance indicators and statistical significance/tests

A binary performance indicator compares two given Pareto front approximations generated for the same model instance. Our selected binary performance indicator is the multiplicative I_ϵ measure [68]. The calculation of $I_\epsilon(P,Q)$ for Pareto front approximations P and Q is shown in the following equation: $I_\epsilon(P,Q) = \inf_{\epsilon\in\mathbb{R}}\{\forall z^2 \in B\ \exists z^1 \in A : z^1 \succeq_\epsilon z^2\}$. The value computed by $I_\epsilon(P,Q)$ represents the minimum factor required to multiply every element in P in order to weekly dominate Q. That is, the minimum ϵ so P ϵ-dominates Q. As our calibration problem constitutes a minimization problem, if $I_\epsilon(P,Q) < I_\epsilon(Q,P)$ then we can assume that P is better than Q.

Tables 3 and 4 present the multiplicative I_ϵ values computed for the resulting Pareto front approximations of each algorithm for every model instance. These values are the average of each possible $I_\epsilon(P,Q)$, with P and Q being any pair of Pareto front approximations of different algorithms, resulting from any of the 30 independent executions (i.e., a pair-wise comparison of every run). The values for Tables 3 and 4 support the previous conclusions drawn from the HVR indicator: MOEA/D outperforms the remaining algorithms for most model instances. MOEA/D obtains a lower average I_ϵ value for every comparison with the other algorithms, with the exception of P2, where GWASF-GA obtains a better indicator value. The Nelder-Mead's simplex method is again outperformed by every EMO algorithm in every model instance.

In addition, we develop a statistical test and study the significance of the I_ϵ values to avoid that isolated results could bias our former analysis. We perform this test following the methodology described in [54, 7]: let N be the number of repetitions of two algorithms A and B; then let A_i and B_j be two arbitrary resulting Pareto front approximations with $1 \leq i \leq N$ and $1 \leq j \leq N$; finally let $p_{A_i}(B_j)$ be 1 if A_i dominates B_j based on the computed I_ϵ value (i.e., $I_\epsilon(A_i,B_j) \leq 1$ and $I_\epsilon(B_j,A_i) > 1$) and 0 otherwise. Using $p_{A_i}(B_j)$, we can define P by the equation $P_{A_i}(B) = 1/N \cdot \sum_{j=1}^{N} p_{A_i}(B_j)$ as the percentage or resulting Pareto front approximations obtained by algorithm B



(a) P1(25) (b) P2(40) (c) P3(46) (d) P4(55)

(e) P5(61) (f) P6(70) (g) P7(76) (h) P8(85)

(i) P9(91) (j) P10(100) (k) P11(115) (l) P12(130)

(m) P13(145) (n) P14(160) (o) P15(175)

Figure 3: Attainment surfaces for the different problem instances. Each attainment surface represents the aggregated Pareto front approximation obtained by each algorithm: MOEA/D (green filled circles), SPEA2 (red squares), SMS-EMOA (yellow crosses), IBEA (brown asterisks), NSGA-II (blue triangles), GWASF-GA (purple blades), MOMBI2 (empty orange circles), and Nelder-Mead (light purple empty circles). Finally, inverted triangles represent the aggregated Pareto front.

P1(25)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0398**	**1.0347**	**1.0026**	**1.0383**	**1.0184**	**1.0118**	**0.7244**
SPEA2	1.3562	-	**1.043**	**1.0423**	1.0525	1.0798	**1.0399**	**0.8211**
SMS-EMOA	1.3804	1.0601	-	**1.0465**	1.0599	1.0934	**1.0395**	**0.8297**
IBEA	1.3821	1.0899	1.0793	-	1.0894	1.1004	**1.0494**	**0.8301**
NSGA-II	1.3571	**1.0507**	**1.0428**	**1.0403**	-	1.0737	**1.0394**	**0.8192**
GWASF-GA	1.2867	**1.0596**	**1.0596**	**1.0475**	**1.0595**	-	**1.0545**	**0.7947**
MOMBI2	1.4394	1.0991	1.085	1.0717	1.0971	1.1315	-	**0.8543**
Nelder-Mead	1.8792	1.8375	1.8273	1.7152	1.8361	1.8014	1.7512	-

P2(40)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0736**	**1.0681**	**1.0518**	**1.0725**	1.0527	**1.0301**	**0.5817**
SPEA2	1.1193	-	**1.0503**	**1.0479**	1.0619	1.0991	**1.0418**	**0.6305**
SMS-EMOA	1.1576	1.0794	-	**1.049**	1.0831	1.1331	**1.0454**	**0.6542**
IBEA	1.1934	1.122	1.0903	-	1.1221	1.1677	**1.0559**	**0.6747**
NSGA-II	1.1091	**1.0488**	**1.0419**	**1.0397**	-	1.0878	**1.0328**	**0.6275**
GWASF-GA	**1.0485**	**1.0557**	**1.0625**	**1.0632**	**1.058**	-	**1.0472**	**0.5933**
MOMBI2	1.2211	1.177	1.1495	1.112	1.1758	1.1985	-	**0.6889**
Nelder-Mead	4.9663	5.1355	5.0527	4.7961	5.1252	4.9992	4.6028	-

P3(46)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0035**	**0.9973**	**0.9812**	**1.0026**	**0.9891**	**0.9937**	**0.728**
SPEA2	1.181	-	**1.0256**	**1.0202**	1.0351	**1.0156**	**1.0276**	**0.8331**
SMS-EMOA	1.1927	1.0428	-	**1.0292**	1.0463	**1.0267**	1.038	**0.8433**
IBEA	1.2139	1.0547	1.0481	-	1.057	1.038	1.0445	**0.8599**
NSGA-II	1.1751	**1.0282**	**1.0231**	**1.0157**	-	**1.0126**	**1.0238**	**0.8316**
GWASF-GA	1.1997	1.0464	1.0414	**1.0343**	1.0504	-	1.0419	**0.847**
MOMBI2	1.1917	1.0367	**1.0315**	**1.0211**	1.0401	**1.0207**	-	**0.8428**
Nelder-Mead	2.5998	2.5702	2.5519	2.5005	2.568	2.5343	2.542	-

P4(55)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.023**	**1.0289**	**1.0094**	**1.0218**	**0.9975**	**1.0206**	**0.7981**
SPEA2	1.1484	-	1.0441	**1.0385**	1.047	**1.0359**	**1.0473**	**0.8805**
SMS-EMOA	1.1563	**1.0386**	-	**1.0305**	1.046	**1.0395**	**1.0401**	**0.8882**
IBEA	1.1619	1.0464	1.0472	-	1.0487	**1.0412**	1.0385	**0.8948**
NSGA-II	1.135	**1.0308**	**1.0361**	**1.0264**	-	**1.0236**	**1.0361**	**0.871**
GWASF-GA	1.1347	1.055	1.0639	1.0534	1.0585	-	1.0641	**0.867**
MOMBI2	1.1695	1.0496	1.049	**1.0325**	1.0545	**1.0492**	-	**0.9001**
Nelder-Mead	1.6515	1.663	1.6692	1.6161	1.654	1.6141	1.6392	-

P5(61)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0121**	**1.0043**	**0.9888**	**1.0093**	**0.9941**	**0.9958**	**0.8466**
SPEA2	1.0986	-	1.0376	**1.0425**	1.0415	**1.0426**	1.0408	**0.9117**
SMS-EMOA	1.088	**1.0314**	-	**1.03**	1.0349	**1.0339**	**1.0282**	**0.9043**
IBEA	1.0901	1.0591	1.0499	-	1.0565	1.0468	1.0394	**0.9013**
NSGA-II	1.0865	**1.026**	**1.0278**	**1.0316**	-	**1.0314**	**1.03**	**0.9011**
GWASF-GA	1.0841	1.0459	1.042	**1.0365**	1.0464	-	1.0398	**0.8908**
MOMBI2	1.0899	**1.0369**	1.033	**1.0305**	1.038	**1.0371**	-	**0.9092**
Nelder-Mead	1.7265	1.7202	1.7024	1.6418	1.7149	1.6832	1.6804	-

P6(70)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0505**	**1.0557**	**1.0344**	**1.0545**	**1.0515**	**1.0206**	**0.7621**
SPEA2	1.0932	-	**1.0582**	**1.0557**	1.0607	1.0632	**1.0689**	**0.8034**
SMS-EMOA	1.0972	1.0597	-	**1.0464**	1.0656	1.0799	**1.0619**	**0.7961**
IBEA	1.1314	1.1011	1.0908	-	1.1031	1.1293	1.0735	**0.7985**
NSGA-II	1.0791	**1.0413**	**1.046**	**1.0424**	-	**1.0538**	**1.0545**	**0.7934**
GWASF-GA	1.1005	**1.0615**	**1.0786**	**1.0836**	1.0759	-	**1.0875**	**0.8036**
MOMBI2	1.106	1.1035	1.1009	**1.0672**	1.1036	1.1072	-	**0.7821**
Nelder-Mead	1.7925	1.803	1.8023	1.7494	1.8046	1.793	1.6763	-

P7(76)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0208**	**1.0117**	**1.002**	**1.0196**	**0.9931**	**1.0003**	**0.7977**
SPEA2	1.1284	-	1.0396	**1.0408**	1.0489	**1.0414**	**1.0422**	**0.8819**
SMS-EMOA	1.1225	**1.0345**	-	**1.0338**	1.0451	**1.0368**	**1.0345**	**0.8756**
IBEA	1.1278	1.0464	1.0438	-	1.0531	**1.044**	**1.0364**	**0.8832**
NSGA-II	1.114	**1.0258**	**1.0282**	**1.0276**	-	**1.0292**	**1.0283**	**0.8698**
GWASF-GA	1.1176	1.063	1.0571	1.0539	1.0666	-	**1.0523**	**0.8629**
MOMBI2	1.1351	1.0567	1.0539	1.0471	1.0634	1.0528	-	**0.8876**
Nelder-Mead	1.6197	1.6104	1.5963	1.5701	1.6085	1.5667	1.5616	-

P8(85)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0283**	**1.0413**	**1.0222**	**1.0284**	**1.0183**	**1.0192**	**0.7381**
SPEA2	1.3567	-	1.1705	1.2018	1.0869	1.1694	1.1974	**0.9229**
SMS-EMOA	1.2166	**1.0091**	-	1.0841	**1.0165**	**1.0596**	**1.0829**	**0.843**
IBEA	1.1851	**1.0543**	**1.073**	-	**1.0565**	**1.0655**	**1.0671**	**0.8257**
NSGA-II	1.3252	**1.0521**	1.146	1.1755	-	1.1457	1.1722	**0.9068**
GWASF-GA	1.2102	**1.0517**	1.0944	1.0999	**1.0587**	-	**1.1006**	**0.8303**
MOMBI2	1.2585	**1.1061**	1.1428	1.1413	**1.1122**	1.1366	-	**0.8745**
Nelder-Mead	2.0949	2.0418	2.0544	1.9747	2.0387	2.0213	1.9565	-

Table 3: Average I-ϵ values for every pair of algorithms for P1, P2, P3, P4, P5, P6, P7, and P8 instances. Best values for each pair of algorithms are shown using bold font.

P9(91)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0308**	**1.0501**	**1.0492**	**1.031**	**1.04**	**1.0588**	**0.8305**
SPEA2	1.1565	-	1.1079	1.1262	1.056	1.1389	1.1444	**0.9295**
SMS-EMOA	1.0859	**1.011**	-	1.062	**1.0135**	1.0656	1.0785	**0.8673**
IBEA	1.0682	**1.0186**	**1.0408**	-	**1.0201**	**1.0495**	1.0614	**0.853**
NSGA-II	1.1536	**1.0548**	1.1069	1.1247	-	1.1336	1.1426	**0.9259**
GWASF-GA	1.0693	**1.0402**	**1.0598**	1.061	**1.0398**	-	1.0732	**0.8345**
MOMBI2	1.0674	**1.0277**	**1.0479**	**1.0478**	**1.0274**	**1.0492**	-	**0.8446**
Nelder-Mead	1.8012	1.7617	1.7964	1.7787	1.7593	1.7794	1.7872	-

P10(100)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.1088**	**1.1079**	**1.0623**	**1.1138**	**1.0753**	**1.0435**	**0.7763**
SPEA2	1.2404	-	**1.0534**	**1.0492**	1.0572	1.1913	**1.0814**	**0.8328**
SMS-EMOA	1.2665	1.0685	-	**1.0494**	1.0669	1.2168	**1.0897**	**0.8389**
IBEA	1.3202	1.116	1.1065	-	1.1145	1.2743	1.1055	**0.8585**
NSGA-II	1.2482	**1.0566**	**1.0501**	**1.0456**	-	1.1981	**1.0814**	**0.8309**
GWASF-GA	1.0917	**1.0929**	**1.0998**	**1.0877**	**1.0988**	-	**1.0787**	**0.8035**
MOMBI2	1.2741	1.1544	1.1499	**1.0901**	1.1566	1.2297	-	**0.8399**
Nelder-Mead	2.0886	1.8768	1.8639	1.8167	1.8787	2.0105	1.8482	-

P11(115)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0383**	**1.0484**	**1.0292**	**1.04**	**1.0362**	**1.0425**	**0.7775**
SPEA2	1.2678	-	1.0932	1.1233	1.0661	1.1961	1.1572	**0.9519**
SMS-EMOA	1.2128	**1.0201**	-	1.0771	**1.0308**	1.1463	1.1084	**0.9136**
IBEA	1.168	**1.0498**	**1.0614**	-	**1.0516**	1.105	1.0747	**0.8769**
NSGA-II	1.2499	**1.0391**	1.0796	1.1074	-	1.1802	1.1403	**0.9395**
GWASF-GA	1.1132	**1.0602**	**1.0769**	**1.0613**	**1.0641**	-	**1.0766**	**0.814**
MOMBI2	1.1473	**1.0463**	**1.0554**	**1.0365**	**1.0471**	1.0872	-	**0.8598**
Nelder-Mead	1.887	1.8816	1.8895	1.8076	1.8823	1.8476	1.8152	-

P12(130)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.022**	**1.0346**	**1.0303**	**1.0213**	**1.0108**	**1.0361**	**0.7534**
SPEA2	1.171	-	1.0747	1.0919	1.0513	1.1252	1.1105	**0.842**
SMS-EMOA	1.1379	**1.0293**	-	1.0585	**1.021**	1.0953	1.0775	**0.8222**
IBEA	1.1081	**1.0215**	**1.0317**	-	**1.0132**	1.0691	1.052	**0.8051**
NSGA-II	1.1676	**1.048**	1.064	1.0816	-	1.1239	1.103	**0.8438**
GWASF-GA	1.0839	**1.0475**	**1.0615**	**1.0621**	**1.0468**	-	1.0692	**0.7818**
MOMBI2	1.0837	**1.0107**	**1.0211**	**1.0265**	**1.0048**	**1.0441**	-	**0.7858**
Nelder-Mead	1.7398	1.6749	1.6962	1.6943	1.6719	1.6915	1.7104	-

P13(145)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0256**	**1.051**	**1.0349**	**1.0384**	**1.0104**	**1.0511**	**0.744**
SPEA2	1.2326	-	1.0922	1.072	1.0866	1.1095	1.1032	**0.862**
SMS-EMOA	1.2015	**1.018**	-	**1.0286**	**1.0488**	1.0832	**1.057**	**0.8441**
IBEA	1.2382	**1.0474**	1.0724	-	1.0812	1.1179	**1.0823**	**0.869**
NSGA-II	1.1809	**1.0192**	1.0542	**1.0384**	-	**1.064**	**1.0643**	**0.8295**
GWASF-GA	1.1481	**1.0468**	**1.0803**	**1.0649**	1.0648	-	**1.0871**	**0.8042**
MOMBI2	1.2741	**1.0877**	1.1095	1.088	1.1179	1.1533	-	**0.8997**
Nelder-Mead	2.0712	1.9167	1.9591	1.9175	1.9574	1.9523	1.9364	-

P14(160)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0142**	**1.0463**	**1.0264**	**1.0259**	**1.0372**	**1.0274**	**0.6615**
SPEA2	1.1558	-	1.1044	1.0754	1.0662	1.1316	1.0993	**0.7218**
SMS-EMOA	1.0959	**1.0029**	-	**1.0235**	**1.0194**	**1.0679**	**1.0415**	**0.6843**
IBEA	1.1351	**1.0331**	1.0857	-	**1.0519**	1.12	1.0707	**0.6961**
NSGA-II	1.1366	**1.0286**	1.0846	1.0557	-	1.1111	**1.0793**	**0.707**
GWASF-GA	1.0794	**1.0332**	1.069	**1.0531**	**1.0462**	-	**1.0567**	**0.6837**
MOMBI2	1.137	**1.0657**	1.1033	**1.0674**	1.0803	1.1184	-	**0.6915**
Nelder-Mead	2.2442	2.1266	2.1995	2.1276	2.155	2.2156	2.0988	-

P15(175)	MOEAD	SPEA2	SMS-EMOA	IBEA	NSGA-II	GWASF-GA	MOMBI2	Nelder-Mead
MOEAD	-	**1.0294**	**1.0428**	**1.0306**	**1.0331**	**1.0278**	**1.0471**	**0.7095**
SPEA2	1.2198	-	1.1024	1.1218	1.0788	1.1289	1.1481	**0.7989**
SMS-EMOA	1.1603	**1.0186**	-	1.062	**1.0373**	1.0869	1.0833	**0.7549**
IBEA	1.147	**1.0476**	**1.0615**	-	**1.0546**	1.0835	1.0665	**0.7434**
NSGA-II	1.1874	**1.0308**	1.0764	1.0922	-	1.1041	1.1166	**0.7783**
GWASF-GA	1.1287	**1.0477**	**1.0701**	**1.0688**	**1.0548**	-	1.0892	**0.7536**
MOMBI2	1.1328	**1.0455**	**1.0557**	**1.0369**	**1.051**	**1.076**	-	**0.7284**
Nelder-Mead	1.9677	1.9101	1.9207	1.8513	1.9125	1.9083	1.8518	-

Table 4: Average I-ϵ values for every pair of algorithms for P9, P10, P11, P12, P13, P14, and P15 instances. Best values for each pair of algorithms are shown using bold font.

Figure 4: Boxplots representing the ϵ dominance percentage values for each model instance and every pair of EMOs.

that are dominated by A_i.

We have included boxplots representing the resulting ϵ dominance percentage values in Figure 4, which contains the computed $P_A(B)$ for every pair of algorithms and each model instance with the exception of Nelder-Mead's simplex method, which has been widely outperformed at this stage of the analysis. In these charts we can observe how MOEA/D generally obtains bigger dominance percentages than the remaining algorithms. On the one hand, the boxplots on the right of the MOEA/D label contain the values of $P_B(MOEA/D)$. In these charts we can observe how both their boxes and whiskers cover a considerable percentage of the interval, implying a big dominance probability. In contrast, boxplots below the MOEA/D label contain the values of $P_{MOEA/D}(A)$, where we can notice that, in general, the values for all instances but P9 are small. This is coherent with the previous conclusions of our analysis as MOEA/D was already outperformed by other algorithms in instances like P9.

Finally, we can consider vector $P_A(B) = P_{A_1}(B), P_{A_2}(B), \ldots, P_{A_N}(B))$ as a random variable representing the percentage of times that algorithm A outperforms algorithm B, since it is the proportion of resulting Pareto approximations of algorithm A dominating the Pareto approximations delivered by algorithm B. Therefore, if the expectation of $P_A(B)$ is greater than the expectation of $P_B(A)$ we can claim that A is better than B because it is more likely that the resulting Pareto approximations of A dominate those obtained by B. Our selected test is the Wilcoxon ranksum test (null hypothesis $E(P_A(B)) = E(P_B(A))$, alternate hypothesis $E(P_A(B)) > E(P_B(A))$), seeing that it has proven to be useful when analyzing the performance of evolutionary algorithms [20]. The significance level considered is 0.05.

Table 5 shows the significance for the resulting $p-$values of the statistical test. These results are again consistent with the previous indicator values, as MOEA/D shows an outstanding and robust behavior, being able to perform significantly better than the remaining algorithms in most instances. Hence, MOEA/D is the best performing decomposition-based algorithm for our problem, since it almost always outperforms GWASF-GA. Regarding the performance of the rest of the algorithms, we can see how the Pareto dominance-based EMO algorithms (NSGA-II and SPEA2) outperform most of the algorithms for the first seven instances. However, if we compare these two algorithms, we can observe that SPEA2 does not significantly outperform NSGA-II in any instance, suggesting that NSGA-II is the best algorithm from this family when dealing with the ABM calibration problem. SMS-EMOA would be the best performing indicator-based EMO algorithm but we can find some instances like P12 where it is outperformed by MOMBI2.

Nevertheless, the behavior of MOEA/D is eroded when dealing with specific instances like P6 or some of the bigger instances like P9, P12, and P15. Although MOEA/D obtains the best HVR values for some of these instances, the statistical tests revealed that it is dominated by other algorithms. In the case of P6, SMS-EMOA and NSGA-II are the best performing algorithms and significantly outperform MOEA/D. As we pointed out, MOMBI2 arises as the best performing algorithm for the P12 model instance. The p-values for the P9 instance, where SMS-EMOA and IBEA perform significantly better than the other algorithms, with SMS-EMOA finally outperforming IBEA. SMS-EMOA shows better convergence for this instance, which explains the better dominance $P_B(A)$ values shown at Figure 4 and corroborated by the statistical test. Finally, SMS-EMOA is the best performing EMO for the P15 instance, as it dominates the remaining six algorithms despite obtaining the fourth average HVR value, which shows the great convergence of SMS-EMOA for this instance. Therefore, we can observe how SMS-EMOA's specific features (such as its combination of hypervolume maximization and its replacement strategy) are effective with

optimization problems with these characteristics.

5.3. Drawbacks of the methods

The previous sections have shown how the EMO algorithms, specially MOEA/D and SMS-EMOA, successfully perform in our ABM calibration problem. However, we also acknowledge that these algorithms present certain drawbacks. For example, we can observe how the increased dimensionality of the bigger instances erodes the behavior of most algorithms (we will analyze the impact of the dimensionality in next Section 5.4). The drawbacks of the EMO algorithms are presented next:

- The main drawback of MOEA/D is its sensitivity to the properties of the different problem instances, further discussed in Section 5.4. In addition, we can observe how the increased dimensionality reduces its performance. This can be identified even if MOEA/D outperforms other algorithms. For example, the values of Table 2 show that its HVR value for instances bigger than P8 do not surpasses the computed average HVR value for every instance, showing a decay of performance even when outperforming other algorithms.

- SPEA2 and NSGA-II present similar drawbacks. According to their p-values, they perform competitively for the instances below P10, where they significantly outperform several EMO algorithms. However, they reduce their performance, specially for the bigger instances, where they are outperformed by IBEA and GWASF-GA. In addition, a direct comparison between SPEA2 and NSGA-II reveals that NSGA-II is better in most instances.

- SMS-EMOA is the most robust EMO algorithm regarding the dimensionality of our calibration problem. This could be related to its combination of hypervolume maximization and the replacement strategy that mainly targets the solutions that contribute poorly to the hypervolume of its respective fronts. However, SMS-EMOA is unsuccessful dealing with the smaller instances since it barely outperforms GWASF-GA and IBEA for the instances below P6, where its strategy seems to obtain a lesser impact.

- IBEA and GWASF-GA also present similar drawbacks. In general, they both show a poor performance for our ABM calibration problem, with the only exceptions of P9 for IBEA and P14 for GWASF-GA. However, this can be related with the loss of performance observed in the rest of the algorithms.

- Finally, MOMBI2 stands out by obtaining the best HVR values for two instances and increasing its performance with the number of decision variables of the instances. Nevertheless, its behavior seems unstable, since it outperforms the remaining algorithms for P12 but fails to outperform any of them for P14.

5.4. Influence of the instances' properties on the algorithms' performance

In view of the results obtained by both the unary and binary indicators, we can observe how specific properties of the problem instances are affecting the performance of the evaluated EMO algorithms. These properties are the shape of the feasible region, the shape of the Pareto front, and the dimensionality of the problem instance. Some studies [26, 36] have pointed out a relationship between the performance of decomposition-based EMO algorithms (such as MOEA/D, the best performing algorithm in our study) and the shapes of both the feasible region and the Pareto front.

MOEA/D	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
SPEA2	-	-	+	-	+	-	+	+	-	-	-	+	+	+	+
SMS-EMOA	-	-	+	+	+	-	+	-	-	-	-	-	+	-	-
IBEA	+	-	+	+	+	-	+	+	-	-	+	-	+	+	+
NSGA-II	-	-	+	-	+	-	+	+	+	-	+	+	+	+	+
GWASF-GA	+	-	+	+	+	+	+	+	-	-	+	+	+	+	+
MOMBI2	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+

SPEA2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SMS-EMOA	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
IBEA	+	-	+	+	+	-	+	-	-	+	-	-	-	-	-
NSGA-II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GWASF-GA	+	-	+	+	+	+	+	-	+	-	-	+	-	-	-
MOMBI2	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-

SMS-EMOA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+
SPEA2	-	-	-	-	-	+	-	-	+	+	+	+	+	+	+
IBEA	+	-	+	+	+	+	+	+	-	+	+	-	+	+	+
NSGA-II	-	-	-	-	-	+	-	+	+	-	+	+	-	+	+
GWASF-GA	+	-	+	+	+	+	+	+	+	-	+	+	-	+	+
MOMBI2	+	+	-	-	-	+	+	+	+	-	-	-	-	+	+

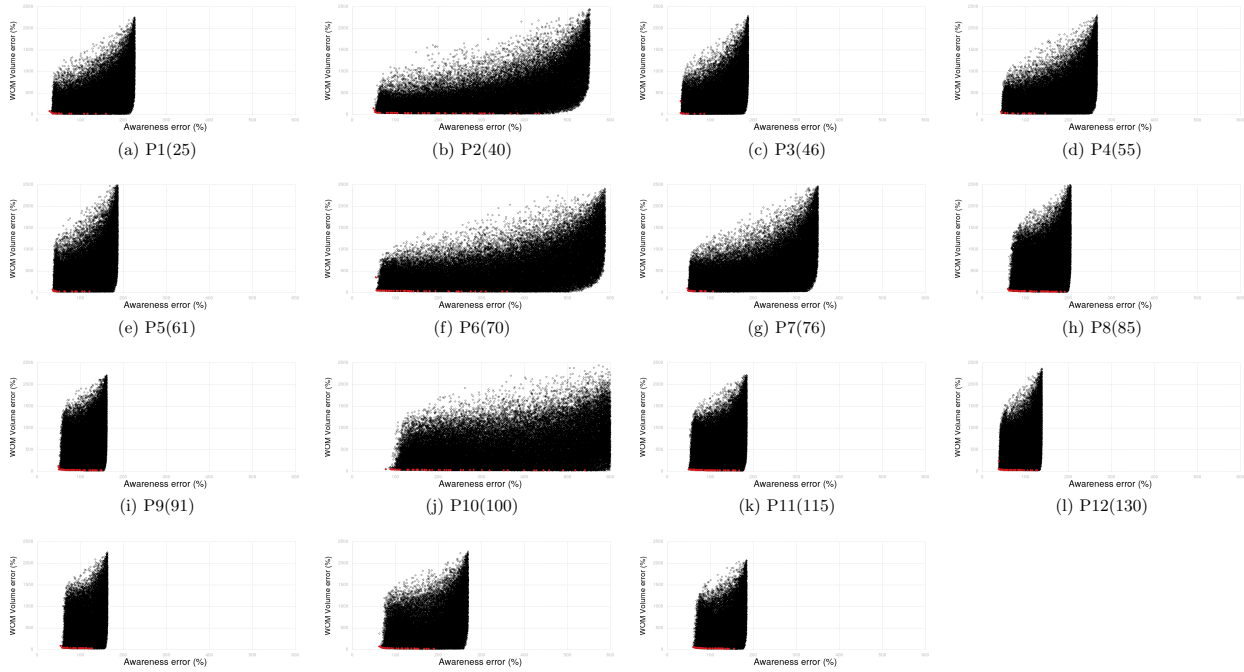
IBEA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
SPEA2	-	-	-	-	-	-	-	-	+	-	-	+	+	+	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NSGA-II	-	-	-	-	-	-	-	-	+	-	-	+	-	+	+
GWASF-GA	-	-	-	+	-	-	+	-	+	-	-	+	-	-	-
MOMBI2	-	+	-	-	-	-	+	+	-	-	-	-	-	+	-

NSGA-II	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
SPEA2	-	-	+	+	-	+	+	-	-	+	+	+	+	+	+
SMS-EMOA	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-
IBEA	+	-	+	+	+	+	+	+	-	+	-	-	+	-	-
GWASF-GA	+	-	+	+	+	+	+	+	+	-	-	+	-	-	-
MOMBI2	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-

GWASF-GA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SPEA2	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IBEA	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
NSGA-II	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
MOMBI2	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-

MOMBI2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
MOEA/D	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
SPEA2	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-
SMS-EMOA	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
IBEA	-	-	+	+	+	-	-	-	-	-	+	+	+	-	+
NSGA-II	-	-	-	-	-	-	-	-	+	-	-	+	+	-	+
GWASF-GA	-	-	+	+	+	-	+	-	+	-	+	+	+	-	+

Table 5: Significance of the results with respect of the computed $p-$values for every EMO algorithm against the other methods for every model instance. Cells marked as + represent statistical significance.



(a) P1(25) (b) P2(40) (c) P3(46) (d) P4(55)

(e) P5(61) (f) P6(70) (g) P7(76) (h) P8(85)

(i) P9(91) (j) P10(100) (k) P11(115) (l) P12(130)

(m) P13(145) (n) P14(160) (o) P15(175)

Figure 5: Sampled solutions for the different instances using 100,000 random configurations. The non-dominated solutions are coloured in red. The axis of the charts have been fixed for comparing the different shapes of the feasible region for each problem instance.

Figure 5 shows an approximation to the search space configuration of the problem instances tackled in the current contribution using scatter-plots. In addition, Figure 6 displays the shape of the Pareto fronts. The shape of the feasible region is approximated by sampling 100,000 random configurations for each problem instance. In the plots in Figure 5, we can observe that the search space extent is considerably bigger for P2, P6, P7, and P10 instances when compared with the rest of the problem instances. Therefore, the shape of the feasible region for these instances can explain their difficulty, specially for the performance of the EMO algorithms that employ reference points [36].

We approximate the shape of the global Pareto fronts by using the aggregated Pareto fronts, which contain the overall non-dominated solutions obtained for each problem instance. In the plot in Figure 6, we can observe that P2, P6, and P10 instances have a long tail shape compared with the rest of instances. These long tail shapes can explain a reduction of performance for the algorithms using weight vectors because these shapes are non-symmetric and mismatch a distribution of uniformly generated weight vectors [26]. Hence, the problem instances with these properties may require a customized set of weight vectors for improving its performance.

We can also observe how these instances' properties produce different effects on the behavior of the EMO algorithms for the identified instances P2, P6, P7, and P10:

- In the case of P2, it was observed how most EMO algorithms obtain HVR values over 0.95 and compete similarly, since MOMBI2 is the only EMO that results dominated by the rest of EMO algorithms. Thus, the long tail shape of the P2 instance is not sufficient for eroding the behavior of the EMO algorithms, but this could be explained by reduced the number of variables considered by this instance (only 40).

- With respect to the P6 instance, we have seen that MOEA/D obtains the best HVR values, closely followed by SMS-EMOA and NSGA-II. Despite that, the results of the statistical test pointed out that MOEA/D is dominated by SMS-EMOA and NSGA-II. A deeper analysis of the Pareto sets obtained by MOEA/D's in its individual runs reveals that for some of these runs MOEA/D performed poorly. This lack of consistency solving P6 explains why is dominated by SMS-EMOA and NSGA-II although it obtains a better average HVR value across the 30 runs.

- It can be observed that the shape of the feasible region for P7 is not so long-tailed as other instances, but it is still remarkable. However, the results for this instance are comparable with those obtained for other regular instances, as MOEA/D is clearly the best performing algorithm (obtains the best HVR values with some margin and significantly outperforms the remaining EMOs). This suggests that the shape of the feasible region for P7 is not wide enough for eroding MOEA/D's behavior.

- In the case of P10, NSGA-II obtains the best HVR value, closely followed by SMS-EMOA and SPEA2. However, the results of the statistical tests for P10 showed that none of the EMO algorithms is able to significantly dominate more than two of the remaining algorithms. Similarly to the P2 instance, it could be argued that most EMO algorithms are performing similarly, but in this case the HVR values are sensibly lower for P10 than for P2. Because MOEA/D is the best performing algorithm for most of the instances in this study, we can argue that its behavior is more influenced by P10's properties than NSGA-II or SMS-EMOA.

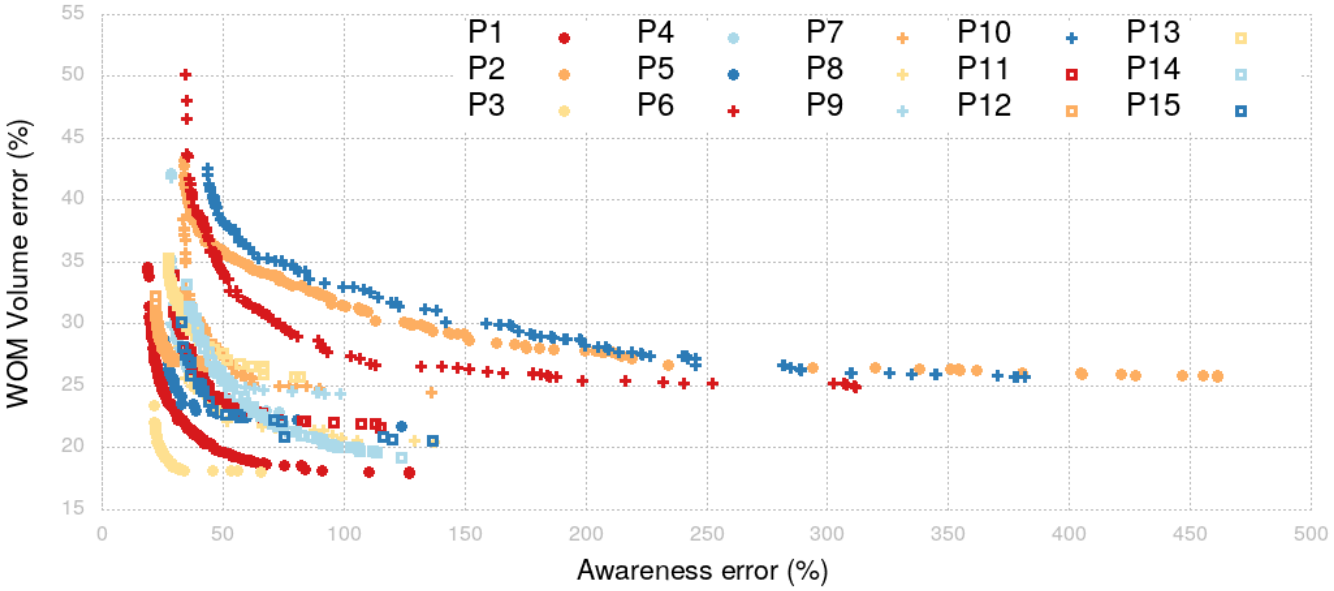


Figure 6: Approximation to the shape of the Pareto fronts using the aggregated Pareto fronts for each problem instance. Each front contains the overall non-dominated solutions obtained for each problem instance.

6. Final remarks

In this paper we have conducted an exhaustive analysis of the performance of several EMO algorithms when calibrating multiple instances of an ABM for marketing jointly considering global awareness and WOM volume as its main outputs. Starting from an initial model instance built with real data, we have synthetically generated 14 additional model instances by changing the market characteristics to achieve a progressive dimensionality increase. Using this set of benchmarks, we have tested the calibration performance of seven EMO algorithms from different families and a classical mathematical method. We have analyzed the calibration results using both unary and binary performance indicators along with Wilcoxon ranksum test for assessing the significance of the results. In addition, we have used attainment surfaces for visually supporting the analysis of the performance indicators.

The results of our experimentation allow to provide the following insights:

- MOEA/D shows outstanding and robust behavior for our problem, being able to perform significantly better than the other EMO algorithms in most instances. Therefore, the decomposition-based strategy proposed by MOEA/D is clearly the best performing for the search space of the analyzed problem.

- We could also observe how the performance of MOEA/D was reduced when dealing with certain instances. A deeper analysis of the shape of the feasible region and the shape of the Pareto front of the instances revealed that some of them have certain characteristics that can affect the performance of decomposition-based algorithms [26, 36].

- The calibration results on the high-dimensional instances have shown that its dimensionality erodes the performance of most of the algorithms. This is not the only characteristic of the instances causing this behavior in some of the algorithms. For instance, SPEA2 and

NSGA-II outperform most of the remaining EMO algorithms for instances having less than 90 decision variables. However, their performance decays for the biggest instances and IBEA and GWASF-GA are able to outperform them.

- SMS-EMOA is the most robust EMO algorithm with respect to both the dimensionality of the instances and the shapes of their feasible region and Pareto front. This suggests that the strategy used by the SMS-EMOA, which combines hypervolume maximization and a replacement strategy targeting those solutions that poorly contribute to the hypervolume of its respective fronts, is effective when dealing with optimization problems having these characteristics.

7. Practical implications and future directions

In view of our results, we conclude as final practical implications that the calibration of similar ABM models (i.e., high-dimensional models using a set of historical data values that the model is intended to reproduce) can be improved by using either MOEA/D or SMS-EMOA for tuning their parameters. Although the performance of NSGA-II is competitive for some of our model instances, we encourage practicioners and modelers to go beyond the use of the most popular EMO algorithm.

Future work will be focused on evaluating the possible improvement of including qualitative pattern features, which could be useful for minimizing the loss of information produced by the fitness functions [59, 58]. Since fitness functions like those employed in our study mainly focus on the distance between series of points, the aggregation of these values can potentially loose the shape of the series in the process. This issue can be solved in multiple ways. For example, the current fitness functions could be modified or additional objectives related to each of the model's output could be incorporated. Further research should clarify which alternative produces the best results. Apart from the use of qualitative patterns, other ABM consumer models may require the calibration of additional key performance indicators, such as the calibration of sales. Calibrating more outputs could be approached by including them as additional objectives, which defines a new scenario where the use of many-objective EMO algorithms will be required. Besides we believe that surrogate fitness functions would be useful for future studies due to the high computational costs of simulating multiple times for every evaluation of a single model configuration [29].

Acknowledgments

This work is supported by the Spanish Agencia Estatal de Investigación, the Andalusian Government, the University of Granada, and European Regional Development Funds (ERDF) under grants EXASOCO (PGC2018-101216-B-I00), SIMARK (P18-TP-4475), and AIMAR (A-TIC-284-UGR18). M. Chica is also supported through the Ramón y Cajal program (RYC-2016-19800). Authors thank the "Centro de Servicios de Informtica y Redes de Comunicaciones" (CSIRC) from the University of Granada for providing the computing resources (Alhambra supercomputer).

References

[1] Atamturktur, S., Liu, Z., Cogan, S., Juang, H., 2015. Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach. Structural and Multidisciplinary Optimization 51, 659–671.

[2] Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

[3] Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. Journal of Hydrology 341, 165–176.
[4] Beume, N., Naujoks, B., Emmerich, M., 2007. SMS-EMOA: Multiobjective selection based on dominated hypervolume. European Journal of Operational Research 181, 1653–1669.
[5] Bidgoli, A.A., Ebrahimpour-Komleh, H., Rahnamayan, S., 2020. An evolutionary decomposition-based multi-objective feature selection for multi-label classification. PeerJ Computer Science 6, e261.
[6] Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99, 7280–7287.
[7] Campomanes-Álvarez, B.R., Cordón, O., Damas, S., 2013. Evolutionary multi-objective optimization for mesh simplification of 3d open models. Integrated Computer-Aided Engineering 20, 375–390.
[8] Chica, M., Barranquero, J., Kajdanowicz, T., Cordón, O., Damas, S., 2017. Multimodal optimization: an effective framework for model calibration. Information Sciences 375, 79–97.
[9] Chica, M., Rand, W., 2017. Building agent-based decision support systems for word-of-mouth programs: A freemium application. Journal of Marketing Research 54, 752–767.
[10] Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A., et al., 2007. Evolutionary algorithms for solving multi-objective problems. volume 5. Springer.
[11] Confesor, R.B., Whittaker, G.W., 2007. Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and Pareto optimization. Journal of the American Water Resources Association 43, 981–989.
[12] Deb, K., 2001. Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Inc., New York, NY, USA.
[13] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197.
[14] Eichfelder, G., 2008. Adaptive Scalarization Methods in Multiobjective Optimization. Springer.
[15] Epstein, J.M., 2006. Generative social science: Studies in agent-based computational modeling. Princeton University Press.
[16] Falcón-Cardona, J.G., Coello, C.A., 2018. Towards a more general many-objective evolutionary optimizer using multi-indicator density estimation, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, Association for Computing Machinery, New York, NY, USA. pp. 1890–1893.
[17] Farhadi, S., Nikoo, M.R., Rakhshandehroo, G.R., Akhbari, M., Alizadeh, M.R., 2016. An agent-based-nash modeling framework for sustainable groundwater management: A case study. Agricultural Water Management 177, 348–358.
[18] Farmer, J.D., Foley, D., 2009. The economy needs agent-based modelling. Nature 460, 685–686.
[19] Fonseca, C.M., Fleming, P.J., 1996. On the performance assessment and comparison of stochastic multiobjective optimizers, in: Voigt, H.M., Ebeling, W., Rechenberg, I., Schwefel, H.P. (Eds.), Parallel Problem Solving from Nature — PPSN IV, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 584–593.
[20] García, S., Molina, D., Lozano, M., Herrera, F., 2008. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. Journal of Heuristics 15, 617–644.
[21] Gong, D., Xu, B., Zhang, Y., Guo, Y., Yang, S., 2020. A similarity-based cooperative co-evolutionary algorithm for dynamic interval multiobjective optimization problems. IEEE Transactions on Evolutionary Computation 24, 142–156.
[22] Gong, M., Li, H., Meng, D., Miao, Q., Liu, J., 2019. Decomposition-based evolutionary multiobjective optimization to self-paced learning. IEEE Transactions on Evolutionary Computation 23, 288–302.
[23] Guo, J., Zhou, J., Lu, J., Zou, Q., Zhang, H., Bi, S., 2014. Multi-objective optimization of empirical hydrological model for streamflow prediction. Journal of Hydrology 511, 242–253.
[24] Hernández Gómez, R., Coello, C.A., 2015. Improved metaheuristic based on the R2 indicator for many-objective optimization, in: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, Association for Computing Machinery, New York, NY, USA. pp. 679–686.
[25] Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688.
[26] Ishibuchi, H., Setoguchi, Y., Masuda, H., Nojima, Y., 2017. Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes. IEEE Transactions on Evolutionary Computation 21, 169–190.
[27] Janssen, M.A., Ostrom, E., 2006. Empirically based, agent-based models. Ecology and Society 11, 37.
[28] Kurek, W., Ostfeld, A., 2013. Multi-objective optimization of water quality, pumps operation, and storage sizing of water distribution systems. Journal of Environmental Management 115, 189–197.

[29] Lamperti, F., Roventini, A., Sani, A., 2018. Agent-based model calibration using machine learning surrogates. Journal of Economic Dynamics and Control 90, 366–389.
[30] Li, H., Zhang, Q., 2009. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. IEEE Transactions on Evolutionary Computation 13, 284–302.
[31] Libai, B., Muller, E., Peres, R., 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. Journal of Marketing Research 50, 161–176.
[32] Liu, Y., 2009. Automatic calibration of a rainfallrunoff model using a fast and elitist multi-objective particle swarm algorithm. Expert Systems with Applications 36, 9533–9538.
[33] Liu, Y., Sun, F., 2013. Parameter estimation of a pressure swing adsorption model for air separation using multi-objective optimisation and support vector regression model. Expert Systems with Applications 40, 4496–4502.
[34] Macal, C.M., North, M.J., 2005. Tutorial on agent-based modeling and simulation, in: Proceedings of the 37th conference on Winter simulation, ACM. pp. 2–15.
[35] Macdonald, E.K., Sharp, B.M., 2000. Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. Journal of Business Research 48, 5–15.
[36] Matsumoto, T., Masuyama, N., Nojima, Y., Ishibuchi, H., 2018. Performance comparison of multiobjective evolutionary algorithms on problems with partially different properties from popular test suites, in: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 769–774.
[37] Mavrotas, G., 2009. Effective implementation of the ϵ-constraint method in multi-objective mathematical programming problems. Applied Mathematics and Computation 213, 455–465.
[38] Meghwani, S.S., Thakur, M., 2017. Multi-criteria algorithms for portfolio optimization under practical constraints. Swarm and Evolutionary Computation 37, 104–125.
[39] Moya, I., Chica, M., Cordón, Ó., 2019. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization. Decision Support Systems 124, 113111.
[40] Moya, I., Chica, M., Sáez-Lozano, J.L., Cordón, Ó., 2017. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections. Knowledge-Based Systems 123, 200–216.
[41] Muraro, D., Dilo, R., 2013. A parallel multi-objective optimization algorithm for the calibration of mathematical models. Swarm and Evolutionary Computation 8, 13–25.
[42] Narzisi, G., Mysore, V., Mishra, B., 2006. Multi-objective evolutionary optimization of agent-based models: An application to emergency response planning, in: Proceedings of the 2nd IASTED International Conference on Computational Intelligence, CI 2006, pp. 224–230.
[43] Nebro, A.J., Durillo, J.J., Vergne, M., 2015. Redesigning the jMetal multi-objective optimization framework, in: Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, ACM, New York, NY, USA. pp. 1093–1100.
[44] Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. The Computer Journal 7, 308–313.
[45] Newman, M., Barabási, A.L., Watts, D.J., 2006. The structure and dynamics of networks. Princeton University Press.
[46] Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research 151, 552–568.
[47] Omran, S.M., El-Behaidy, W.H., Youssif, A.A.A., 2020. Decomposition based multi-objectives evolutionary algorithms challenges and circumvention, in: Arai, K., Kapoor, S., Bhatia, R. (Eds.), Intelligent Computing, Springer International Publishing, Cham. pp. 82–93.
[48] Rand, W., Rust, R.T., 2011. Agent-based modeling in marketing: Guidelines for rigor. International Journal of Research in Marketing 28, 181–193.
[49] Read, M.N., Alden, K., Rose, L.M., Timmis, J., 2016. Automated multi-objective calibration of biological agent-based simulations. Journal of The Royal Society Interface 13.
[50] Rong, M., Gong, D., Zhang, Y., Jin, Y., Pedrycz, W., 2019. Multidirectional prediction approach for dynamic multiobjective optimization problems. IEEE Transactions on Cybernetics 49, 3362–3374.
[51] Saborido, R., Ruiz, A.B., Luque, M., 2017. Global WASF-GA: An evolutionary algorithm in multiobjective optimization to approximate the whole Pareto optimal front. Evolutionary Computation 25, 309–349.
[52] Sargent, R.G., 2005. Verification and validation of simulation models, in: Proceedings of the 37th conference on Winter simulation, pp. 130–143.
[53] Stonedahl, F., Rand, W., 2014. When does simulated data match real data? Comparing model calibration functions using genetic algorithms, in: Advances in Computational Social Science. Springer, Japan. volume 11 of *Agent-Based Social Systems*, pp. 297–313.

[54] Snchez, L., Villar, J.R., 2008. Obtaining transparent models of chaotic systems with multi-objective simulated annealing algorithms. Information Sciences 178, 952–970.
[55] Waldrop, M.M., 2018. Free agents. Science 360, 144–147.
[56] Wu, F., Huberman, B.A., 2007. Novelty and collective attention. Proceedings of the National Academy of Sciences 104, 17599–17601.
[57] Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 599–608.
[58] Yücel, G., Barlas, Y., 2015. Pattern recognition for model testing, calibration, and behavior analysis. Analytical methods for dynamic modelers , 173–206.
[59] Ycel, G., Barlas, Y., 2011. Automated parameter specification in dynamic feedback models based on behavior pattern features. System Dynamics Review 27, 195–215.
[60] Zambrano-Vega, C., Nebro, A.J., García-Nieto, J., Aldana-Montes, J.F., 2017. Comparing multi-objective metaheuristics for solving a three-objective formulation of multiple sequence alignment. Progress in Artificial Intelligence 6, 195–210.
[61] Zhang, Q., Li, H., 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on Evolutionary Computation 11, 712–731.
[62] Zhang, X., Srinivasan, R., Liew, M.V., 2010. On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the swat model. Hydrological Processes 24, 955–969.
[63] Zhang, X., Srinivasan, R., Van Liew, M., 2008. Multi-site calibration of the SWAT model for hydrologic modeling. Transactions of the ASABE 51, 2039–2049.
[64] Zhang, Y., Shao, Q., Taylor, J.A., 2016. A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model. Journal of Hydrology 538, 802–816.
[65] Zitzler, E., Künzli, S., 2004. Indicator-based selection in multiobjective search, in: Parallel Problem Solving from Nature - PPSN VIII, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 832–842.
[66] Zitzler, E., Laumanns, M., Thiele, L., 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. TIK-report 103.
[67] Zitzler, E., Thiele, L., 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation 3, 257–271.
[68] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., 2003. Performance assessment of multiobjective optimizers: An analysis and review. IEEE Transactions on Evolutionary Computation 7, 117–132.

5 A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization

- I. Moya, M. Chica, O. Cordón. A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization, Decision Support Systems, vol. 124, 113111, 2019. DOI: 10.1016/j.dss.2019.113111.

  - State: Published.
  - Impact Factor (JCR 2019): 4.721.
  - Category: OPERATIONAL RESEARCH & MANAGEMENT SCIENCE. Order: 8/83. Q1, D1.

A multicriteria integral framework for agent-based model calibration using evolutionary multiobjective optimization and network-based visualization

Ignacio Moya[*,a], Manuel Chica[a,b], Óscar Cordón[a]

[a] *Andalusian Research Institute DaSCI "Data Science and Computational Intelligence", University of Granada, 18071 Granada, Spain*

[b] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia*

Abstract

Automated calibration methods are a common approach to agent-based model calibration as they can estimate those parameters which cannot be set because of the lack of information. The modeler requires to validate the model by checking the parameter values before the model can be used and this task is very challenging when the model considers two or more conflicting outputs. We propose a multicriteria integral framework to assist the modeler in the calibration and validation of agent-based models that combines evolutionary multiobjective optimization with network-based visualization, which we believe is the first integral approach to model calibration. On the one hand, evolutionary multiobjective optimization provides several sets of calibration solutions (i.e., parameter values) with different trade-offs for the considered objectives in a single run. On the other hand, network-based visualization is used to better understand the decision space and the set of solutions from the obtained Pareto set approximation. To illustrate our proposal, we face the calibration of three agent-based model examples for marketing which consider two conflicting criteria: the awareness of the brand and its word-of-mouth volume. The final analysis of the calibrated solutions shows how our proposed framework eases the analysis of Pareto sets with high cardinality and helps with the identification of flexible solutions (i.e., those having close values in the design space).

Keywords— Agent-based modeling, model calibration, evolutionary multiobjective optimization, information visualization.

1. Introduction

Model simulation is useful for representing and analyzing complex systems, but validating a model is not straightforward, specially if the modeling technique involves the definition and setting of many parameters. This is the case of agent-based modeling (ABM), a model simulation technique that has become highly relevant in the recent years [14, 43]. The ABM methodology [5, 13, 24] relies on a population of autonomous entities called agents which behave according to simple rules and by social interactions with other agents. The aggregation of these simple rules and interactions allows us to represent complex and emerging dynamics as well as to define what-if scenarios and

[*]Corresponding author

Email addresses: imoya@ugr.es (Ignacio Moya), manuelchica@ugr.es (Manuel Chica), ocordon@decsai.ugr.es (Óscar Cordón)

to forecast hypothetical scenarios [19]. However, creating and configuring a model for a specific problem from scratch can be difficult for designers and decision makers. If some of the model parameter values cannot be specified using the available information and knowledge, the modeler needs to manually estimate them for properly simulating the desired dynamics. The process of adjusting the values is known as the calibration of the model and it is a crucial step during the model validation [7, 8, 31].

A common calibration approach is automated calibration, a data-rich and computationally intensive process that compares real-world data to model outputs and tunes a set of model's parameters to match the data [31, 34]. Automated calibration requires a set of historical data, an error measure, and an optimization method for modifying the parameters in a systematic way by minimizing the error measure. However, after the application of the optimization method, the resulting parameter values need to be carefully reviewed and validated, since a good fitting of the historical data does not ensure the validity of the model. Additionally, typical parameters of computational models exhibit non-linear interactions and usually the best approach is to use a non-linear optimization algorithm such as metaheuristics [38] that can search across a large span of the model parameter space [7, 26, 37]. Metaheuristics are a family of approximate non-linear optimization techniques that provide high quality solutions in a reasonable time for solving complex problems in science and engineering [38]. In addition, often modelers design their models considering two or more conflicting criteria or key performance indicators (KPIs). In this scenario, the existence of an optimal solution is replaced by a set of Pareto-optimal solutions with the best trade-off between the different criteria. Therefore, calibrating those models involve a multicriteria decision making process since the modeler needs to select a model configuration between different solutions that satisfy the multiple criteria at different levels.

The validation of these models becomes even more difficult as there are multiple sets of parameter values that could be considered valid. In this paper, we propose to improve the validation of ABM systems and other discrete-event simulation models with multiple conflicting KPIs by introducing a multicriteria integral framework. This framework combines two elements: evolutionary multiobjective optimization (EMO) [10] algorithms for automatically calibrating the model parameters from historical data and an advanced visualization method that enhances the understanding of the calibration process and its results. To the best of our knowledge, ours is the first integral approach dealing with the calibration of simulation models from a multicriteria perspective with the aid of visualization tools. Our integrated approach shows that visualization is a key issue for automated calibration as it increases the understanding of the calibrated model, assisting the designer on the model validation [31, 34]. This way the modeler can adapt the automatic calibration process to consider the mentioned conflicting outputs replacing the underlying optimization algorithm by a multiobjective metaheuristic. Although there are previous efforts using EMO for calibrating ABMs [28, 32], none of them considered an integral framework incorporating novel visualization methods for easing the validation of the calibrated models.

We propose the use of *moGrams* [40] to represent the set of calibration results. It is a methodology that combines the visualization of non-dominated solutions in both the design and the objective spaces. A moGram is a weighted network where the nodes represent the solutions of a Pareto set approximation and each edge represents a similarity relationship between two solutions in the design space. By using moGrams, the modeler will improve her/his understanding of the calibration problem by being able to identify clusters of solutions, to detect the most flexible ones (i.e., those that can be exchanged with another solution with minimum changes in their decision variables),

and to conveniently validate their selection of parameter values based on the relationships between solutions (i.e., model parameter configurations).

Two different experimental setups have been designed to validate our proposal. First, we introduce a controlled experiment where we show how the framework properly identifies the optimal values in a design environment. Second, we present its application to a practical problem calibrating two instances of a real banking marketing scenario which considers real data and different dimensionality (i.e., number of decision variables). In both cases we consider an ABM for marketing modeling two conflicting criteria: brand awareness and word-of-mouth volume. Marketing and word-of-mouth programs are one of the fields with the greatest number of ABM applications [9], although our framework is not restricted to this area nor any specific modeling technique.

NSGA-II [12], the most extended EMO algorithm, is used in our experimentation although the framework allows us to consider any EMO algorithm. Finally, the behavior of the resulting solutions for the practical problem is analyzed by visualizing their parameter settings using moGrams, which allows us to uncover additional insights about the problem decision space and help with the validation process.

The structure of this paper is as follows. We introduce our approach for multiobjective model calibration in Section 2. Then, we present the used ABM for marketing in Section 3. We describe our experimentation and its results in Section 4, including the analysis and validation of the calibrated solutions. Finally, in Section 5 we discuss our conclusions and final remarks.

2. Multicriteria integral framework for model calibration

In this section we describe our integral multicriteria framework for model calibration using EMO algorithms and network-based visualization. A diagram illustrating the components of our framework is shown in Figure 1. We start by presenting how we handle the problem objectives and the parameters considered during calibration in Section 2.1. Then, Section 2.2 explains the role of the EMO algorithm, which is a core component of the framework. Section 2.3 elaborates the importance of visualization for model calibration and validation. Finally, Section 2.4 introduces moGrams, the selected visualization methodology to analyze and better understand the calibrated solutions, which is the other core component of the proposed framework.

2.1. Calibration objectives and parameters

In an automated calibration process, the values of the model parameters are adjusted to match the model outputs with the data reality of the modeled scenario. We define each parameter configuration $X = (x_1, ..., x_n)$ as a vector of n decision variables. The modeler should carefully select the model parameters that will be estimated by automated calibration, since the difficulty of validating the calibrated configurations increases with the number of calibrated parameters. On the one hand, the modeler should consider those parameters being the most uncertain and the hardest to define by her/him according to the available information. On the other hand, sensitive parameters should also be considered for calibration since small changes in their values can significantly affect the model's response and the global output. Our approach considers the calibration of parameters using either integer or real values.

In our multiobjective approach, we assess the quality of a given model configuration regarding two or more conflicting criteria which are considered as calibration objectives. We evaluate the quality of the model regarding the fitting of its output to the provided real, historical data for the multiple defined KPIs. In order to avoid over-fitting in the resulting calibrated parameters,

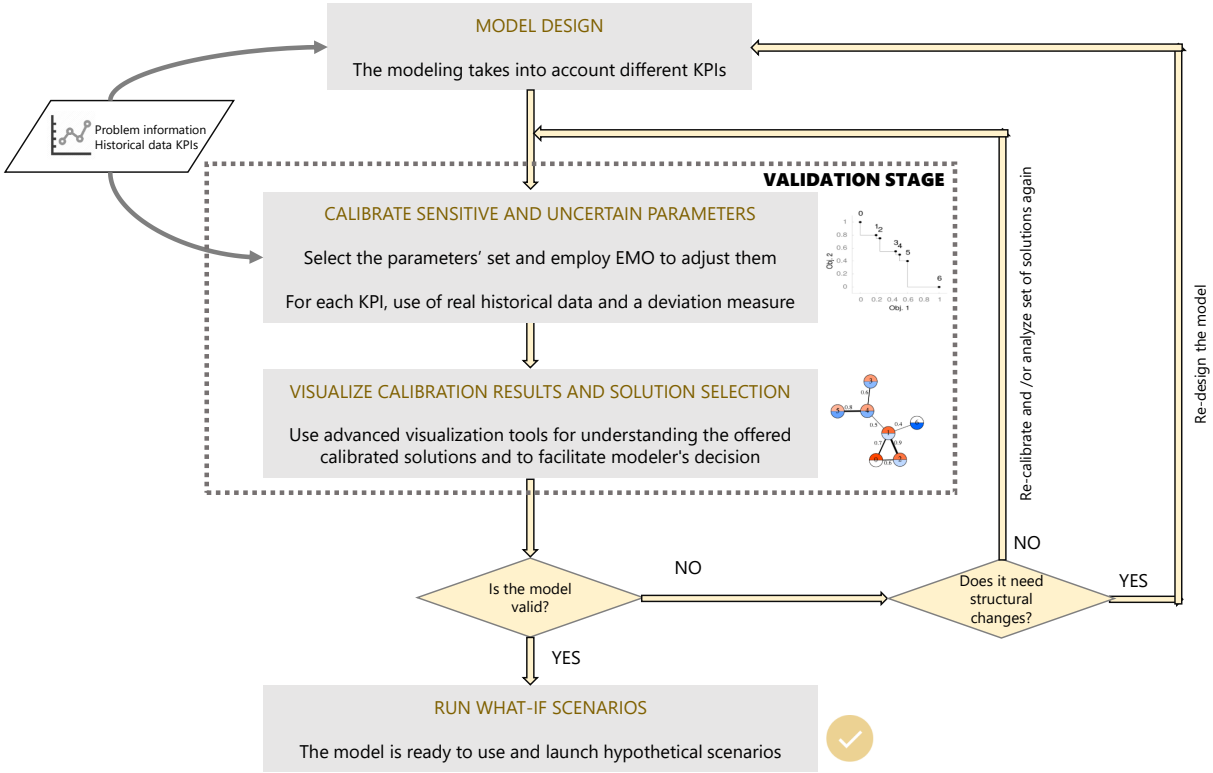


Figure 1: Diagram illustrating the components and the flow of our multicriteria integral framework for model calibration.

the historical data can be split into training and hold-out. Hold-out sets are encouraged to be considered when having real data for calibration purposes in order to check if the model is general enough and to avoid over-fitting (see guidelines when building agent-based decision support systems in [9]). These hold-out sets are then useful for testing the results of the calibration procedure and improving the model confidence. However, since normally ABM implementations are employed as explanatory models instead of forecasting ones [17], the use of hold-out sets can be unnecessary in some cases. The fitting over the training historical data is computed using a deviation measure ϵ that guides the optimization algorithm calculating the error between the provided ground-truth data and the model output for a given objective. This distance can be computed using any of the standard deviation measures (RMSE, MAPE, or MARE [18], for instance). The modeler should choose the most appropriate for each objective. The goal of the optimization algorithm is to minimize $F(X) = (f_1(X), ..., f_M(X))$, where M represents the number of objectives. Each fitness function $f_j(X)$ computes the error associated to objective j, and is defined by Equation 1, where $\widetilde{o}^j$ represents the target ground-truth values for the j-th output and $o^j(X)$ represents the simulated values of the model using the parameter configuration X. Note that ϵ^j is independent for each objective $j \in [1, M]$ and different deviation measures can be used. Our framework does not impose any kind of restriction on that issue.

$$\min F(X) = \min (f_1(X), ..., f_j(X), ..., f_M(X)), \text{ where } f_j(X) = \epsilon^j(o^j(X), \widetilde{o}^j). \tag{1}$$

2.2. EMO algorithm

The EMO algorithm is a core component of the framework. EMO algorithms are population-based metaheuristics that represent the solutions of the problem as individuals of a population. They can provide different model configurations (i.e., sets of parameter values) in a single run. In our design for the calibration problem, each individual of the population has n genes that corresponds with the n decision variables that represent each model configuration, with these genes being either real-coded or integer-coded values. The model configurations obtained by EMO algorithms have different values in the objective space and comprise a Pareto-optimal set approximation.

Any EMO algorithm can be selected for performing this process and should be chosen depending on the characteristics of the model being calibrated. For instance, if the calibration problem considers less than four objectives, any well-known EMO algorithm such as NSGA-II [12], SPEA2 [53], or MOEA/D [21] can be considered. Otherwise, the calibration problem should be treated as a many-objective optimization problem and the selected EMO algorithm should perform properly in this environment. Some examples of EMO algorithms for many-objective optimization would be NSGA-III [11], HypE [2], GrEA [48], or KnEA [51]. In addition, the modeler should consider the use of additional tools and methods for ensuring the good performance of the selected EMO algorithm. For instance, the modeler could use approaches such as irace [23] for automatically adjusting the EMO parameters. However, this decision should be addressed by the modeler since it depends on the specific search space of the defined calibration problem.

2.3. Visualization for model calibration

The effective use of an ABM simulation for representing a complex system heavily relies on the transparency of the underlying model. ABM modelers and stakeholders require to understand how the model recreates a given behavior, since the simulation of ABMs is frequently used for defining *what-if* scenarios and forecast hypothetical scenarios [19, 42]. This can be achieved from

a white-box perspective [33], where both modelers and stakeholders can make use of visualization tools for increasing the *explainability* of the model.

Improving the understanding of artificial intelligence-based models is one of the goals of the emerging area of explainable artificial intelligence [33]. It encourages modelers and researchers to *open* black-box models so their behavior can be easily understood and their output can be better explained. Explainable artificial intelligence also empowers the solutions delivered by white-box models, since boosting the transparency of the delivered solutions should increase the trust in the behavior and performance of these solutions.

This highlights the role of visualization methods for model calibration, since they are powerful tools that increase the understanding of the modeler on the calibrated model and its parameter settings [7]. The use of visualization increases the transparency of the quality indicators (mostly focused on the fitting of the model to real data) for the validation of the calibrated model [4, 7, 20]. Thus, visually showing the underlying relationships between an input configuration and its corresponding model output becomes a critical component of the validation process.

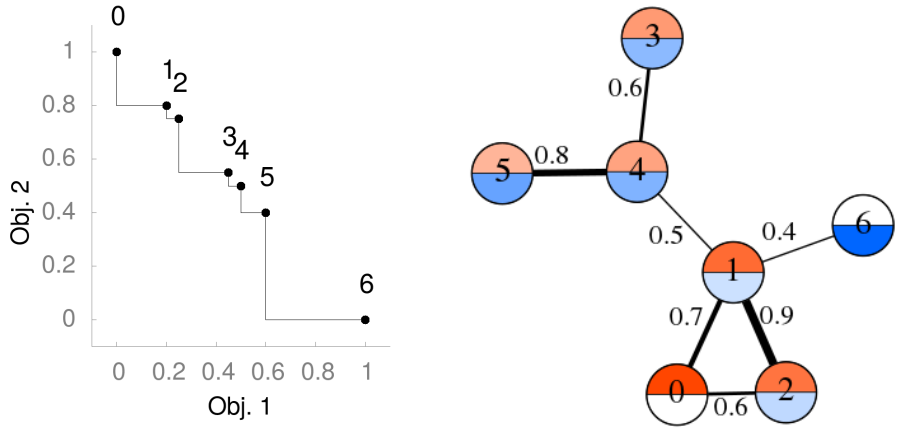
When the model considers two or more conflicting outputs, multiobjective visualization methods are specifically required for the validation of the model. Most of the available contributions in the literature focus on the visualization of the non-dominated solutions in the objective space [41, 44]. In contrast, only a few contributions tackle the visualization of the solutions in the design space, which is the most interesting for discovering knowledge about the parameter values, and even less proposals derive joint visualizations for both the objective and design spaces [40]. One of the few existing approaches of such kind is the *moGrams* methodology [40], which mutually analyzes and visualizes the solutions obtained by EMO algorithms in the decision and objective spaces.

2.4. moGrams

moGrams [40] represents the non-dominated solutions in a Pareto set approximation as nodes in a weighted network where each edge stands for a relationship between the connected solutions in the design space. The weights of the edges are computed using a similarity metric specifically defined for each problem by the designer. In order to improve the readability of the network, moGrams employs the Pathfinder network pruning algorithm [36] for reducing the edges of the network leaving only the most relevant ones from a global viewpoint. In addition, each node is divided into sectors of the same size, each of them associated to a different objective, which are colored differently with its opacity proportional to the quality of the solution for the respective objective. For example, if a problem considers four objectives, the node is divided into four sectors with different colors. This way, the modeler has access to the whole information of both the objective and the design spaces at the same time. For our problem (i.e., the validation of a model) it provides additional information regarding the parameter settings of the different calibrated model configurations, highlighting similarities between them.

Figure 2 shows an example of the moGram generated for the Pareto set approximation of a problem with two objectives. The Pareto front approximation obtained for the generated moGrams network is also shown, since the joint visualization of both elements is suggested for better understanding the relationships in the design space. In this network example, we can identify two separate clusters connected by the edge between Solution 4 and Solution 1, with the latter being the most connected solution. From the neighbors of Solution 1, we can observe that Solution 2 has the highest similarity relationship. This means that the parameter configuration of Solution 2 is highly similar to the one of Solution 1.

However, the high similarity between Solutions 1 and 2 could lead us to think that both solutions are close in the objective space. In this regard, moGrams provides the user with other relationships that are not intuitive, such as the relationship between Solution 1 and Solution 6. This relationship reveals that the closer configuration to Solution 6 is the one defined by Solution 1, which is located at the other end of the Pareto front. Thus, the decision maker can detect parameters that drastically change the behavior of the solutions using this relationship. In addition, the topology of the moGrams network allows the modeler to identify Solution 1 as the most flexible solution, which can be swapped with other solutions with minimum changes on its decision variables.



(a) Pareto front. (b) moGram representation.

Figure 2: Generated moGram network example for a given Pareto set approximation corresponding to a problem with two objectives.

3. Description of the agent-based model for marketing

This section describes the main features of the considered ABM for marketing scenarios. First, the general structure of the model and behavior of the agents are presented in Section 3.1. Then, Section 3.2 introduces the artificial social network and its features. Section 3.3 presents how the mass media channels are modeled. Finally, in Section 3.4 the parameters of the model selected for calibration are summarized and the conflict of adjusting both KPIs of the model with their corresponding fitting functions is described in Section 3.5.

3.1. ABM general structure and agent's state and update rule

Our proposed model considers a terminating simulation with T weeks of a market that comprises a set of brands B. Using a time-step of a week, the model simulates the behavior of N agents and their reaction to social influence through a social network in a word-of-mouth (WOM) process; and external influences (advertisement) through a set of C mass media channels. The model has two main outputs or KPIs: brand awareness and WOM volume (i.e., the number of WOM interactions among the consumers). We select these KPIs because they have an important role in market expansion [22, 25]. A general scheme of our model is presented in Figure 3.

The awareness values of the agents are modeled using a state variable called $a_i^b \in \{0, 1\}$. If $a_i^b(t) = 1$, then the agent i is aware of brand b at time step t. Otherwise, the agent does not

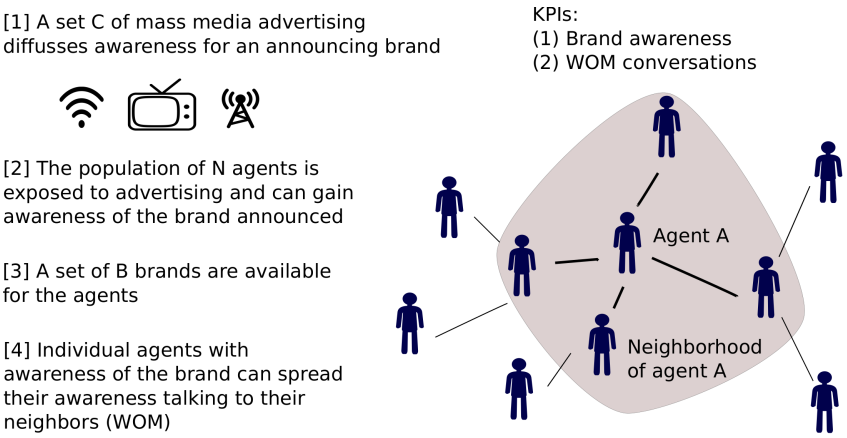


Figure 3: General scheme and structure of the ABM with an example of a brand advertised using mass media. The agents exposed to advertising can gain awareness of the brand announced and talk about it to their neighbors.

have awareness of brand b. This state variable is initialized using an initial awareness parameter set for each brand ($a^b(0) \in [0,1]$) which is the global awareness of the population and fulfills $a^b(0) = \frac{1}{N}\sum_{i=1}^{N} a_i^b(0)$. Therefore, the initial awareness parameter for each brand specifies the percentage of agents that have awareness of that brand at the beginning of the simulation. This process is carried out during the initialization of the agents, where they activate their awareness of each brand b with probability $a^b(0)$.

The awareness values of the agents do not remain static but change during the simulation: agents may loose or gain awareness of any brand at each step of the simulation. On the one hand, agents may gain awareness of a brand due to advertising or due to interacting with other agents through a WOM diffusion process. On the other hand, if the awareness of a brand is not reinforced, it may be lost over time because of a deactivation process [46, 47].

We model these losing/gaining effects with additional parameters. The parameter regulating the rate at which awareness is lost over time is called awareness deactivation probability ($d \in [0,1]$). This parameter is modeled as follows. At the beginning of each step t, the agent i checks all the brands the agent is aware of ($a_i^b(t) = 1, \forall b \in B$). Each of these awareness values will be deactivated with a probability d by setting $a_i^b(t) = 0$. If the deactivation takes effect, the agent could still re-gain awareness due to the WOM diffusion and/or the mass media channels during the same simulation step, but it will not check for deactivation until the next simulation step. The modeling processes of the awareness obtained by WOM diffusion and mass media channels are explained in detail in Sections 3.2 and 3.3, respectively.

In addition, each agent stores the number of conversations produced during its diffusion process (depicted in Section 3.2) in order to compute the WOM volume for each brand ($\omega_i^b(t)$). This way, every time an agent starts a diffusion process and talks with its neighborhood, the variable $\omega_i^b(t)$ will be updated by incrementing it with the number of agents' neighbors (i.e., conversations). Finally, it will update the global $\omega^b(t)$ variable for the respective brand and time step.

3.2. Social network of agents and their word-of-mouth interactions

Our agents populate an artificial social network [3, 45]. We model this social network using an artificial scale-free network [3] because the most real networks match with this network model [3, 30]. In these kinds of networks, the degree distribution follows a power law [3]. This means

that few nodes have a significantly large number of connections (hubs of the social network) and most nodes have a very low number of connections. We generate our scale-free network using the Barabasi-Albert preferential attachment algorithm [3]. This algorithm has a main parameter m which regulates the network growth rate and its final density. The generation process starts with a small clique (a completely connected network) with m_0 nodes. At each generation step, a new node is added and connected to m different existing nodes. When a new node is included, the probability of choosing an existing node is proportional to its degree (preferential attachment). After t iterations, the Barabasi-Albert algorithm results in a social network with $m \cdot t$ edges. Finally, the average degree of the social network is $\langle k \rangle = 2 \cdot m$.

The agents of the model can spread their awareness values during the simulation through the artificial social network. We model this social interaction as a contagion process which allows information diffusion through the nodes of the social network depending on their connectivity [30, 35, 49, 50]. Every agent i has a talking probability ($p(t)_i^b \in [0, 1]$) to spread the brands it is aware of at time step t (i.e., for every brand b where $a_i^b = 1$). This probability p_i^b specifies when the agent i talks with all of its neighbors in the artificial social network, having the chance of transferring its awareness (i.e., a contagion process). We model this contagion effect using a parameter called WOM awareness impact ($\alpha^{WOM} \in [0, 1]$), which represents the probability for an agent in the neighborhood to be aware of a brand after having a conversation about it.

3.3. Modeling mass media channels

We model external influences like brand advertising as global mass media [16] using a similar approach to the one applied in the social network. The external influences are parameterized to define the differences between the channels (i.e., press, radio, and television). The set of C mass media channels can influence any number of agents at random depending on the channel potential for reaching the population and the investment of each brand.

The maximum population percentage that can be reached by a mass media channel is bounded by the nature of the channel itself. In this sense, some media are able to reach more people than others. For example, the maximum population percentage that can be reached by a campaign scheduled in the radio is bounded by the maximum population percentage that listens to the radio. We model this behavior with a reach parameter ($r_c \in [0, 1], \forall c \in C$), which defines the maximum number of people a channel c is able to hit during a single step.

The advertising campaigns of the mass media channels are modeled using gross rating points (GRPs). In advertising [15], a GRP is a measure of the magnitude of the impressions scheduled for a mass media channel. Specifically, we use the convention that one GRP means reaching 1% of the target population. The investment units in GRPs for channel c by brand b and time step t is modeled by the variable $\chi_c^b(t)$. Each channel has different costs for GRP and the brands need to carefully choose their investment since increasing the population awareness using mass media channels implies a monetary cost. Using both the supplied GRPs for a given brand and the reach values for a mass media channel, we are able to model brand advertising. Algorithm 1 shows the scheduling algorithm for modeling impacts of the media channels over the population.

Each mass media channel has an awareness impact parameter ($\alpha_c \in [0, 1], \forall c \in C$) that defines the probability of the agent becoming aware of the brand after one impact. If the agent is not aware of the brand at a given time step t ($a_i^b(t) = 0$), this probability α_c will activate the awareness of the agent for brand b.

Moreover, the advertising transmitted by mass media channels can produce a viral buzz effect in the reached agent, as done in [27]. This buzz effect increases the number of conversations about

Algorithm 1: Pseudo-code of the advertising scheduling of the model for a given brand, time step and channel.

```
begin
    reach_step = 0;
    total_hits = χ_c^b(t) · 0.01 · N;
    reach_increment = 1 / N ;
    i = 0;
    while i < total_hits do
        select agent randomly;
        if selected agent was already hit then
            impact agent;
            i++;
        else if reach_step + reach_increment ≤ r_c then
            impact agent;
            i++;
            reach_step += reach_increment;
```

the announced brand, modifying the talking probability (p_i^b) of the reached agents. We model this effect through a variable called buzz increment (τ_c) for each channel $c \in C$. This increment of the agents talking probability is computed as a percentage increment over the initial talking probability ($p_i^b(0)$) of the agent. However, if the generated buzz is not reinforced, its effect could decay over time as previous interactions are forgotten. We model this effect with a variable called buzz decay ($d\tau_c$). The action of buzz decay reduces the previous increment of talking probability (σ_c) applied to the agent through channel c. The update process for the talking probability value of agent i for brand b due to both buzz increment and decay effects of channel c is shown in Equation 2.

$$p_i^b(t+1) = p_i^b(t) - \sigma_{i\,c}^b(t) \cdot d\tau_c + p_i^b(0) \cdot \tau_c, \text{ where } \sigma_{i\,c}^b(t) = \sum_{i=1}^{t} (p_i^b(t) - p_i^b(0) \cdot \tau_c). \tag{2}$$

3.4. Parameters selected for calibration

A summary of the complete set of model parameters is listed in Table 1. From those, we select for calibration the parameters that either modify the agent awareness values or their number of conversations for the automated calibration process since they are the most uncertain and the hardest to estimate. These parameters regulate the awareness and talking probability gained by mass media and social interactions with the addition of the awareness deactivation probability (d) and the social network generation parameter (m). The range of possible parameter values during the calibration process is limited to $[0, 1]$ for the real-coded parameters and to $\{2, ..., 8\}$ for the social network generation parameter (m), which is the only integer-coded parameter of the model. Notice that, the density of the agents social network is a parameter that is always difficult to identify. This way, each of the selected model calibration parameters corresponds with one decision variable in the coding scheme of the EMO algorithm. The final set of parameters to be calibrated for each model instance is determined by the size of the modeled scenario: three parameters for each mass media channel plus four fixed social parameters. Briefly, those parameters are the following:

Name	Description
N	Number of agents running in the model
$\|B\|$	Number of brands contained in the model
$\|C\|$	Number of mass media channels in the model
T	Number of steps of the model
$a^b(0)$	Initial awareness for brand b
d	Awareness deactivation probability in the model
m	Parameter for social network generator
$p_i^b(0)$	Initial talking probability, same value for each brand b
α^{WOM}	Awareness impact for social interactions
χ_c^b	GRP units invested by brand b in channel c
r_c	Reach for mass media channel c
α_c	Awareness impact for mass media channel c
τ_c	Buzz increment for mass media channel c
$d\tau_c$	Buzz decay for mass media channel c

Table 1: List of parameters of our proposed marketing model.

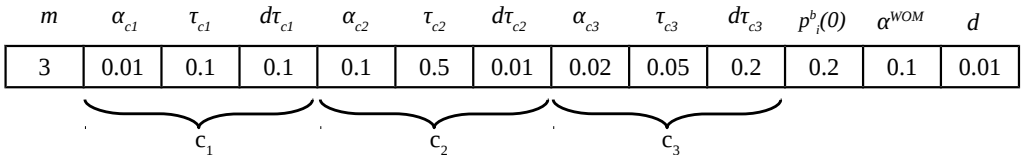


Figure 4: Coding scheme for a model instance with three mass media channels. The first gene (m parameter) is an integer value bounded to $\{2, ..., 8\}$. The rest of genes in the chromosome represent the real-coded parameters and are defined in $[0, 1]$.

- Social network parameters. We calibrate the initial talking probability ($p_i^b(0)$), social awareness impact (α^{WOM}), awareness deactivation probability (d), and social network generation parameter (m).
- Mass media parameters. For each defined mass media channel $c \in C$, we calibrate its awareness impact (α_c), buzz increment (τ_c), and buzz decay ($d\tau_c$).

Therefore, the number of calibrated parameters is $3 \cdot |C| + 4$. Figure 4 shows the coding scheme for a model instance using three mass media channels, that is, composed of 13 genes.

3.5. KPI fitting functions

Equations 3 and 4 define the objective fitting functions for the two KPIs, f_1 (awareness deviation error) and f_2 (WOM volume deviation error), respectively. These functions compute the deviation error between the provided series of target data and the model outputs for each objective using the

standard mean absolute percentage error (MAPE) function, where $\widetilde{a}$ and $\widetilde{\omega}$ represent the ground-truth target values of awareness and WOM volume of the whole population, respectively. The simulated values are generated using Monte-Carlo simulations by computing the average of those independent runs.

$$f_1 = \frac{100}{T \cdot |B|} \sum_{b=1}^{|B|} \sum_{t=1}^{T} \left| \frac{a^b(t) - \widetilde{a}^b(t)}{\widetilde{a}^b(t)} \right|, \tag{3}$$

$$f_2 = \frac{100}{T \cdot |B|} \sum_{b=1}^{|B|} \sum_{t=1}^{T} \left| \frac{\omega^b(t) - \widetilde{\omega}^b(t)}{\widetilde{\omega}^b(t)} \right|. \tag{4}$$

We can observe that both outputs of the model are correlated. Thus, it is not advisable to calibrate the outputs independently because modifying the parameters of the model implies changes for both outputs. On the one hand, mass media channels activate brand awareness using advertisement but this effect also modifies the WOM volume generated via the buzz effect of the campaign, which can deviate its value beyond the target data. Additionally, as WOM spreads brand awareness via the agent interactions through the social network, adjusting the number of conversations will modify the awareness of the population, deviating its value beyond the target data. Therefore, for most model instances there will be no model configuration able to satisfy the fitting of both KPIs to the historical data at the same time.

Finally, a sampling of the decision space is shown in Figure 5. These values were obtained generating 10,000 random calibration solutions for a given instance of the model and plotting their fitting for both KPIs. In this figure, we can visually confirm the fact that both optimization objectives are in conflict, since the solutions that have the lowest deviations for awareness (f_1) have the highest deviations for WOM volume (f_2) and *vice versa*. This can be observed in the extreme locations of the decision space, where the greater concentration of sampled values are located. In addition, several sampled solutions scatter diagonally from one concentration of solutions to the other, showing the difficulty of balancing the optimization of these two criteria.

4. Experimentation

This section presents the experimentation developed and the analysis of results. Sections 4.1 and 4.2 explain the experimental setup, describing the considered problem instances and the algorithmic configuration. Then, Sections 4.3 and 4.4 present the application of our framework both to a designed environment and to real calibration scenarios. Finally, Section 4.5 illustrates the use of the visualization method on the calibration results and analyzes the composition of the non-dominated solutions from the decision maker's point of view.

4.1. Experimental setup

Our experimentation considers three different model instances. These instances are generated starting from an initial *baseline* instance (called P-25 because of its 25 decision variables), which is based on real data from a banking marketing scenario. Then, we synthetically generate two additional instances: a simplified instance for conducting a controlled experiment and an additional instance with increased dimensionality. The simplified instance considers a reduced set of five model parameters, which are selected because of their sensitivity on the model's behavior. The instance with increased dimensionality includes 10 additional mass media channels (which are generated

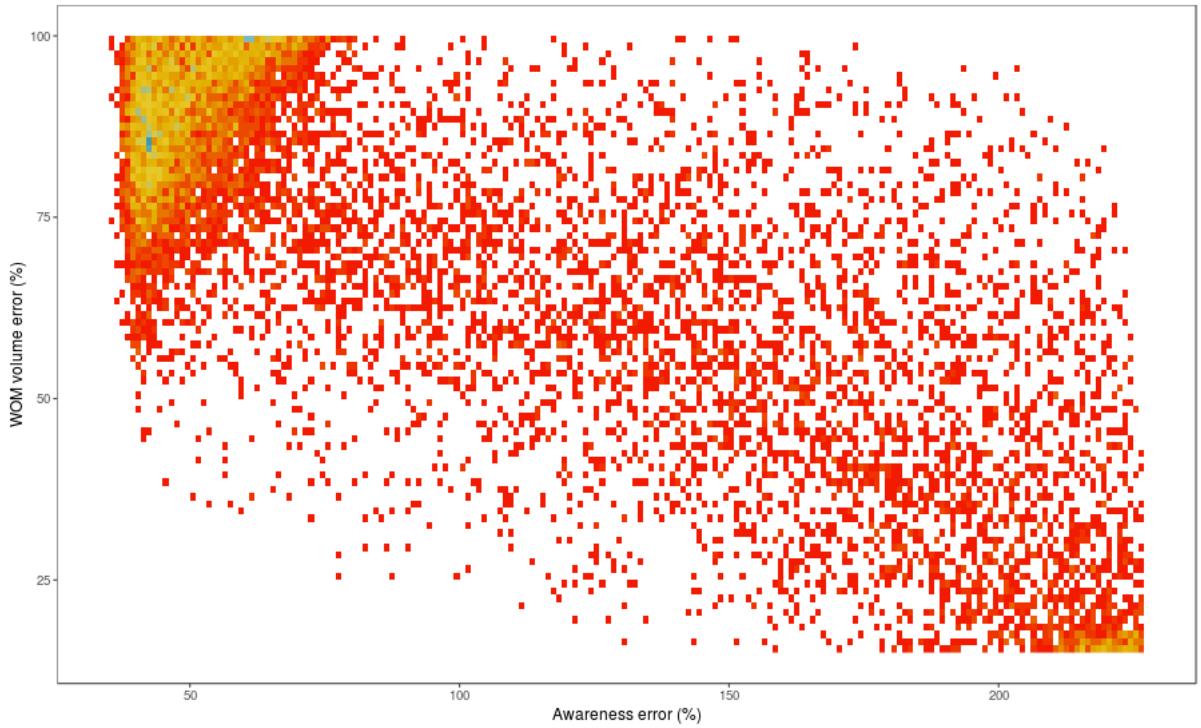


Figure 5: Sampling of the decision space for the two objectives. The values are shown as a scatter plot, where locations gathering multiple individuals have more intense colors.

perturbing the GRP investment of the existing ones) and modifications of the target series of historical values for both objectives: awareness and WOM volume. Thus, P-55 will increase the dimensionality of the baseline real-world instance by adding 30 new decision variables to enable a more complete analysis of the algorithms' performance.

Table 2 shows the parameters' values that are manually set using data for the baseline instance P-25. This instance belongs to a real marketing scenario of banking in Spain. These parameters define the initial conditions of our terminating simulation, such as the delimitation of the market (number of brands B, number of channels C, and total number of agents N) and the initial awareness conditions. Each of these agents initially activate their awareness for each brand b with probability $a^b(0)$ which is an input parameter of the model. These awareness values are taken from a brand tracking study used by one of the competing brands. In addition, all the media channels are initialized using their input parameters (i.e., mass media parameters either manually set or automatically calibrated). For instance, we set the values for the reach parameters from a mass media study in Spain[1]. Notice that our terminating simulation takes $T = 52$ time steps, matching the weeks of a year of simulation. This period is selected because of the historical data, which considers the behavior of the model KPIs during one year.

The generated P-55 instance shares the previous parameter configuration with the addition of the corresponding reach parameter r_c for the additional mass media channels. The reach parameter values of these new mass media channels take the value of the original ones employed

[1] `https://es.slideshare.net/ZenithES/libro-de-medios-2015`

Baseline instance P-25									
Name	**Value**	**Name**	**Value**	**Name**	**Value**	**Name**	**Value**	**Name**	**Value**
N	1000	$\|B\|$	8	$\|C\|$	7	T	52	$a^{b_1}(0)$	0.709
$a^{b_2}(0)$	0.757	$a^{b_3}(0)$	0.589	$a^{b_4}(0)$	0.2559	$a^{b_5}(0)$	0.081	$a^{b_6}(0)$	0.429
$a^{b_7}(0)$	0.395	$a^{b_8}(0)$	0.34	r_{c_1}	0.928	r_{c_2}	0.579	r_{c_3}	0.548
r_{c_4}	0.035	r_{c_5}	0.432	r_{c_6}	0.382	r_{c_7}	0.696		

Table 2: Summary of the configuration of the baseline model instance P-25 for the parameters that are manually set. These values are set from the real available data.

for its generation. For example, if the additional mass media channel c_9 was generated from the original channel c_5, then $r_{c_9} = r_{c_5}$. The generated simplified instance shares the reach parameter configuration with P-25. Those parameters that were not selected for calibration are set to 0 to avoid adding more complexity in the controlled experiment. Finally, the model's historical target values are generated using the output of two known parameter configurations for each objective, which are identified as the best values for the designed environment.

4.2. NSGA-II as the EMO algorithm

We select NSGA-II as our EMO algorithm during our experiments. NSGA-II [12] has become one of the most well-known EMO algorithms and there are several applications in model calibration [1, 32, 52]. It has been proven computationally fast while maintaining good levels of diversity by using a search strategy based on non-dominated sorting in problems with 2 or 3 objectives. During each generation, NSGA-II creates an offspring population Q_t from the previous parents population P_t. These two sets are joined into a temporary population R_t of size $2 \cdot |P|$ ranking every solution according to its non-dominance level, that is, how many solutions it is dominated by. The new population P_{t+1} is created including the solutions with the best rank, which belongs to the best non-dominated front. Then, the solutions from the following ranks are included iteratively until $|P|$ individuals are selected. This way the algorithm is guided to non-dominated regions and the solutions from the best non-dominated front are always kept in the population. In order to enhance diversity, the first non-dominated front that cannot be fully included in P_{t+1} is filtered using a crowding distance calculated for only including the most diverse individuals.

We include simulated binary crossover (SBX) [10] as our crossover strategy, which is applied with probability $p_c = 1.0$. SBX performs the crossover operation on real-coded decision variables emulating the behavior of a single-point crossover from binary-encoding. This operator works as follows: given two parents $P_1 = (p_{11}, ..., p_{1n})$ and $P_2 = (p_{21}, ..., p_{2n})$, SBX generates two springs $C_1 = (c_{11}, ..., c_{1n})$ and $C_2 = (c_{21}, ..., c_{2n})$ as $c_{1i} = \bar{X} - \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$ and $c_{2i} = \bar{X} + \frac{1}{2} \cdot \bar{\beta} \cdot (p_{2i} - p_{1i})$, where $\bar{X} = \frac{1}{2}(p_{1i} + p_{2i})$ and $\bar{\beta}$ is the spread factor, a random number fetched from a probability distribution generated using a given distribution index. Regarding the mutation strategy, we choose polynomial mutation [10]. It modifies the values of an individual genes using a polynomial distribution. It is applied with probability $p_m = 1/n$, where n is the number of decision variables (i.e., parameters of the model to be calibrated).

Since EMO algorithms are stochastic algorithms, they must be ran several times for ensuring a good performance for any optimization problem they are applied to. Therefore, we run NSGA-II 20 times using different seeds for each run. The NSGA-II has a population of 100 individuals ($|P| = 100$) and evolves during 100 generations using 10,000 evaluations as stopping criteria. In

addition, the first 39 steps of the historical data (i.e., 75% of total) are used for training, leaving the remaining values as hold-out. Each evaluation of an individual is a Monte-Carlo simulation of 15 runs. This value showed an acceptable trade-off between the stability of the model's output and the required runtime. The mutation operator uses a distribution index value of 10 and a different mutation probability value depending on the number of parameters of each model instance. We implemented NSGA-II using the jMetal framework [29].

4.3. Controlled experiment in a designed environment

We define a controlled experiment by employing an instance with a reduced set of sensitive parameters. Therefore, this calibration scenario defines a controlled environment where the optimal parameters values for the defined historical data for each objective are known. The selected parameters are α_{c_1}, α_{c_2}, α_{c_3}, α_{c_5}, and α_{c_7}. The latter channels are those with the highest investment in the model and can provide changes into the model's response without being affected by other parameter values (e.g., α_{c_4} and α_{c_6} belong to mass media channels with lower investment and thus are excluded from the experiment). By selecting these parameters we see if the calibration algorithm can find solutions to perfectly match each KPI in isolation or a trade-off combination: when the parameters are set to 1, the model perfectly matches the historical awareness values; and when the parameters are set to 0, the model perfectly matches the historical WOM volume values. Therefore, this controlled experiment will test if the calibration process is able to search across the whole solution space by setting the values of the parameters to their limits.

Figure 6 shows the resulting attainment surface [54] of applying our framework to the designed instance. Attainment surfaces are surfaces uniquely determined by a set of non-dominated points that separates the region of the objective space dominated by the set from the region not dominated by it [54]. The surface of Figure 6 is obtained by merging all the obtained Pareto set approximation from every execution of NSGA-II and by removing the dominated solutions. It shows how the framework identifies the best result for each of the objectives (i.e., the parameter configuration obtaining a perfect match), which are represented by red boxes in Figure 6. Additionally, the resulting attainment surface provides interesting insights regarding the behavior of the WOM volume dynamics in the model. For example, there are several solutions at the right end of the attainment surface that achieve similar/equal awareness error while decreasing WOM error. This can be explained by the sensitivity of WOM volume dynamics to Monte-Carlo variability.

This simple experiment allows us to properly validate the performance of the first stage of our framework. On the one hand, it shows how the EMO algorithm is actually able to obtain the optimal parameter values for each considered KPI in the specific model calibration instance. On the other hand, it also illustrates the capability of the EMO algorithm to generate a Pareto set composed of a large number of parameter configurations showing different trade-offs between the two KPIs being optimized.

4.4. Application to a practical problem

We begin the analysis and validation of the model instances P-25 and P-55 by visualizing the solutions of the generated Pareto set approximations with respect to the two conflicting objectives (see Figure 7 with the two Pareto front approximations). We have selected three of the most representative solutions for the two instances: a) the solution with lowest awareness error, b) the solution with lowest WOM volume error, and c) the solution with the best trade-off for both objectives. In order to select the best trade-off solution we use the procedure followed in [6]. Briefly, we generate 1,000 random weights $w \in [0, 1]$ and compute the average value of the aggregation

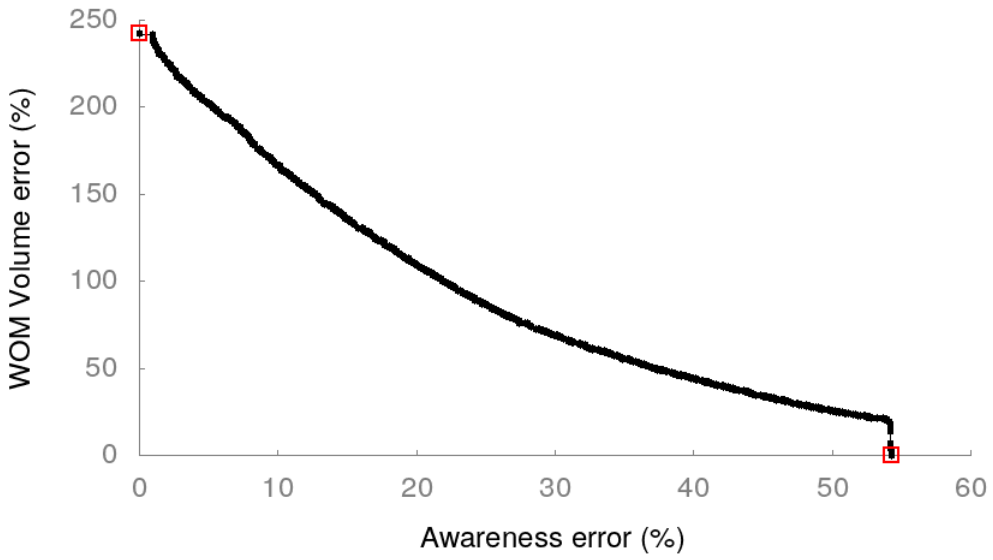


Figure 6: Attainment surface for the controlled experiment. Best solutions for each objective are located at the left end and the right end of the front respectively. These solutions are highlighted in red.

function of both objectives f_1 and f_2. Since the values of f_1 are much bigger than the values of f_2, we apply a normalization factor δ in order to scale them $\delta = \frac{1}{|s|}\sum_{i=1}^{|s|}\frac{f_2(s_i)}{f_1(s_i)}$, where s is the set of solutions in the Pareto front. We formulate this process as $\overline{F}(s_i) = \frac{1}{1000}\sum_{j=1}^{1000}\delta \cdot w_j \cdot f_1(s_i) + (1 - w_j) \cdot f_2(s_i)$. The selected solutions are highlighted in their respective Pareto front approximations in Figure 7.

We visualize the outputs of the model using the calibrated configurations setting the focus on some specific brands in order to carry out an understandable analysis of the behavior of the selected solutions. The KPI evolution along the simulation steps for brands 3 and 6 is shown in Figures 8 and 9 respectively. These brands were chosen because their behavior is a good resemblance of the rest of the brands for both objectives. These charts show the model output for both the training and the hold-out sets, where we can note than the latter obtain similar fitting than the former and therefore, the model can generalize well. Both figures show that adjusting the behavior the dynamics of the awareness evolution over time is harder than the WOM volume dynamics. In contrast, as already identified in the controlled experiment, WOM volume dynamics are more sensible to Monte-Carlo variability for both model instances (as seen in the blurred areas in Figures 8b, 8d, 9b, and 9d).

This is specially relevant for the P-25 instance, as shown in Figures 8a and 9a, since the best calibrated solutions only capture the trend of the target values. In the case of the P-55 instance, the awareness output of the solutions is wavier but the resulting values are far from the target data and the final awareness error for this model instance ends up being greater (as seen in Figure 7). However, this could be a consequence of the synthetically generated target values, that could be too difficult to match. We extract similar conclusions for the WOM volume objective. Although trade-off and best awareness solutions achieve reasonable WOM volume outputs for the P-25 instance (Figures 8b and 9b), final WOM volume errors are higher for the P-55 instance even when considering the fittest solutions (as shown in Figures 8d and 9d).

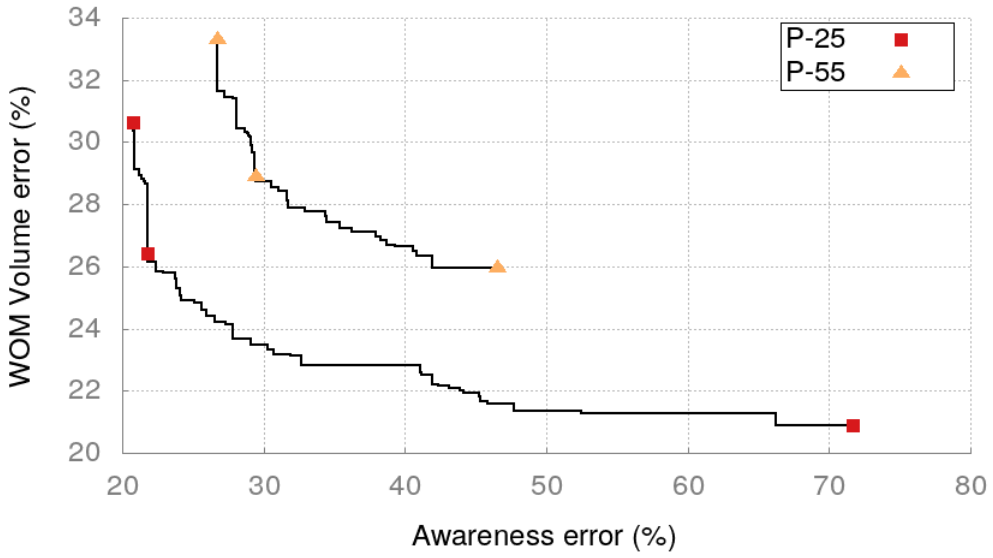
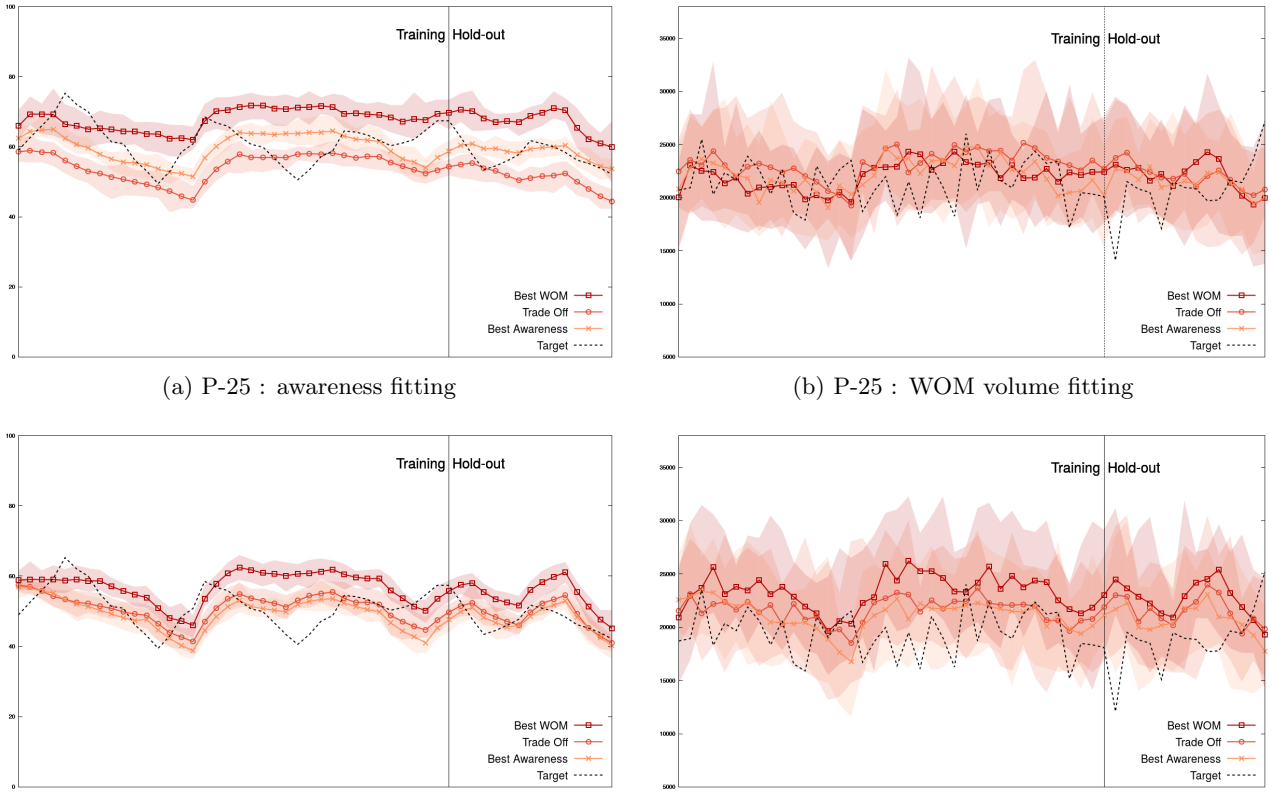


Figure 7: Selected solutions for the two model instances (P-25 and P-55): lowest awareness error, lowest WOM volume error, and best trade-off. Solutions from P-25 are represented using squares and solutions from P-55 are displayed using triangles.

Finally, Table 3 shows a sample of the parameter values of the selected solutions for P-25 and P-55 instances. This sample includes the values of the social network parameters and the different mass media channels. In these solutions we can observe how higher values for the awareness parameters benefit the fitting of WOM dynamics, since more awareness involves more conversations in the social network. However, these values also reduce awareness fitting. We can see that these values are consistent with the behavior shown in Figures 8 and 9. For example, the solution with best trade-off and the solution with lowest WOM volume error (*Best WOM* in Table 3) for the P-25 instance have higher buzz increment values (τ_{c_1} and τ_{c_4}) and higher social awareness impact (α^{WOM}). In contrast, the selected solutions for the P-55 instance show similar buzz increment values for all the selected solutions. In addition, these solutions also have similar values for the awareness deactivation probability (d). This also shows how configurations with high awareness values reduce WOM error (improves WOM adjustment) while reducing awareness fitting.

4.5. Visual and qualitative multicriteria analysis using moGrams

We continue the analysis of the calibrated model instances using moGrams, which composes the second stage of our framework. As said, moGrams is a visualization methodology that combines the visualization of both the design and the objective spaces that aids the decision maker enhancing her understanding of the problem [40]. Our approach is similar to the one followed during the behavior analysis: we apply moGrams to two Pareto set approximations obtained by NSGA-II for both P-25 and P-55 model instances. In order to perform moGrams generation, we need to define a similarity metric for our calibration problem. Our similarity metric $Sim(X_i, X_j) \in [0, 1]$ compares two solutions (i.e., set of calibrated parameters) X_i and X_j using the normalized Euclidean distance, since our calibration problem considers many independent decision variables. The similarity metric is defined in Equation 5.



(a) P-25 : awareness fitting

(b) P-25 : WOM volume fitting

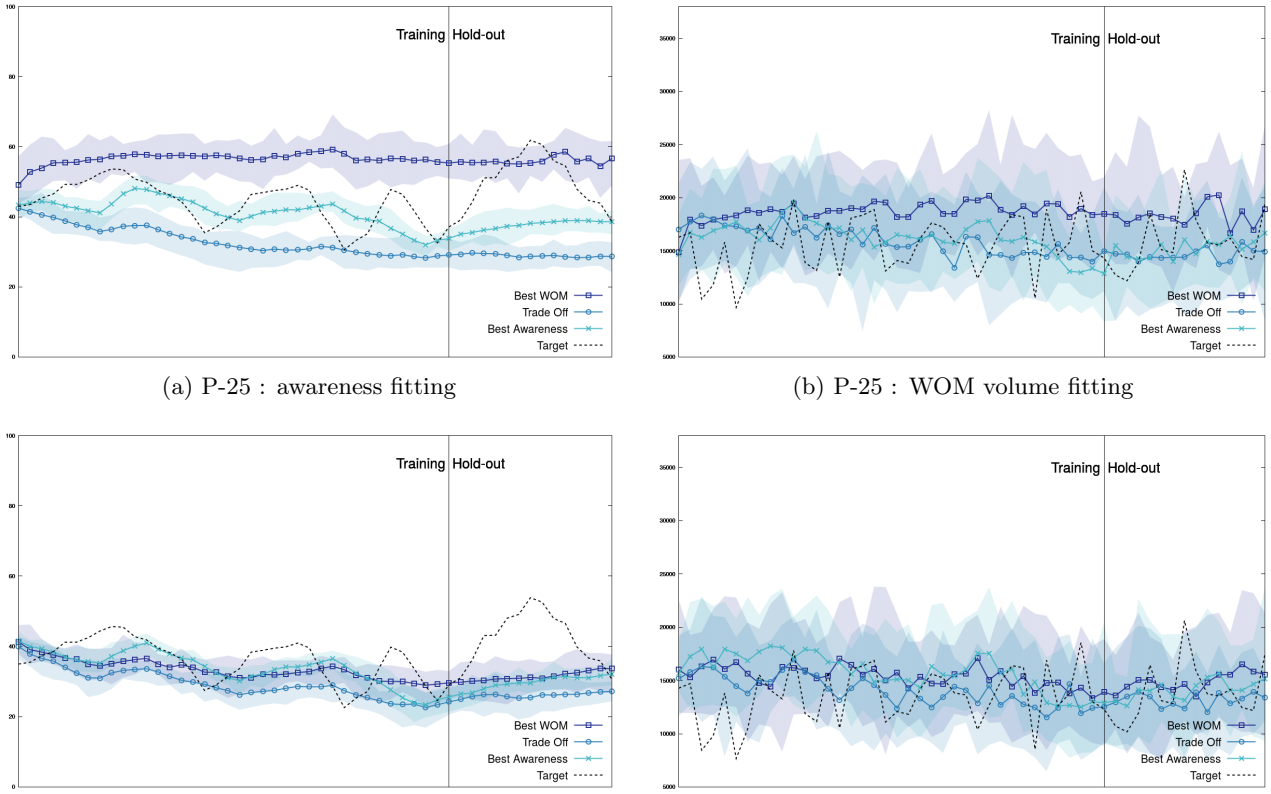
(c) P-55 : awareness fitting

(d) P-55 : WOM volume fitting

Figure 8: Awareness output and WOM volume over time for P-25 and P-55 regarding brand 3. In these charts, the central line represents the average of the Monte-Carlo simulation and the blurred areas represent the minimum and maximum values obtained for all the Monte-Carlo 15 independent runs. In addition, the dashed lines represent target values. Best WOM and best awareness (lowest error) solutions are represented with pointed lines containing squares and crosses respectively. Trade-off solutions are represented using lines with circles.

$$Sim(X_i, X_j) = 1 - \sqrt{\frac{(x_{i1} - x_{j1})^2 + ... + (x_{in} - x_{jn})^2}{n}}. \quad (5)$$

The generated moGram for the P-25 model instance is shown in Figure 10 and its associated Pareto front approximation is displayed in Figure 11. We can see that, given the relatively high cardinality of the Pareto front approximation (46 solutions), the decision making process for this model instance seems too complex to deal without a visualization method. Following the moGrams methodology, each node in the generated network is associated to an individual solution (i.e., model parameter setting) from the Pareto set approximation. We draw each node as a pie where the upper pie segment represents the awareness error objective using degradation between orange and white while the lower pie segment represents the WOM volume error objective using degradation between blue and white. For both objectives, a more intense color means a better value with a white color being the worst possible value. In addition, we have included indexes for the solutions in both the network and the Pareto front approximation for making their relation clearer (see Figures 10 and 11). We provide several observations from the moGrams visualization:



(a) P-25 : awareness fitting

(b) P-25 : WOM volume fitting

(c) P-55 : awareness fitting

(d) P-55 : WOM volume fitting

Figure 9: Awareness output and WOM volume over time for P-25 and P-55 regarding brand 6. In these charts, the central line represents the average of the Monte-Carlo simulation and the blurred areas represent the minimum and maximum values obtained for all the Monte-Carlo 15 independent runs. In addition, the dashed lines represent target values. Best WOM and best awareness (lowest errors) solutions are represented with pointed lines containing squares and crosses respectively. Trade-off solutions are represented using lines with circles.

- Regarding the structure of the network, we can identify multiple clusters of solutions (i.e., groups of solutions) in the design space. Two of these subsets of solutions, located in the left side of the network, are connected to the general network by Solution 31, which bridges with another subset through Solution 28. In addition, Solution 15 can be identified as another hub that connects to another subset of solutions located in the right side of the network.

- From those clusters we can identify Solution 31 and Solution 28 as the most connected ones, since they are the only solutions with degree 4. Due to their connectivity and the additional information provided by moGrams, these solutions could be interesting configurations for the modeler. In terms of similarity, both solutions have values of 0.9, which suggests they have good flexibility and can be swapped by other solutions with minimum parameter changes.

- The moGrams visualization methodology assists us in validating the best trade-off solution (Solution 9, located in the right side of the map), which could be a suitable model configuration due to its good balance for both objectives. This solution is connected with Solutions 5 and 16 with similarity values of 0.8. However, these solutions are located in the same region

P-25	τ_{c_1}	α_{c_3}	τ_{c_4}	$d\tau_{c_4}$	α_{c_7}	$d\tau_{c_7}$	$p_i^b(0)$	α^{WOM}
Best Aw.	0.159	0.212	0.207	0.728	0.2	0.09	0.29	0.152
Trade-Off	0.579	0.492	0.273	0.672	0.25	0.679	0.326	0.138
Best WOM	0.935	0.625	0.655	0.762	0.709	0.023	0.289	0.833
P-55	τ_{c_1}	$d\tau_{c_2}$	α_{c34}	$d\tau_{c_4}$	$\alpha_{c_{15}}$	$\tau_{c_{16}}$	α^{WOM}	d
Best Aw.	0.6	0.187	0.127	0.128	0.003	0.904	0.15	0.201
Trade-Off	0.629	0.043	0.404	0.004	0.02	0.988	0.175	0.2
Best WOM	0.542	0.078	0.976	0.514	0.245	0.997	0.243	0.207

Table 3: Sample of the parameter values for the selected solutions for P-25 and P-55 instances. This sample includes a selection of parameters for the social network and different mass media channels.

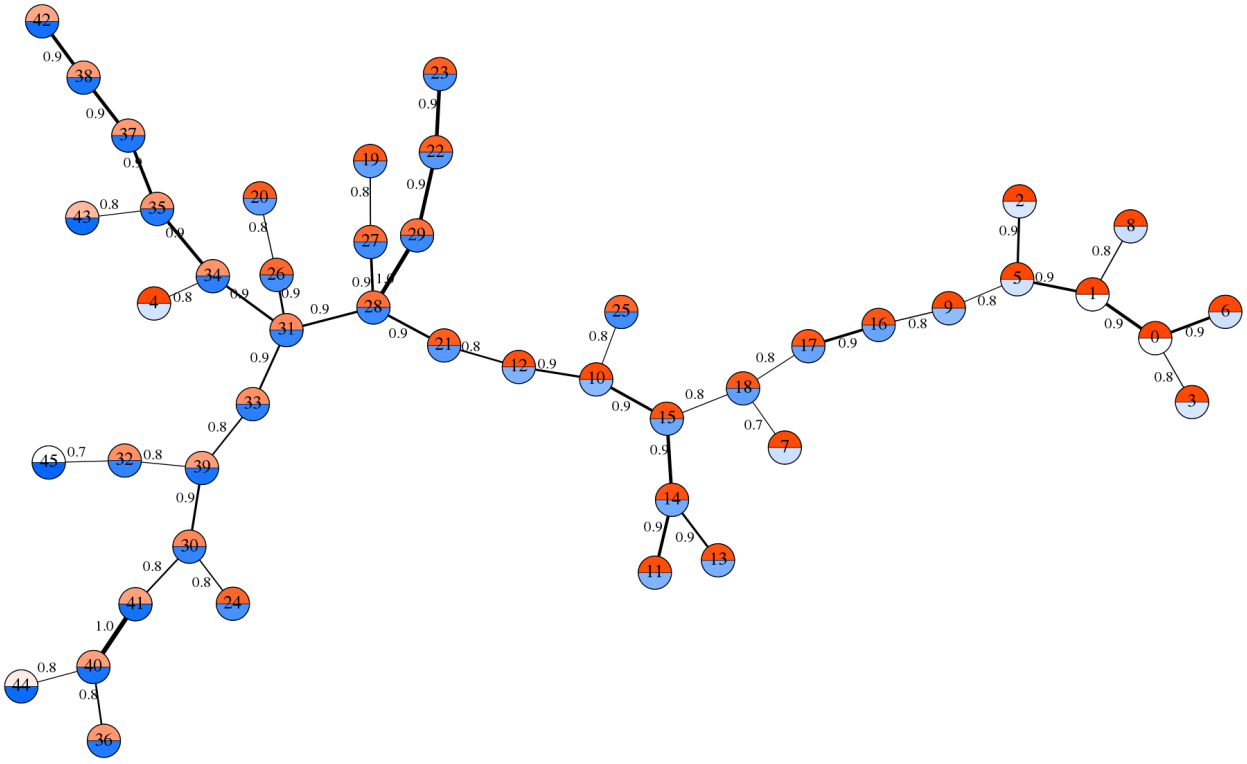


Figure 10: moGrams network representing the non-dominated calibration solutions for P-25 model instance.

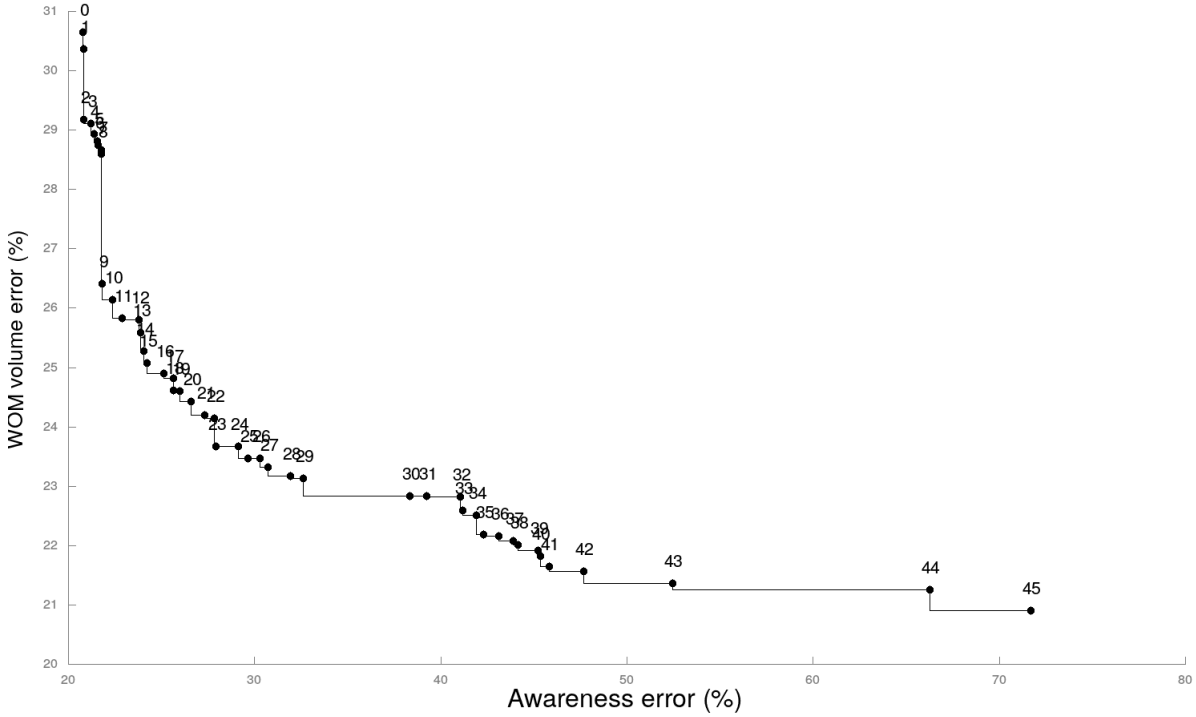


Figure 11: Pareto front approximation for P-25 model instance associated to the moGram of Figure 10.

of the Pareto front approximation, reducing the interest of swapping Solution 9 with any of its neighbors.

- With respect to the best solutions for each objective (Solutions 0 and 45), both of them are located in opposite regions of the network. Solution 0 has three neighbors with similarity values beyond 0.8, meanwhile Solution 45 has a single connection to Solution 32 with a low similarity value (0.7). However, the neighborhood of Solution 0 may not be really interesting, since all the solutions are close in the Pareto front approximation.

Figure 12 shows the generated moGram for the P-55 model instance while its associated Pareto front approximation is displayed in Figure 13. Similarly to the previous moGram, its associated Pareto front approximation has a relatively high cardinality (32 solutions). Again, we can provide the following observations and interesting insights for the modeler from a validation point of view:

- Due to the star topology of the network, three main subsets of solutions in the design space arise, which grow as tree-shaped subnetworks from Solution 13. We can identify Solution 16 as the most connected, since it is the only with four neighbors. This solution could be interesting for the modeler since its connections include solutions from opposite regions of the Pareto front approximations (i.e., Solution 31, the solution with the lowest WOM error and Solution 2, which is close to the solution with the lowest awareness error).

- In addition, several other solutions have three connections. From those, we could highlight Solution 2 and Solution 12 since their connections are more diverse and include solutions from other regions of the Pareto front approximation. Solution 12 can also be identified as

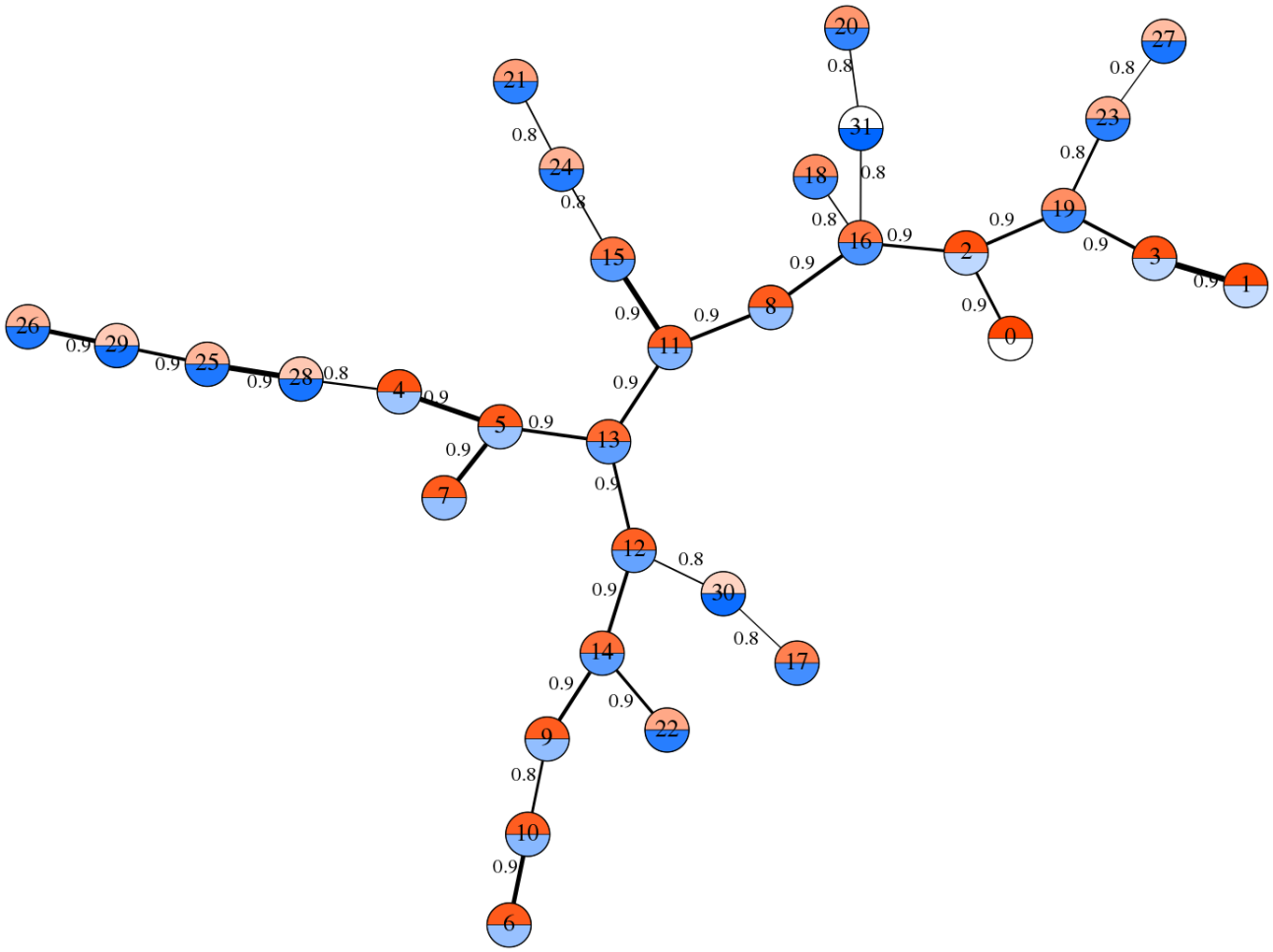


Figure 12: moGrams network representing the non-dominated calibration solutions for P-55 model instance.

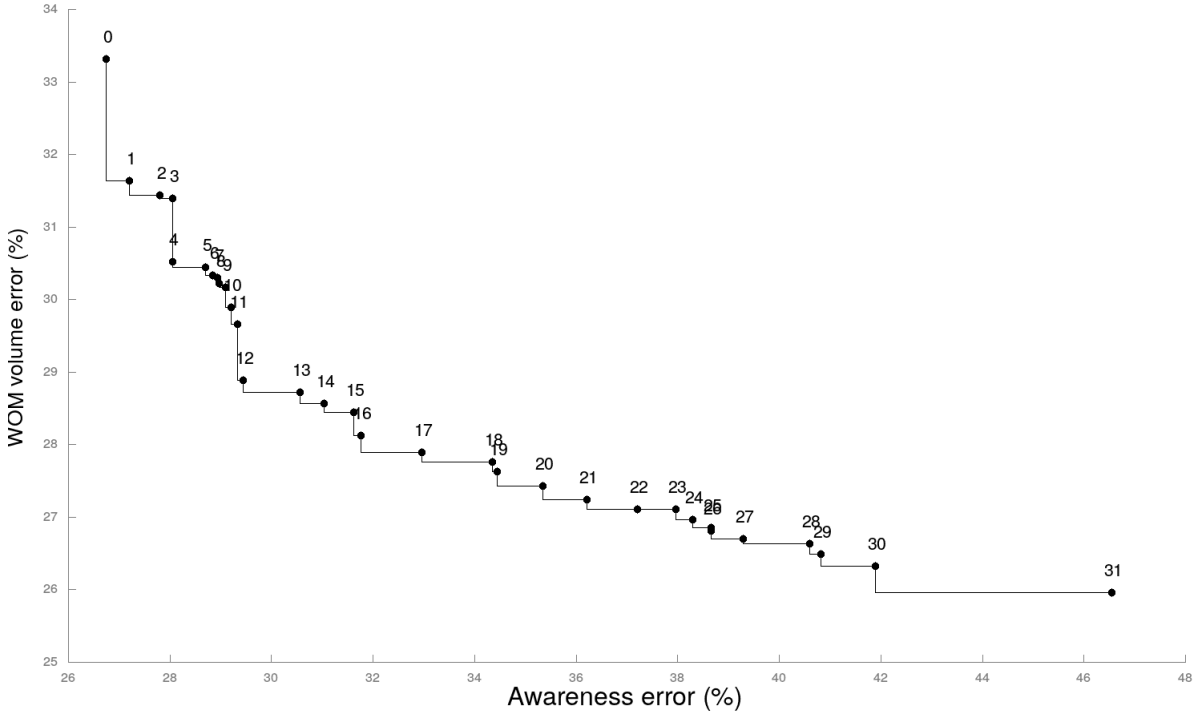


Figure 13: Pareto front approximation for P-55 model instance associated to the moGram of Figure 12.

the trade-off solution, which could be a plus for the modeler. The most interesting neighbor of Solution 12 is Solution 30 with a similarity value of 0.8.

- Finally, the best solutions for each objective are located in the same subset of solutions. As already pointed out, Solution 31 is connected to the hub defined by Solution 16. However, Solution 0 is isolated with a single connection to Solution 2, which belongs to the same region of the Pareto front approximation. Additionally, we can observe that the proximity in the network for both solutions could be relevant for the modeler, since it suggests that modifying the value of some sensible parameters can drastically change the behavior of the model.

The previous statement can be confirmed by sampling the parameters of Solutions 0 and 16, shown in Table 4. Additionally, we also included the parameters of the solutions with best trade-off and the solution with lowest WOM error. These parameters were selected because they show the largest differences among the selected configurations. It can be observed that most of these parameters are related with buzz increment and buzz decay. Solution 0 has lower increment and decay values of several mass media channels, suggesting that these parameters are really important to control the behavior of the model. Finally, we should note that the awareness impact values of Solution 31 are higher than the values the other solutions, supporting our previous insight that higher awareness can benefit WOM fitting.

5. Final remarks

In this paper we have introduced a multicriteria integral framework for the calibration and validation of ABMs considering multiple objectives. The framework comprises an EMO method to

	$d\tau_{c_1}$	α_{c_3}	τ_{c_3}	τ_{c_4}	$d\tau_{c_4}$	$d\tau_{c_6}$	τ_{c_9}	$\alpha_{c_{10}}$
Solution 0	0.041	0.127	0.172	0.407	0.128	0.370	0.954	0.045
Solution 2	0.626	0.563	0.718	0.468	0.112	0.440	0.794	0.071
Solution 12	0.133	0.404	0.852	0.855	0.004	0.456	0.492	0.079
Solution 16	0.602	0.560	0.199	0.873	0.010	0.473	0.958	0.095
Solution 31	0.555	0.976	0.221	0.922	0.514	0.918	0.417	0.469

Table 4: Sample of the parameter values for the identified solutions for P-55 using its moGram at Figure 12.

search for the best set of configuration parameters and a visualization method to help the modeler in the decision making of the best model parameter setting. We have applied the novel multicriteria framework to an ABM for marketing scenarios, driven by awareness and WOM volume as KPIs.

We have designed and implemented our calibration approach using NSGA-II. The proposed framework has been validated in two different experiments: through a controlled experiment where we show that our approach identifies the optimal model configurations for each objective, and through its application to two model instances using real data. The first experiment has allowed us to demonstrate that the multiobjective optimization considered in the first stage is actually able to obtain high-quality calibrated models. In the second experiment, we have analyzed the latter resulting Pareto front approximations by selecting three solutions: the solution with best awareness error (f_1), the solution with best WOM volume error (f_2), and the solution with best trade-off for both objectives. Our analysis suggested that awareness dynamics were more difficult to adjust than the WOM volume for the calibrated instances, specially for P-55, the instance with highest dimensionality. Due to the reasonable fitting behavior for the baseline model, we can conclude that the increasing dimensionality of the problem influences the fitting of the resulting models.

Finally, we have applied the second stage of our framework by analyzing the design space for our calibration problem using moGrams on individual Pareto front approximations from P-25 and P-55 instances. This analysis has shown the usefulness of our framework when validating relevant solutions and assessing their flexibility (i.e., the solution with best trade-off for both objectives) from the Pareto front approximation. We have concluded that the solutions with the best trade-off had good flexibility but they did not have interesting neighboring solutions in the decision space. In contrast, other solutions with higher degree had the potential of being more relevant for the modeler. We could also notice that analyzing a Pareto set approximation with this high cardinality (31 and 46 solutions, respectively) without a visualization methodology such as moGrams can be difficult for modelers and stakeholders. For example, it would have been tricky to identify and validate alternative solutions (Solution 16 from Figure 12 would be an example). Thus, we have shown moGrams is a powerful resource for aiding modelers when dealing with multiobjective model calibration problems.

As future works, we consider extending our designed consumer model for including brand sales as an additional KPI. In addition, including new objectives could involve replacing NSGA-II with another EMO algorithm that could be able to successfully deal with many-objective optimization. Due to the high cardinality of the Pareto set approximations delivered during our experiments, we also consider interesting to extend our calibration approach to evaluate the impact of including the modeler's preferences during the calibration process [39].

Acknowledgments

This work is supported by Spanish Ministerio de Economía y Competitividad under the EX-ASOCO project (ref. PGC2018-101216-B-I00), including European Regional Development Funds (ERDF). M. Chica is also supported through the Ramón y Cajal program (RYC-2016-19800).

References

[1] Atamturktur, S., Liu, Z., Cogan, S., Juang, H., 2015. Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach. Structural and Multidisciplinary Optimization 51, 659–671.

[2] Bader, J., Zitzler, E., 2011. Hype: An algorithm for fast hypervolume-based many-objective optimization. Evolutionary Computation 19, 45–76.

[3] Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

[4] Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling & Software 40, 1–20.

[5] Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99, 7280–7287.

[6] Campomanes-Álvarez, B.R., Cordón, O., Damas, S., 2013. Evolutionary multi-objective optimization for mesh simplification of 3D open models. Integrated Computer-Aided Engineering 20, 375–390.

[7] Chica, M., Barranquero, J., Kajdanowicz, T., Cordón, O., Damas, S., 2017. Multimodal optimization: an effective framework for model calibration. Information Sciences 375, 79–97.

[8] Chica, M., Óscar Cordón, Damas, S., Iglesias, V., Mingot, J., 2016. Identimod: Modeling and managing brand value using soft computing. Decision Support Systems 89, 41–55.

[9] Chica, M., Rand, W., 2017. Building agent-based decision support systems for word-of-mouth programs: A freemium application. Journal of Marketing Research 54, 752–767.

[10] Deb, K., 2001. Multi-objective optimization using evolutionary algorithms. volume 16. John Wiley & Sons.

[11] Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. IEEE Transactions on Evolutionary Computation 18, 577–601.

[12] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197.

[13] Epstein, J.M., 2006. Generative social science: Studies in agent-based computational modeling. Princeton University Press.

[14] Farmer, J.D., Foley, D., 2009. The economy needs agent-based modelling. Nature 460, 685–686.

[15] Farris, P.W., Bendle, N.T., Pfeifer, P.E., Reibstein, D.J., 2010. Marketing metrics: The definitive guide to measuring marketing performance. Wharton School Publishing. 2nd edition.

[16] González-Avella, J.C., Cosenza, M.G., Klemm, K., Eguíluz, V.M., San Miguel, M., 2007. Information feedback and mass media effects in cultural dynamics. Journal of Artificial Societies and Social Simulation 10, 9.

[17] Hassan, S., Arroyo, J., Galán, J.M., Antunes, L., Pavón, J., 2013. Asking the oracle: Introducing forecasting principles into agent-based modelling. Journal of Artificial Societies and Social Simulation 16, 13.

[18] Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688.

[19] Janssen, M.A., Ostrom, E., 2006. Empirically based, agent-based models. Ecology and Society 11, 37.

[20] Lee, J.S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, J.G., Sun, Z., Parker, D.C., 2015. The complexities of agent-based modeling output analysis. Journal of Artificial Societies and Social Simulation 18, 4.

[21] Li, H., Zhang, Q., 2009. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. IEEE Transactions on Evolutionary Computation 13, 284–302.

[22] Libai, B., Muller, E., Peres, R., 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. Journal of Marketing Research 50, 161–176.

[23] López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L.P., Birattari, M., Stützle, T., 2016. The irace package: Iterated racing for automatic algorithm configuration. Operations Research Perspectives 3, 43–58.

[24] Macal, C.M., North, M.J., 2005. Tutorial on agent-based modeling and simulation, in: Proceedings of the 37th conference on Winter simulation, ACM. pp. 2–15.
[25] Macdonald, E.K., Sharp, B.M., 2000. Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. Journal of Business Research 48, 5–15.
[26] Miller, J.H., 1998. Active nonlinear tests (ANTs) of complex simulation models. Management Science 44, 820–830.
[27] Moya, I., Chica, M., Sáez-Lozano, J.L., Cordón, Ó., 2017. An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections. Knowledge-Based Systems 123, 200–216.
[28] Narzisi, G., Mysore, V., Mishra, B., 2006. Multi-objective evolutionary optimization of agent-based models: An application to emergency response planning, in: Proceedings of the 2nd IASTED International Conference on Computational Intelligence, CI 2006, pp. 224–230.
[29] Nebro, A.J., Durillo, J.J., Vergne, M., 2015. Redesigning the jMetal multi-objective optimization framework, in: Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, ACM, New York, NY, USA. pp. 1093–1100.
[30] Newman, M., Barabási, A.L., Watts, D.J., 2006. The structure and dynamics of networks. Princeton University Press.
[31] Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research 151, 552–568.
[32] Read, M.N., Alden, K., Rose, L.M., Timmis, J., 2016. Automated multi-objective calibration of biological agent-based simulations. Journal of The Royal Society Interface 13.
[33] Samek, W., Wiegand, T., Müller, K., 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. CoRR abs/1708.08296. 1708.08296.
[34] Sargent, R.G., 2005. Verification and validation of simulation models, in: Proceedings of the 37th conference on Winter simulation, pp. 130–143.
[35] Schramm, M.E., Trainor, K.J., Shanker, M., Hu, M.Y., 2010. An agent-based diffusion model with consumer and brand agents. Decision Support Systems 50, 234–242.
[36] Schvaneveldt, R.W., Durso, F.T., Dearholt, D.W., 1989. Network structures in proximity data, Academic Press. volume 24 of *Psychology of Learning and Motivation*, pp. 249–284.
[37] Stonedahl, F., Rand, W., 2014. When does simulated data match real data? Comparing model calibration functions using genetic algorithms, in: Advances in Computational Social Science. Springer, Japan. volume 11 of *Agent-Based Social Systems*, pp. 297–313.
[38] Talbi, E.G., 2009. Metaheuristics: from design to implementation. John Wiley & Sons.
[39] Thiele, L., Miettinen, K., Korhonen, P.J., Molina, J., 2009. A preference-based evolutionary algorithm for multi-objective optimization. Evolutionary computation 17, 411–436.
[40] Trawiński, K., Chica, M., Pancho, D.P., Damas, S., Cordón, O., 2018. moGrams: A network-based methodology for visualizing the set of nondominated solutions in multiobjective optimization. IEEE Transactions on Cybernetics 48, 474–485.
[41] Tušar, T., Filipič, B., 2015. Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. IEEE Transactions on Evolutionary Computation 19, 225–245.
[42] Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders–next generation. Environmental Modelling & Software 77, 196–220.
[43] Waldrop, M.M., 2018. Free agents. Science 360, 144–147.
[44] Walker, D.J., Everson, R., Fieldsend, J.E., 2013. Visualizing mutually nondominating solution sets in many-objective optimization. IEEE Transactions on Evolutionary Computation 17, 165–184.
[45] Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.
[46] Wu, F., Huberman, B.A., 2007. Novelty and collective attention. Proceedings of the National Academy of Sciences 104, 17599–17601.
[47] Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 599–608.
[48] Yang, S., Li, M., Liu, X., Zheng, J., 2013. A grid-based evolutionary algorithm for many-objective optimization. IEEE Transactions on Evolutionary Computation 17, 721–736.
[49] Zaffar, M.A., Kumar, R.L., Zhao, K., 2011. Diffusion dynamics of open source software: An agent-based computational economics (ACE) approach. Decision Support Systems 51, 597–608.
[50] Zaffar, M.A., Kumar, R.L., Zhao, K., 2018. Using agent-based modelling to investigate diffusion of mobile-based branchless banking services in a developing country. Decision Support Systems (in press) .

[51] Zhang, X., Tian, Y., Jin, Y., 2015. A knee point-driven evolutionary algorithm for many-objective optimization. IEEE Transactions on Evolutionary Computation 19, 761–776.

[52] Zhang, Y., Shao, Q., Taylor, J.A., 2016. A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model. Journal of Hydrology 538, 802–816.

[53] Zitzler, E., Laumanns, M., Thiele, L., 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. Technical Report 103. Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zurich.

[54] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., 2003. Performance assessment of multiobjective optimizers: An analysis and review. IEEE Transactions on Evolutionary Computation 7, 117–132.

Bibliography

[ABL+] G. Agrawal, C. Bloebaum, K. Lewis, K. Chugh, C.-H. Huang, and S. Parashar. *Intuitive Visualization of Pareto Frontier for Multiobjective Optimization in n-Dimensional Performance Space.*

[AH05] A. Auger and N. Hansen. A restart CMA evolution strategy with increasing population size. In *2005 IEEE Congress on Evolutionary Computation*, vol. 2, pp. 1769–1776, 2005.

[ALCJ15] S. Atamturktur, Z. Liu, S. Cogan, and H. Juang. Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach. *Structural and Multidisciplinary Optimization*, 51(3):659–671, 2015.

[AM08] J. Adams and A. K. Mayer. Condorcet efficiency with adaptive parties in a spatial model. *Mathematical and Computer Modelling*, 48(9):1298–1307, 2008.

[Aso04] Asociación para la Investigación de Medios de Comunicación (AIMC). AIMC February to November 2004. General Recap. EMG. `http://www.aimc.es/-Datos-EGM-Resumen-General-.html`, 2004. Accessed: 2016-03-28.

[BA99] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[BABR17] J. Branke, M. Asafuddoula, K. S. Bhattacharjee, and T. Ray. Efficient use of partially converged simulations in evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 21(1):52–64, 2017.

[BAHT17] D. Brockhoff, A. Auger, N. Hansen, and T. Tušar. Quantitative performance assessment of multiobjective optimizers: The average runtime attainment function. In H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme, editors, *Evolutionary Multi-Criterion Optimization*, pp. 103–119, Cham, 2017. Springer International Publishing.

[Bal07] V. A. Bali. Terror and elections: Lessons from Spain. *Electoral Studies*, 26(3):669–687, 2007.

[BCD+18] E. Bermejo, M. Chica, S. Damas, S. Salcedo-Sanz, and O. Cordón. Coral reef optimization with substrate layers for medical image registration. *Swarm and Evolutionary Computation*, 42:138–159, 2018.

[BCG+13] N. D. Bennett, B. F. Croke, G. Guariso, J. H. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, L. T. Newham, J. P. Norton, C. Perrin, S. A. Pierce, B. Robson, R. Seppelt, A. A. Voinov, B. D. Fath, and V. Andreassian.

Characterising performance of environmental models. *Environmental Modelling & Software*, 40:1–20, 2013.

[BDRD+20] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[Ber54] B. Berelson. *Voting; a study of opinion formation in a presidential campaign.* University of Chicago Press, Chicago, 1954.

[BFM97] T. Back, D. B. Fogel, and Z. Michalewicz. *Handbook of evolutionary computation.* IOP Publishing Ltd., Bristol (UK), 1997.

[Bha13] M. Bhattacharya. Evolutionary approaches to expensive optimisation. *International Journal of Advanced Research in Artificial Intelligence*, 2(3), 2013.

[BHSM08] X. Blasco, J. Herrero, J. Sanchis, and M. Martínez. A new graphical visualization of n-dimensional pareto front for decision-making in multiobjective optimization. *Information Sciences*, 178(20):3908–3924, 2008. Special Issue on Industrial Applications of Neural Networks.

[BK07] C. Berrebi and E. F. Klor. *The Impact of Terrorism on Voters' Preferences.* RAND Santa Monica, CA, 2007.

[BK08] C. Berrebi and E. F. Klor. Are voters sensitive to terrorism? direct evidence from the Israeli electorate. *The American Political Science Review*, 102(3):279–301, 2008.

[BN07] E. G. Bekele and J. W. Nicklow. Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology*, 341(3):165–176, 2007.

[BN19] S. Bandaru and A. H. C. Ng. Trend mining: A visualization technique to discover variable trends in the objective space. In K. Deb, E. Goodman, C. A. Coello Coello, K. Klamroth, K. Miettinen, S. Mostaghim, and P. Reed, editors, *Evolutionary Multi-Criterion Optimization*, pp. 605–617, Cham, 2019. Springer International Publishing.

[BNE07] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

[Bon02] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287, 2002.

[BS89] R. A. Brody and C. R. Shapiro. Policy failure and public support: The Iran-Contra affair and public assessment of President Reagan. *Political Behavior*, 11(4):353–369, 1989.

[BTAH19] D. Brockhoff, T. Tušar, A. Auger, and N. Hansen. Using well-understood single-objective functions in multiobjective black-box optimization test suites, 2019.

[BZ11] J. Bader and E. Zitzler. Hype: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011.

[CÁCD13a] B. R. Campomanes-Álvarez, O. Cordón, and S. Damas. Evolutionary multi-objective optimization for mesh simplification of 3d open models. *Integrated Computer-Aided Engineering*, 20(4):375–390, 2013.

[CACD13b] B. R. Campomanes-Álvarez, O. Cordón, and S. Damas. Evolutionary multi-objective optimization for mesh simplification of 3D open models. *Integrated Computer-Aided Engineering*, 20(4):375–390, 2013.

[CBK+17] M. Chica, J. Barranquero, T. Kajdanowicz, O. Cordón, and S. Damas. Multimodal optimization: an effective framework for model calibration. *Information Sciences*, 375:79–97, 2017.

[CC15] E. Canessa and S. Chaigneau. Calibrating agent-based models using a genetic algorithm. *Studies in Informatics and Control*, 24(1):79–90, 2015.

[CCD+16] M. Chica, O. Cordón, S. Damas, V. Iglesias, and J. Mingot. Identimod: modeling and managing brand value using soft computing. *Decision Support Systems*, 89:41–55, 2016.

[CD07] D. Chong and J. N. Druckman. Framing theory. *Annual Review of Political Science*, 10:103–126, 2007.

[Cen04] Centro de Investigaciones Sociológicas. Estudio 2555. CIS Data bank, 2004.

[CGMMGGSS19] C. Camacho-Gómez, I. Marsa-Maestre, J. M. Gimenez-Guzman, and S. Salcedo-Sanz. A coral reefs optimization algorithm with substrate layer for robust Wi-Fi channel assignment. *Soft Computing*, 23(23):12621–12640, 2019.

[CH06] B. Calvez and G. Hutzler. Automatic tuning of agent-based models using genetic algorithms. In J. S. Sichman and L. Antunes, editors, *Multi-Agent-Based Simulation VI*, pp. 41–57, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[Cho11] C. Chowanietz. Rallying around the flag or railing against the government? Political parties' reactions to terrorist acts. *Party Politics*, 17(5):673–698, 2011.

[CJB+20] M. Chica, A. Juan, C. Bayliss, O. Cordon, and D. Kelton. Why simheuristics? benefits, limitations, and best practices when combining metaheuristics with simulation. *SORT*, 22(2):1–35, 2020.

[CLG14] E. César Trejo Zúñiga, I. L. López Cruz, and A. R. García. Parameter estimation for crop growth model using evolutionary and bio-inspired algorithms. *Applied Soft Computing*, 23:474–482, 2014.

[Clo08] E. Clough. Still converging? A downsian party system without polls. *Journal of Theoretical Politics*, 20(4):461–476, 2008.

[CLVV+07] C. A. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al. *Evolutionary algorithms for solving multi-objective problems*, vol. 5. Springer, 2007.

[Con04] Consell de l'Audiovisual de Catalunya. 11-14M: la construcció televisiva (in catalonian). *Quaderns del CAC*, (19-20):89–204, 2004.

[CR17] M. Chica and W. Rand. Building agent-based decision support systems for word-of-mouth programs: A freemium application. *Journal of Marketing Research*, 54(5):752–767, 2017.

[CS08] J. Chan and W. Suen. A Spatial Theory of News Consumption and Electoral Competition. *The Review of Economic Studies*, 75(3):699–728, 2008.

[CW07] R. B. Confesor and G. W. Whittaker. Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and Pareto optimization. *Journal of the American Water Resources Association*, 43(4):981–989, 2007.

[DA94] K. Deb and R. B. Agrawal. Simulated Binary Crossover for Continuous Search Space. Technical report, Departement of Mechanical Enginering, Indian Institute of Technology, Kanpur, India, 1994.

[Deb01] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[DJ14] K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014.

[DJBJ10] S. A. Delre, W. Jager, T. H. A. Bijmolt, and M. A. Janssen. Will it spread or not? the effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management*, 27(2):267–282, 2010.

[DM11] J. Duggan and C. Martinelli. A Spatial Theory of Media Slant and Voter Choice. *The Review of Economic Studies*, 78(2):640–666, 2011.

[DOM+19] J. Del Ser, E. Osaba, D. Molina, X.-S. Yang, S. Salcedo-Sanz, D. Camacho, S. Das, P. N. Suganthan, C. A. Coello Coello, and F. Herrera. Bio-inspired computation: Where we stand and what's next. *Swarm and Evolutionary Computation*, 48:220–250, 2019.

[Dow57] A. Downs. An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150, 1957.

[DPAM02] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[Dun61] O. J. Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64, 1961.

[DYX+09] C. Dai, M. Yao, Z. Xie, C. Chen, and J. Liu. Parameter optimization for growth model of greenhouse crop using genetic algorithms. *Applied Soft Computing*, 9(1):13–19, 2009.

[EH84] J. M. Enelow and M. J. Hinich. *The spatial theory of voting: An introduction*. CUP Archive, 1984.

[EH94] J. M. Enelow and M. J. Hinich. A test of the predictive dimensions model in spatial voting theory. *Public Choice*, 78(2):155–169, 1994.

[Eic08] G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization.* Springer, 2008.

[Eps06] J. M. Epstein. *Generative social science: Studies in agent-based computational modeling.* Princeton University Press, 2006.

[ES93] L. J. Eshelman and J. D. Schaffer. Real-coded genetic algorithms and interval-schemata. In L. D. WHITLEY, editor, *Foundations of Genetic Algorithms*, vol. 2 of *Foundations of Genetic Algorithms*, pp. 187–202. Elsevier, 1993.

[Fab13] A. Fabretti. On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, 8(2):277–293, 2013.

[FBPR10] P. W. Farris, N. T. Bendle, P. E. Pfeifer, and D. J. Reibstein. *Marketing metrics: The definitive guide to measuring marketing performance.* Wharton School Publishing, 2nd edition, 2010.

[FCAM19] J. E. Fieldsend, T. Chugh, R. Allmendinger, and K. Miettinen. A feature rich distance-based many-objective visualisable test problem generator. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, pp. 541–549, New York, NY, USA, 2019. Association for Computing Machinery.

[FCC18] J. G. Falcón-Cardona and C. A. Coello. Towards a more general many-objective evolutionary optimizer using multi-indicator density estimation. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '18, pp. 1890–1893, New York, NY, USA, 2018. Association for Computing Machinery.

[FE13] J. Fieldsend and R. Everson. Visualising high-dimensional pareto relationships in two-dimensional scatterplots. In R. C. Purshouse, P. J. Fleming, C. M. Fonseca, S. Greco, and J. Shaw, editors, *Evolutionary Multi-Criterion Optimization*, pp. 558–572, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[FF96] C. M. Fonseca and P. J. Fleming. On the performance assessment and comparison of stochastic multiobjective optimizers. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature — PPSN IV*, pp. 584–593, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

[FF09] J. D. Farmer and D. Foley. The economy needs agent-based modelling. *Nature*, 460(7256):685–686, 2009.

[Fis05] R. Fishman. On the continuing relevance of the Weberian methodological perspective (with applications to the Spanish case of elections in the aftermath of terrorism). *Working Paper of the Helen Kellogg Institute for International Studies*, pp. 1–28, 2005.

[FLPEC05] E. Fieldhouse, L. Lessard-Phillips, B. Edmonds, and N. Crossley. The complexity of turnout: An agent-based simulation of turnout cascades. *Electoral Studies*, 24(2), 2005.

[FM08] D. Floreano and C. Mattiussi. *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies.* The MIT Press, 2008.

[FNR+16] S. Farhadi, M. R. Nikoo, G. R. Rakhshandehroo, M. Akhbari, and M. R. Alizadeh. An agent-based-nash modeling framework for sustainable groundwater management: A case study. *Agricultural Water Management*, 177:348–358, 2016.

[Fon95] C. M. M. d. Fonseca. *Multiobjective genetic algorithms with application to control engineering problems.* PhD thesis, University of Sheffield, 1995.

[Fri40] M. Friedman. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11(1):86–92, 1940.

[FS05] J. H. Fowler and O. Smirnov. Dynamic parties and social turnout: An agent-based model. *American Journal of Sociology*, 110(4):1070–1094, 2005.

[FT18] B. Filipič and T. Tušar. A taxonomy of methods for visualizing pareto front approximations. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '18, pp. 649–656, New York, NY, USA, 2018. Association for Computing Machinery.

[FT20a] B. Filipič and T. Tušar. Visualization in multiobjective optimization. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, GECCO '20, p. 775–800, New York, NY, USA, 2020. Association for Computing Machinery.

[FT20b] B. Filipič and T. Tušar. Visualization in multiobjective optimization. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, GECCO '20, pp. 775–800, New York, NY, USA, 2020. Association for Computing Machinery.

[GACK+07] J. C. González-Avella, M. G. Cosenza, K. Klemm, V. M. Eguíluz, and M. San Miguel. Information feedback and mass media effects in cultural dynamics. *Journal of Artificial Societies and Social Simulation*, 10(3):9, 2007.

[Gar05] R. Garcia. Uses of agent-based modeling in innovation/new product development research. *Journal of Product Innovation Management*, 22(5):380–398, 2005.

[GCCCH20] J. Giráldez-Cru, M. Chica, O. Cordón, and F. Herrera. Modeling agent-based consumers decision-making with 2-tuple fuzzy linguistic perceptions. *International Journal of Intelligent Systems*, 35(2):283–299, 2020.

[GdFFH01] V. Grunert da Fonseca, C. M. Fonseca, and A. O. Hall. Inferential performance assessment of stochastic optimisers and the attainment function. In E. Zitzler, L. Thiele, K. Deb, C. A. Coello Coello, and D. Corne, editors, *Evolutionary Multi-Criterion Optimization*, pp. 213–225, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[GG20] T. A. González and G. D. Granic. Spatial voting meets spatial policy positions: An experimental appraisal. *American Political Science Review*, 114(1):285—290, 2020.

[GHGHSM+20] L. Garcia-Hernandez, J. Garcia-Hernandez, L. Salas-Morera, C. Carmona-Muñoz, N. Alghamdi, J. V. de Oliveira, and S. Salcedo-Sanz. Addressing unequal area facility layout problems with the coral reef optimization algorithm with substrate layers. *Engineering Applications of Artificial Intelligence*, 93:103697, 2020.

[GHSMCM+20] L. Garcia-Hernandez, L. Salas-Morera, C. Carmona-Muñoz, J. Garcia-Hernandez, and S. Salcedo-Sanz. A novel island model based on coral reefs optimization algorithm for solving the unequal area facility layout problem. *Engineering Applications of Artificial Intelligence*, 89:103445, 2020.

[GMLH08] S. García, D. Molina, M. Lozano, and F. Herrera. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. *Journal of Heuristics*, 15(6):617–644, 2008.

[GNL19] H. Gao, H. Nie, and K. Li. Visualisation of pareto front approximation: A short survey and empirical comparisons. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1750–1757, 2019.

[Gro85] B. Grofman. The neglected role of the status quo in models of issue voting. *The Journal of Politics*, 47(1):230–237, 1985.

[GZL+14] J. Guo, J. Zhou, J. Lu, Q. Zou, H. Zhang, and S. Bi. Multi-objective optimization of empirical hydrological model for streamflow prediction. *Journal of Hydrology*, 511:242–253, 2014.

[HGC15] R. Hernández Gómez and C. A. Coello. Improved metaheuristic based on the R2 indicator for many-objective optimization. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, GECCO '15, pp. 679–686, New York, NY, USA, 2015. Association for Computing Machinery.

[HGM+97] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. Dna visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pp. 437–441, 1997.

[HK06] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

[HL13] M. Haenlein and B. Libai. Targeting revenue leaders for a new product. *Journal of Marketing*, 77(3):65–80, 2013.

[HLQS18] H.-H. Hu, J. Lin, Y. Qian, and J. Sun. Strategies for new product diffusion: Whom and how to target? *Journal of Business Research*, 83:111–119, 2018.

[Hol79] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[Hol96] T. Holbrook. *Do campaigns matter?* Contemporary American Politics. SAGE publications, 1996.

[HS15] J. Herrmann and I. Savin. Evolution of the electricity market in germany: Identifying policy implications by an agent-based model. Vfs annual conference 2015 (muenster): Economic development - theory and policy, Verein für Socialpolitik / German Economic Association, 2015.

[HSV12] U. Hakala, J. Svensson, and Z. Vincze. Consumer-based brand equity and top-of-mind awareness: a cross-country analysis. *Journal of Product & Brand Management*, 21(6):439–451, 2012.

[HY16] Z. He and G. G. Yen. Visualization and performance metric in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(3):386–402, 2016.

[Ins85] A. Inselberg. The plane with parallel coordinates. *The visual computer*, 1(2):69–91, 1985.

[ISMN17] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima. Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes. *IEEE Transactions on Evolutionary Computation*, 21(2):169–190, 2017.

[IYM03] H. Ishibuchi, T. Yoshida, and T. Murata. Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling. *IEEE Transactions on Evolutionary Computation*, 7(2):204–223, 2003.

[JJ03] M. A. Janssen and W. Jager. Simulating Market Dynamics: Interactions between Consumer Psychology and Social Networks. *Artificial Life*, 9(4):343–356, 2003.

[JO06] M. A. Janssen and E. Ostrom. Empirically based, agent-based models. *Ecology and Society*, 11(2):37, 2006.

[Ked05] O. Kedar. When moderate voters prefer extreme parties: Policy balancing in parliamentary elections. *American Political Science Review*, 99(2):185–199, 2005.

[Ked09] O. Kedar. *Voting for policy, not parties: How voters compensate for power sharing*. Cambridge University Press, 2009.

[KG17] P. Kerschke and C. Grimme. An expedition to multimodal multi-objective optimization landscapes. In H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme, editors, *Evolutionary Multi-Criterion Optimization*, pp. 329–343, Cham, 2017. Springer International Publishing.

[KMP92] K. Kollman, J. H. Miller, and S. E. Page. Adaptive parties in spatial elections. *American Political Science Review*, 86(04):929–937, 1992.

[KMP98] K. Kollman, J. H. Miller, and S. E. Page. Political parties and electoral landscapes. *British Journal of Political Science*, 28(01):139–158, 1998.

[KO13] W. Kurek and A. Ostfeld. Multi-objective optimization of water quality, pumps operation, and storage sizing of water distribution systems. *Journal of Environmental Management*, 115:189–197, 2013.

[KR03] K.-O. Kim and L. R. Rilett. Simplex-based calibration of traffic microsimulation models with intelligent transportation systems data. *Transportation Research Record*, 1855(1):80–89, 2003.

[KS00] N. Krasnogor and J. Smith. A memetic algorithm with self-adaptive local search: TSP as a case study. In *Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation*, pp. 987–994. Morgan Kaufmann Publishers Inc., 2000.

[KY12] F. Kudo and T. Yoshikawa. Knowledge extraction in multi-objective optimization problem based on visualization of pareto solutions. In *2012 IEEE Congress on Evolutionary Computation*, pp. 1–6, 2012.

[Lav05] M. Laver. Policy and the dynamics of political competition. *American Political Science Review*, 99(02):263–281, 2005.

[Lav11] M. Laver. Why vote-seeking parties may make voters miserable. *Irish Political Studies*, 26(4):489–500, 2011.

[LBG65] P. F. Lazarsfeld, B. Berelson, and H. Gaudet. *The people's choice; how the voter makes up his mind in a presidential campaign.* Columbia Univ. Press, New York, 1965.

[LDV+18a] A. Liefooghe, B. Derbel, S. Verel, M. López-Ibáñez, H. Aguirre, and K. Tanaka. On pareto local optimal solutions networks. In A. Auger, C. M. Fonseca, N. Lourenço, P. Machado, L. Paquete, and D. Whitley, editors, *Parallel Problem Solving from Nature – PPSN XV*, pp. 232–244, Cham, 2018. Springer International Publishing.

[LDV+18b] A. Liefooghe, B. Derbel, S. Verel, M. López-Ibáñez, H. Aguirre, and K. Tanaka. On pareto local optimal solutions networks. In A. Auger, C. M. Fonseca, N. Lourenço, P. Machado, L. Paquete, and D. Whitley, editors, *Parallel Problem Solving from Nature – PPSN XV*, pp. 232–244, Cham, 2018. Springer International Publishing.

[LFLZ+15] J.-S. Lee, T. Filatova, A. Ligmann-Zielinska, B. Hassani-Mahmooei, F. Stonedahl, I. Lorscheid, A. Voinov, J. G. Polhill, Z. Sun, and D. C. Parker. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, 18(4):4, 2015.

[LG04] G. López García. El 11-M y el consumo de medios de comunicación (in Spanish). *Sala de Prensa*, 71, 2004.

[LHKM04] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina. Real-coded memetic algorithms with crossover hill-climbing. *Evolutionary Computation*, 12(3):273–302, 2004.

[LIPS10] M. López-Ibáñez, L. Paquete, and T. Stützle. *Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization*, pp. 209–222. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[Liu07] F. C. Liu. Constrained opinion leader influence in an electoral campaign season: Revisiting the two-step flow theory with multi-agent simulation. *Advances in Complex Systems*, 10(02):233–250, 2007.

[Liu09] Y. Liu. Automatic calibration of a rainfall–runoff model using a fast and elitist multi-objective particle swarm algorithm. *Expert Systems with Applications*, 36(5):9533–9538, 2009.

[LK14] D. Lee and H. Kim. The effects of network neutrality on the diffusion of new internet application services. *Telematics and Informatics*, 31(3):386–396, 2014.

[LKB04] A. V. Lotov, G. K. Kamenev, and V. A. Bushenkov. *Interactive decision maps: approximation and visualization of Pareto frontier.* Springer, 2004.

[LKGG+19] A. LaTorre, M. T. Kwong, J. A. García-Grajales, R. Shi, A. Jérusalem, and J.-M. Peña. Model calibration using a parallel differential evolution algorithm in computational neuroscience: simulation of stretch induced nerve deficit. *Journal of Computational Science*, p. 101053, 2019.

[LM06] I. Lago and J. R. Montero. *The 2004 election in Spain: Terrorism, accountability, and voting.* Institut de Ciències Polítiques i Socials, Barcelona, 2006.

[LMP13] B. Libai, E. Muller, and R. Peres. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *Journal of Marketing Research*, 50(2):161–176, 2013.

[LRS18] F. Lamperti, A. Roventini, and A. Sani. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389, 2018.

[LS07] M. Laver and M. Schilperoord. Spatial models of political competition with endogenous political parties. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1711–1721, 2007.

[LS13] Y. Liu and F. Sun. Parameter estimation of a pressure swing adsorption model for air separation using multi-objective optimisation and support vector regression model. *Expert Systems with Applications*, 40(11):4496–4502, 2013.

[LS18] R. Lehrer and G. Schumacher. Governator vs. hunter and aggregator: A simulation of party competition with vote-seeking and office-seeking rules. *PLOS ONE*, 13(2):1–23, 2018.

[LSG+04] M. J. Landau, S. Solomon, J. Greenberg, F. Cohen, T. Pyszczynski, J. Arndt, C. H. Miller, D. M. Ogilvie, and A. Cook. Deliver us from evil: The effects of mortality salience and reminders of 9/11 on support for President George W. Bush. *Personality and Social Psychology Bulletin*, 30(9):1136–1150, 2004.

[LZ09] H. Li and Q. Zhang. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation*, 13(2):284–302, 2009.

[Mav09] G. Mavrotas. Effective implementation of the ϵ-constraint method in multi-objective mathematical programming problems. *Applied Mathematics and Computation*, 213(2):455–465, 2009.

[MD13] D. Muraro and R. Dilão. A parallel multi-objective optimization algorithm for the calibration of mathematical models. *Swarm and Evolutionary Computation*, 8:13–25, 2013.

[Mic05] N. Michavila. Guerra, terrorismo y elecciones: incidencia electoral de los atentados islamistas en madrid. *Boletín Elcano*, 1(63):34, 2005.

[MiC08] F. Martínez i Coma. *¿Por qué importan las campañas electorales? (in Spanish).* Centro de Investigaciones Sociológicas, 2008.

[Mil98] J. H. Miller. Active nonlinear tests (ANTs) of complex simulation models. *Management Science*, 44(6):820–830, 1998.

[ML09] J. R. Montero and I. Lago. Voto, terrorismo y rendición de cuentas: las elecciones de 2004 en España (in Spanish). *Il cittadino-elettore in Europa e America. Studi in onore di Giacomo Sani. Bologna: Il Mulino*, pp. 137–175, 2009.

[MMNI18] T. Matsumoto, N. Masuyama, Y. Nojima, and H. Ishibuchi. Performance comparison of multiobjective evolutionary algorithms on problems with partially different properties from popular test suites. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 769–774, 2018.

[MN05] C. M. Macal and M. J. North. Tutorial on agent-based modeling and simulation. In *Proceedings of the 37th conference on Winter simulation*, pp. 2–15. ACM, 2005.

[Mon11] J. G. Montalvo. Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. *The Review of Economics and Statistics*, 93(4):1146–1154, 2011.

[Mos89] P. Moscato. On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Technical Report 826, Caltech Concurrent Computation Program, Pasadena, USA, 1989.

[MS00] E. K. Macdonald and B. M. Sharp. Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of Business Research*, 48(1):5–15, 2000.

[MS06] G. Mckeown and N. Sheehy. Mass media and polarisation processes in the bounded confidence model of opinion dynamics. *Journal of Artificial Societies and Social Simulation*, 9(1):11–42, 2006.

[MSEH14] N. Malleson, L. See, A. Evans, and A. Heppenstall. Optimising an agent-based model to explore the behaviour of simulated burglars. In *Theories and Simulations of Complex Social Systems*, pp. 179–204. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[MT17] S. S. Meghwani and M. Thakur. Multi-criteria algorithms for portfolio optimization under practical constraints. *Swarm and Evolutionary Computation*, 37:104–125, 2017.

[Mue73] J. E. Mueller. *War, presidents, and public opinion.* John Wiley & Sons, 1973.

[Mui10] J. Muis. Simulating political stability and change in the netherlands (1998-2002): an agent-based model of party competition with media effects empirically tested. *Journal of Artificial Societies and Social Simulation*, 13(2):4, 2010.

[NBW06] M. Newman, A.-L. Barabási, and D. J. Watts. *The structure and dynamics of networks.* Princeton University Press, 2006.

[NM65] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

[NM12] D. Ngoduy and M. Maher. Calibration of second order traffic models using continuous cross entropy method. *Transportation Research Part C: Emerging Technologies*, 24:102–121, 2012.

[NMA+10] M. J. North, C. M. Macal, J. S. Aubin, P. Thimmapuram, M. Bragen, J. Hahn, J. Karr, N. Brigham, M. E. Lacy, and D. Hampton. Multiscale agent-based consumer market modeling. *Complexity*, 15(5):37–47, 2010.

[NMM06] G. Narzisi, V. Mysore, and B. Mishra. Multi-objective evolutionary optimization of agent-based models: An application to emergency response planning. In *Proceedings of the 2nd IASTED International Conference on Computational Intelligence, CI 2006*, pp. 224–230, 2006.

[Obs04] Observatorio Político Autonómico. Resultados encuesta postelectoral elecciones generales 14 de marzo de 2004 (in Spanish), 2004.

[Oli03] R. Oliva. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151(3):552–568, 2003.

[Olm05] J. A. Olmeda. Miedo o engaño: el encuadramiento de los atentados terroristas del 11-M en Madrid y la rendición de cuentas electoral. *Boletín Elcano*, 1(70):47, 2005.

[OS03] S. Obayashi and D. Sasaki. Visualization and data mining of pareto solutions using self-organizing map. In C. M. Fonseca, P. J. Fleming, E. Zitzler, L. Thiele, and K. Deb, editors, *Evolutionary Multi-Criterion Optimization*, pp. 796–809, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[Pap13] E. I. Papageorgiou. *Fuzzy cognitive maps for applied sciences and engineering: from fundamentals to extensions and learning algorithms*, vol. 54. Springer Science & Business Media, 2013.

[PBP17] P. Perez, A. Baños, and C. Pettit. Agent-based modelling for urban planning current limitations and future trends. In M.-R. Namazi-Rad, L. Padgham, P. Perez, K. Nagel, and A. Bazzan, editors, *Agent Based Modelling of Urban Systems*, pp. 60–69, Cham, 2017. Springer International Publishing.

[PM08] T. Plümper and C. W. Martin. Multi-party competition: A computational model with abstention and memory. *Electoral Studies*, 27(3):424–441, 2008.

[PMN07] A. Pryke, S. Mostaghim, and A. Nazemi. Heatmap visualization of population based multi objective algorithms. In S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, editors, *Evolutionary Multi-Criterion Optimization*, pp. 361–375, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[PS15] A. Piolatto and F. Schuett. Media competition and electoral politics. *Journal of Public Economics*, 130:80–93, 2015.

[PSV01] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200, 2001.

[PYS17] S. Picascia and N. Yorke-Smith. Towards an agent-based simulation of housing in urban beirut. In M.-R. Namazi-Rad, L. Padgham, P. Perez, K. Nagel, and A. Bazzan, editors, *Agent Based Modelling of Urban Systems*, pp. 3–20, Cham, 2017. Springer International Publishing.

[Ran18] D. Randahl. Terrorism and public opinion: The effects of terrorist attacks on the popularity of the president of the united states. *Terrorism and Political Violence*, 30(3):373–383, 2018.

[RART16] M. N. Read, K. Alden, L. M. Rose, and J. Timmis. Automated multi-objective calibration of biological agent-based simulations. *Journal of The Royal Society Interface*, 13(122), 2016.

[RCC20] J. F. Robles, M. Chica, and O. Cordon. Evolutionary multiobjective optimization to target social network influentials in viral marketing. *Expert Systems with Applications*, 147:113183, 2020.

[RHM13] J. Robbins, L. Hunter, and G. R. Murray. Voters versus terrorists: Analyzing the effect of terrorist events on voter turnout. *Journal of Peace Research*, 50(4):495–508, 2013.

[RM89] G. Rabinowitz and S. E. Macdonald. A directional theory of issue voting. *American Political Science Review*, 83(1):93–121, 1989.

[RMA07] W. Rose, R. Murphy, and M. Abrahms. Does terrorism ever work? The 2004 Madrid train bombings. *International Security*, 32(1):185–192, 2007.

[RNI95] S. Russell, P. Norvig, and A. Intelligence. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27, 1995.

[Rog10] E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[RR11] W. Rand and R. T. Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3):181–193, 2011.

[SACM21] A. Sayyed-Alikhani, M. Chica, and A. Mohammadi. An agent-based system for modeling users' acquisition and retention in startup apps. *Expert Systems with Applications*, 176:114861, 2021.

[Sar05] R. G. Sargent. Verification and validation of simulation models. In *Proceedings of the 37th conference on Winter simulation*, pp. 130–143, 2005.

[SBM11] V. K. Singh, S. Basak, and N. Modanwal. Agent based modeling of individual voting preferences with social influence. In *Trends in Computer Science, Engineering and Information Technology*, pp. 542–552. Springer, 2011.

[SDD89] R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt. Network structures in proximity data. In *The Psychology of Learning and Motivation*, vol. 24 of *Psychology of Learning and Motivation*, pp. 249–284. Academic Press, 1989.

[Shi10] T. A. Shimp. *Integrated Marketing Communication in Advertising and Promotion*. South-Western Cengage Learning Australia, 2010.

[SI16] E. Serrano and C. A. Iglesias. Validating viral marketing strategies in Twitter via agent-based social simulation. *Expert Systems with Applications*, 50:140–150, 2016.

[SKM13] Y. Sudo, S. Kato, and A. Mutoh. The impact of exchanging opinions in political decision-making on voting by using multi-agent simulation. In *PRIMA 2013: Principles and Practice of Multi-Agent Systems*, pp. 340–354. Springer, 2013.

[Sör15] K. Sörensen. Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18, 2015.

[SP97] R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[SR14] F. Stonedahl and W. Rand. When does simulated data match real data? Comparing model calibration functions using genetic algorithms. In *Advances in Computational Social Science*, vol. 11 of *Agent-Based Social Systems*, pp. 297–313. Springer, Japan, 2014.

[SRL17] R. Saborido, A. B. Ruiz, and M. Luque. Global WASF-GA: An evolutionary algorithm in multiobjective optimization to approximate the whole Pareto optimal front. *Evolutionary Computation*, 25(2):309–349, 2017.

[SS16] S. Salcedo-Sanz. Modern meta-heuristics based on nonlinear physics processes: A review of models and design procedures. *Physics Reports*, 655:1–70, 2016.

[SS17] S. Salcedo-Sanz. A review on the coral reefs optimization algorithm: new development lines and current applications. *Progress in Artificial Intelligence*, 6(1):1–15, 2017.

[SSCGM+17] S. Salcedo-Sanz, C. Camacho-Gómez, A. Magdaleno, E. Pereira, and A. Lorenzana. Structures vibration control via Tuned Mass Dampers using a co-evolution Coral Reefs Optimization algorithm. *Journal of Sound and Vibration*, 393:62–75, 2017.

[SSCGMH16] S. Salcedo-Sanz, C. Camacho-Gomez, D. Molina, and F. Herrera. A coral reefs optimization algorithm with substrate layers and local search for large scale global optimization. *2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3574–3581, 2016.

[SSCGMP+16] S. Salcedo-Sanz, C. Camacho-Gómez, R. Mallol-Poyato, S. Jiménez-Fernández, and J. Del Ser. A novel Coral Reefs Optimization algorithm with substrate layers for optimal battery scheduling optimization in micro-grids. *Soft Computing*, 20(11):4287–4300, 2016.

[SSDSLT+14] S. Salcedo-Sanz, J. Del Ser, I. Landa-Torres, S. Gil-López, and J. A. Portilla-Figueras. The Coral Reefs Optimization algorithm: A novel metaheuristic for efficiently solving optimization problems. *The Scientific World Journal*, 2014:1–15, 2014.

[SSGHCG+19] S. Salcedo-Sanz, R. García-Herrera, C. Camacho-Gómez, E. Alexandre, L. Carro-Calvo, and F. Jaume-Santero. Near-optimal selection of representative measuring points for robust temperature field reconstruction with the cro-sl and analogue methods. *Global and Planetary Change*, 178:15–34, 2019.

[SSMBV17] S. Salcedo-Sanz, J. Muñoz-Bulnes, and M. J. Vermeij. New coral reefs-based approaches for the model type selection problem: a novel method to predict a nation's future energy demand. *International Journal of Bio-inspired Computation*, 10(3):145–158, 2017.

[Str04] D. Strömberg. Mass Media Competition, Political Competition, and Public Policy. *The Review of Economic Studies*, 71(1):265–284, 2004.

[Str15] D. Strömberg. Media and politics. *Annual Review of Economics*, 7(1):173–205, 2015.

[STSH10] M. E. Schramm, K. J. Trainor, M. Shanker, and M. Y. Hu. An agent-based diffusion model with consumer and brand agents. *Decision Support Systems*, 50(1):234–242, 2010.

[SV08] L. Sánchez and J. R. Villar. Obtaining transparent models of chaotic systems with multi-objective simulated annealing algorithms. *Information Sciences*, 178(4):952–970, 2008.

[SWM17] W. Samek, T. Wiegand, and K. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *CoRR*, abs/1708.08296, 2017.

[Sys89] G. Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pp. 2–9, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[Tal09] E.-G. Talbi. *Metaheuristics: from design to implementation.* John Wiley & Sons, 2009.

[tBvVL16] G. ten Broeke, G. van Voorn, and A. Ligtenberg. Which sensitivity analysis method should I use for my agent-based model? *Journal of Artificial Societies & Social Simulation*, 19(1):5, 2016.

[TCP+18] K. Trawiński, M. Chica, D. P. Pancho, S. Damas, and O. Cordón. moGrams: A network-based methodology for visualizing the set of nondominated solutions in multiobjective optimization. *IEEE Transactions on Cybernetics*, 48(2):474–485, 2018.

[TD18] A. K. A. Talukder and K. Deb. A topologically consistent visualization of high dimensional pareto-front for multi-criteria decision making. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1579–1586, 2018.

[TD20] A. K. A. Talukder and K. Deb. PaletteViz: A visualization method for functional understanding of high-dimensional Pareto-optimal data-sets to aid Multi-Criteria Decision Making. *IEEE Computational Intelligence Magazine*, 15(2):36–48, 2020.

[TF] T. Tušar and B. Filipič. Visualizing exact and approximated 3D empirical attainment functions. *Mathematical Problems in Engineering*, 2014:18.

[TF13] R. Tanabe and A. Fukunaga. Success-history based parameter adaptation for differential evolution. In *2013 IEEE Congress on Evolutionary Computation*, pp. 71–78, 2013.

[TF14] R. Tanabe and A. S. Fukunaga. Improving the search performance of shade using linear population size reduction. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1658–1665, 2014.

[TF15a] T. Tušar and B. Filipič. Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. *IEEE Transactions on Evolutionary Computation*, 19(2):225–245, 2015.

[TF15b] T. Tušar and B. Filipič. Visualization of pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. *IEEE Transactions on Evolutionary Computation*, 19(2):225–245, 2015.

[TKG14] J. C. Thiele, W. Kurth, and V. Grimm. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using Netlogo and 'R'. *Journal of Artificial Societies and Social Simulation*, 17(3):11, 2014.

[TMKM09] L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina. A preference-based evolutionary algorithm for multi-objective optimization. *Evolutionary computation*, 17(3):411–436, 2009.

[TR04] M. Torcal and G. Rico. The 2004 Spanish General Election: In the Shadow of Al Qaeda. *South European Society and Politics*, 9(3):107–121, 2004.

[TR18] A. Trivedi and S. Rao. Agent-based modeling of emergency evacuations considering human panic behavior. *IEEE Transactions on Computational Social Systems*, 5(1):277–288, 2018.

[TRJ13] M. Trusov, W. Rand, and Y. V. Joshi. Improving prelaunch diffusion forecasts: Using synthetic networks as simulated priors. *Journal of Marketing Research*, 50(6):675–690, 2013.

[VB07] J. J. Valdes and A. J. Barton. Visualizing high dimensional objective spaces for multi-objective optimization: A virtual reality approach. In *2007 IEEE Congress on Evolutionary Computation*, pp. 4199–4206, 2007.

[vdH19] S. van der Hoog. Surrogate Modelling in (and of) Agent-Based Models: A Prospectus. *Computational Economics*, 53(3):1245–1263, 2019.

[Ver05] M. Vermeij. Substrate composition and adult distribution determine recruitment patterns in a caribbean brooding coral. *Marine Ecology Progress Series*, 295:123–133, 2005.

[VKM+16] A. Voinov, N. Kolagani, M. K. McCall, P. D. Glynn, M. E. Kragt, F. O. Ostermann, S. A. Pierce, and P. Ramu. Modelling with stakeholders–next generation. *Environmental Modelling & Software*, 77:196–220, 2016.

[VNKT19] V. Volz, B. Naujoks, P. Kerschke, and T. Tušar. Single- and multi-objective game-benchmark for evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, pp. 647–655, New York, NY, USA, 2019. Association for Computing Machinery.

[WA08] R. Willer and N. Adams. The threat of terrorism and support for the 2008 presidential candidates: Results of a national field experiment. *Current Research in Social Psychology*, 14(1):1–22, 2008.

[Wal15] D. J. Walker. Visualising multi-objective populations with treemaps. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, GECCO Companion '15, pp. 963–970, New York, NY, USA, 2015. Association for Computing Machinery.

[Wal18] M. M. Waldrop. Free agents. *Science*, 360(6385):144–147, 2018.

[WC20] D. J. Walker and M. J. Craven. Identifying good algorithm parameters in evolutionary multi- and many-objective optimisation: A visualisation approach. *Applied Soft Computing*, 88:105902, 2020.

[WD07] D. J. Watts and P. S. Dodds. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34(4):441–458, 2007.

[WEF10] D. J. Walker, R. M. Everson, and J. E. Fieldsend. Visualisation and ordering of many-objective populations. In *IEEE Congress on Evolutionary Computation*, pp. 1–8, 2010.

[WEF13] D. J. Walker, R. Everson, and J. E. Fieldsend. Visualizing mutually nondominating solution sets in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 17(2):165–184, 2013.

[WH07] F. Wu and B. A. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601, 2007.

[WM97] D. Wolpert and W. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.

[WMS19] G. Wu, R. Mallipeddi, and P. N. Suganthan. Ensemble strategies for population-based optimization algorithms – A survey. *Swarm and Evolutionary Computation*, 44:695–711, 2019.

[WR15] U. Wilensky and W. Rand. *Introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. MIT Press, 2015.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[WS15] M. Wright and P. Sengupta. Modeling oligarchs' campaign donations and ideological preferences with simulated agent-based spatial elections. *Journal of Artificial Societies and Social Simulation*, 18(2):3, 2015.

[YB11] G. Yücel and Y. Barlas. Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review*, 27(2):195–215, 2011.

[YB15] G. Yücel and Y. Barlas. Pattern recognition for model testing, calibration, and behavior analysis. *Analytical methods for dynamic modelers*, pp. 173–206, 2015.

[YL10] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *2010 IEEE International Conference on Data Mining*, pp. 599–608. IEEE, 2010.

[YLLZ13] S. Yang, M. Li, X. Liu, and J. Zheng. A grid-based evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 17(5):721–736, 2013.

[Zad65] L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[Zak14] N. Zakaria. Modeling political belief and its propagation, with Malaysia as a driving context. *Open Journal of Political Science*, 4:58–75, 2014.

[ZC15] J. Zhong and W. Cai. Differential evolution with sensitivity analysis and the Powell's method for crowd model calibration. *Journal of Computational Science*, 9:26–32, 2015.

[ZK04] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Parallel Problem Solving from Nature - PPSN VIII*, pp. 832–842, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[ZKZ11] M. A. Zaffar, R. L. Kumar, and K. Zhao. Diffusion dynamics of open source software: An agent-based computational economics (ACE) approach. *Decision Support Systems*, 51(3):597–608, 2011.

[ZKZ19] M. A. Zaffar, R. L. Kumar, and K. Zhao. Using agent-based modelling to investigate diffusion of mobile-based branchless banking services in a developing country. *Decision Support Systems*, 117:62–74, 2019.

[ZL07] Q. Zhang and H. Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

[ZLT01] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK-report*, 103, 2001.

[ZSL10] X. Zhang, R. Srinivasan, and M. V. Liew. On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrological Processes*, 24(8):955–969, 2010.

[ZST16] Y. Zhang, Q. Shao, and J. A. Taylor. A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model. *Journal of Hydrology*, 538:802–816, 2016.

[ZSVL08] X. Zhang, R. Srinivasan, and M. Van Liew. Multi-site calibration of the SWAT model for hydrologic modeling. *Transactions of the ASABE*, 51(6):2039–2049, 2008.

[ZT99] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.

[ZTJ15] X. Zhang, Y. Tian, and Y. Jin. A knee point-driven evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 19(6):761–776, 2015.

[ZTL+03] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

[ZVNGNAM17] C. Zambrano-Vega, A. J. Nebro, J. García-Nieto, and J. F. Aldana-Montes. Comparing multi-objective metaheuristics for solving a three-objective formulation of multiple sequence alignment. *Progress in Artificial Intelligence*, 6(3):195–210, 2017.

[ZZWZ13] X. Zhang, J. Zhu, Q. Wang, and H. Zhao. Identifying influential nodes in complex networks with community structure. *Knowledge-Based Systems*, 42:74 – 84, 2013.

"I've seen things you people wouldn't believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in rain. Time to die."

– Rutger Hauer (Roy), *Blade Runner*