Universidad de Granada

Knowledge representation in the domain of environmental dynamics

Pedro Javier Magaña Redondo

Advisor: Miguel Ortega Sánchez

Programa de Doctorado de Dinámica de Flujos Biogeoquímicos y sus Aplicaciones

January 2015

Editor: Universidad de Granada.Tesis Doctorales
Autor: Pedro Javier Magaña Redondo
ISBN: 978-81-9125-058-6
URI: http://hdl.handle.net/10481/39954

El doctorando **Pedro Javier Magaña Redondo** y el director de la tesis **Miguel Ortega Sánchez** 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.

Granada, 14 de enero de 2015

Director de la Tesis	Doctorando
Fdo.: Miguel Ortega Sánchez	Fdo.: Pedro Javier Magaña Redondo

Abstract

The management of environmental areas, with a particular focus in coastal and marine zones, has become as an essential mechanism in ensuring sustainable development. These areas has proved to be not only an invaluable ecological and economic asset but also especially fragile because of the artificial and human activities taking place in them. The degeneration of these areas has increased dramatically in recent years. These changes had created the need to establish effective new forms of organization to ensure the sustainability of these areas.

Artificial intelligence techniques could contribute significantly to the development of a novel approach of integral management within environmental dynamic field. Some of these methods had been applied successfully in other scientific fields. Nevertheless, their application is significantly lower in environmental works. In particular, knowledge representation may exhibit an extraordinary and valuable resource to structure complex systems.

Knowledge representation is a subfield of artificial intelligence which main aim is modeling a domain of discourse in an efficient and computer understandable manner. Knowledge representation field emerged from the need of enhancing tradicional procedural methods, which had been proven insufficient to solve certain complex problems using computers. Although logicians was already confronted with the problem of formalizing knowledge declaratively before the inception of computers, their efforts were focused on mathematical knowledge. Non-mathematical knowledge representation is a scientific field based on more expressive logics.

With the emergence of computers and the evolution of the non-mathematical knowledge representation, another artificial intelligence subfield called **automatic reasoning** began to assume greater importance. Automatic reasoning allows inferring non-explicit knowledge, make new assertions or verify the consistency of declared facts. At the same time, knowledge bases size was reduced because it was no longer needed to represent the whole knowledge explicitly.

Nevertheless, many works within dynamic environment field start frequently from previous research, or their data source is not public available. For these scenarios it is not possible to either formalize knowledge nor apply automatic reasoning. Thus, **knowledge discovery** methods need to be employed. Knowledge discovery usually involves several steps. The first one is selecting and pre-processing data. After that, the core step is taken place, which is generally called **data mining**. Eventually, results are validated and interpreted.

In other circumstances, data has to be obtained from non-numerical sources. Techniques widely applied in environmental dynamics like remote sensing or video-monitoring frequently need an additional feature extraction step. In these situations, another subfield of the artificial intelligence, the **computer vision**, becomes relevant to acquire new kind of data.

While the application of these methods can be very beneficial for environmental dynamic field, the selection of the suitable technique and its implementation is out of the work scope of most environmental engineers. It is essential to carry out an interdisciplinary work to overcome these difficulties and make artificial intelligence tools easily accesible for environmental researchers.

The main contribution of this Thesis is to provide a framework to environmental researchers who want to apply artificial intelligence techniques to their works. This Thesis was conducted in collaboration with several research groups. The work was mainly done at the multidisciplinary research group of Environmental Fluid Dynamics of the University of Granada. The main field of expertise of the group is the integral management of natural resources and the related infrastructures for their exploitation, with specific focusing on coastal areas, ports and river basins. Those topics are investigated through a combined methodology of theoretical analysis, numerical modeling, measurements both at the field and at the laboratory, and data from different sources. This group cooperated with the LexiCon Research Group to produce a lexical environmental resource called EcoLexicon, which is the basis of some of the chapters in the Thesis. Finally, as a result of a collaboration with the Terrestrial Ecology Research Group, an ecological indicator system in Sierra Nevada was also developed. In summary, the data used in this Thesis come from a variety of projects and works developed in collaboration with theses research groups.

Resumen

La gestión de áreas ambientalmente sensibles, con particular interés en la zonas costeras y marítimas, se ha convertido en mecanismo esencial para asegurar un desarrollo sostenible. Se ha comprobado que estas áreas no solo constituyen un importante activo ecológico y económico, sino que además son especialmente frágiles a las actividades artificiales y humanas que en esas se desarrollan. El deterioro sufrido por estas zonas se ha visto incrementado de forma preocupante en los últimos años. Estos cambios han creado la necesidad de establecer nuevas formas efectivas de organización para garantizar la sostenibilidad de estas zonas.

Las técnicas de **inteligencia artificial** pueden contribuir de forma significativa al desarrollo de nuevos enfoques para la gestión integral dentro del campo de la dinámica ambiental. Algunos de estos métodos han sido aplicados en otros campos científicos. Sin embargo, su aplicación es sensiblemente menor en trabajos medioambientales. En particular, la representación del conocimiento puede aportar un extraordinario y valioso recurso para estructurar complejos sistemas.

La representación del conocimiento es un campo de la inteligencia artificial cuyo principal objetivo es modelar un dominio de discurso de una manera eficiente y entendible por un ordenador. La representación del conocimiento surge ante la necesidad de resolver problemas complejos mediante el uso de ordenadores para los que los formalismos convencionales procedimentales no resultan prácticos. Aunque los lógicos ya habían afrontado el problema de formalizar el conocimiento de forma declarativa antes de la invención de los ordenadores, sus esfuerzos se centraban principalmente en la formalización de las matemáticas. La representación del conocimiento no matemático es una rama de estudio basada en la creación de nuevas lógicas más expresivas.

Con la aparición de los ordenadores y la representación del conocimiento no matemático se empieza a avanzar en paralelo en el campo del **razonamiento automático** con gran éxito. El razonamiento automático permite inferir conocimiento no explícito, realizar nuevas afirmaciones sobre el conocimiento o comprobar la consistencia del mismo. De forma complementaria, se reduce el tamaño de la base de conocimiento, ya que no es necesario representar toda la información de forma explícita.

No obstante, en numerosas ocasiones los trabajos dentro del ámbito de la dinámica ambiental parten de trabajos previos, o incluso de datos a los que no se tiene acceso directo a su fuente. En estas situaciones no resulta factible la formalización de dicho conocimiento ni es posible utilizar razonamiento automático, por lo que resulta más interesante la aplicación de técnicas para el **descubrimiento del conocimiento**. Este proceso generalmente incluye una primera fase de selección y preprocesamiento de los datos, una etapa principal consistente en **minería de datos**, y una última fase de validación e interpretación de los resultados.

En otras circunstancias, los datos deben ser obtenidos de fuentes no numéricas. Técnicas ampliamente aplicadas en la dinámica ambiental como la teledetección o la video-monitorización requieren a menudo de un paso extra para extraer las características relevantes de las imágenes. En estas situaciones, otro campo de la inteligencia artificial, la visión artificial, es fundamental a la hora de adquirir nuevos tipos de datos.

Mientras que la aplicación de estas técnicas puede ser muy beneficiosa para el campo de la dinámica ambiental, la selección de la técnica adecuada y su implementación se encuentra a menudo fuera del ámbito de trabajo de la mayoría de los ingenieros medioambientales. Es esencial llevar a cabo un trabajo interdisciplinar para superar estas dificultades y hacer las herramientas de inteligencia artificial accesible para los investigadores medioambientales.

La mayor contribución de esta Tesis es proporcionar un marco de trabajo a ingenieros medioambientales que deseen aplicar técnicas de inteligencia artificial a sus trabajos. Esta Tesis ha sido desarrollada en colaboración con otros grupos de investigación. El trabajo principal ha sido llevado a cabo en el grupo multidisciplinar de Dinámica de Flujos Ambientales de la Universidad de Granada. El principal campo de trabajo del grupo es la gestión integral de los recursos naturales y las infraestructuras relacionadas para su explotación, con especial atención a las zonas portuarias, costeras y fluviales. La investigación es aplicada a través de una metodología que combina análisis teórico, modelado numérico, medidas tanto en campo como en laboratorio, y datos procedentes de diferentes fuentes. Este grupo colaboró con el Grupo de investigación LexiCon para implementar un recurso léxico dentro del dominio del medioambiente llamado EcoLexicon, que es la base para alguno de los capítulos de la Tesis. Finalmente, como resultado de una colaboración con el Grupo de Ecología Terrestre, se desarrollo un sistema de indicadores ecológicos para Sierra Nevada. En resumen, los datos empleados en esta Tesis proceden de una gran variedad de proyectos y trabajos llevados a cabo en colaboración con otros grupos de investigación.

Published papers derived from the Thesis

Imagery for geographic studies

Pedro Magaña, Alejandro López-Ruiz, Andrea Lira, Miguel Ortega-Sánchez, Miguel A. Losada, A public, open Western Europe database of shoreline undulations based on imagery, *Applied Geography*, Volume 55, December 2014, Pages 278–291, ISSN 0143–6228, http://dx.doi.org/10.1016/j.apgeog.2014.09.018.

Ontological indicators system

A.J. Pérez-Luque, R. Pérez-Pérez, F.J. Bonet-García, **P.J. Magaña**, An ontological system based on MODIS images to assess ecosystem functioning of Natura 2000 habitats: A case study for Quercus pyrenaica forests, *International Journal of Applied Earth Observation and Geoinformation*, Available online 26 September 2014, ISSN 0303–2434, http://dx.doi.org/10.1016/j.jag.2014.09.003.

Port management

García-Morales, Rafael; **Magaña, Pedro**; Solari, Sebastián; García-Contreras, Darío; Baquerizo, Asunción; Losada, Miguel. 2011. Optimization of port operational and safety. In *Book of Abstract Risk & Extreme Values in Insurance and Finance*. Risk & Extreme Values in Insurance and Finance. Lisboa, Portugal.

García-Morales, Rafael; **Magaña, Pedro**; Solari, Sebastián; García-Contreras, Darío; Baquerizo, Asunción; Losada, Miguel. 2011. Optimización de la Operatividad y Seguridad Portuaria. En *Libro de Resúmenes de las Jornadas de Costas y Puertos*. Jornadas Españolas de Puertos y Costas. Las Palmas de Gran Canaria, España.

Multilingual knowledge resource on the environment

Faber, Pamela, Ricardo Mairal, and **Pedro Javier Magaña**. 2011. Linking a Domain-Specific Ontology to a General Ontology. In *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*, 564–569.

León Araúz, Pilar, **Pedro Magaña**, and Pamela Faber. 2011. Integrating Environment into the Linked Data Cloud. In *Proceedings of the 25th International Conference Environmental Informatics*. EnviroInfo Ispra 2011, ed. W. Pillman, S. Schade, and P. Smits, 370–379. Shaker Verlag.

León Araúz, Pilar, Pamela Faber, and **Pedro Javier Magaña**. 2011. Linking Domain-Specific Knowledge to Encyclopedic Knowledge: an Initial Approach to Linked Data. In *2nd Workshop on the Multilingual Semantic Web (The 10th International Semantic Web Conference)*. Bonn.

León Araúz, Pilar, and **Pedro Javier Magaña**. 2010. EcoLexicon: contextualizing an environmental ontology. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

Reimerink, Arianne, Pilar León Araúz, and **Pedro Javier Magaña**. 2010.EcoLexicon: an environmental TKB. In *Proceedings of the Seventh conference on International Language Resources*

and Evaluation (LREC'10), ed. Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner,and Daniel Tapias, 2322–2329. Valletta: European Language Resources Association (ELRA).

León Araúz, Pilar, **Pedro Javier Magaña**, and Pamela Faber. 2009. Building the SISE : an environmental ontology. In *European conference of the Czech Presidency of the Council of the EU TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*, ed. J Hrebícek, J Hradec, E Pelikán, O Mírovský, W Pillmann, I Holoubek, and T Bandholtz, 16–23. Brno: Masaryk University.

León Araúz, Pilar, **Pedro Javier Magaña**, and Pamela Faber. 2009. Managing inner and outer overinformation in EcoLexicon: an environmental ontology. In *8ème conférence internationale Terminologie et Intelligence Artificielle*. Toulouse.

Senso Ruiz, José, **Pedro Javier Magaña**, Pamela Faber, and María Amparo Vila Miranda. 2007. Metodología para la estructuración del conocimiento de una disciplina: el caso de PuertoTerm. *El Profesional de la Información 16*, no. 6 (January): 591–604. doi:10.3145/epi.2007.nov.06.

Papers in preparation

A Tintoré, P Magaña, MA Reyes-Merlo and M Ortega-Sánchez, A Dredging Operations Database: How to Create Databases of Civil Engineering Systematic Activities, *Environmental Modelling & Software*

P Magaña, M Ortega-Sánchez, A Moñino and MA Losada. A versatile, low-cost and lightweight research-oriented video-based monitoring system. *Computers & Geosciences*

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Context of the work

The management of environmental areas, with a particular focus in coastal and marine zones, has become as an essential mechanism in ensuring sustainable development. These areas has proved to be not only an invaluable ecological and economic asset but also especially fragile because of the artificial and human activities taking place in them. The degeneration of these areas has increased dramatically in recent years. These changes had created the need to establish effective new forms of organization to ensure the sustainability of these areas.

Artificial intelligence techniques could contribute significantly to the development of a novel approach of integral management within environmental dynamic field. Some of these methods had been applied successfully in other scientific fields. Nevertheless, their application is significantly lower in environmental works. In particular, knowledge representation may exhibit an extraordinary and valuable resource to structure complex systems.

Knowledge representation is a subfield of artificial intelligence which main aim is modeling a domain of discourse in an efficient and computer understandable manner. Knowledge representation field emerged from the need of enhancing tradicional procedural methods, which had been proven insufficient to solve certain complex problems using computers. Although logicians was already confronted with the problem of formalizing knowledge declaratively before the inception of computers, their efforts were focused on mathematical knowledge. Non-mathematical knowledge representation is a scientific field based on more expressive logics.

With the emergence of computers and the evolution of the non-mathematical knowledge representation, another artificial intelligence subfield called **automatic reasoning** began to assume greater importance. Automatic reasoning allows inferring non-explicit knowledge, make new assertions or verify the consistency of declared facts. At the same time, knowledge bases size was reduced because it was no longer needed to represent the whole knowledge explicitly.

Nevertheless, many works within dynamic environment field start frequently from previous research, or their data source is not public available. For these scenarios it is not possible to either formalize knowledge nor apply automatic reasoning. Thus, **knowledge discovery** methods need to be employed. Knowledge discovery usually involves several steps. The first one is selecting and pre-processing data. After that, the core step is taken place, which is generally called **data mining**. Eventually, results are validated and interpreted.

In other circumstances, data has to be obtained from non-numerical sources. Techniques widely applied in environmental dynamics like remote sensing or video-monitoring frequently need an additional feature extraction step. In these situations, another subfield of the artificial intelligence, the **computer vision**, becomes relevant to acquire new kind of data

While the application of these methods can be very beneficial for environmental dynamic field, the selection of the suitable technique and its implementation is out of the work scope of most environmental engineers. It is essential to carry out an interdisciplinary work to overcome these difficulties and make artificial intelligence tools easily accesible for environmental researchers.

Purpose of the research

The main contribution of this Thesis is to provide a framework to environmental researchers who want to apply artificial intelligence techniques to their works. This global aim is structured in different specific targets:

- To differentiate data types employed usually within environmental dynamics field
- To determine knowledge representation and extraction methods more suitable
- To identify additional methods required to build fully functional applications
- To develop case studies prototypes applying the previous techniques to manage the information derived from environmental processes
- To propose a set of recommendations to help develop new applications and enhance previous ones

Outline of the document

The document is structured as follows. Firstly, data types exhibits in environmental dynamics are studied and characterized. Subsequently, the foundation of the different artificial intelligence techniques employed is introduced. Afterwards, several study cases are presented together with the methodology applied and its rationale, with an emphasis on the singularities exhibited by the environmental discipline. Finally, the closing part of the document consists in a discussion of the work and a proposal of best practices. The proposal includes recommendations for both knowledge management and publishing data within environmental domain with the objetive of enhance a collaborative knowledge sharing between scientific communities.

Chapter 2

Applications

In the first application, EcoLexicon is presented. EcoLexicon is a **multilingual knowledge resource on the environment** and provides an internally coherent information system covering a wide range of specialized linguistic and conceptual needs. Data in our terminological knowledge base (TKB) are primarily hosted in a relational database which was subsequently linked to an ontology in order to apply reasoning techniques and enhance user queries. The advantages of ontological reasoning can only be obtained if conceptual description is based on systematic criteria and a wide inventory of non-hierarchical relations, which confer dynamism to knowledge representation. Thus, our research has mainly focused on conceptual modeling and providing a user-friendly multimodal interface. The dynamic interface, which combines conceptual (networks and definitions), linguistic (contexts, concordances) and graphical information offers users the freedom to surf it according to their needs. Furthermore, dynamism is also present at the representational level. Contextual constraints have been applied to re-conceptualize versatile concepts that cause a great deal of information overload.

Thereafter, the following application consists on the development of a **methodology to create databases of civil engineering systematic activities**. Many of those activities are constantly taking place around us with an important social, environmental and economical impact. This research provides value added in accounting literature given the scarcity of works dealing with databases or systematic information on dredging activities that can be used to gain deeper insight into management decisions. This can be due to the complexity of the documentation involved in civil engineering projects that are frequently modified during the construction phase. This methodology was specifically apply to study the dredging operations at Andalusian ports during since 1993 until present. The gathered data was finally incorporated into a PostGIS database, which allows both spatial queries and advanced GIS analysis to be carried out.

Subsequently, it is introduced an **ontological system** called Savia that can describe both the ecosystem functioning and the behavior of abiotic factors in a Natura 2000 habitat. In Savia, a methodological approach that combines the use of (satellite) Earth observation with ontologies to monitor Natura 2000 habitats and assess their functioning is implemented. This system is able to automatically download images from MODIS products, create indicators and compute temporal trends for them. We have developed an ontology that takes into account the different concepts and relations about indicators and temporal trends, and the spatio-temporal components of the datasets. All the information generated from datasets and MODIS images, is stored into a knowledge base according to the ontology. Eventually, users can formulate complex questions using a SPARQL end-point.

Finally, the last work explore the potential of using **imagery for geographic studies** implicating large extensions, with a particular emphasis on coastal studies. To that end, we developed a database of identified and characterized shoreline undulations created after reviewing a total of approximately 50,000 km of coast in Western Europe and Northwestern Africa. This database is free, public and available for the scientific community and can be used to gain deeper insight into coastal morphodynamic processes. This work also explored the potential of using data mining techniques for this type of studies. Although the number of sites (294) and the total number of individual undulations (901) identified was not huge, the total amount of information (17 fields for every undulation) was sufficient to start applying data mining to analyze the data. This type of analysis will become more significant if more observations in time are available and if the number of sites or variables increases. With this work we also encourage other researchers to perform similar analyses and improve and/or create new geomorphic large extension databases with public access.

2.1 Visual terminological knowledge base

Introduction

Currently, there are many environmental data sets available; however, they lack homogeneity, since they have been created for very different purposes and in very different formats. In this work we propose a model capable of integrating some of these data sets with EcoLexicon, an e-environment knowledge base. It facilitates learning and communication and is also integrated in an ontological model. This facilitates its integration in the Semantic Web, and, thus supports the development of implementation tools for external initiatives. Our initial proposal within the framework of Linked Data is to integrate EcoLexicon with DBpedia, GeoNames and GEMET.

One of the main goals of the our work is to provide some sort of integrated and shared environmental information space. This can be accomplished with a knowledge framework capable of managing and integrating information from different sources. This task has much in common with the Semantic Web initiative, from which the e-environment community can benefit, especially when it comes to the interoperability of previously developed technologies (León Araúz et al., 2009). Currently, there are many environmental data sets available, such as the General Multilingual Environmental Thesaurus (GEMET), Environmental Applications Reference Thesaurus (EARTh), Umwelt Thesaurus (UMTHES), SilvaTerm, and ontologies, such as EnVo, SWEET, and the Earth and Planetary Ontology. It is thus our challenge to propose a model capable of integrating some of these data sets with EcoLexicon (http://ecolexicon.ugr.es), an e-environment knowledge base.

EcoLexicon: an e-environment resource

EcoLexicon is a multilingual terminological knowledge base (TKB) on the environment. This TKB was initially implemented in Spanish, English, and German. Currently, four more languages are being added: Modern Greek, Russian, French, and Dutch. So far it has a total of 3,271 concepts and 14,644 terms. EcoLexicon can contribute to the development of the SEIS and SISE from both a linguistic and knowledge representation perspective. It is conceived as a knowledge acquisition tool for a wide range of agents involved in environmental communication (i.e. environmental experts, specialized translators, technical writers, lay users, etc.). According to Hrebícek and Pillmann (2009), the aim of the SEIS is to provide environmental information to all citizens. In this sense, EcoLexicon facilitates learning and communication and eliminates terminological confusion. Undoubtedly, one of its main assets is its multilinguality. However, it also includes semantic networks, graphical resources, and contextual information that enhance the representation of conceptual and terminological knowledge. These features also help to raise public awareness of environmental issues and contribute to the standardization of designations in different languages, which also promotes shared knowledge at an international level. EcoLexicon is primarily hosted in a relational database (RDB), but at the same time it is integrated in an ontological model. This facilitates its integration in the Semantic Web, and, thus supports the development of implementation tools for SEIS, the main objective of the SISE (Hrebícek and Pillmann, 2009).

User interface

Each entry in EcoLexicon provides multiple interrelated modules, such as those shown in Figure 2.1.1 for the concept ebb current. Users are not obliged to view all this information at the same time, but can browse through the interface, depending on their needs. Users can perform both single and combined term searches and obtain different results that can be retrieved from the Search results tab. Alternatively, they can also browse the semantic networks. Furthermore, in the left margin of the screen, users can visualize and expand four modules: Terms, Resources, Definition and Domains.

Under the tag *Terms*, different multilingual choices are shown to the users, who can then click on any of them and obtain terminological information, such as whether a linguistic designation is the main entry term or if it is a synonym, acronym or register and style variant. Information is also provided regarding how terms are used in real contexts. Figure 2.1.1 shows that ebb current has a total of 16 different designations in Spanish, English, German, and Greek since all registers and linguistic varieties are accounted for. For instance, in this case, *ebb tide* is a non-technical variant for the concept ebb current. Even in large term bases, multilingual variety is rarely represented in an exhaustive way. However, this type of information is invaluable because not only does it provide users with multiple options for specialized text comprehension and production, but it is also useful for conceptual disambiguation (see section 3).

When users click on the tag *Resources*, they are provided with informative URLs or graphical information (images, graphics, charts, etc.) specifically selected, depending on the information contained in *Definitions*, the next tag. Definitions are modelled in terms of the constraints imposed by the domain-specific conceptual categories and relations in EcoLexicon (Faber et al., 2007). This produces a set of templates based on category membership, which is also shown under the tag *Domains*. Figure 2.1.1 shows that the definition of ebb current has the genus tide,

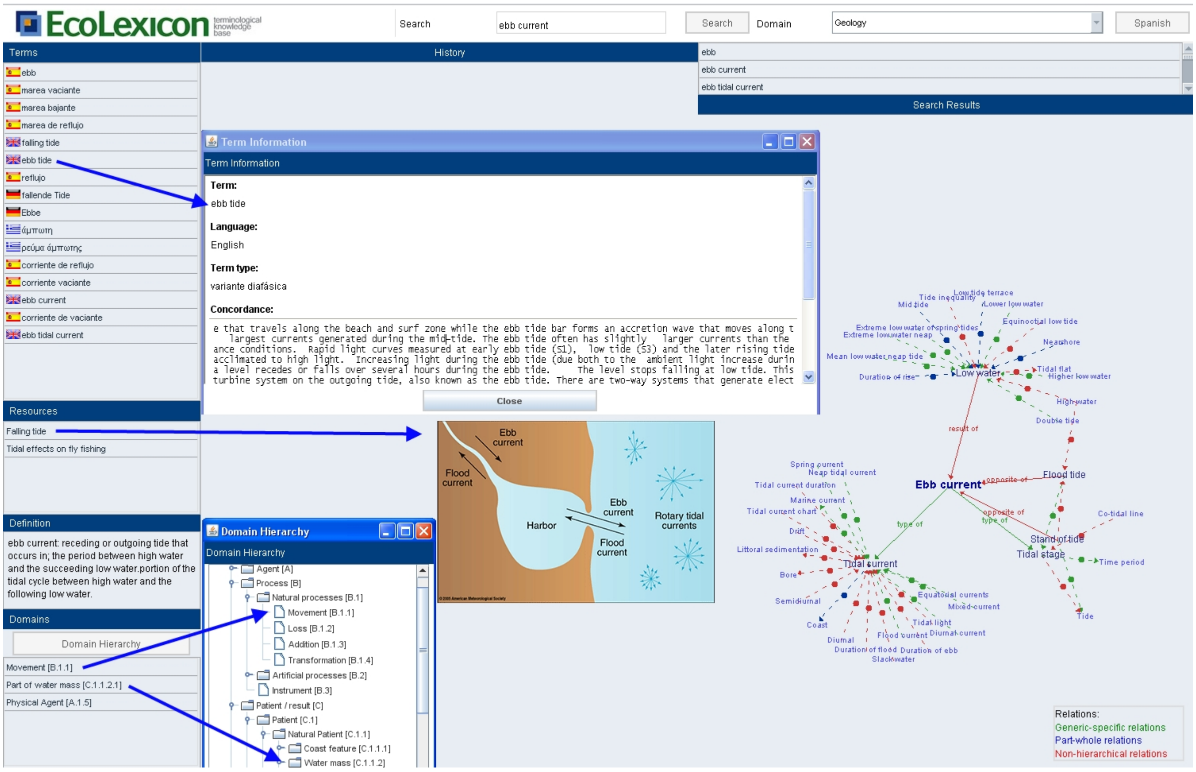

Figure 2.1.1: The EcoLexicon user interface: entry for ebb current

which, because of its multidimensionality, can potentially belong to three domains: movement, part of water mass, and physical agent.

At a more fine-grained level, concepts are displayed in a dynamic network linked to other concepts (right-hand side of the window). As shown in the lower righthand corner, conceptual relations are color-coded and classified into generic-specific, part-whole, and non-hierarchical relations, which contribute to the representation of knowledge natural dynamism (León Araúz and Faber, 2010). Users can click on any of these concepts and thus further expand their knowledge of this domain sector. Nevertheless, problems can arise when it is a question of browsing networks of very general concepts, which are linked to too many other concepts and thus carry an excessive load of information.

The environmental domain has many concepts that can be represented from very diverse perspectives since it is such a vast knowledge area. This is known as multidimensionality (Rogers, 2004; Kageura, 1997) and is commonly regarded as a way of enriching conventional static knowledge representations. However, this can also lead to an information overload that is a serious obstacle to knowledge acquisition. General versatile concepts, such as water, share multiple relations with many other concepts, but they rarely, if ever, activate all those relations at the same time since this would evoke completely different and incompatible scenarios. In this sense, although concepts are entrenched cognitive routines which are interrelated in various ways that facilitate their co-activation, they actually retain sufficient autonomy so that the activation of one does not necessarily entail the activation of all of the rest (Langacker 1987: 162). In line with situated cognition, our claim is that any specialized domain contains sub-domains in which conceptual dimensions become more or less salient, depending on the activation of specific contexts (León Araúz and Faber, 2010; León Araúz et al., 2013). Thus, context is a dynamic construct that triggers or restricts knowledge. For instance, the proposition, water treatment plant *affects* water, would only be informative for users whose knowledge search activated a wastewater scenario, whereas water *causes* erosion would only be useful for users that wished to situate the concept in a geological scenario. It is evident that the same query would never lead to the retrieval of both propositions unless users performed their search in a context-free mode.

The area of environmental knowledge was thus divided into a set of contextual domains (e.g. hydrology, geography, oceanography, civil engineering, environmental engineering, etc.) and the relational power of concepts was constrained accordingly. Contextual domains were allocated in a similar way as in the European General Multilingual Environmental Thesaurus, whose structure is based on themes and descriptors reflecting a systematic, category or discipline-oriented perspective (GEMET 2004).

In EcoLexicon, contextual constraints are neither applied to individual concepts nor to individual relations. Instead, they are applied to conceptual propositions, each of which may be assigned to more than one domain. This constrains versatile concepts, such as water, as well as other concepts that are linked to more general ones. For instance, erosion takes the following shape in a context-free network (Figure 2.1.2), which appears overloaded mainly because it is closely linked to water, one of the most important agents of this process.

When contextual constraints are applied, however, erosion is only linked to other concepts by propositions applicable in geology (Figure 2.1.3) or hydrology (Figure 2.1.4).

Relational database and ontology

EcoLexicon can undoubtedly benefit from the Semantic Web initiative. Ontologies are a powerful representational model since they add the semantic expressiveness lacking in RDBs. This enriches potential queries because reasoning techniques can be applied to extract implicit information. Nevertheless, EcoLexicon, in the same way as many other resources, was initially not conceived as an ontology. This means that legacy systems (RDB stored information) must be linked to an ontological system (León Araúz and Magaña Redondo, 2010). This is not an easy task since both representational models are very different. In contrast to relational databases, ontologies are highly expressive relational structures that strive to describe concepts in very similar terms to those used by humans. EcoLexicon stores semantic information in the ontology, while leaving the rest in the relational database. In this way, we can continue using the new ontological system, while at the same time feeding the database.

As seen in Figure 2.1.5, contextual domains have inspired the design of our ontology classes. The ontology is automatically retrieved from the data stored in our RDB, according to the following assumption: if a concept c is part of one or more propositions allocated to a contextual domain C, c will be an instance of the class C. This contextual category structure makes user queries more dynamic since they can perform different searches through the union and/or intersection of our domains. In this way, they can obtain new but still cognitively-sound knowledge networks. For instance, the intersection of hydrology and geology would restrict the conceptual structure to only hydrogeological propositions, whereas the union of environmental, mining, transport, hydraulic

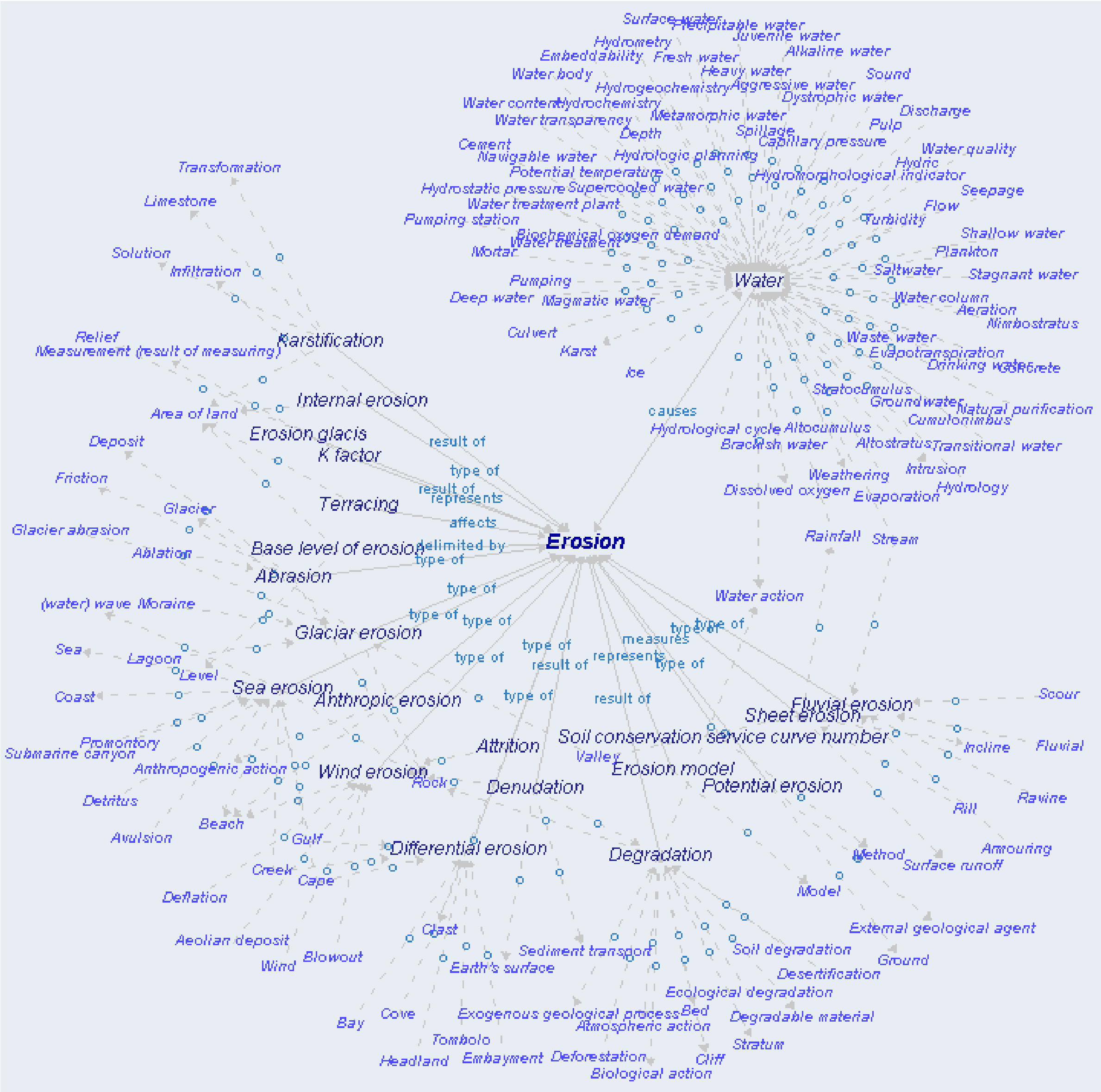

Figure 2.1.2: Erosion context-free network

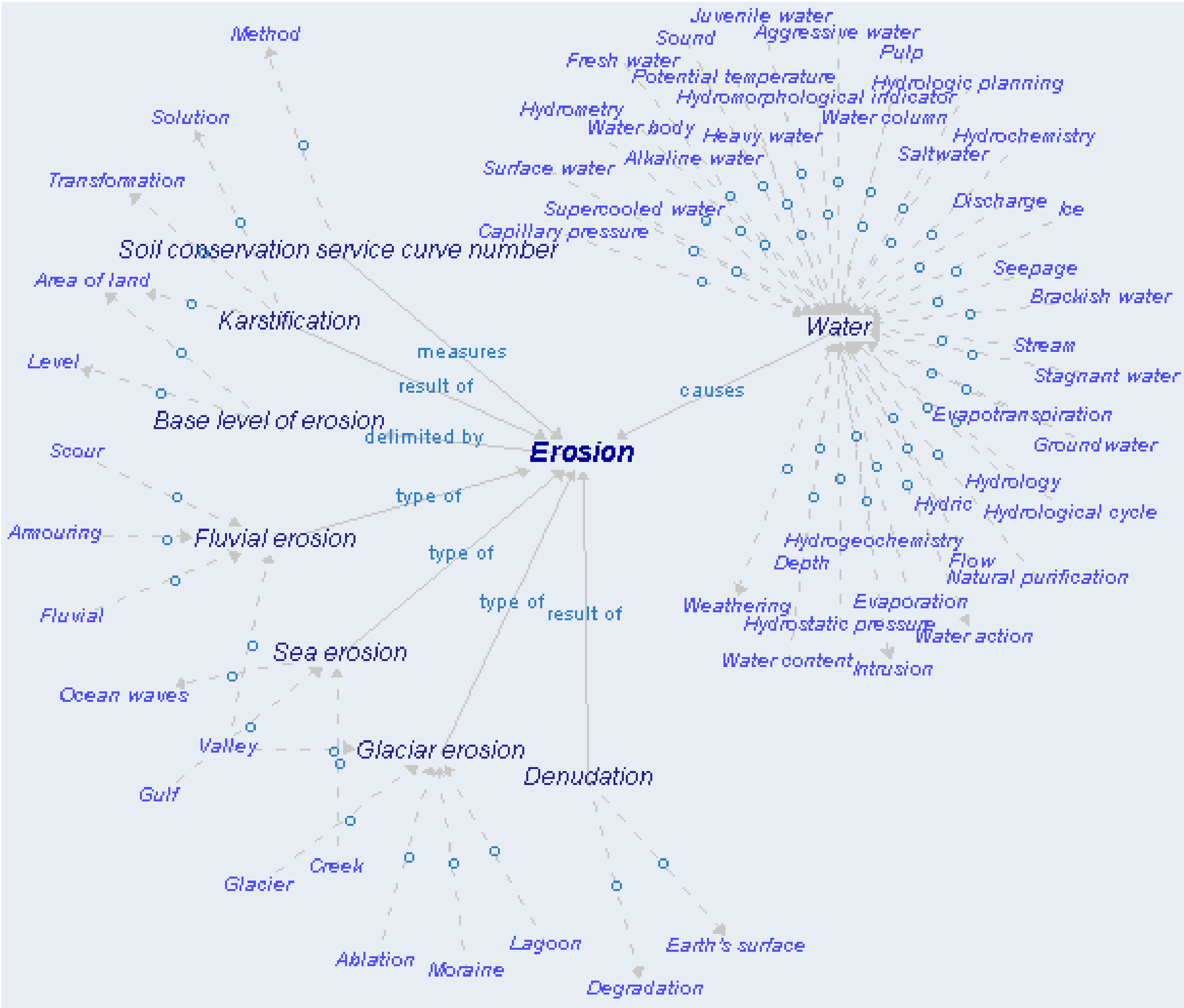

Figure 2.1.3: Erosion in the geology contextual domain

and coastal engineering would make up the whole domain of civil engineering.

Furthermore, our conceptual relations can be enhanced by an additional degree of OWL semantic expressiveness provided by property characteristics, such as transitivity. For this purpose, partonymy has been split up into six different relations (*part_of, composed_of, takes_place_in, phase_of, located_at, delimited_by*) because not all parts interact in the same way with their wholes. For example, if *located_at* were regarded as a *part_of* relation, that would cause fallacious transitivity (Murphy 2003). If a gabion is *part_of* a groyne and a groyne *part_of* the sea, an ontology would infer that gabions are *part_of* the sea, which is not a plausible example. Thus, the only relation that can be considered fully transitive is *part_of*, which is only used for physical objects that are sharply bounded in space. For example, from our ontology we can retrieve berm and beach as two wholes of which berm crest is a part, despite the fact that the only explicit propositions stored in the ontology are berm crest *part_of* berm and berm *part_of* beach. However, transitivity does not extend to higher levels of the ontology since beach is not *part_of* anything but is *located_at* the coast. The reason for this is that evidently not all coasts need beaches to

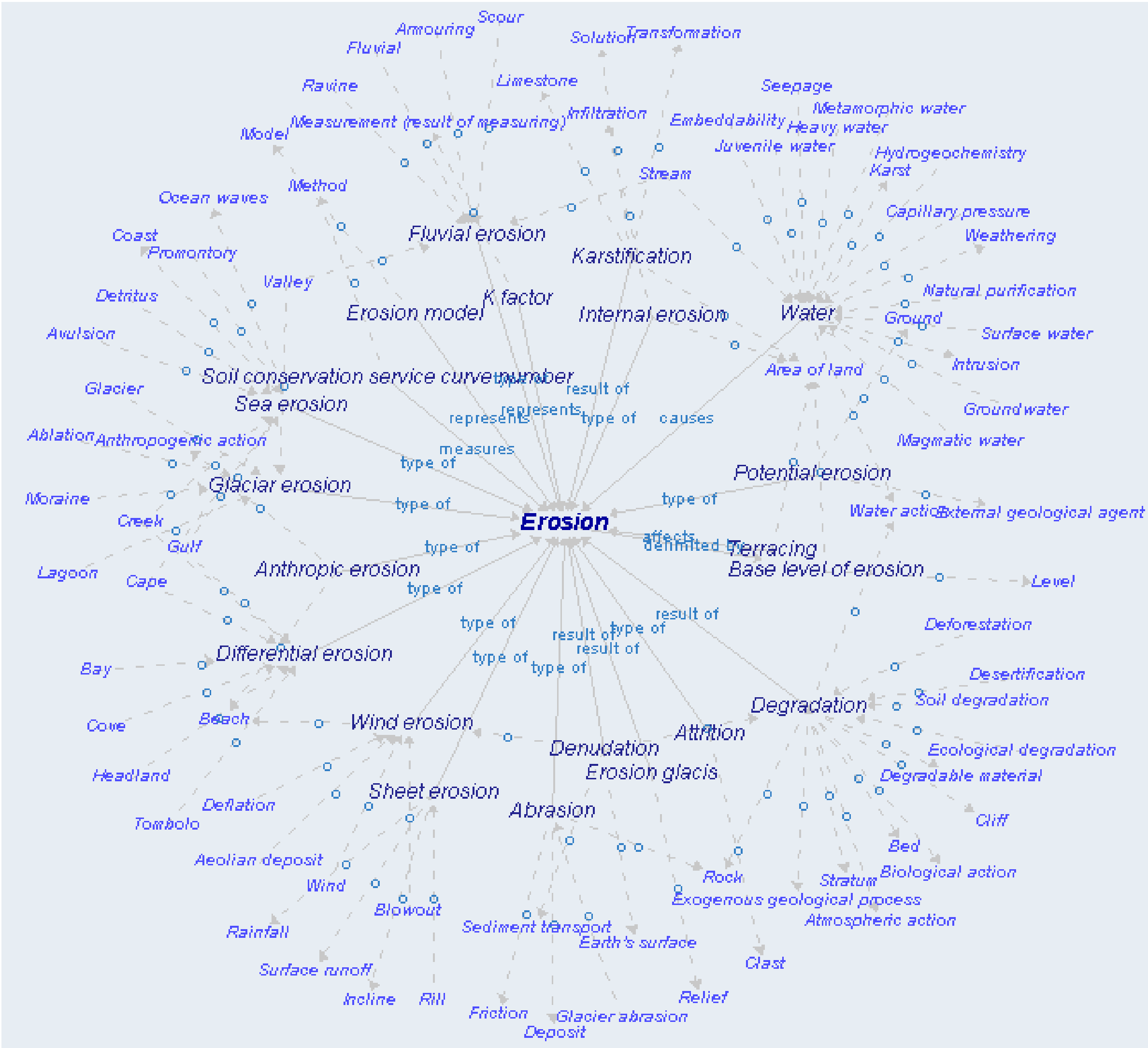

Figure 2.1.4: Erosion in the hydrology contextual domain

exist. In this way, the proposition berm crest *part_of* coast is avoided and is thus not inferred.

Linked data

Once information can be accessed by using ontological resources, it is also easier to connect it with other environmental resources. Even though reusability is often based on data merging, this has the disadvantage of leading to a heterogeneous blending of very diverse data that were originally stored for very different purposes. Various techniques have been proposed for linking data. Previous attempts use automatic mediation algorithms to map and merge ontology schemas (de Brujin et al., 2006). However, an important drawback is that the schemas do not always remain public. However, many systems provide interfaces to interact with their structured data, namely, APIs (Application Programming Interfaces). These interfaces allow developers to combine information from different data sources and create new services known as mashups

Figure 2.1.5: Ontology classes

(Zang et al., 2008). Nevertheless, given the fact that many APIs are proprietary, it is thusnot possible to set links between data objects.

Linked Data (Berners-Lee, 2006) is an innovative approach to this problem. It uses Semantic Web technologies to publish structured data and, at the same time, set links between data items in one data source and data items in other data sources (Heath et al. 2008). This makes it possible to connect data items in different resources, and at the same time keep the resources independent. In our opinion, this methodology can be successfully applied to EcoLexicon to link its data to other semantically-related data sources. This would also enable other environmental resources to enhance their systems with our information, which would help to build a real environmental community of shared data within the Linked Data framework.

Linked data provides a framework to set RDF links pointing to other data sources on the Web. Technically, an external RDF link is an RDF triple in which the subject of the triple is a URI reference in the namespace of one data set, while the predicate and/or object of the triple are URI references pointing to the namespaces of other data sets (Heath and Bizer, 2011). DBpedia is the linked data version of Wikipedia and is at the core of such an initiative, given that it is one of the central knowledge sources of the web. Connecting the concept wastewater from EcoLexicon to that of DBpedia results in the following statement:

```
<http://ecolexicon.ugr.es/resource/wastewater>
<http://www.w3.org/2002/07/owl#sameAs>
<http://dbpedia.org/resource/Wastewater>
```

Integrating EcoLexicon into the linked data cloud

Our initial proposal is to integrate EcoLexicon with DBpedia and GeoNames, which are already part of the Linked Data cloud, and GEMET, the GEneral Multilingual Environmental Thesaurus. The GeoNames database integrates data, such as place names in various languages, geographical features, population, and other related information from various sources (i.e. the National Geospatial-Intelligence Agency, the US Board of Geographic Names, the US Geological Survey Geographic Names Information System, etc.). All the information in GeoNames is organized in nine classes called features. In contrast, GEMET was developed as an indexing, retrieval, and control tool for the European Topic Centre on the Catalogue of Data Sources and the European Environment Agency. One of its aims is to define a core general terminology for the environment in 17 European languages and structure it in a set of semantic fields called themes and groups. We decided to link EcoLexicon to these three resources in order to cover a wide scope: (1) DBpedia is the core of Linked Data; (2) GEMET is more connected to environmental institutions and policies; (3) GeoNames opens up the possibility of exploring a geolocalization line in EcoLexicon.

Linking data sources from DBpedia can be quite straightforward since different tools, such as the ontology editor TopBraid Composer, can automatically suggest links. However, due to lexical variation and the lack of univocity in both general and specialized knowledge, automatic mappings are not always viable. This means that manual work is still necessary and desirable to a certain extent. For example, an automatic mapping of the geological concept cape in EcoLexicon to DBpedia resources could be misleading since *cape* is a highly polysemous term. There are fewer problems related to polysemy in GeoNames and GEMET because all concepts are domain-specific, but lexical variation may impair the string matching process.

Nevertheless, instead of mapping one-to-one manual correspondences, we can take advantage of the semantics contained in our resource. Term strings from EcoLexicon can be compared with those from the other three resources, enhanced by EcoLexicon data sets that include multilingual choices and variants, category membership, and semantic relations, such as *is_a* and *part_of*. As previously mentioned, the use of multilingual choices is a powerful method for conceptual disambiguation. Nevertheless, monolingual variants also ensure a systematic matching procedure since not all concepts are designated by their canonical form. This means that a *sameAs* relationship can still be semi-automatically established even between concepts that are designated by different terms in each data set.

The first step in the data linking process is the comparison of the string of all our English variants with the DBpedia, Geonames, and GEMET entries. Since these strings may match various entries in DBpedia and lead to erroneous mappings, conceptual disambiguation is then performed by comparing other language equivalents. In this way, cape in EcoLexicon would only be linked to the concept in DBpedia with the Spanish equivalent of *cabo* (geographical landform) and not *capa* (clothing article). Nevertheless, in those cases in which polysemy also occurs at a cross-linguistic level, category membership information must be added to the linking algorithm. If any term belonging to the same contextual class of the search concept appears in any of the RDF properties, then concepts are equivalents.

Contextual classes are also used when mapping EcoLexicon concepts to GEMET and GeoNames. Accordingly, our classes were manually compared to those in both resources in order to restrict the number of concepts to be mapped and ensure context-based correspondences. Since our domains were designed in a similar way to those in GEMET, all themes in the latter coincide with one or more of our classes. For example, GEMET's Chemistry theme matches our classes Chemistry and Chemical Engineering. As for GeoNames, we have ruled out some of the nine features, such as city and village since EcoLexicon does not as yet cover instances. Rather our focus is on both natural landforms (peak, cape, beach) and artificial constructions (dam, breakwater, jetty) which, based on the data in EcoLexicon, are the most interesting ones to be shown on a map. In our ontology, natural landforms and artificial constructions may only be contained in five of our classes. Thus, the GeoNames features Spot, building, and farm are mapped onto our civil engineering class, whereas Mountain, hill, rock and stream, lake and undersea are mapped onto geography, geology, oceanography, and hydrology.

Finally semantic relations are also used in GeoNames to display more concepts on the map. For instance, once a concept has been matched in the two resources, all of its wholes (if any) and/or its immediate superordinate concept will also be included by transitivity. This means that not only will the concept dock be shown on the map, but also its hyperonym coastal structure and its whole harbor. Furthermore, certain categories in GeoNames refer to parts of other concepts, such as *section of stream*, *section of lake*, and *section of harbor*. In these cases, the string after *section of* will be compared with the concepts in EcoLexicon. If there is any match, then all of its parts (if any) would be included.

Semantic relations can also be used in the rare cases of ambiguity in GEMET, given that this is the only resource of the three that is hierarchically organized. Apart from the themes and groups, GEMET is organized in broader, narrower and related terms. In those cases where multilingual choices are not sufficient, all related terms can be mapped onto our networks and be used for disambiguation.

The case of spit, bank, and wastewater treatment plant

To illustrate our data linking proposal, we have chosen three concepts: spit, bank, and wastewater treatment plant. Although all of them are included in the three resources, there are certain differences. In EcoLexicon, English and Spanish terms for the concept spit are *spit*, *sand spit*, *offshore beach*, *barrier beach*, *playa barrera* and *cordón litoral*. These lexical variations allow us to identify and link the same concept in the three resources, since in Geonames and DBpedia spit is designated by *spit,*whereas in GEMET, it is designated by *barrier beach*. However, in DBpedia, *spit* is a polysemic term. It may refer to *rain*, *saliva*, *rapping* or a *coastal landform,*among other things. In this case, contextual categories are not necessary for disambiguation, since comparing the property *label* of all possible forms with our Spanish equivalents leads us to the term *cordón litoral*, which is not ambiguous. Thus, the underlying algorithm of this linking rule uses our data sets related to linguistic variations and multilingual choices.

The case of bank is similar to that of spit. Nevertheless, it is necessary to add other parameters to the linking rule since *bank* is also polysemic in Spanish. In the same way as in English, *banco* can refer to a geographic landform or a financial institution, and there are not many other common multilingual equivalents in DBpedia for disambiguation. In DBpedia, this domain-specific entry is named, and differentiated from others, such as bank (geography). In order to match this entry and not any of the others it is necessary to add a context-based rule. Therefore, this match will occur in the following contexts: (1) when the word in brackets matches the string of any of our classes; (2) when any term associated with any concept belonging to the same contextual category as the search concept appears in one or more of the following properties: *dbpedia-owl:abstract*, *dcterms:subject*, *rdfs:comment,* or *dbpedia-owl:wikiPageRedirects*. In this case bank in EcoLexicon belongs to the classes, geography, geology and oceanography, as do many other concepts, such as shoreline, estuary, reservoir, slope, river, marsh, etc, all of which are contained in the properties *dbpedia-owl:abstract*, *dcterms:subject* and *rdfs:comment*. Furthermore, since the disambiguating word in brackets coincides with the EcoLexicon class geography, the second step is not required. In GeoNames, there is only one bank concept. As for GEMET, the concept is also distinguished from others, such as bank (land). Nevertheless, these steps are not required, since the Spanish equivalent *ribera* matches one of the variations in EcoLexicon.

Finally, wastewater treatment plant does not show ambiguity problems because it is a very specialized concept. However, it is a clear example that reinforces the need for term bases to store not only multilingual choices but also monolingual variations. In EcoLexicon there are 17 terms in English, Spanish, and Modern Greek associated with this concept. If lexical variations were not so exhaustively collected in EcoLexicon, we would not be able to link it to the three resources, since each resource gives a different term to designate the concept. In DBpedia, wastewater treatment plant is named *wastewater treatment plant*, the same as in EcoLexicon. However, the term in GEMET is an orthographic variation (*waste water treatment plant*), whereas the GeoNames term (*sewage treatment plant*) clearly reflects register variation.

This is a good example of how linking data does not always ensure knowledge acquisition by the user, since conceptual modeling does not necessarily follow a concrete pattern in all resources. Consequently, there is no assurance that the content is well structured. The definition of *wastewater treatment plant* in DBpedia does not describe the concept at all. In fact, it is wrongly ascribed to a disambiguation category and it redirects users to different types of wastewater treatment. Nevertheless, it does not offer a proper definition of the plant itself The Spanish version of Wikipedia has a good entry for its equivalent (*estación depuradora de aguas residuales*), but there is no link between them. The pseudocode of the different matching algorithms is shown in Table 2.1.1.

Conclusions and future work

Even though Linked Data is conceived as a simple and efficient way for the integration of heterogeneous information in the Web, making links is not a trivial task when it comes to set correspondences between general and specialized knowledge concepts or concepts designated by different linguistic variations. Different EcoLexicon's features, such as, multilinguality, category membership and semantic relations are useful data sets to ensure concept disambiguation and an accurate string matching process.

References

Berners-Lee, T. 2006. Linked Data. *W3C Design Issues*.

De Bruijn, J., Ehrig, M., Feier, C., Martín Recuerda, F., Scharffe, F., and Weiten, M. (2006). "Ontology Mediation, Merging, and Aligning". *Semantic Web Technologies*.

Faber, P., P. León Araúz ,J. A. Prieto Velasco and A. Reimerink. (2007). "Linking Images and Words: the description of specialized concepts (extended version)". *International Journal of Lexicography* 20 (1), 39-65.

Langacker, R. W. 1987. *Foundations of cognitive grammar: theoretical prerequisites*, Vol 1. Stanford: Stanford University Press.

León Araúz, P., Reimerink, A., Aragón, A.G. 2013. "Dynamism and context in specialized knowledge". *Terminology*

León Araúz, P. and P. Faber. 2010. "Natural and contextual constraints for domain-specific relations". In *Proceedings of Semantic relations. Theory and Applications*. Malta.

León Araúz, P. and Magaña Redondo, P.J. (2010). EcoLexicon: contextualizing an environmental ontology. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

León Araúz, P., Magaña Redondo, P. and Faber, P. (2009). Building the SISE: an environmental ontology. In *TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*. Brno: Masaryk University.

Murphy, M. L. 2003. *Semantic Relations and the Lexicon.* Cambridge: Cambridge University Press.

```
for each c:context in ecolexicon
for each t:theme_list in gemet
  if context_matching_gemet(c, t) > context_threshold
    for each w:word in c
    for each cp:concept in t
      all_words = language_variants(w)
      for each v:variant in all_words
        v' = stem(v); cp' = stem(cp)
        if str_compare(v', cp') > word_threshold
          multi_e = multilingual_variants(v)
          multi_g = multilingual_variants(cp)
          if multingual_compare(multi_e, multi_g) > multilingual_threshold
             result.add(pair(v, cp))
          n_terms = narrower_terms(cp); t_terms = type_of_terms(v)
          for each n:n_term in n_terms
          for each t:t_term in t_terms
            if narrower_compare(n, t) > narrower_threshold
              result.add(pair(n, t))
```

```
for each c:context in ecolexicon
for each f:feature in geonames
  if context_matching_geonames(c, f) > context_threshold
    for each w:word in c
    for each cp:concept in f
      w' = stem(w); cp' = stem(cp)
      if str_compare(v', cp') > word_threshold
        result.add(pair(v, cp))
        result.add(make_pairs(super(v), cp))
        result.add(make_pairs(part_of(v), cp))
        result.add(make_pairs(v, part_of(section_of(cp)))
```

```
for each w:word in ecolexicon
for each cp:concept in dbpedia
    w' = stem(w); cp' = stem(cp)
    if str_compare(v', cp') > word_threshold
      multi_e = multilingual_variants(v)
      multi_g = multilingual_variants(cp)
      if multingual_compare(multi_e, multi_g) > multilingual_threshold
        result.add(pair(v, cp))
        related_instances = instances_of(context(v))
        for each i:instance in related_instances
          if look_for_text(comment_properties(cp), i) > text_threshold
            result.add(pair(v, i))
```

Table 2.1.1: Pseudocode

GEMET. 2004. "About GEMET. General Multilingual Environmental Thesaurus". Available at: http://www.eionet.europa.eu/gemet/about

Heath, T. and Bizer, C. 2011. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Hrebícek, J., Pillmann, W. 2009. Shared Environmental Information System and Single Information Space in Europe for the environment: Antipodes or Associates? In *TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*. Brno: Masaryk University.

Zang, N., Rosson, M. B. and Nasser, V. 2008. Mashups: who? what? why? In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, p. 3171-3176.

2.2 How to Create Databases of Civil Engineering Systematic Activities

Introduction

Environmental management has become increasingly complex in recent decades, with higher demands for consideration of multiple economic, environmental and ecological factors. The environmental management industry has become increasingly sophisticated in the use of technology, as more stakeholders become involved, better tools are developed, and requirements for openness and accountability increase (Argent et al., 2009). Many complex environmental problems can benefit from a multi-disciplinary analysis provided by integrated modeling and many researchers indicate the need to solve increasingly complex real-world problems involving the environment and its relationship to human systems and social and economic activities (Whelan et al., 2014).

During last decades there has been an overall shipping expansion with a significant contribution to the worldwide economy. Recent data indicate that around 90% of world trade is being carried by the international shipping industry (Puig et al., 2014). Port facilities have been enlarged with multiple examples in many countries ((Ramos, 2014),(Lee et al., 2014)). For the Spanish case, the main commercial ports renewed or built new infrastructures during the last 20 years and some of them are still under progress. In addition, countries with a high percentage of coastline also develop a strong regional network of fishing ports or leisure harbors. This is the case of the Spanish coast, where more than 200 ports (including commercial, fishing and leisure) are distributed along its 7800 kms of coastline.

Although the growth of maritime transport contribute to the economic development of coastal countries and to provide both direct and indirect employment to the region (Puig et al., 2014), operations and activities at ports may also have adverse consequences on the environment. Many examples have been analyzed in the bibliography (Donnelly et al., 2007), (Lee et al., 2014), (Puig et al., 2014)) with most of them focusing on the water quality (Bedri et al., 2014), gas emissions (L. Goldsworthy and B. Goldsworthy, 2015), air pollution (B. Lin and C.-Y. Lin, 2006) and sediment contamination ((Hedge et al., 2009),(Guerra et al., 2009),(Casado-Martínez et al., 2009)). Particular attention has also been focused to accidental or intentional spills from moving ships since they can be a significant cause of marine pollution and a major source of environmental risk (Mestres et al., 2010). As a result, many ports tried to develop system of indicators for sustainable management ((Darbra et al., 2004), (Peris-Mora et al., 2005), (Donnelly et al., 2007), (Puig et al., 2014)).

Among the different port operations, dredging is of primary importance to overcome the siltation problem with sediments. They consist in removing sediments to maintain the water depths and thus the ship traffic with the maximum safety and operational conditions. Their annual worldwide volumes largely exceeds 600Mm3 every year (Zentar et al., 2009), although growing and land reclamation in developing countries increases annual dredging volumes to 9000Mm3. In 2012 the turnover of the global dredging market was 11.3 billions € (data from the International Association of Dredging Companies, http://www.iadc.com). There are historical evidences showing repeated dredging operations since the third century BC (Marriner and Morhange, 2007) that highlights their importance.

(Darbra et al., 2009) published a study where a total of 26 European ports were interviewed to understand their requirements for environmental information. Despite the diversity amongst European ports and their environmental performances, “dredging“ and “dredging disposal“ were at the second and third positions of the top ten port environmental issues (Darbra et al., 2005). Contemporary ports are required to operate in a more environmentally friendly manner due to the increasing awareness and concerns of their stakeholders (Chang, 2013). Also, nowadays many regulations such as the Water Framework Directive (2000/60/EC), the Priority Substances Directive (Directive 2008/105/EC) on Environmental Quality Standards and the EU Marine Strategy Framework Directive (2008/56/ EC) involve potential constraints on dredging projects (Sheehan and Harrington, 2012). This is even more demanded for leisure harbors (marinas), where being awarded with the Blue flag logo promote their development. Hence, common urgent or regular dredging operations performed every year have to deal with the existing restrictions and regulations (Wasserman et al., 2013).

Dredging activities frequently have a high environmental impact and economical costs (Ramos, 2014). The dredged sediment is generally polluted (Mamindy-Pajany et al., 2010) and should be treated with caution. Awareness relies on the fact that contaminated sediments resuspension may cause the release of otherwise precipitated material in inert chemical forms, and also because of the effects of the hydrodynamics during dredging and in the disposal site (Wasserman et al., 2013). Furthermore, the deepening of channels and construction of ports increases deposition of fine-grained sediments and has as consequence an increase of maintenance dredging and a local and instantaneous increase of suspended particle matter concentration in and around the disposal site (Fettweis et al., 2010). The complexity of this problem was highlighted in the results of the European Sediment Research Network EU funded project that covered many of the problems derived from dredging. Despite of this complexity, recent studies are focusing on the feasibility of producing manufactured topsoil (Sheehan et al., 2010) and other solutions from dredge material to be used in civil engineering (Zentar et al., 2009). (Yozzo et al., 2004) also analyzed the benefits of using the dredge material classified as unsuitable for open-ocean disposal for other uses such as habitat creation.

The region of Andalusia (Southern Spain) has more than 40 ports distributed along the 1100 kms of coastline; the Regional Government directly managed 20 of them (Figure 2.2.1). They are located both at the Atlantic or Mediterranean coast, and some of them are inside estuaries. To conserve the maritime access to them frequent dredging operations are needed. The investment in management dredging is excessive and many environmental impacts are difficult to control. Hence, managers are demanding integral strategies where the benefits can be optimized, the negative impacts minimized and the material re-utilized whenever is possible. However, presently there are not enough processed data to make a diagnosis and analyze possible strategies, as happens in other disciplines such as fishing (Watson et al., 2006), forests (Song, 2013) and soil mapping (Dewitte et al., 2012). Moreover, developing large databases and the movement towards open-source databases is not straightforward (Magaña et al., 2014).

This work provides value added in accounting literature given the scarcity of works dealing with databases or systematic information on dredging activities that can be used to gain deeper insight into management decisions. This can be due to the complexity of the documentation involved in civil engineering projects that are frequently modified during the construction phase. Also, the documentation changes between countries without standards that facilitate the analysis of the information. The main objective of this work is to present the methodology that we developed to create a database of the dredging operations that have been done since 1993 by the Spanish Andalusian Regional Government. Our methodology is straightforward applicable to create similar databases of other civil engineering systematic activities. The database contains more than 80 fields of information and has been implemented in PostgreSQL, using PostGIS extension for spatial data, and therefore GIS integration. This database is free, public and available for the scientific community and can be used to gain deeper insight into dredging processes. This work also presents analyses that can be obtained using the collected data.

Study Area

The region of Andalusia (Southern Spain) has 1100kms of coastline. It is surrounded by the Mediterranean Sea and by the Atlantic Ocean, being the Strait of Gibraltar the connection between them (Figure 2.2.1). Atlantic Andalusian ports are located on Huelva, Sevilla and Cádiz (partially) provinces, whereas the rest face to the Mediterranean Sea (Malaga, Granada and Almeria provinces).

Presently, there are 53 ports along the coastline of Andalucia (Table 2.2.1), averaging 1 port every 21 kms of coastline, and with 30% of them located inside estuaries, rivers and/or river mouths. Their distribution between the Atlantic and the Mediterranean coasts is very similar (47 – 53%, respectively), even though the length of the Mediterranean coast is significantly larger. The morphological complexity of the Mediterranean coast, with many cliffs and a significant proportion of rocky coast (Ortega-Sánchez et al., 2014), partially explains this behavior.

The main commercial operations take place at the 13 ports owned and operated by the Spanish National Government, leaded by the Port of Algeciras located in the Strait of Gibraltar (Figure 2.2.1). The rest are managed directly or indirectly (administrative concession system to a company) by the Regional Government through a public agency called APPA (Agencia Pública de Puertos de Andalucía). These ports are mainly focused in providing leisure or fishing services. APPA functions consist on the maintenance, reparation, facilities and infrastructure expansion and integral management. Our study focuses on the analysis of the 20 ports that are directly managed by the APPA and where at least one dredging operations has been done.

Methodology

This work is based on the compilation, organization and analysis of several fields of information regarding the dredging operations that have been done by the APPA. We developed an easy-to-follow methodology that can be applied to any other type of civil engineering systematic activities (i.e. road maintenance). The resulting dataset was integrated in an open and public database including its incorporation into GIS systems for further analysis. Other complementary data, such as maritime climate information, was also compiled and included in the database.

Phases of the methodology

We have analyzed all the dredging operations that have been performed or are under project phase since 1993 when APPA was created. The information of these operations is contained in more than 90 paper-format projects and other documents that are stored in APPA's offices. A detailed review of every single operation was required because the type and content of the

PROVINCE AND NUMBERING OF PORTS	PORT NAME/HARBOUR NAME	MANAGEMENT	FUNCTION		
			LEISURE	FISHING	COMMERCIAL
HUELVA					
1	Sanlúcar de Guadiana	APPA D	X		
2	Ayamonte	APPA D	X	X	
3	Isla Cristina	APPA D	X	X	
4	Isla Canela	APPA I	X		
5	El Terrón	APPA D	X	X	
6	El Rompido	APPA D	X	refuge	
7	Marina de Nuevo Portil	APPA I	X		
8	Punta Umbría	APPA D	X	X	
9	Huelva	PE	X	X	X
10	Mazagón	APPA D	X		
SEVILLA					
11	Sevilla	PE	X		X
12	Gelves	APPA I	X		
CÁDIZ					
13	Pesquero de Bonanza	APPA D		X	
14	Chipiona	APPA D	X	X	
15	Rota	APPA D	X	X	
16	Sherry	PE	X		
17	La Cabezuela	PE			X
18	Santa María	PE	X	X	X
19	América	APPA D	X		
20	Cádiz	PE	X	X	X
21	de Gallineras	APPA D	X		
22	Sancti Petri	APPA D	X	refuge	
23	Conil	APPA D	X	X	
24	Barbate	APPA D	X	X	X
25	Tarifa	PE		X	X
26	Algeciras	PE	X	X	X
27	La Alcaidesa (Línea de la Concepción)	PE	X		
28	La Atunara	APPA D		X	
29	Sotogrande	APPA I	X		
MÁLAGA					
30	La Duquesa	APPA I	X		
31	Estepona	APPA D	X	X	
32	José Banús	APPA I	X		
33	Deportivo de Marbella	APPA I	X		
34	Marina la Bajadilla	APPA I	X	X	
35	Cabopino	APPA I	X		
36	Fuengirola	APPA D	X	X	
37	Benalmádena	APPA I	X		
38	Málaga	PE	X	X	X
39	El Candado	APPA I	X		
40	Caleta de Vélez	APPA D	X	X	
GRANADA					
41	Punta de la Mona	APPA I	X		
42	Puerto de Motril	PE	X	X	X
ALMERÍA					
43	Adra	APPA D	X	X	
44	Almerimar	APPA I	X		
45	Roquetas de Mar	APPA D	X	X	
46	Aguadulce	APPA I	X		
47	Almería	PE	X	X	X
48	San José	APPA I	X		
49	Comercial de Carboneras	PE			X
50	Pesquero de Carboneras	APPA D	X	X	
51	Garrucha	APPA D	X	X	X
52	Villaricos la Esperanza	APPA D	X	X	
53	Villaricos la Balsa	APPA D	X	X	

Table 2.2.1: Name, ownership and function of the Andalusian ports. Ports managed by the National Government or by the Andalusian Regional Government (directly or indirectly) are denoted as PE, APPA-D and APPA-I, respectively. Numbering is in correspondence with the information included in Figure 2.2.1

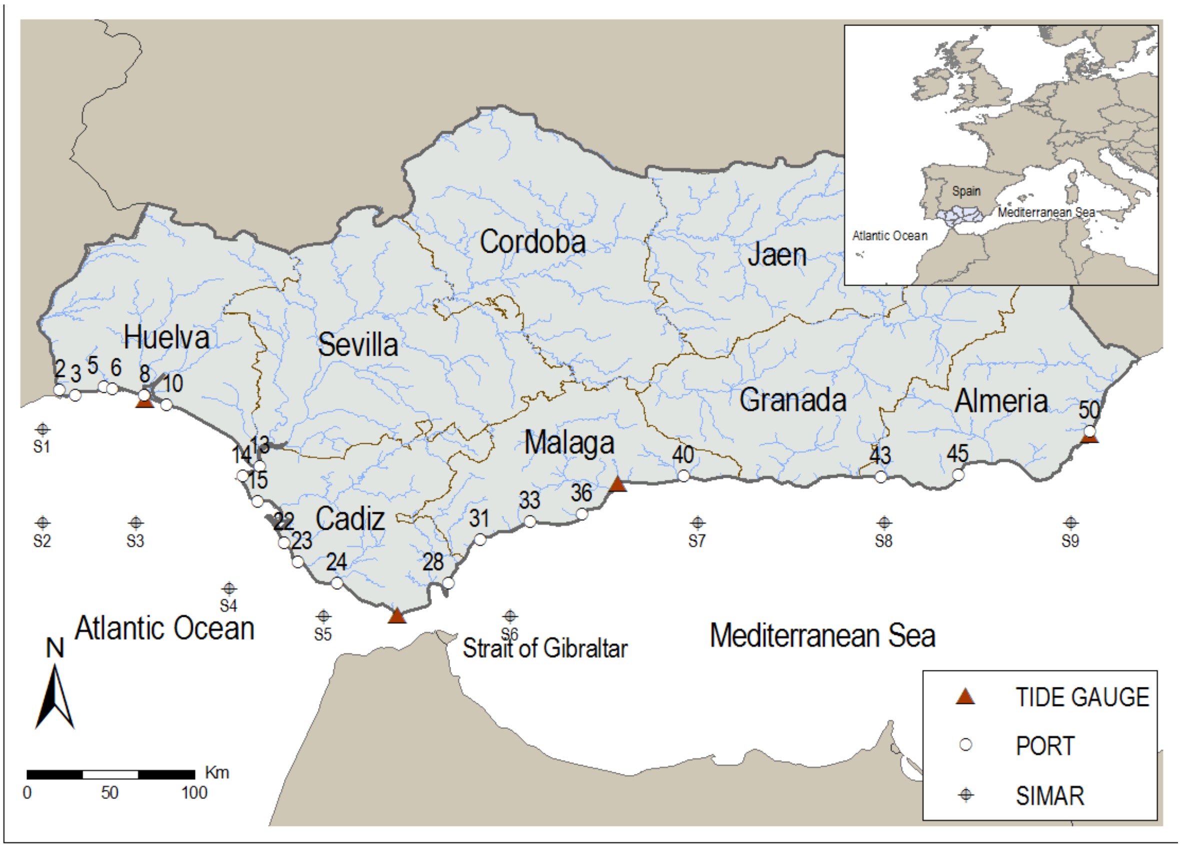

Figure 2.2.1: Study area and location of the ports managed by the Andalusian Regional Government studied in this work. The numbers associated with each port are in correspondence with the information included in Table 2.2.1

documentation vary and change over time due to the increasing technical demands, stricter security measures, technological advances and legislative amendments.

Overall, in our case every dredging activity consists of two main documents, in a similar manner to many other civil engineering activities: (1) a basic design project with the technical detailed information and a corresponding budget, and (2) a construction project containing the revised (constructed) technical information and a lower budget. The decrease in the budget was due to the public procurement system: besides the technical aspects, the lower cost offer is generally selected. Furthermore, although ideally the initial construction budget and the liquidation budget (when the dredging is completed) should be very similar, the truth is that the second frequently increases due to substantial unexpected modifications. Hence, new documents frequently enlarged the documentation of a given project.

To unify, clarify, simplify and collect the information in a unique and interchangeable way, we developed a strict methodology to ensure the consistency of the results (Figure 2.2.2).

The procedure of the methodology (Figure 2.2.2) is as follows:

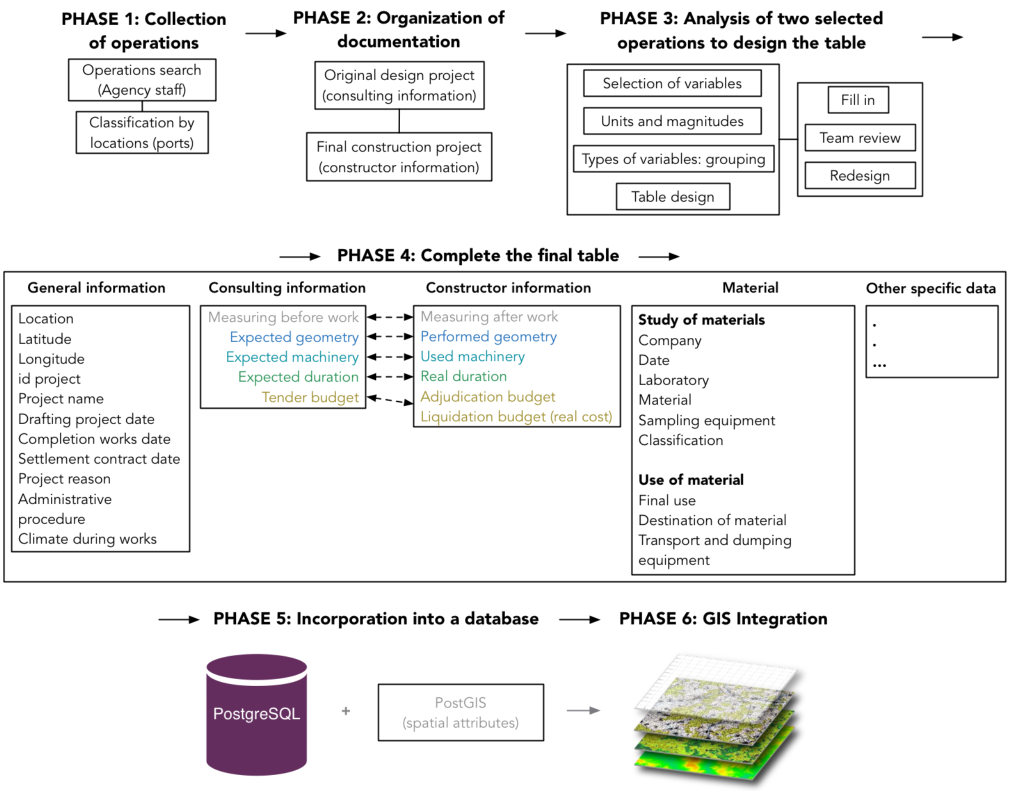

Figure 2.2.2: General scheme of the developed methodology to create databases from systematic civil engineering activities

- *Phase 1: Collection of operations*. Firstly, the Organism in charge of the management of the operations is contacted to collect all the information. Then, the provided information is classified by locations (in our case ports).
- *Phase 2: Organization of the documentation*. For every location, the information is organized in two groups: the consulting (associated to the initial design project of the activity) and the constructor information (associated to the final construction project and executed works). Although ideally both projects should be similar, frequently they significantly differ.
- *Phase 3: Analysis of two selected operations to design the guideline table*. Two specific activities are selected as examples and benchmarks to analyze the data that will be collected and included in the database. They should be relevant operations that contain the main fields of information; the advice of the Organism staff becomes fundamental for this task. Then, the main variables characterizing the activity are selected including their units and orders of magnitude. Next, those variables are grouped and organized into a first Table 2.2.design to easily collect the information. The following step consists on completing the Table 2.2.for the two selected activities and reviewing the results with the rest of the member of the working team. If necessary, the Table 2.2.is redesigned and the procedure repeated until a final Table 2.2.and set of variables are achieved.
- *Phase 4: Complete the final table*. This step consists on collecting the data for all the interventions to complete the table. Although the variables that we included in Figure 2.2.2 are in common with many other civil engineering systematic activities, there could

be other specific or complementary data that should be collected for further analysis of the information and that specifically depends on the type of activity; in this work, this information is the maritime climate that will be described in subsection 3.3. The variables highlighted in color appear both at the consulting information and constructor information; thus, they provide the basis for further comparison and the development of management strategies. Periodic revisions by the working team are necessary to avoid inconsistencies during this step.

- *Phases 5 and 6: Incorporation into a database and GIS integration*. When the information is collected, the final steps consist on the incorporation into a database and the integration into a GIS. They will be further described in the next section.

The time to revise and compile all the information (Phases 1-4) mainly depends on the number of operations and variables, and the format of the information. The development of the database of this work took 5 months; during this time weekly meetings were done to revise the progress of the tasks and to update the methodology. Also, the APPA staff provides a strong support during this time.

Development of the database and GIS integration

The object-relational database PostgreSQL was selected to storage the information. This database management system exhibits several noTable 2.2.features. PostgreSQL can handle spatial attributes such as points, polygons or lines by using the PostGIS extension that makes it possible to undertake spatial queries to the database (Carrera-Hernandez, 2008). In addition, it can be easily linked to the statistical language R (R Development Core Team, 2005) that provides libraries for statistical analysis. Another advantage of PostgreSQL is that it can be linked to the open source GIS Quantum GIS (QGIS) which provides tools for raster, vector and point analysis as well as tools for image processing.

In the present work, the data were incorporated into a PostgreSQL database with the PostGIS spatial extension (Zhang & Yi, 2010). The structure of the database is shown in Figure 2.2.3. QGIS conveniently integrates these representational models and was used to depict the data and develop the different types of maps (QGIS Development Team, 2009). Digital geodatabases allow the accumulation of vast amounts of information that can be readily accessed with simple tools and applied in many fields of applied geography (Jaäppinen, Toivonen & Salonen, 2013; McCool, 2014; Rozenstein & Karnieli, 2011).

Some of these tools have been apply successfully in a previous project (Magaña et al., 2014). This project presents a public, open database with a total of 294 sites showing shoreline undulations identified using Google Earth Imagery along 50,000 km of Western European and Northwestern African coasts. This is significant, because one of the main objectives of this work is to present a methodology employed to develop decision support database for dredging activities, providing a framework that can be applied to any other activities. Besides, this framework should be accessed using software freely available, and open source where possible.

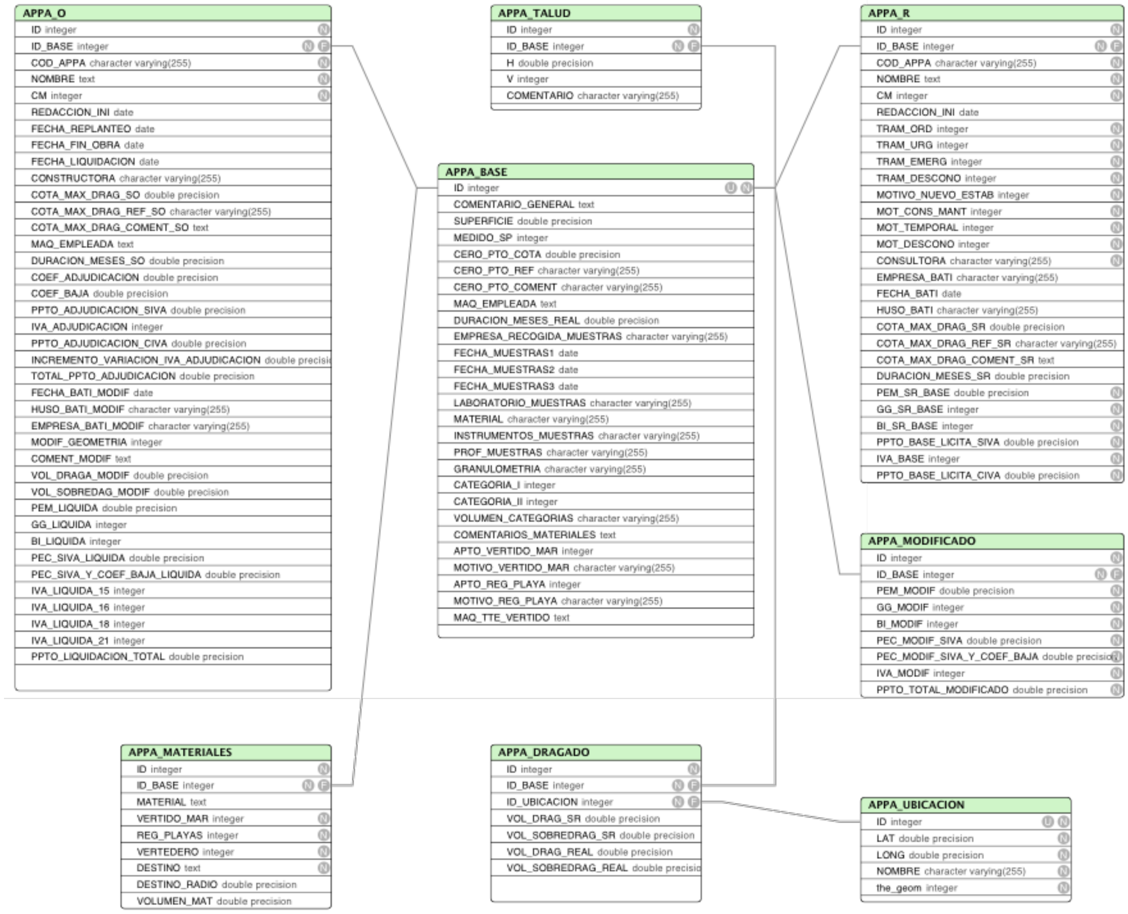

Figure 2.2.3: Relational diagram

Complementary data: maritime climate

Since the dynamics of coastal sediments depends on the maritime climate, we collected wave, wind and tidal data for further analysis. Puertos del Estado (Ministry of Public Works, Spanish Government) has developed one of the most complete worldwide measurement networks for oceanographic and meteorological variables. It includes real time measurements through the most advance instruments such as radar or deep-water buoys, historical data and also simulations and predictions provided by numerical models calibrated and validated for the Spanish coast. Information can be requested at www.puertos.es.

For our study we have selected a set of nine SIMAR points for wave and wind data, and four tidal gauges for sea level (Figure 2.2.1). SIMAR dataset contains daily forecast wave (WAM wave model) and wind data (high-resolution limited-area model, HIRLAM) collected since 1958 and consists of spectral wave height, spectral peak period, mean wave direction, wind velocity and wind direction every 3 h. Complementary information regarding the wave energy flux was obtained from C3E (2014). This information was used to analyze the wave climate and define the predominant forcing conditions, and to establish correlations with the dredging operations.

Results

A total of 70 dredging operations were performed by the APPA from 1993 to 2014. The information included in the global database and the stored geographical information is offered as a PostGIS layer for the scientific community. Although the information is freely available, registration is required for statistical purposes only (http://gdfa.ugr. es/dredging_data).

Following the procedure described in the methodology and after the analysis of some selected projects, a total of 87 terms were defined to concentrate and coordinate the most relevant information provided by each dredging project (Table 2.2.2). Although many of these fields are in common to many other civil engineering systematic activities, some of them may change (i.e. geometrical parameters). Establishing the range of values for each field was proven to be a nice quality control mechanism.

Because information is well-structured and organized in the database, it is possible to query about data using a query language, such as SQL. It allows obtaining a sub group of the information that fits certain criteria. Indeed, data can be gathered and expressed in a summary form for analysis purposes. Thus, the processed information can be used to build trends in data or discover patterns within data.

In the following sections we present some analysis of the information contained in the database to exhibit its potential.

Global analysis

The majority of the interventions were done at the Atlantic ports (43, 61%), whereas the rest corresponds to the Mediterranean's (27, 39%). The port with the highest number of interventions (11) was Conil de la Frontera, followed by Estepona (7). 48% of the interventions are concentrated only on 5 ports (Conil, Barbate, Estepona, Marbella and Fuengirola), whereas at the rest very few operations were done. Focusing on the interventions by provinces, Cádiz exhibits the major problems with at least 2 interventions at each port.

Figures 2.2.4 and 2.2.5 relate the number of interventions with the average wave energy flux and H_{12}, respectively. These variables represent the severity of the mean wave climate. Cádiz province is exposed to the highest wave energy fluxes, whereas the lower values are attained at Granada and Almeria; similar results are obtained for the wave height. Results indicate the existence of a clear trend between the severity of the wave climate and the number of dredging operations: ports exhibiting the higher number of interventions (i.e. Conil and Estepona) are subjected to more energetic waves, whereas the lower number of interventions are attained at low wave energy ports (i.e. Adra and Roquetas de Mar). Although other variables such as the mobility of the sediment fractions are also fundamental to explain the volumes and frequency of the dredging activities, wave climate is playing a key role in the dynamics of the Andalusian ports.

The purpose of the dredging activities was to guarantee the safety access to the ports and to return the bathymetries to their original designs. A total of 38% of the dredging works can be classified as regular maintenance operations, whereas 62% corresponds to urgent interventions. The latest generally took place after significant storms that mobilized a remarkable amount of

	TERM	UNITS	RANGE
General information			
	Location (city, village)	Nominal	-
	Latitude	Coordinates	36.189958-37.22505
	Longitude	Coordinates	-1.899244-7.411683º
	Harbor name	Nominal	-
	id project	-	-
	Project name	Nominal	-
	Drafting project date	-	1993-2014
	Start works date	-	1993-2011
	End works date	-	1993-2011
	Settlement of the contract date	-	1994-2012
	Project purpose	Nominal	-
	Administrative procedure	Nominal	-
	Machinery (used)	Nominal	-
	Duration of the work (real duration)	Months	1-12
Geometry			
	Dredging area	m^2	1900-6937672
	Reference level	-	-
	Gradient of the slope (H-V)	-	0:1-10:1
Consulting information			
	Consultancy firm	Nominal	-
	Machinery (suggested)	Nominal	-
	Expected duration	Months	0,25-12
Previous bathymetry			
	Company	Nominal	-
	Date	-	1992-2013
	Time zone	-	29-30
	Projection system	-	-
Geometry			
	Expected maximum dredging depth	BMVE	-2,5/-10
	Expected dredging volume	m^3	3000-924359,05
	Expected extra dredging volume	m^3	1318,68-45000
Budget			
	Material execution budget	€	7954,46-6053158,45
	Overhead cost	%	13
	Industrial profit	%	6
	Tender budget (before VAT)	€	9465,80-23093074,08
	VAT	%	15-21
	Tender budget	€	10980,33-8355779,93
Constructor information			
	Construction company	Nominal	-
	Machinery used	Nominal	-
	Expected duration	Months	0,25-19
Bathymetry after works			
	Company/enterprise	Nominal	-
	Date	-	1993-2013
	Time zone	-	29-30
	Projection system	-	-
Geometry			
	Maximum dredging depth	BMVE	-2,5/-10
Modifications			
	Geometry and dredging level modifications	Nominal	-
	Real dredging volume	m^3	5040-299328,03
Budget			
	Material execution budget after modifications	€	139916,61-2272680
	Allocation coefficient	-	0,58-1
	Reduce of tender	%	0-100
	Adjudication budget (before VAT)	€	9015,18-5491869,50
	VAT	%	15-21
	Adjudication budget	€	10457,61-6370568,62
	Budget increase	€	0-41784,25
	Adjudication budget (total)	€	10457,61-6370568,62
Final cost			
	Liquidation material execution budget	€	7434,77-2501556,40
	Overhead cost	%	13
	Industrial profit	%	6
	Liquidation budget (before VAT)	€	8426,19-2051014,25
	VAT	%	15-21
	Liquidation budget	€	9774,38-2420196,82
Materi als			
Sediment samples			
	Company	Nominal	-
	Date	-	1993-2013
	Laboratory	Nominal	-
	Material	Nominal	-
	Sampling equipment	Nominal	-
	Sampling depth	Nominal	-
	Granulometry	Nominal	-
	Classification	-	I-II-III*
	Volume according to the classification	m^3	-
Uses of dredged material			
	Final use	Nominal	-

Table 2.2.2: Detailed information included in the dredging database obtained following the methodological procedure summarized in Figure 2.2.2. Maritime climate variables were included at the end of the table

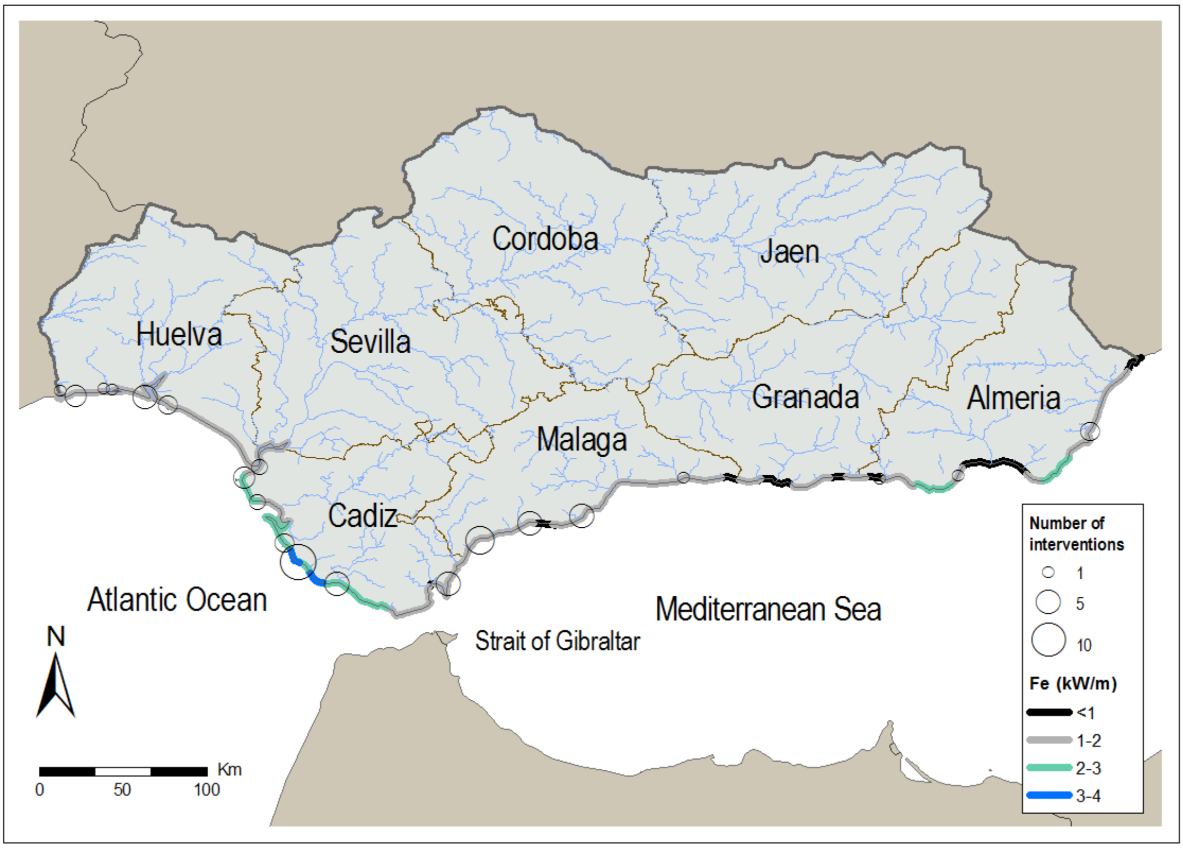

Figure 2.2.4: Relation between the number of interventions and the module of the wave energy flux (kW/m)

sediment. Overall, the dredging depths were lower (below 10m) for all the interventions.

Figure 2.2.6 depicts the monthly averaged dredging volume (ADVm) at every port. This parameter represents the monthly dredging efficiency calculated as the ratio between the real dredging volume and the real duration of the activities: the higher the parameter at a given port, the more efficient the dredging operations. Ports at Cadiz and Huelva provinces exhibit the greatest efficiency, even though the wave climate frequently limits dredging works during many days of the year. This can be due to the lower sediment size and less cohesive material. In fact, when this efficiency is evaluated for the Atlantic (23,000 m^3/month) and Mediterranean ports (13,000 m^3/month), this trend is clearly remarkable. Also, efficiency at Mediterranean ports is in general more homogenous (2,200 m^3/month in Roquetas-17,500m^3/month in Marbella) than at the Atlantic side (6,000m^3/month in El Terrón-43,000 m^3/month in Rota). One factor than partially explains this behavior is that many Atlantic ports (70%) are located inside estuaries or river mouths, whereas at the Mediterranean side coastal ports are predominant.

Volumes and costs

The total amount of sediment mobilized during the 70 interventions is 5,847,593.81 m^3 that gives an average of 83,537 m^3 per intervention. The maximum total volume was dredged at

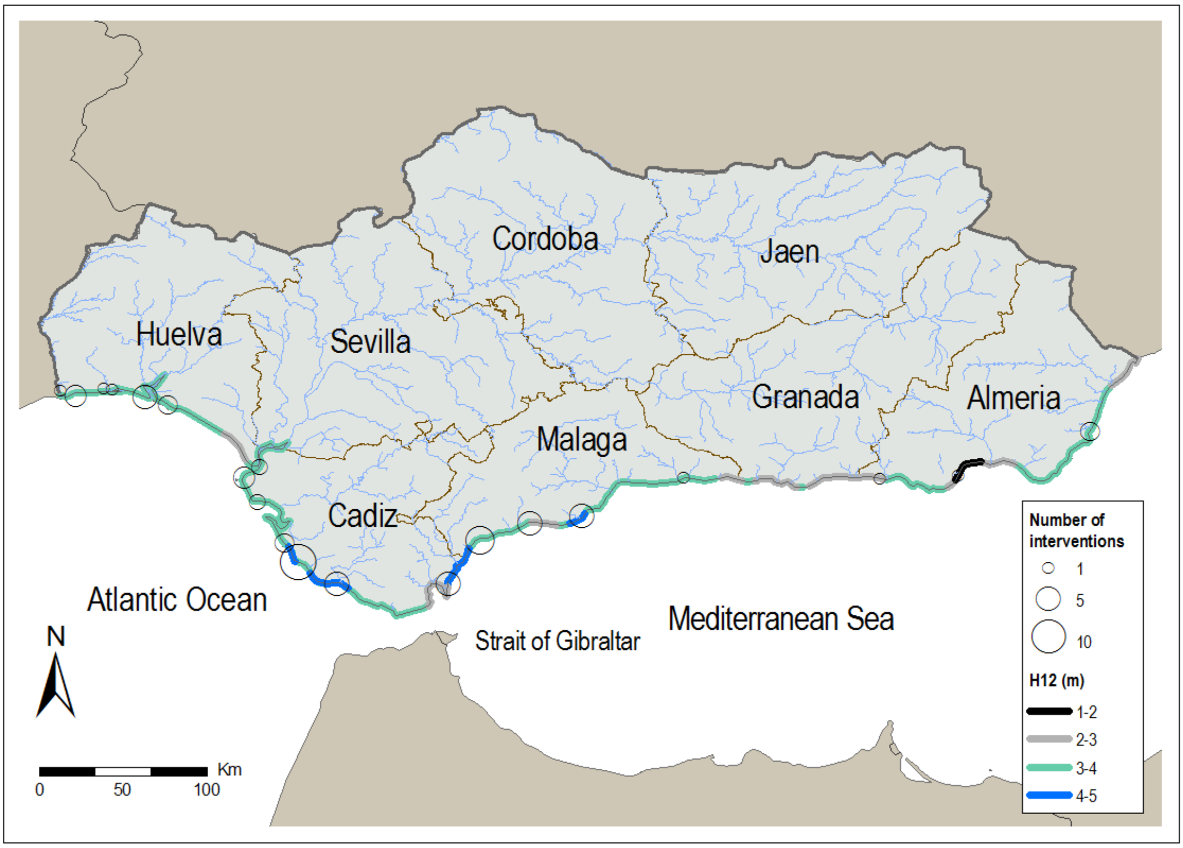

Figure 2.2.5: Relation between the number of interventions and H_{12} (m)

Conil (1,536,380 m^3-26%) followed by Punta Umbria (874,590 m^3-15%), whereas the minimum was dredged in Roquetas de Mar (19,440 m^3- <1%). In average, 2008 was the year in which more volume was dredged (1,160,771.633 m^3). This increase was primarily due to a dredging intervention in Conil (928,033.10 m^3-80%) to reach a deeper bathymetry than usual (-10m).

Figure 2.2.7 depicts the averaged dredging volume (ADVi) per intervention at every port calculated as the ratio between the real dredging volume and the number of interventions. Huelva's values are significantly higher than those for the rest of provinces, with El Rompido exhibiting the maximum value (196,625.9 m^3)by intervention, followed by Punta Umbria (174,917.58 m^3), Ayamonte (162316.4 m^3) and Mazagon (152,655.10 m^3). On the contrary, Roquetas de Mar (19,440 m^3) and Marbella (32,327.99 m^3) have the minimum values. Comparing the Atlantic and Mediterranean ports, the values of ADVi are significantly higher at the former (especially for Huelva's ports), whereas at the later those values are lower but more uniform. In general, the amount of sediment mobilized at the Atlantic side is greater, that allow us to conclude that regarding the volumes, attention should focus on improving the management of these ports.

Regarding the costs of the activities, the total inversion was 39,420,534.10 €. The highest inversion was in Conil (5,397,481.85 €-14%) followed by Punta Umbria (4,558,214.14 €-11%) and the lowest was in Caleta de Velez (448,260.66 €-1%). The largest expenditure in dredging interventions was in 2011, 8,136,443.09 € (20% of the total). This can be related to the number

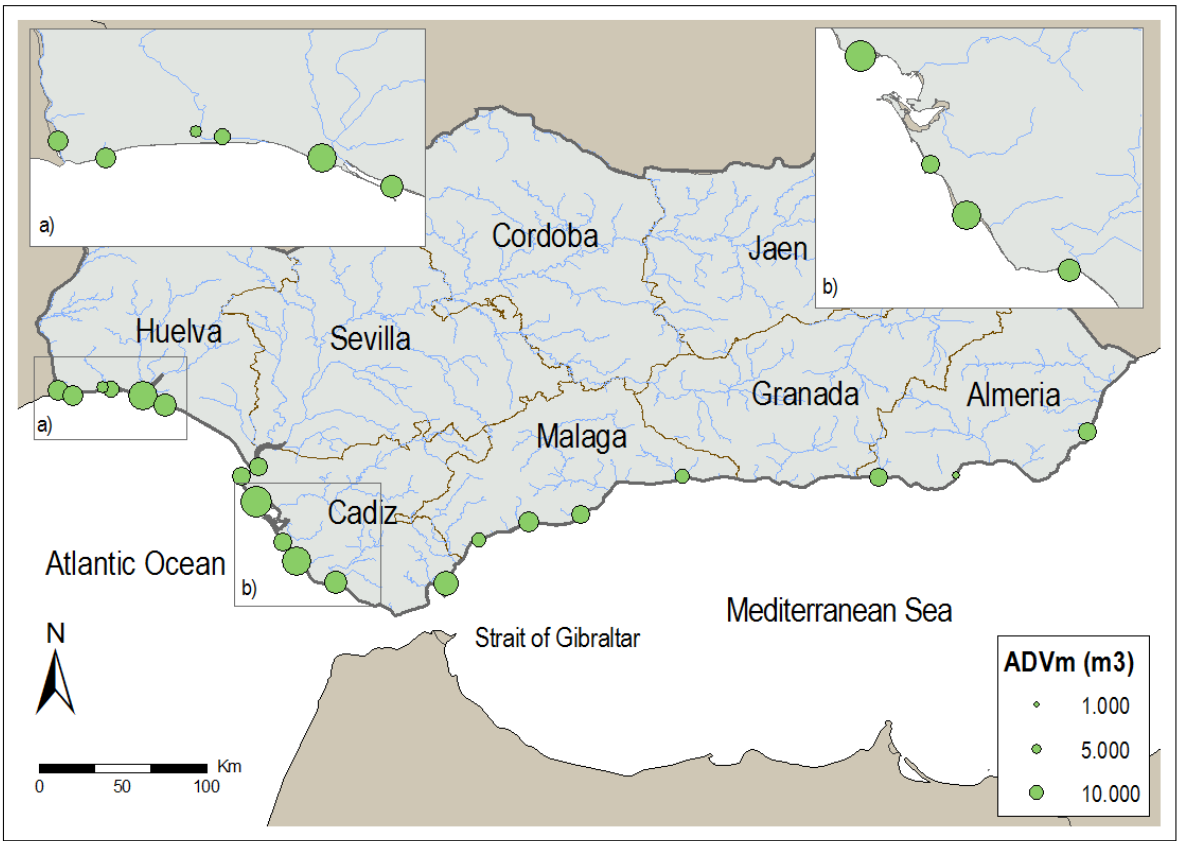

Figure 2.2.6: verage dredging volume (ADVm, m^3) per month

of interventions since many of them started in 2010 (10) but the payment was made during 2011. This total amount destined for dredging interventions is relevant considering the small size of the ports and that all of these interventions were only for maintenance purposes.

Figure 2.2.8 shows the averaged liquidation budget (ALBi) per intervention at every port calculated as the ratio between the liquidation budget and the number of interventions. El Rompido represents the highest value (2,238,577.42 €) followed by Roquetas de Mar (1,525,133.01 €). In both cases costs are increased due to particular situations: El Rompido is located in an environmental protected area that slows down the works. On the other side, Roquetas de Mar has a particular problem with the final destination of dredged material for environmental reasons. By contrast, Fuengirola is the port that shows the lowest value (168,494.60 €). From an economic point of view, results indicate that attention should focus on improving the management of Atlantic ports, which is in agreement with the volumes results. However, the remarkable high ALBi value at Roquetas de Mar (Figure 2.2.8) leads to the necessity of deeper analysis of the dredging activities at this port.

The average cost of dredging 1m^3 of material was calculated as the ratio between liquidation budget and dredging volume in each port (Figure 2.2.9). Roquetas de Mar results by far in the largest value (78.45 €) because material treatments are often necessary before final destination. The second highest value is found at El Terron (18.21 €), whereas the minimum value is obtained

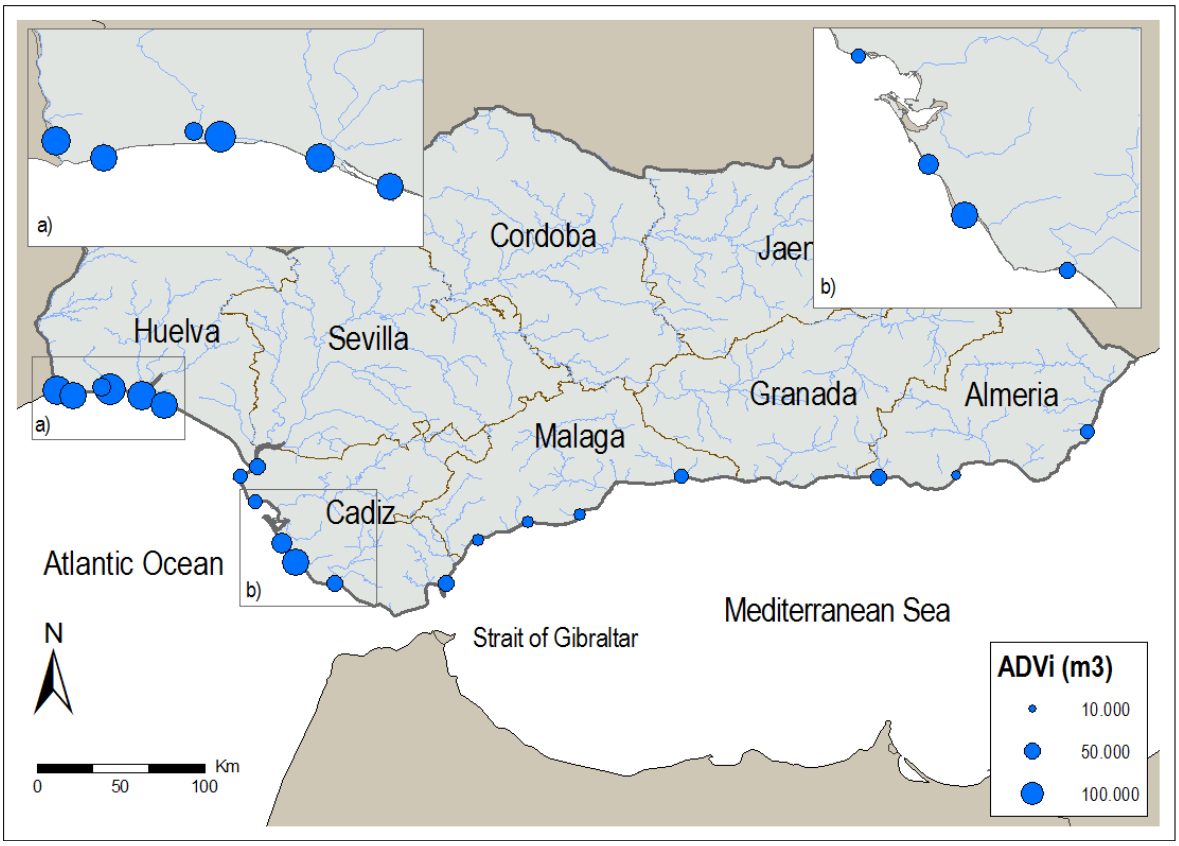

Figure 2.2.7: Average dredging volume (ADVi, m^3) per intervention

in Conil (3.51 €). From a global perspective, while dredging volumes and costs by intervention (Figures 2.2.7 and 2.2.8) are higher in the Atlantic watershed, the mean cost of dredging 1 m^3 is lower (Atlantic - 6.28 €/m^3 vs Mediterranean - 8.28€/m^3). The analysis of the information contained in the database shows the complexity of the management of these types of periodic activities.

Dredged material

One of the major economic engines of Andalusia's region is the tourist sector, and thus the quality of the beaches is a priority in coastal municipalities. Increasing erosion problems during last years can reduce the impact of this activity. The tendency over the years has been trying to harness the dredging material with appropriate quality for regenerating beaches. During the 1990s, reusing the dredging material was practically non-existing; in contrast, in the 90% of the operations carried out during the last five years dredging material was reused to replenish nearby beaches. Furthermore, in 95% of the interventions, the quality of the dredging material is suitable to regenerate beaches (Figure 2.2.10). One factor that is forcing the Administration to reutilize this material is the situation of the global economy, and the reduction in the budget for beach restoration.

Current local regulations are permissive enough to allow regeneration of beaches, except

Figure 2.2.8: Average liquidation budget (ALBi, €) per intervention

where other restrictions apply. This is the case of Carboneras, which is located in a protected area and dumping at both sea and beach regeneration is strongly restricted. Also, materials have to cover long distances because to avoid protected areas. For other cases the materials are sometimes subject to special treatments. This applies to Roquetas de Mar, where material needs to be dried before dumping. In general, all these factors significantly increase the cost of the interventions at those sites (Figure 2.2.9).

Overall tendency of the Andalusian Ports

The number of interventions per year (Figure 2.2.11) does not exhibit a clear trend. Except in 1997, in the rest of the years at least one intervention was done, with average yearly values around 3 being the maximum in 2010 with 10 interventions. The sedimentation occurring in ports are generally quite a constant process unless any major intervention is done, and hence systematic interventions are expected to be quite regular. The variability of this number of interventions is generally due to economic reasons and management decisions (i.e. urgent interventions), and not mainly based on physical (sediment deposition rates) reasons.

Figure 2.2.12 shows the dredging expenditure per year. These data corresponds to the database information but transformed to real prices. 2011 was the year with the higher expenditures in interventions, whereas 2002 was the lower. The trend in this case indicates a progressive

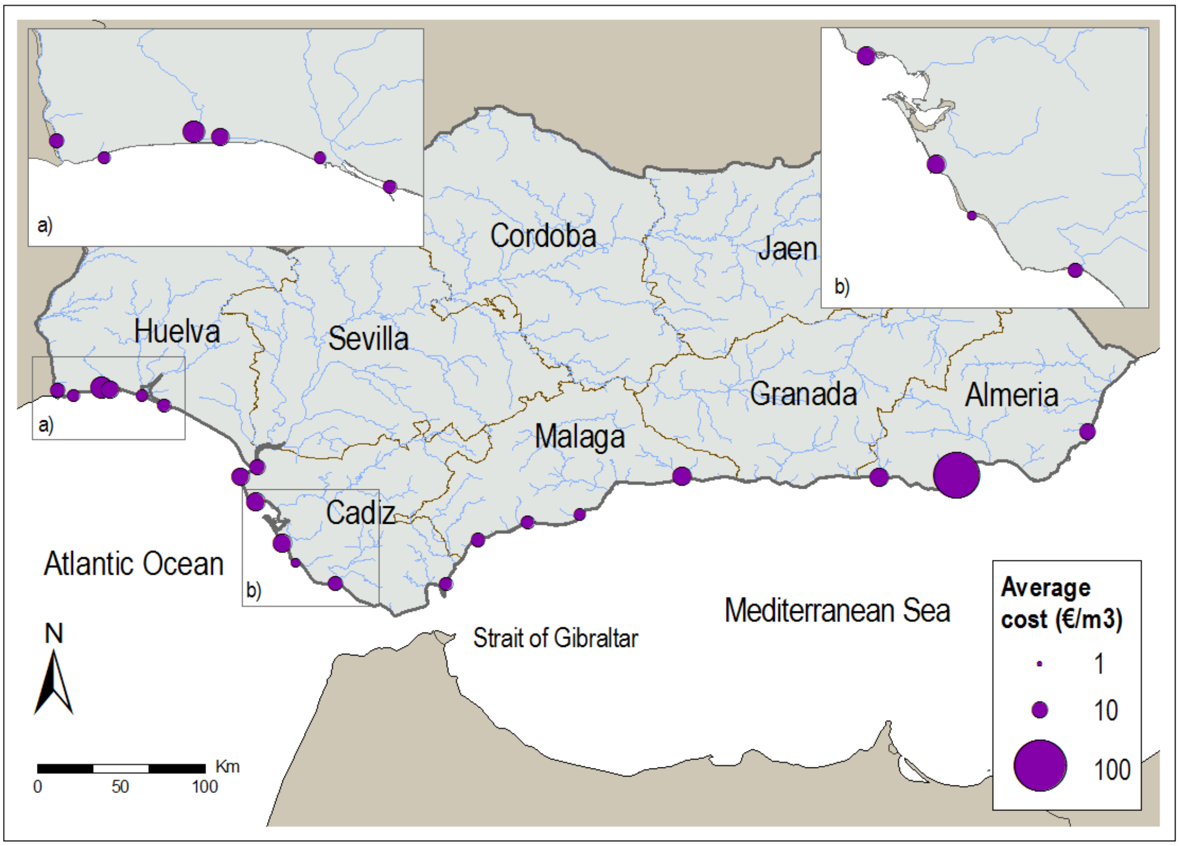

Figure 2.2.9: Average cost (€) of dredging $1m^3$

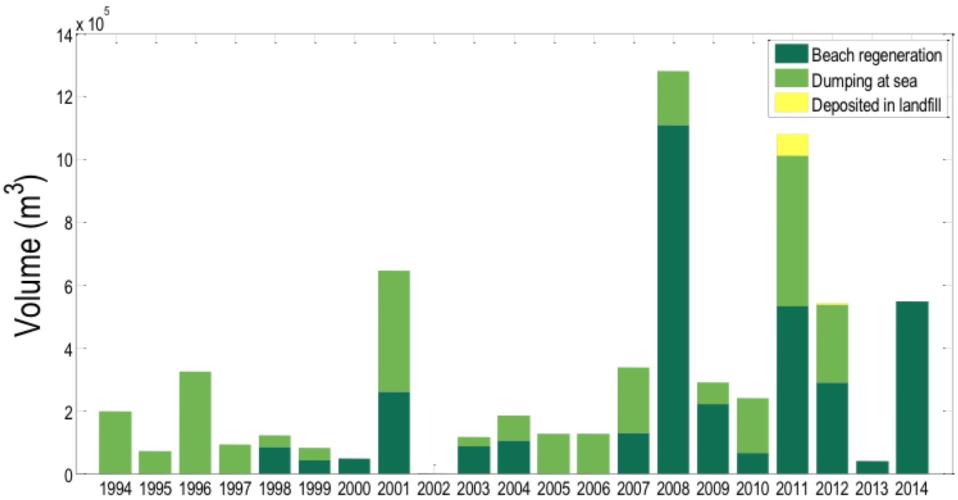

Figure 2.2.10: Evolution of the destination of dredging material over the years

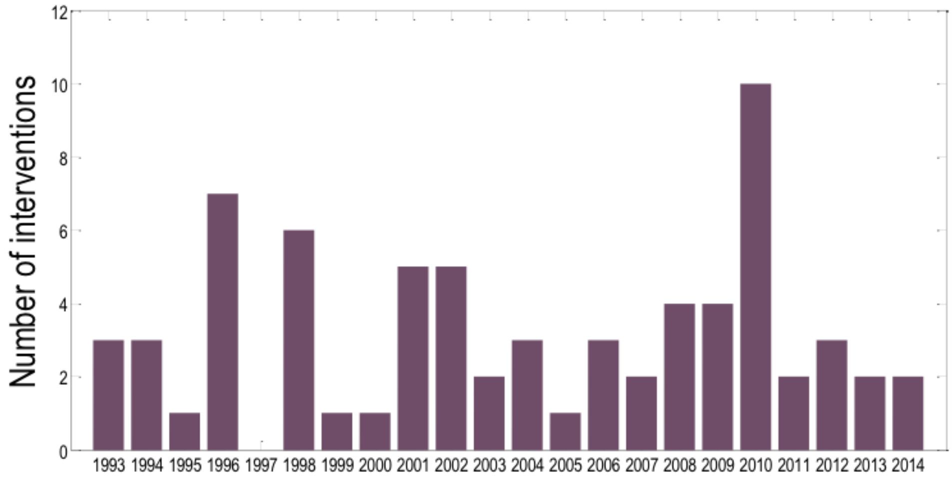

Figure 2.2.11: Number of dredging interventions per year

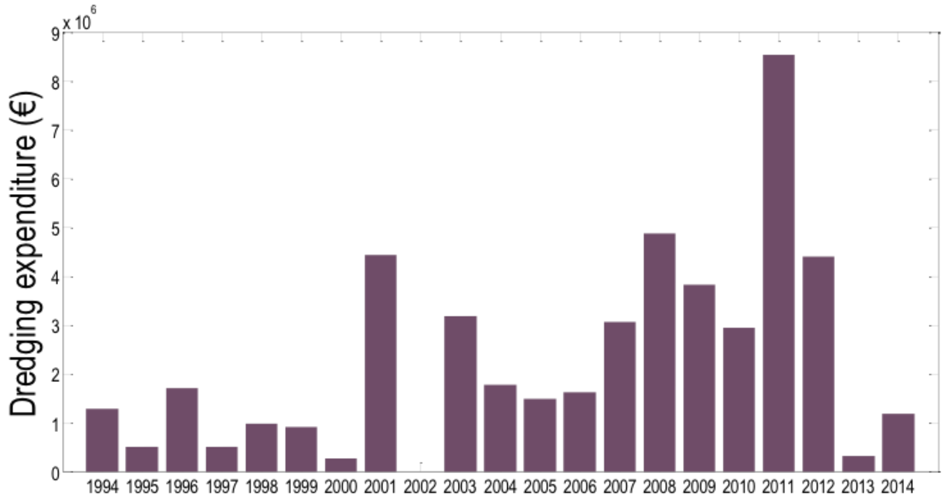

Figure 2.2.12: Dredging expenditure (real prices) (€) per year

increase in the amount expend in dredging interventions, with fluctuations associated both to the number of interventions and to global economic reasons. The sudden reduction in 2013 is due to the decrease in the volumes mobilized in the interventions (mainly due to economic reasons). Nevertheless, what can be extracted from the results considering how the environmental variables will evolve is that dredging problems will maintain (and even increase) in the upcoming years.

Punta Umbría (Huelva): an Example of the Database Potential

Punta Umbria is located inside the River Piedras estuary (Figure 2.2.1, port number 8), facing to the Gulf of Cádiz. Two ports managed by the APPA are located inside the estuary, at a distance of 2kms from the mouth of the estuary. Since the late 50's many new industrial and touristic activities developed in the surrounding area, with the consequent occupation of natural landscapes and the alteration of the natural behavior of the ecosystem. The result is a reduction in the magnitude of the natural agents with capacity to maintain the morphology at the entrance of the estuary. Hence, during last decades there has been a progressive increase in the magnitude and frequency of the sedimentations at the mouth of the estuary. This ongoing tendency of infilling estuaries is common in the Atlantic Southern Spanish coast.

To maintain the navigation channel and guarantee safety conditions periodic and urgent dredging's are necessary. According to the information of the database, 4 interventions have been done in the period 2004-2014 with a total cost of 4,558,214.14€. The volume of sediment mobilized at Punta Umbria is the second highest of Andalucía (874,587.92 m^3), representing 15% of the total volume dredged in Andalucía since 1993, only surpassed by Conil (26%).

Comparing with the other Andalusian ports with a significant number of interventions, the average price of dredging 1m^3 in Punta Umbria is 10.4€, one of the highest average prices (Figure 2.2.13-a) which provides an average magnitude per intervention of 1,140,000€ (Figure 2.2.8). Also, results indicate that the average price in Punta Umbria was similar at different interventions with values between 8.88-11.03 €/m^3. However, for other sites such as Conil deviations are significantly higher.

Figure 2.2.13-b shows liquidation budgets of the interventions at the main ports. Punta Umbria is one of the ports where more budgets were destined for dredging interventions and the costs of every intervention were higher than in others. In this case there are very noTable 2.2.differences between the budgets of the different interventions in Punta Umbria: the deviations are in average in the middle of the main ports, whereas significantly higher than for the results depict in Figure 2.2.13-a.

One of the main problems in Punta Umbria is the amount of sediment that the marine dynamics can mobilize due to both its availability and the severity of the wave climate. At least one intervention is required every 2 (maintenance) or 4 years (urgent) to maintain the navigability of the channel. Figure 2.2.13-c shows dredging volumes through the different interventions in the main ports. In Punta Umbria volumes are similar in each intervention with low deviations. Box-plot analyses show that Punta Umbria spends in dredging interventions above average in spite of dredging lower volumes. Nevertheless, there are others ports where the deviations are greater and where managers should focus first to improve efficiency and reduce costs.

Since most of the sediment is alongshore transported under storm conditions, we searched for correlations between the storms (frequency, number of storms and associated energy flux) and the dredging operations (dates, volumes and reasons). Figure 2.2.14 shows the relation between the number of storms, their energy and main angle and when the dredging operation was done. After a stormy winter such as those that occurred in 2002 and 2003, a subsequent dredging intervention was done (summer 2004). The winters that followed (2004, 2005 and 2006) were of low energy and no interventions were necessary. However, during the winter of 2007 the number

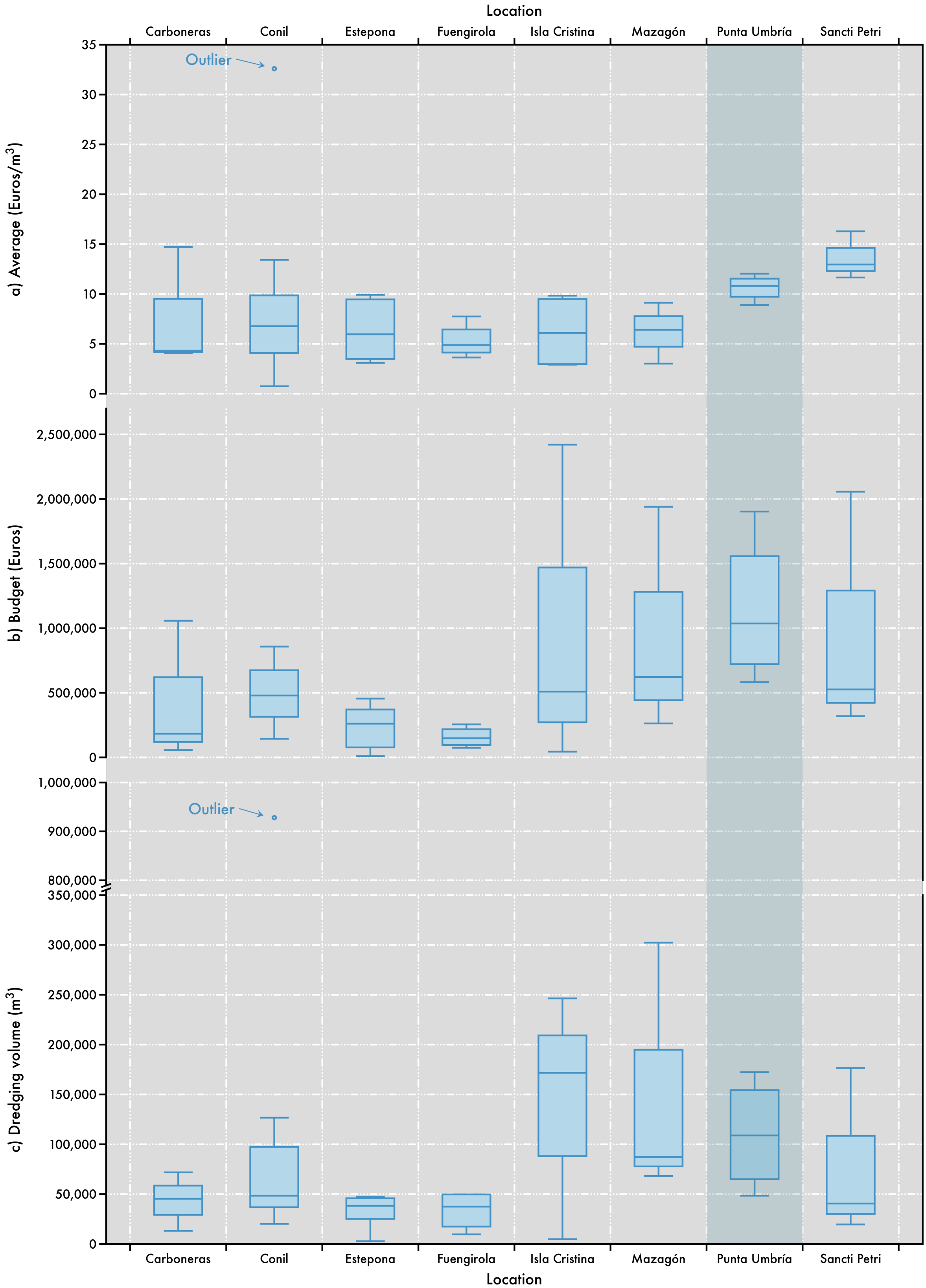

Figure 2.2.13: Box-plot showing the analysis of (a) average price of dredging 1m^3 of material, (b) liquidation budget of dredging interventions and (c) dredging volume in each intervention. The main ports were depicted

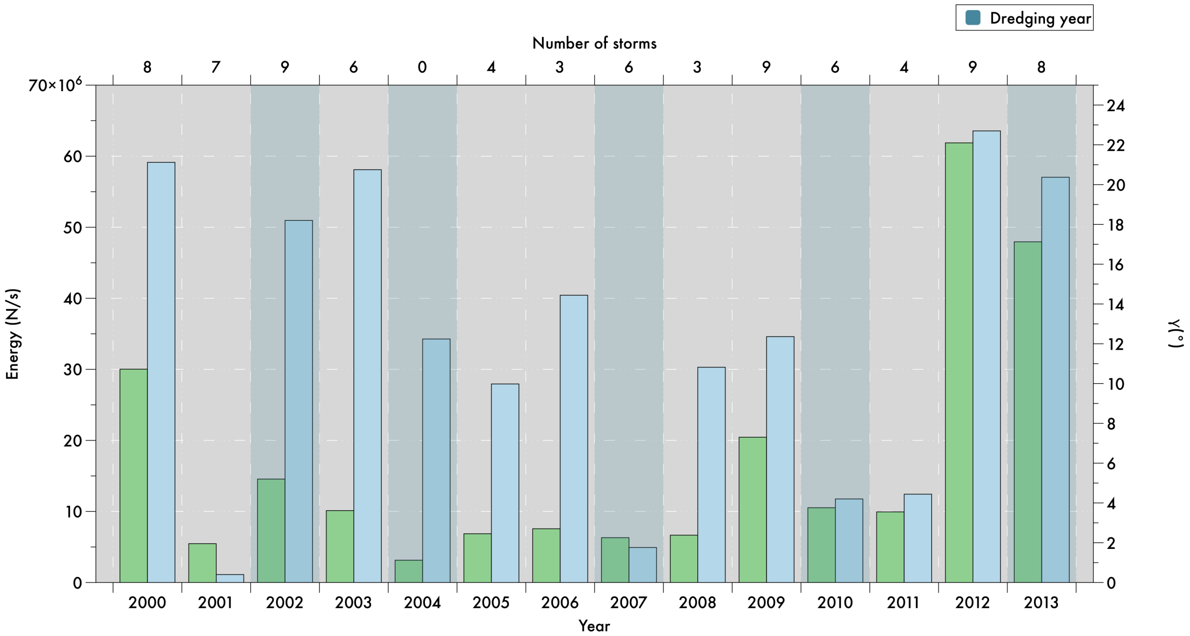

Figure 2.2.14: Relation between the number of storms, wave energy and wave angle and years of the dredging operations

of storms slightly increases and finally a dredging was done in summer 2008. This operation was followed by a very energetic winter in 2009 that results in a maintenance intervention in 2010. The winter of 2011 was mild, whereas in the winter of 2012-2013 a total of 9 storms arrived. Since summer of 2013 the estuary is having navigation problems and another operation is presently ongoing. For that specific case, an additional problem is the economic crisis and how the projects are being dealt; the delay between the design project (summer 2013) and the final construction project (December 2014) produced that more storms arrive and the sedimentation was higher than expected. Thus, the cost of the project increases and the administrative procedure is even more difficult.

According to Figure 2.2.14 and to the previous analysis, one operation is required every 4 years at Punta Umbria; this frequency reduces to 2 years if stormy winters arrive. For this site the main forcing factor is the littoral drift: alternative solutions should focus on reducing this rate.

Discussion

Dredging operations at ports are common engineering works to ensure both their security and functionality. Besides the dredging associated to new constructions or enlargement projects, they are one of the most common maintenance operations at ports with elevated costs, both economical and environmental ((Ramos, 2014)). Although the studied region is relatively small (87,268 km2), not highly populated (8,440,000 people) and without a strong industrial sector, it is characterized by a large coastline with a high number of ports. The development of a database to gain deeper insight into this problem was proven to be a valuable tool. The alteration of the natural behavior of many ecosystems, in combination with an incorrect management of the littoral zone and flood areas and a strong wave climate, results in a situation with many excessive silting's at ports.

Importance of databases for systematic civil engineering problems

Technical and scientific advances in many engineering disciplines are dependent on the availability of high quality data. The service level of a port is a crucial factor in building customer loyalty and managers demand actions that enhance the performance of given port (García-Morales et al., 2014). However, given the complexity of the interventions, port management generally focuses on optimizing the benefits derived from goods, whereas other management activities are marginalized.

Similar problems are found on other type of civil engineering activities. One of the most useful estimations obtained from the database is the relation between expected data (before works) and real data (after works) of a given magnitude. To analyze these deviations and anticipate how much the original expectations might change during works at each site, we have defined the percentage deviation (PD) of a given variable contained in the database. PD is calculated as:

$$PD = \frac{D_R - D_E}{D_e} * 100(\%)$$

where D_R and D_E corresponds to the real and expected data, respectively.

Figure 2.2.15 plots PD for the budget. Green (red) color indicates that the final cost was lower (higher) than projected, whereas the size of the symbol is proportional to this variation. The major differences are found at Huelva province: at all ports expected costs were underestimated, being the major differences variations in Ayamonte (53%) and El Rompido (31%). In contrast, Cadiz province in general and Conil port in particular indicate that the previsions were overestimated; final cost was around 50% of the initial prevision.

Similar analyses have been done for volume and duration (Figures 2.2.16 and 2.2.17). In this case, Huelva shows remarkable results; all their ports needed dredging volumes higher than expected. The main variation was found at El Rompido (around 76% higher than expected) followed by Mazagon with 44% of variation, whereas minimum variation are in Chipiona (6%) and Sancti Petri(8%), both located in Cadiz. Regarding the duration of the works, this analysis shows a global tendency: in general interventions are longer than expected. The highest variation is in El Rompido that shows a real duration 500% longer than expected duration, followed by far by Punta Umbria with 88%, Carboneras with 69% and Estepona 65%. La Atunara and Bonanza have the smallest differences, 2% and 14%.

The analyses showed above have been obtained working with the information contained in the database. Those results would help predicting the performance and behavior of future interventions based on real data and previous experiences in similar works. Hence, project managers could find those results useful at different stages of a project's development to make more accurate estimations.

6.2. Variables affecting dredging

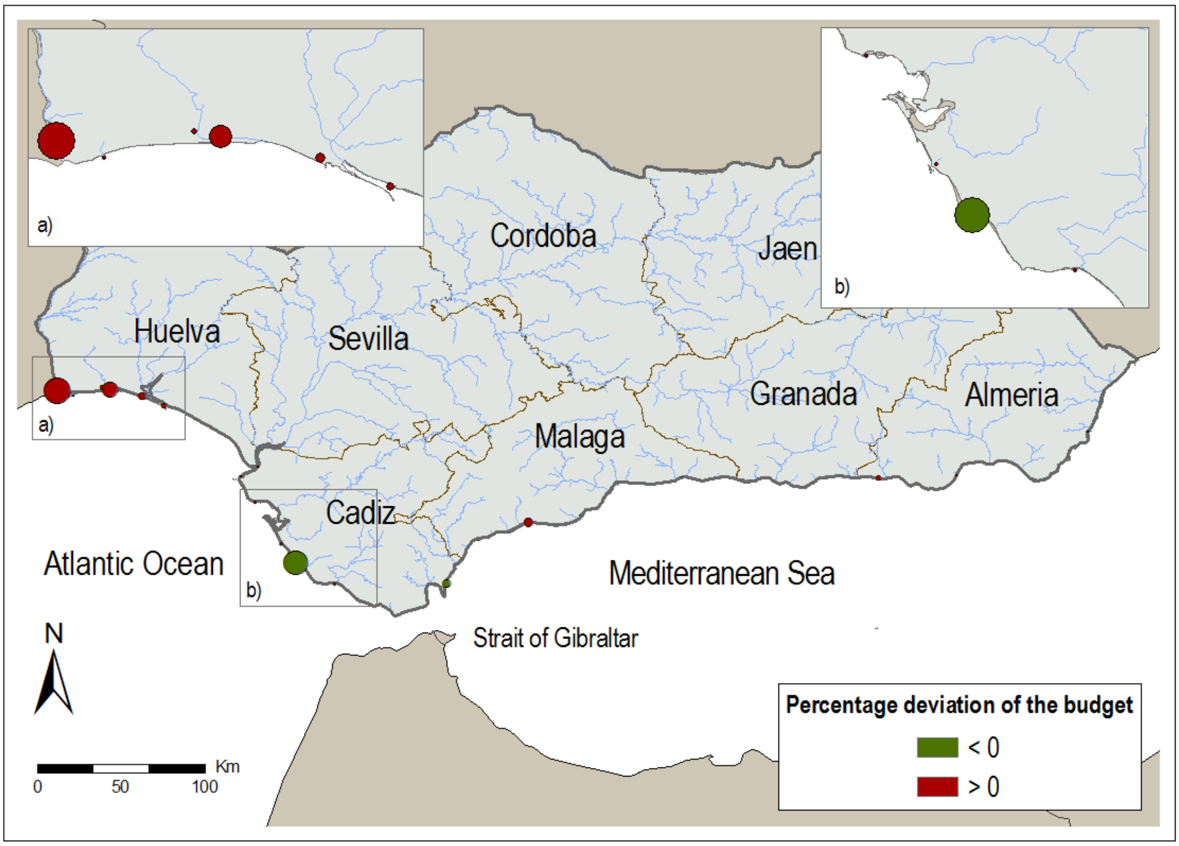

Figure 2.2.15: Map represents the different between liquidation budget and adjudication budget

As an engineering work, the dredging process can be characterized by a set of variables, being the most important the volume of material to be mobilized, the cost of the operations and the estimated time until the next intervention. The analysis of these variables is not trivial because those variables are not always directly correlated, particularly the volume and the cost. Analysis reveals that whereas the highest volume of dredging was at Cadiz, one of the highest average budget was inverted in Huelva interventions (Figure 2.2.8). The reasons are twofold: the maritime climate is much severer and the sediment is finer; both factors contribute to higher rates of sediment transport along the coast. Results indicate that the major problems are concentrated on the Atlantic side. The analysis of the storms along the coast from West to East (Figure 2.2.18) reveals that the number of storms along the Atlantic watershed generally is higher close to Conil, whereas decreases for the rest of the sites. According to this analysis, storms are more frequent in the central part just where more interventions have been done and the dredging volumes are higher.

Many environmental issues at ports are related with the maritime and atmospheric climate. Particularly, as we showed in the work, sedimentation at or near ports is strongly related with the sediment transport processes along the coast. The amount of sediment that is transported by the littoral drift is directly proportional to the energy of the incoming waves, resulting in the highest amount of sediments being advected under storms. The ability to predict when those storms will arrive, as well as the development of engineering tools for management that includes

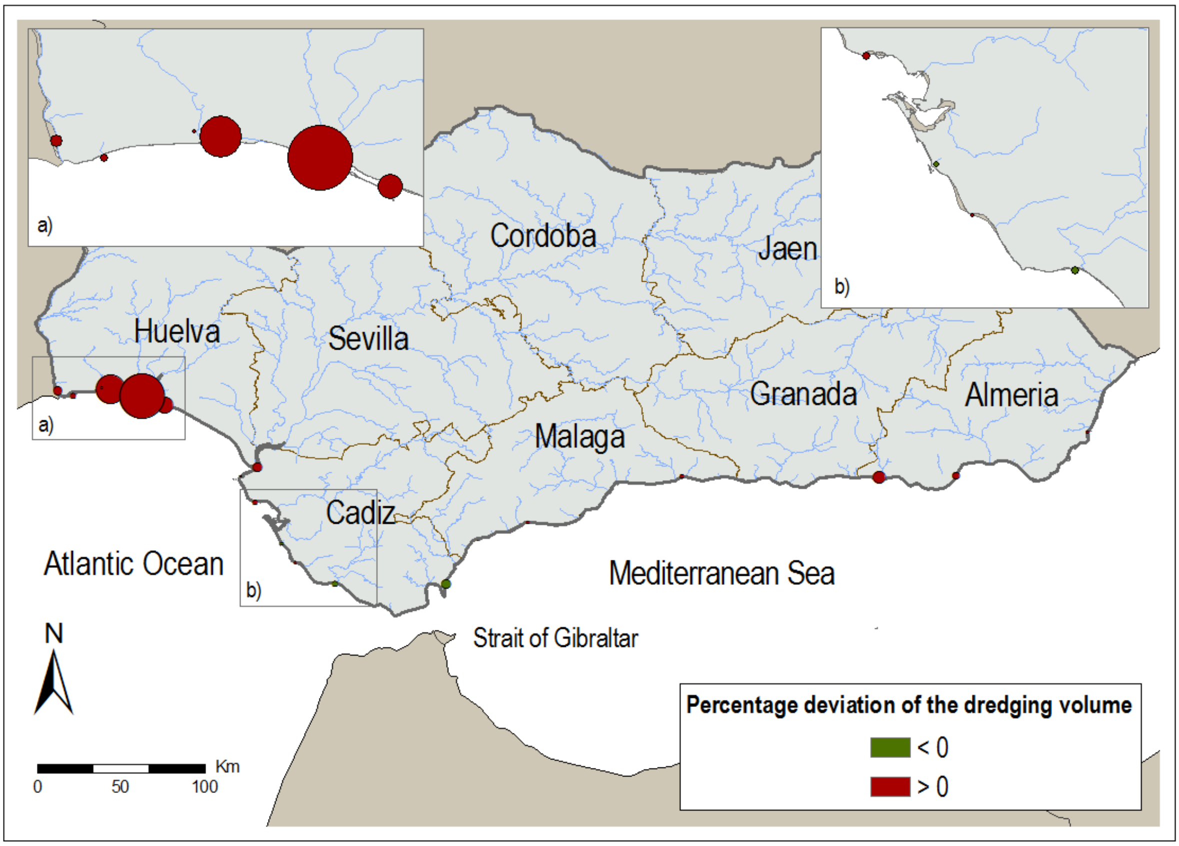

Figure 2.2.16: Percentage deviation of the dredging volume

simulations of the maritime climate (Solari et al., 2011) are key factors for an improvement in port management in general, and in sedimentation problems in particular.

The management of the dredging operations is mainly based on when to perform the operation, and what amount of sediment should be mobilized. According to our results, several factors need to be considered when planning dredging operations. First, the adequation between the machinery and the type of sediment to be mobilized is crucial. When the material is consolidated, the dredge performance diminishes and the cost of the operations increases. As an example of this, the last dredging performed in Punta Umbria has increased the price 20% because the dredging material was more consolidate than expected, in spite of dredging volume is less than initial data volume. Second, the reutilization of the material: only 28 of the dredging interventions analyzed in this work could have been reutilized. For the rest of the cases, the material was simply disposed at the sea or to the landfill, but the data show a clear trend in favor of reutilization in beach regeneration. Third, when the previous factor applies, for some cases there are no allowed disposal sites close to the port due to environmental restrictions. That is the case of Roquetas de Mar port, which has the highest individual price for every cubic meter (Figure 2.2.9). Fourth, administrative procedure influence work schedule, urgent procedure has the advantages in terms of faster procedures in comparison with normal procedure; in fact, the slowness of the bureaucracy delays especially beginning works. The previous factors can in general be planned in advance, whereas others such as the wave climate present a high level of uncertainty.

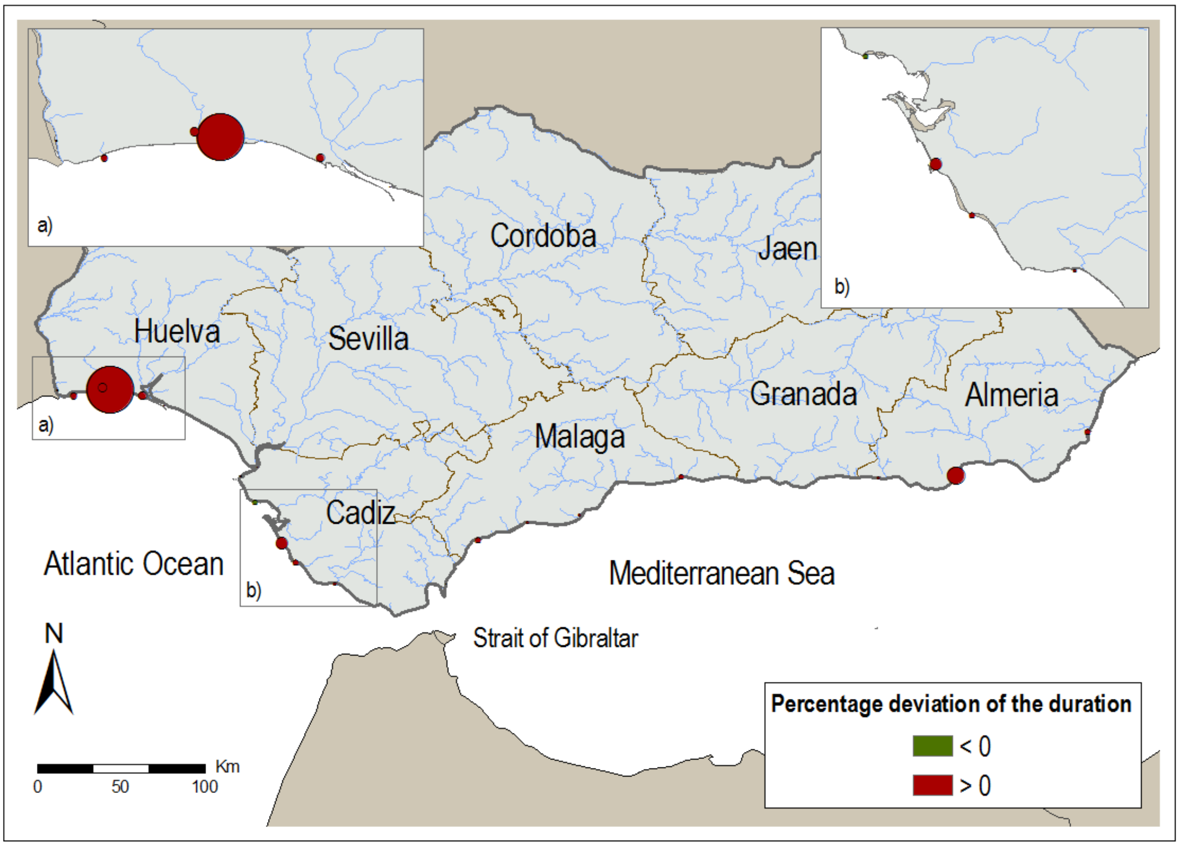

Figure 2.2.17: Percentage deviation of the duration of works

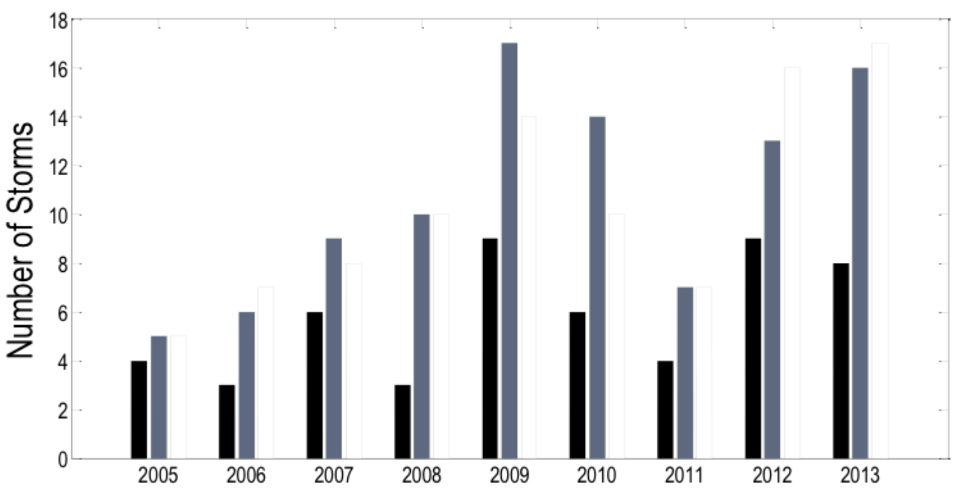

Figure 2.2.18: Graph shows the number of storms per year

Conclusions

Civil engineering infrastructures required systematic maintenance activities with an important impact on the economy. Despite their importance, few studies have focused on analyzing this type of interventions. Within this work we have developed a general methodology to create databases of civil engineering systematic activities. It consists on the achievement of six sequential phases to gather and organize the information into a database. The developed database can be implemented in PostgreSQL, using PostGIS extension for spatial data, and therefore GIS integration, which is a valuable tool for the civil engineering discipline.

The methodology was applied to the case of the dredging interventions in the 20 ports that are managed by the Andalusian Regional Government (Spain). The database contains 87 fields of information collected after the analysis of the 70 interventions performed since 1994. This database is free, public and available for the scientific community. Some analyses of the database information were done to highlight the potential of the database for this type of studies. The following conclusions can be extracted:

- Because information is well-structured and organized in the database, it is possible to query about data using a query language, such as SQL. It allows obtaining a sub group of the information that fits certain criteria. Indeed, data can be gathered and expressed in a summary form for analysis purposes. Thus, the processed information can be used to build trends in data or discover patterns within data.
- The dredging interventions were highly correlated with the severity of the wave climate. The majority of the interventions, dredging volumes and costs were at the Atlantic part of the Andalusian coast, where both the wave energy flux and the wave height but also the frequency and intensity of the storms were higher. On the contrary, minor interventions took place on many ports of the Mediterranean, mainly at locations were the wave climate was significantly milder. Considering both the volumes and costs of the interventions, improving strategies should be focused on the Atlantic side.
- In terms of efficiency, Atlantic ports exhibit higher values even though the wave climate limits the works during many time of the year. This is due to the non-cohesive and lower sediment size (sand), and also because the mobilized volume in each intervention is higher at this part of the coast.
- With the developed database analysis of specific sites are easy to implement. Results for the case of Punta Umbría indicate deviations from the rest of the main ports; deviations of the average costs, budget and dredging volumes are easily estimated and compared with other cases. Also, site-specific analysis regarding the correlation between the number of storms and the interventions provide results that can be used to improve the manager's decisions.
- Many civil engineering works and interventions are characterized by strong deviations between the initial and final projects. Our results show that for some sites deviations in terms of budget and volumes exceed 50% between projection and execution, whereas for some others differences are lower than 5%. This database was proven to be a nice tool to identify and analyze sites and interventions that exhibit the major deviations.

References

Argent, R.M., Perraud, J.M., Rahman, J.M., Grayson, R.B., Podger, G.M., 2009. A new approach to water quality modelling and environmental decision support systems. Environmental Modelling & Software 24, 809–818. doi:10.1016/j.envsoft.2008.12.010

Bedri, Z., Corkery, A., O'Sullivan, J.J., Alvarez, M.X., Erichsen, A.C., Deering, L.A., Demeter, K., O'Hare, G.M.P., Meijer, W.G., Masterson, B., 2014. An integrated catchment-coastal modelling system for real-time water quality forecasts. Environmental Modelling & Software 61, 458–476. doi:10.1016/j.envsoft.2014.02.006

Casado-Martínez, M.C., Forja, J.M., DelValls, T.A., 2009. A multivariate assessment of sediment contamination in dredged materials from Spanish ports. Journal of Hazardous Materials 163, 1353–1359. doi:10.1016/j.jhazmat.2008.07.106

Chang, Y.-T., 2013. Environmental efficiency of ports: a Data Envelopment Analysis approach. Maritime Policy & Management 40, 467–478. doi:10.1080/03088839.2013.797119

Darbra, R.M., Pittam, N., Royston, K.A., Darbra, J.P., Journee, H., 2009. Survey on environmental monitoring requirements of European ports. Journal of Environmental Management 90, 1396–1403. doi:10.1016/j.jenvman.2008.08.010

Darbra, R.M., Ronza, A., Casal, J., Stojanovic, T.A., Wooldridge, C., 2004. The Self Diagnosis Method. Marine Pollution Bulletin 48, 420–428. doi:10.1016/j.marpolbul.2003.10.023

Darbra, R.M., Ronza, A., Stojanovic, T.A., Wooldridge, C., Casal, J., 2005. A procedure for identifying significant environmental aspects in sea ports. Marine Pollution Bulletin 50, 866–874. doi:10.1016/j.marpolbul.2005.04.037

Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., Montanarella, L., 2012. Satellite remote sensing for soil mapping in Africa: An overview. Progress in Physical Geography 36, 514–538. doi:10.1177/0309133312446981

Donnelly, A., Jones, M., O'Mahony, T., Byrne, G., 2007. Selecting environmental indicator for use in strategic environmental assessment. Environmental Impact Assessment Review 27, 161–175. doi:10.1016/j.eiar.2006.10.006

Fettweis, M., Francken, F., Van den Eynde, D., Verwaest, T., Janssens, J., Van Lancker, V., 2010. Storm influence on SPM concentrations in a coastal turbidity maximum area with high anthropogenic impact (southern North Sea). Continental Shelf Research 30, 1417–1427. doi:10.1016/j.csr.2010.05.001

Goldsworthy, L., Goldsworthy, B., 2015. Modelling of ship engine exhaust emissions in ports and extensive coastal waters based on terrestrial AIS data – An Australian case study. Environmental Modelling & Software 63, 45–60. doi:10.1016/j.envsoft.2014.09.009

Guerra, R., Pasteris, A., Ponti, M., 2009. Impacts of maintenance channel dredging in a northern Adriatic coastal lagoon. I: Effects on sediment properties, contamination and toxicity. Estuarine, Coastal and Shelf Science 85, 134–142. doi:10.1016/j.ecss.2009.05.021

Hedge, L.H., Knott, N.A., Johnston, E.L., 2009. Dredging related metal bioaccumulation in oysters. Marine Pollution Bulletin 58, 832–840. doi:10.1016/j.marpolbul.2009.01.020

Lee, T., Yeo, G.-T., Thai, V.V., 2014. Environmental efficiency analysis of port cities: Slacks-based measure data envelopment analysis approach. Transport Policy 33, 82–88. doi:10.1016/j.tranpol.2014.02.009

Lin, B., Lin, C.-Y., 2006. Compliance with international emission regulations: Reducing the air pollution from merchant vessels. Marine Policy 30, 220–225. doi:10.1016/j.marpol.2005.01.005

Magaña, P., López-Ruiz, A., Lira, A., Ortega-Sánchez, M., Losada, M.A., 2014. A public, open Western Europe database of shoreline undulations based on imagery. Applied Geography 55, 278–291. doi:10.1016/j.apgeog.2014.09.018

Mamindy-Pajany, Y., Libralato, G., Roméo, M., Hurel, C., Losso, C., Ghirardini, A.V., Marmier, N., 2010. Ecotoxicological evaluation of Mediterranean dredged sediment ports based on elutriates with oyster embryotoxicity tests after composting process. Water Research 44, 1986–1994. doi:10.1016/j.watres.2009.11.056

Marriner, N., Morhange, C., 2007. Geoscience of ancient Mediterranean harbours. Earth-Science Reviews 80, 137–194. doi:10.1016/j.earscirev.2006.10.003

Mestres, M., Sierra, J.P., Mösso, C., Sánchez-Arcilla, A., 2010. Modelling the sensitivity to various factors of shipborne pollutant discharges. Environmental Modelling & Software 25, 333–343. doi:10.1016/j.envsoft.2009.08.006

Peris-Mora, E., Orejas, J.M.D., Subirats, A., Ibáñez, S., Alvarez, P., 2005. Development of a system of indicators for sustainable port management. Marine Pollution Bulletin 50, 1649–1660. doi:10.1016/j.marpolbul.2005.06.048

Puig, M., Wooldridge, C., Darbra, R.M., 2014. Identification and selection of Environmental Performance Indicators for sustainable port development. Marine Pollution Bulletin 81, 124–130. doi:10.1016/j.marpolbul.2014.02.006

Ramos, S.J., 2014. Planning for competitive port expansion on the U.S. Eastern Seaboard: the case of the Savannah Harbor Expansion Project. Journal of Transport Geography 36, 32–41. doi:10.1016/j.jtrangeo.2014.02.007

Sheehan, C., Harrington, J., 2012. Management of dredge material in the Republic of Ireland – A review. Waste Management 32, 1031–1044. doi:10.1016/j.wasman.2011.11.014

Sheehan, C., Harrington, J., Murphy, J.D., 2010. An environmental and economic assessment of topsoil production from dredge material. Resources, Conservation and Recycling 55, 209–220. doi:10.1016/j.resconrec.2010.09.011

Solari, S., Losada, M.A., 2011. Non-stationary wave height climate modeling and simulation. J. Geophys. Res. 116. doi:10.1029/2011JC007101

Song, C., 2013. Optical remote sensing of forest leaf area index and biomass. Progress in Physical Geography 37, 98–113. doi:10.1177/0309133312471367

Wasserman, J.C., Barros, S.R., Lima, G.B.A., 2013. Planning dredging services in contaminated sediments for balanced environmental and investment costs. Journal of Environmental Management 121, 48–56. doi:10.1016/j.jenvman.2013.02.024

Watson, R., Revenga, C., Kura, Y., 2006. Fishing gear associated with global marine catches. Fisheries Research 79, 97–102. doi:10.1016/j.fishres.2006.01.010

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Babendreier, J., Galvin, M., 2014. Design of a component-based integrated environmental modeling framework. Environmental Modelling & Software 55, 1–24. doi:10.1016/j.envsoft.2014.01.016

Yozzo, D.J., Wilber, P., Will, R.J., 2004. Beneficial use of dredged material for habitat creation, enhancement, and restoration in New York–New Jersey Harbor. Journal of Environmental Management 73, 39–52. doi:10.1016/j.jenvman.2004.05.008

Zentar, R., Abriak, N.E., Dubois, V., Miraoui, M., 2009. Beneficial use of dredged sediments in public works. Environmental Technology 30, 841–847. doi:10.1080/09593330902990139

2.3 Ontology-based ecological system

Introduction and rationale

European Union has developed a set of environmental directives focused on nature conservancy (Evans, 2012). Their main aims are: 1) to halt the biodiversity loss according to the Convention on Biological Diversity (CBD, 2005), 2) to promote the implementation of policies for achieving sustainable development in a context of global change.

The Birds (79/409/EEC; 2009/147/EU) as well as the Habitats Directives (92/43/EEC) seek a favourable conservation status for all listed habitats and species all throughout the European territory (Louette et al., 2011). For these objectives, it is mandatory to implement methods to assess the conservation status of habitats and species. This is a challenging task that requires taking into consideration the concept of monitoring (Lindenmayer and Likens, 2010; Pereira and Cooper, 2006). According to Lindenmayer and Likens, the protocols used to satisfy legislation requirements must be focused on identifying trends in structural and functional features of habitats. These authors assert that “mandated monitoring” (required by legislation) can help in assessing the changes in the conservation status of habitats (Lindenmayer and Likens, 2010: 1325)

Satellites gather huge amounts of information that could be useful to monitor and to assess the conservation status of habitats (Vanden Borre et al., 2011). Such information would be adequate to assess both structural (distribution) and functional changes (productivity, phenology, etc.) in the Natura 2000 habitats. For example, a wide set of products derived from MODIS (Moderate Resolution Imaging Spectroradiometer) sensor are useful for monitoring ecosystem function at a landscape scale (250–1000 m resolution) (Hall et al., 2002; Huete et al., 2002; Justice et al., 2002). Other satellites such as Quickbird or IKONOS provide information at a finely detailed spatial resolution (0.5–4 m resolution), which is useful to monitor habitat distribution and structure (Förster et al., 2008; Hyde et al., 2006; Wang et al., 2004). The most important advantage of satellite Earth observation in relation to habitat monitoring could be its capacity to allow comparisons among different locations (Vanden Borre et al., 2011). The temporal homogeneity (the same information is gathered with a predefined periodicity) is also a key feature to implement monitoring protocols using (satellite) Earth observation. However, the information collected from satellites cannot be processed and interpreted straightforwardly by most scientists and decision makers (Kalluri et al., 2003). Both the overwhelming amount of data to process/analyse as well as the inherent complexity of the variables measured makes it difficult to create an operational system for assessing habitat functioning (Xue et al., 2011).

Ontologies are knowledge-representation techniques defined as a specification of a conceptualization (Gruber, 1993) within a domain of interest (habitat functioning in our case). A conceptualization is “an abstract, simplified view of the world that we wish to represent for some purpose” (Gruber, 1993:199). A computer can “understand” an ontology, because ontologies are structured according to concepts and relationships on which a computer can “reason”, as opposed to unstructured files like documents (Antoniou and van Harmelen, 2004). The use of ontologies can foster comprehensive data discovery and integration (Gruber, 1993; Jones et al., 2006), adding semantic meaning to data. Thus, these techniques can promote the use of remote sensing by environmental managers and ecologists (Silva et al., 2005).

While ontologies help to represent the domain, knowledge bases are used to store facts and complex information defined according to ontologies. Consequently, an inference engine, a software tool that applied logical rules to the knowledge base, can reason about those facts, deduce implicit facts, or resolve semantic queries (Hayes-Roth et al., 1983). Although ontologies are commonly used in different disciplines (Bard and Rhee, 2004; Renear and Palmer, 2009), they are not common in Ecology (Madin et al., 2007, 2008; Williams et al., 2006), or Earth observation (Arvor et al., 2013; Fallahi et al., 2008; Hashimoto et al., 2011; Larin Fonseca and Garea Llano, 2011; Oliva-Santos et al., 2014; Wiegand and García, 2007).

In this work, we describe the design and implementation of an ontological system (called *Savia,* http://obsnev.es/ontologia/index) that combines the advantages of (satellite) Earth observation with the knowledge-representation capabilities of ontologies to create a tool that displays indicators and trends regarding habitat functioning. This work had two objectives: a) to assess the functioning of a Natura 2000 habitat and its relationships with abiotic factors (thematic objective), and b) to use ontologies to create a operational system that satisfies the first objective (methodological objective). Our work provides a novel case study to the body of knowledge regarding the use of ontologies in Earth observation. It is also of value because we compute temporal indicators and trends to assess the conservation status of habitats. Finally, we show how ontologies can help to bridge the gap between ecologists and remote-sensing experts.

Study area and data

Study area

Sierra Nevada (SE Spain) is a mountainous area (ranging from 860 m to 3482 m a.s.l.) covering more than 2000 km2 (Figure 2.3.1a). The climate is Mediterranean, characterized by cold winters and hot summers, with a pronounced summer drought.

Sierra Nevada is considered one of the most important biodiversity hotspots in the Mediterranean region (Blanca et al., 1998) and has several types of legal protection: Biosphere Reserve, National and Natural Park, and Nature 2000 site. Sierra Nevada is also a LTER (Long-Term Ecological Research) site.

We have focused this work on one habitat of Sierra Nevada: forests dominated by *Quercus pyrenaica* Willd. This habitat (EU habitat code 9230) is included in the Annex I of the Habitats Directive and its conservation status is not well known (EIONET, 2013), partly due to lack of detailed ecological studies (García and Jiménez, 2009). The Pyrenean oak forests extend from southwestern France to the Iberian Peninsula (Franco, 1990) (Figure 2.3.1a), reaching their southernmost European limit in Sierra Nevada, where nine oak patches (2400 ha) have been identified (Figure 2.3.1b), ranging between 1100–2000 m a.s.l.

Q. pyrenaica is considered as vulnerable in southern Spain (Blanca and Mendoza, 2000) and the populations inhabiting Sierra Nevada are considered relict forests (Melendo and Valle, 1996). They have undergone intensive anthropic use in recent decades (Camacho-Olmedo et al., 2002). They are also expected to suffer the impact of climate change, due to their climate requirements (wet summers): *Q. pyrenaica* requires between 650 and 1200 mm of annual precipitation and minimal summer precipitation between 100 and 200 mm. Thus, simulations of the climate-change effects on this habitat point to a reduction in suitable habitat for Sierra Nevada (Benito, 2009;

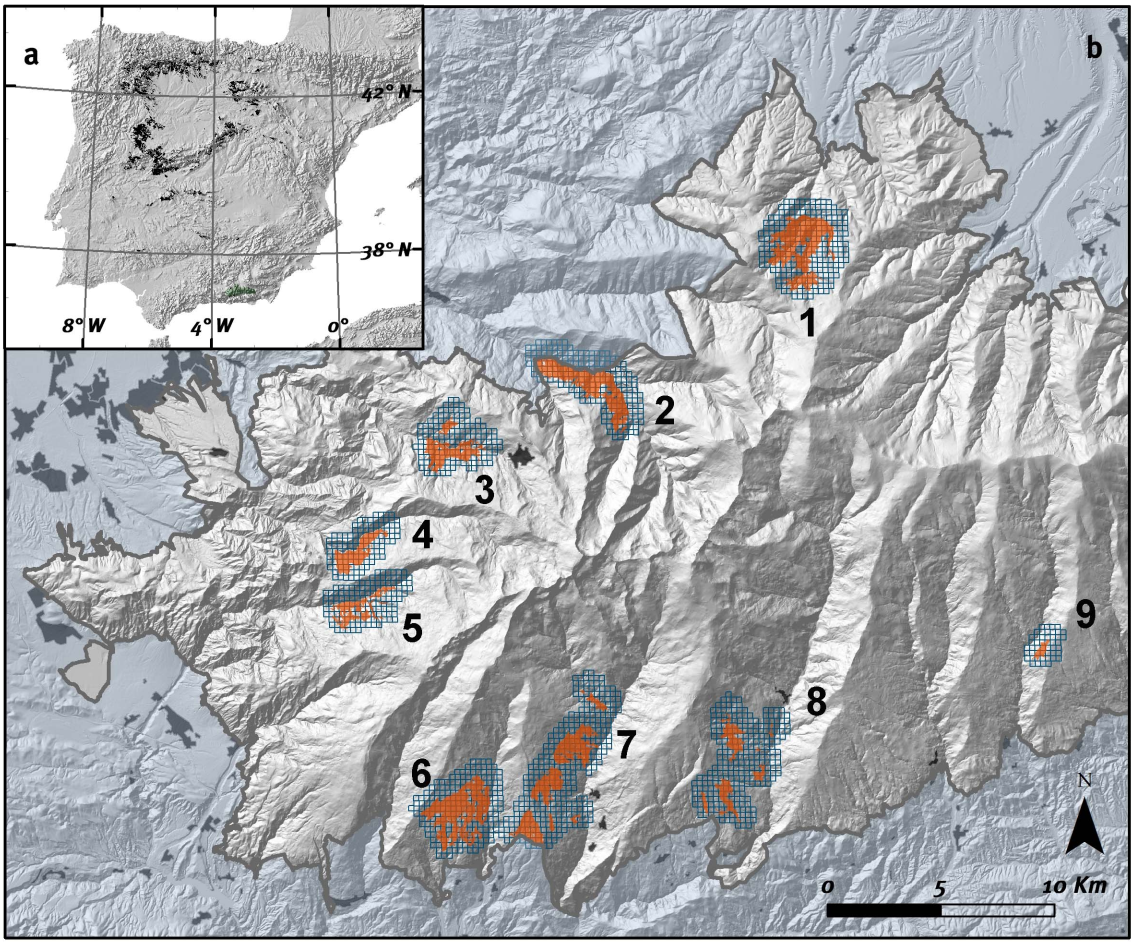

Figure 2.3.1: Location of Sierra Nevada mountains. The distribution of *Q. pyrenaica* in the Iberian Peninsula is shown in black (a). The nine patches of *Q. pyrenaica* in Sierra Nevada are shown in orange (b). The grey line shows the boundary of the natural protected area of Sierra Nevada. The pixels used to compute the vegetation and snow indicators are included (blue grid)

Benito et al., 2011).

Data sets and derived information

We have selected two MODIS products: MOD13Q1 to assess the habitat functioning and MOD10A2 to study the behaviour of an abiotic factor (snow cover). MOD13Q1 provides information on vegetation indices NDVI (Normalized Difference Vegetation Index). The spatial resolution of this product is 250 m and the temporal resolution is 16 days. MOD10A2 provides information about snow cover extent (Hall et al., 2002). It has a periodicity of 8 days and a spatial resolution of 500 m. Each MOD10A2 pixel is labelled as snow if it has had snow on one of the previous 8 days. We selected MODIS products because both their spatial resolution and temporal resolutions are appropriate for the scope of this study.

We homogenized the different spatial and temporal resolutions in these two products to produce the final data at 500 m of spatial resolution and 16 days of temporal resolution. For the spatial resolution, we intersected the two grids to assign the identifier of any MOD10A2 pixel to its overlapping one in MOD13Q1. For temporal homogenization, we aggregated the data from MOD10A2 (8 days) to gain information regarding at least MOD13Q1 scale (i.e. more than 16 days). We used the MODIS time series from 2000 to 2012.

NDVI seasonal measurements (aggregation of NDVI values by season) are suitable tools to quantify productivity and biomass (Running et al., 2004; Turner et al., 2006), seasonality (Piñeiro et al., 2006; Potter and Brooks, 1998) and other phenological measurements (Cleland et al., 2007). These measurements have been used to characterize ecosystem functioning (Cabello et al., 2012). We have calculated indicators regarding these ecological functions using the mean NDVI profiles provided by MODIS (Figure 2.3.2) *sensu*Alcaraz-Segura et al. (2009):

- **annual and seasonal mean (NDVI-I)** which can be used to estimate fAPAR (Fraction of Absorbed Photosynthetically Active Radiation) (Sellers et al., 1996) and thus net primary production (Paruelo et al., 1997; Sellers et al., 1992; Tucker et al., 1985).
- **annual relative range (RREL)**; difference between maximum and minimum NDVI divided by annual mean. This variable provides an indicator of the seasonality of the photosynthetic activity (Paruelo and Lauenroth, 1995).
- **maximum and minimum NDVI values (MAX and MIN)**and**months (MMAX and MMIN)**in which they occur. They provide an additional description of phenology, indicating the intra-annual distribution of the periods with maximum and minimum photosynthetic activity (Hoare and Frost, 2004; Lloyd, 1990).

NDSI (Normalized Difference Snow Index) is a “spectral band ratio that takes advantage of the fact that snow reflectance is high in the visible wavelengths and low in the shortwave infrared region” (Salomonson and Appel, 2006: 351). This index has proven to be a robust indicator of snow cover using MODIS images (Rittger et al., 2013). We have calculated several indicators from MOD10A2 images (Wang and Xie, 2009) (Figure 2.3.2):

- **snow-cover duration (SCD)**: is defined as the number of days covered by snow per hydrological year (describe a time period of 12 months for which precipitation totals are measured).
- **snow-cover onset dates (SCOD)**: is defined as the first date in the hydrological year that the pixel has snow. This indicator is useful to identify shifts in the starting of snow season.
- **snow-cover melting dates (SCMD)**: is the last date in the hydrological year that the pixel has snow. This indicator provides useful information about the melting process.
- **snow-cover melting cycles (SCMC)**: number of melting cycles in each pixel per hydrological year.

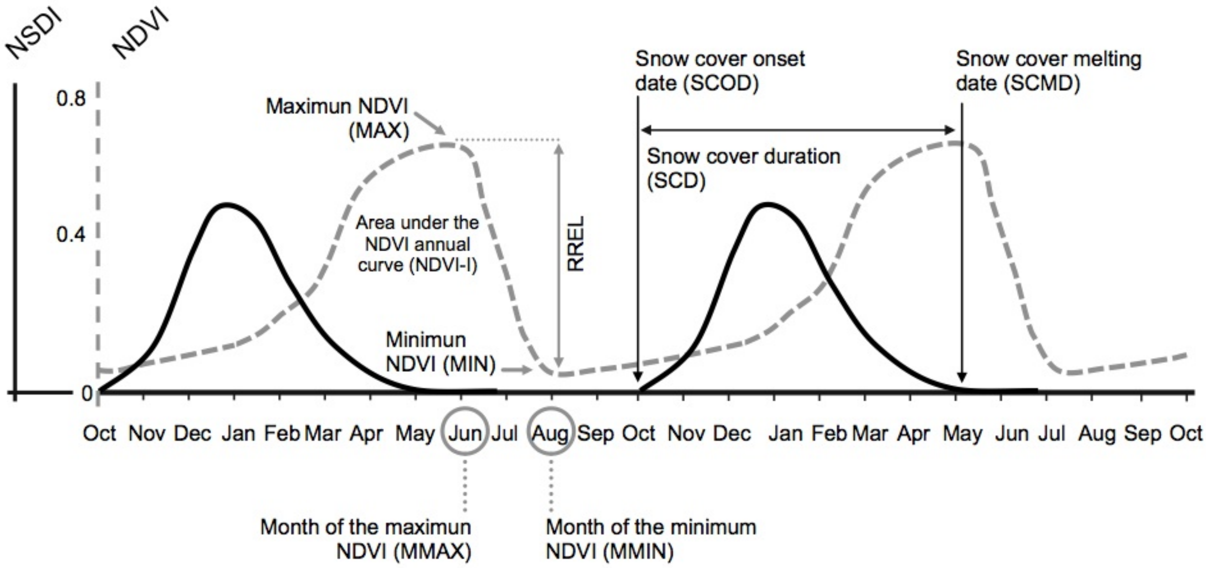

Figure 2.3.2: Attributes derived of Normalized Difference Vegetation Index (NDVI) and snow-cover profiles. Modified from Alcaraz-Segura et al. (2009) and Wang and Xie (2009)

Knowledge retrieval: ontologies and semantic processing

Savia was designed taking into account a client-server architecture (Figure 2.3.3). The system contains different modules that extract relevant knowledge from the raw data. These modules act in a user-transparent way and are detailed in the following subsections, highlighting image processing, the development of the ontology, how instances are generated, and the final query system.

Embedding MODIS images in a database and calculating thematic indicators

HDF (Hierarchical Data Format) files are downloaded from NASA servers and processed using a workflow that makes the process automatic and reproducible. This workflow is stored and documented in a model repository called ModeleR (Bonet et al., 2014; Pérez-Pérez et al., 2012). The workflow extracts information contained in any HDF files and stored it in a relational database (see structure in Figure 2.3.4). NDVI and NDSI values are stored in a table that is linked to a vector layer containing the centroids of MODIS pixels. These raw data are used to aggregate and calculate the different indicators in *Savia*. The results are integrated again into the relational database, that is part of the Sierra Nevada LTER site information system (Bonet et al. 2011).

The indicators described in Section 2.3.2.2 were calculated for each pixel and temporal stage (by hydrological year, i.e. the period between October 1st of one year and September 30th of the next; and by season) using SQL queries. The temporal trend for each pixel was calculated using the nonparametric Mann-Kendall trend test (Kendall, 1975; Mann, 1945). The analyses were computed in R (R Core Team, 2013) with Kendall package (McLeod, 2011). We set 0.05 the alpha level for the test, and slopes with p-values > 0.05 were considered significant.

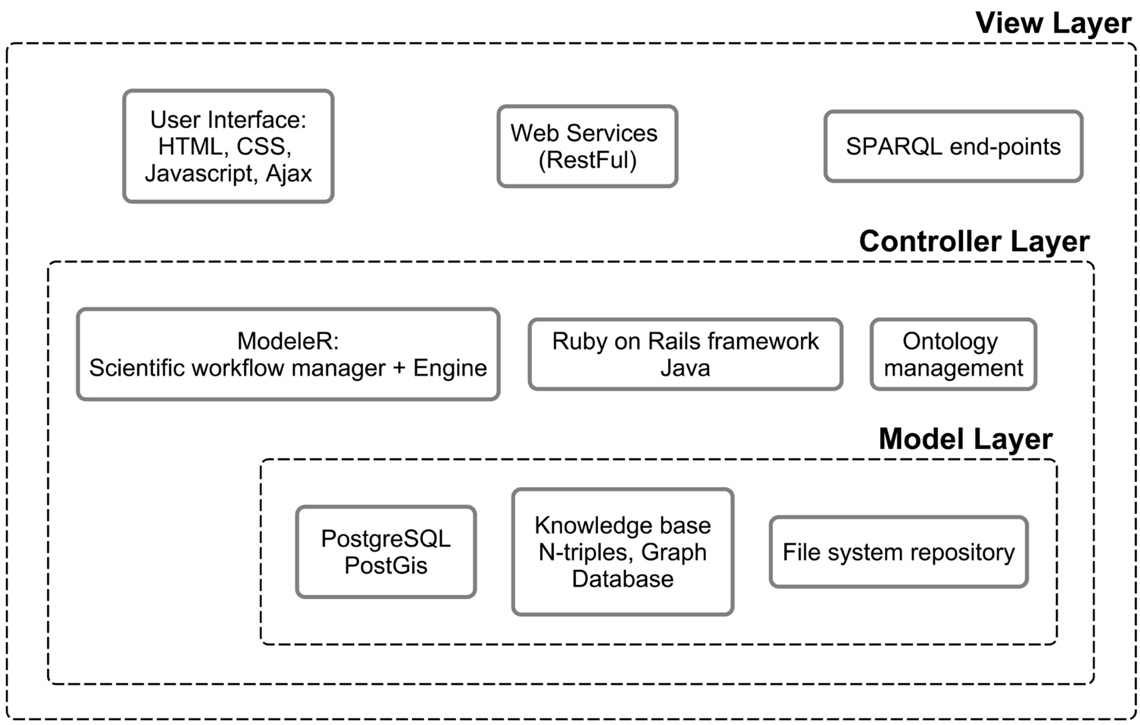

Figure 2.3.3: System architecture

Creating the ontology

The ontology must represent both the information (MODIS products, indicators, and temporal trends) and the concepts used to add ecological meaning to the data (Figure 2.3.5). To build the ontology, we used Time Ontology in OWL (Web Ontology Language) (Hobbs and Pan, 2004) and Basic Geo (WGS84 lat/long) Vocabulary (Brickley, 2003) external ontologies. The OWL-Time ontology promoted by W3C (World Wide Web Consortium) (W3C, 2013), provides a vocabulary for expressing instants and intervals, together with information concerning durations and date/time information (Hobbs and Pan, 2006). The Basic Geo is an RDF (Resource Description Framework) vocabulary for representing latitude, longitude, altitude information as well as other information related to spatial-located items.

Thus, the ontology takes into account three different parts (Figure 2.3.5):

Representing spatial information. The main concept is the *Pixel*, which represents a pixel from a MODIS image. Some pixels that share similar functions (i.e. be covered by the same habitat) may belong to a *Patch*. Finally, some patches sharing the same dynamics may belong to a *Group*. The properties called *PixelBelongsToPatch* and *PatchBelongsToGroup* help to define the relationships between the previously defined concepts. *PixelIsNearTo*is another useful property that adds the functionality of proximity to any pixel. The distance threshold used was 500 m between pixels (500 m is the spatial resolution of MODIS snow products). This property is symmetric because when a pixel A is near B, B is also near A.

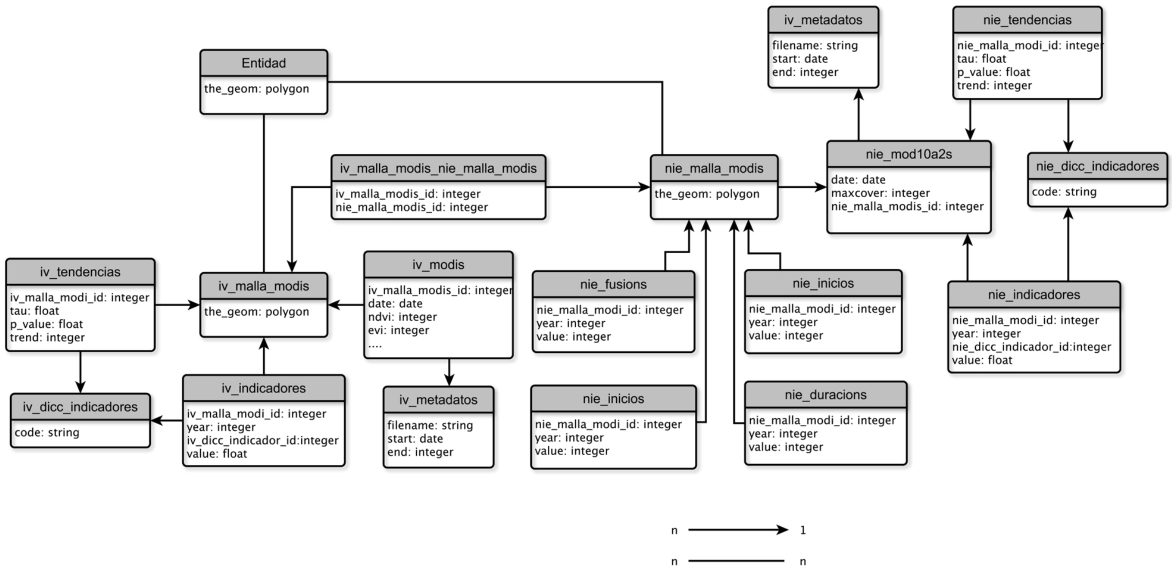

Figure 2.3.4: Database schema. For each MODIS product the relational model stores three types of information: i) spatial distribution of the pixels; ii) values of NDVI and NDSI from original HDF files; and iii) the metadata associated with each original image. The database also contains an auxiliary table to manage spatial entities (i.e. *Q. pyrenaica* patches). Finally, there was a set of tables containing the aggregated information and indicators obtained after processing the raw data

Indicators. This part contains a concept (*IndicatorValues*) that represents the different values that take an indicator (see Section 2.3.2.2) at a given time point (through the concept called *time: Year*and the property *HasYear*) and in a given place (through the concept *Pixel*and the property *IndicatorValuesLocateInPixel*). We have also included a concept to describe all the indicators (*Snow-cover duration, Snow-cover onset date, NDVI_i annual, Maximum NDVI,*etc.). These concepts are grouped according to their thematic area (*Snow*and *Vegetation*). Each indicator has a property called *value* that is measured using a given specific unit.

The temporal trends are described in a concept called *IndicatorTrend.*This concept shows the temporal trend of a single point for the whole time series (it is linked to *Pixel*via *PixelHasIndicatorTrends*). We have also created a concept for each temporal trend calculated for the previously described indicators (*Trend of Snow cover duration*, *Trend NDVI_i annual*, etc.). These concepts are also grouped according to their thematic area (*Snow Trend*, *Vegetation Trend*). All these concepts have the following properties:

- *value_tau* and *p_value*: These properties contain the statistic (*value_tau*) and the significance (*p_value*) reached by the Mann-Kendall trend analysis.
- *value_trend*: Categorical property ranges from –1 (significant negative trend) to 1 (significant positive trend). It is calculated according to the values of *value_tau*and *p_value.*

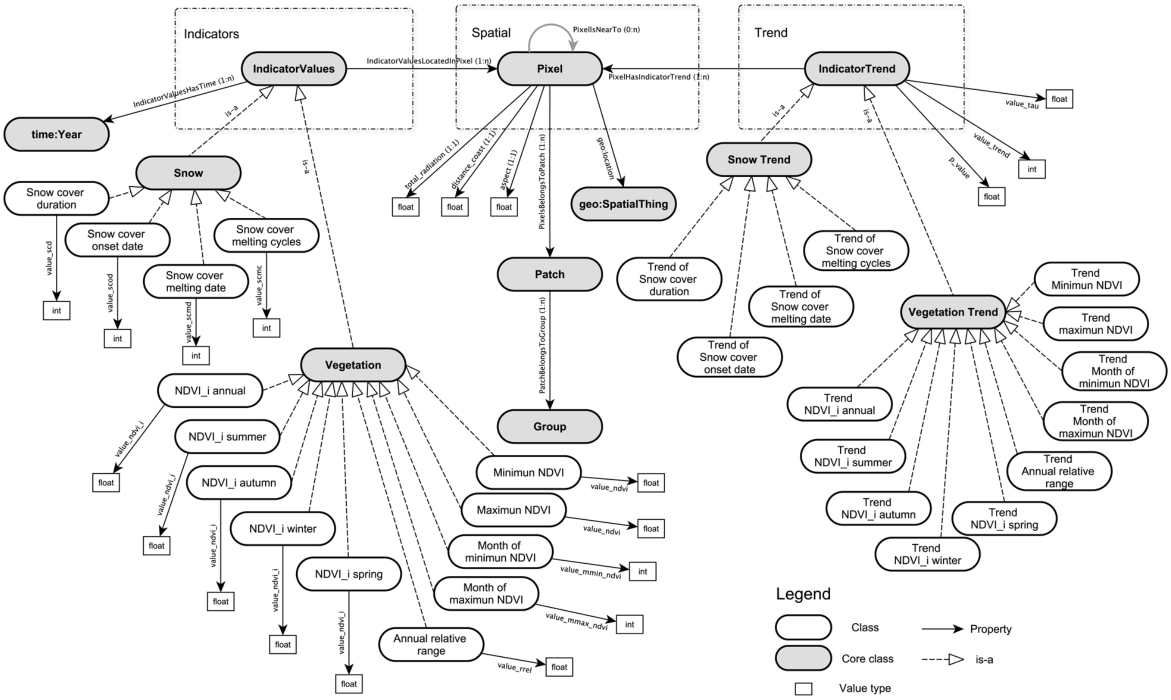

Figure 2.3.5: Detailed representation of the ontology created. Three main parts are considered: spatial information, indicators, and temporal trends of the indicators

This schema was implemented using OWL DL (Description logic) that allows an enhanced expression level and does not limit the values for cardinality (Smith et al., 2004). The structure of the ontology created can be downloaded following this link: http://iecolab.es/indicators.rdf

Knowledge base, SPARQL endpoint and inference

The next step after creating the ontology is to map the records in the database that contain the data to the ontology. Firstly, we used D2RQ (Bizer et al., 2004) to map the relational database to OWL ontology. This software allows instance data to be retrieved from relational databases on-the-fly during the execution of SPARQL queries. Nevertheless, this procedure is time consuming and demands powerful computational capabilities. Thus, we dumped the mapping created with D2RQ into an intermediate N-triples file to avoid this drawback (Sarkar et al., 2011). This file was created with the data existing in the database and has all the triplets contained in the knowledge base.

To store the knowledge base, our tests with the open-source Apache Fuseki and Jena (http://jena.apache.org/) frameworks yielded unsuccessful results as soon as the data volume started to grow. Because we need an efficient implementation that can be scaled to large, enterprise-class data (Wilkinson et al., 2003), we also conducted some tests with AllegroGraph (http://www.franz.com/agraph/allegrograph/) and Virtuoso (http://virtuoso.openlinksw.com/), choosing the former option because of its capabilities and user-friendly management environment. This software is a triplestore that uses a graph database and it has the ability to encode values directly into its triples.

To enhance the results of the queries, a reasoning task can be also triggered within the generation of the system output process. AllegroGraph provides a built-in inference engine that derives implicit information from the knowledge base. Thus, users can easily turn it on by toggling that option in the query builder interface to enrich their queries. The inference engine is useful to find relations on different types of indicators and other implicit properties such as *PixelIsNearTo*. For example, *Savia* can answer questions concerning implicit knowledge of pixels with a positive trend on seasonal mean of NDVI near others with a negative trend in snow-cover melting dates.

Study Case

For the improvement of the conservation status of habitats, it is necessary to implement management plans according to the Annex 6 of the Habitats Directive. Our system provides knowledge useful to design those management plans. We have used *Quercus pyrenaica* forests in Sierra Nevada (Spain) to explore the importance of snow duration in the functioning of *Q. pyrenaica* forests. We have chosen this habitat as a case study for two reasons: a) its interesting ecological dynamics (deciduous forest in a Mediterranean mountain) and b) the need to manage these forests in a global-change context.

We have structured the case study according to three questions that will provide two types of results. Some of them will help in the understanding of the ecological functioning of the target habitat. And others will demonstrate how ontologies are useful tools to make remote sensing information more accessible for non-expert users.

Which pixels show a trend towards higher productivity in summer?

Q. pyrenaica forests show a well-defined growth season centred in summer (Alcaraz-Segura et al., 2006; Dionisio et al., 2012). Some works have pointed out changes in habitat functioning: increase in annual vegetation greenness in Sierra Nevada (Alcaraz-Segura et al., 2008, 2010) and seasonal functional changes in *Q. pyrenaica* woodlands (Martínez and Gilabert, 2009), during the last decade.

This question aims to explore whether our target habitat is undergoing changes in summer productivity, specifically which *Q. pyrenaica* forests of Sierra Nevada have shown a positive trend of the value of summer productivity (summer NDVI).

Which pixels show a trend towards an earlier snowmelt?

Several studies have pointed out a trend towards higher temperatures and lower precipitation for the Mediterranean area (García-Ruiz et al., 2011; Giorgi and Lionello, 2008). Significant declines in snow-cover extent and duration has been reported in some European mountains (Marty, 2008; Moreno-Rodríguez, 2005; Nikolova et al., 2013; Scherrer et al., 2004). Climate projections forecast an increase of +4.8ºC at the end of the 21st century (Benito et al., 2011) for Sierra Nevada and it expected that snowmelt will occur earlier in the year and will be more rapid (García-Ruiz et al., 2011).

The second question that we raised it concerns the observed changes in snowpack in Sierra Nevada. We are interested specifically in which pixels show a trend towards earlier snowmelt during spring-summer. This question is crucial, given that *Q. pyrenaica* forests need water in summer for growth.

Which* Q. pyrenaica *patches show a trend towards a more productive summer and earlier snowmelt?

This question explores the relationships and co-occurrence between biological production and snow-cover features.

Snow-related variables can explain the distribution of plant communities in the landscape (Jones et al., 2001). This causal relationship is more important at high elevation (Bonet and Cayuela, 2009) in Sierra Nevada. But snow cover also explains part of the ecosystem functioning. Trujillo et al. (2012) observed that vegetation greenness increases with snow accumulation. This relationship varies with elevation, reaching a maximum between 2000–2600 m.

Some works have pointed out the influence of snow on greenness in Pyrenean oak forests (Alcaraz-Segura et al., 2009; Dionisio et al., 2012), but to date we have found no studies that analyse the coupling between snow cover and forest greening. Water availability is a key issue on the distribution of *Q. pyrenaica* (del Río et al., 2007; Gavilán et al., 2007). This combination of plant growth and water scarcity makes summer a critical season for the functioning of this habitat.

The third question assesses the capacity of our ontology to show relationships similar to those described above. We have explored the co-occurrence of significant trends in biological production and snow-cover melting date in *Q. pyrenaica* forests. In other words, we have analysed which *Q. pyrenaica* forests show a trend towards higher productivity and earlier melting date in summer.

Results

We translated the above questions from natural language into ontology. For the first and second questions (Sections 4.1 and 4.2) we used two concepts (*Pixel, IndicatorTrend*) and some properties describing these concepts (*value_trend, value_tau, PixelHasIndicatorTrends, Trend NDVI_i summer, Trend of Snow cover melting date*) included in the ontology. Specifically:

- **"select all *Pixel* where *IndicatorTrend* is positive for summer NDVI-I indicator"**for question 4.1 (Figure 2.3.6a)
- **"select all *Pixel* where *IndicatorTrend* is negative for snow-cover melting date"**for question 4.2 (Figure 2.3.6b)

We used SPARQL language to query the knowledge base.

Regarding the first question, we found that 75% of pixels had a positive significant trend for summer NDVI (Figure 2.3.6a). For these, more than 80% showed a strong or very strong positive trend. In general, *Q. pyrenaica* patches located on the north face of Sierra Nevada showed a higher amount of significant pixels than the southern ones did (see map in Figure 2.3.6a).

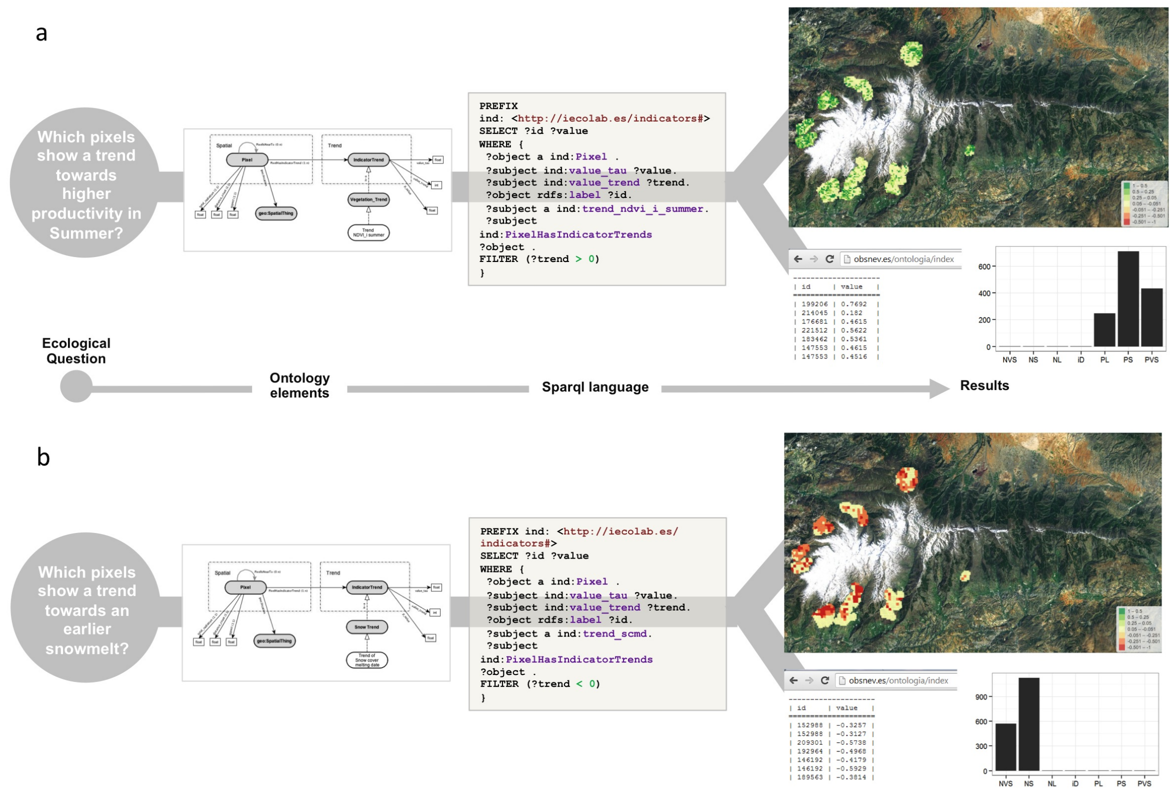

Figure 2.3.6: Scheme showing how the ontology solves questions regarding habitat functioning (a) and the behaviour of an abiotic factor (b). For each ecological question, different ontology elements are used to answer it. Then SPARQL language is used to query the knowledge base. Finally, results can be shown in different formats: map, csv or histogram

The second question showed that almost 70% of the pixels covered by *Q. pyrenaica* forests had a strong or very strong negative and significant trend towards an earlier melting date (Figure 2.3.6b). Similar to NDVI, the northern patches showed a higher amount of significant pixels than the southern ones.

The third question is more difficult to translate to the ontology because it takes into account two datasets and more concepts than the previous questions. We have included a concept called *Patch*, being a subset of pixels that share some ecological features (they belong to the same *Q. pyrenaica* population). This question also includes other concepts already mentioned (*Pixel, IndicatorTrend*) and properties describing those concepts (*value_trend, value_tau, PixelHasIndicatorTrends*). We also calculated the percentage of pixels per *Patch* that showed trends towards more productive summers and earlier snowmelt. These elements were used to translate the original question to another one that was more suitable for the ontology: **select all Pixels where the *IndicatorTrend* is positive for the summer NDVI-I indicator**(*Trend NDVI_i summer*) **and negative for snow-cover melting date***(*Trend of Snow-cover melting date). We used SPARQL language to query the knowledge base. The results can be displayed both in a map and table format (Figure 2.3.7).

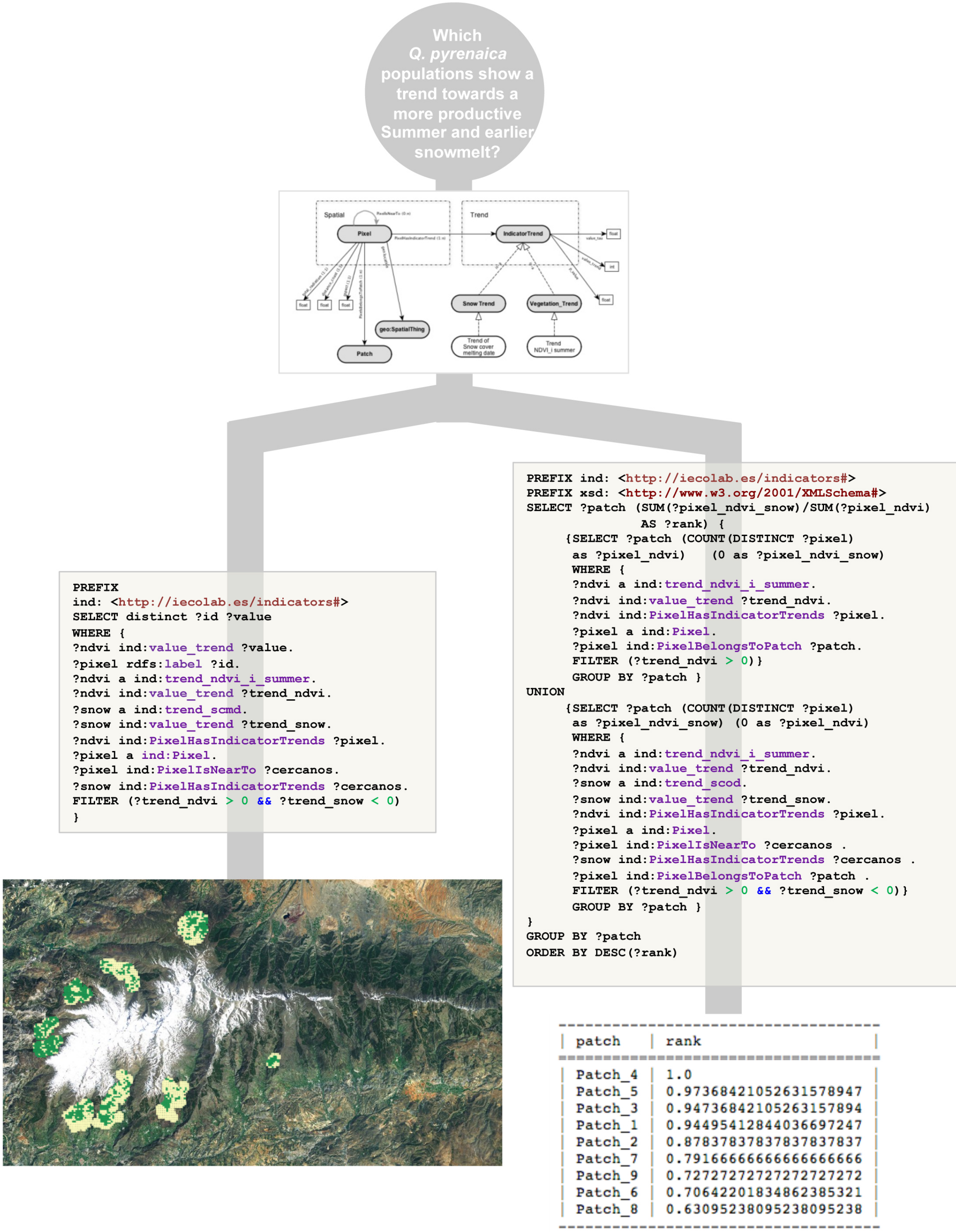

Figure 2.3.7: Scheme showing the process of answering a complex query by the ontology

Savia provides two types of answers for this question: a) A table (Figure 2.3.7) shows the different *Q. pyrenaica* patches ranked according to the percentage of pixels having the described trends in summer productivity and snow-cover melting date. b) A binary map showing the pixels (grouped by *Patch*) that satisfy both conditions (Figure 2.3.7). All the patches that share the same behaviour are considered as *Groups*.

Discussion and Conclusions

The system that we have created adds a semantic component to remote-sensing images using ontologies to describe this information. *Savia* is an operational system that is available for any user via the web (http://obsnev.es/ontologia/index).

Our system implements a query builder user interface that allows users to build questions using SPARQL. It also includes a set of predefined questions to show its capabilities. Furthermore, users can select different output file formats to display results (csv, text or map). All the analytical procedures needed to run this system have been documented using a model repository called ModeleR (Bonet et al., 2014; Pérez-Pérez et al., 2012). The ontology created reuses and extends public ontologies like OWL-Time and Basic Geo (WGS84 lat/long) Vocabulary. The database containing MODIS images was translated into facts within a knowledge base. This requires a mapping between the database and the concepts contained in the ontology. The dynamical queries to knowledge base, using the mapping tool, were one of the most relevant bottlenecks that we have found during the implementation of the system, and we finally used enterprise-ready software to optimise queries to the knowledge base. We also used an inference engine to solve complex queries that require using advanced properties in the ontology (transitivity and symmetry, mainly).

We tested the ontological system in a case study focusing on *Q. pyrenaica* habitat in Sierra Nevada. We identified significant trends in summer NDVI for 75% of pixels covered by the target habitat. These pixels were located mainly in northern-faced patches (aspect was calculated using DEM). These results could be explained by a different pattern of summer productivity among the *Q. pyrenaica* patches. We have also described similar trends in snow patterns: 70% of pixels show a significant and negative trend towards an earlier melting date. Most of those pixels are also located in northerly facing patches. This result could have several hydrological and ecological implications: a) water from the melted snow is available for vegetation earlier each year, which could help deciduous trees to overcome the summer drought, b) the ground is free of snow during a longer period each year, which could provide extra area to treeline communities for altitudinal shifts.

The ontology has also helped to unveil the co-occurrence of significant trends both in snow cover (abiotic factor) and ecosystem functioning (NDVI). Thus, western patches display a high percentage of pixels showing this co-occurrence. The ecological implications of this co-occurrence can be explained by arguing that the earlier snowmelt provides water to *Q. pyrenaica* trees when they are in the middle of their growing season. This earlier amount of water supply encourages trees to be more productive in summer. On the other hand, the southern patches also show this co-occurrence in the opposite way: The lack of significant trends in summer productivity for southern patches could be explained by the lack of pixels with trends towards earlier snowmelt in these areas. Although these results are still preliminary, we have established a link between the status of an abiotic factor and the functioning of ecosystems. Some forest activities can be scheduled according to the trends observed. It could be useful, for example, to reinforce the western patches by planting *Q. pyrenaica* trees. These new trees could take advantage of the productive summers in order to create denser forests. These ecological results are similar to others found in different habitats (Trujillo et al. 2012).

The results (both ecological and methodological) demonstrate that the information in the MODIS time series is useful to assess the functioning of a terrestrial Natura 2000 habitat. We have described the temporal behaviour of *Q. pyrenaica* forests in Sierra Nevada, distinguishing among patches located in areas with different environmental conditions. We have also showed temporal trends in several functioning indicators. The trends discovered would help managers to assess the conservation status of this habitat. They can also build management plans using the knowledge provided by our ontology (i.e. to decide where to locate plantations taking into account the productivity trends). We have also described the behaviour of a key abiotic factor: snow cover; and we calculated trends for several snow-cover related indicators (snow duration, snow-cover melting date, etc.). Those could help managers to identify places where snow-cover trends could change in the coming years. Finally, we have detected relationships between trends in habitat productivity and snow-cover melting date for the target habitat. All this knowledge is offered to users (mainly managers and scientists) through a web portal, the use of which does not require expertise in remote sensing. Thus, we believe that this work is a worthwhile example of a web-based expert system created using an interdisciplinary approach.

References

Alcaraz-Segura, D., Cabello, J., Paruelo, J., 2009. Baseline characterization of major Iberian vegetation types based on the NDVI dynamics. Plant Ecol. 202, 13–29, http://dx.doi.org/10.1007/s11258–008–9555–2.

Alcaraz-Segura, D., Cabello, J., Paruelo, J.M., Delibes, M., 2008. Trends in the surface vegetation dynamics of the national parks of Spain as observed by satellite sensors. Appl. Veg. Sci. 11, 431–440, http://dx.doi.org/10.3170/2008–7–18522.

Alcaraz-Segura, D., Liras, E., Tabik, S., Paruelo, J., Cabello, J., 2010. Evaluating the consistency of the 1982–1999 NDVI trends in the Iberian Peninsula across four time-series derived from the AVHRR sensor: LTDR, GIMMS, FASIR, and PAL-II. Sensors 10, 1291–1314, http://dx.doi.org/10.3390/s100201291.

Alcaraz-Segura, D., Paruelo, J., Cabello, J., 2006. Identification of current ecosystem functional types in the Iberian Peninsula. Glob. Ecol. Biogeogr. 15, 200–212, http://dx.doi.org/10.1111/j.1466–822X.2006.00215.x.

Antoniou, G., van Harmelen, F., 2004. A semantic web primer. The MIT Press, Cambridge.

Arvor, D., Duriex, L., Andrés, S., Laporte, M.A., 2013. Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective. ISPRS J. Photogramm. 82, 125–137, http://dx.doi.org/10.1016/j.isprsjprs.2013.05.003.

Bard, J.B.L., Rhee, S.Y., 2004. Ontologies in biology: design, applications and future challenges.

Nat. Rev. Genet. 5, 213–222, http://dx.doi.org/10.1038/nrg1295.

Benito, B., 2009. Ecoinformática apicada a la conservación: simulación de efectos del cambio global en la distribución de la flora de Andalucía. Thesis. Univerisity of Granada.

Benito, B., Lorite, J., Peñas, J., 2011. Simulating potential effects of climatic warming on altitudinal patterns of key species in mediterranean-alpine ecosystems. Clim. Change 108, 1–13, http://dx.doi.org/10.1007/s10584–010–0015–3.

Bizer, C., Seaborne, A., Berlin, F.U., Labs, H., 2004. D2RQ-treating non-RDF databases as virtual RDF graphs, In: Proceedings of the 3rd International Semantic Web Conference (ISWC2004).

Blanca, G., Cueto, M., Martínez-Lirola, M.J., Molero-Mesa, J., 1998. Threatened vascular flora of Sierra Nevada (Southern Spain). Biol. Conserv. 85, 269–285, http://dx.doi.org/10.1016/S0006–3207(97)00169–9.

Blanca, G., Mendoza, R., 2000. Libro rojo de la flora silvestre amenazada de Andalucía: especies vulnerables. Junta de Andalucía, Consejería de Medio Ambiente, Sevilla.

Bonet, F.J., Aspizua, R., Zamora, R., Sánchez, F.J., Cano-Manuel, F.J., Henares, I., 2011. Sierra Nevada Observatory for monitoring global change: towards the adaptive management of natural resources, In: MaB, A.M.A. (Ed.), Biosphere Reserves in the Mountains of the World. Excellence in the Clouds? Austrian Academy of Sciences Press, pp. 48–52.

Bonet, F.J., Cayuela, L., 2009. Seguimiento de la cubierta de nieve en Sierra Nevada: tendencias en la última década y posibles implicaciones ecológicas de las mismas, In: IX Congreso Nacional de La Asociación Española de Ecología Terrestre: La Dimensión Ecológica Del Desarrollo Sostenible: Ecología, Del Conocimiento a La Aplicación. Asociación Española de Ecología Terrestre, Úbeda.

Bonet, F.J., Pérez-Pérez, R., Benito, B.M., de Albuquerque, F.S., Zamora, R., 2014. Documenting, storing, and executing models in Ecology: a conceptual framework and real implementation in a global change monitoring program. Environ. Model. Softw. 52, 192–199, http://dx.doi.org/10.1016/j.envsoft.2013.10.027.

Brickley, D., 2003. Basic Geo (WGS84 lat/long) vocabulary, http://www.w3.org/2003/01/geo/ (accessed November 2013).

Cabello, J., Fernández, N., Alcaraz-Segura, D., Oyonarte, C., Piñeiro, G., Altesor, A., Delibes, M., Paruelo, J., 2012. The ecosystem functioning dimension in conservation: insights from remote sensing. Biodivers. Conserv. 21, 3287–3305, http://dx.doi.org/10.1007/s10531–012–0370–7.

Camacho-Olmedo, M.T., Jiménez-Olivencia, Y., García-Martínez, P., Mentor, J., Paniza, A., 2002. La alta alpujarra granadina en la segunda mitad del siglo XX a través de la cartografía evolutiva de su paisaje: dinámica vegetal y repoblación forestal. Cuad. Geogr. 32, 25–42, http://www.ugr.es/\~cuadgeo/docs/articulos/032/032–002.pdf.

CBD (Secretariat of the Convention on Biological Diversity), 2005. Handbook of the Convention on Biological Diversity, Biological Conservation.

Cleland, E.E., Chuine, I., Menzel, A., Mooney, H.A., Schwartz, M.D., 2007. Shifting plant phenology in response to global change. Trends Ecol. Evol. 22, 357–365, http://dx.doi.org/10.1016/j.tree.2007.04.003.

del Río, S., Herrero, L., Penas, Á., 2007. Bioclimatic analysis of the Quercus pyrenaica forests in Spain. Phytocoenologia 37, 541–560, http://dx.doi.org/10.1127/0340–269X/2007/0037–0541.

Dionisio, M., Alcaraz-Segura, D., Cabello, J., 2012. Satellite-based monitoring of ecosystem functioning in protected areas: recent trends in the oak forests (Quercus pyrenaica Willd.) of Sierra Nevada (Spain), In: Young, S. (Ed.), International Perspectives on Global Environmental Change. pp. 355–374.

EIONET, 2013. Online report on Article 17 of the Habitats Directive: conservation status of habitats and species of Community interest (2001–2006). European Environment Agency, http://bd.eionet.europa.eu/article17/index_html/habitatsummary, (accessed December 2013).

Evans, D., 2012. Building the European Union's Natura 2000 network. Nat. Conserv. 1, 11–26, http://dx.doi.org/10.3897/natureconservation.1.1808.

Fallahi, G.R., Frank, A.U., Mesgari, M.S., Rajabifard, A., 2008. An ontological structure for semantic interoperability of GIS and environmental modeling. Int. J. Appl. Earth Obs. Geoinf. 10, 342–357, http://dx.doi.org/10.1016/j.jag.2008.01.001.

Förster, M., Frick, A., Walentowski, H., Kleinschmit, B., 2008. Approaches to utilising QuickBird data for the monitoring of NATURA 2000 habitats. Community Ecol. 9, 155–168, http://dx.doi.org/10.1556/ComEc.9.2008.2.4.

Franco, J., 1990. Quercus L., In: Castroviejo, A., Laínz, M., López-González, G., Montserrat, P., Muñoz-Garmendia, F., Paiva, J., Villar, L. (Eds.), Flora Iberica. Vol. II. Real Jardín Botánico CSIC, Madrid, pp. 15–36.

García, I., Jiménez, P., 2009. 9230 Robledales de Quercus pyrenaica y robledales de Quercus robur y Quercus pyrenaica del Noroeste ibérico, In: Ministerio de Medio Ambiente, y Medio Rural y Marino (Ed.), Bases Ecológicas Preliminares Para La Conservación de Los Tipos de Hábitat de Interés Comunitario En España. Ministerio de Medio Ambiente, y Medio Rural y Marino. Madrid.

García-Ruiz, J.M., López-Moreno, J.I., Vicente-Serrano, S.M., Lasanta–Martínez, T., Beguería, S., 2011. Mediterranean water resources in a global change scenario. Earth-Science Rev. 105, 121–139, http://dx.doi.org/10.1016/j.earscirev.2011.01.006.

Gavilán, R., Sánchez Mata, D., Vilches, B., Entrocassi, G., 2007. Modeling current distribution of Spanish Quercus pyrenaica forests using climatic parameters. Phytocoenologia 37, 561–581, http://dx.doi.org/10.1127/0340–269X/2007/0037–0561.

Giorgi, F., Lionello, P., 2008. Climate change projections for the mediterranean region. Glob. Planet. Change 63, 90–104, http://dx.doi.org/10.1016/j.gloplacha.2007.09.005.

Gruber, T., 1993. A translation approach to portable ontology specifications. Knowl. Acquis. 5 (2), 199–220, http://dx.doi.org/10.1006/knac.1993.1008.

Hall, D., Riggs, G., Salomonson, V.V., DiGirolamo, N., Bayr, K., 2002. MODIS snow-cover products. Remote Sens. Environ. 83, 181–194, http://dx.doi.org/10.1016/S0034–4257(02)00095–0.

Hashimoto, S., Tadono, T., Onosato, M., Hori, M., Moriyama, T., 2011. A framework of ontology-based knowledge information processing for change detection in remote sensing data, In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 3927–3930.

Hayes-Roth, F., Waterman, D.A., Lenat, D.B., 1983. Building Expert Systems. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Hoare, D., Frost, P., 2004. Phenological description of natural vegetation in southern Africa using remotely-sensed vegetation data. Appl. Veg. Sci. 7(1), 19–28, http://dx.doi.org/10.1111/j.1654–109X.2004.tb00591.x.

Hobbs, J.R., Pan, F., 2004. An ontology of time for the semantic web. ACM Trans. Asian Lang. Inf. Process. 3, 66–85, http://dx.doi.org/10.1145/1017068.1017073.

Hobbs, J.R., Pan, F., 2006. Time Ontology in OWL. http://www.w3.org/TR/owl-time/, (accessed November 2013).

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213, http://dx.doi.org/10.1016/S0034–4257(02)00096–2.

Hyde, P., Dubayah, R., Walker, W., Blair, J.B., Hofton, M., Hunsaker, C., 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sens. Environ. 102, 63–73, http://dx.doi.org/10.1016/j.rse.2006.01.021.

Jones, H., Pomeroy, J., Walker, D., Hoham, R. (Eds.), 2001. Snow ecology: an interdisciplinary examination of snow-covered ecosystems, Cambridge, United Kingdom.

Jones, M.B., Schildhauer, M.P., Reichman, O.J., Bowers, S., 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annu. Rev. Ecol. Evol. Syst. 37, 519–544, http://dx.doi.org/10.1146/annurev.ecolsys.37.091305.110031.

Justice, C., Townshend, J.R., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., Morisette, J., 2002. An overview of MODIS Land data processing and product status. Remote Sens. Environ. 83, 3–15, http://dx.doi.org/10.1016/S0034–4257(02)00084–6.

Kalluri, S., Gilruth, P., Bergman, R., 2003. The potential of remote sensing data for decision makers at the state, local and tribal level: experiences from NASA's Synergy program. Environ. Sci. Policy 6, 487–500, http://dx.doi.org/10.1016/j.envsci.2003.08.002.

Kendall, M., 1975. Rank correlation methods. London.

Larin Fonseca, R., Garea Llano, E., 2011. Automatic representation of geographical data from a semantic point of view through a new ontology and classification techniques. Trans. GIS 15, 61–85, http://dx.doi.org/10.1111/j.1467–9671.2010.01242.x/.

Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. Biol. Conserv. 143, 1317–1328, http://dx.doi.org/10.1016/j.biocon.2010.02.013.

Lloyd, D., 1990. A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. Int. J. Remote Sens. 11, 2269–2279, http://dx.doi.org/10.1080/01431169008955174.

Louette, G., Adriaens, D., Adriaens, P., Anselin, A., Devos, K., Sannen, K., Van Landuyt, W., Paelinckx, D., Hoffmann, M., 2011. Bridging the gap between the Natura 2000 regional conservation status and local conservation objectives. J. Nat. Conserv. 19, 224–235, http://dx.doi.org/10.1016/j.jnc.2011.02.001.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F., 2007. An ontology for describing and synthesizing ecological observation data. Ecol. Inform. 2, 279–296, http://dx.doi.org/10.1016/j.ecoinf.2007.05.004.

Madin, J.S., Bowers, S., Schildhauer, M.P., Jones, M.B., 2008. Advancing ecological research with ontologies. Trends Ecol. Evol. 23, 159–68, http://dx.doi.org/10.1016/j.tree.2007.11.007.

Mann, H., 1945. Nonparametric tests against trend. Econometrica 13, 245–259.

Martínez, B., Gilabert, M.A., 2009. Vegetation dynamics from NDVI time series analysis using the wavelet transform. Remote Sens. Environ. 113, 1823–1842, http://dx.doi.org/10.1016/j.rse.2009.04.016.

Marty, C., 2008. Regime shift of snow days in Switzerland. Geophys. Res. Lett. 35, L12501, http://dx.doi.org/10.1029/2008GL033998.

McLeod, A., 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. http://CRAN.R-project.org/package=Kendall, (accessed July 2013).

Melendo, M., Valle, F., 1996. Estudio comparativo de los melojares nevadenses, In: Chacón, J., Rosúa, J.L. (Eds.), 1a Conferencia Internacional Sierra Nevada, Conservación y Desarrollo Sostenible. p. vol. 2. 463–479.

Moreno-Rodriguez, J. (Ed.), 2005. Evaluación preliminar de los impactos en España por efecto del cambio climático. Ministerio de Medio Ambiente, Madrid.

Nikolova, N., Faško, P., Lapin, M., Švec, M., 2013. Changes in snowfall/precipitation-day ratio in Slovakia and their linkages with air temperature and precipitation. Contrib. to Geophys. Geod. 43, 141–145, http://dx.doi.org/10.2478/congeo–2013–0009.

Oliva-Santos, R., Maciá-Pérez, F., Garea-Llano, E., 2014. Ontology-based topological representation of remote-sensing images. Int. J. Remote Sens. 35, 16–28, http://dx.doi.org/10.1080/01431161.2013.858847.

Paruelo, J., Lauenroth, W., 1995. Regional patterns of normalized difference vegetation index in

North American shrublands and grasslands. Ecology 76, 1888–1898, http://dx.doi.org/10.2307/1940721.

Paruelo, J.M., Epstein, H.E., Lauenroth, W.K., Burke, I.C., 1997. ANPP estimates from NDVI for the central grassland region of the United States. Ecology 78, 953–958, http://dx.doi.org/10.2307/2266073.

Pereira, H.M., David Cooper, H., 2006. Towards the global monitoring of biodiversity change. Trends Ecol. Evol. 21, 123–129, http://dx.doi.org/10.1016/j.tree.2005.10.015.

Pérez-Pérez, R., Benito, B.M., Bonet, F.J., 2012. ModeleR: an enviromental model repository as knowledge base for experts. Expert Syst. Appl. 39, 8396–8411, http://dx.doi.org/10.1016/j.eswa.2012.01.180.

Piñeiro, G., Oesterheld, M., Paruelo, J., 2006. Seasonal variation in aboveground production and radiation-use efficiency of temperate rangelands estimated through remote sensing. Ecosystems 9, 357–373, http://dx.doi.org/10.1007/s10021–005–0013-x.

Potter, C., Brooks, V., 1998. Global analysis of empirical relations between annual climate and seasonality of NDVI. Int. J. Remote Sens. 19, 2921–2948, http://dx.doi.org/10.1080/014311698214352.

R Core Team, 2013. R: a language and environment for statistical computing. http://www.r-project.org/, (accessed July 2013)

Renear, A.H., Palmer, C.L., 2009. Strategic reading, ontologies, and the future of scientific publishing. Science 325, 828–832, http://dx.doi.org/10.1126/science.1157784.

Rittger, K., Painter, T.H., Dozier, J., 2013. Assessment of methods for mapping snow cover from MODIS. Adv. Water Resour. 51, 367–380, http://dx.doi.org/10.1016/j.advwatres.2012.03.002.

Runnning, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience 54, 547–560, http://dx.doi.org/10.1641/0006–3568(2004)054%5B0547:ACSMOG%5D2.0.CO;2.

Salomonson, V., Appel, I., 2006. Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. IEEE Trans. Geosci. Remote Sens. 44, 1747–1756, http://dx.doi.org/10.1109/TGRS.2006.876029.

Sarkar, A., Marjit, U., Biswas, U., 2011. Linked data generation for the university data from legacy database. Int. J. Web Semant. Techn. 2, 21–31, http://dx.doi.org/10.5121/ijwest.2011.2302.

Scherrer, S.C., Appenzeller, C., Laternser, M., 2004. Trends in Swiss Alpine snow days: the role of local- and large-scale climate variability. Geophys. Res. Lett. 31, L13215, http://dx.doi.org/10.1029/2004GL020255.

Sellers, P.J., Berry, J.A., Collatz, G.J., Field, C.B., Hall, F.G., 1992. Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. Remote Sens. Environ. 42, 187–216, http://dx.doi.org/10.1016/0034–4257(92)90102-P.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D., Bounoua, L., 1996. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: model formulation. J. Clim. 9, 676–705, http://dx.doi.org/10.1175/1520–0442(1996)009<0676:ARLSPF>2.0.CO;2.

Silva, M.P.S., Camara, G., Souza, R.C.M., Valeriano, D.M., Escada, M.I.S., 2005. Mining patterns of change in remote sensing image databases. Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05). IEEE. Computer Society, http://dx.doi.org/10.1109/ICDM.2005.98.

Smith, M.K., Welty, C., McGuinness, D.L., 2004. OWL web ontology language. http://www.w3.org/TR/owl-guide/, (accessed November 2013).

Trujillo, E., Molotch, N.P., Goulden, M.L., Kelly, A.E., Bales, R.C., 2012. Elevation-dependent influence of snow accumulation on forest greening. Nat. Geosci. 5, 705–709, http://dx.doi.org/10.1038/ngeo1571.

Tucker, C.J., Townshend, J.R.G., Goff, T.E., 1985. African land-cover classification using satellite data. Science 227, 369–375, http://dx.doi.org/10.1126/science.227.4685.369.

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M.H., Kirschbaum, A.A., Ham, J.M., Saleska, S.R., Ahl, D.E., 2006. Evaluation of MODIS, NPP and GPP products across multiple biomes. Remote Sens. Environ. 102, 282–292, http://dx.doi.org/10.1016/j.rse.2006.02.017.

Vanden Borre, J., Paelinckx, D., Mücher, C.A., Kooistra, L., Haest, B., De Blust, G., Schmidt, A.M., 2011. Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. J. Nat. Conserv. 19, 116–125, http://dx.doi.org/10.1016/j.jnc.2010.07.003.

W3C, 2013. Large triple stores. http://www.w3.org/wiki/LargeTripleStores, (accessed November 2013).

Wang, X., Xie, H., 2009. New methods for studying the spatiotemporal variation of snow cover based on combination products of MODIS Terra and Aqua. J. Hydrol. 371, 192–200, http://dx.doi.org/10.1016/j.jhydrol.2009.03.028.

Wang, L., Sousa, W.P., Gong, P., Biging, G.S., 2004. Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. Remote Sens. Environ. 91, 432–440, http://dx.doi.org/10.1016/j.rse.2004.04.005.

Wiegand, N., García, C., 2007. A task-based ontology approach to automate geospatial data retrieval. Trans. GIS 11, 355–376, http://dx.doi.org/10.1111/j.1467–9671.2007.01050.x.

Wilkinson, K., Sayers, C., Kuno, H. A., Reynolds, D., 2003. Efficient RDF storage and retrieval in Jena2, In: Proceedings of First International Workshop on Semantic Web and Databases. Vol. 3, pp. 131–150.

Williams, R.J., Martinez, N.D., Golbeck, J., 2006. Ontologies for ecoinformatics. Web Semant. Sci. Serv. Agents World Wide Web 4, 237–242, http://dx.doi.org/10.1016/j.websem.2006.06.002.

Xue, Y., Chen, Z., Xu, H., Ai, J., Jiang, S., Li, Y., Wang, Y., Guang, J., Mei, L., Jiao, X., He, X., Hou, T., 2011. A high throughput geocomputing system for remote sensing quantitative retrieval and a case study. Int. J. Appl. Earth Obs. Geoinf. 13, 902–911, http://dx.doi.org/10.1016/j.jag.2011.06.006.

2.4 Pattern recognition in shoreline dataset

Introduction

Coasts represent some of the most dynamic environments on earth, and 84% of the countries of the world have a coastline adjacent to open oceans, inland seas or both (Martínez et al., 2007). Coastlines develop a wide range of morphologies depending on many factors, such as forcing conditions, characteristics of the forming materials and regional geology (Carter and Woodroffe, 1997; Woodroffe, 2002). Bird (2010) recently published an extensive updated revision of worldwide coastal morphologies. Global and regional studies have identified and analyzed the main characteristics of such coastal morphologies (e.g., cliffs, beaches, estuaries, lagoons). Among them, beaches have been the most studied because of their social, economic and environmental interest (Davis and Fitzgerald, 2009). Beaches are frequently found on rectilinear or slightly curved coasts, although they can also appear in other environments, such as estuaries, deltas or river mouths (Bird, 2011). Beaches have been intensively analyzed morphologically, with special attention given to both the beach form and profile (Dean and Dalrymple, 2002) and the morphological state (Masselink and Short, 1993).

Many studies have also analyzed the generation, evolution and characteristics of secondary shoreline features (Coco and Murray, 2007) that many beaches exhibit. Secondary features can be parallel or perpendicular to the shoreline and may have different geometrical characteristics (Pethick, 1984). Recent attention has focused on shoreline undulations (hereafter referred to as SUs), which can be defined as medium to large spatial-scale shoreline features that have longshore dimensions ranging from hundreds to thousands of meters and cross-shore widths ranging from tens to hundreds of meters (Ortega-Sánchez et al., 2015; see Figure 2.4.3). SUs are generally classified as rhythmic coastline features, although some are neither periodic nor regularly spaced (López-Ruiz et al., 2012b). SUs are frequently associated with sudden changes in the orientation of the coast, such as at spits (Kaergaard and Fredsoe, 2012), and are often located in proximity to human infrastructures (López-Ruiz et al., 2012a). Many authors have referred to SUs as shoreline sand waves; however, shoreline sand waves are generally considered to be rhythmically spaced, and they migrate alongshore (Stewart and Davidson-Arnott, 1988; Davidson-Arnott and Van Heyningen, 2003).

The mechanism(s) behind the formation and the dynamics of SU are not well understood, but the working hypothesis adopted in recent years is that coastlines with a wave climate dominated by very oblique incidence waves may be unstable and commonly feature large-scale undulations (Ashton et al., 2001; Ashton and Murray, 2006; Medellín et al., 2008). Recent advances reveal that variation in the alongshore sediment transport with the angle formed by the wave crests and the coastline plays a major role in the development of SU (López-Ruiz et al., 2012b). In this respect, the curvature of the coastline seems to play a key role in the formation of such features (López-Ruiz et al., 2014).

Compared to other general beach characteristics, such as the morphological state (Scott et al., 2011), or other shoreline features, such as beach cusps (Coco et al., 1999), a global compilation of sites that contains SU information for scientific and management purposes is not available. Stakeholders, managers and politicians are all involved in preserving coastal features. From an operational perspective, a massive sample collection of shoreline features can aid in understanding how coastlines form and developing new tools.

The advances in modeling the natural processes responsible for geographic processes are strongly related to the quality, availability and temporal and spatial scale of the data, such as in the cases of forests (Song, 2013), benthic habitats (Godet et al., 2009) and soil mapping (Dewitte et al., 2012). They are generally studied by remote sensing and by the analysis of statistical surveys and field data integrated with historical maps (Carretero et al., 2014; Messerli et al., 2014). Satellite remote sensing is valuable for providing cost-effective information (Kuenzer et al., 2014; Emel et al. 2014; Borrelli et al., 2014). Nevertheless, the satellite spatial resolution can be too coarse for detailed mapping and for distinguishing local variability, while very high-resolution satellite imagery is very expensive. Additionally, the inclusion of this type of data into large and readily accessible geo-referenced databases is rapidly growing, but the methodological frontiers are still advancing.

Developing large databases is not a trivial task because raw data are frequently unavailable, and accumulating and verifying the data are lengthy and tedious tasks (Gajewski, 2008). Additionally, the movement towards open-source databases is not straightforward (Goff and Chague-Goff, 2014). Although databases related to paleoenvironmental, hydrological or land-use are growing (MacDonald et al., 2008; Maetens et al., 2012; Wan et al., 2014), many works demonstrate the frequent absence of databases related to morphological and terrain features, both at regional and global scales (Bhambri and Bolch, 2009; Rozenstein and Karnieli, 2011; Butler et al., 2014). Although many advances have been made during the last few years at local, regional or country scales (i.e., Brown, 2012; Messerli et al., 2014), such databases are frequently based on satellite images covering very few years and commonly have inconsistencies that have to be revised (Portillo-Quintero et al., 2012). Despite these issues, their potential and interest for the scientific community keeps growing with many different examples and applications (i.e., Yu et al., 2013; Delmelle et al., 2014; McCool, 2014; Widener et al., 2014).

The main objective of this work is to explore the potential of using imagery for geographic studies implicating large extensions, with a particular emphasis on coastal studies. To that end, we present a database of identified and characterized SUs created after reviewing a total of approximately 50,000 km of coast in Western Europe and Northwestern Africa. This database is free, public and available for the scientific community and can be used to gain deeper insight into coastal morphodynamic processes. This work also presents analyses that can be obtained using the collected data; this information constitutes a valuable dataset for coastal areas.

Methodology

Region of study

We analyzed the coastlines of Western Europe and Northwestern Africa to identify SUs (Figure 2.4.1). A total of approximately 50,000 km of coast was investigated. For the analysis, we used images from the Google Earth imagery database. Google Earth maps the earth by the superimposition of images obtained from satellite imagery, aerial photography and GIS 3D globe and is freely available. The resolution for the sites analyzed in this work is < 1 meter/pixel; certain images had a resolution on the order of cm/pixel. For sites where the visibility or quality of the images was inadequate, satellite images from other virtual globes such as Bing or Nokia were reviewed.

Google Earth hosts high-resolution imagery and allows the development of practical methods

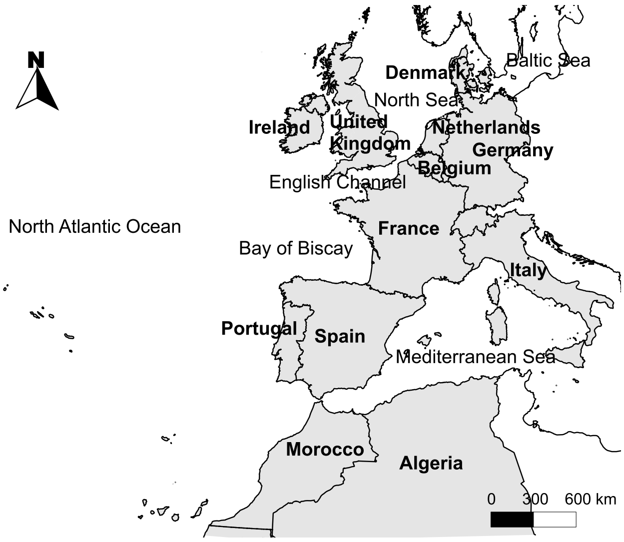

Figure 2.4.1: Coastline of the Western European and Northwestern African countries that were analysed

for studying the regions of interest in which a coarser resolution is insufficient. Google Earth has been recognized for its potential to significantly improve the visualization and dissemination of scientific data since its origin in the year 2005 (Butler, 2006).

Thus, several scientific works have been published that use and compare this new tool with existing systems. Yu and Gong (2012) presented a review of Google Earth in earth sciences research. However, several concerns related to the limitations of Google Earth (Thenkabailc et al., 2007) were briefly noted (i.e., the absence of very-high resolution imagery for every location in the world or the presence of images from varying dates).

Although Google has been unwilling to release detailed information on any of these aspects of their holdings, Potere (2008) stated that the positional accuracy is more than sufficient for medium-scale features and that no geo-correction is required. This study addressed the trustworthiness issue in Google Earth's horizontal positional information via a comparison with that of Landsat GeoCover. The study was carried out in 2008 with 436 control points. Our area of study (i.e., Europe) had an accuracy of 25.7 meters RMSE.

Furthermore, due to the constant evolution of this platform, the imagery has been improved several times over the last few years. Google Earth has attempted to overcome some of the initial limitations. For example, Google has attempted to enhance the imagery with new providers and better resolution. Most of the current providers offer spatial resolutions of less than 10 m. A more recent work studied Google Earth's accuracy compared with that of high-precision field measurements in a region of Texas (Benker et al., 2011), and a horizontal position accuracy of 2.46 m RMSE was determined. This value is an order of magnitude greater than the smaller dimensions of SUs (see the definition of SUs in Section 2.4.1).

Regarding the selection of images, we followed a very strict methodology to ensure that the results were consistent. All of the images used were taken after 2011. Also, with the release

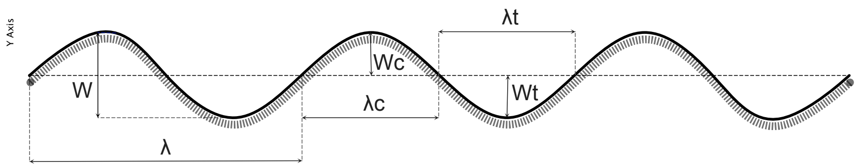

Figure 2.4.2: General geometry of a series of shoreline undulations. W - width; λ - length; Wc - width of the crest; Wt – width of the trough; λc – length of the crest; λt – length of the trough

Reference	λ (m)	W (m)	λ / W
Bruun (1954)	300–2000	60–80	~ 5-25
Verhagen (1989)	2500–22000	30–500	~ 83-44
Thevenot and Kraus (1995)	750	40	~ 19
Schwartz (2006)	500–750	15–25	~ 30-33
Medellín et al. (2008)	125–150	15	~ 8-10
López-Ruiz et al. (2012)	500–2000	150	~ 3-13

Table 2.4.1: Geometrical characteristics of shoreline undulations and/or shoreline sand waves identified in previous works found in the literature

of Google Earth 5 in 2009, historical imagery became a key feature of the platform, allowing users to browse through past and archived images of a given area. Although the available time windows remain quite limited, we were able to compare at least 3 historical images for the sites to verify the data collected about shoreline undulations.

Identification and global information

Because images from Google Earth are processed for visualization purposes, precise quantitative measurements and analyses are lost (Yu and Gong, 2012). In the current work, shorelines were visually analyzed and the undulations were identified based on the ratio λ /W, where λ and W are the length and width of the undulations, respectively (Figure 2.4.2). In this work, any undulation with λ /W ϵ (3–83) was considered an SU (Table 2.4.1). In certain cases, only one undulation was present; however, numerous sites contained series of undulations characterizing the shape of the shoreline (Figure 2.4.3). Because other variables such as the presence of obstacles (constraints) near the SU were identified, this task could not be automated; therefore, it was highly time-consuming and prone to error. Two co-authors of this work independently verified the coastlines of the study regions to minimize possible errors.

For each site where an SU was identified, different data were gathered and included in a general table (Table 2.4.2). We divided the data into four main categories: (1) general information, (2) geometry, (3) constraints and (4) maritime climate. Constraints are the presence of any artificial obstacles at the shoreline located near the SU. López-Ruiz et al. (2012a) demonstrated that these obstacles stop the littoral drift and might trigger the formation of the undulations when they are at a distance lower than 5λ measured from the boundary of the SU closest to the constraint.

Figure 2.4.3: Examples of sites exhibiting shoreline undulations along the European coast. (a) Coutances (France); (b) Crotone (Italy); (c) Catanzaro (Italy); (d) Reggio Calabria (Italy); (e) Vannes (France); (f) Stodov (Denmark); (g) Málaga (Spain); (h) Málaga (Spain)

	Term	Units	Range
General information	Name (Id)	Nominal	-
	Latitude	Coordinates	31.46° -57.93°
	Longitude	Coordinates	-10.21° -18.27°
	Location (city/village, region, country)	Nominal	-
	Ocean/sea	Nominal	-
	Year (of the image)	-	2000 to 2012
Geometry	Number of undulations	-	1 to > 15
	Length	m	30 - 5000
	Width (total)	m	4 - 500
	Length crest/sine	m	10 - 3200
	Width crest/sine	m	2 - 300
Constraints	Type of coast	Nominal	-
	Obstacle (natural or artificial)	Nominal	-
	Distance to the obstacle	m	20 – 8000
Maritime climate	Significant wave height (H_s)	m	0.43 – 1.71
	Peak period (T_p)	s	3.75 – 10.39
	Tidal range	m	0.31 – 4.90
	Wave energy flux (module)	kW/m	0.53 – 13.91
	Wave energy flux (direction)	(°)	-

Table 2.4.2: Main fields included in the general table that characterise the undulations. Maritime climate data only corresponds to sites where information was available to the authors. Expanded fields of the table include: (1) Ocean/Sea: Mediterranean Sea, Atlantic Ocean, Bay of Biscay, English Channel, Irish Sea, North Sea, Skagerrak, Kattegat and Baltic Sea; (2) Type of coast: rectilinear, curvilinear and river mouth or estuary; (3) Obstacle: channelised river mouth, breakwater, pier, harbour and rocks

Data analysis

The data were incorporated into a PostgreSQL database with the PostGIS spatial extension (Zhang and Yi, 2010). QGIS conveniently integrates these representational models and was used to depict the data and develop the different types of maps (QGIS Development Team, 2009). Digital geodatabases allow the accumulation of vast amounts of information that can be readily accessed with simple tools and applied in many fields of applied geography (Rozenstein and Karnieli, 2011; Jäppinen et al., 2013; McCool, 2014). The data for maritime climate included in Table 2 refer to mean or average values of the main characteristic variables (i.e., wave height). Table 2.4.3 summarizes the maritime climate sources depicted in this work.

The information collected was pre-processed using software for data analysis. Intervals were defined for certain variables (λ, W and distance to the constraint), and correlations between groups of two and three variables were attended to, with special attention given to λ/W, with the results compared with published data. Subsequent analyses performed with the collected information were anticipated, and the most representative variables were analyzed (i.e., geometrical characteristics of the undulations).

	Source	Region
Wave Data	Puertos del Estado (2013); Ortega-Sánchez et al. (2008)	Spain and Northern Africa
	Levoy et al. (2000)	French coast (English Channel)
	Beels et al. (2007); Falqués (2006)	German and UK coasts (North Sea)
	Kaergaard et al. (2012)	Danish coast (North Sea)
	Scott et al. (2011)	UK coast (Atlantic Ocean)
	Waters et al. (2009); Soomere et al. (2012)	Danish and German coasts (Kattegat, Skagerrak and Baltic Sea)
	ISPRA - www.idromare.it (2013)	Italian coast
Wind Data	Puertos del Estado (2013)	Spain and Northern Africa
	Windfinder.com (2013)	Rest of countries
Tidal Ranges	Cambio Climático en la Costa Española: Visor C3E (2013)	Spain
	Average values found through internet	Remaining countries

Table 2.4.3: List of the main references and other sources of information that have been used to characterise the maritime climate at the study sites

Beyond basic statistical analysis, a data mining analysis was applied to explore possible similarities between data groups. Certain data mining techniques such as clustering are usually applied in similar contexts. Nevertheless, we have found association rules to be more suitable for our purposes. The main purpose of association rules is to discover significant patterns and regularities in the data. This type of data analysis consists of a non-supervised learning method that does not require a priori knowledge. Association rules can be generally defined as the data-mining task of finding frequently co-occurring items in transactional sets. Association rules obtained from the analysis can be easily interpreted and are widely applied in a myriad of research fields (Rajak and Gupta, 2008).

The R framework was used to perform the data mining process. Despite the many alternatives that have been studied, such as Weka, RapidMiner and Orange, we found the R environment to be the most convenient for our purposes, allowing us to easily pre-process data and retrieve results in a graphical and comprehensive way. In particular, the arules package (Hahsler et al., 2005) was used to discover association rules in our data set, and the arulesviz package (Hahsler and Chelluboina, 2011) was used to visualize and interpret the results.

Association rules were derived from a Apriori algorithm. The algorithm is presented in Agrawal and Srikant (1994) and in order to understand the algorithm completely, some previous background knowledge of association rules is required (transactional databases, frequent item sets, confidence and support of a rule, etc.). Frequent patterns are patterns (e.g., item sets, subsequences, or substructures) that appear frequently in a data set. Finding frequent patterns plays an essential role in mining associations, correlations, and many other interesting relationships among data. The main objetive of Apriori algorithm is mining those frequent item sets efficiently in order to discover significant patterns and regularities in the data. The idea behind the algorithm is attempting to find incrementally frequent item sets, which are common

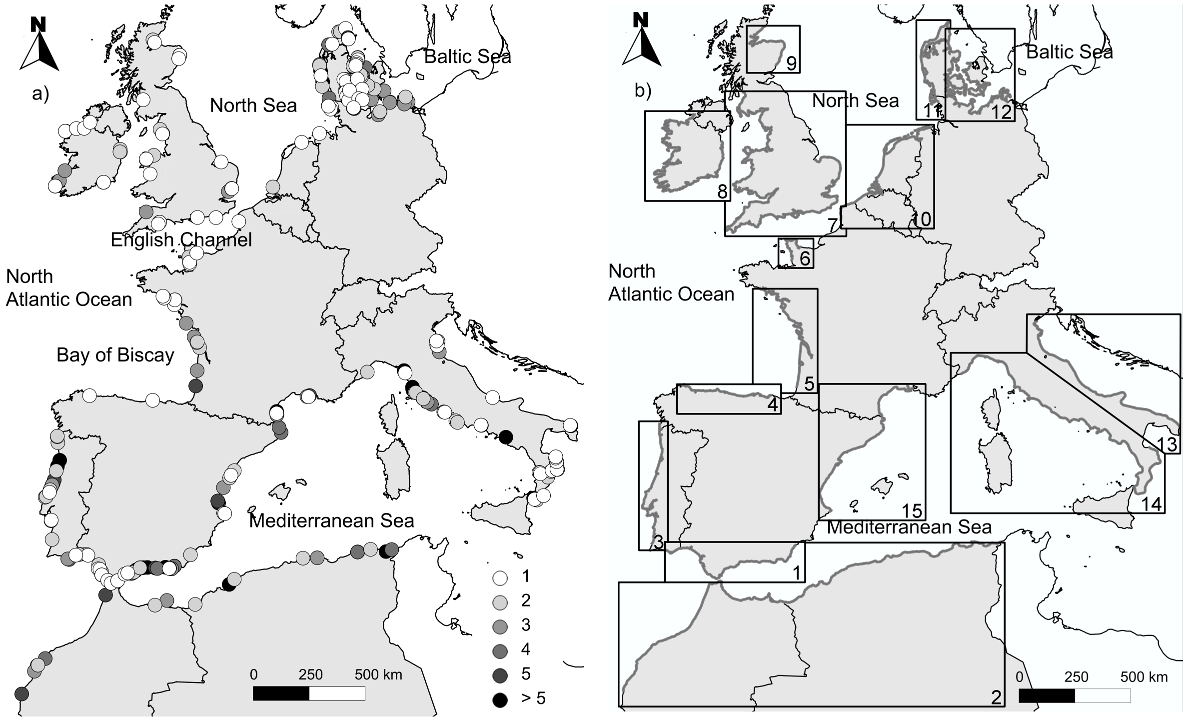

Figure 2.4.4: Maps showing: a) sites where shoreline undulations were identified, where the legend corresponds to the number of individual shoreline undulations at each site; b) Regions (1 to 15) within the area of study that were divided to describe the collected data

to at least a minimum threshold (Borgelt and Kruse, 2002).

General Results

A total of 294 sites exhibiting SUs were identified (Figure 2.4.4a). The information included in the global database and the stored geographical information is offered as a PostGIS layer for the scientific community. Although the information is freely available, registration is required for statistical purposes only (http://gdfa.ugr.es/su_data).

Regional analysis

The majority of the sites exhibiting SUs were found in the Mediterranean Sea (107, 36.4%), whereas 55 (18.7%) were found in the Atlantic Ocean and 50 (17%) at the Kattegat Strait (between the North Sea and the Baltic Sea). The remaining sites face local seas or straits, with proportions lower than 6%. The country with the most sites is Denmark (64, 21.8%), followed by Spain (56, 19%) and Italy (51, 17.3%). For the remaining countries, less than 9% of the sites are identified as exhibiting SUs. To describe the collected data, we divided the study area into regions depending on the number of sites, the global orientation and length of the coastline and the country boundary (Figure 2.4.4b).

Table 2.4.3 shows the average density (d) for each region, defined as the number of sites divided by the length of the coastline (d=sites/km). The greater densities were obtained for

Region	d (sites/km)
1	1/23
2	1/135
3	1/59
4	1/319
5	1/125
6	1/59
7	1/269
8	1/251
9	1/135
10	1/494
11	1/103
12	1/110
13	1/110
14	1/44

Table 2.4.4: The average SU density for each region defined in Figure 2.4.4b

Regions 1 and 14 (both on the Mediterranean Sea), indicating that similar maritime climate conditions and geomorphic characteristics could be responsible for the high number of SUs along both coastlines. In contrast, smaller densities were obtained for Regions 4, 7, 8 and 10, with d<1/250. These regions are located in the Atlantic Ocean and in the North Sea, and the majority of the SU sites in them have only one individual undulation (Figure 2.4.4a).

Series of undulations and type of coast

Certain sites only contained one individual SU, whereas series of undulations were found at numerous sites, some of them exhibiting up to 15 individual SUs (Figures 2.4.4a and 2.4.5). Sites with only one undulation were predominant (99, 34%), followed by sites with two undulations (81, 28%). The number of sites exhibiting more than two SUs decreases rapidly until it reaches 5–6 SU when the number stabilizes. A cubic function fits the data with a high correlation (Figure 2.4.5).

SUs with the greatest lengths are found at sites characterized by only one or two undulations, whereas an increasing number of undulations generally indicates a shorter length. Figure 2.4.6 depicts sites where the SUs have average lengths lower than 500 m. In general, sites with a higher number of undulations have lengths lower than 250 m and widths that are usually below 30 m.

We also analyzed the number of undulations as a function of the type of coast (Figure 2.4.7). This division was based on the importance of the curvature in the generation of shoreline features (Kaergaard et al., 2012; López-Ruiz et al., 2014). A similar number of observations were identified for the three coastal types. However, when this variable is analyzed in combination with the

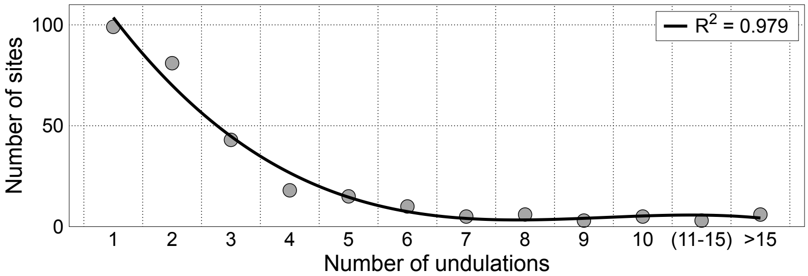

Figure 2.4.5: Relationship between the number of undulations and the number of sites. The line corresponds to the fitting of the data to a cubic function

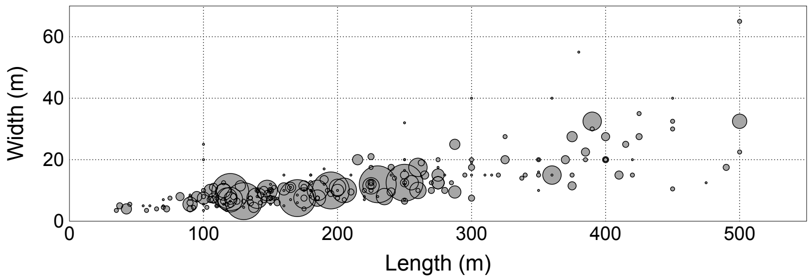

Figure 2.4.6: Sites where SU are characterised by lengths lower than 500 m. The diameter of the circles is proportional to the number of undulations at a given site

number of undulations (Figure 2.4.7), sites showing more than 6 undulations are not found at outlets or river mouths but are predominantly located along rectilinear coasts, and sites with only one undulation are predominantly found at river mouths (43%) followed by curvilinear (36%) and rectilinear (20%) coasts.

Constraints

In theory, a quasi-rectilinear coastline with a constant alongshore sediment transport rate should be in equilibrium, and no undulations would be expected (López-Ruiz et al., 2014). However, modifications of such an ideal situation produced by different mechanisms such as bi-modal wave distributions (Ashton et al., 2006) or by an artificial constraint at the coastline (López-Ruiz et al., 2012a) could lead to the formation of SUs (López-Ruiz et al., 2014).

Constraints located at a distance smaller than 5λ from the nearest boundary of the SU were considered. A total of 135 SUs (45.9%) could have been influenced by constraints, with coastal jetties the predominant obstacle (39, 29%). Rocky shores are also found along with a high number of SUs (34, 25%), followed by ports (29, 22%) and channeling at river mouths (24, 18%). When the distance to the obstacle is analyzed, 34% are located at a distance smaller than 1λ, with almost 60% at a distance smaller than 2λ. These results indicate the strong influence constraints have on the morphodynamic processes that generate SUs, as already shown by López-Ruiz et al. (2012a).

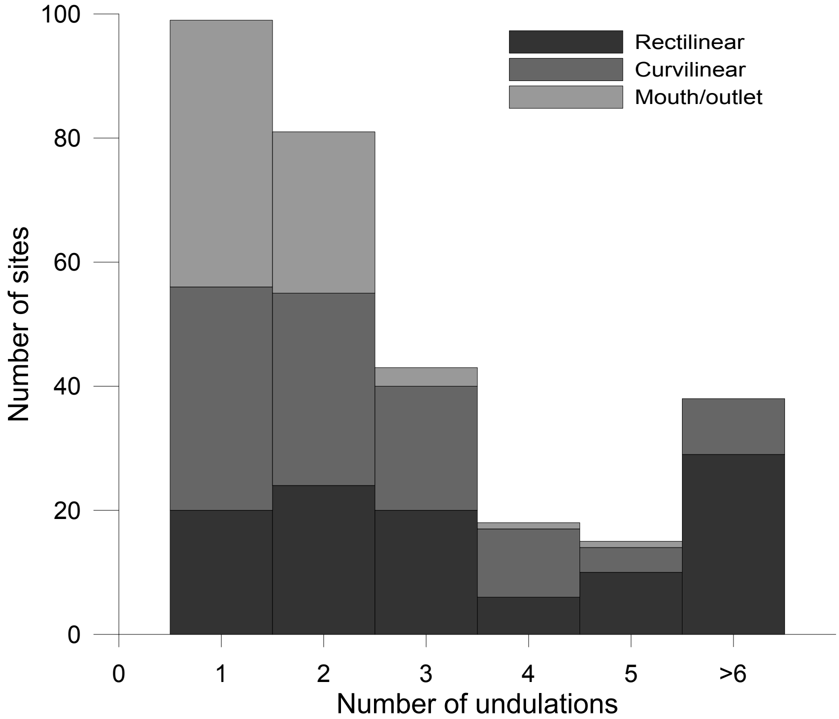

Figure 2.4.7: Number of sites as a function of the number of undulations and the type of coast

Geometric analysis

Relationship between the length and width of the SU. In this section, we investigate whether there is a predominant relationship between the length and width of the SU. Figure 2.4.9 plots the fitting of λ and W using a linear relation. The best linear correlation gives a relation of λ/W ~ 20, with a fitting skill of 0.984 (Wilmott 1981). When the fitting is performed for different wavelength intervals (Figure 2.4.10), lower correlations are obtained because of the increasing dispersal of the data (fitting skills are lower than 0.9, with a minimum of 0.428). Similar results are obtained if series with the same number of SUs are analyzed. Our value of λ/W is the same as that obtained by Thevenot and Kraus (1995) and is within the range of the value obtained by Bruun (1954). However, this result differs slightly from those of the other authors included in Table 1 who refer only to one specific location. Nevertheless, we consider our results more representative because a higher number of sites and samples were included.

Shape of the shoreline undulations: symmetry

. Similar to other typical shoreline morphological features that have been widely studied, such as beach cusps (Coco et al., 1999), SUs can be symmetric or asymmetric. If the difference between the lengths (widths) of the different individual SUs in a series of undulations was ≤20 m (≤5 m), the SU was considered symmetric with respect to its length (width). The same criteria were applied for the crest and trough of an individual SU.

When studying the symmetry between the different individual undulations that are part of a series of SUs, only sites that exhibited more than one undulation (195 sites) should be

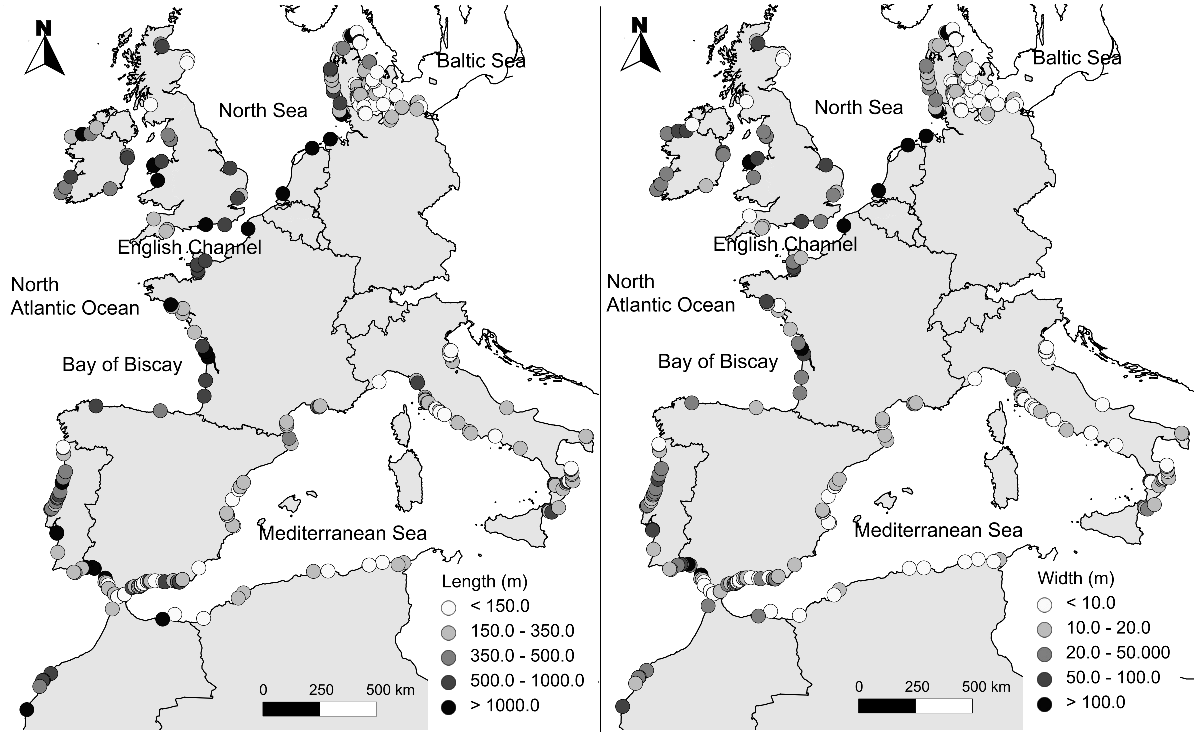

Figure 2.4.8: SU classified according to the range of lengths (a) and widths (b)

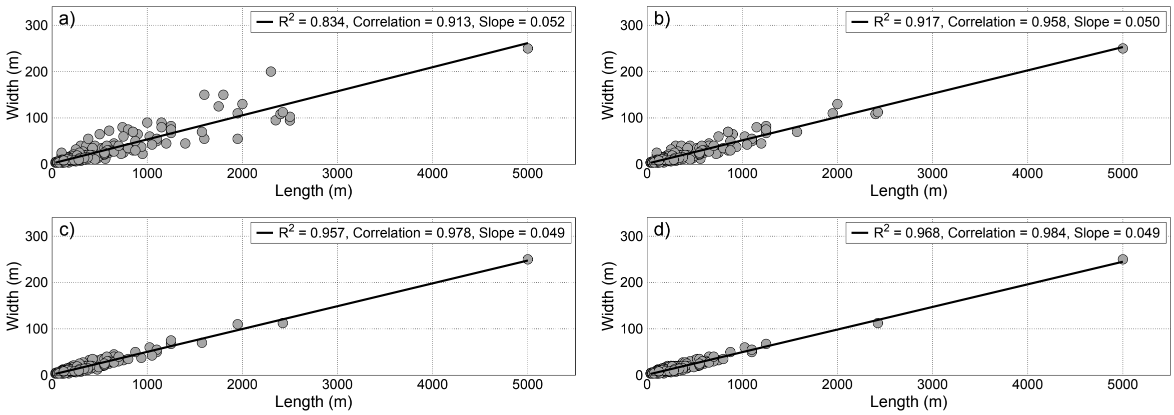

Figure 2.4.9: Fitting of average length and width: (a) all sites included, Skill (S) = 0.9047; (b) filtered $2sigma$, 276 sites, S = 0.9566; (c) filtered $2sigma$ from (b), 255 sites, S = 0.9779; (d) filtered $2sigma$ from (c), 240 sites, S = 0.9835. $sigma$ - standard deviation

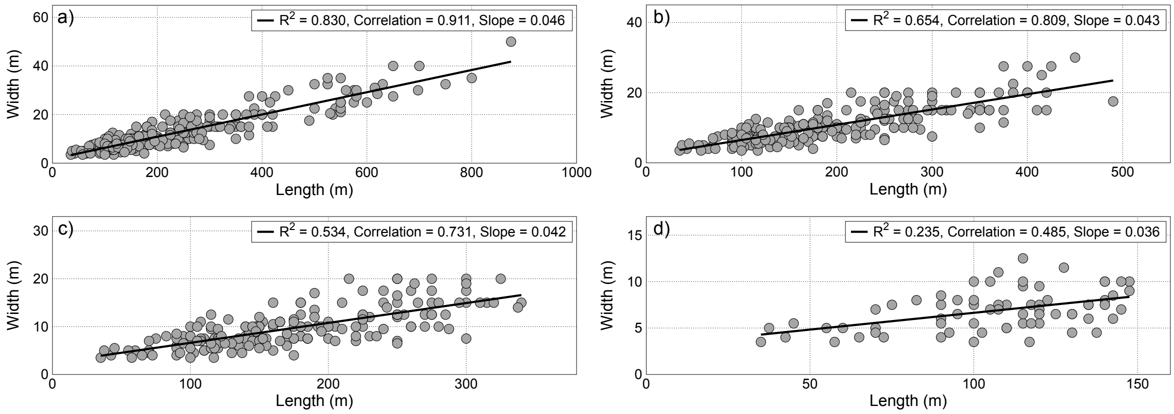

Figure 2.4.10: Fitting of average length and width for different intervals of λ: (a) λ<1000 m, 234 sites, S = 0.9074; (b) λ<500 m, 208 sites, S = 0.7906; (c) λ<350 m, 187 sites, S = 0.7107; (d) λ<150 m, 79 sites, S = 0.4277. Sites included in the analysis are derived from the fitting from Figure 11(d)

analyzed. Asymmetry with respect to length was found for 83% of sites, whereas 62% showed asymmetry with respect to width, and 61% showed asymmetry with respect to both length and width. When studying the symmetry with respect to the crest and trough, 64% and 73% of the sites showed asymmetry with respect to length and width, respectively. In sites where only one SU was present (99 sites), 54.5% and 73% were symmetric with respect to length and width, respectively; symmetry with respect to both length and width was found for 46.5% of the sites. Finally, in studying the symmetry with respect to crest and trough for the entire set of SUs (294 sites), 58% and 73% showed asymmetry with respect to length and width, respectively.

Spain: a country-scale example of the database potential

Spain has a complete dataset of maritime information managed by Puertos del Estado (Ministry of Public Works, Spain) that is available for research and management purposes (Table 3). Within the framework of different research projects, a complete recompilation of maritime climate along the Spanish coastline was gathered and has been published on the web at http://www.c3e.ihcantabria.com. Figure 2.4.11 depicts the main climate variables along the Spanish coastline in combination with the location and main characteristics of the SUs.

Sites exhibiting SUs are predominant along the Mediterranean and south Atlantic Spanish coasts, whereas they are very scarce along the northern coast (Figure 2.4.11). As previously described, the density decreases from d=1/23 at the southern coast of Spain to d=1/319 at the northern coast. The lengths and widths of SUs are similarly distributed along the Spanish coast, with exceptions for the longest SUs found along the south Atlantic coast (Figures 2.4.11(a) and 2.4.11(b)).

Our results show that the distribution of sites along the Spanish coast appears to be related to the values of the main maritime variables: significant wave height, peak period, predominant wave angle of incidence and tidal range (Figures 2.4.11, maps c to f, respectively). The average values of significant wave heights are higher for the northern coast, whereas they are significantly

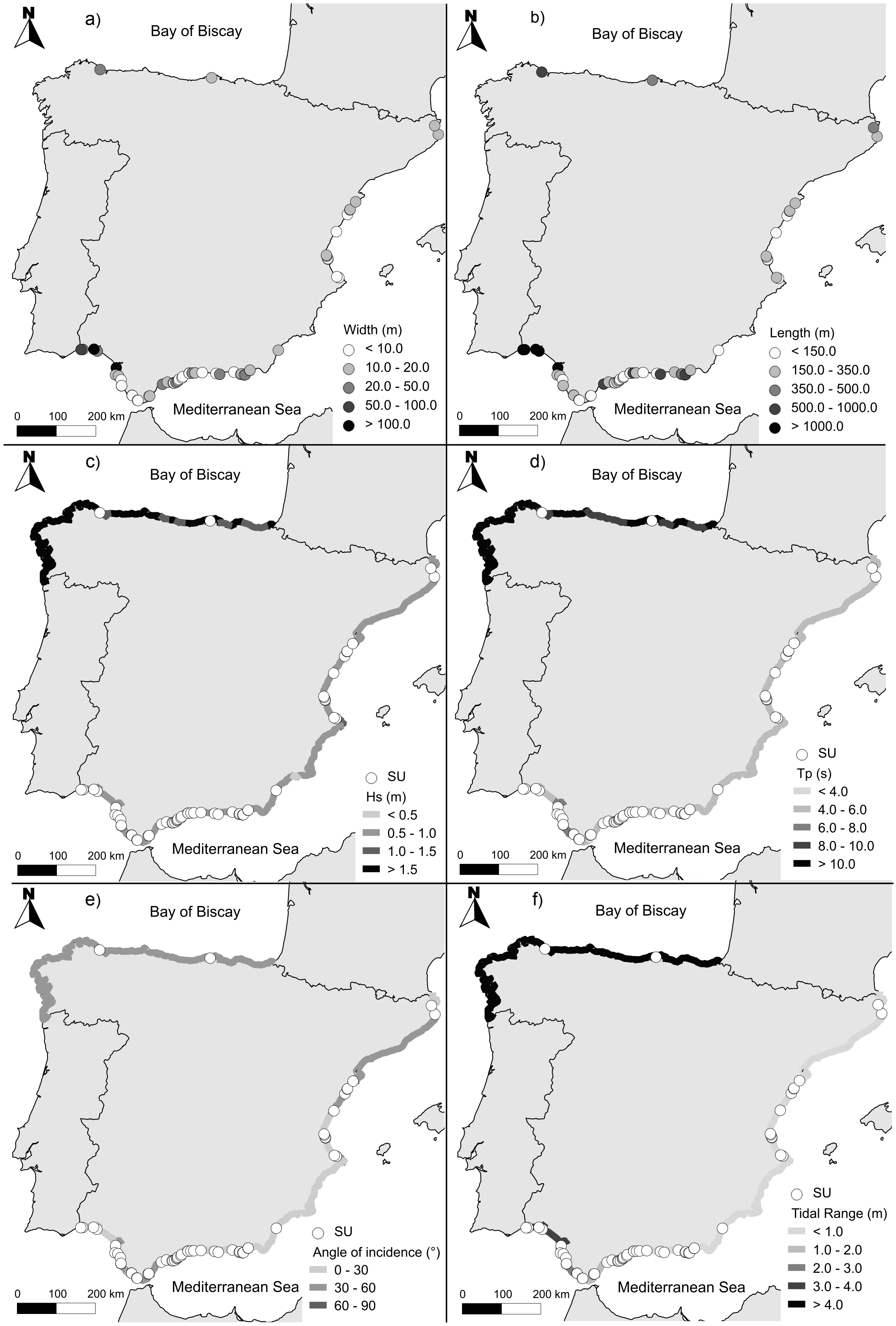

Figure 2.4.11: Maps showing different geometric (a, b) and maritime (c-f) information for the Spanish coast. Hs = significant wave height; Tp = wave peak period

reduced for the Mediterranean coast (Figure 2.4.11(c)). This behavior appears to be correlated with the number of identified sites, with a minimum (maximum) number of sites with higher (lower) wave energy. Waves with the largest peak periods arrive along the northern coast, whereas the shortest waves arrive along the south Mediterranean coast (Figure 2.4.11(d)). These characteristics are directly related to the number of sites with SUs, with more sites found where the peak period is shorter.

A similar trend is observed for the two remaining variables analyzed: the wave angle of incidence (Figures 2.4.11(e)) and the tidal range (Figure 2.4.11(f)). The wave angle of incidence was calculated as the difference between the shore normal and the wave fronts. Thus, the greater the angle, the more obliquely the waves arrive at the coast. For coastlines where the waves impinge more obliquely, fewer SUs are identified. Conversely, when waves approach more perpendicularly, the development of SUs is favored. Regarding the tidal range, more sites with SUs are found where the tidal range is lower. The previous results indicate that a relationship between the maritime wave climate and sites exhibiting SUs can be established. These results would be enriched by the incorporation of topographic and bathymetric data into the analysis; however, this is out of the scope of the present work.

Data mining analysis: relationships between the different variables

In this section, we present patterns obtained from our data by data mining techniques. Although data mining is usually applied to huge datasets involving thousands of items to achieve significant results (Witten and Frank, 2005), we have obtained consistent results in our particular case despite our dataset containing just 294 items. Enriching our data or using a larger data set should drastically improve the results. Consequently, this approach has great potential and could be easily expanded to other geographic applications with similar purposes. Specifically, association rules (Agrawal et al., 1993) are being used because of their summarization capability (Wong et al., 2005).

We have defined two case studies according to the available data. The first case study only takes into account the Spanish sample. Thus, local and focused regularities are the main objective in this country-scale study. The second case considers the whole of Europe and attempts to ascertain general patterns and widespread trends. Although this case involves more data, not every variable has a value available for every location.

Below, we present some of the results obtained from applying an Apriori algorithm to our cases studies. For each rule, its support, confidence and lift are shown. To run the algorithm, three parameters are specified, leaving the remaining parameters at default values. The first two arguments establish minimum thresholds for the support and confidence of the association rules. The third parameter represents the maximum number of items per item set.

In the Spanish case study, greater values for the minimum support and confidence could be established (Table 2.4.5). This case was executed with minimum support=0.136, minimum confidence=0.9 and maximum length=10. Because richer and more detailed information was available, associations with higher values of support and confidence were obtained. Figure 2.4.12 shows the rules that were obtained synthetically. The association rules obtained in the Europe case study are displayed in Table 2.4.6. The parameters for this case were minimum support=0.051, minimum confidence=0.8 and maximum length=10. Because of their size and heterogeneity, the

Id	Rule	Support	Confidence	Lift
1	If the country is Portugal then the tidal range is between 3 m and 3.5 m	0.05782	0.8095	7.212
2	If the width of the undulation is greater than 60 m then the length of the undulation is greater than 500 m	0.08844	0.9630	4.103
3	If the number of undulations is 1 and the width of the undulation is less than 20 m and the tidal range is less than 0.5 m then the country is Denmark	0.06122	0.8182	3.759
4	If the country is Italy and the length of the undulation is between 100 m and 200 m then the tidal range is between 0.5 m and 1 m	0.05782	0.8500	3.204
5	If the length of the undulation is between 100 m. and 200 m. then the width of the undulation is less than 20 m	0.31293	1	1.434

Table 2.4.5: Association rules obtained for the Europe case study

Id	Rule	Support	Confidence	Lift
1	If the length of the undulation is between 100 m and 200 m and the peak wave period is between 4 s and 6 s then the tidal range is less than 1 m	0.3214	1	1.556
2	If the number of undulations is greater than 6 then the tidal range is less than 1 m	0.1429	1	1.556
3	If the length of the undulation is between 100 m and 200 m then the width of the undulation is less than 20 m	0.4286	1	1.273
4	If the peak wave period of the undulation is between 4 s and 6 s then the significant wave height is between 0.5 m and 1 m	0.75	1	1.120
5	If the length of the undulation is between 100 m and 200 m then the significant wave height is between 0.5 m and 1 m	0.4107	0.9583	1.073

Table 2.4.6: Association rules obtained for the Spain case study

minimum support and confidence parameters had to be decreased to obtain association rules. Nevertheless, the results show an interesting relationship between hydrodynamic variables, number of undulations and location.

Discussion and conclusions

This work presents a database of sites exhibiting shoreline undulations along the coast of Western Europe and Northwestern Africa identified by visual interpretation of Google Earth images. Although public geographic databases already exist (i.e., Messerli et al., 2014), ours constitute the first developed database characterizing coastal features at a multi-country scale. Many uncertainties are inherent in the analysis, with the primary uncertainty being the temporal scale and spatial extension of the problem. Regarding the spatial extension, the analysis of 50,000 km of coast is a highly time-consuming and error-prone task; thus, certain sites could have been disregarded unintentionally. For the temporal scale, the only analyzed images correspond to

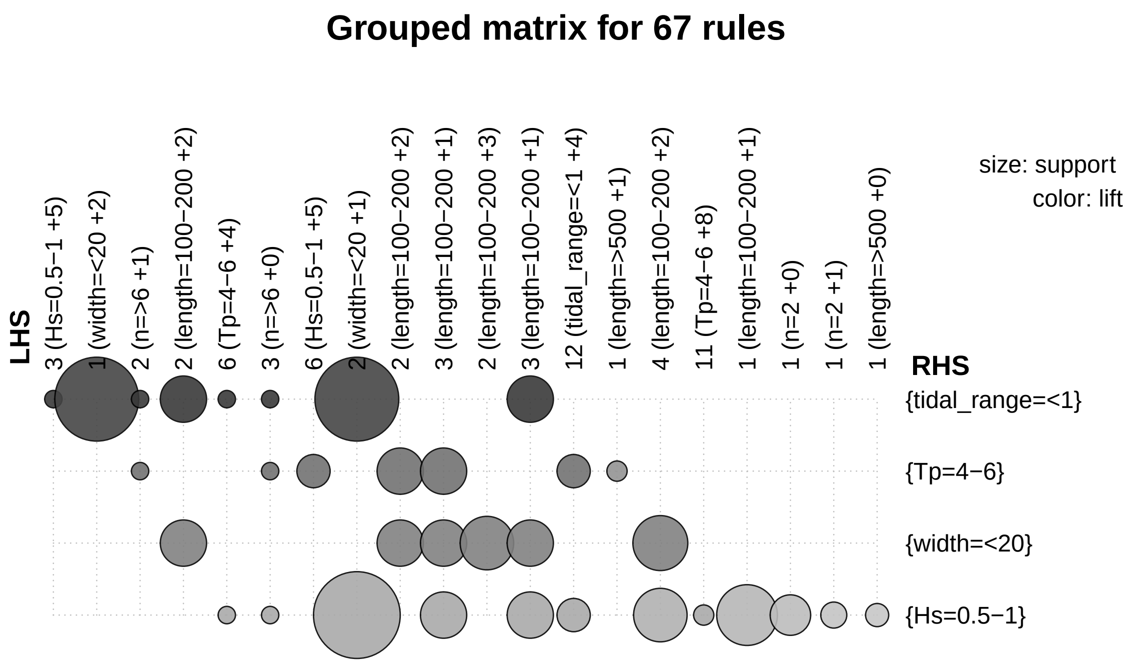

Figure 2.4.12: Grouped matrix-based visualisation of rules for the Spanish case study. Grouped matrix uses a balloon plot with antecedent groups as columns and consequents as rows. The number of rules and the most important items in the group are displayed as the labels for the columns followed by the number of other items in that antecedent group. Columns and rows in the plot are reordered such that the aggregated interest measure is decreasing from the top down and from left to right. Groups with the most interesting rules according to lift are shown in the top-left corner of the plot. Support of the rules is represented with the size of the balloons

the present configuration of the coast. If this work could have been performed in the past, the results would have been different because of natural and/or human interventions. However, our work has shown the potential for this type of analysis, and it provides the scientific community with access to an open, free database of shoreline undulations that can be smoothly integrated in the scientific workflow. We encourage other researchers to perform similar analyses and improve and/or create new geomorphic large extension databases with public access, as already performed in works such as Borrelli et al. (2014).

Geometry and shape of the shoreline undulations

We identified the country and location with the highest (Denmark and Spain) and lowest (Northern Ireland) frequency of SUs. Our results indicate that a dominant order of magnitude of the relationship between wavelength and amplitude for SU is ~20. This order of magnitude is similar to that of additional smaller shoreline features, such as beach cusps (Coco et al., 1999; Holland, 1998). We found that tides discourage the formation of SUs but do not similarly affect smaller-scale features. One possible explanation is that alongshore processes on the coast dominate how SUs are formed, with cross-shore processes being of minor importance (López-Ruiz et al., 2012b). When the tidal range is important, the breaking zone moves cross-shore, changing the wave dissipation and sediment transport patterns (López-Ruiz et al., 2014).

While the reason for SU development is out of the scope of our work, recent works indicate that one of the factors that trigger their formation is the presence of artificial constraints (obstacles) along the coast. López-Ruiz et al. (2012a) applied a one-line-type model to ascertain the importance of a channelized river mouth in the development of SU at the adjacent beach. The results indicate that the barrier-effect on the alongshore sediment transport from the channelization plays a key role in SU formation and growth. To analyze the statistical importance of this phenomenon, the existence of these constraints was also identified in the present work. Applying basic statistical analysis to our data, the results show that approximately 56% of the identified SUs could be influenced by constraints because 60% were located at a distance smaller than 2λ from a constraint, which reinforces the conclusions of López-Ruiz et al. (2012a) concerning the importance of these structures in SU development.

Our data show that SUs can be either symmetrical or asymmetrical with respect to width and length, although asymmetrical shoreline undulations are more likely to occur, with more than 50% of SUs presenting some type of asymmetry. Some authors have proposed mechanisms for the formation of SUs that generate symmetrical features (Ashton and Murray, 2006) but are not capable of reproducing asymmetrical features (Ortega-Sánchez et al., 2008). Similarly, Ortega-Sánchez et al. (2014) presented a shoreline evolution model that accounts for alongshore wave energy variations induced by the complex offshore bathymetry that generates asymmetrical features for Carchuna beach (Southern Spain) that are similar to the features identified in our work.

Influence of the tides and waves

Because of the length scale of the area of study, collecting and analyzing wave and climate information and finding correlations with the shoreline undulations is not an easy task. For some areas, information is scarce, and for other areas, information collection is difficult. Our results indicate that undulations develop easily under low wave heights (protected coasts), which could explain the high number of undulations along the coast of Denmark or the southern Mediterranean coast of Spain. However, the presence of low wave height is insufficient for SU development, because for sites along the English Channel or the Adriatic Sea, the wave energy is generally low and SUs appear at very low frequencies. In contrast, the greater the fetch, the lower the number of undulations, which could be the case for Spain and Portugal, where a number of SUs follow this trend (Figure 2.4.11); however, this is not verified at the Adriatic site. While wave energy plays a significant role in the development of SUs, other factors are also important, mainly the character of the wave climate and the type of waves (Ashton et al., 2006). Thus, further analysis is recommended to support our results.

Tidal range is also relevant in the development of SUs. Figure 2.4.13 depicts the location of the sites as a function of the tidal range. In general terms, SUs are predominant in areas where the tidal range is lower (Mediterranean Sea and along the Denmark coast); conversely, SUs are infrequent along the northern Spanish coast, the English Channel and in some tracks of

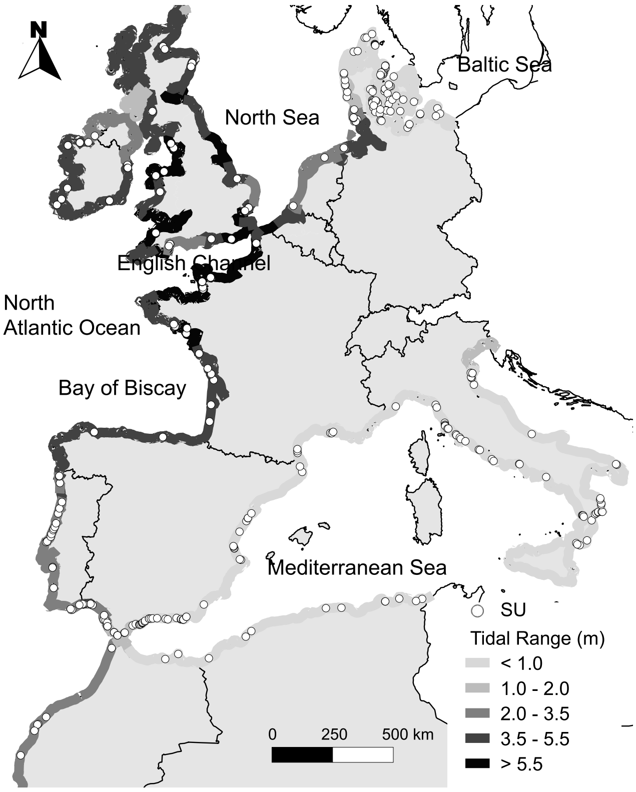

Figure 2.4.13: Distribution of the shoreline undulations superimposed over the tidal range

the French and English coast where the tidal range is higher. SUs are not found at Bretagne, where tidal ranges surpass 7 m, and the same pattern is observed along the English Channel, where tidal ranges easily exceed 5 m. This hypothesis is generally verified except for sites along Portugal's coast, where other factors most likely play a significant role in the distribution of the features. The results show that the highest number of undulations is found at sites with tidal ranges lower than 1 m and that the number of SUs decreases as the tidal range increases. The results are significant, considering that the lengths of coasts with tidal ranges lower than 1 m and between 3.5 and 5.5 m are ~15,000 km and ~18,000 km, respectively.

The influence of tidal range in the dimensions of the undulations is also clearly observed when analyzing their wavelengths (Figures 2.4.8 and 2.4.13). In the Mediterranean Sea, only two undulations are over 1000 m (1.9%), whereas this ratio increases to 14.5 and 20% for the Atlantic Ocean and North Sea, respectively. The highest proportion of long undulations is found along the English Channel, where a total of 36% are over 1000 m and the tidal ranges vary more intensively. For the Baltic Sea and the Straits of Skagerrak and Kattegat, the undulations are predominantly short. Nevertheless, for more conclusive results, wave climate should also be considered in this analysis.

The influence of tides on coastal rhythmic features such as beach cusps has been studied (Coco et al., 2004), and it was concluded that the height and cross-shore extent of beach cusps are tidally modulated. However, this refers to the short-term scale and small spatial-scale features, which are outside the range of SUs. During recent years, different works have established the relationship between the formation and characteristics of large-scale shoreline features such as SUs and wave climate characteristics (Ashton et al., 2001; Ortega-Sánchez et al., 2003; López-Ruiz et al., 2012b). However, the influence of tides in their formation has not been explored, most likely because short-term variables were not significant for the generation of medium- to long-term shoreline features. Our results show that this variable should be included in the analysis.

Final remarks

Forming materials, characteristics of the topography-bathymetry and tidal range may contribute to a high percentage of the sites exhibiting SUs. Different fractions of sand and gravel coexist at Mediterranean Spanish beaches (Bramato et al., 2012). It has been intensively reported in the bibliography that beach cusps are best formed when two different grain sizes are present at the beach (Nolan et al., 1999). In a similar manner, it seems that the redistribution of sediment of different sizes alongshore would facilitate the formation and development of undulations. However, the low number of identified sites where the coast is dominated by cliff or embayed beaches (i.e., coasts of England or the north of Spain) does not favor the genesis of shoreline undulations. Our results would be significantly enriched by the incorporation of a detailed bathymetry-topography of the study area into the analysis.

This work also explored the potential of using data mining techniques for the development of coastal studies occupying large extensions. Our compilation of information about shoreline undulations was restricted to the coasts of Western Europe, which represent a small portion of the total length of the world's coasts. We have also collected information on a selected number of characteristic variables (Table 2). Although the number of sites (294) and the total number of individual undulations (901) identified was not huge, the total amount of information (17 fields for every undulation) was sufficient to start applying data mining to analyze the data. This type of analysis will become more significant if more observations in time are available and if the number of sites or variables increases.

Association rules provide new data analysis instruments to study information from new perspectives, and the results for both the Western European and Spanish cases are significant. The discovered relationships between variables can be more or less complex depending on the amount of data and the combination of variables. Certain rules obtained reflect information inherited from data directly and allow us to verify that the method was working properly (i.e., Rule 1, Tables 5 and 6), whereas the remainder quantify different relationships between the variables that were previously unknown (Tables 5 and 6).

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in:. Presented at the Proceedings of the ACM SIGMOD International Conference on Management of Data, ACM, pp. 207–216. doi:10.1145/170036.170072

Agrawal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules in Large Databases, in:. Presented at the Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 487–499.

Ashton, A.D., Brad Murray, A., 2006. High-angle wave instability and emergent shoreline shapes: 1. Modeling of sand waves, flying spits, and capes. Journal of Geophysical Research 111. doi:10.1029/2005JF000422

Ashton, A.D., Brad Murray, A., Arnault, O., 2001. Formation of coastline features by large-scale instabilities induced by high-angle waves. Nature 414, 296–300.

Beels, C., Henriques, J., De Rouck, J., Pontes, M., De Backer, G., Verhaeghe, H., 2007. Wave energy resource in the North Sea, in:. Presented at the Proceedings of the Seventh European Wave and Tidal energy Conference EWTEC.

Benker, S.C., Langford, R.P., Pavlis, T.L., 2011. Positional accuracy of the Google Earth terrain model derived from stratigraphic unconformities in the Big Bend region, Texas, USA. Geocarto International 26, 291–303. doi:10.1080/10106049.2011.568125

Bhambri, R., Bolch, T., 2009. Glacier mapping: a review with special reference to the Indian Himalayas. Progress in Physical Geography 33, 672–704. doi:10.1177/0309133309348112

Bird, E., 2010. Encyclopedia of the World's Coastal Landforms. Springer.

Bird, E., 2011. Coastal Geomorphology: An Introduction. Wiley.

Borgelt, C., Kruse, R., 2002. Induction of Association Rules: Apriori Implementation, in:. Presented at the Proceedings of the 15th Conference on Computational Statistics (COMPSTAT), Physica Verlag, pp. 395–400. doi:10.1007/978–3–642–57489–4_59

Borrelli, P., Modugno, S., Panagos, P., Marchetti, M., Schütt, B., Montanarella, L., 2014. Detection of harvested forest areas in Italy using Landsat imagery. Applied Geography 48, 102–111. doi:10.1016/j.apgeog.2014.01.005

Bramato, S., Ortega-Sánchez, M., Mans, C., Losada, M.A., 2012. Natural recovery of a mixed sand and gravel beach after a sequence of a short duration storm and moderate sea states. Journal of Coastal Research 28, 89–101. doi:10.2112/JCOASTRES-D–10–00019.1

Brown, G., 2012. An empirical evaluation of the spatial accuracy of public participation GIS (PPGIS) data. Applied Geography 34, 289–294. doi:10.1016/j.apgeog.2011.12.004

Bruun, P., 1954. Migrating Sand Waves or Sand Humps, with Special Reference to Investigations carried out on the Danish North Sea Coast, in:. Presented at the Proceedings of 5th Conference on Coastal Engineering, pp. 269–295.

Butler, B.J., Hewes, J.H., Liknes, G.C., Nelson, M.D., Snyder, S.A., 2014. A comparison of techniques for generating forest ownership spatial products. Applied Geography 46, 21–34. doi:10.1016/j.apgeog.2013.09.020

Butler, D., 2006. Virtual globes: The web-wide world. Nature 439, 776–778. doi:10.1038/439776a

Cambio Climático en la Costa Española Visor C3E, 2014. Acción Estratégica De Energía y Cambio Climático (in Spanish). Available from: http://www.c3e.ihcantabria.es/.

Carretero, S., Braga, F., Kruse, E., Tosi, L., 2014. Temporal analysis of the changes in the sand-dune barrier in the Buenos Aires Province, Argentina, and their relationship with the water resources. Applied Geography 54, 169–181. doi:10.1016/j.apgeog.2014.08.004

Carter, R.W.G., Woodroffe, C.D., 1997. Coastal Evolution: Late Quaternary Shoreline Morphodynamics. Cambridge University Press.

Coco, G., Brad Murray, A., 2007. Patterns in the sand: From forcing templates to self-organization. Geomorphology 91, 271–290. doi:10.1016/j.geomorph.2007.04.023

Coco, G., Burnet, T.K., Werner, B.T., Elgar, S., 2004. The role of tides in beach cusp development. Journal of Geophysical Research C: Oceans 109, C04011 1–12. doi:10.1029/2003JC002154

Coco, G., O'Hare, T., Huntley, D.A., 1999. Beach cusps: A comparison of data and theories for their formation. Journal of Coastal Research 15, 741–749.

Davidson-Arnott, R.G.D., Van Heyningen, A.G., 2003. Migration and sedimentology of longshore sandwaves, Long Point, Lake Erie, Canada. Sedimentology 50, 1123–1137. doi:10.1046/j.1365–3091.2003.00597.x

Davis, R., Fitzgerald, D., 2009. Beaches and Coasts. Wiley-Blackwell.

Dean, R.G., Dalrymple, R.A., 2002. Coastal Processes with Engineering Applications. Cambridge University Press.

Delmelle, E.M., Zhu, H., Tang, W., Casas, I., 2014. A web-based geospatial toolkit for the monitoring of dengue fever. Applied Geography 52, 144–152. doi:10.1016/j.apgeog.2014.05.007

Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., Montanarella, L., 2012. Satellite remote sensing for soil mapping in Africa: An overview. Progress in Physical Geography 36, 514–538. doi:10.1177/0309133312446981

Emel, J., Plisinski, J., Rogan, J., 2014. Monitoring geomorphic and hydrologic change at mine sites using satellite imagery: The Geita Gold Mine in Tanzania. Applied Geography 1–7. doi:10.1016/j.apgeog.2014.07.009

Falqués, A., 2006. Wave driven alongshore sediment transport and stability of the Dutch coastline, in:. Presented at the Coastal Engineering, pp. 243–254. doi:10.1016/j.coastaleng.2005.10.012

Gajewski, K., 2008. The Global Pollen Database in biogeographical and palaeoclimatic studies. Progress in Physical Geography 32, 379–402. doi:10.1177/0309133308096029

Geng, L., Hamilton, H.J., 2006. Interestingness measures for data mining: A survey. ACM Computing Surveys (CSUR) 38, 9–es. doi:10.1145/1132960.1132963

Godet, L., Fournier, J., Toupoint, N., Olivier, F., 2009. Mapping and monitoring intertidal benthic habitats: a review of techniques and a proposal for a new visual methodology for the European coasts. Progress in Physical Geography 33, 378–402. doi:10.1177/0309133309342650

Goff, J., Chague-Goff, C., 2014. The Australian Tsunami Database: A review. Progress in Physical Geography 38, 218–240. doi:10.1177/0309133314522282

Hahsler, M., Chelluboina, S., 2011. Visualizing Association Rules in Hierarchical Groups, in:. Presented at the Computing Science and Statistics, Vol. 42, 42nd Symposium on the Interface: Statistical, Machine Learning, and Visualization Algorithms (Interface 2011), The Interface Foundation of North America.

Hahsler, M., Grün, B., Hornik, K., 2005. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. Journal of Statistical Software 14, 1–25.

Holland, K., 1998. Beach cusp formation and spacings at Duck, USA. Continental Shelf Research 18, 1081–1098.

ISPRA - Istituto Superiore per la Protezione e la Ricerca Ambientale, 2014. APAT - Servizio Mareografico. Available from: http://www.idromare.it/.

Jäppinen, S., Toivonen, T., Salonen, M., 2013. Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. Applied Geography 43, 13–24. doi:10.1016/j.apgeog.2013.05.010

Kaergaard, K., Fredsoe, J., 2013. Numerical modeling of shoreline undulations part 2: Varying wave climate and comparison with observations. Coastal Engineering 75, 77–90. doi:10.1016/j.coastaleng.2012.11.003

Kaergaard, K., Fredsoe, J., Knudsen, S.B., 2012. Coastline undulations on the West Coast of Denmark: Offshore extent, relation to breaker bars and transported sediment volume. Coastal Engineering 60, 109–122. doi:10.1016/j.coastaleng.2011.09.002

Kuenzer, C., van Beijma, S., Gessner, U., Dech, S., 2014. Land surface dynamics and environmental challenges of the Niger Delta, Africa: Remote sensing-based analyses spanning three decades (1986–2013). Applied Geography 53, 354–368. doi:10.1016/j.apgeog.2014.07.002

Levoy, F., Anthony, E.J., Monfort, O., Larsonneur, C., 2000. The morphodynamics of megatidal beaches in Normandy, France. Marine Geology 171, 39–59. doi:10.1016/S0025–3227(00)00110–9

López-Ruiz, A., Ortega-Sánchez, M., Baquerizo, A., Losada, M.A., 2012a. Short and medium-term evolution of shoreline undulations on curvilinear coasts. Geomorphology 159, 189–200.

López-Ruiz, A., Ortega-Sánchez, M., Baquerizo, A., Losada, M.A., 2014. A note on alongshore sediment transport on weakly curvilinear coasts and its implications. The CoastView Project: Developing coastal video monitoring systems in support of coastal zone management 88, 143–153. doi:10.1016/j.coastaleng.2014.03.001

López-Ruiz, A., Ortega-Sánchez, M., Baquerizo, A., Navidad, D., Losada, M.Á., 2012b. Nonuniform alongshore sediment transport induced by coastline curvature. Int. Conf. Coastal. Eng. 1. doi:10.9753/icce.v33.sediment.29

MacDonald, G.M., Bennett, K.D., Jackson, S.T., Parducci, L., Smith, F.A., Smol, J.P., Willis, K.J., 2008. Impacts of climate change on species, populations and communities: palaeobiogeographical insights and frontiers. Progress in Physical Geography 32, 139–172. doi:10.1177/0309133308094081

Maetens, W., Vanmaercke, M., Poesen, J., Jankauskas, B., Jankauskiene, G., Ionita, I., 2012. Effects of land use on annual runoff and soil loss in Europe and the Mediterranean: A meta-analysis of plot data. Progress in Physical Geography 36, 599–653. doi:10.1177/0309133312451303

Martínez, M.L., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., Landgrave, R., 2007. The coasts of our world: Ecological, economic and social importance. Ecological Economics 63, 254–272. doi:10.1016/j.ecolecon.2006.10.022

Masselink, G., Short, A.D., 1993. The Effect of Tide Range on Beach Morphodynamics and Morphology - a Conceptual Beach Model. Journal of Coastal Research 9, 785–800.

McCool, J.-P.P., 2014. PRAGIS: a test case for a web-based archaeological GIS. Journal of Archaeological Science 41, 133–139. doi:10.1016/j.jas.2013.07.037

Medellín, G., Medina, R., Falqués, A., González, M., 2008. Coastline sand waves on a low-energy beach at "El Puntal" spit, Spain. Marine Geology 250, 143–156. doi:10.1016/j.margeo.2007.11.011

Messerli, P., Giger, M., Dwyer, M.B., Breu, T., Eckert, S., 2014. The geography of large-scale land acquisitions: Analysing socio-ecological patterns of target contexts in the global South. Applied Geography 53, 449–459. doi:10.1016/j.apgeog.2014.07.005

Nolan, T.J., Kirk, R.M., Shulmeister, J., 1999. Beach cusp morphology on sand and mixed sand and gravel beaches, South Island, New Zealand. Marine Geology 157, 185–198.

Ortega-Sánchez, M., Baquerizo, A., Losada, M.A., 2003. On the development of large-scale cuspate features on a semi-reflective beach: Carchuna beach, Southern Spain. Marine Geology 198, 209–223.

Ortega-Sánchez, M., Fachin, S., Sancho, F., Losada, M.A., 2008a. Relation between beachface morphology and wave climate at Trafalgar beach (Cádiz, Spain). Geomorphology 99, 171–185. doi:10.1016/j.geomorph.2007.10.013

Ortega-Sánchez, M., Lobo, F.J., López-Ruiz, A., Losada, M.A., Fernandez-Salas, L.M., 2014. The influence of shelf-indenting canyons and infralittoral prograding wedges on coastal morphology: The Carchuna system in Southern Spain. Marine Geology 347, 107–122.

Ortega-Sánchez, M., López-Ruiz, A., Baquerizo, A., Losada, M.A., 2015. Shoreline undulations. Encyclopedia of Estuaries (Ed. Michael J. Kennish). Encyclopedia of Earth Sciences Series, Springer. DOI 10.1007/978–94–017–8801–4.

Ortega-Sánchez, M., Quevedo, E., Baquerizo, A., Losada, M.A., 2008b. Comment on "High-angle wave instability and emergent shoreline shapes: 1. Modeling of sand waves, flying spits, and capes" by Andrew D. Ashton and A. Brad Murray. Journal of Geophysical Research-Earth Surface 113. doi:10.1029/2007JF000860

Pethick, J., 1989. An Introduction to Coastal Geomorphology. Edward Arnold.

Portillo-Quintero, C.A., Sanchez, A.M., Valbuena, C.A., Gonzalez, Y.Y., Larreal, J.T., 2012. Forest cover and deforestation patterns in the Northern Andes (Lake Maracaibo Basin): A synoptic assessment using MODIS and Landsat imagery. Applied Geography 35, 152–163. doi:10.1016/j.apgeog.2012.06.015

Potere, D., 2008. Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive. Sensors 8, 7973–7981. doi:10.3390/s8127973

Puertos del Estado, 2014. Ministerio de Fomento. Available from: http://www.puertos.es/.

QGIS Development Team, 2009. QGIS Geographic Information System.

Rajak, A., Gupta, M.K., 2008. Association rule mining-applications in various areas, in:. Presented at the Proceedings of International Conference on Data Management, Ghaziabad, India, pp. 3–7.

Rozenstein, O., Karnieli, A., 2011. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. Applied Geography 31, 533–544. doi:10.1016/j.apgeog.2010.11.006

Schwartz, M. (Ed.), 2006. Encyclopedia of Coastal Science. Springer.

Scott, T., Masselink, G., Russell, P., 2011. Morphodynamic characteristics and classification of beaches in England and Wales. Marine Geology 286, 1–20. doi:10.1016/j.margeo.2011.04.004

Song, C., 2013. Optical remote sensing of forest leaf area index and biomass. Progress in Physical Geography 37, 98–113. doi:10.1177/0309133312471367

Soomere, T., Viška, M., 2014. Simulated wave-driven sediment transport along the eastern coast of the Baltic Sea. Journal of Marine Systems 129, 96–105. doi:10.1016/j.jmarsys.2013.02.001

Stewart, C.J., Davidson-Arnott, R.G.D., 1988. Morphology, Formation and Migration of Longshore Sandwaves - Long Point, Lake Erie, Canada. Marine Geology 81, 63–77.

Thenkabailc, P.S., Biradar, C.M., Noojipady, P., Cai, X., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Pandey, S., 2007. Sub-pixel Area Calculation Methods for Estimating Irrigated Areas. Sensors 7, 2519–2538. doi:10.3390/s7112519

Thevenot, M.M., Kraus, N.C., 1995. Longshore Sand Waves at Southampton Beach, New-York - Observation and Numerical-Simulation of Their Movement. Marine Geology 126, 249–269. doi:10.1016/0025–3227(95)00081–9

Verhagen, H., 1989. Sand Waves Along the Dutch Coast. Coastal Engineering 13, 129–147.

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D., Tang, G., 2014. A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration. Environmental Modelling & Software 58, 86–94. doi:10.1016/j.envsoft.2014.04.007

Waters, R., Engstroem, J., Isberg, J., Leijon, M., 2009. Wave climate off the Swedish west coast. Renewable Energy 34, 1600–1606. doi:10.1016/j.renene.2008.11.016

Widener, M.J., Li, W., 2014. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. Applied Geography 54, 189–197. doi:10.1016/j.apgeog.2014.07.017

Willmott, C.J., 1981. On the validation of models. Physical Geography 2, 184–194.

Windfinder.com, 2014. Wind & weather statistics. Available from: http://www.windfinder.com/.

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Wong, K.W., Zhou, S., Yang, Q., Yeung, J.M., 2005. Mining Customer Value: From Association Rules to Direct Marketing. Data Min. Knowl. Discov. 11, 57–79. doi:10.1007/s10618–005–1355-x

Woodroffe, C.D., 2002. Coasts: Form, Process and Evolution. Cambridge University Press.

Yu, G., Yu, Q., Hu, L., Zhang, S., Fu, T., Zhou, X., He, X., Liu, Y., Wang, S., Jia, H., 2013. Ecosystem health assessment based on analysis of a land use database. Applied Geography 44, 154–164. doi:10.1016/j.apgeog.2013.07.010

Yu, L., Gong, P., 2012. Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives. International Journal of Remote Sensing 33, 3966–3986. doi:10.1080/01431161.2011.636081

Zhang, L., Yi, J., 2010. Management methods of spatial data based on PostGIS, in:. Presented at the 2010 Second Pacific-Asia Conference on Circuits,Communications and System (PACCS), IEEE, pp. 410–413. doi:10.1109/PACCS.2010.5626962

CHAPTER 3

CONCLUSIONS AND FUTURE WORK

The main contribution of this Thesis is to provide a framework to environmental researchers who want to apply artificial intelligence techniques to their works. Through the document, several applications has been shown using different artificial techniques.

EcoLexicon targets different user groups, such as translators, technical writers, environmental experts, etc., who wish to expand their knowledge of the environment for the purpose of text comprehension or generation. In the case of EcoLexicon, environmental dynamic knowledge has been represented using ontological model. These users can freely access EcoLexicon and are able to find the information needed, thanks to a user-friendly visual interface with different modules for conceptual, linguistic, and graphical data. Also, a first attempt to integrate EcoLexicon in Linked Data has been conducted. Even though Linked Data is conceived as a simple and efficient way for the integration of heterogeneous information, it provides an invaluable resource, specially within the environmental research field.

We have developed a general methodology to create databases of civil engineering systematic activities. It consists on the achievement of six sequential phases to gather and organize the information into a database. The developed database can be implemented in PostgreSQL, using PostGIS extension for spatial data, and therefore GIS integration, which is a valuable tool for the civil engineering discipline. The methodology was applied to the case of the dredging interventions in the 20 ports that are managed by the Andalusian Regional Government (Spain). The database contains 87 fields of information collected after the analysis of the 70 interventions performed since 1994. This database is free, public and available for the scientific community. Some analyses of the database information were done to highlight the potential of the database for this type of studies. Because information is well-structured and organized in a database, it is possible to query about data using a query language, such as SQL. It allows obtaining a sub group of the information that fits certain criteria. Indeed, data can be gathered and expressed in a summary form for analysis purposes. Thus, the processed information can be used to build trends in data or discover patterns within data.

The system that we have created adds a semantic component to remote-sensing images using ontologies to describe this information. Savia is an operational system that is available for any user via the web (http://obsnev.es/ontologia/index). The system implements a query builder user interface that allows users to build questions using SPARQL. It also includes a set of predefined questions to show its capabilities. Furthermore, users can select different output file formats to display results (csv, text or map). The ontology has also helped to unveil the co-occurrence of significant trends. The results (both ecological and methodological) demonstrate that the information in the MODIS time series is useful to assess the functioning of a terrestrial Natura 2000 habitat.

The latest application presents a database of sites exhibiting shoreline undulations along the coast of Western Europe and Northwestern Africa identified by visual interpretation of Google Earth images. This work also explored the potential of using data mining techniques for the development of coastal studies occupying large extensions. Our compilation of information about shoreline undulations was restricted to the coasts of Western Europe, which represent a small portion of the total length of the world's coasts. This type of analysis will become more significant if more observations in time are available and if the number of sites or variables increases. Association rules provide new data analysis instruments to study information from new perspectives, and the results for both the Western European and Spanish cases are significant. The discovered relationships between variables can be more or less complex depending on the amount of data and the combination of variables. Certain rules obtained reflect information inherited from data directly and allow us to verify that the method was working properly, whereas the remainder quantify different relationships between the variables that were previously unknown.

Conclusiones y trabajo futuro

La mayor aportación de esta tesis es la de proporcionar un marco de trabajo a los investigadores ambientales que deseen aplicar técnicas de inteligencia artificial a sus trabajos. A lo largo de la tesis se han mostrado varias aplicaciones donde se han utilizado diferentes técnicas de inteligencia artificial.

EcoLexicon tiene como objetivo diferentes grupos de usuarios, como traductores, escritores técnicos, expertos en medioambiente, etc., que deseen expandir su conocimiento del medioambiente para la compresión o la generación de conocimiento. En el caso de EcoLexicon, se ha representado el conocimiento mediante un modelo ontológico. Estos usuarios pueden acceder libremente y encontrar la información que necesiten, gracias a una interfaz de usuario amigable, con diferentes módulos para información conceptual, lingüística o gráfica. Además, se ha intentado hacer una primera aproximación para integrarlo dentro del Linked Data. Esta nueva técnica, a pesar de su sencillez, proporciona un modo muy eficaz para la integración información heterogénea. Algo de especial relevancia en el entorno de la dinámica ambiental.

Se ha desarrollado una metodología para desarrollar bases de datos en ingeniería civil. Consiste en la puesta en práctica de seis fases y la organización de su información en una base de datos. La base de datos utilizada ha sido PostgreSQL, utilizando la extensión PostGIS para almacenar datos espaciales, y por tanto facilitar la integración con los Sistemas de Información Geográfica (SIG), fundamentales en la disciplina de la ingeniería civil. La metodología ha sido aplicada a las intervenciones de dragado de los 20 puertos gestionados por el gobierno regional andaluz (España). La base de datos contiene 87 campos de información recopilados a lo largo de 70 intervenciones realizadas desde 1994. Esta base de datos es libre, pública y está disponible para toda la comunidad científica. Dado que la información está bien estructurada en una base de datos espacial, es posible utilizar lenguajes de consultas como SQL para obtener información. Así es posible filtrar información utilizando algún criterio. También es posible resumir la información para poder analizarla. Este mismo proceso puede ser utilizado para descubrir tendencias en los datos.

El sistema creado añade un componente semántico a las imágenes obtenidas mediante teledetección usando ontologías para describir dicha información. Savia es un sistema que se encuentra disponible para cualquier usuario via web (http://obsnev.es/ontologia/index). El sistema implementa un interfaz de usuario para la realización de consultas utilizando el lenguaje SPARQL. Para demostrar su funcionamiento, incluye un conjunto de preguntas predefinidas. Además, los usuarios pueden seleccionar el formato de salir para mostrar los resultados (csv, texto o visualmente en un mapa). La ontología ha contribuido además al descubrimiento de tendencias significativas en los datos. Los resultados, tanto ecológicos como metodológicos, han demostrado que la información de las series de MODIS son útiles para gestionar el funcionamiento del habitat Natura 2000.

La última aplicación presenta un base de datos de lugares con ondulaciones costeras a lo largo del oeste de Europa y el noroeste de África, identificándolas visualmente mediante imágenes de Google Earth. En este trabajo también se explora el potencial de usar técnicas de data mining para el desarrollo estudios que abarcan gran cantidad de datos. Nuestra compilación de información se limitó a dos zonas, que representan una parte pequeña de las costas mundiales. Conforme más datos se dispongan, este tipo de análisis genera resultados mucho más significativos. Las reglas de asociación proporcionan nuevos instrumentos para estudiar los datos desde diferentes perspectivas. Esto nos ha permitido encontrar relaciones desconocidas, así como confirmar que el sistema funciona al derivar relaciones triviales a partir de los datos.