UNIVERSIDAD
DE GRANADA

Patricio Yeste Donaire

Hydrological modelling and its application to the study of the impacts of climate change in the Spanish catchments

Tesis doctoral / PhD Thesis

Memoria presentada para optar al título de Doctor por la Universidad de Granada dentro del Programa de Doctorado en Física y Ciencias del Espacio

Directoras

Dra. Yolanda Castro Díez

Dra. María Jesús Esteban Parra

Departamento de Física Aplicada

Granada, Noviembre 2022

Editor: Universidad de Granada. Tesis Doctorales
Autor: Patricio Yeste Donaire
ISBN: 978-84-1117-638-5
URI: https://hdl.handle.net/10481/79169

Financiación / Funding

Esta tesis ha sido posible gracias a la concesión de una beca para la Formación de Profesorado Universitario (FPU) referencia FPU17/02098 del Ministerio de Educación, Cultura y Deporte de España. Este trabajo fue parcialmente financiado por el Ministerio de Economía y Competitividad de España y el Fondo Europeo de Desarrollo Regional (FEDER) por medio de los proyectos CGL2017-89836-R y CGL2013-48539-R, con ayuda adicional a través del proyecto P20_00035, financiado por FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades, y el proyecto B-RNM-336-UGR18, financiado por FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento. Estos proyectos proporcionaron acceso al servicio de Supercomputación de la Universidad de Granada (https://supercomputacion.ugr.es/) para llevar a cabo las simulaciones requeridas en esta tesis. Financiación adicional fue otorgada por medio de una Ayuda para Estancias Breves de tres meses referencia EST19/00169 asociada a la beca FPU en el Hydrology and Quantitative Water Management Group de la Universidad de Wageningen (Países Bajos), y una beca Erasmus+ del programa Eramus+ de la Unión Europea para hacer una estancia de investigación de dos meses en el Institute of Environmental Science and Geography de la Universidad de Potsdam (Alemania).

/

This thesis has been possible through the granting of a Faculty Training Programme (FPU) grant reference FPU17/02098 from the Ministry of Education, Culture and Sport of Spain. This work was partially funded by the Spanish Ministry of Economy and Competitiveness and the European Community Funds (FEDER) projects CGL2017-89836-R and CGL2013-48539-R, with additional support from project P20_00035, funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades, and project B-RNM-336-UGR18, funded by FEDER/Junta de

Andalucía-Consejería de Economía y Conocimiento. These projects provided access to the High-Performance Computing service at University of Granada (https://supercomputacion.ugr.es/) to carry out the required simulations for this thesis. Additional funding was provided through a three-month Aid for Research Stays reference EST19/00169 associated to the FPU grant in the Hydrology and Quantitative Water Management Group at Wageningen University (The Netherlands), and an Erasmus+ grant from the Erasmus+ funding programme of the European Union for a two-month research stay in the Institute of Environmental Science and Geography at University of Potsdam (Germany).

Summary

Large-sample hydrology is intended to provide a reliable characterization of the spatial variability in hydrologic processes at the global, continental and regional scales. With a primary focus being placed on large-scale streamflow processes and streamflow datasets, large-sample approaches are increasingly supplemented by the profusion of remote sensing data containing key information on the different components of the water balance at different spatial and temporal resolutions.

In this thesis, a large-sample application of the Variable Infiltration Capacity (VIC) model has been carried out for a representative number of headwater catchments located in Spain in order to 1) assess the model performance against streamflow observations and satellite-based evaporation data across the Spanish domain and 2) to perform hydrologic projections and analyse the impacts of climate change on the various water storages and fluxes that integrate the water balance.

The application of the VIC model in the Spanish catchments has been divided into four study cases to uncover the potential of combining streamflow and evaporation data into the different stages of the hydrological modelling exercise. This effort extends to the sensitivity analysis stage, the calibration of model parameters, the evaluation of model performance for the streamflow and evaporation simulations, and the obtention of hydrologic projections for the water balance components.

Firstly, the VIC model was calibrated for 31 headwater catchments belonging to the Duero River Basin against monthly streamflow observations, and its performance was evaluated for monthly simulations of streamflow and evaporation. The VIC model produced satisfactory adjustments to both variables and largely improved the benchmark performance. A sensitivity analysis was carried out on the basis of a Monte Carlo experiment to quantify the parameter sensitivities for each component of the water balance implemented in VIC and understand their strong interdependency. A final equifinality

assessment revealed that a considerable number of parameter combinations yielded similar performance estimates to the optimum determined during calibration for monthly streamflow.

The second study case included those 24 headwater catchments from Duero where the VIC model produced better performance estimates for monthly streamflow during calibration and evaluation to study the impacts of climate change on atmospheric and land surface variables. The VIC model simulations were conducted using a Euro-CORDEX multi-model ensemble of 18 members under the Representative Concentration Pathways (RCP) scenarios RCP4.5 and RCP8.5. The projected hydrologic changes evidenced a shift towards an enhanced evaporative regime for all the catchments, attaining reductions of annual streamflow of up to 40% for several catchments. The changes in precipitation and evaporation were subject to a strong intra-annual variability and could adequately explain the seasonal streamflow detriments.

The next study case presents a framework for integrating streamflow and evaporation data into the sensitivity analysis and calibration stages based on the mathematical definition of the Multi-Objective Optimization Problem (MOOP). The framework was applied in three studied catchments, two located in the Guadalquivir River Basin and one in Duero. The VIC soil parameters and the routing parameters were identified as influential to the daily streamflow performance, whereas the vegetation parameters were important to the performance for monthly evaporation. The multi-objective calibration experiments produced an adequate adjustment to both variables simultaneously and showed the specific benefits of Pareto-based approaches compared to the long-established streamflow-only calibration exclusively performed against streamflow data.

Lastly, a large-sample application of the VIC model was carried out for a total of 189 headwater catchments encompassing the main River Basin Districts in Spain. Daily streamflow and monthly evaporation data were combined to study parameter sensitivities and calibrate and evaluate the VIC model using meteorological observations. Results manifested an improvement in the joint performance for daily streamflow and monthly evaporation if their metrics are combined into a composite single objective during calibration. Hydrologic projections were performed to analyse the impacts of climate change in the studied catchments using regional climate simulations conducted with the Weather Research and Forecasting (WRF) model for the Iberian Peninsula under the RCP scenarios RCP4.5 and RCP8.5, and evinced that most of the

catchments will likely experience a more evaporative regime with marked reductions of streamflow and increasing soil drying conditions under future climate.

This thesis constitutes an important contribution to the field of large-sample hydrology and the application of multi-criteria approaches focused on the integration of streamflow and evaporation data into modelling frameworks. This thesis provides an insightful characterization of the VIC model performance and the VIC predictive capabilities for a large and representative number of Spanish catchments. Results from this investigation will help steer future developments to advance the process representation in hydrologic models and increase the reliability of hydrologic projections in the context of global warming.

Resumen

La hidrología macroescala tiene como objetivo proporcionar una caracterización fidedigna de la variabilidad espacial de los procesos hidrológicos a escala global, continental y regional. Siendo su principal centro de atención los procesos de caudal a gran escala y las bases de datos de caudal, los enfoques macroescala son complementandos cada vez más con datos de teledetección que contienen información clave sobre los diferentes componentes del balance de agua a distintas resoluciones espaciales y temporales.

En esta tesis se ha llevado a cabo una aplicación macroescala del modelo hidrológico Variable Infiltration Capacity (VIC) para un número representativo de cuencas de cabecera localizadas en España con objeto de 1) determinar el ajuste del modelo frente a observaciones de caudal y datos de evaporación basados en información de satélite sobre el dominio español, y 2) obtener proyecciones hidrológicas y analizar los impactos del cambio climático en los flujos y los almacenamientos de agua que integran el balance de agua.

La aplicación del modelo VIC en las cuencas españolas se ha dividido en cuatro casos de estudio para analizar el potencial de combinar datos de caudal y evaporación en las distintas etapas del ejercicio de modelización hidrológica. Este esfuerzo se extiende a la etapa de análisis de sensibilidad, la calibración de los parámetros del modelo, la evaluación del ajuste del modelo para las simulaciones de caudal y evaporación, y la obtención de proyecciones hidrológicas para los componentes del balance de agua.

En primer lugar, el modelo VIC se calibró en 31 cuencas de cabecera ubicadas en la Demarcación Hidrográfica del Duero frente a observaciones mensuales de caudal, y el ajuste se evaluó para las simulaciones mensuales de caudal y evaporación. El modelo VIC produjo ajustes satisfactorios para ambas variables y mejoró el ajuste seleccionado como referencia comparativa. Se realizó un análisis de sensibilidad en base a una simulación de Monte Carlo para cuantificar las sensibilidades de los parámetros correspondientes a cada componente del balance de agua implementado en VIC y así entender la fuerte

interdependencia entre ellos. Un análisis final de la equifinalidad del modelo reveló que un número considerable de combinaciones de parámetros produjo un ajuste similar al óptimo determinado durante la calibración del caudal mensual.

El segundo caso de estudio incluyó aquellas 24 cuencas pertenecientes a la Demarcación Hidrográfica del Duero donde el modelo VIC produjo un mejor ajuste al caudal mensual durante la calibración y la evaluación para así estudiar los impactos del cambio climático en las variables atmosféricas y del suelo. Las simulaciones del modelo VIC fueron conducidas usando un ensamblado de 18 miembros de la base de datos Euro-CORDEX bajo los escenarios de emisión RCP4.5 y RCP8.5. Los cambios hidrológicos proyectados evidenciaron un cambio hacia un régimen evaporativo más pronunciado para todas las cuencas, alcanzando reducciones del caudal anual de hasta un 40% en varias cuencas. Los cambios de la precipitación y la evaporación estuvieron sujetos a una fuerte variabilidad intra-anual y pudieron explicar adecuadamente los descensos estacionales de caudal.

El siguiente caso de estudio presenta un marco teórico para la integración de datos de caudal y evaporación en las etapas de análisis de sensibilidad y calibración tomando como base la definición matemática del problema de optimización multi-objetivo (MOOP). El marco fue aplicado en tres cuencas de estudio, dos localizadas en la Demarcación Hidrográfica del Guadalquivir y una en la Demarcación Hidrográfica del Duero. Los parámetros del suelo del modelo VIC y los parámetros de la propagación del caudal fueron influyentes para el ajuste del caudal diario, mientras que los parámetros de la vegetación fueron importantes para el ajuste de la evaporación mensual. Los experimentos de calibración multi-objetivo produjeron un ajuste adecuado para ambas variables simultáneamente y mostraron los beneficios específicos de una optimización de Pareto en comparación con la tradicional calibración frente a datos de caudal exclusivamente.

Por último, se llevó a cabo una aplicación macroescala del modelo VIC para un total de 189 cuencas abarcando las principales Demarcaciones Hidrográficas de España. Se combinaron datos diarios de caudal y datos mensuales de evaporación para estudiar la sensibilidad de los parámetros y calibrar y evaluar el ajuste del modelo empleando observaciones meteorológicas. Los resultados manifestaron una mejora del ajuste simultáneo del caudal diario y la evaporación mensual si sus métricas eran combinadas dentro de una sola función objetivo durante la calibración. Las proyecciones hidrológicas

permitieron analizar los impactos del cambio climático en las cuencas de estudio usando simulaciones del modelo climático regional Weather Research and Forecasting (WRF) para la Península Ibérica bajo los escenarios de emisión RCP4.5 y RCP8.5, y mostraron que la mayoría de las cuencas experimentarán probablemente un régimen más evaporativo con marcados descensos del caudal y crecientes condiciones de aridez en el futuro.

Esta tesis constituye una contribución importante para el campo de la hidrología macroescala y la aplicación de enfoques mulicriterio centrados en la integración de datos de caudal y evaporación en la modelización. Este trabajo proporciona una caracterización detallada del ajuste del modelo VIC y sus capacidades predictivas para un número importante y significativo de cuencas en España. Los resultados de esta investigación ayudarán a encaminar futuros desarrollos para la representación de procesos en modelos hidrológicos y permitirán obtener proyecciones hidrológicas más fiables en el contexto del calentamiento global.

Contents

List of Figures

List of Tables

Acronyms

AEMET Spanish Meteorological Agency

CDF Cumulative Distribution Function

CEDEX Spanish Center for Public Work Experimentation and Study

CORDEX Coordinated Downscaling Experiment

DELSA Distributed Evaluation of Local Sensitivity Analysis

GCM General Circulation Model

GHG Greenhouse gas

GLDAS Global Land Data Assimilation System

GLEAM Global Land Evaporation Amsterdam Model

GSA Global Sensitivity Analysis

IPCC Intergovernmental Panel on Climate Change

MOEA Multi-Objective Evolutionary Algorithm

MOOP Multi-Objective Optimization Problem

NSE Nash-Sutcliffe Efficiency

RCM Regional Climate Model

RCP Representative Concentration Pathway

ROEA Official Network of Gauging Stations in Spain

RSA Regional Sensitivity Analysis

SAIH Automatic Hydrological Information System in Spain

SCE-UA Shuffled Complex Evolutionary algorithm - University of Arizona

SIMPA Integrated System for Rainfall-Runoff Modelling

SPREAD Spanish PREcipitation At Daily scale

SRC Standardized Regression Coefficient

STEAD Spanish TEmperature At Daily scale

UH Unit Hydrograph

UMD-GLCC University of Maryland Global Land Cover Classification

USDA United States Department of Agriculture

VIC Variable Infiltration Capacity

WRF Weather Research & Forecasting

1 Introduction

1.1 Overview of modelling approaches and datasets in large-sample hydrology

The vulnerability of the water cycle to global warming constitutes a major concern and a key challenge for the hydrologic community (Clark et al., 2016; Blöschl et al., 2019a; Nathan et al., 2019). Climate change is likely to increase the frequency and intensity of hydrologic extremes (Blöschl et al., 2019a; Yang et al., 2019), such as droughts (Tomas-Burguera et al., 2020) and floods (Vormoor et al., 2015), as well as the alteration of the freshwater availability and the snow dynamics in mountainous systems (Viviroli et al., 2011; Mankin et al., 2015; Thackeray et al., 2019). This situation has resulted in an increasing need to develop effective adaptation strategies that mitigate the different hydrologic stresses (Clark et al., 2016; Garrote et al., 2016) and guarantee future water security (Lehner et al., 2019).

The scientific underpinning for understanding and monitoring the water cycle relies upon simulations of water stores and fluxes ranging from local to global scale (Fersch et al., 2020; Pokhrel et al., 2021). Our ability to reproduce existing hydroclimatic conditions and anticipate their future changes depends on an accurate representation of hydrologic processes within both climate and hydrologic model structures (Clark et al., 2015a; Koppa et al., 2022). From a hydrological perspective, this situation brings about the need to enhance physical realism in process-based hydrologic models by developing approaches focused on a significant number of catchments across different hydroclimatic regions (Addor et al., 2020) and on calibration strategies that integrate multivariate data (Clark et al., 2017, 2021; Efstratiadis and Koutsoyiannis, 2010).

In order to enhance the robustness of the findings and conclusions stemming from the analysis of hydrologic datasets and the application of hydrologic models, large-sample hydrology (Addor et al., 2020) and large-scale hydrology

(e.g., Bierkens, 2015; Wood et al., 2011) aim to promote the transferability of knowledge between regions and assess the applicability of hydrologic models and theories at regional, continental and global scales. Large-sample hydrology involves large sets (tens to thousands) of catchments, and its main focus is to provide generalizable knowledge of hydrological processes and models based on a large sample of catchments representing different hydroclimatic conditions with a particular emphasis on streamflow (Addor et al., 2020).

Similarly, large-scale hydrology relies on simulations from land-surface models carried out at the so-called spatial hyper-resolution (> 1 km) to quantify and monitor the terrestrial water cycle at the regional, continental and global scales (Bierkens, 2015; Bierkens et al., 2015; Wood et al., 2011). Land-surface models were initially developed as the land-surface scheme to be coupled to General Circulation Models (GCMs) and Regional Climate Models (RCMs) in order to incorporate the biophysical processes governing the land-atmosphere interaction (Clark et al., 2015a). However, land-surface models have been extensively used as hydrologic models in stand-alone (i.e., uncoupled) applications (e.g., Hamman et al., 2018; Melsen et al., 2016a; Mendoza et al., 2015, 2016; Sepúlveda et al., 2022), and the distinction between hydrologic models and land-surface models has become blurred over time (Clark et al., 2015b).

Large-sample and large-scale hydrologic studies play an important role in supporting water resources planning and quantifying hydrologic changes across scales in the context of a changing climate (Addor et al., 2020; Wood et al., 2011). Large-sample and large-scale applications can help elucidate the different sources of uncertainty in hydrologic projections and assess the relative contribution to the total uncertainty of the various modelling components, such as the effect of the combination of multiple GCM/RCM simulations of precipitation and temperature for different emission scenarios (Marx et al., 2018), the use of different hydrologic and/or land-surface models to determine the projected changes of annual runoff and discharge timing (Melsen et al., 2018), or the extent to which climate change controls parameter sensitivities (Melsen and Guse, 2019) and its implications for model calibration and evaluation (Melsen and Guse, 2021).

The gap between both hydrological disciplines is becoming increasingly reduced, and both fields can be considered as two complementary approaches that attempt to provide a solid understanding of the spatial variability of hydrologic processes and to facilitate the intercomparison of model structures across climates (Addor et al., 2020). This is also manifested in the greater areas

covered by hydrologic models (Beck et al., 2016, 2020), the development of gridded runoff observations (Gudmundsson and Seneviratne, 2016; Ghiggi et al., 2019), the tendency towards finer resolutions in land-surface models (Bierkens, 2015; Wood et al., 2011), and the use of macroscale hydrologic models in large-sample studies (e.g., Mizukami et al., 2017; Newman et al., 2017; Rakovec et al., 2016a,b, 2019; Sepúlveda et al., 2022).

From a data perspective, large-sample hydrology encompasses hydrologic studies founded on large-sample datasets of streamflow observations, hydrometeorological data and hydroclimatic and landscape attributes (Addor et al., 2020). This includes investigations on extreme events (e.g., Blöschl et al., 2017; Do et al., 2017; Gudmundsson et al., 2019), climate change impacts (e.g., Marx et al., 2018; Melsen et al., 2018), variations in terrestrial water storage (e.g., Zhang et al., 2017), model evaluation and benchmarking (e.g., Aerts et al., 2022; Newman et al., 2017; Rakovec et al., 2019), data and modelling uncertainties (e.g., Beck et al., 2017; Coxon et al., 2015), parameter estimates during calibration (e.g., Beck et al., 2016, 2020; Mizukami et al., 2017; Rakovec et al., 2016a,b), and transferability of parameters in space based on parameter regionalization techniques (e.g., Beck et al., 2020; Pool et al., 2021; Rakovec et al., 2019).

But over and above the extensive hydroclimatic characterization commonly provided in large-sample datasets, there is no doubt that streamflow is considered a category of its own (Addor et al., 2020). Streamflow datasets are primarily based on individual contributions from national hydrological services, which constitute the building blocks of continental and global streamflow repositories currently available. The role of national water archives is of capital importance in this respect, and ultimately, it is the international collaboration among national authorities worldwide which makes it possible to tackle this complex challenge (Addor et al., 2020).

On a global scale, the Global Runoff Data Base[1] (GRDB) represents the biggest community effort and the most widely used streamflow dataset in hydrologic studies. The GRDB dataset includes daily and monthly streamflow observations for more than 9000 stations and is maintained by the Global Runoff Data Centre (GRDC), an initiative supported by voluntary contributions from national authorities around the world. Notably, GRDB has been recently expanded by means of the Global Streamflow Indices and Metadata (GSIM)

[1] https://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html

archive (Do et al., 2018; Gudmundsson et al., 2018) to include time series of streamflow indices in stations where raw streamflow data are not publicly available.

At the continental and national scales, the European Water Archive (EWA) is a comprehensive dataset comprising streamflow observations for more than 3000 gauging stations in Europe provided by 29 national hydrological services. The EWA archive has not been updated since October 2014, and the former EWA stations have been progressively included in the GRDB dataset and are now regularly updated there[2]. Moreover, the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) dataset (Newman et al., 2015) gathers hydrometeorological and landscape information for 671 catchments in the contiguous United States. A similar approach to that followed in CAMELS has been extended to produce large-sample datasets in Australia (CAMELS-AUS, Fowler et al., 2021), Great Britain (CAMELS-GB, Coxon et al., 2020), Chile (CAMELS-CL, Alvarez-Garreton et al., 2018) and Canada (CANOPEX, Arsenault et al., 2016).

In Spain, two national initiatives stand out as the most important efforts in the context of large-sample hydrology: first, the streamflow observations monitored in Automatic Hydrological Information Systems (SAIHs, Sistemas Automáticos de Información Hidrológica) and in the Official Network of Gauging Stations (ROEA, Red Oficial de Estaciones de Aforo) are jointly provided in the SAIH-ROEA dataset[3], a national archive annually updated that collects daily and monthly streamflow observations for the main River Basin Districts in Spain; secondly, the Integrated System for Rainfall-Runoff Modelling (SIMPA, Sistema Integrado de Modelación Precipitación-Aportación; Estrela and Quintas, 1996; Alvarez et al., 2005) stands as the greatest modelling effort available for the Spanish domain through a set of monthly rasterized maps for different hydrologic variables including runoff, actual evaporation and other relevant water storages and fluxes[4]. Both initiatives are maintained and continuously improved by the Spanish Center for Public Work Experimentation and Study (CEDEX, Centro de Estudios y Experimentación de Obras Públicas), and constitute the basis for water planning purposes in the country.

[2]https://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa.html

[3]https://www.miteco.gob.es/en/cartografia-y-sig/ide/descargas/agua/anuario-de-aforos.aspx

[4]https://www.miteco.gob.es/es/agua/temas/evaluacion-de-los-recursos-hidricos/evaluacion-recursos-hidricos-regimen-natural/

The field of large-sample hydrology is now faced with the profusion of satellite remote sensing products. What are the most common uses of remote sensing data in hydrologic studies and how does this affect the application of large-sample approaches? These questions will be addressed in next section.

1.2 Hydrologic applications of satellite remote sensing data

Satellite remote sensing data have existed since the launch of Sputnik 1 and Explorer 1 in the 1950s (Lettenmaier et al., 2015). However, it has not been until recent decades that satellite remote sensing has proven to have a significant impact on hydrologic studies (Cui et al., 2019; Dembélé et al., 2020a). Most of the progress in remote sensing applications in hydrological sciences can be attributed to the possibility of extending the knowledge of hydrologic processes from field experiments to the regional, continental and global scales, as well as to data-scarce areas with little or no access to hydroclimatic information (Lettenmaier et al., 2015). This implies the use of satellite sensors to retrieve important variables from a water balance perspective, including precipitation, streamflow, evaporation rates, terrestrial water storage, soil moisture, snow and surface water storage (Dembélé et al., 2020b).

Global precipitation products are available at multiple spatial and temporal resolutions, and are of capital importance for regions lacking in-situ measurements (Lettenmaier et al., 2015). For instance, the NASA's Global Precipitation Measurement (GPM) mission[5] provides gridded precipitation data at 10 km resolution and 30-minute time step. Gridded estimates of precipitation from satellite sensors and their ever increasing accuracy have fostered the development of global flood and drought forecasting systems, such as the Global Flood Monitoring System[6] (GFMS) developed at the University of Maryland (Wu et al., 2014).

In-situ observations of streamflow are commonly performed using rating curves that relate river stage to discharge. Based on this approach, the challenges of satellite streamflow estimates are obtaining cross-sectional velocities in rivers to define rating curves and producing accurate measures of river

[5]https://gpm.nasa.gov/
[6]http://flood.umd.edu/

stage. Current satellite remote sensing techniques are unable to measure cross-sectional velocities, and the goal of estimating streamflow from satellite sensors remains elusive (Lettenmaier et al., 2015). However, the field of global hydrology and in particular global streamflow estimates are expected to significantly advance thanks to the Global River Widths from Landsat (GRWL) database (Allen and Pavelsky, 2018) and the Surface Water and Ocean Topography[7] (SWOT) mission, which is planned to launch on December 12, 2022 to provide an inventory of water bodies and rivers with more than 100 m width and to measure global storage changes at sub-monthly, seasonal, and annual time scales.

Evaporation represents the nexus between the water and energy budgets, and is the second largest component of the global water balance (Konapala et al., 2020; Koppa et al., 2022). Although there is a wide variety of methods to estimate evaporation from remote sensing data, the two major variants regarding techniques that are commonly used are represented in Vinukollu et al. (2011) and Miralles et al. (2011). Vinukollu et al. (2011) estimated global evaporation from satellite measurements using the Penman-Monteith formulation, the Priestley-Taylor approach and the Surface Energy Budget System (SEBS) method (Su, 2002). On its part, Miralles et al. (2011) reported global evaporation estimates using passive microwave remote sensing data and the algorithms in the Global Land Evaporation Amsterdam Model[8] (GLEAM). These two variants have been extensively used in global climate and hydrologic research (Lettenmaier et al., 2015), and GLEAM has become an important reference in hydrologic studies to calibrate and evaluate hydrologic models (e.g., Bouaziz et al., 2021; Dembélé et al., 2020b,a; Koppa et al., 2019).

Terrestrial Water Storage (TWS) estimates for the globe are available from the Gravity Recovery and Climate Experiment[9] (GRACE) mission, a collaboration between NASA and the German Space Agency (DLR). The GRACE dataset is based on measurements of the gravitational field in the north-south direction for each orbit, and is provided at 1º resolution and a monthly time step. GRACE has allowed the hydroclimatic community to better quantify the water balance at large scales, and has enabled the indirect estimation of continental and global mass fluxes when combined with other satellite products (Lettenmaier et al., 2015).

[7] https://swot.jpl.nasa.gov/
[8] https://www.gleam.eu/
[9] https://grace.jpl.nasa.gov/

On the other hand, soil moisture is a central pillar of land-surface hydrology and is an important boundary condition regarding the land-atmosphere interactions (Seneviratne et al., 2010). Satellite soil moisture products are available only for the first few centimeters of the soil profile, and are usually combined with models to estimate complete soil moisture profiles (Lettenmaier et al., 2015). Relevant satellite missions to retrieve soil moisture data include the Soil Moisture and Ocean Salinity (SMOS) mission[10] from the European Space Agency (ESA) and the Soil Moisture Active and Passive (SMAP) mission[11]. SMOS provides soil moisture information at ~ 35 km resolution and a periodicity of 2-3 days, while SMAP improves the resolution of SMOS and provides soil moisture estimates at 3, 9 and 36 km for the same time step.

Finally, satellite measurements of snow-related variables and surface water storages (i.e., lakes, reservoirs) are currently evolving. Snow has been the first hydrologic variable measured through remote sensing techniques, and although snow cover information can be accurately estimated based on retrievals from the Moderate Resolution Imaging Spectroradiometer[12] (MODIS), snow water equivalent measurements are still elusive, particularly in mountainous areas with an heterogeneous distribution of snow (Lettenmaier et al., 2015). MODIS provides a large suite of products including land, atmosphere, cryosphere and ocean information at 250 m, 500 m and 1 km resolutions. MODIS has also been used to estimate storage variations in large reservoirs (e.g., Gao et al., 2012), although surface water storage retrievals are expected to greatly improve in the light of results of the SWOT mission.

Satellite remote sensing data have the potential to improve land-surface models and hydrologic models via data assimilation or via model calibration (Dembélé et al., 2020a; López López et al., 2017). Data assimilation is used to continuously update model states with observations and compensate structural deficiencies, and is a common procedure to merge satellite and ground-based observations with land-surface models (Wood et al., 2011) and to generate global and regional climate reanalysis products (Hersbach et al., 2020). To calibrate model parameters, satellite data are normally combined with in-situ observations of streamflow following either a single-objective or a multi-objective optimization approach (Dembélé et al., 2020a; Efstratiadis and Koutsoyiannis, 2010; López López et al., 2017).

[10]https://www.esa.int/Applications/Observing_the_Earth/FutureEO/SMOS
[11]https://smap.jpl.nasa.gov/
[12]https://modis.gsfc.nasa.gov/data/

It is also important to highlight the growing role played by satellite information in relation to the detection and accurate quantification of the impacts of climate change on water resources availability. The use of satellite-based algorithms to retrieve evaporation and terrestrial water storage information represents an unprecedented opportunity to monitor the dynamics and the climate-driven changes in evaporative fluxes (Konapala et al., 2020; Koppa et al., 2022) and subsurface water storage (Pokhrel et al., 2021). Moreover, evaporation is expected to increase as a consequence of global warming (IPCC, 2021), and the projected declines for terrestrial water storage will likely lead to more severe and frequent droughts (Pokhrel et al., 2021). These changes can pose a challenge for future water security and water resources availability from regional to global scale (Lehner et al., 2019; Konapala et al., 2020; Koppa et al., 2022; Pokhrel et al., 2021).

Therefore, large-sample hydrology can strongly benefit from the inclusion of satellite remote sensing data into modelling frameworks in order to calibrate and evaluate models for more than one hydrologic variable (traditionally streamflow) and draw more robust conclusions on catchment functioning for a large number of catchments (Clark et al., 2017; Rakovec et al., 2016a,b, 2019). This is a promising solution to overcome the limitations of the so-called streamflow-only calibration (Dembélé et al., 2020a), and a step forward in the direction of "What is my model good for?" rather than "How good is my model?" (Clark et al., 2021). This will help attain a more robust characterization of the water cycle, providing a solid foundation for an improved water planning and greater resilience to the impacts of climate change.

1.3 Objectives and thesis outline

The main goal of this thesis is to develop a hydrological modelling framework in order to examine the streamflow and evaporation dynamics and study the impacts of climate change for a large set of Spanish catchments. As part of the Iberian Peninsula, the Spanish domain constitutes a region where the effects of climate change are already noticeable and are expected to be much more pronounced by the end of the 21th century (IPCC, 2021). The Iberian Peninsula has been previously identified as a hotspot (Diffenbaugh and Giorgi, 2012), and has manifested recurrent droughts and an increasing tendency towards

aridity conditions for the last decades (García-Valdecasas Ojeda et al., 2021; Páscoa et al., 2017).

From a hydrological perspective, the Spanish catchments has shown a marked reduction of streamflow during the last half century (Lorenzo-Lacruz et al., 2012, 2013), and evaporative fluxes play a dominant role concerning the water balance in the entire region (García-Valdecasas Ojeda et al., 2020b; Vicente-Serrano et al., 2014). These changes are expected to exacerbate under climate climate change (García-Valdecasas Ojeda et al., 2020a, 2021), and can pose an important threat for the future water planning and management in the country. Thus, the simultaneous integration of streamflow and evaporation data into modelling frameworks represents a promising solution to achieve a more reliable quantification of the water balance and a more robust approach to assess the future impacts of climate change in Spain.

The modelling framework developed in this thesis will be focused on the Variable Infiltration Capacity (VIC) model (Liang et al., 1994, 1996), one of the most widely used hydrologic models in hydrologic studies (Addor and Melsen, 2019). The VIC model has been successfully implemented in many previous large-sample studies and large-scale applications (e.g., Melsen et al., 2018; Mizukami et al., 2017; Rakovec et al., 2019; Sepúlveda et al., 2022). This, together with the large number of developments since its first appearance in 1994 (Liang et al., 1994), makes VIC an excellent choice for the purpose of this thesis. The VIC model capabilities to integrate streamflow observations and satellite-based evaporation data will be thoroughly investigated, and its predictive capabilities will be leveraged to perform hydrologic projections for the Spanish catchments. The application of the VIC model will be carried out in a step-by-step manner for four study cases focused on different modelling approaches and catchments belonging to the Spanish domain. The four study cases have been specifically design to assess the potential of integrating streamflow and evaporation data into the different stages of the hydrological modellling exercise: sensitivity analysis, calibration and evaluation, and the predictive stage using climate change information (i.e., hydrologic projections).

The rest of this thesis is structured as follows: Chapter 2 presents the Spanish catchments and the hydroclimatic data that are to be used in the modelling framework. Chapter 3 sets out the VIC model structure and the spatial configuration approach for the modelling exercise. Chapter 3 also introduces the four study cases representing the central core of the VIC model application, and the results and insights deriving from them are extensively described in

Chapters 4 to 7. Finally, Chapter 8 summarizes the main conclusions from this thesis.

2 Study area and Data

This chapter presents the Spanish catchments constituting the study area of this research and the different hydroclimatic datasets that will be used during the hydrological modelling exercise. Hydroclimatic information includes streamflow, precipitation and temperature observations, satellite-based evaporation data and two datasets of climate change projections.

2.1 Headwater catchments and streamflow observations

This thesis is focused on a large sample of headwater catchments distributed across the main River Basin Districts in Spain (Fig. 2.1, Table 2.1). 441 headwater catchments were defined according to the physical location of a large number of reservoirs and gauging stations gathered from SAIH-ROEA[1], an integrated network of streamflow observations administrated by the Spanish Center for Public Work Experimentation and Study (CEDEX, Centro de Estudios y Experimentación de Obras Públicas[2]) that includes hydrologic information for more than 350 reservoirs and 900 gauging stations belonging to the main River Basin Districts in Spain.

The headwater catchments were selected according to two criteria: 1) the absence of upstream reservoirs, and 2) the largest possible catchment was kept in those cases where several gauging stations fell within the same catchment area. Streamflow observations from SAIH Hidrosur[3] and the Catalan Water

[1]https://www.miteco.gob.es/en/cartografia-y-sig/ide/descargas/agua/anuario-de-aforos.aspx

[2]https://ceh.cedex.es/anuarioaforos/default.asp

[3]http://www.redhidrosurmedioambiente.es/saih/

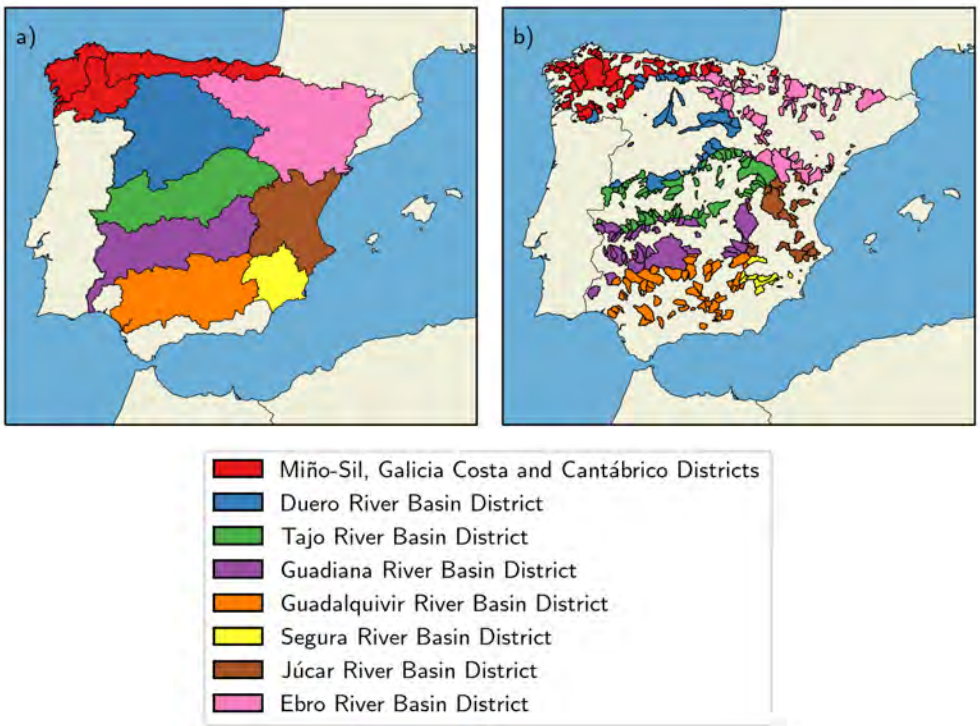


Fig. 2.1: River Basin Districts in Spain and the 441 headwater catchments considered in this thesis. (a) River Basin Districts. (b) Topographic boundaries of the 441 headwater catchments.

Agency[4] (ACA, Agencia Catalana del Agua) are not integrated into the SAIH-ROEA network and were not considered in this investigation.

The SAIH-ROEA database is delivered at daily and monthly time scales and includes streamflow time series starting as early as 1911 for some catchments. Streamflow estimates in gauging stations are retrieved from water depth measurements using rating curves, and a daily water balance of water storages and releases is applied in reservoirs to calculate inflow data, which are later aggregated into monthly values. A data quality control of the number of gaps in the streamflow observations will be applied in each particular study case to select a subset of headwater catchments with enough data available. Specific details on the quality control for streamflow data will be addressed later in this thesis for each study case.

[4]https://aca.gencat.cat/ca/inici/

Table 2.1: Number of headwater reservoirs and gauging stations per River Basin District considered in this thesis.

River Basin District	Reservoirs	Gauging stations	Total
Miño-Sil, Galicia Costa and Cantábrico	28	64	92
Duero	18	15	33
Tajo	31	49	80
Guadiana	18	29	47
Guadalquivir	36	23	59
Segura	13	0	13
Júcar	20	6	26
Ebro	42	49	91
Total	**206**	**235**	**441**

2.2 Precipitation and temperature observations

Precipitation, maximum temperature and minimum temperature data were extracted from two high-resolution (~ 5 km) daily gridded datasets: the Spanish PREcipitation At Daily scale (SPREAD) dataset[5] (Serrano-Notivoli et al., 2017) for precipitation data and the Spanish TEmperature At Daily scale (STEAD) dataset[6] (Serrano-Notivoli et al., 2019) for temperature data. Both datasets cover the whole territory of Spain and were built using information from an extensive net of observatories (> 12000 for SPREAD and > 5000 for STEAD) of precipitation and maximum and minimum temperature provided by several administrations including the Spanish Meteorological Agency (AEMET, Agencia Estatal de Meteorología) and some River Basin Districts.

Precipitation series span the period 1950-2012 in peninsular Spain and the period 1971-2012 in the Balearic and Canary Islands, whereas maximum and minimum temperature series are longer and cover the periods 1901-2014 and 1971-2014 for peninsular Spain and the Canary and Balearic Islands, respectively. The raw meteorological data were subject to an in-depth quality control, and gaps were independently filled on each day and location to produce a complete dataset. Both gridded products were finally constructed through the calculation of Reference Values (RVs) and were accompanied by an estimation of uncertainty for all days and locations (see Serrano-Notivoli et al.,

[5]https://digital.csic.es/handle/10261/141218
[6]https://digital.csic.es/handle/10261/177655

2017, 2019, for a detailed description of the methodological steps integrating the quality control and gridding stages).

2.3 Satellite-based evaporation data

The Global Land Evaporation Amsterdam Model (GLEAM, Martens et al., 2017; Miralles et al., 2011) comprises a set of algorithms for the estimation of the different components of land evaporation (i.e., evaporation from bare soil and from open water bodies, transpiration, interception loss and sublimation). GLEAM evaporation is calculated using microwave remote sensing data and a data assimilation infrastructure that has been optimized to operate at the global scale.

GLEAM version 3 (Martens et al., 2017) is delivered as two global gridded datasets (version 3a and version 3b) at 0.25º resolution that differ in the time period covered and in their forcings: version 3a spans the period from 1980 to the present day and is based on satellite and reanalysis data, whereas data in version 3b are available from 2003 onward and are exclusively based on satellite information. GLEAM version 3a has been selected in this thesis as the reference dataset for evaporation given its greater temporal coverage, which makes it suitable for the modelling purpose of this thesis.

The GLEAM dataset is gaining increasing attention in the hydrologic community given the prevalent role of evaporative fluxes in the water balance and the sparse availability of in-situ observations of evaporation. GLEAM has been successfully used in many previous hydrologic studies including evaporation as an evaluation objective during a post-calibration phase (e.g., Beck et al., 2021; Bouaziz et al., 2021) or as calibration target itself (e.g., Dembélé et al., 2020a,b; Koppa et al., 2019; López López et al., 2017).

2.4 Climate change projections

In this thesis, two datasets of climate change projections were chosen to perform hydrologic projections for the future climate: 1) an ensemble of 18 Regional Climate Model (RCM) experiments from the Euro-CORDEX database (Table 2.2), and 2) simulations carried out with the Weather Research and

Table 2.2: Ensemble of combinations RCM+GCM chosen from the Euro-CORDEX database.

RCM	GCM
RCA4	CNRM-CERFACS-CNRM-CM5
	ICHEC-EC-EARTH, r12i1p1
	IPSL-IPSL-CM5A-MR
	MOHC-HadGEM2-ES
	MPI-M-MPI-ESM-LR
CCLM4-8-17	CNRM-CERFACS-CNRM-CM5
	ICHEC-EC-EARTH
	MOHC-HadGEM2-ES
	MPI-M-MPI-ESM-LR
HIRHAM5	ICHEC-EC-EARTH, r3i1p1
	NCC-NorESM1-M
RACMO22E	CNRM-CERFACS-CNRM-CM5
	ICHEC-EC-EARTH, r12i1p1
	ICHEC-EC-EARTH, r1i1p1
	MOHC-HadGEM2-ES
REMO2009	MPI-M-MPI-ESM-LR, r1i1p1
	MPI-M-MPI-ESM-LR, r2i1p1
WRF331F	IPSL-IPSL-CM5A-MR

Forecasting (WRF) model for the Iberian Peninsula in García-Valdecasas Ojeda et al. (2020a,b).

The Euro-CORDEX project[7] (Jacob et al., 2014) is established as the largest climate modelling effort for the European region (Herrera et al., 2020), with a plethora of RCM simulations available at 0.11º and 0.44º that have been the basis of a great number of hydrological impact studies for many European catchments (e.g., Gampe et al., 2016; Papadimitriou et al., 2016; Meresa and Romanowicz, 2017; Hakala et al., 2018; Hanzer et al., 2018; Vieira et al., 2018; Fonseca and Santos, 2019; Pastén-Zapata et al., 2020). Daily climate data were gathered at a spatial resolution of 0.11º (EUR-11, ~ 12.5 km) for eight atmospheric variables: precipitation, maximum and minimum temperature, near-surface wind speed, incoming shortwave and longwave radiation, atmospheric pressure and vapour pressure. The multi-model ensemble consists of 18 RCM+GCM combinations for the Representative Concentration Pathways

[7] https://euro-cordex.net/

(RCP) scenarios RCP4.5 and RCP8.5. Data were extracted for the historical period of 1975-2005 and for 2021-2100 as the future period considering the hydrologic year (i.e., from October to September) and the associated hydrologic seasons. The latter was divided into three sub-periods: short-term (2021-2050), mid-term (2041-2070) and long-term (2071-2100) future periods.

The WRF model simulations were driven by the Fourth Version of the Community Climate System Model (WRFCCSM hereafter) and the Max Plank Earth System Model at Low Resolution (WRFMPI hereafter) at 0.088º (~ 10 km) resolution. Daily precipitation and maximum and minimum temperature data were extracted for three study periods based on climatic years (i.e., from December to November): the historical period 1980-2015, the short-term future period 2021-2050 and the long-term future period 2071-2100. The future projections of climate change were carried out under RCP4.5 and RCP8.5. The use of climatic years were preferred over the definition of hydrologic years in order to facilitate the discussion and comparison of results with the findings reported in García-Valdecasas Ojeda et al. (2020a,b).

3 Methods

In this chapter, the VIC model representation of the water balance is introduced. The surface runoff, baseflow and evaporation equations are described, as well as the soil and vegetation parameters and the spatial configuration approach adopted for VIC. The application of VIC in four study cases is finally outlined. Specific methods corresponding to each study case will be further addressed in the next four chapters.

3.1 The VIC model

The VIC model (Liang et al., 1994, 1996) is a semi-distributed macroscale hydrologic model that was initially conceived as a Land-Surface Model (LSM) to be coupled to General Circulation Models (GCMs) and Regional Climate Models (RCMs) as a the land scheme that allows for simulating the biophysical processes involved in the land-atmosphere interaction. Since then, the VIC model has played the role of both a LSM and a hydrologic model and has become one of the most widely used models in hydrological modelling studies (Addor and Melsen, 2019).

Applications of VIC range from a global to a local scale (Hamman et al., 2018; Melsen et al., 2016a) and are performed at the spatial hyper-resolution (≥ 1 km, Bierkens, 2015; Bierkens et al., 2015; Wood et al., 2011). The VIC model is predominantly used in the United States (US) as its development started and continues to be carried out in US research institutions (Addor and Melsen, 2019). This includes the ongoing real-time drought monitoring and forecasting systems at University of California, Los Angeles[1], and the NASA's Land Data Assimilation System[2] (LDAS).

[1] http://www.hydro.ucla.edu/SurfaceWaterGroup/forecast/monitor/index.shtml
[2] https://ldas.gsfc.nasa.gov/

The VIC model implements a wide variety of biophysical processes at the grid-cell scale and computes the water and the energy balances at the land surface. Such processes encompass, among others, the runoff generation, evaporative fluxes into the atmosphere, thermal processes in the soil profile, snow and frozen soil dynamics, the impoundment of surface water in lakes and wetlands, and specific processes regulating the carbon cycle (e.g., photosynthesis, autotrophic respiration and heterotrophic respiration). Its extensive parameterization is distributed into different modules or submodels to provide scalability and a varying degree of complexity depending on the modelling purpose.

In this thesis, the VIC model version 4.2.d[3] was selected to carry out hydrologic simulations of water storages and fluxes at the catchment scale choosing the water balance mode of operation and a daily time step. Hydrologic processes in VIC are modelled for a total of three soil layers vertically interconnected (Fig. 3.1). The sub-grid variability in land cover classes is evaluated statistically, and the spatially heterogeneous structure of the infiltration capacity is conceptualized according to the Xinanjiang model (Zhao et al., 1980). This approach takes into account the sub-grid variability in the soil moisture storage capacity.

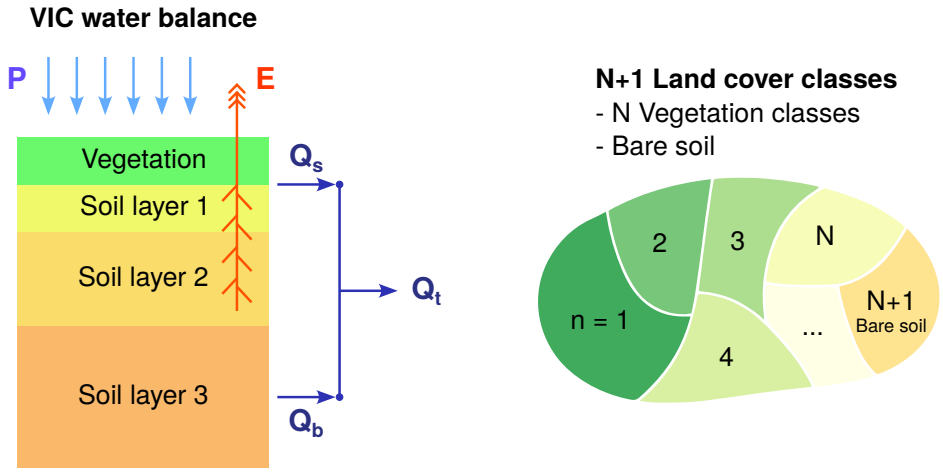


Fig. 3.1: VIC model conceptualization.

[3]https://vic.readthedocs.io/en/vic.4.2.d/

Surface runoff is generated through an infiltration excess applying the Xinanjiang formulation (Zhao et al., 1980) to the upper two soil layers:

$$Q_s = \begin{cases} P - z_2 \cdot (\theta_S - \theta_2) + z_2 \cdot \theta_S \cdot \left(1 - \frac{i_0 + P}{i_m}\right)^{1+b_i}, & P + i_0 \leq i_m \\ P - z_2 \cdot (\theta_S - \theta_2), & P + i_0 > i_m \end{cases} \tag{3.1}$$

For each time step Q_s [L] is the surface (direct) runoff, P [L] is the precipitation, z_2 [L] is the depth of the upper two soil layers, θ_2 is their volumetric soil moisture content, θ_S is their porosity, i_m [L] is the maximum infiltration capacity, i_0 [L] is the infiltration capacity that corresponds to the soil moisture at that time step and b_i is the infiltration shape parameter.

Baseflow is generated in the third soil layer following the Arno formulation (Franchini and Pacciani, 1991), and is expressed as:

$$Q_b = \begin{cases} \frac{D_S D_m}{W_S \theta_S} \cdot \theta_3, & 0 \leq \theta_3 \leq W_S \theta_S \\ \frac{D_S D_m}{W_S \theta_S} \cdot \theta_3 + \left(D_m - \frac{D_S D_m}{W_S}\right) \cdot \left(\frac{\theta_3 - W_S \theta_S}{\theta_S - W_S \theta_S}\right), & \theta_3 > W_S \theta_S \end{cases} \tag{3.2}$$

Here, Q_b [L] is the baseflow for each time step, D_m [L] is the maximum baseflow, D_S is a fraction of D_m, θ_3 is the volumetric soil moisture content of the soil layer 3, θ_S is the porosity in this layer and W_S is a fraction of θ_S. The baseflow recession curve is divided into two parts: a linear part for lower values of θ_3 and a non-linear (quadratic) part for higher values of θ_3. The total runoff generated within a catchment (Q_t) is finally calculated as the sum of Q_s and Q_b (Fig. 3.1).

The water balance in the VIC model considers three types of evaporation depending on each land cover class (N vegetation classes + Bare soil, see Fig. 3.1): evaporation from the canopy layer (E_c), transpiration from the different types of vegetation (E_t) and evaporation from bare soil (E_1). Potential evapotranspiration (E_p) is calculated using the Food and Agriculture Organization (FAO) Penman-Monteith equation (Allen et al., 1998), and represents the atmospheric demand for water vapor. Actual evapotranspiration (E) is obtained as the sum of the three components for each grid cell:

$$E = E_c + E_t + E_1 \tag{3.3}$$

The VIC snow model divides each grid cell into snow bands, thereby accounting for the sub-grid variability in topography, land uses and precipitation. The snow model is applied separately to each snow band and land cover class to calculate the snow depth and the snow water equivalent. The snowpack is represented as a two-layer model that solves the energy and the mass balance and determines whether the snowpack is subject to accumulation or ablation, thus making VIC suitable for applications in any part in the world (Liang et al., 1994, 1996).

3.2 Soil and vegetation parameters

While most of the soil and vegetation parameters required for the application of the VIC model can be obtained from measurable quantities of soil and vegetation characteristics, the conceptual parameters modulating the surface runoff (b_i, see Eq. 3.1) and baseflow generation (D_S, W_S and D_m, see Eq. 3.2) do not have a measurable physical meaning, and their values are to be determined during calibration together with the soil depth. These parameters are the most common calibration parameters according to the recommendations for the calibration of the VIC model[4], and are also in agreement with previous studies using VIC for sensitivity analysis and/or calibration purposes (e.g., Chawla and Mujumdar, 2015; Liang et al., 2004; Mizukami et al., 2017; Melsen et al., 2016b; Oubeidillah et al., 2014; Rakovec et al., 2014, 2019).

Table 3.1 summarizes the soil and vegetation datasets used to extract the VIC soil and vegetation parameters considered observable. Soil properties were taken from SoilGrids1km (Hengl et al., 2014) and EU-SoilHydroGrids ver1.0 (Tóth et al., 2017), all of them at 1 km resolution. In both datasets the different soil properties are provided for seven soil depths up to 2 m (0, 5, 15, 30, 60, 100 and 200 cm). These soil parameters are: (1) bulk density and soil textural classes of the United States Department of Agriculture (USDA) from SoilGrids1km; and (2), field capacity, saturated hydraulic conductivity, porosity and wilting point from EU-SoilHydroGrids ver1.0. The depth of the

[4] https://vic.readthedocs.io/en/vic.4.2.d/Documentation/Calibration/

Table 3.1: Soil and vegetation datasets selected to extract the VIC parameters considered observable.

Dataset	Variable	Resolution	Data source
SoilGrids1km	- Bulk density	1 km	Hengl et al. (2014)
	- USDA textural classes		
EU-SoilHydroGrids ver1.0	- Field capacity	1 km	Tóth et al. (2017)
	- Hydraulic conductivity		
	- Porosity		
	- Wilting point		
UMD-GLCC	- Land cover classes	1 km	Hansen et al. (2000)
GLDAS	- 13 vegetations parameters (e.g., Leaf-Area Index, albedo, minimum stomatal resistance) for each land cover class	1 km	Rodell et al. (2004)

complete soil profile implemented in VIC was established at 2 m to match the depth covered by the soil datasets. The thickness of the first soil layer (d_1) was fixed at 10 cm according to Liang et al. (1996), and the thickness of the second soil layer (d_2) was adjusted during calibration.

The VIC model handles land uses information as a set of vegetation parameters for the different vegetation classes specified in a vegetation library. Here, the University of Maryland Global Land Cover Classification (UMD-GLCC, Hansen et al., 2000) was chosen with a spatial resolution of 1 km, and a total of 13 vegetation parameters (Table 3.2) were defined for each vegetation class in agreement with the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) specifications for VIC[5]. The vegetation parameters remain mostly fixed as default values and are rarely varied during calibration (although see Melsen et al., 2016a; Mendoza et al., 2015; Sepúlveda et al., 2022, for some exceptions).

3.3 Spatial configuration and routing procedure

A semi-distributed implementation of the VIC model was carried out following a catchment-by-catchment approach and choosing a spatial resolution of 0.05º (~ 5 km). This resolution is a common choice in previous studies using VIC

[5]https://ldas.gsfc.nasa.gov/gldas/vegetation-parameters

Table 3.2: VIC vegetation parameters for each vegetation class.

Parameter	Description
$depth_1$	Thickness of the first root zone
$depth_2$	Thickness of the second root zone
rarc	Architectural resistance
rmin	Minimum stomatal resistance
LAI	Leaf-Area Index
albedo	Albedo
rough	Vegetation roughness
disp	Vegetation displacement
wind_h	Height of wind speed measures
RGL	Minimum incoming shortwave radiation for transpiration
rad_atten	Radiation attenuation
wind_atten	Wind speed attenuation through overstory
trunk_ratio	Ratio of total tree height that is trunk

(e.g., Melsen et al., 2016b; Sepúlveda et al., 2022) and other semi-distributed hydrologic model such as the mesoscale Hydrologic Model (mHM, see Marx et al., 2018) and the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (e.g., Beck et al., 2020). In addition, the chosen resolution improves the resolution of RCM simulations available in Euro-CORDEX (0.11º) and the resolution of the WRF model simulations carried out in Argüeso et al. (2012) and in García-Valdecasas Ojeda et al. (2017, 2020a,b) for the Iberian Peninsula (0.088º), providing an increase in quality and reliability for the process representation in space. The semi-distributed implementation of the VIC model was carried out averaging the soil parameters for each grid cell and keeping the vegetation parameters at the original resolution of 1 km as the model considers the sub-grid variability in land cover classes.

The distributed hydrologic variables simulated with VIC for a given catchment were averaged according to the fractional area of each cell falling inside the catchment in order to respect its topographic area (Fig. 3.2). Since the conceptualization of the VIC model does not include the horizontal fluxes between contiguous grid cells, the runoff generated within a catchment needs to be propagated to the catchment outlet in order to determine simulated values of streamflow that can be compared with existing observations. In this thesis, streamflow time series were calculated following two approaches (see

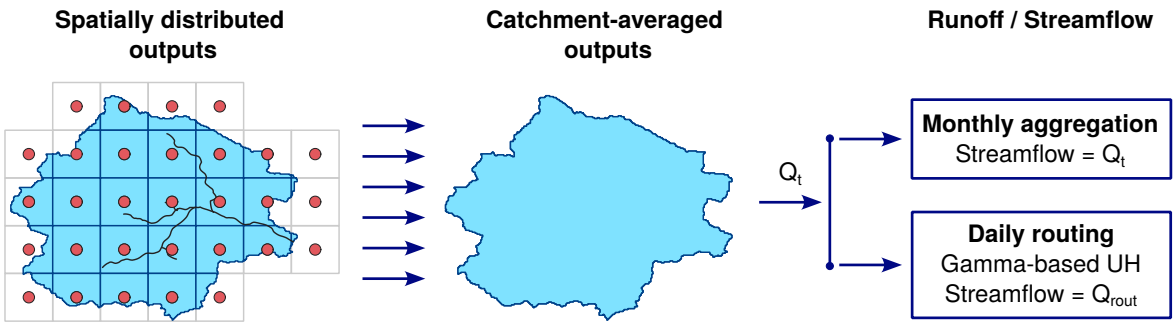


Fig. 3.2: Spatial configuration for the VIC model outputs.

Fig. 3.2): 1) a monthly aggregation of the total runoff generated for a catchment assuming that all the runoff drains by the end of the month; and 2) a Unit Hydrograph (UH) approach (Te Chow et al., 1988) to perform a daily routing of the runoff simulations.

The UH describes the response of a linear system to an input time function according to a certain unit impulse response function $U(t)$ such that $\int_0^\infty U(t)dt = 1$. Then, streamflow data can be calculated as the convolution integral of the runoff time series and $U(t)$:

$$Q_{\text{rout}}(t) = \int_0^t Q_{\text{t}}(\tau)U(t-\tau)d\tau \tag{3.4}$$

Where Q_{t} represents the total runoff, Q_{rout} is the streamflow and $t - \tau$ is the time delay since the impulse was applied (Te Chow et al., 1988). For the purpose of this thesis, a gamma distribution function with two parameters was chosen as the UH:

$$U(t) = \gamma(t : k, \theta) = \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-t/\theta} \tag{3.5}$$

Where k is a shape parameter [-], θ is a timescale parameter [T] and $\Gamma(k)$ is the gamma function. The gamma-based UH constitutes a parsimonious way to represent the time delay between the runoff generated within the catchment

and the streamflow draining through the catchment outlet, and has been widely used to calibrate daily streamflow simulations in previous hydrological modelling studies (e.g., Mizukami et al., 2017; Rakovec et al., 2014, 2019).

3.4 Study cases

The core of this thesis has consisted of the application of the VIC model in four study cases. The four study cases involve the sensitivity analysis, calibration, evaluation and predictive stages of the hydrological modelling exercise, and constitute the content of the next four chapters. For the sake of completeness and clarity, a brief outline for each study case is provided below:

1. **The Duero River Basin under present climate: sensitivity analysis, calibration and evaluation** (Chapter 4; Yeste et al., 2020). The VIC model was calibrated for 31 headwater catchments belonging to the Duero River by following a single-objective optimization approach focused on the monthly simulations of streamflow. The VIC model performance was evaluated for monthly streamflow and monthly evaporation, and a cross-evaluation of the calibrated parameters was carried out to study parameter transferability in space. The Standardized Regression Coefficients (SRC) method (Saltelli et al., 2008) was applied to quantify the sensitivity of the water balance components to the calibration parameters, and an equifinality assessment (Beven, 2006, 2012) was finally performed for monthly streamflow.

2. **The Duero River Basin under future climate: projected hydrologic changes** (Chapter 5; Yeste et al., 2021). The second study case is based on the calibration results attained in the previous study case for the Duero River Basin. The VIC model was applied in 24 headwater catchments where the model produced satisfactory performance estimates for monthly streamflow to obtain hydrologic projections for the water balance variables using the calibrated VIC and an ensemble 18 members from Euro-CORDEX (Jacob et al., 2014). The Quantile Mapping procedure (Themeßl et al., 2011) was adopted as the bias correction method for the meteorological forcings from Euro-CORDEX during the historical period, and the projected hydrologic changes were calculated for the short-term, mid-term and long-term future periods.

3. **Streamflow and evaporation trade-offs: sensitivity analysis and calibration based on Pareto optimization** (Chapter 6). In this study case we propose a framework to integrate streamflow and evaporation data into the sensitivity analysis and calibration stages of the hydrological modelling exercise. The framework was implemented for two catchments located in the Guadalquivir River Basin and one catchment located in Duero. The Distributed Evaluation of Local Sensitivity Analysis (DELSA) method (Rakovec et al., 2014) was applied to identify influential parameters from among the VIC soil and vegetation parameters to the performance against daily streamflow and monthly evaporation data. A Pareto-based sensitivity analysis was subsequently carried out to further refine the previous parameter selection, and the optimum joint performance for daily streamflow and monthly evaporation was finally determined on the basis of a constrained multi-objective calibration approach.

4. **Hydrologic simulations for the Spanish catchments under present and future climate** (Chapter 7). The last study case builds into the experience gained in the rest of the study cases to carry out a large-sample application of the VIC model for 189 Spanish headwater catchments. A Regional Sensitivity Analysis (RSA) method (Hornberger and Spear, 1981) was conducted to select the most important soil and vegetation parameters to represent the streamflow and evaporation dynamics, and the selected parameters were calibrated following two single-objective calibration strategies against daily streamflow and monthly evaporation data. The model performance was evaluated for daily streamflow and monthly evaporation using meteorological observations and WRF historical data García-Valdecasas Ojeda et al. (2020a,b), and the hydrologic projections were performed for the short-term and long-term future periods using regional climate change simulations from WRF.

Table 3.3 summarizes the modelling steps for the application of VIC in the four study cases. Specific methods will be further described for each particular study case in next chapters.

Table 3.3: The four study cases and modelling steps for the application of VIC in this thesis.

Study case	No. catchments	Sensitivity analysis	Calibration and evaluation approach	Hydrologic projections
1) The Duero River Basin under present climate (Chapter 4; Yeste et al., 2020)	31	SRC method	- Calibration for monthly streamflow; evaluation for monthly streamflow and monthly evaporation; cross-evaluation of the calibrated parameters; equifinality assessment for monthly streamflow	—
2) The Duero River Basin under future climate (Chapter 5; Yeste et al., 2021)	24	—	- Calibration based on previous study case; evaluation for monthly streamflow and monthly evaporation using Euro-CORDEX data	- Quantile Mapping method; historical simulations and short-term, mid-term and long-term future projections using Euro-CORDEX data
3) Streamflow and evaporation trade-offs (Chapter 6)	3	DELSA, Pareto-based sensitivity analysis	- Multi-objective calibration for daily streamflow and monthly evaporation	—
4) Hydrologic simulations for the Spanish catchments (Chapter 7)	189	RSA	- Calibration for daily streamflow and monthly evaporation; evaluation for daily streamflow and monthly evaporation using various meteorological observations and WRF data	- Historical simulations and short-term and long-term future projections using WRF data

4 The Duero River Basin under present climate: sensitivity analysis, calibration and evaluation

The content of this chapter is based on the following publication:

Integrated sensitivity analysis of a macroscale hydrologic model in the north of the Iberian Peninsula

Patricio Yeste, Matilde García-Valdecasas Ojeda, Sonia R. Gámiz-Fortis, Yolanda Castro-Díez and María Jesús Esteban-Parra

Published in *Journal of Hydrology*[1]

Abstract

Process-based hydrologic models allow to identify the behaviour of a basin providing a mathematical description of the hydrologic processes underlying the runoff mechanisms that govern the streamflow generation. This chapter focuses on a macroscale application of the Variable Infiltration Capacity (VIC) model over 31 headwater catchments belonging to the Duero River Basin, located in the Iberian Peninsula, through a three-part approach: (1) the calibration and evaluation of the VIC model performance for all the catchments; (2) an integrated sensitivity analysis concerning the soil parameters chosen

[1]DOI: https://doi.org/10.1016/j.jhydrol.2020.125230

for the calibration, and (3) an assessment of equifinality and the efficiency of the calibration algorithm. The temporal evaluation of the model was done for the calibration and the subsequent validation periods, and showed good results for most of the catchments that largely improved the benchmark performance. The spatial performance reflected a high transferability for most parameter combinations, and the least transferable were related to catchments located in the northern mountains. An additional evaluation of the simulated actual evapotranspiration produced satisfactory adjustments to two selected data products. The sensitivity measures were obtained with the Standardized Regression Coefficients method through a post-process of the outputs of a Monte Carlo simulation carried out for 10 000 parameter samples for each catchment. This allowed to quantify the sensitivity of the water balance components to the selected parameters for the calibration and understanding the strong dependencies between them. The final assessment of the equifinality hypothesis manifested that there are many parameter samples with performances as good as the optimum, calculated using the Shuffled-Complex-Evolution Algorithm. For almost all the analyzed catchments the calibration algorithm resulted efficient, reaching the optimal fit. Both the Monte Carlo simulation and the use of a calibration algorithm will be of interest for other feasible applications of the VIC model in other river basins.

4.1 Introduction

Water resources in the Mediterranean Basin have undergone dramatic changes during the 20th century as a consequence of the rising temperatures and the significant decrease of precipitation (García-Ruiz et al., 2011). The effects of climate change in this region are already noticeable and are expected to be much more pronounced by the end of the 21th century (IPCC, 2014). This fact, together with the increasing water demand for agriculture, industry and urban supply, makes the water scarcity problem of paramount importance, being its accurate identification essential for adopting adequate water management strategies and mitigation measures that ensure the sustainability of the water resources (Chavez-Jimenez et al., 2013; Garrote et al., 2016).

Being able to identify the hydrologic behaviour of a basin is necessary in order to assess the effects of climate and land changing conditions, and therefore a profound description of the main hydrologic processes governing the response of the basin is required. In this way, process-based hydrologic models are powerful tools that represent the underlying runoff mechanisms governing the streamflow generation for a given basin, and therefore constitute mathematical hypotheses on how the hydrologic system functions, characterizing the potential changes of the water resources using precipitation and temperature data as inputs variables.

Calibration and validation of hydrologic models are required in order to develop reliable models (Savenije, 2009), and sensitivity analysis should be carried out for a better knowledge of complex models (Song et al., 2015). Moreover, the recognition of the equifinality concept, that is, the existence of many sets of parameters conducive to good adjustments to some target observations (Beven, 2006, 2012), is unavoidable and necessary (Beven and Freer, 2001). The correct identification of modelling uncertainties remains a fundamental question for hydrologic models after several decades of continuous progress, with a particular focus on structural uncertainties as they are more elusive than input and parametric uncertainties (Blöschl et al., 2019a).

In the context of climate modelling, these models are usually called Land-Surface Models (LSMs, Wood et al., 2011). Although there is a subtle difference between a hydrologic model and a LSM, this distinction has become blurred over time (Clark et al., 2015b). In this respect, the Variable Infiltration Capacity (VIC) macroscale hydrologic model (Liang et al., 1994, 1996) has played a major role in the hydroclimatic community according to evidende provided

in Melsen et al. (2016b) in a meta-analysis of 192 peer-reviewed studies where the VIC model was calibrated and validated. Current research has aimed at the calibration of the VIC model over the contiguous United States (CONUS) domain and constitutes a promising approach for the calibration of large-domain hydrologic models (Mizukami et al., 2017; Rakovec et al., 2019; Yang et al., 2019). These studies involve a large-scale parameter estimation through spatially distributed techniques such as the multiscale parameter regionalization (MPR) method (Samaniego et al., 2010) for calibrating the VIC model parameters using transfer functions that relate them to certain physical properties (Mizukami et al., 2017; Rakovec et al., 2019), or the regionalization of various streamflow characteristics for calibrating the model everywhere (Yang et al., 2019).

Since its first development many efforts have been made in order to study the sensitivity of the VIC model, which has been explored in a broad sense in the following terms:

- *Sensitivity to spatio-temporal variability*: the impacts of the implemented spatial resolution in the simulated runoff and other water fluxes have been addressed in various studies, suggesting that there is a high influence of the sub-grid variability of the precipitation on the model performance (Haddeland et al., 2002; Liang et al., 2004). However, a critical spatial resolution under which a better model performance is not necessarily achieved (Liang et al., 2004) could exist. These impacts are also noticeable in calibration and validation exercises with an increase of the model accuracy at higher resolutions (Oubeidillah et al., 2014), although a high transferability of the calibrated parameters across the different resolutions may be an indicator of a poor representation of the spatial variability (Melsen et al., 2016a). Unfortunately, the time step of the calibration and validation has not kept up with the increasing spatial resolution, and this is a crucial aspect for the correct representation of the involved hydrologic processes (Melsen et al., 2016b). The fact that it is more difficult to transfer parameters across temporal resolutions than across the spatial dimension brings the need of a better representation of the spatial variability in macroscale hydrologic models (Melsen et al., 2016a).

- *Sensitivity to boundary conditions*: understanding the boundary conditions as the meteorological forcings that drive the simulations, the VIC model sensitivity to the boundary conditions has been studied through the application of different climate change scenarios and the analysis of the

impacts of changing precipitation and temperature on the hydrology and water resources of several continental river basins (Nijssen et al., 2001), the Pacific Northwest (Vano et al., 2015), or more locally in the Colorado River Basin (Christensen et al., 2004; Bennett et al., 2018) and in the upper Ganga Basin (Chawla and Mujumdar, 2015). Also, these kind of studies sometimes are carried out in conjunction with other sensitivity analysis, i.e. the combined and segregated effects of climate change and land use changes on streamflow (Chawla and Mujumdar, 2015) or the parameter sensitivity under a changing climate (Bennett et al., 2018).

- *Sensitivity to initial conditions*: the question about if the hydrologic predictions are affected by the hydrologic initial conditions (i.e. the initial moisture state at the snowpack and the soil profile) or if the boundary conditions constitute the main contributor to the model simulations was studied in detail in Cosgrove et al. (2003); Wood and Lettenmaier (2008); Li et al. (2009). It is known that the soil moisture content for the bottom soil layer of the VIC grid cells is the variable that commonly takes the longest time to reach the equilibrium, and although there is not a general agreement in how long the model spin-up period should be, since it highly depends on each particular application, it has been found that wetter states lead to faster spin-up times (Cosgrove et al., 2003; Melsen et al., 2016a). This issue is of relevance for a proper calibration and validation of the VIC model and is usually avoided by fixing a long-enough spin-up period previous to the simulation together with a wet initialization of the soil layers.

- *Sensitivity to soil and vegetation parameters*: since its generalization as a three-layer soil model (VIC-3L) in Liang et al. (1996) after its two-layer predecessor (VIC-2L, Liang et al., 1994), the sensitivity of the model to soil parameters has been studied at basin-scale and at global-scale through a cell-based approach. The basin-scale approach of Demaria et al. (2007) allowed to estimate the sensitivity of the simulated streamflows to the parameters that control the surface and subsurface runoff generation, and the global-scale study of Chaney et al. (2015) evaluated the efficiency of the VIC model for monitoring global floods and droughts under a parameter uncertainty framework. The sensitivity to land use changes and the vegetation parameters associated to the different vegetation classes (i.e. Leaf Area Index and albedo) have also been explored and expressed for the different components of the water and energy balances from the VIC

model (VanShaar et al., 2002; Chawla and Mujumdar, 2015; Bennett et al., 2018).

This work aims to contribute to the knowledge of the VIC model in a macroscale application over the headwater catchments of an important basin located in the north of the Iberian Peninsula, the Duero River Basin. For this end, the hydrologic modelling exercise here developed has been divided into three interrelated parts:

- The calibration of the VIC model for the selected catchments of the study area and the subsequent evaluation of its performance against a benchmark performance using the results of a well-calibrated model in Spain.
- An integrated sensitivity analysis for all the catchments focused on the soil parameters chosen for the calibration.
- A final assessment of equifinality and the efficiency of the calibration algorithm that links the calibration and sensitivity analysis results.

The Duero River Basin has been investigated in various previous studies and the main issues addressed are: the temporal trend of water supply and its relation to precipitation, temperature and plant cover changes (Ceballos-Barbancho et al., 2008); the hydrologic response to land-cover changes (Morán-Tejeda et al., 2010, 2012a), the impacts of different climate oscillations (Morán-Tejeda et al., 2010) and its response to the North Atlantic Oscillation (Morán-Tejeda et al., 2011a); the characteristics of the different existing river regimes (Morán-Tejeda et al., 2011b) and the effects of reservoirs on them (Morán-Tejeda et al., 2012b). All these studies were based on statistical analyses of different hydroclimatic and land-surface variables and contributed to a better understanding of the hydrologic behaviour of the Duero River Basin. The hydrologic modelling analysis carried out in this chapter can then provide an added value to this set of issues since the potentialities of a macroscale hydrologic model such as the VIC model have been examined in detail for this river.

In Sections 4.2 and 4.3 the study area and the methods are described. Section 4.4 gathers the results of the three-part approach and Section 4.5 corresponds to the discussion of the key results. Finally, the main conclusions of this chapter are provided in Section 4.6.

4.2 Study area

The Duero River Basin constitutes the largest basin of the Iberian Peninsula with a surface of 98073 km^2. It is a shared territory between Spain and Portugal, characterized by a high water contribution (~ 15000 hm^3/year). The study is focused on the Spanish part of the basin (Fig. 4.1), which represents the 80% of the area (78859 km^2). Most of this territory constitutes a plain

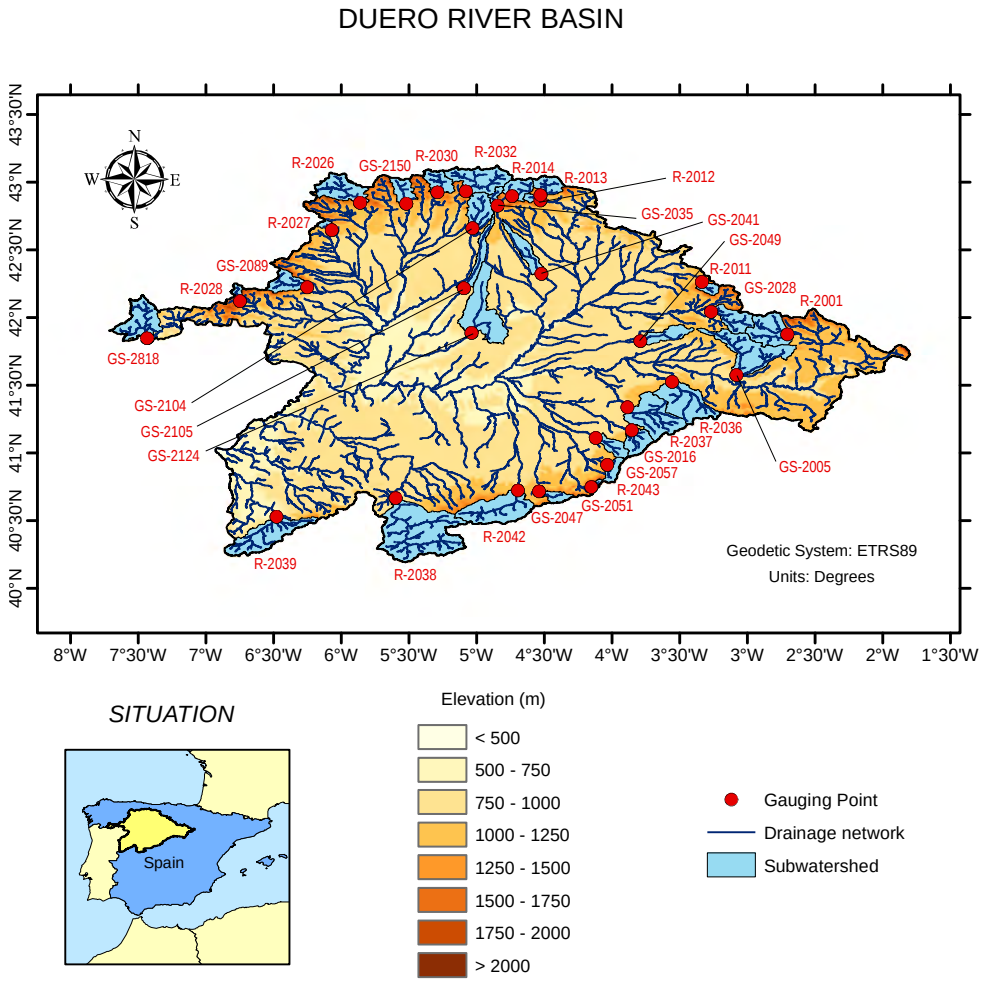


Fig. 4.1: Duero River Basin and the 31 studied catchments. The prefix "R-" denotes "Reservoir" and the prefix "GS-" denotes "Gauging Station".

surrounded by mountainous chains, thus configuring two topographic areas well differentiated. The large depression is filled with sediments of the Tertiary and the Quaternary, constituting a complex hydrogeologic environment. The lithology of the northern mountains consists of siliceous, calcareous and carbonated rocks with local small aquifers to the west part and aquifers of greater capacity to the east. The south system harbors rocks of low permeability and is dominated by a granite batholith. Finally, the eastern mountainous areas hold a silicic core enclosed by carbonated rocks with a high presence of karstic aquifers.

The basin presents a predominant Mediterranean climate with a mean annual precipitation volume of 50000 hm^3 which is mostly lost into the atmosphere through evaporative fluxes (~ 35000 hm^3/year). Most precipitation is concentrated in the mountainous areas reaching values above 1500 mm/year to the north of the basin and values slightly below 1000 mm/year to the south and east. As for the most part of the Iberian Peninsula, precipitation exhibits a very irregular intra-annual distribution, being concentrated in spring and fall and almost nonexistent during summer. Winter months are cold with a mean temperature of 2ºC in January, while summer is soft with maximum temperatures occurring in July (~ 20.5ºC). The Duero River Basin is regulated by a total of 31 reservoirs, and additional streamflow monitoring is also carried out in a large network of ~ 200 gauging stations.

4.3 Methods

4.3.1 Hydrologic dataset

The streamflow records were gathered in a monthly basis from the SAIH-ROEA dataset for 33 headwater catchments belonging to the Duero River Basin (see Section 2.1 in Chapter 2). An analysis of the percentage of gaps in the time series revealed that 31 out of the 33 headwater catchments presented less than 5% of missing values for the period Oct 2000 - Sep 2011, and were selected as the study catchments for the modelling exercise (Fig. 4.1). Table 4.1 collects the main characteristics of the 16 reservoirs and the 15 gauging stations included in the final selection: area (km^2), mean elevation (m), averaged annual precipitation (P_{an}, mm/year), potential evapotranspiration (PET_{an}, mm/year), and streamflow (Q_{an}, hm^3/year).

Table 4.1: Main characteristics of the 31 catchments.

Code	Name	Area (km^2)	Mean elevation (m)	$P_{\mathbf{an}}$ (mm/year)	$PET_{\mathbf{an}}$ (mm/year)	$Q_{\mathbf{an}}$ (hm^3/year)
R-2001	Cuerda del Pozo	546.7	1319	1122	900	188.7
R-2011	Arlanzón	106.7	1440	1293	798	77.6
R-2012	Cervera de Ruesga	54.3	1284	1005	859	86.1
R-2013	La Requejada	220.7	1378	1022	780	153.2
R-2014	Camporredondo	229.6	1673	1457	747	223.9
R-2026	Barrios de Luna	482.9	1496	1468	757	400.1
R-2027	Villameca	45.8	1180	946	941	34.1
R-2028	Moncabril (Sistema)	62.9	1712	1242	744	93.5
R-2030	Porma Juan Benet	250.3	1412	1251	753	304.5
R-2032	Riaño	592.3	1451	1569	748	623.7
R-2036	Linares del Arroyo	761.3	1111	515	1222	52.4
R-2037	Burgomillodo	803.1	1097	573	1232	87.0
R-2038	Santa Teresa	1845.4	1326	882	1235	734.8
R-2039	Águeda	788.4	895	1189	1333	392.7
R-2042	Castro de las Cogotas	853.3	1279	527	1231	89.0
R-2043	Pontón Alto	150.4	1582	990	993	78.3
GS-2005	Osma	893.1	1090	728	1084	121.0
GS-2016	Pajares de Pedraza	281.3	1298	634	1196	61.6
GS-2028	Salas de los Infantes	353.5	1257	988	926	104.7
GS-2035	Otero de Guardo	69.2	1492	1702	838	29.3
GS-2041	Villalcázar de Sirga	307.7	929	664	1011	35.5
GS-2047	Mediana de Voltoya	130.4	1347	518	1128	15.4
GS-2049	Cabañes de Esgueva	270.2	995	658	1046	24.7
GS-2051	El Espinar	36.7	1610	905	1032	18.8
GS-2057	Villovela de Pirón	202.0	1183	615	1212	33.0
GS-2089	Morla de la Valdería	281.1	1369	1001	945	146.1
GS-2104	Villaverde de Arcayos	371.1	1146	1041	942	135.3
GS-2105	Santervas de Campos	277.1	900	591	1107	28.8
GS-2124	Medina de Rioseco	908.4	790	450	1315	37.2
GS-2150	Pardavé	223.8	1448	1356	787	220.8
GS-2818	Rabal	597.9	678	1328	1146	305.7

Precipitation, maximum temperature and minimum temperature data were extracted from SPREAD and STEAD (Serrano-Notivoli et al., 2017, 2019, see Section 2.2 in Chapter 2). Atmospheric pressure, incoming shortwave radiation, incoming longwave radiation, vapor pressure and wind speed data were taken from daily outputs of high resolution (0.088º, ~ 10 km) simulations carried out with the Weather Research and Forecasting (WRF) model driven by the ERA-Interim Reanalysis data (WRFERA hereafter) for the spatial domain of the Iberian Peninsula (García-Valdecasas Ojeda et al., 2017). The WRFERA simulation for the Iberian Peninsula constitutes a suitable tool for studying the *AET* behaviour, and its performance was recently evaluated showing a good ability to represent the land-surface processes in this region (García-Valdecasas Ojeda et al., 2020a). The meteorological forcings were interpolated to the VIC grid cells for the 31 studied catchments (Fig. 4.2) following a nearest neighbour assignment.

4.3.2 Parameter calibration

The model was calibrated for the period from October 2000 to September 2009 choosing the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) as the objective function. The NSE was calculated by comparing the monthly observations of streamflow with the monthly aggregated total runoff simulated by the VIC

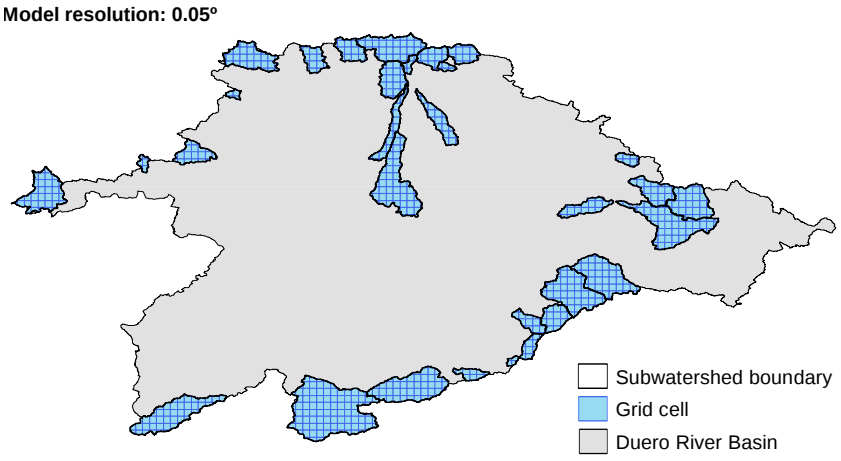


Fig. 4.2: VIC model implementation.

model. The calibration was carried out for the five soil parameters indicated in Table 4.2 using the Shuffled-Complex-Evolution Algorithm (SCE-UA) of Duan et al. (1994). A spin-up period of ten years previous to the calibration period was simulated in order to ensure that the soil moisture content of the three soil layers reached an equilibrium, and therefore the initial conditions did not affect the calibration process.

4.3.3 Model evaluation

The temporal evaluation of the VIC model capability to appropriately simulate the streamflow was carried out by calculating four skill measures commonly selected for this end (Mizukami et al., 2017; Rakovec et al., 2019; Yang et al., 2019): NSE, the coefficient of correlation (r), and two ratios here called $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$. NSE evaluates the predictive skill of the VIC model comparing the monthly observations with the monthly simulations, and was chosen as the objective function for the calibration exercise. r measures the degree of agreement between the dynamics of the simulated and observed time series. $alpha_{\mathrm{d}}$ is the ratio between the standard deviation of the simulations and the standard deviation of the observations. $alpha_{\mathrm{m}}$ is the ratio of the mean of the simulations to the mean of the observations.

The four skill metrics were calculated for the calibration period and were validated for the period October 2009 - September 2011 (validation period) in order to evaluate the model predictive skill outside the calibration years.

Table 4.2: Selected parameters for the calibration.

Parameter	Units	Lower bound	Upper bound	Description
b_i	-	10^{-5}	0.4	Infiltration shape parameter (see Eq. 3.1 in Chapter 3)
D_S	-	10^{-9}	1	Fraction of D_m where non-linear baseflow starts (see Eq. 3.2 in Chapter 3)
W_S	-	10^{-9}	1	Fraction of the porosity of soil layer 3 where non-linear baseflow starts (see Eq. 3.2 in Chapter 3)
D_m	mm/day	10^{-9}	30	Maximum baseflow (see Eq. 3.2 in Chapter 3)
d_2	m	0.1	0.9	Thickness of soil layer 2

The model performance was compared to a benchmark performance based on the streamflow simulations carried out with the Integrated System for Rainfall-Runoff Modelling (SIMPA) model (Estrela and Quintas, 1996; Alvarez et al., 2005) for the domain of Spain. SIMPA simulations are periodically updated and used among the different water districts of Spain as a tool for water planning and water resources management purposes[3].

The spatial assessment of the VIC model streamflow simulations was done through a cross-evaluation exercise that aimed to analyze the spatial transferability of the five soil parameters chosen for the calibration. The VIC model was run for each catchment using the calibrated parameters of the remaining 30 catchments, and the NSE was then calculated for the different model runs. The spatial evaluation was done for the complete study period (calibration + validation period).

Finally, the VIC model performance was also evaluated for the actual evapotranspiration (AET) simulations comparing the monthly outputs for each catchment with two AET products: GLEAM version 3.3a (Martens et al., 2017; Miralles et al., 2011, see Section 2.3 in Chapter 2), and the WRFERA simulation carried out in García-Valdecasas Ojeda et al. (2017). The monthly AET data for both products were spatially aggregated following a similar aggregation method to that applied to the simulated total runoff. The same four skill metrics posed for the streamflow were calculated for the AET in the calibration + validation period, and the SIMPA model simulations were again taken as the benchmark performance.

4.3.4 Sensitivity analysis

The Standardized Regression Coefficients (SRC) method (Saltelli et al., 2008) aims to study the propagation of uncertainty from model inputs to outputs. The SRC method is focused on the behaviour of the model outputs in relation to a certain set of parameters once the boundary conditions (i.e. the meteorological forcings) and the initial conditions (i.e. the soil moisture content of the three soil layers) have been fixed. This sensitivity analysis method requires two elements: first, a Monte Carlo simulation where the model is run with a specified number

[3]https://www.miteco.gob.es/en/agua/temas/evaluacion-de-los-recursos-hidricos/evaluacion-recursos-hidricos-regimen-natural/

of parameter samples; and second, a multiple linear regression of each model output of interest as a linear function of the parameters.

The sensitivity analysis was carried out for each catchment considered in the study area. As in the calibration process, the period October 2000 - September 2009 was chosen and ten years of spin-up prior to the study period were run.

Monte Carlo simulation

A parametric space is defined through the selection of several parameters and their upper and lower bounds. Here, a 5-dimensional parametric space was established for the five calibration parameters and their corresponding upper and lower bounds. A sampling method is then applied to the parametric space, extracting a large-enough sample of parameters for the Monte Carlo simulation. The Latin Hypercube Sampling (LHS) method (Iman and Conover, 1982) was applied for this step extracting a total of 10000 random samples. This process allowed to define a sampling matrix, $\mathbf{\Theta}$, of order m x n, where m represents the number of samples (m = 10000) and n the number of parameters for the analysis (n = 5). The model was finally run for each parameter combination (i.e. row) of $\mathbf{\Theta}$.

The Monte Carlo simulation was also used for assessing equifinality and the efficiency of the calibration algorithm by studying the response given by each parameter sample in terms of the NSE. The results were compared with the NSE determined during the calibration period.

Multiple linear regression

The outputs of interest from the VIC model were those components included in the water balance: surface runoff (Q_s), baseflow (Q_b), total runoff (Q_t), actual evapotranspiration (AET) and the soil moisture content of the three soil layers (SM_1, SM_2 and SM_3). For each component and for each run of the Monte Carlo simulation, the mean value of the simulated series was calculated, and a multiple linear regression model was then adjusted relating the mean values of each component with the sampling parameters:

$$\mathbf{y} = a_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \sum_{i=1}^{n} a_{\mathrm{i}} \mathbf{\Theta}_{\mathrm{i}} \tag{4.1}$$

Where $\mathbf{y}$ is a column vector with the m mean values of the component, a_0 is the intercept of the hyperplane, a_{i} is the regression coefficient of the parameter i and $\mathbf{\Theta}_{\mathrm{i}}$ is the column of the sampling matrix corresponding to the parameter i. The standardized regression coefficients, β_{i}, are then calculated for each parameter:

$$\beta_{\mathrm{i}} = \frac{\sigma_{\Theta_{\mathrm{i}}}}{\sigma_{y_{\mathrm{p}}}} a_{\mathrm{i}} \tag{4.2}$$

Here, $\sigma_{\Theta_{\mathrm{i}}}$ and $\sigma_{y_{\mathrm{p}}}$ are the standard deviations of $\mathbf{\Theta}_{\mathrm{i}}$ and the predicted values of $\mathbf{y}$, respectively. ${\beta_{\mathrm{i}}}^2$ represents the relative contribution of the parameter i to the variance of the model output of interest, being $\sum_{i=1}^{n} {\beta_{\mathrm{i}}}^2 \leq 1$ and equal to the coefficient of determination r^2 of the adjustment. A threshold of $r^2 \geq 0.7$ is usually defined for assuming that the fitted model has a good linear behaviour, and therefore the coefficients β_{i} are valid measures of the sensitivity (Saltelli et al., 2006), although they can be robust and reliable measures even for nonlinear models (Saltelli et al., 2008). β_{i} can take values between -1 and 1. A high absolute value of β_{i} implies that the component is sensitive to the parameter and its sign indicates whether the effect is positive or negative.

4.4 Results

4.4.1 Calibration results

The values of the four skill metrics for the calibration and validation periods are shown in Table 4.3. Fig. 4.3 depicts the simulated streamflows during both periods together with the observed streamflows for six selected catchments (R-2011, R-2037, R-2038, GS-2005, GS-2089 and GS-2150) located in different parts of the basin. The NSE for the calibration period presents values above 0.75 in 19 out of the 31 catchments and reaches values above 0.85 in 10 catchments, and the corresponding r values are high too. For the validation period, both NSE and r values are predominantly high, and generally lower than the

Table 4.3: Values of the four skill metrics for the calibration and validation periods.

Code	NSE_{cal}	r_{cal}	$alpha_{d\text{-}cal}$	$alpha_{m\text{-}cal}$	NSE_{val}	r_{val}	$alpha_{d\text{-}val}$	$alpha_{m\text{-}val}$
R-2001	0.8520	0.9494	1.0968	1.2299	0.4813	0.8006	1.0882	1.2763
R-2011	0.9237	0.9696	1.0798	1.0731	0.7606	0.8947	1.0328	0.8599
R-2012	0.2723	0.8983	0.3737	0.3895	0.3758	0.8733	0.5686	0.5270
R-2013	0.8263	0.9263	0.7619	0.9274	0.8141	0.9152	0.7636	1.0211
R-2014	0.8777	0.9370	0.9422	1.0152	0.8736	0.9347	0.9382	1.0076
R-2026	0.8865	0.9645	1.1449	1.1055	0.7904	0.8958	1.0049	1.0082
R-2027	0.8254	0.9477	0.7769	0.7314	0.7003	0.9278	0.6790	0.6937
R-2028	0.7475	0.8812	0.9708	0.8514	-0.3382	0.8078	0.2941	0.2746
R-2030	0.6478	0.8774	0.6893	0.7401	0.5780	0.8730	0.7258	0.7209
R-2032	0.9409	0.9705	0.9618	0.9707	0.8354	0.9220	0.9881	0.9222
R-2036	0.6662	0.8217	0.8436	1.1285	0.8070	0.9177	0.7531	0.9172
R-2037	0.8222	0.9114	0.9971	0.9726	0.3201	0.7912	1.3354	0.9508
R-2038	0.9145	0.9587	0.9241	1.0785	0.8301	0.9188	0.9737	0.8973
R-2039	0.9521	0.9760	0.9766	1.0413	0.2494	0.8033	1.3851	1.2843
R-2042	0.8903	0.9500	1.0336	1.1031	0.8309	0.9559	1.1520	1.2374
R-2043	0.8322	0.9237	1.0287	1.0923	0.8960	0.9476	0.9033	0.9925
GS-2005	0.8307	0.9156	1.0006	1.0305	0.5003	0.8161	0.6139	0.6316
GS-2016	0.8350	0.9173	0.8664	1.0787	0.7798	0.9126	0.7777	0.8281
GS-2028	0.6290	0.8320	1.0762	1.0830	0.6980	0.8465	0.7543	0.8810
GS-2035	0.5047	0.8850	0.9824	2.1179	-2.1931	0.7974	1.4392	3.1461
GS-2041	0.7106	0.8497	0.9556	1.0304	0.1784	0.8344	0.3184	0.3503
GS-2047	0.6277	0.8148	0.9591	1.2317	0.4813	0.7479	1.0270	1.0153
GS-2049	0.7333	0.8738	1.0208	1.1350	-0.0513	0.8926	1.5098	1.5572
GS-2051	0.7743	0.8909	0.7924	0.8814	0.7591	0.9226	0.6616	0.8019
GS-2057	0.7230	0.8864	1.1062	1.1746	0.6515	0.9176	0.7842	0.6004
GS-2089	0.9116	0.9564	1.0038	1.0377	0.9026	0.9561	0.8707	0.9415
GS-2104	0.8137	0.9208	1.0955	1.0734	0.6694	0.8985	0.5954	0.7873
GS-2105	0.6204	0.8032	0.9603	1.0057	0.3723	0.8010	0.5803	0.6466
GS-2124	0.7397	0.8604	0.8403	0.9855	0.3129	0.8246	1.3097	0.8383
GS-2150	0.7909	0.9118	0.8582	0.7828	0.8507	0.9436	1.1388	0.9702
GS-2818	0.8724	0.9403	1.0319	1.0953	0.8975	0.9720	0.8782	0.7920

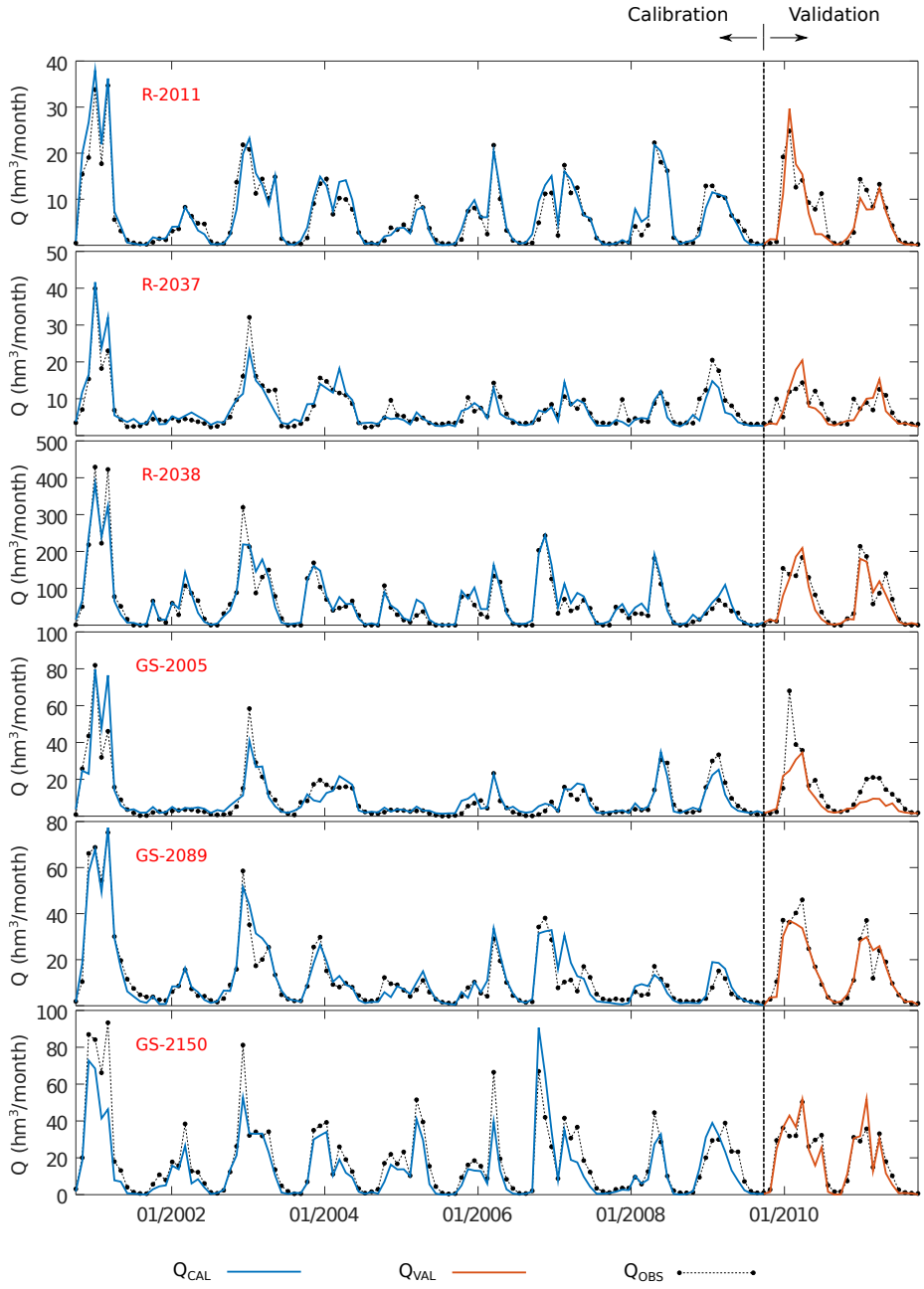


Fig. 4.3: Time series of the observed streamflows along with the simulated ones for the calibration and the validation periods for six example catchments.

corresponding ones for the calibration process, although minimum values of NSE below 0 were attained for 3 catchments. Note that the results of the calibration and the validation processes are slightly better for the reservoirs, what could indicate a quality difference between the streamflow databases from the reservoirs and from gauging stations.

Some of the high r values are obtained for low NSE estimations, which indicates that the model is able to capture the intra-annual variability of the streamflow observations but is not able to reach a good fit for the peaks of streamflow (Table 4.3, Fig. 4.3). For instance, stations R-2012 and GS-2035 present NSE values for the calibration period of 0.2723 and 0.5047, respectively, being the corresponding r values high and equal to 0.8983 and 0.8850, respectively. This can be better understood analyzing the values of $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$ in such cases. The values of $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$ are below 0.4 for station R-2012, and this evidences a clear underestimation of both the variability of the streamflow (i.e. low $alpha_{\mathrm{d}}$) and its total volume during the period (i.e. low $alpha_{\mathrm{m}}$). In the case of station GS-2035, the variability of the simulated streamflow is almost identical to the observations ($alpha_{\mathrm{d}} \sim 1$), while the streamflow volume is greatly overestimated ($alpha_{\mathrm{m}} \sim 2$). It is interesting to note that high NSE values were obtained for catchments with varying sizes, with good fits for both small-sized (e.g. R-2011 and GS-2089) and medium-sized (e.g. R-2038 and GS-2005) catchments, emphasizing the ability of the VIC model to provide accurate predictions of the streamflow across different spatial scales.

Fig. 4.4 shows the cumulative distribution functions (CDFs) of the four skill metrics together with the benchmark performance (i.e. SIMPA model performance) for the calibration and validation periods separately. The NSE and r values (Fig. 4.4a, b) are notably higher for the VIC model both in calibration and validation conditions, being their corresponding CDFs steeper and closer to the upper bounds of NSE and r. The CDFs for $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$ (Fig 4.4c, d) show that most of the catchments present values close to 1 for the VIC model in the calibration period, being its performance slightly deteriorated in the validation years. The simulated streamflow variability and volume are generally lower than the observations for the two model structures. However, the streamflow variability becomes markedly overestimated by SIMPA for cumulative probabilities superior to 0.8 and reaches $alpha_{\mathrm{d}}$ values above 1.5. The overestimation is also evident for the streamflow volume, and in this case the presence of values of $alpha_{\mathrm{m}}$ above 1.5 is noticeable both for VIC and SIMPA.

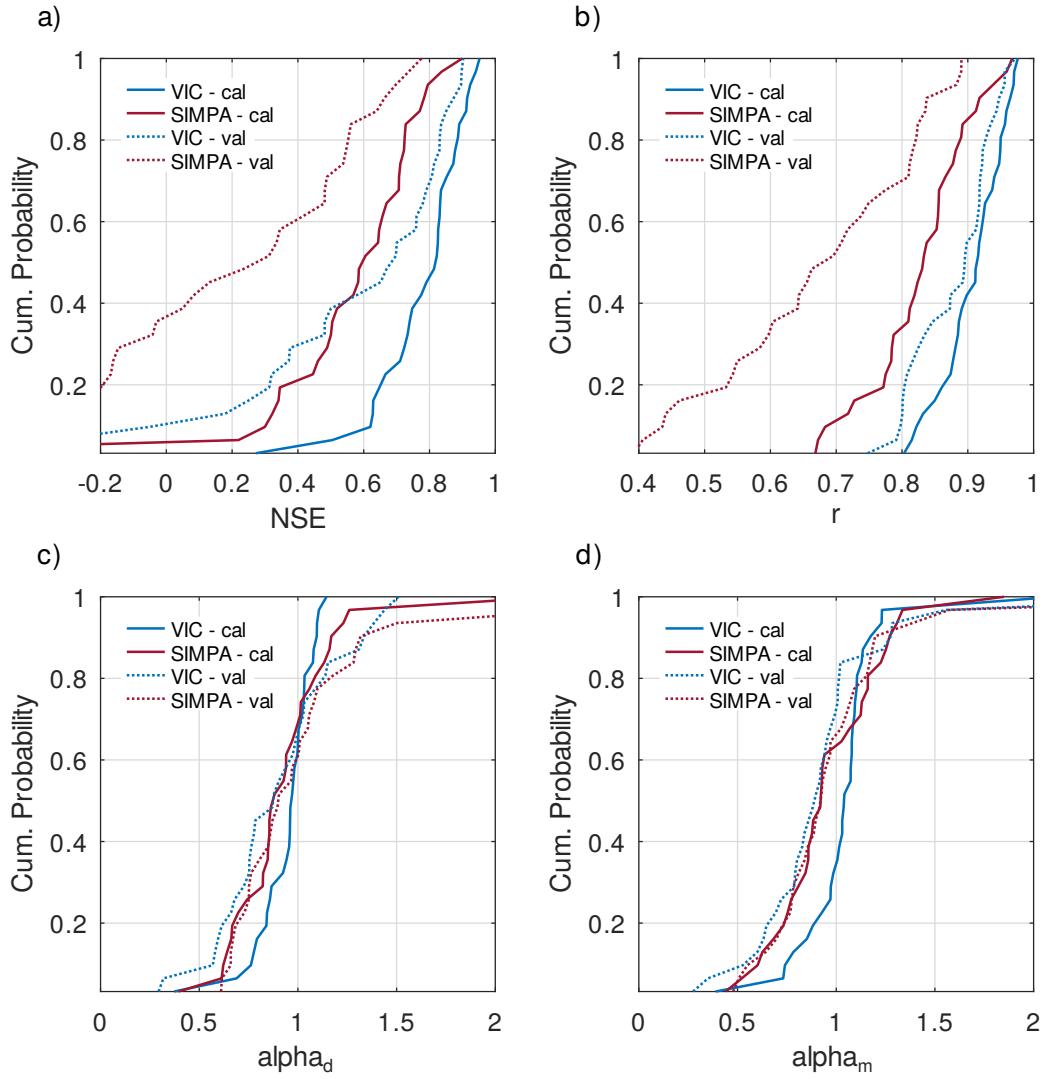


Fig. 4.4: CDFs of the four skill metrics for the streamflow simulations: a) NSE; b) r; c) $alpha_{\mathrm{d}}$, and d) $alpha_{\mathrm{m}}$. Blue lines represent the VIC model performance and red lines represent the SIMPA model (benchmark) performance. Straight lines correspond to the calibration period and dotted lines correspond to the validation period.

4.4.2 Cross-evaluation of the calibrated parameters

Fig. 4.5a depicts the spatial distribution of the optimal NSE values reached for the calibration + validation period, presenting values above 0.8 in 14 out of the 31 catchments and values between 0.6 and 0.8 in 12 of them. There

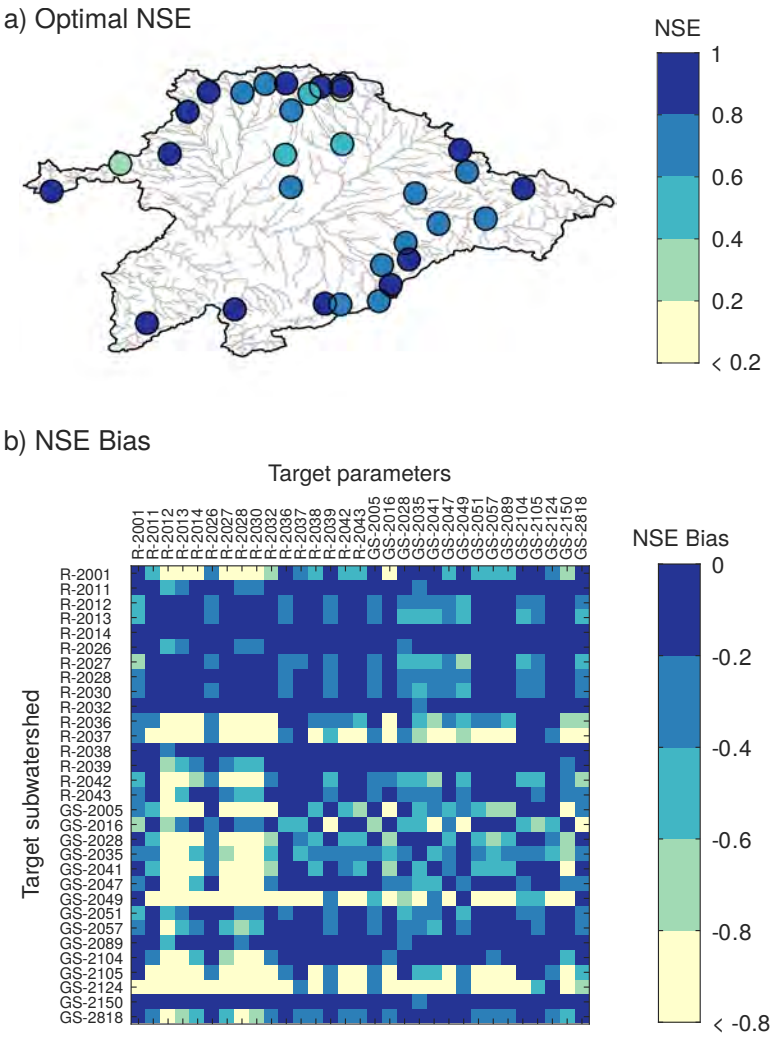


Fig. 4.5: Cross-evaluation of the calibrated parameters for the 31 catchments: a) spatial distribution of the optimal NSE values, and b) cross-performance of the different parameter combinations for each catchment.

is a spatial pattern for the optimal NSE distribution, and most of the lowest values were obtained for the catchments located in the headwater areas of the central depression of the basin, being mainly associated to gauging stations (e.g. GS-2041 and GS-2105).

Fig. 4.5b shows the distribution of the differences between the NSE value calculated for each experiment of the cross-evaluation exercise and the optimal

NSE corresponding to each catchment (i.e. NSE bias). The results can be interpreted in two different directions: firstly, each row reflects the degree of receptivity of a given catchment to the calibrated parameters obtained for every catchment including itself. The presence of darker blue tones indicates a greater parameter receptivity for the target catchment (i.e. NSE biases close to 0). Secondly, each column represents the degree of transferability of the different calibrated parameters when they are used to run the VIC model for all the catchments of the study area. In this case the presence of darker blue tones indicates a greater transferability for the target parameter combination. The maximum degrees of receptivity and transferability correspond to a null NSE bias and are obtained when the target catchment and the target parameters coincide (i.e. main diagonal of the pixel plot).

The horizontal dimension of the pixel plot reveals a predominance of low NSE bias sequences, and therefore many of the catchments present an elevated degree of receptivity to different parameter combinations. The greatest degrees of receptivity were independent from the model performance for the streamflow simulations and were obtained for catchments with both high optimal NSE estimations (e.g. R-2011 and GS-2089) and low optimal NSE values (e.g. R-2012 and GS-2057). The lowest degrees of receptivity followed a similar behaviour but were much less abundant (e.g. R-2001 and GS-2105), The vertical dimension confirms a prevalence of low NSE bias sequences as well, suggesting a high degree of transferability for the majority of the calibrated parameters. In this case, the lowest parameter transferability occurs for stations R-2012 to R-2014 and for stations R-2027 to R-2032, all of them located in the northern mountains of the Duero River Basin (see Fig. 4.1).

4.4.3 Model performance for the AET simulations

Fig. 4.6 evaluates the AET predictability comparing VIC and the benchmark AET simulations against the GLEAM and WRFERA datasets. There is a large gap between the performance of VIC and SIMPA concerning the NSE metric (Fig. 4.6a), yielding a median NSE value around 0.5 in the comparison of VIC with both AET products, and a slower median NSE value of approximately 0.2 in the case of SIMPA. Each model structure reflects similar performances against GLEAM and WRFERA when evaluated separately, although VIC produces higher NSE values for cumulative probabilities above 0.5 in the comparison with WRFERA.

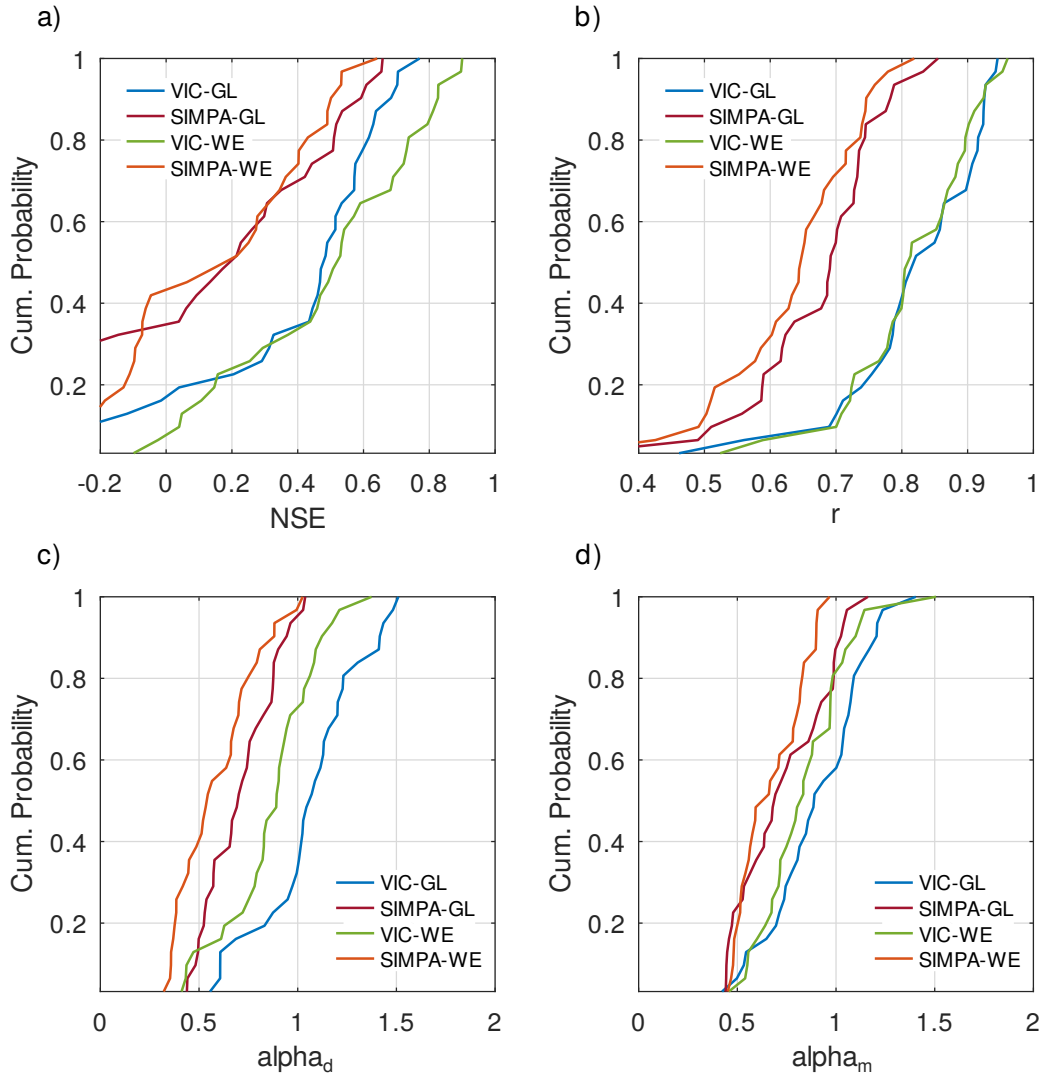


Fig. 4.6: CDFs of the four skill metrics for the AET simulations: a) NSE; b) r; c) $alpha_{\mathrm{d}}$, and d) $alpha_{\mathrm{m}}$. Blue and green lines represent the VIC model performance against GLEAM and WRFERA, respectively. Red and orange lines represent the SIMPA model (benchmark) performance against GLEAM and WRFERA, respectively.

A comparable behaviour is observed in relation to r. The AET dynamics are better captured by VIC with 60% of the population presenting values above 0.8 for both GLEAM and WRFERA, and being the performances against them almost identical (Fig. 4.6b). The SIMPA model produces r values slightly higher when compared against GLEAM and hardly reaches a value of 0.8 for

very few catchments.

Lastly, there is a generalized underestimation of the AET variability and the total volume for both model structures, with $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$ values predominantly below 1 and markedly slower for SIMPA (Fig. 4.6c, d). The wide distance between the CDFs corresponding to $alpha_{\mathrm{d}}$ supposes that the different performances can be better defined and distinguished for this skill metric. The greatest closeness to 1 is observed for VIC, and while the variability is underestimated when compared to WRFERA, it turns out to be particularly overestimated in comparison to GLEAM. Concerning $alpha_{\mathrm{m}}$, the spacing between the different CDFs becomes narrower and the underestimation of the mean AET is also noticeable in the VIC model comparison to GLEAM. The greatest closeness to 1 now corresponds to the VIC model performance against GLEAM, thus producing more similar AET volumes.

4.4.4 Integrated sensitivity analysis

Through the application of the SRC method the β coefficients for the five calibration parameters of the water balance components in the VIC model were obtained, and the results are shown in Fig. 4.7 for all the catchments. The r^2 value obtained from the multiple linear regression and the estimation of r^2 as the sum of the squares of β coefficients are also depicted in Fig. 4.7 (h, i), reflecting very similar values. The results of the sensitivity analysis for the selected components to the parameters are given below in a component-by-component basis providing the necessary explanations when there is a strong dependency between them:

- Q_{s}: the values of r^2 are above 0.7 for all the catchments (Fig. 4.7), fulfilling the criterion of enough linearity for interpreting the results of the sensitivity analysis. The strongest positive effect corresponds to the parameter b_{i}, which means that a higher value of b_{i} leads to more surface runoff. This is clearly evidenced in Eq. 3.1 in Chapter 3, where a relation of exponential type between Q_{s} and b_{i} is established. D_{m} produces a negative effect on the surface runoff, suggesting that an increase of the maximum baseflow brings a reduction of the surface component under the assumption of the same meteorological forcings. d_2 also yields a negative effect on Q_{s}, and this effect is related to an increase in AET.

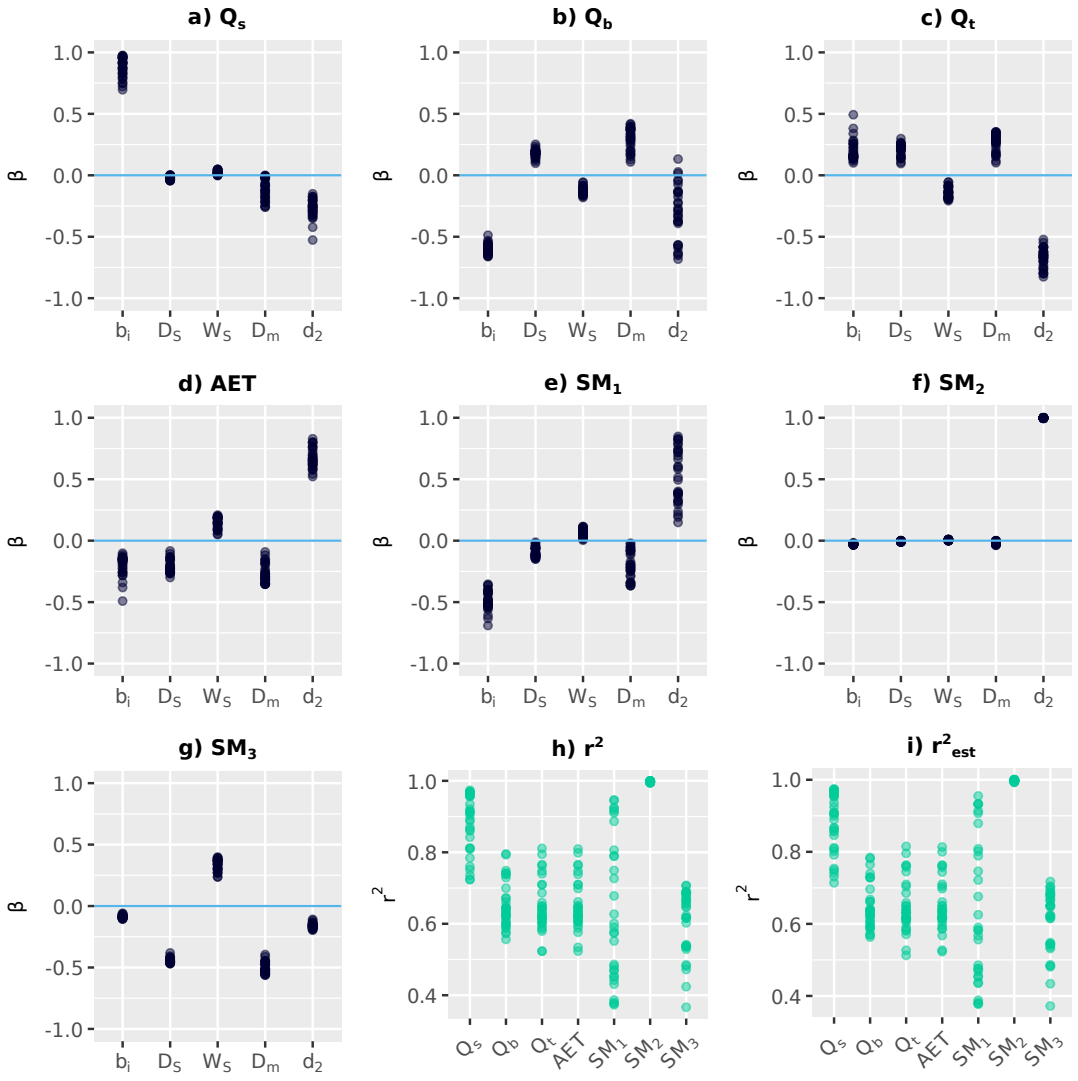


Fig. 4.7: (a to g) β coefficients for the 31 catchments. h) r^2 value from the multiple linear regression. i) r^2 estimated from the β coefficients.

- Q_b: the values of r^2 mostly range from 0.5 to 0.7, with some values close to 0.8 (Fig. 4.7). In this case the linearity criterion is hardly reached and therefore it is difficult to interpret the β coefficients. However, it is interesting to note that, with the exception of d_2, the β coefficients of the parameters are characterized by a low dispersion. This is an indicator of

the robustness of the VIC model response, and although the threshold of linearity is not always achieved, the dependency of Q_b with respect to these parameters can be accepted. As expected from the previous analysis of the surface component of the runoff, b_i reflects a strong negative effect. The positive effects now correspond to D_S and D_m. This is obvious in the case of D_m but not so evident for D_S since a higher value of D_S only means that the baseflow law tends to be more linear (see Eq. 3.2 in Chapter 3). The amplitude of the β coefficients for d_2 is broader than for the rest of the parameters but always negative except for two catchments.

- Q_t: the total runoff exhibits an additive effect of the previous components for both r^2 and the β coefficients as it is computed through the sum of the surface runoff and the baseflow (Fig. 4.7). Thus, higher values of b_i, D_S and D_m lead to an increase of Q_t and higher values of d_2 produce a negative effect on Q_t due to a rise of AET. This component is of particular interest given that it is the component subject to calibration in this work. In order to provide a better understanding on its behaviour the spaghetti plots of the Monte Carlo simulation for an example catchment (R-2038) are depicted in Fig. 4.8. As shown there, the time series of the observed streamflow (Fig. 4.8c) falls into the range of responses of the model for almost all the study period and therefore one or more sets of parameters will afford a god fit with the observations.

- AET: this component and Q_t are linked through the law of conservation of mass applied to the system defined by each catchment, being the precipitation equal to the sum of Q_t, AET and the variation of the storage in the hydrologic system. Moreover, the study period for the sensitivity analysis is long-enough to neglect the last term of the water balance equation, and the precipitation is fixed for each catchment as a boundary condition. In consequence, the linearity of both Q_t and AET with respect to the parameters must be similar and the β coefficients for AET are essentially identical to the corresponding ones for Q_t but with opposite signs (Fig. 4.7). Fig. 4.8d shows the spaghetti plots of this component together with the potential evapotranspiration (PET) profile. The reason of the existence of some values of AET above the PET curve responds to the internal handling of the Penman-Monteith equation used in the VIC model because various different approaches are considered when computing the potential evapotranspiration, and the curve presented in Fig. 4.8d corresponds to the current vegetation parameters.

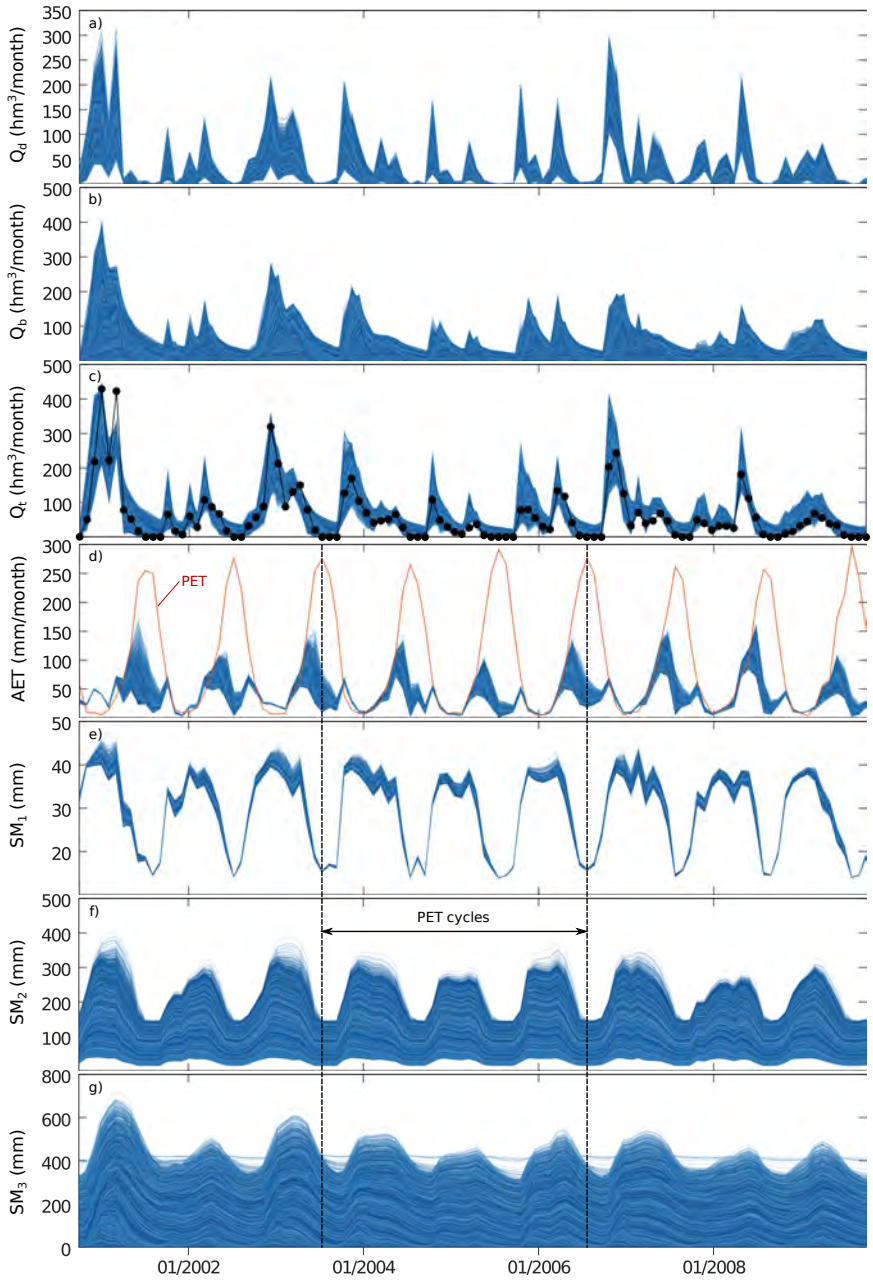


Fig. 4.8: Spaghetti plots of the water balance components resulting from the Monte Carlo simulation for the catchment R-2038.

- SM_1: the values of r^2 are widely scattered and range from 0.35 to values above 0.9 (Fig. 4.7). The nature of such a scattered distribution may be

an outcome of the closeness between the soil moisture profiles in this layer, making it difficult to adjust a multiple linear regression model to its mean values. Similarly to the case of Q_b, most of the β coefficients present a relatively low dispersion and subsequently the results of the sensitivity analysis can be interpreted. The negative effects mainly concern to b_i and D_m, demonstrating that a higher exponent in the surface runoff equation and a higher maximum baseflow are related to lower soil moisture values for the upper soil layer. On the other hand, the positive effects are associated with increasing values of d_2 despite of revealing highly dispersed β coefficients. The spaghetti plots of SM_1 (Fig. 4.8e) seem to reproduce the PET cycles and this results from the evaporative fluxes themselves as the transpiration process occurs from the roots of the vegetation.

- SM_2: it is the component with the highest linearity with regard to the calibration parameters, proffering values of r^2 very close to 1 for all the catchments (Fig. 4.7). In this case d_2 dominates the sensitivity of SM_2 with a noticeable positive effect (i.e. values of β near to 1). The PET cycles are also markedly represented in the soil moisture profiles of this layer (Fig. 4.8f).

- SM_3: even though the values of r^2 lie between 0.6 and 0.7 predominantly, some of them fall below 0.6 with minimum values close to 0.4 (Fig. 4.7). Once again the dispersion of the β coefficients is relatively low and in this occasion this is also true for d_2. Baseflow takes place from this layer and this is reflected in the β coefficients corresponding to D_S, W_S and D_m, which present opposite signs and similar absolute values to the calculated ones for Q_b. As for the previous soil layers, the PET cycles are present too but here there is a lag in the valleys of the soil moisture profiles due to the delay in the baseflow generation process (Fig. 4.8g).

4.4.5 Equifinallity and efficiency of the calibration algorithm

Equifinality and the efficiency of the calibration algorithm were assessed through the evaluation of the NSE values for the Monte Carlo simulations of all the catchments by comparing the total runoff of each simulation with the observed streamflow during the calibration period. For this purpose, two counts of the number of simulations satisfying certain criteria were carried out: first, the number of simulations for each catchment presenting NSE values

above the NSE determined during the calibration (NSE_{cal}) minus 0.05 was used as indicator of equifinality of the VIC model and the parameter samples; and second, the number of simulations with NSE values above NSE_{cal} hinted at the efficiency of the SCE-UA algorithm in finding the optimal set of parameters producing the best fit with the streamflow records. The results of this exercise are expressed in Table 4.4.

It is clear that for the majority of the catchments there are many simulations with NSE values very close to the optimal model, and in some cases the number of simulations is fairly high (> 3000). This can be also appreciated when the columns of the sampling matrix are plotted against the NSE of each simulation in a "dotty plot". Fig. 4.9 shows the dotty plots of the five parameters of the calibration for two catchments (R-2038 and GS-2089) as an example of this analysis. For both catchments the NSE values of the simulations are above 0 and the points clouds are concentrated on the top of the diagrams, suggesting that a high number of them are close to the optimal fit (see also Table 4.4). The shape of the dotty plots supplies useful information about the behaviour of the parameter samples as a set. For example, the dotty plot of d_2 for the catchment R-2038 reflects a trend to produce high NSE values when the parameter values are near to the upper bound. The optimum was reached for $d_2 = 0.8995$ m, while the rest of the fitted parameters were located between the fixed limits. Also, most of the NSE values were below NSE_{cal} when the second count was executed, implying that the SCE-UA algorithm was highly efficient in searching for the optimal set of parameters.

4.5 Discussion

4.5.1 VIC model performance for the streamflow simulations

The results of the calibration and the validation suggest that the macroscale application of the VIC model carried out in this chapter performs well for a large number of catchments in the Duero River Basin. The VIC model showed a considerably better performance in the prediction of the streamflow in comparison to the SIMPA model according to the four skill metrics calculated. The VIC performance for the streamflow simulations in the calibration and validation periods is also comparable to other studies using hydrologic models

Table 4.4: Behaviour of NSE in the Monte Carlo simulations for assessing equifinality and the efficiency of the calibration algorithm.

Code	$\mathbf{NSE_{cal}}$	$\mathbf{NSE > NSE_{cal} - 0.05}$	$\mathbf{NSE > NSE_{cal}}$
R-2001	0.8520	96	0
R-2011	0.9237	1970	5
R-2012	0.2723	426	0
R-2013	0.8263	1259	0
R-2014	0.8777	3463	0
R-2026	0.8865	2804	8
R-2027	0.8254	1863	56
R-2028	0.7475	1364	2
R-2030	0.6478	1132	0
R-2032	0.9409	3568	0
R-2036	0.6662	10	3
R-2037	0.8222	0	0
R-2038	0.9145	2808	0
R-2039	0.9521	227	0
R-2042	0.8903	368	0
R-2043	0.8322	1660	5
GS-2005	0.8307	32	5
GS-2016	0.8350	1758	0
GS-2028	0.6290	103	0
GS-2035	0.5047	373	0
GS-2041	0.7106	0	0
GS-2047	0.6277	267	0
GS-2049	0.7333	0	0
GS-2051	0.7743	2446	0
GS-2057	0.7230	518	3
GS-2089	0.9116	3578	3
GS-2104	0.8137	93	0
GS-2105	0.6204	8	0
GS-2124	0.7397	1	0
GS-2150	0.7909	3484	0
GS-2818	0.8724	26	0

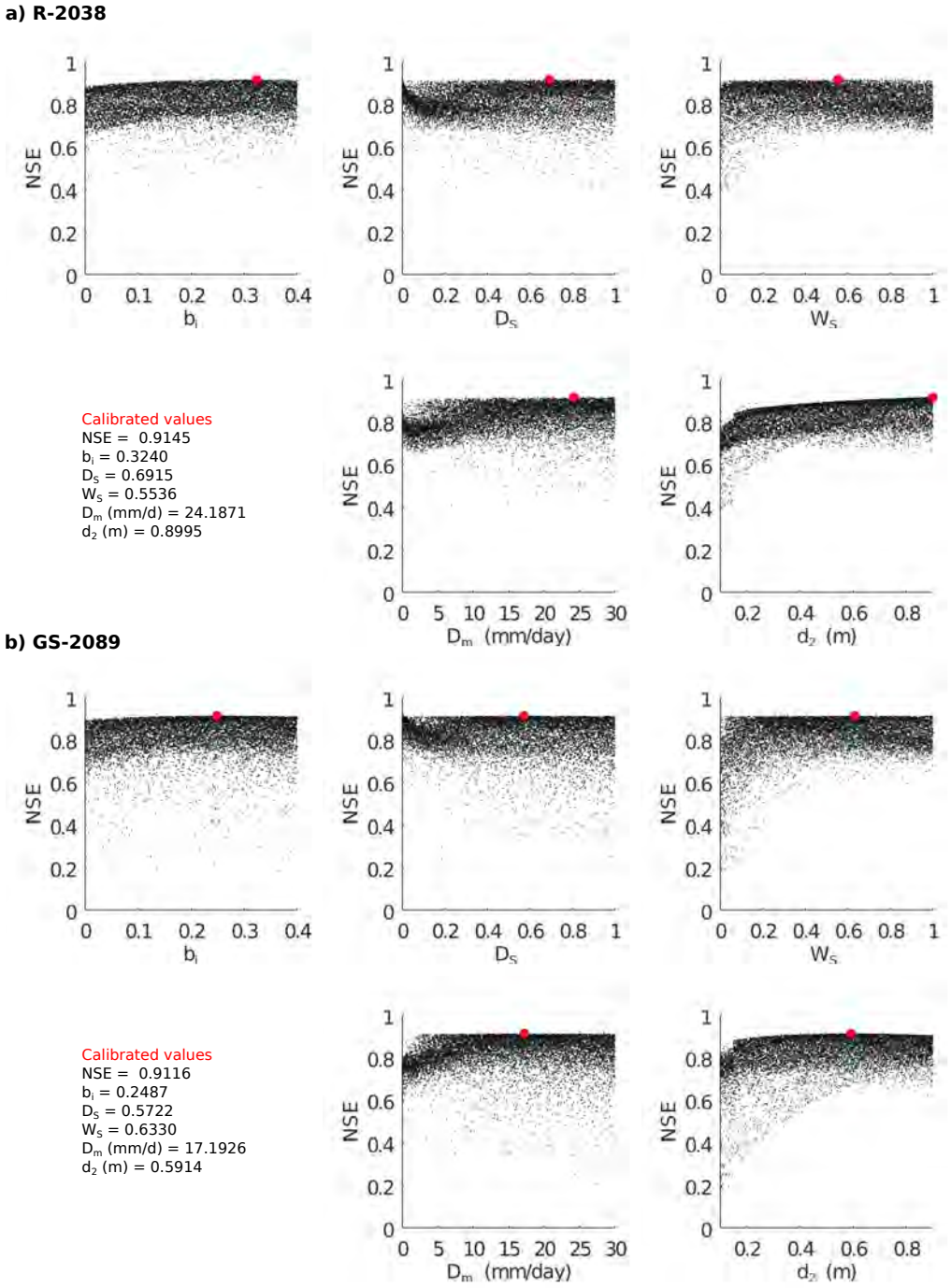


Fig. 4.9: Dotty plots for two catchments: a) R-2038 and b) GS-2089. Red dot corresponds to the calibrated value for the corresponding parameter using the SCE-UA algorithm.

developed in northern Spain (Morán-Tejeda et al., 2014) and recently in the south of Spain (Pellicer-Martínez and Martínez-Paz, 2018; Yeste et al., 2018).

Other studies also aimed at the calibration and the evaluation of the VIC model performance over the CONUS domain (Mizukami et al., 2017; Rakovec et al., 2019; Yang et al., 2019) applying different techniques for the estimation of spatially distributed parameters. The application of the MPR method (Samaniego et al., 2010) to the VIC parameter estimations (Mizukami et al., 2017; Rakovec et al., 2019) and the regionalization of various key streamflow characteristics (Yang et al., 2019) have shown to be consistent manners for estimating parameters both in gauged and ungauged basins. This supposes a clear gain of information with respect to the calibration of individual basins as the spatial discontinuities and limitations inherent to a basin-by-basin approach are effectively avoided. However, this is always achieved at the expense of a large loss of accuracy when compared to the individual basin calibration (Mizukami et al., 2017; Rakovec et al., 2019).

The CDFs corresponding to the skill metrics (Fig 4.4) showed higher NSE and r values than those obtained for the catchments of the CONUS domain, as well as $alpha_{\mathrm{d}}$ and $alpha_{\mathrm{m}}$ values closer to 1. This is also noticeable for the individual basin calibration (Mizukami et al., 2017; Rakovec et al., 2019). The large gap between both VIC performances could be connected with the selected calibration time step, and while the model was calibrated in a monthly basis in this work, the other studies implemented a daily calibration.

This is supported by the results of the VIC performance obtained in Yang et al. (2019) for five large basins, being the daily NSE values considerably lower than the corresponding monthly estimations. Moreover, while the CDFs here were calculated for 31 catchments, the CDFs in Mizukami et al. (2017) and Rakovec et al. (2019) represented approximately 500 basins belonging to a much bigger and heterogeneous domain that combines humid and dry regions. Within the climatic variability of the Iberian Peninsula, the Duero River Basin is representative of a wetter climate. The tendency for VIC to show poorer results in drier regions (Mizukami et al., 2017; Yang et al., 2019) could thus also explain its good performance in our study area.

Nevertheless, the calibration results were not exempt from poor skill metrics estimations. It is to be expected that the application of a single model structure over a heterogeneous spatial domain, such as the Duero River Basin, does not conduct to a good adjustment of the simulated streamflow with the

observations for all the studied catchments. Furthermore, the existence of other potential pressures over the water resources may be responsible for those cases where the evaluation exercise showed poor results, and therefore further research is required in order to identify the origin of the biases with respect to the observations of the simulated streamflows for these catchments.

Finally, the aggregation method has proven to be accurate and efficient, permitting its application in other studies using hydrologic models that operate over the grid cell in a similar mode to the VIC model. However, its applicability may become limited as the catchment size increases, but it is expected that for catchments with a similar size to these studied in this work, this only could happen for a shorter calibration time step (e.g. daily time step).

4.5.2 Spatial evaluation and AET predictability

The spatial evaluation exercise explored the parameter transferability and evinced the VIC performance for the 961 experiments resulting from crossing each catchment with each parameter combination. Many catchments produced low NSE biases for most experiments, as well as almost all the parameter combinations exhibited a high degree of transferability. The lowest degree of receptivity was evident for those catchments where the Monte Carlo analysis showed a lesser number of simulations close to the optimum (see Table 4.4 and compare with Fig. 4.5b).

Meanwhile, the parameter combinations reflecting the highest NSE biases were found for various catchments located in the north of the basin. This latitudinal gradient requires further research in order to study the main drivers of its behaviour (e.g. forcing dependency, relation to soil properties). Notably, these results could be the basis for a future application of VIC in both gauged and ungauged parts of the Duero River Basin using those parameters with greater transferability.

The VIC performance for the AET simulations was evaluated through the CDFs of the same for skill metrics calculated for the streamflow. Since the streamflow was the only variable subject to calibration, the VIC performance was slightly lower for the AET simulations, but broadly improved the benchmark performance. The VIC model showed slightly higher NSE values when compared to WRFERA, and this could be related to the use of some WRFERA

data as meteorological forcings. The underestimation of the variability and the AET volume are consistent with the VIC performance for the AET simulations over the CONUS domain (Rakovec et al., 2019). The values of alpham for the VIC-GLEAM comparison were higher than those for the VIC-WRFERA performance, and this is also in line with previous results exhibiting a predominantly positive annual relative bias of WRFERA compared with GLEAM for the study area (García-Valdecasas Ojeda et al., 2020a).

4.5.3 Sensitivity analysis and equifinallity assessment

The application of the SRC method allowed a deep understanding of the existing relationships between the components of the water balance in the VIC model and the selected parameters for the calibration as long as the linearity criterion was fulfilled. Even when the coefficient of determination of the fitted model did not satisfy the linearity criterion, the relatively low dispersion of the β coefficients permitted the interpretation of the results. Special attention deserves the component Q_t since it is the component that was compared with the streamflow observations during the calibration and validation processes. The sensitivity of this component to the five soil parameters reflected an additive effect of the sensitivity measures of Q_s and Q_b as Q_t is calculated as the sum of the surface and the subsurface components. Q_t was mainly sensitive to b_i and d_2, and this is consistent with the sensitivity measures for the simulated streamflow carried out in Demaria et al. (2007) for four studied catchments.

At the sight of the results of the equifinality assessment, it is unavoidable accepting that no parameter set leads to a single optimal model, or in other words, that there are many parameter samples with performances as good as the optimum calculated with the calibration algorithm. As in the GLUE method (Beven and Binley, 1992; Beven, 2012), this fact could be the starting point of the calibration process, in which a measure of belief is associated to each parameter set according to the degree of proximity to the optimum. This will be an interesting research line for further investigation in the Duero River Basin.

It is worth noting that if the calibration were repeated using different initial conditions, one could expect a similar spread of the calibrated parameters to that identified with the Monte Carlo analysis. However, fixing the number of samples for the Monte Carlo exercise ensured that the sensitivity analysis

could be done under equal conditions for all the catchments, and therefore independently from the number of trials of the SCE-UA algorithm for each individual calibration. In any case, we consider that the use of a calibration algorithm provides a first-look into the goodness-of-fit response surface of the hydrologic model in a computationally more efficient way than the Monte Carlo experiment, serving as a sign of the goodness-of-fit of the overall parameter samples.

4.6 Conclusions

The main conclusions of this chapter can be summarized as follows:

- The VIC performance for the streamflow simulations reflected good results for most of the studied catchments in the Duero River Basin, largely improving the benchmark performance. The results were slightly better for the reservoirs than for the gauging stations and this may be a consequence of a quality difference between the streamflow databases. The poor results found in a few subwaterheds may be caused by the existence of pressures over the water resources that have not been taken into account in the modelling exercise. However, this is out of the scope of this work since the main interest is placed on the macroscale application of the VIC model, which has shown to perform well for a great part of the Duero River Basin.

- An additional evaluation of the model performance was carried out for the transferability of the calibrated parameters and for the actual evapotranspiration simulations. Most parameter combinations exhibited a high degree of transferability, and the least transferable were associated to catchments located in the north of the basin. The VIC performance was evaluated for two actual evapotranspiration products, yielding satisfactory results with higher skill levels than the benchmark evaluation.

- The β coefficients calculated during the sensitivity analysis allowed to quantify the sensitivity of the water balance components to the selected parameters for the calibration. The surface runoff and the soil moisture content of the soil layer 2 were the components with the highest linearity and were mainly dominated by the values of the infiltration shape

parameter and the thickness of soil layer 2, respectively, both with a positive effect. The total runoff presented a combined behaviour from the surface runoff and the baseflow components, and the sensitivity analysis yielded similar results to other sensitivity measures previously reported in the literature. The potential evapotranspiration cycles were noticeable in the whole soil profile and more evidently in the upper two soil layers.

- A final exercise for assessing equifinality and the efficiency of the calibration algorithm was carried out, finding that there are many parameter sets with NSE values as high as the NSE determined during the calibration. The calibration algorithm was efficient and reached the optimal fit for almost all the studied catchments. The use of a calibration algorithm is also in line with other possible practical applications of the VIC model for studying the impacts of climate change on water resources in the Duero River Basin, where a parameter set must be chosen prior to the simulations using climate change data.

5 The Duero River Basin under future climate: projected hydrologic changes

The content of this chapter is based on the following publication:

Projected hydrologic changes over the north of the Iberian Peninsula using a Euro-CORDEX multi-model ensemble

Patricio Yeste, Juan José Rosa-Cánovas, Emilio Romero-Jiménez, Matilde García-Valdecasas Ojeda, Sonia R. Gámiz-Fortis, Yolanda Castro-Díez and María Jesús Esteban-Parra

Published in *Science of the Total Environment*[1]

Abstract

This chapter explores the impacts of climate change on the hydrology of the headwater areas of the Duero River Basin, the largest basin of the Iberian Peninsula. To this end, an ensemble of 18 Euro-CORDEX model experiments was gathered for two periods, 1975-2005 and 2021-2100, under two Representative Concentration Pathways (RCP4.5 and RCP8.5), and were used as the meteorological forcings of the Variable Infiltration Capacity (VIC) during the hydrological modelling exercise. The projected hydrologic changes for the future period were analyzed at annual and seasonal scales using several evaluation metrics, such as the delta changes of the atmospheric and

[1]DOI: https://doi.org/10.1016/j.scitotenv.2021.146126

land variables, the runoff and evapotranspiration ratios of the overall water balance, the snowmelt contribution to the total streamflow and the centroid position for the daily hydrograph of the average hydrologic year. Annual streamflow reductions of up to 40% were attained in various parts of the basin for the period 2071-2100 under the RCP8.5 scenario, and resulted from the precipitation decreases in the southern catchments and the combined effect of the precipitation decreases and evapotranspiration increases in the north. The runoff and the evapotranspiration ratios evinced a tendency towards an evaporative regime in the north part of the basin and a strengthening of the evaporative response in the south. Seasonal streamflow changes were mostly negative and dependent on the season considered, with greater detriments in spring and summer, and less intense ones in autumn and winter. The snowmelt contribution to the total streamflow was strongly diminished with decreases reaching -80% in autumn and spring, thus pointing to a change in the snow regime for the Duero mountains. Finally, the annual and seasonal changes of the centroid position accounted for the shape changes of the hydrograph, constituting a measure of seasonality and reflecting high correlations degrees with the streamflow delta changes.

5.1 Introduction

Global water resources are expected to undergo vital changes as a consequence of the increasing temperatures and the varying precipitation regimes projected for the future climate (IPCC, 2014). The global water cycle is governed by the partitioning of precipitation into evapotranspiration and runoff (Saha et al., 2020), and despite the importance of the future changes in precipitation, the changes in evapotranspiration and runoff can play an even more meaningful role for the assessment of future water security (Lehner et al., 2019).

Although many efforts have been made to identify the emergence of the climate change signal under a wide range of climate scenarios (Taylor et al., 2012; O'Neill et al., 2016; Lehner et al., 2019), its effects are already evident in certain regions and are expected to become stronger with the increase of greenhouse gas (GHG) emissions (IPCC, 2014). This is the case of the Mediterranean Basin, where water scarcity and the occurrence of extreme events have strengthened over the 20th century (García-Ruiz et al., 2011; Garrote et al., 2016). For instance, Tramblay et al. (2020) indicate a growing frequency and severity for Mediterranean droughts. Floods have shown a downward trend in many catchments of southern Europe over the last decades (Blöschl et al., 2019b; Tramblay et al., 2020), presumably due to decreasing precipitation and increasing evaporation ratios, and resulting in diminutions of up to 23% per decade (Blöschl et al., 2019b). However, catchments belonging to north-western Europe have manifested increasing floods of about 11% per decade (Blöschl et al., 2019b). This tendency is also noticeable for small catchments of few squared kilometers in south-western Europe (Amponsah et al., 2018), where enhanced convective storms and land-cover changes may cause flash floods to increase (Blöschl et al., 2019b).

Approximately one half of the water scarcity areas of the Mediterranean Basin are located in southern Europe (Iglesias et al., 2007; Garrote et al., 2016), where the runoff reductions can present a threat for meeting the water supply needs of the agricultural, industrial and urban water demands (García-Ruiz et al., 2011). Notably, the south-western sector of the Mediterranean region, represented by the Iberian Peninsula, has been identified as a hotspot particularly vulnerable to the climate change impacts (Diffenbaugh and Giorgi, 2012; Marx et al., 2018; García-Valdecasas Ojeda et al., 2020a, 2021a). Precipitation is expected to decrease over this region under climate change scenarios, with marked projected reductions in autumn, spring and summer for Spain (Argüeso

et al., 2012; García-Valdecasas Ojeda et al., 2020a,b) and Portugal (Soares et al., 2017). Projected evapotranspiration changes over the Iberian Peninsula reflect considerable spatio-temporal variability (García-Valdecasas Ojeda et al., 2020a,b) and result from the interplay between the future water availability in the soil and a higher atmospheric demand driven by increasing temperatures (Jerez et al., 2012), with a trend towards soil-drying conditions by the end of the 21st century (García-Valdecasas Ojeda et al., 2020a). On the other hand, there is substantial evidence for the decrease of Iberian streamflows during the last half of the 20th century (Lorenzo-Lacruz et al., 2012), with a strong consensus about this trend for different Iberian catchments under present and future climate conditions (Salmoral et al., 2015; Gampe et al., 2016; Pellicer-Martínez and Martínez-Paz, 2018; Yeste et al., 2018; Fonseca and Santos, 2019).

Climate change impact studies are mainly based on the analysis and application of the projections carried out with General Circulation Models (GCMs) and Regional Climate Models (RCMs) (Pastén-Zapata et al., 2020). While the conceptualization of the Earth System processes is common for both modelling approaches, their primary difference lies in the spatial resolution of the implemented domain, typically set at 2.5º for GCMs (Tapiador et al., 2020) and allowing a more accurate representation of the regional and local characteristics in the case of RCMs (Rummukainen, 2010; Teutschbein and Seibert, 2010). Nonetheless, the increasing computing power has led to a progressive refinement of the spatial resolution of GCMs, sometimes exceeding and improving the RCMs resolution, and are expected to be superseded by high-resolution GCMs in the next generation of climate model simulations (Tapiador et al., 2020). Anyhow, RCMs projections still remain as a valuable and suitable data source for impact studies given the lack of widespread availability of high-resolution GCM projections, and constitute an appropriate tool for the evaluation of hydrologic changes at the basin scale (Pastén-Zapata et al., 2020).

Using an ensemble of RCM simulations from the Euro-CORDEX initiative (Jacob et al., 2014), this chapter aims to identify and analyse the projected hydrologic changes for an important basin located in the north of the Iberian Peninsula, the Duero River Basin. The Duero basin has been previously studied mainly from a statistical perspective focused on various hydroclimatic and land-surface variables under present climate. For instance, Ceballos-Barbancho et al. (2008) and Morán-Tejeda et al. (2010) reported the impacts of land-cover changes on water availability and water resources management for the basin

during the last half of the 20th century. Morán-Tejeda et al. (2011a) provided useful insights on the different river regimes characterizing the Duero and its tributaries. More recently, Fonseca and Santos (2019) studied the impacts of climate change for the Tâmega River, a northern tributary of the Duero River located in Portugal, using a Euro-CORDEX ensemble as well.

In this chapter, the Variable Infiltration Capacity (VIC) model (Liang et al., 1994, 1996) has been implemented for various headwater catchments of the Duero basin based on the previous calibration exercise carried out in Yeste et al. (2020) for the study area. The VIC model largely improved the benchmark performance against streamflow observations and two actual evapotranspiration products, ensuring the further applicability of the calibrated parameters for the modelling exercise here developed. The main objective of this work consists of evaluating the future changes of the different hydrologic variables involved in the water balance at annual and seasonal scales, adopting different interrelated approaches that accurately highlight many fundamental features of the future hydrologic behaviour of the basin.

5.2 Study area and Data

The Duero River Basin is an international basin located in the north of Spain and Portugal and represents the largest basin of the Iberian Peninsula (98073 km^2). The focus of this work is placed on the 80% of its area, corresponding to the Spanish territory (Fig. 5.1). The topography of the basin is mainly constituted by a large central depression and the surrounding mountain chains that configure the headwater areas of the hydrologic network. The mean annual precipitation volume is around 50000 hm^3 and mostly evaporates into the atmosphere (~ 35000 hm^3), representing the remaining volume the water contribution of the basin as natural runoff. With a predominant Mediterranean climate, most of the precipitation occurs in the mountainous systems, exceeding 1000 mm/year in the northern mountains and showing values below 1000 mm/year in the southern part of the basin. It is concentrated in the autumn, winter and spring months, with a dry period affecting the majority of the area during summer, with a warmer temperature (~ 20.5 ºC in July).

The selection of the catchments for this chapter (Fig. 5.1, Table S5.1 in Supporting Information) was based on the implementation of the VIC model carried out in Yeste et al. (2020) for the headwaters of the Duero River Basin (see

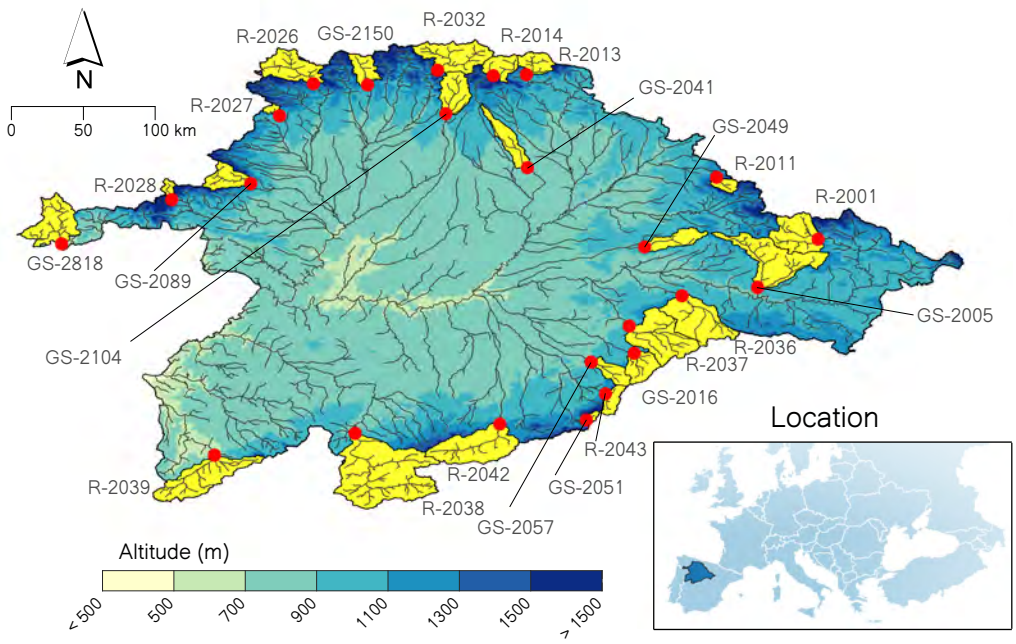


Fig. 5.1: Duero River Basin and the 24 studied catchments. The prefix "R-" denotes "Reservoir" and the prefix "GS-" denotes "Gauging Station".

Chapter 4). The Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) was selected there as the main evaluation metric, and despite the good performance for the majority of the studied catchments, some of them showed poor NSE estimations. A threshold NSE value of 0.67 was set in this work as an acceptable model performance based on previous studies (e.g., Martinez and Gupta, 2010; Ritter and Muñoz-Carpena, 2013; Her et al., 2019). This criterion reduced the number of catchments considered to 24 out of the 31 originally included in Yeste et al. (2020) (see Table S5.1).

Daily climate data were gathered from the Euro-CORDEX project for 18 RCM experiments (Jacob et al., 2014, see Section 2.4 in Chapter 2) at a spatial resolution of 0.11º (EUR-11, ~ 12.5 km). The dataset was regridded to 0.05º (~ 5 km) using the Climate Data Operator (CDO) software (Schulzweida, 2019) and choosing a nearest neighbour assignment for the subsequent hydrological modelling exercise.

5.3 Methods

5.3.1 Bias correction

The straightforward application of raw RCM data for hydrological impact studies is inadequate given the emerging systematic errors (i.e. biases) during the dynamical downscaling of GCM outputs (Gudmundsson et al., 2012; Hanzer et al., 2018). These uncertainties are usually managed with the use of ensembles of RCM simulations and the application of bias correction techniques (Déqué, 2007; Teutschbein and Seibert, 2012). Within the different bias correction methods, the Quantile Mapping (QM) (e.g., Wood et al., 2002; Déqué, 2007; Themeßl et al., 2011) has shown to produce better results (Themeßl et al., 2011; Teutschbein and Seibert, 2012; Hakala et al., 2018) and allows the correction of daily precipitation and temperature data (Meresa and Romanowicz, 2017; Hakala et al., 2018; Pastén-Zapata et al., 2020).

In this work we used the R package 'qmap' (Gudmundsson et al., 2012) in order to fit the cumulative distribution functions (CDFs) of the meteorological time series to the CDFs of the observations. Precipitation, maximum temperature and minimum temperature were the only bias-corrected variables given the absence of observations for the rest of the meteorological fields. For this end, daily precipitation, maximum temperature and minimum temperature data were gathered from the observational datasets SPREAD and STEAD (Serrano-Notivoli et al., 2017, 2019, see Section 2.2 in Chapter 2) for the historical period. The QM method was then applied for each month of the year using pooled daily data. In the case of daily temperature, the QM technique was applied to the diurnal temperature range (DTR) and the maximum daily temperature. This approach effectively avoids the occurrence of negative DTR values and improves the posterior estimation of minimum daily temperature (Thrasher et al., 2012).

5.3.2 Snowmelt contribution to the total streamflow

The total runoff simulated with the VIC model feeds on water arising from both rainfall and snowmelt that infiltrates into the soil. However, the proportion of runoff corresponding to each of them is not explicitly accounted for (Siderius et al., 2013), and can be calculated for each month as follows:

$$Q_{snow,i} = \min \left\{ Q_i \cdot \left[\frac{melt_i}{rain_i + melt_i} \right], melt_i \right\} \tag{5.1}$$

For the month i, $Q_{\text{snow,i}}$ [L^3/T] is the streamflow arising from the snowmelt (i.e. melt streamflow), $Q_{\text{snow,i}}$ [L^3/T] is the total streamflow, $melt_{\text{i}}$ [L^3/T] is the snowmelt rate and $rain_{\text{i}}$ [L^3/T] the rainfall rate. This approach is analogous to that applied in Siderius et al. (2013) and Li et al. (2019), and constitutes an appropriate method for the estimation of the snowmelt contribution to the total runoff. Eq. 5.1 assumes that $Q_{\text{snow,i}}$ cannot exceed the melt-to-rain ratio nor the total snowmelt occurring for a given month, thus not introducing an imbalance conducive to unrealistic values. Note that in this work we use the terms "runoff" and "streamflow" interchangeably due to the aggregation method applied in Yeste et al. (2020) to the raw gridded outputs from VIC for obtaining the hydrologic time series at the subwateshed scale.

5.3.3 Evaluation metrics and projected hydrologic changes

The projected hydrologic changes were firstly analysed applying the delta change approach (Hay et al., 2000) to the mean annual and seasonal values of precipitation (P), potential evapotranspiration (PET), actual evapotranspiration (AET), total streamflow (Q) and melt streamflow (Q_{snow}). The statistical significance of the delta-changes was evaluated using the Mann-Whitney U test at 95% confidence level.

The future changes were then evaluated using five hydrologic signature measures (Stewart et al., 2005; Rasmussen et al., 2014; Mendoza et al., 2015, 2016). The first two measures provide information about the overall water balance for a certain region and can be derived from the water balance equation normalized by P:

$$1 = \frac{Q}{P} + \frac{AET}{P} + \frac{\Delta S}{P} \tag{5.2}$$

Where ΔS is the variation of the storage in the hydrological system for a given period, and all the variables are expressed in units of volume. For long periods,

the storage component can be neglected (Rasmussen et al., 2014; Mendoza et al., 2015, 2016), and Eq. 5.2 can be rewritten as:

$$1 = \frac{Q}{P} + \frac{AET}{P} = i_Q + i_E \tag{5.3}$$

Where i_{Q} is the runoff ratio and i_{E} is the evapotranspiration ratio. Eq. 5.3 represents the steady-state for the water balance, and implies that both measures are complementary. Therefore, i_{Q} and i_{E} will be used to evaluate the present and future partitioning of precipitation into runoff and actual evapotranspiration at annual scale.

The third signature measure is the snowmelt contribution ratio to the total streamflow, $Q_{\mathrm{snow}}\ ratio$, and can be calculated both at annual and seasonal scales as follows:

$$Q_{snow}ratio = \frac{Q_{snow}}{Q} \tag{5.4}$$

Lastly, the centroid position for the daily hydrograph of the average hydrologic year was selected for the diagnosis of the projected Q changes. The most common metric is the X coordinate of the centroid for the entire hydrologic year or "center time" of runoff (Stewart et al., 2005; Mendoza et al., 2015, 2016), and evaluates the seasonality of runoff. In this work we have calculated both the X and Y coordinates (C_{X}, C_{Y}) at annual and seasonal scales, as shown in Fig. 5.2. The annual and seasonal centroids together provide a more accurate picture about the daily hydrograph. In addition, C_{Y} is a valuable source of information about the shape of the hydrograph, and its future changes are related to the annual and seasonal delta-changes of Q. The correlation between the C_{Y} changes and the projected Q delta-changes was tested through a linear adjustment calculating the coefficient of correlation r and the corresponding slope and intercept values. In this case the statistical significance of the r values was calculated using the Student's-t test at 95% confidence level.

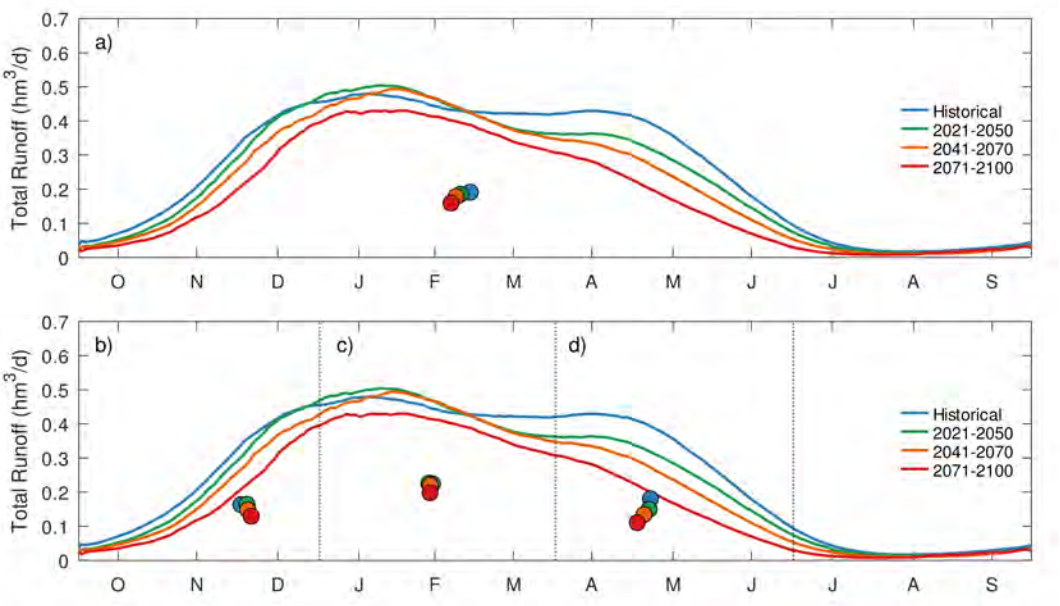


Fig. 5.2: X and Y coordinates of the centroid for the daily hydrograph (C_X, C_Y) at annual and seasonal scales: a) annual centroid, b) fall centroid, c) winter centroid and d) spring centroid.

5.3.4 Model validation

The predictive capability of VIC was firstly tested through the NSE values corresponding to the monthly Q and AET ensemble simulations for the period Oct 2000 - Sep 2011. This period was selected based on the prior calibration exercise of Yeste et al. (2020), and spans the last part of the historical period and the first years of the two RCP scenarios (historical+RCP validation periods hereafter). The analysis was carried out for both historical+RCP validation periods. Similarly to Yeste et al. (2020), Q simulations were validated against monthly streamflow observations gathered from SAIH-ROEA (see Section 2.1 in Chapter 2), and AET was compared to the monthly outputs of the GLEAM version 3.3a (Martens et al., 2017; Miralles et al., 2011, see Section 2.3 in Chapter 2).

Lastly, the suitability of VIC to simulate the evaluation metrics previously described was analyzed for the two historical+RCP validation periods. The annual partitioning of precipitation into runoff and evapotranspiration was only evaluated for i_Q using streamflow observations and SPREAD precipita-

tion, assuming the steady-state for the water balance represented in Eq. 5.3. The annual and seasonal C_X and C_Y values were compared to those obtained with daily streamflow observations. The validation for $Q_{snow}\ ratio$ was not possible given the limited amount of observations for snow-related variables in the Iberian Peninsula.

5.4 Results

5.4.1 Validation results

The NSE values of Q and AET for the historical+RCP8.5 validation period are shown in Fig. 5.3a and Fig. 5.3b, respectively, and are almost identical to those calculated for the historical+RCP4.5 validation period (Fig. S5.1a and Fig. S5.1b in Supporting Information). NSE(Q) values are highly underestimated using the Euro-CORDEX multi-model ensemble in comparison to the calibration results in Yeste et al. (2020). This indicates that, while being an appropriate measure for calibrating hydrologic models, NSE(Q) constitutes a high-end performance extremely difficult to achieve when using climate model outputs.

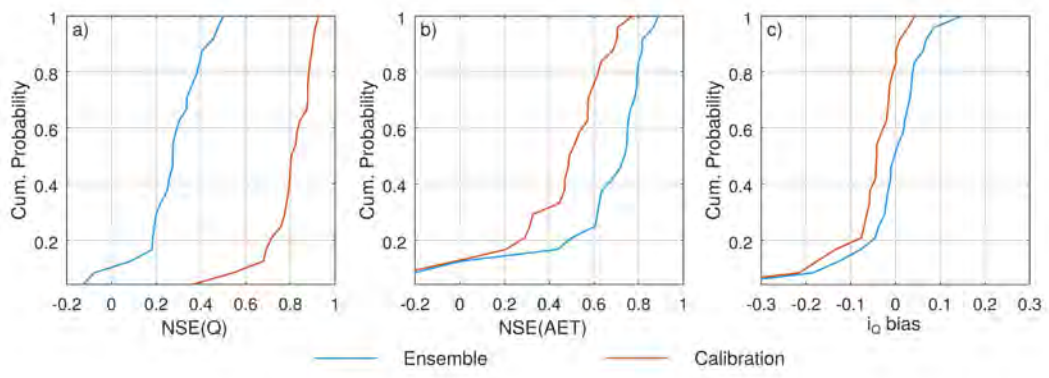


Fig. 5.3: CDFs of the VIC model performance for the period Oct 2000 - Sep 2011 corresponding to the combination historical+RCP8.5: a) NSE for the streamflow simulations against streamflow observations; b) NSE for the AET simulations against GLEAM; c) i_Q bias with respect to the ratio of SPREAD precipitation to streamflow observations. Blue lines represent the ensemble simulation, and orange lines correspond to the calibration results from Yeste et al. (2020).

The runoff performance is commonly checked for less demanding metrics such as i_Q, which is calculated from mean values of runoff and precipitation (Eq. 5.3) and plays a major role in assessing future water security (Lehner et al., 2019). Contrarily to NSE(Q), the i_Q biases of the ensemble and the calibration presented similar distributions with most values falling in the ±0.1 range (Fig. 5.3c and Fig. S5.1c in Supporting Information).

On the other hand, the ensemble clearly improved the VIC performance for the AET simulations, presumably due to the model-nature of GLEAM. Further calibration efforts for the Duero basin will aim to improve the VIC performance for AET targeting both Q and AET simulations simultaneously.

Finally, the annual and seasonal C_X and C_Y biases for the historical+RCP8.5 validation period (Fig. 5.4) resemble the ones estimated for the historical+RCP4.5 validation period (Fig. S5.2 in Supporting Information). Overall, the CDFs corresponding to the ensemble suggest an acceptable performance when compared to the calibration results from Yeste et al. (2020). The ensemble reflected a higher presence of positive biases, while the calibration presented slightly steeper CDFs closer to 0. Notably, the C_X biases at annual scale are comparable to those showed in Mendoza et al. (2015, 2016) for three headwater catchments in the Colorado River Basin.

5.4.2 Annual delta changes of P, PET, AET and Q

The mean annual values of P, PET and AET for the historical period are depicted in Fig. 5.5a, and the mean annual values of Q for this period are gathered in Table 5.1. A marked latitudinal gradient was found for the atmospheric variables, with P values generally above 1000 mm/year and PET values below 1000 mm/year for the northern catchments, and reaching minimum P and maximum PET in the south. AET shows a narrower range of variability with the majority of values falling between 400 and 700 mm/year. The latitudinal gradient is also noticeable for this variable and reflects an opposite spatial distribution to PET.

Fig. 5.6 collects the annual delta changes of P, PET, AET and Q for the period 2071-2100 under the RCP8.5 scenario. The changes corresponding to RCP4.5 for the three future periods, and to the RCP8.5 for the 2021-2050 and 2041-2070 are shown in Supporting Information (Fig. S5.3 to Fig. S5.7). A generalized decrease of annual P is expected for all the future study periods presenting

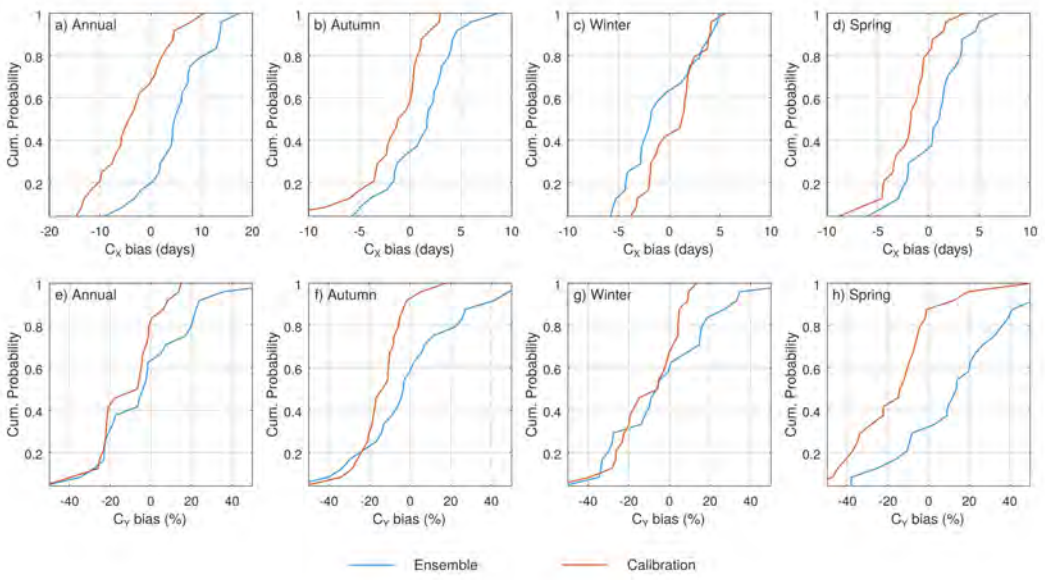


Fig. 5.4: CDFs of the VIC model performance for the annual and seasonal values of C_X (a to d) and C_Y (e to h) corresponding to the period Oct 2000-Sep 2011 and the combination historical+RCP8.5. C_X biases are calculated as the difference between simulated and observed values. C_Y biases represent fractional changes calculated as [(simulations - observations)/observations · 100]. Blue lines represent the ensemble simulation, and orange lines correspond to the calibration results from Yeste et al. (2020).

maximum decreases of up to 40% in the south for the period 2071-2100 under the RCP8.5 scenario. PET is subject to significant increases for all the future study periods and RCPs, with maximum increases, above 40%, taking place in the northern catchments by the end of the century under the RCP8.5 scenario. The annual delta changes of P and PET suggest that the latitudinal gradient noticed for both atmospheric fields in the historical period (Fig. 5.5a) tends to fade away, and thus P and PET become more homogeneous over the entire study area in the future periods.

The delta changes of AET range from significant increases of up to 30% in the northern catchments to significant decreases of about 30% in the south for the period 2071-2100 under the RCP8.5 scenario (Fig. 5.6). The AET changes reflect a greater heterogeneity than the P and PET ones, and although the historical values follow a north-south distribution (Fig. 5.5a), the future AET changes do not compensate the latitudinal gradient. On the contrary, they exacerbate it, leading to a widening on the range of values of AET.

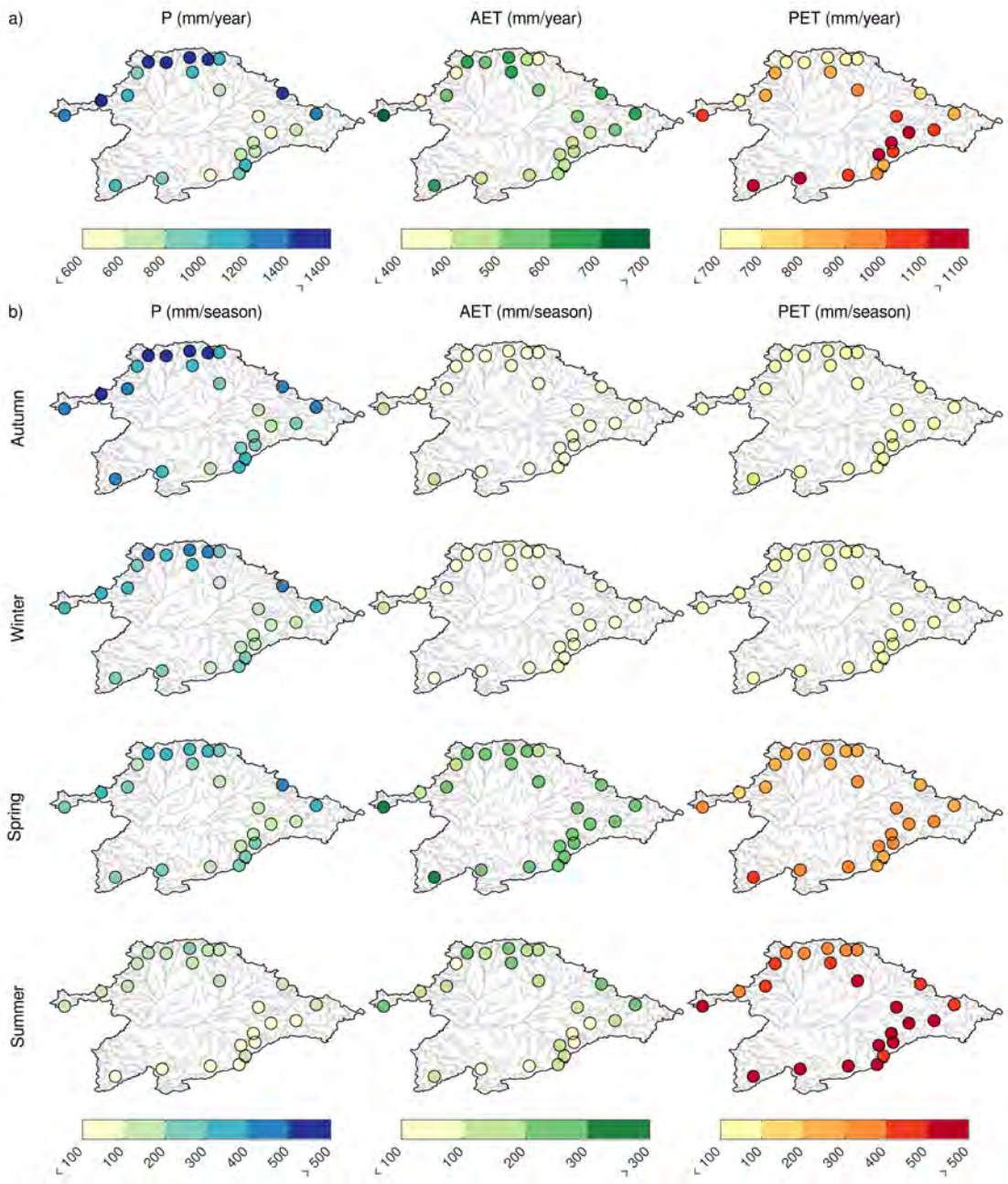


Fig. 5.5: Basin-averaged annual and seasonal values of P, AET and PET for the historical period.

Table 5.1: Basin-averaged annual and seasonal values of Q for the historical period.

Code	Annual (hm^3/year)	Autumn (hm^3/season)	Winter (hm^3/season)	Spring (hm^3/season)	Summer (hm^3/season)
R-2001	312.22	65.43	121.62	101.05	24.12
R-2011	93.70	21.41	40.31	29.22	2.76
R-2013	150.37	42.42	57.24	41.88	8.83
R-2014	229.93	60.93	84.57	69.71	14.71
R-2026	445.30	110.75	158.98	138.26	37.31
R-2027	23.32	7.44	10.90	4.36	0.62
R-2028	71.66	25.41	24.70	17.98	3.57
R-2032	658.43	184.00	266.30	181.26	26.86
R-2036	46.80	11.21	17.89	15.00	2.70
R-2037	99.36	22.37	40.16	25.90	10.93
R-2038	773.35	237.64	337.55	178.98	19.18
R-2039	299.90	88.51	151.97	55.34	4.07
R-2042	122.79	22.17	64.59	32.49	3.55
R-2043	87.76	16.84	33.61	31.56	5.75
GS-2005	113.82	19.94	55.21	30.53	8.13
GS-2016	81.50	17.42	36.68	22.49	4.90
GS-2041	16.69	2.98	9.56	3.79	0.35
GS-2049	19.12	2.89	7.64	7.18	1.41
GS-2051	18.38	3.66	8.43	5.72	0.58
GS-2057	47.05	8.83	21.39	15.20	1.63
GS-2089	182.53	46.51	79.67	46.39	9.96
GS-2104	147.60	36.24	64.15	36.22	11.00
GS-2150	191.03	57.94	73.14	53.74	6.20
GS-2818	270.69	52.50	131.59	74.83	11.77

The delta changes of Q are prevalently negative and statistically significant for all periods and RCPs, with changes below -40% in some of the southern catchments for the period 2071-2100 under the RCP8.5 scenario (Fig. 5.6). In this respect, two main driving mechanisms for the annual Q detriments were identified: 1) the future P decreases in the southern mountains; and 2) the combined effect of the future P reductions and the AET increases in the north part of the basin. The former mechanism supposes that both the runoff generation and the evapotranspiration processes become limited by the less abundant precipitations under the future scenarios. The latter corresponds to those areas where there is still enough water availability for satisfying the higher atmospheric demand for water vapour (i.e. higher PET) in the future, and therefore represents a two-fold limiting factor for the runoff generation.

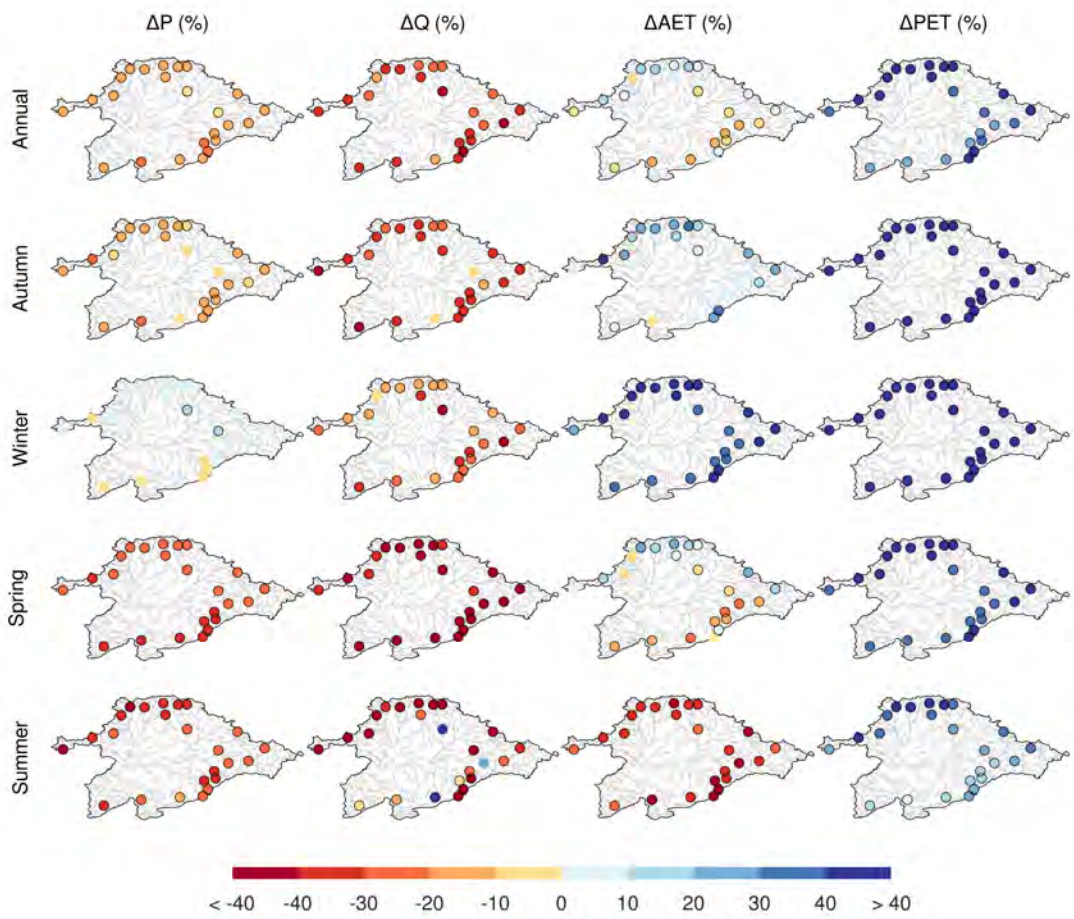


Fig. 5.6: Delta changes of annual and seasonal P, Q, AET and PET for the period 2071-2100 under the RCP8.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

5.4.3 Seasonal delta changes of P, PET, AET and Q

The mean seasonal values of P, PET and AET for the historical period are mapped in Fig. 5.5b, and Table 5.1 includes the mean seasonal values of Q. Seasonal P evidences a latitudinal gradient for all the seasons comparable to that observed in Fig. 5.5a for annual scale. The highest values are reached during autumn and are above 500 mm/season for various northern catchments. The minimum values correspond to summer and are below 100 mm/season in the south. Seasonal PET is broadly below 100 mm/season in autumn and winter, being the latitudinal gradient almost inexistent. The PET values start

to be noticeable in spring and achieve their maximum in summer with values above 500 mm/season in the southern catchments. The maximum AET values were obtained for the spring months and are above 200 mm/season for the majority of the subwatesheds. The summer AET values are somewhat lower than the spring ones, and they are broadly below 200 mm/season in autumn and winter. Finally, the highest Q volumes take place in winter and are lower in autumn and spring, with minimums attained in summer (Table 5.1).

Fig. 5.6 also shows the seasonal delta changes of P, PET, AET and Q for the period 2071-2100 under the RCP8.5 scenario, being the changes associated to the rest of the future periods and scenarios shown together with their annual counterparts in Supporting Information (Fig. S5.3 to Fig. S5.7). The smallest decreases of seasonal P correspond to autumn, while they predominantly exceed 20% in spring and summer. Delta changes for winter, in turn, are mostly positive, and notably significant under the RCP4.5 scenario for 2071-2100 (Fig. S5.7) and the RCP8.5 scenario for 2041-2070 (Fig. S5.6). Little difference was found for the seasonal PET changes with respect to the annual changes, though the significant increments are slightly larger in autumn and winter, and less severe in spring and summer.

The significant AET increases detected for autumn and winter (Fig. 5.6) manifest that the evapotranspiration process is limited by the atmospheric demand of water vapour (i.e. PET) for the first half of the hydrologic year. During summer, however, the AET changes are negative and significant in the entire region, suggesting that the water availability constrains the evaporative fluxes. The spring AET changes lie between those extremes and show significant increases in the northern part of the basin and significant decreases over the south, with similar results for the rest of the future periods and scenarios (Fig S5.3 to Fig. S5.7). Hence, the seasonal AET changes can explain the annual AET delta changes as follows: 1) the increases identified for the northern catchments are due to the increments occurring in autumn, winter and spring, without a noteworthy effect of the summer diminutions on the annual differences; and 2) the projected detriments in the southernmost areas are promoted by the spring and summer decreases, whilst the autumn and winter increases do not cause a flip in the sign of the delta changes.

The seasonal delta changes of Q are mostly negative and more pronounced in spring and summer, reaching reductions above 40% in a great number of the catchments for the RCP8.5 scenario during the period 2071-2100 (Fig. 5.6). There is a strong interplay between the seasonal P and AET as the driving

forces of the seasonal Q detriments. Thus, P constitutes the limiting factor for both the runoff generation and the evapotranspiration processes in summer, and over the southern part of the basin in spring. The combined effect of P decreases and AET increments is relevant in autumn, as well as for the northern catchments in spring. The negative delta changes of Q for winter are related to the sharp and significant AET increments.

5.4.4 Delta changes of Q_{snow}

The application of Eq. 5.1 allowed obtaining the Q_{snow} monthly time series for each catchment of the study area that subsequently were aggregated at annual and seasonal time scales. Table 5.2 collects the mean annual and seasonal values of Q_{snow} for the historical period, showing higher Q_{snow} values for the northern catchments as they are characterized by a greater elevation (see Fig. 5.1 and Table S5.1). Fig. 5.7 shows the delta changes of annual and seasonal Q_{snow} for the period 2071-2100 under the RCP8.5 scenario. The remaining results for Q_{snow} are shown in Supporting Information (Fig. S5.8). The summer months were excluded from this analysis due to the absence of a snowmelt component for the streamflow values in this season.

The delta changes of Q_{snow} are always negative and significant, mostly exceeding 80% in the long-term future period (Fig. 5.7), and being more pronounced in autumn and spring. Q_{snow} constitutes the hydrological variable for which the impact of climate change is more evident, evincing a generalized tendency

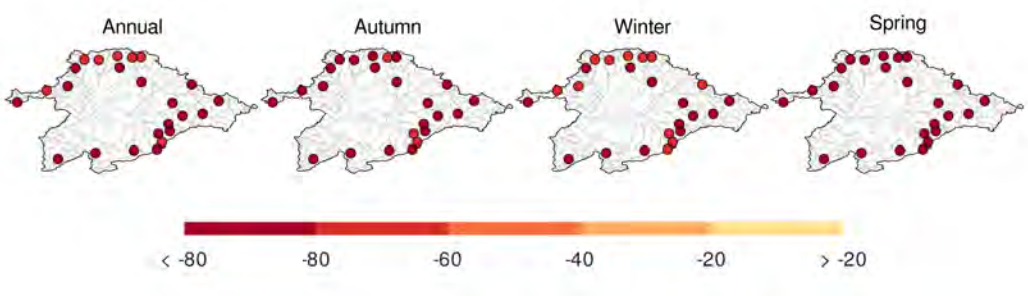


Fig. 5.7: Delta changes (excluding summer) of annual and seasonal Q_{snow} for the period 2071-2100 under the RCP8.5 scenario in the studied catchments. Significant changes at a 95% confidence level have been marked with solid borders.

Table 5.2: Basin-averaged annual and seasonal values (excluding summer) of Q_{snow} for the historical period.

Code	Annual (hm^3/year)	Autumn (hm^3/season)	Winter (hm^3/season)	Spring (hm^3/season)
R-2001	69.44	11.00	44.61	13.82
R-2011	22.25	4.24	14.44	3.56
R-2013	35.75	7.88	22.96	4.91
R-2014	72.94	15.34	40.34	17.26
R-2026	141.16	27.52	75.29	38.28
R-2027	2.81	0.67	2.04	0.09
R-2028	23.06	6.37	11.70	4.99
R-2032	195.16	40.66	114.07	40.42
R-2036	5.42	1.10	3.78	0.54
R-2037	10.83	1.93	7.86	1.03
R-2038	122.19	27.18	82.45	12.56
R-2039	6.80	1.67	5.07	0.05
R-2042	19.50	2.73	15.42	1.35
R-2043	27.99	4.71	16.09	7.19
GS-2005	9.77	1.30	8.13	0.34
GS-2016	11.47	2.18	8.27	1.02
GS-2041	1.29	0.15	1.12	0.02
GS-2049	0.87	0.14	0.69	0.04
GS-2051	5.02	0.85	3.37	0.80
GS-2057	11.24	1.50	7.17	2.58
GS-2089	39.49	7.80	26.75	4.94
GS-2104	22.28	4.02	16.62	1.64
GS-2150	60.15	13.67	32.81	13.63
GS-2818	6.87	0.87	5.75	0.25

for the headwaters of the Duero River Basin to be much less snow dominated as a consequence of climate change.

5.4.5 Future changes of i_Q, i_E and Q_{snow} $ratio$

Fig. 5.8a shows the values of the signature measures i_Q and i_E for the historical period, and their future changes are collected in Fig. 5.8b for the period 2071-

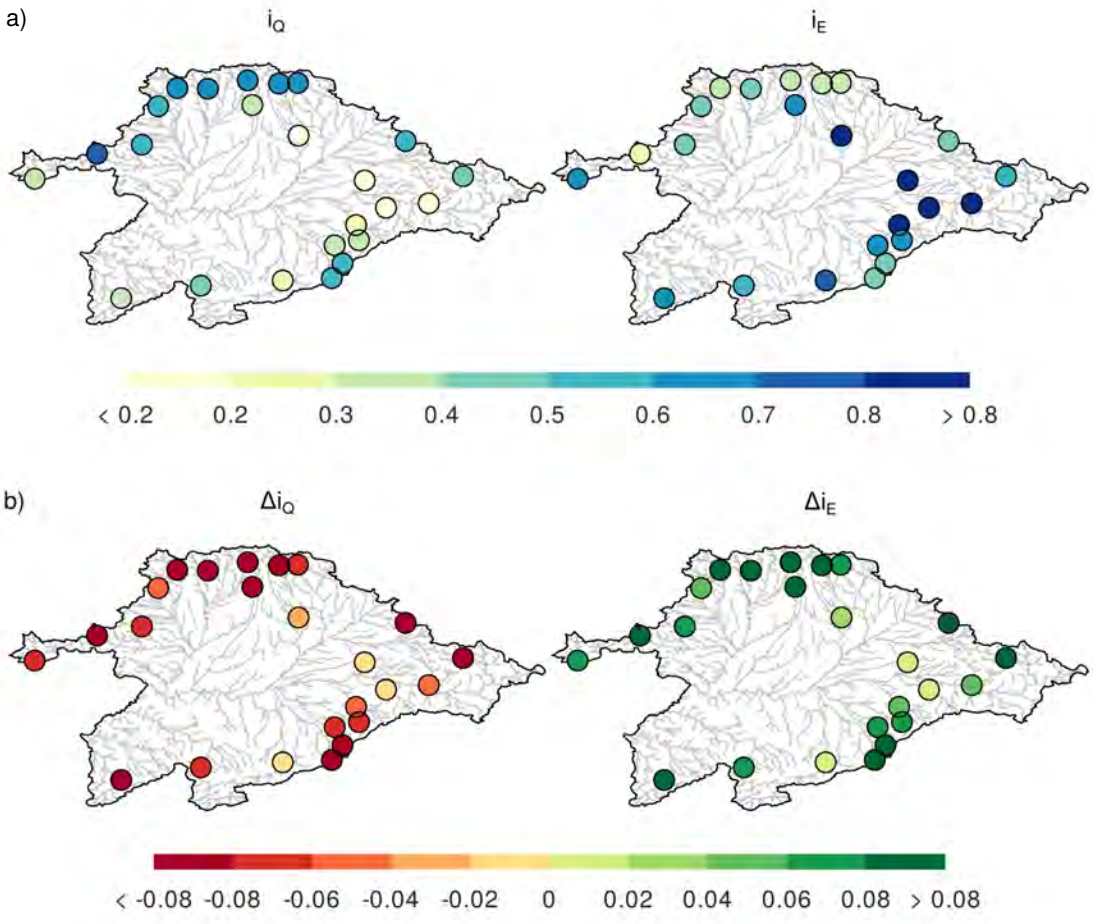


Fig. 5.8: (a) Values of i_Q and i_E for the historical period in the studied catchments. (b) Projected changes of i_Q and i_E calculated as the difference between future and historical values for the period 2071-2100 under the RCP8.5 scenario.

2100 under the RCP8.5 scenario (Fig. S5.9 in Supporting Information depicts the results of i_Q for the rest of the periods and RCP scenarios). The highest values of i_Q and i_E occur for the northern and for the southern catchments, respectively, being the sum of both ratios always close to 1. This complementary assumption is also applicable to the i_Q and i_E changes, being their sum close to 0. The future changes of i_Q and i_E suggest that the northern catchments tend towards the evaporative regime range, and the evaporative response becomes stronger in the south.

The Q_{snow} $ratio$ values for the historical period are presented in Fig. 5.9a

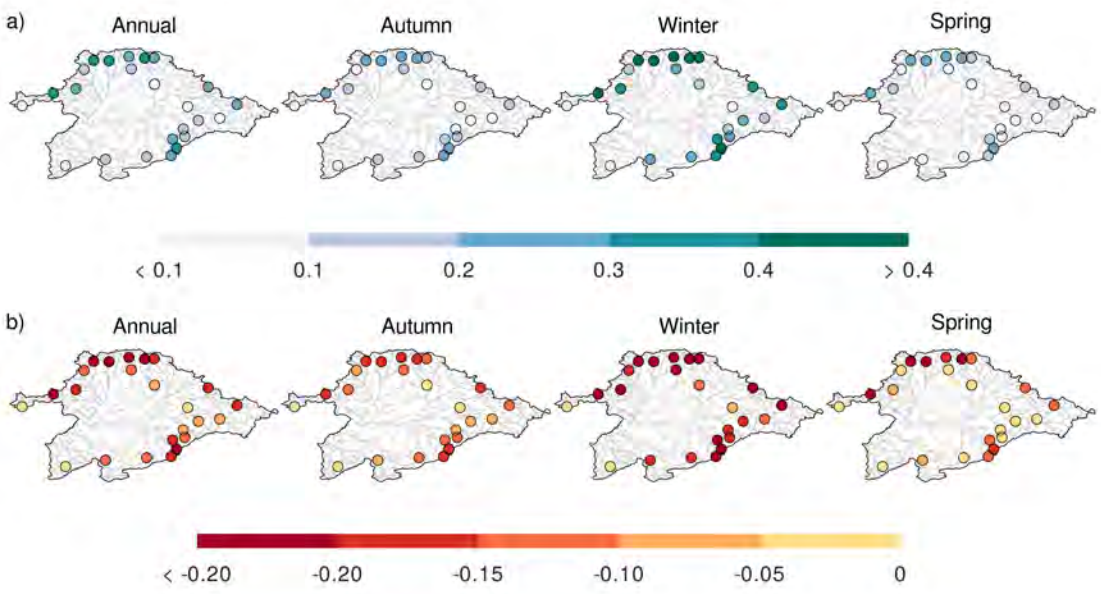


Fig. 5.9: (a) Annual and seasonal values (excluding summer) of Q_{snow} $ratio$ for the historical period in the studied catchments. (b) Projected changes of annual and seasonal Q_{snow} $ratio$ (excluding summer) calculated as the difference between future and historical values for the period 2071-2100 under the RCP8.5 scenario.

at both annual and seasonal scales, excluding summer. Most of the annual Q_{snow} $ratio$ values are above 0.3 for the northern basins, being weaker in the southern mountains, with values that mainly range from 0.1 to 0.3. The seasonal distribution reveals that the highest Q_{snow} $ratio$ are concentrated in the winter months and exceeds 0.4 in many northern catchments. Autumn and spring show similar results with values always below 0.3. The future changes of the Q_{snow} $ratio$ at annual and seasonal time scales are depicted in Fig. 5.9b for the period 2071-2100 under the RCP8.5 scenario (the rest of the changes are shown in Fig. S5.10 in Supporting Information), and manifest a clear predominance of negative values broadly below -0.1.

5.4.6 Future changes of the centroid position

Table 5.3 collects the coordinate pairs (C_X, C_Y) of the annual and seasonal centroids for the daily hydrograph for the average hydrologic year in the historical period. The annual C_X values present a mean value of 150.4 days

Table 5.3: Centroid position at annual and seasonal time scales in the daily hydrograph for the average hydrologic year corresponding to the historical period. C_X is expressed in days since Oct 1 and C_Y in hm^3/day.

Code	Annual		Autumn		Winter		Spring	
	C_X	C_Y	C_X	C_Y	C_X	C_Y	C_X	C_Y
R-2001	162.3	0.560	57.2	0.422	139.3	0.684	221.7	0.597
R-2011	150.6	0.191	62.9	0.164	136.2	0.224	219.4	0.181
R-2013	148.1	0.267	56.4	0.266	136.7	0.318	222.5	0.241
R-2014	152.7	0.405	56.4	0.381	137.5	0.469	221.6	0.411
R-2026	158.5	0.763	56.3	0.689	138.7	0.884	220.9	0.827
R-2027	131.2	0.051	61.4	0.054	133.5	0.062	219.4	0.027
R-2028	137.0	0.128	52.9	0.149	135.7	0.138	221.3	0.107
R-2032	144.6	1.254	58.0	1.198	136.8	1.478	218.7	1.137
R-2036	155.2	0.086	54.9	0.069	139.4	0.100	218.6	0.095
R-2037	160.3	0.169	55.5	0.137	137.2	0.224	219.4	0.158
R-2038	134.3	1.578	56.6	1.502	134.8	1.900	215.8	1.207
R-2039	129.6	0.738	64.4	0.716	131.9	0.890	213.9	0.395
R-2042	150.1	0.278	62.8	0.170	137.5	0.361	216.3	0.215
R-2043	164.8	0.164	60.9	0.120	138.8	0.187	222.3	0.184
GS-2005	157.8	0.233	58.9	0.136	137.0	0.314	217.1	0.199
GS-2016	154.1	0.157	61.0	0.124	136.5	0.204	221.1	0.133
GS-2041	144.3	0.043	66.0	0.027	134.0	0.056	214.6	0.027
GS-2049	171.5	0.037	60.9	0.021	139.9	0.043	222.9	0.041
GS-2051	153.6	0.038	64.4	0.029	137.3	0.047	219.4	0.035
GS-2057	156.1	0.097	59.0	0.059	140.3	0.120	218.1	0.097
GS-2089	147.5	0.352	59.9	0.320	135.9	0.445	219.2	0.285
GS-2104	150.6	0.272	56.2	0.225	136.7	0.358	218.0	0.229
GS-2150	141.6	0.367	58.0	0.379	136.2	0.407	219.6	0.337
GS-2818	152.8	0.574	64.9	0.435	136.3	0.736	219.2	0.459

and a difference of 41.9 days between the maximum and minimum values (i.e. dispersion). The lowest annual C_X values were generally obtained for northern catchments (e.g. R-2027, R-2028 and GS-2150), being the highest ones mainly located in the south (e.g. R-2037, R-2043 and GS-2057). The seasonal C_X values are less dispersed, with a maximum difference of 13.1 days in autumn and not exceeding 10 days in winter and spring. The spatial distribution of C_X for autumn and spring is comparable to that obtained for the entire year, while

spring exhibits an opposite pattern with minimums attained in the south. On the other hand, a strong correlation ($r > 0.99$) was found between the C_Y measures and the average Q values for the historical period (Table 5.1) in all cases, thus implying that the highest C_Y values take place in winter and are lower in autumn and spring.

The projected changes of the centroid position are shown in Fig. 5.10. The annual C_X changes (Fig. 5.10a) are predominantly negative and more pronounced for the RCP8.5 scenario, being the differences mostly below 10 days. A similar behaviour is also noticeable, but to a lesser degree, in winter and spring. The autumn C_X changes are, in turn, prevailingly positive and present increases of up to 5 days. This represents an important feature of the future behaviour of the autumn streamflow that is not well-captured in the annual C_X changes.

The C_Y changes (Fig. 5.10b) are negative and broadly below 20% both at annual and seasonal time scales. However, the spring C_Y decreases are more noticeable and usually exceed 20%, reaching decreases above 40% for RCP8.5. The correlation between the C_Y changes and the delta changes of Q (Table S5.2 in Supporting Information) is characterized by r values statistically significant at a 95% confidence level, with a varying degree of correlation depending on the time scale and RCP considered. The C_Y and the Q changes practically show a 1:1 relationship for winter and spring and both RCPs. Conversely, the correlation is less marked for autumn and for the complete hydrologic year, with r values below 0.9 and slightly above 0.9 at annual scale for the RCP8.5 scenario. Therefore, as the delta changes of Q, the projected changes of C_Y pinpoint a generalized decrease of the streamflow for the study area, being in some cases interchangeable measures (i.e. in winter and spring).

5.5 Discussion

5.5.1 Projected annual hydrologic changes

A similar spatial pattern for the annual P changes over the Duero Basin (Fig. 5.6) was detected using WRF simulations over Spain in Argüeso et al. (2012), concluding that the changes tend to be larger in the southern half of the domain, particularly over the mountainous areas. Likewise, the annual P changes are in agreement with those obtained in Fonseca and Santos (2019)

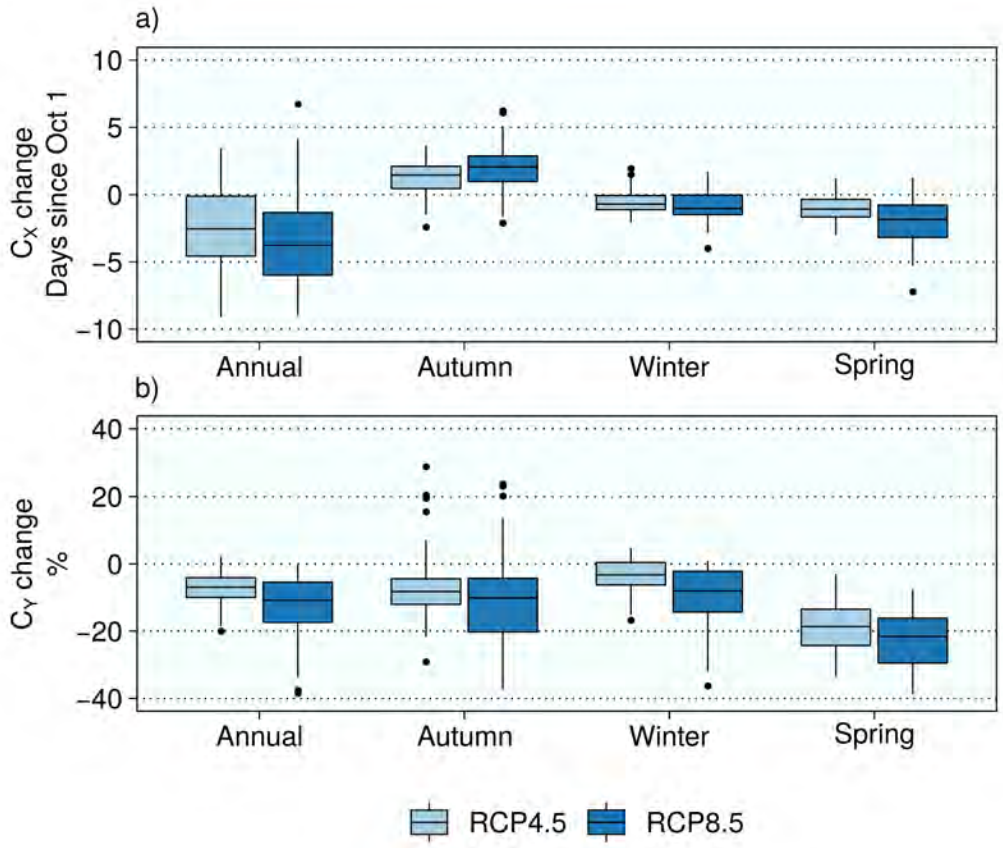


Fig. 5.10: Projected changes of the centroid position at annual and seasonal time scales for the different RCPs: a) changes of C_X calculated as the difference between future and historical values, and b) fractional changes [(future - historical)/historical · 100] of C_Y. Boxplots in a) and b) represent the 24 studied catchments.

for the Tâmega River using also a Euro-CORDEX ensemble. The projected annual PET changes are coherent with the findings of Moratiel et al. (2011) for the Duero Basin, where increases between 5% and 11% are expected for the annual PET by the end of the first half of the 21st century.

The annual delta changes of Q and AET (Fig. 5.6) corroborate the sign and the magnitude of the projected changes of annual Q and AET for the Tâmega Basin in Fonseca and Santos (2019), as well as they further extend the conclusions reached there given the greater number of catchments considered in this chapter and the higher number of members included in the Euro-CORDEX

ensemble. Nonetheless, this work is focused in the Spanish part of the basin, and even though it can be considered as representative of the entire area, future research will encompass the totality of the basin in order to overcome this limitation.

On the other hand, and similarly to this work, negative i_Q changes were found in Mendoza et al. (2015, 2016) in the Colorado River Basin. The steady-state assumption reflected in Eq. 5.3 is one well-known and widespread approach taken for the quantitative analysis of the water balance equation (e.g., Rasmussen et al., 2014; Xu et al., 2014; Liang et al., 2015; Mendoza et al., 2015, 2016; Hasan et al., 2018; Li et al., 2018). However, this assumption is rarely checked and can lead to a long-term imbalance when the storage component is not considered even for long periods (i.e. 10 to 30 years), particularly in arid and semi-arid regions (Han et al., 2020). In order to avoid feasible inaccuracies in the application of this approach, the steady-state assumption was tested for all the studied catchments, remaining the sum of i_Q and i_e close to 1 for the historical period and for all the future scenarios.

5.5.2 Projected seasonal hydrologic changes

The statistically significant positive delta changes corresponding to winter P (Fig. 5.6) point to an important feature for the future precipitations in the central catchments of the Duero Basin. Similarly, other studies using WRF projections concluded that, contrarily to the rest of the seasons, winter precipitation is projected to increase over certain areas of the Iberian Peninsula due to climate change, remarkably over the Northern Plateau, but the increases are generally non-significant (Argüeso et al., 2012; García-Valdecasas Ojeda et al., 2020a). In the same vein, Soares et al. (2017), using WRF and Euro-CORDEX ensembles, found both significant and non-significant increases of winter precipitation for some areas in the north of Portugal including the Portuguese part of the Duero River Basin.

The findings for the seasonal AET changes (Fig. 5.6) partially agree with the results reported in García-Valdecasas Ojeda et al. (2020a), where comparable changes were found during winter and summer over the study area. However, the WRF simulations carried out there diverged from our results for the rest of the hydrologic year: in autumn, the WRF projections led to significant decreases for almost all the simulations, and in spring, the partitioning between

significant increases in the northern headwaters and significant decreases in the south were not captured.

It is expected that the VIC implementation of this work reproduces more realistically the water balance and the future hydrologic changes of the Duero headwaters since it was built upon the calibration exercise developed in Yeste et al. (2020). Although the model was calibrated only using streamflow observations, its performance was also evaluated against two AET products, producing good adjustments as well and improving the benchmark performance in all cases.

Other feasible explanation can be related to the high number of models included in the Euro-CORDEX ensemble in comparison to the two GCMs that drove the WRF simulations in García-Valdecasas Ojeda et al. (2020a). Lastly, the choice of the climatological year (i.e. from December to November) in García-Valdecasas Ojeda et al. (2020a) could also conduct to some differences when compared to the results for the hydrologic year (i.e. from October to September) presented here.

The projected seasonal changes for Q (Fig. 5.6) were similar to those obtained in Fonseca and Santos (2019) for the Tâmega Basin, with downward trends for all the seasons except for winter, where a slight increase was projected. The strongest diminutions were projected for summer, where water scarcity and the increasing frequency of droughts may pose a serious threat in the future in agreement with García-Valdecasas Ojeda et al. (2021a).

Although most of the summer Q changes were characterized by marked and significant decreases, it is important to note the existence of a few significant increases (Fig. 5.9). This responds to an atypical behaviour and is presumably driven by two factors: firstly, the very nature of the low Q values in summer supposes that a higher streamflow, though remaining in the low range, produces a markedly positive delta change; and secondly, the averaging of all the Q time series when the mean of the ensemble was calculated could introduce small biases that finally led to a positive delta change in rare cases.

The results for Q_{snow} (Fig. 5.7) and $Q_{\text{snow}}\ ratio$ (Fig. 5.9) resemble the relative contribution of the snowmelt component to the generated runoff in the Ganges basin applying an identical method for estimating Q_{snow} (Siderius et al., 2013), which was expected to change as climate warms. Similarly, Ceballos-Barbancho et al. (2008) and Morán-Tejeda et al. (2010, 2011a) pointed to a change of the snow regime in the Duero River Basin during the last half of the 20th century,

with important implications for water management that already led to the adoption of different management practices in other parts of Spain (López-Moreno et al., 2004). The marked reductions observed for winter and spring streamflow were likely caused by the decrease of winter precipitation and the increase of winter and spring temperatures. The latter implies a decrease of snow accumulation in winter and an earlier snowmelt presence during spring, therefore affecting the amount and timing of the streamflow (Morán-Tejeda et al., 2010, 2011a).

This downward tendency driven by a warmer climate has being previously identified for the mountainous areas in Spain (López-Moreno et al., 2009; Morán-Tejeda et al., 2017; Collados-Lara et al., 2019) and other parts of the world (e.g., Bhatti et al., 2016; Majone et al., 2016; Coppola et al., 2018; Ishida et al., 2018, 2019; Liu et al., 2018), thus suggesting a critical role of the snowmelt component for the future management of mountain water resources (Viviroli et al., 2011; Mankin et al., 2015).

5.5.3 Annual and seasonal changes of the hydrograph centroid

The projected annual changes of C_X (Fig. 5.10a) suggest that a time shift in the hydrologic year towards earlier streamflow volumes takes place for the future scenarios. The "center time" of runoff is considered a measure of the streamflow seasonality, and is usually calculated for the average hydrologic year as a single metric for the entire hydrograph (Stewart et al., 2005; Mendoza et al., 2015, 2016). Mendoza et al. (2015) suggested that the negative sign of the projected annual changes of C_X are linked to a lesser presence of snow under climate change conditions. This is consistent with the findings of this chapter for the studied catchments, where Q_{snow} is expected to suffer the greatest burden of the impacts of climate change.

It is important to note that with the only use of the annual C_X position as a signature measure there are other important characteristics of the average daily hydrograph that remain hidden and not completely depicted. This limitation has been overcome by calculating the seasonal centroid position and its future changes (Fig. 5.10), revealing additional information about the seasonal timing of the streamflow that is expected to have a large impact on the future water management strategies for the basin.

Finally, the annual and seasonal C_Y changes (Fig. 5.10b) constitute another valuable metric that can be related to the delta changes of Q both at annual and seasonal scales (Table S5.2). The degree of correlation between them responds to the question of to which extent the changes of the shape of the hydrograph (i.e. C_Y changes) are related to the changes of the mean values (i.e. delta changes of Q). Thus, the 1:1 relationship observed for winter and spring indicates that the changes of shape are mainly driven by the delta changes of Q. This is also supported by the closeness to 0 of the C_X changes for these seasons (Fig. 5.10a), being the shape of the hydrograph directly related to the vertical shifting of the centroid. The annual and autumn changes of the centroid position show greater complexity as the C_X changes become a contributing factor to the changes of shape. In these cases the C_Y changes and the delta changes of Q are less correlated but still manifest a linear relationship with statistically significant r values. This approach generalizes the common usage of the "center time" of runoff as a measure of seasonality, and further research will explore the implications of the changes of the C_X and C_Y and their relation to the corresponding delta changes for the rest of atmospheric and land variables involved in the hydrology of the headwaters of the Duero River Basin.

5.6 Conclusions

The multi-model ensemble approach has shown to be an effective tool for the analysis of the impacts of climate change in the headwater areas of the Duero River Basin both at annual and seasonal time scales. The simulations carried out with the VIC model driven by a large number of Euro-CORDEX RCM+GCM combinations and two RCPs has permitted a posterior analysis applying the delta change method and estimating various signature measures for the different land and atmospheric variables enmeshed in the modelling exercise. The former evaluated the future changes of the mean values, and the latter addressed other important hydrologic features including the relative contribution of runoff and actual evapotranspiration to the overall water balance, the snowmelt contribution to the total streamflow and the centroid position for the daily hydrograph of the average hydrologic year. The main findings of this chapter are as follows:

- The annual streamflow decreases were driven by two different mecha-

nisms depending on the mountainous system considered. The precipitation decreases in the south part of the basin imposed a limit to the runoff and evapotranspiration processes. The streamflow reductions for the northern mountains were the outcome of a combined effect of the precipitation decreases and evapotranspiration increases in the future scenarios. The future changes of the runoff and the evapotranspiration ratios revealed a tendency towards an evaporative regime for the northern catchments, while the evaporative response strengthened in the south. The sum of both ratios remained close to 1 for all the studied cases, thus confirming the steady-state assumption usually non-tested in many previous studies.

- The precipitation and evapotranspiration changes evinced a strong intra-annual variability, and were directly related to the seasonal streamflow detriments: the precipitation decreases constituted the limiting factor for the runoff and evapotranspiration processes in summer for all the studied catchments, and over the southern part of the basin in spring; the compound effect of the precipitation reductions and the evapotranspiration increments was noticeable in autumn for the entire basin, and over the north in spring; lastly, the winter streamflow changes were mostly negative and non-significant as a consequence of the non-significant changes projected for the precipitation in this season.

- The snowmelt contribution to the total generated runoff was the hydrologic variable most affected by the climate warming over the study area. The projected changes indicated a downward tendency towards the practically non-existence of snow dominated hydrologic regimes for the headwaters of the Duero River Basin. This behaviour exacerbates the previous findings for the mountainous areas in Spain during the last half of the 20th century, and suggests a major role of this component for the future water management practices.

- The projected changes of the centroid position were estimated for the average daily hydrologic year at annual and seasonal scales, and accounted for the variations of the streamflow seasonality (i.e. horizontal shifts) and the streamflow volumes (i.e. vertical shifts). Particularly, the vertical shifts showed a strong degree of correlation to the corresponding delta changes of the streamflow, being interchangeable measures in winter and spring. This approach generalized the widespread use of the "center time" of runoff as a signpost of seasonality as many other key features

were well-captured and fully explained, and can be further applied for the rest of atmospheric and land variables involved in the modelling exercise.

6 Streamflow and evaporation trade-offs: sensitivity analysis and calibration based on Pareto optimization

The content of this chapter is based on a manuscript submitted for publication:

A Pareto-based sensitivity analysis and calibration approach for integrating streamflow and evaporation data

Patricio Yeste, Lieke A. Melsen, Matilde García-Valdecasas Ojeda, Sonia R. Gámiz-Fortis, Yolanda Castro-Díez and María Jesús Esteban-Parra

Submitted to *Water Resources Research*

Abstract

Evaporation is gaining increasing attention as a calibration and evaluation variable in hydrologic studies that seek to improve the physical realism of hydrologic models and go beyond the long-established streamflow-only calibration. However, this trend is not yet reflected in sensitivity analyses aimed at determining the relevant parameters to calibrate, where streamflow has traditionally played a leading role. On the basis of a Pareto optimization approach, we propose a framework to integrate the temporal dynamics of streamflow and evaporation into the sensitivity analysis and calibration stages of the hydrological modelling exercise, here referred to as "Pareto-based

sensitivity analysis" and "multi-objective calibration". The framework is successfully applied to a case study using the Variable Infiltration Capacity (VIC) model in three catchments located in Spain as representative of the different hydroclimatic conditions within the Iberian Peninsula. Several VIC vegetation parameters were identified as important to the performance estimates for evaporation during sensitivity analysis, and therefore were suitable candidates to improve the model representation of evaporative fluxes. Sensitivities for the streamflow performance, in turn, were mostly driven by the soil and routing parameters, with little contribution from the vegetation parameters. The multi-objective calibration experiments were carried out for the most parsimonious parameterization after a comparative analysis of the performance gains and losses for streamflow and evaporation, and yielded optimal adjustments for both hydrologic variables simultaneously. Results from this chapter will help develop a better understanding of the trade-offs resulting from the joint integration of streamflow and evaporation data into modelling frameworks.

6.1 Introduction

Terrestrial water storage and water fluxes modulating the land-atmosphere interaction are key components of the global water cycle and energy budget. Water storage is in direct connection with the partitioning of precipitation into runoff and evaporation (Lehner et al., 2019; Pokhrel et al., 2021), the latter currently accounting for two thirds of global precipitation (Good et al., 2015). The intensification of the global water cycle driven by an already changing climate represents a challenge for future water security (Lehner et al., 2019), and is likely to affect the occurrence of extreme events such as droughts and floods (Blöschl et al., 2020; Miralles et al., 2019; Peterson et al., 2021). Moreover, evaporation is projected to increase in the context of global warming (IPCC, 2021), and its changes can pose an important threat for water resources availability and the biosphere (Konapala et al., 2020; Koppa et al., 2022).

In this regard, the use of satellite-based and reanalysis evaporation data as well as in-situ measurements allows for a better understanding of other relevant spatio-temporal processes, particularly in data-scarce areas (Dembélé et al., 2020a; López López et al., 2017) and in arid and semi-arid regions where evaporative fluxes are dominant (Dembélé et al., 2020b; Koppa et al., 2022). The growing availability of evaporation products provides an unprecedented opportunity to improve the representation of the land-atmosphere interactions in hydrologic models (Lettenmaier et al., 2015; Minville et al., 2014; Puertes et al., 2019; Széles et al., 2020), effectively overcoming the inherent limitations of the streamflow-only calibration (Dembélé et al., 2020a,b; Koppa et al., 2019). The incorporation of multiple data sources into the calibration stage permits to identify meaningful parameter combinations and helps achieve a more realistic representation of the hydrologic cycle (Fowler et al., 2018; Gharari et al., 2013; Koppa et al., 2019; Nijzink et al., 2018; Rakovec et al., 2016a; Zhang et al., 2020). This is particularly relevant for a field where streamflow has dominated calibration endeavors (Becker et al., 2019; Dembélé et al., 2020a,b) and evaporation has been considered a subproduct of the steady-state assumption for closed hydrologic systems (Han et al., 2020; Liang et al., 2015; Yang et al., 2017).

Although the use of evaporation data is becoming increasingly extended as an evaluation variable during a post-calibration phase (e.g., Bouaziz et al., 2021; Rakovec et al., 2016b,a, 2019; Yeste et al., 2020, 2021) or as a calibration objective itself (e.g., Dembélé et al., 2020a,b; Demirel et al., 2018; Koppa

et al., 2019; López López et al., 2017; Nijzink et al., 2018; Széles et al., 2020), evaporation-related parameters controlling the vegetation processes have remained mostly unattended during sensitivity analysis. This is particularly relevant in the light of results reported in Sepúlveda et al. (2022) for the Variable Infiltration Capacity (VIC) model (Liang et al., 1994, 1996), one of the most widely used models by the hydrological community (Addor and Melsen, 2019), and suggest that, among others, VIC vegetation parameters have strong potential to improve model performance despite their sensitivities have been largely unknown for decades.

In this chapter we propose a framework focused on the integration of the temporal dynamics of streamflow and evaporation into the sensitivity analysis and calibration stages, which is then applied to a case study using VIC in three Spanish catchments. The hydroclimatic characterization of these catchments (Yeste et al., 2018, 2020, 2021) together with the aridity conditions specific of the Iberian Peninsula (García-Valdecasas Ojeda et al., 2020a, 2021a) make them ideal candidates for the matter in question.

First, we revisit the definition of the multi-objective optimization problem and outline how it can be addressed in practice. This will help prepare the ground for introducing the followed approach, which is then applied to the case study. The sound performance of VIC and the broad implications derived from this work suggest a clear contribution to the body of multiple-criteria applications in hydrology, as well as to the knowledge of evaporation and its integration into the hydrological modelling exercise.

6.2 Multi-objective optimization problem

Multi-objective optimization aims to optimize multiple objectives at the same time. Without loss of generality, the multi-objective optimization problem (MOOP) is commonly formulated as a minimization problem (Miettinen, 1998), and can be formally written as follows (Deb, 2001):

$$\left.\begin{array}{lll} \text{Minimize} & f_m(\mathbf{x}) & m = 1, 2, ..., M \\ \text{subject to} & g_j(\mathbf{x}) \leq 0 & j = 1, 2, ..., J \\ & h_k(\mathbf{x}) = 0 & k = 1, 2, ..., K \\ & x_i \in [x_i^{(L)}, x_i^{(U)}] & i = 1, 2, ..., N \end{array}\right\} \tag{6.1}$$

where M is the number of objective functions f_m to minimize. Each component x_i of the n-dimensional decision vector $\mathbf{x} = (x_1, x_2, ..., x_N)$ must belong to the interval defined by the corresponding upper and lower bounds $x_i^{(L)}$ and $x_i^{(U)}$. The MOOP can be subject to J inequality and K equality constraints applied to the constraint functions g_j and h_k, respectively.

By definition, a conflict between the objectives leads to a trade-off in Pareto-optimal solutions that are equally important or likely from an optimization perspective, conforming the so-called Pareto front (Deb, 2001). As none of the solutions can be considered to be superior when there is more than one objective to minimize, Pareto-optimal solutions are also referred to as non-dominated solutions. The MOOP then becomes a multi-criteria decision making problem where only a small subset of solutions are selected for the specific application based on expert knowledge or the use of decision making techniques (Miettinen, 1998).

Classical approaches to solve a MOOP rely on the scalarization of the multiple objectives into a composite single-objective function. The most common form of scalarization involves the introduction of some preference information for the objectives through weighting coefficients, which can be changed in order to obtain alternative Pareto-optimal solutions (Deb, 2001; Miettinen, 1998). Among these techniques, the weighted metric method represents the most general problem for minimizing distances (Deb, 2001):

$$\text{Minimize} \quad \left(\sum_{m=1}^{M} w_m |f_m(\mathbf{x}) - z_m^*|^p \right)^{1/p} \tag{6.2}$$

where w_m is the weight for the objective function f_m, z_m is the m-th component of the ideal vector z^* (usually the null vector) and p is a power parameter. The weights are typically selected such that $w_m \geq 0$ and their sum is one. The problem can be subject to the same constraints as in (6.1), and the parameter p can take values between 1 and ∞. Notably, when $p = 2$ the problem corresponds to minimizing the weighted Euclidean distance to the ideal solution z^*. In the particular case of $p = 2$ and M equal weights, the problem is equivalent to minimizing the pure Euclidean distance.

A major disadvantage of classical methods is that only one solution is handled at a time (Deb, 2001). In the event of non-convex and discontinuous Pareto fronts, this unavoidably results in an incomplete representation of the optimal solutions region (Miettinen, 1998). By contrast, multi-objective optimization

evolutionary algorithms (MOEAs) follow a metaheuristic approach that effectively handles a population of candidate solutions in a single run, making them suitable for nonlinear optimization problems regardless of any continuity or convexity condition (Deb, 2001; Coello Coello, 2005). MOEAs are then left with the task of converging to the Pareto front and achieving a well-distributed set of solutions (Abraham and Jain, 2005; Deb, 2001; Laumanns, 2005; Yang et al., 2014). Metaheuristic approaches have become the way forward to tackle multi-objective optimization challenges in many areas of science and engineering, and the efforts to improve and develop them continue to the present day (Blank and Deb, 2020; Falcón-Cardona et al., 2021; Hernandez-Suarez et al., 2021; LaTorre et al., 2021).

6.3 Approach definition

In this section, the methodological steps towards our goal of integrating streamflow and evaporation data into the model optimization process are presented. These steps are schematized in Fig. 6.1, and their description is provided in the following subsections. The core idea underlying the rationale of this framework consists in the application of the MOOP definition to two objectives that characterize the overall goodness-of-fit of the model regarding the temporal dynamics of the streamflow and evaporation data, respectively. In particular, we propose a multi-objective optimization approach that can serve the purpose of both a sensitivity analysis and a calibration procedure that are performed in tandem.

But first, it is essential to verify if the problem is affordable from a physical perspective. As we will show below, the integration of both data sources into the modelling framework is largely dependent on the law of conservation of mass and its steady-state representation for the hydrologic system under examination.

6.3.1 Verification for streamflow and evaporation data

Over long periods, the water balance equation can be written as follows after normalizing by precipitation (Han et al., 2020; Mendoza et al., 2015, 2016; Rasmussen et al., 2014; Yeste et al., 2021):

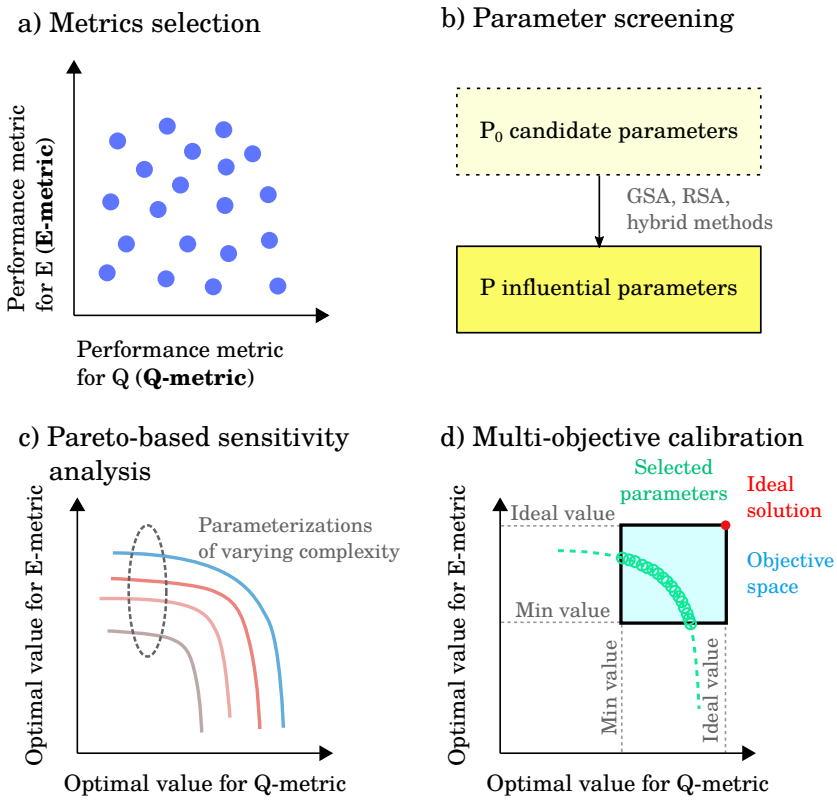


Fig. 6.1: Schematic overview of the framework. (a) Metrics selection for streamflow and evaporation based on model performance. (b) Parameter screening to identify influential parameters. (c) Pareto-based sensitivity analysis for parameterizations of varying complexity. (d) Multi-objective calibration for the final selection of parameters.

$$1 = \frac{\overline{Q}}{\overline{P}} + \frac{\overline{E}}{\overline{P}} \tag{6.3}$$

where $\overline{P}$ [L/T] is the mean annual precipitation for the period, $\overline{Q}$ [L/T] the mean annual streamflow and $\overline{E}$ [L/T] is the mean annual evaporation.

Eq. 6.3 shows that different model simulations forced with same input P data can reach the steady-state for different contribution levels of Q and E. Hence, for a long enough period, an imbalance in the above equation when

calculated for the reference data constitutes a potential condition for a trade-off between the objectives (i.e., a Q-E trade-off) since the model will likely produce a steady-state solution. If the imbalance is significant, the integration of the streamflow and evaporation datasets could become unsuitable due to a failure to comply with the law of conservation of mass. For small to negligible imbalances, however, there is no a priori condition for a Q-E trade-off. In that case, the trade-off will generally depend on the particular application and the (in)ability of the model to simultaneously reproduce both datasets. Eq. 6.3 will be thoroughly checked for the case study.

6.3.2 Metrics selection and parameter screening

The *metrics selection* step (Fig. 6.1a) is based on an preliminary evaluation of model performance by means of several candidate performance metrics for Q and E. These metrics are expressed as single scalar values aggregating different statistical properties of model performance, and can be calculated at different time scales (e.g., daily, monthly). This stage requires a large-enough and representative sample of model responses, which can be achieved via a Monte Carlo experiment specifically designed to explore parameter space. As a result, a Q-metric and an E-metric are selected from the proposed candidates according to their adequacy to the case study.

Next, a *parameter screening* procedure (Fig. 6.1b) is carried out to identify influential and non-infuential parameters to the selected metrics. Sensitivity analysis methods play a major role in this respect as they are traditionally used to this end. In this way, global sensitivity analysis (GSA), regional sensitivity analysis (RSA) and hybrid local-global methods are equally valid approaches that can be conducted for that purpose.

6.3.3 Pareto-based sensitivity analysis and multi-objective calibration

The last two steps are founded on a MOOP implementation for the selected metrics. In the first place, the parameter screening procedure is further refined in a Pareto setting. We call this step *Pareto-based sensitivity analysis* (Fig. 6.1c) since it relies on the MOOP definition and is a continuation of the previously

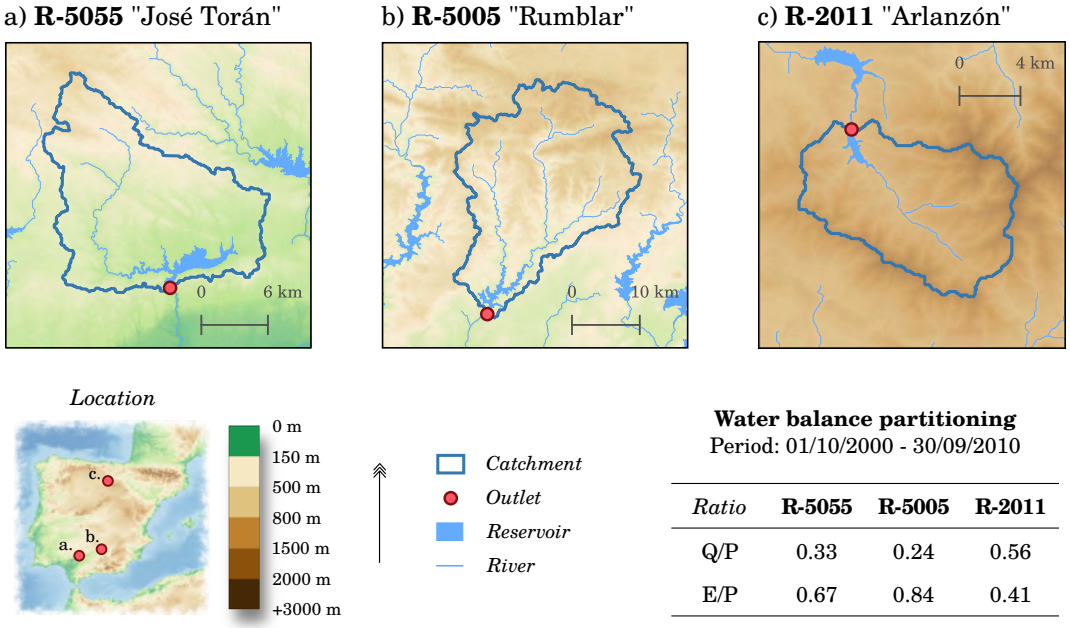
applied sensitivity analysis that aims to find the most parsimonious parameterization leading to satisfactory results. This is attained through unconstrained multi-objective optimization experiments targeted at parameterizations of varying complexity as follows: starting from a reference parameterization that covers the minimum number of parameters to be analyzed, the rest of the parameters are included one at a time. All the parameters are then incorporated into a full parameterization that reflects the maximum achievable complexity for the given implementation, and one of the parameterizations is selected based on the relative gain/loss in performance for both objectives. Optimal model performance is finally determined for the selected parameterization through a constrained *multi-objective calibration* exercise (Fig. 6.1d) focused on a specific region of the objective space in order to discard unfeasible solutions leading to poor performance estimates.

6.4 Case study

This section begins with the presentation of the area and data and the hydrologic model that integrate the case study, and is then followed by a description of how the four steps constituting our framework are implemented for this particular case.

6.4.1 Area and data

In order to assess the applicability of the proposed framework, the approach was implemented for three catchments in Spain corresponding to the following stations (Fig. 6.2): station R-5055 "José Torán" (surface 247.4 km^2, annual P 790 mm/year), R-5005 "Rumblar" (surface 573.3 km^2, annual P 550 mm/year), and R-2011 "Arlanzón" (surface 106.8 km^2, annual P 1320 mm/year). Stations R-5055 and R-5005 belong to the Guadalquivir River Basin, the southernmost and one of the most arid basins in Spain, while station R-2011 is located in the Duero River Basin, a transboundary basin that flows into northern Portugal, and the biggest basin of the Iberian Peninsula. The Guadalquivir and Duero River Basins perfectly represent the latitudinal gradient of the climate variability for the Iberian Peninsula, and therefore are appropriate candidates to test the spatial consistency of the applied methods. We refer to Yeste et al.



Water balance partitioning
Period: 01/10/2000 - 30/09/2010

Ratio	**R-5055**	**R-5005**	**R-2011**
Q/P	0.33	0.24	0.56
E/P	0.67	0.84	0.41

Fig. 6.2: Location of the three studied catchments.

(2018, 2020, 2021) and references therein for a comprehensive hydroclimatic characterization of these basins.

Hydrologic data in this chapter were assembled for a period of 10 hydrologic years spanning from Oct 2000 to Sep 2010. Daily precipitation and temperature observations were extracted from SPREAD and STEAD (Serrano-Notivoli et al., 2017, 2019, see Section 2.2 in Chapter 2). Daily streamflow time series for the studied catchments were acquired from the SAIH-ROEA dataset (see Section 2.1 in Chapter 2), and evaporation data were gathered from the daily of GLEAM version 3.5a (Martens et al., 2017; Miralles et al., 2011, see Section 2.3 in Chapter 2). Additional meteorological forcings (i.e., wind speed, radiation, atmospheric pressure, vapor pressure), were collated from WRFERA simulations for the spatial domain of the Iberian Peninsula in García-Valdecasas Ojeda et al. (2017).

The ratios of streamflow and evaporation to precipitation ($\overline{Q}/\overline{P}$ and $\overline{E}/\overline{P}$) corresponding to the study period were computed for all the catchments and produced the following estimates: 0.33 and 0.67 for R-5055; 0.24 and 0.84

for R-5005, and 0.56 and 0.41 for R-2011. These values do not introduce an artificial imbalance as their sum remains close to 1, and therefore guarantee the further applicability of the framework (see Section 6.3.1). The high values of $\overline{E}/\overline{P}$ for the three catchments are symptomatic of the aridity conditions of the Guadalquivir and Duero River Basins, and thus reinforce the potential of including evaporation information into the modelling framework (Dembélé et al., 2020b; Koppa et al., 2022).

6.4.2 Parameter selection

A set of 20 parameters representing the soil, vegetation and routing processes was subsequently selected for the parameter screening procedure, and comprised 5 soil parameters, the 2 routing parameters and 13 vegetation parameters (Table 6.1). Although the soil and routing parameters considered here are commonly used in other studies for sensitivity analysis and/or calibration purposes (e.g., Chawla and Mujumdar, 2015; Mizukami et al., 2017; Oubeidillah et al., 2014; Rakovec et al., 2014, 2019; Yeste et al., 2020), the vegetation parameters remain mostly fixed as default values and are rarely varied during the hydrological modelling exercise (see Melsen et al., 2016a; Mendoza et al., 2015; Sepúlveda et al., 2022, for some exceptions).

The values of the vegetation parameters were modified by means of multiplication factors varying from 0.5 to 1.5. Multiplication factors confer greater flexibility on the VIC parameterization of the vegetation processes, and were applied to both the single-valued parameters and those with monthly variations (see Table 6.1). This strategy is comparable to the use of spatial multipliers to calibrate distributed hydrologic models (e.g., Bandaragoda et al., 2004; Francés et al., 2007; Pokhrel and Gupta, 2010; Mendoza et al., 2015), and allows for studying the potential of the VIC vegetation parameters as calibration targets (Sepúlveda et al., 2022).

6.4.3 Performance metrics

In this work, the Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) was chosen to evaluate the goodness-of-fit of the model, and the metrics for streamflow and evaporation were selected according to the relative performance of NSE at daily and monthly time scales (Fig. 6.1a). NSE can be expressed as:

Table 6.1: Parameters included in the parameter screening procedure. Parameters denoted with subscript "f" correspond to the dimensionless multiplication factors used to modify VIC vegetation parameters.

Parameter	Units	Min value	Max value	Description
b_i	-	10^{-5}	0.4	Variable infiltration shape parameter
D_S	-	10^{-9}	1	Fraction of D_m where non-linear baseflow begins
W_S	-	10^{-9}	1	Fraction of the porosity of the bottom soil layer where non-linear baseflow begins
D_m	mm/day	10^{-9}	30	Maximum baseflow
d_2	m	0.1	0.9	Thickness of soil layer 2
$rout_1$	-	10^{-9}	10	Shape parameter of gamma function
$rout_2$	days	10^{-9}	2	Scale parameter of gamma function
$depth_{1f}$	-	0.5	1.5	Thickness of root zone for layer 1
$depth_{2f}$	-	0.5	1.5	Thickness of root zone for layer 2
$rarc_f$	-	0.5	1.5	Architectural resistance
$rmin_f$	-	0.5	1.5	Minimum stomatal resistance
LAI_f	-	0.5	1.5	Leaf-area index
$albedo_f$	-	0.5	1.5	Albedo
$rough_f$	-	0.5	1.5	Vegetation roughness
$disp_f$	-	0.5	1.5	Vegetation displacement
$wind\_h_f$	-	0.5	1.5	Height of wind speed measures
RGL_f	-	0.5	1.5	Minimum incoming shortwave radiation for transpiration
rad_atten_f	-	0.5	1.5	Radiation attenuation
$wind\_atten_f$	-	0.5	1.5	Wind speed attenuation through overstory
$trunk\_ratio_f$	-	0.5	1.5	Ratio of total tree height that is trunk

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}(y_\mathrm{t} - y_\mathrm{ref,t})^2}{\sum_{t=1}^{T}(y_\mathrm{ref,t} - \bar{y}_\mathrm{ref})^2} \tag{6.4}$$

where y_t is the model output and $y_\mathrm{ref,t}$ is the reference value of variable y for time step t, and $\bar{y}_\mathrm{ref}$ is the mean value of reference data for the study period. NSE is a multiple-criteria estimator of model performance, and can be also formulated as an algebraic decomposition into three components representing different statistical properties of model behaviour (Clark et al., 2021; Gupta et al., 2009; Knoben et al., 2019; Lamontagne et al., 2020):

$$\mathrm{NSE} = 2\alpha r - \alpha^2 - \frac{(\beta - 1)^2}{\mathrm{CV}_{\mathrm{ref}}^2} \tag{6.5}$$

here, r is the coefficient of correlation between simulations and reference data, α is the ratio of the standard deviations, β the ratio for mean values, and $\mathrm{CV}_{\mathrm{ref}}$ is the coefficient of variation of the reference data. α and β report, respectively, on the overestimation (> 1) or underestimation (< 1) of the variability and the mean value of the analyzed variable. $\mathrm{CV}_{\mathrm{ref}}$ is model-independent, and consequently NSE can be seen as an interplay between r, α and β (see Gupta et al., 2009; Knoben et al., 2019, for a detailed description of these statistics and their relationships). From the definition of β, two simulations satisfying Eq. 6.3 will produce the following equality for the relative changes of β_{Q} and β_{E}:

$$\frac{\Delta\beta_{\mathrm{E}}}{\Delta\beta_{\mathrm{Q}}} = \frac{\Delta\overline{E}/\overline{E}_{\mathrm{ref}}}{\Delta\overline{Q}/\overline{Q}_{\mathrm{ref}}} = -\frac{\overline{Q}_{\mathrm{ref}}}{\overline{E}_{\mathrm{ref}}} \tag{6.6}$$

where $\overline{Q}_{\mathrm{ref}}$ and $\overline{E}_{\mathrm{ref}}$ are the mean annual values of the Q and E reference data, respectively. Then, simulations achieving the steady-state will manifest a data-driven trade-off for β_{Q} and β_{E} as a straight line with negative slope that only depends on reference data. The implications of selecting NSE as a performance metric and its decomposition into r, α and β will be discussed in the light of results.

6.4.4 Monte Carlo experiment and DELSA sensitivity analysis

A Monte Carlo experiment was conducted for each of the studied catchments using a Latin Hypercube sample (LH sample) of 10,000 parameter combinations for the 20 parameters listed in Table 6.1. NSE was calculated at daily and monthly time scales for the streamflow and evaporation simulations, and the best performing metric in each case was selected based on a comparative study. Parameter sensitivities with respect to the two selected metrics were studied by applying the Distributed Evaluation of Local Sensitivity Analysis (DELSA) method (Rakovec et al., 2014), and the most influential parameters were selected for each catchment.

DELSA is a hybrid local-global method that estimates local gradients at several locations across parameter space. Each parameter is individually perturbed by a fixed amount at every location (usually 1%, see Melsen et al. (2016a); Melsen and Guse (2019); Rakovec et al. (2014); Sepúlveda et al. (2022)) so as to estimate the first-order partial derivatives, which are an indicator of sensitivity. In this work, 1% perturbations were carried out for a population of 500 points (hereafter referred to as DELSA sample) that were extracted from the LH sample based on their proximity to the ideal case of NSE = 1 for both streamflow and evaporation. This initialization strategy is convenient insofar as it allows for estimating parameter sensitivities in the region of interest from an optimization perspective, thus making additional use of the outputs from the Monte Carlo experiment.

6.4.5 Multi-objective optimization strategy and single-objective experiments

In terms of eligibility, the approach followed in this chapter is open to any multi-objective optimization algorithm and customization strategy one may apply. In our case study, we selected two optimization algorithms: the Non-Dominated Sorted Genetic Algorithm II (NSGA-II, Deb et al., 2002) and the Multi-Objective Flower Pollination Algorithm (MOFPA, Yang et al., 2014). NSGA-II is a well-established evolutionary algorithm that has been extensively used in many optimization endeavors given its fast and efficient handling of a evolving population of solutions (Ercan and Goodall, 2016). MOFPA is a nature-inspired algorithm that mimics self-pollination effects of flowering plants and cross-pollination processes over long distances carried out by pollinators obeying a Lévy flight behaviour (Pavlyukevich, 2007; Yang and Deb, 2010; Yang, 2012).

The combination of two or more algorithms to approximate the Pareto front attempts to overcome the inability of one single optimizer to invariably perform well regardless of the problem to be solved (also known as the "No Free Lunch" theorem, see Wolpert and Macready, 1997), and constitutes an effective measure to ameliorate this problem (Faris et al., 2016; Sun et al., 2020; Vrugt et al., 2009) and a suitable method to explore complex objective spaces (Maier et al., 2014; Vrugt and Robinson, 2007).

On the other hand, the nature of such an approach implies that there is a

fine line between the Pareto-based sensitivity analysis and the multi-objective calibration, it being the use (or not) of constraints. Constraints applied during the multi-objective calibration exercise were handled through a parameterless penalty approach (Deb, 2000), and were defined according to the results obtained in the prior step.

Despite being commonly ignored in multi-criteria applications, constraints remain a key aspect of optimization (Blank and Deb, 2020). Indeed, setting limits of acceptability for the objectives ensures that optimal solutions cover exclusively the region of interest of the objective space, and has shown to improve the streamflow representation under changing climate conditions (Fowler et al., 2018) as well as the possibility to effectively incorporate different water fluxes in the calibration framework (Koppa et al., 2019).

Finally, five single-objective experiments were conducted for station R-5055 using the Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al., 1994) in order to benchmark the relative performance of different calibration strategies. These experiments were based on the weighted metric method (Eq. 6.2) and included: a streamflow-only calibration, an evaporation-only calibration, and three weighted Euclidean distance experiments. Table 6.2 summarizes the different calibration strategies evaluated, and the implementational details of the employed algorithms are provided in Text S6.1 in Supporting Information.

Table 6.2: Multi-objective and single-objective calibration experiments. Single-objective experiments were implemented only for station R-5055.

Experiment	**Number of objectives**	**Objective(s) to minimize**
Multi-objective calibration	2	$1-\text{NSE}(Q_d)$, $1-\text{NSE}(E_m)$
Evaporation-only calibration	1	$1-\text{NSE}(E_m)$
Weighted Euclidean distance	1	$\sqrt{0.75 \cdot (1-\text{NSE}(E_m))^2 + 0.25 \cdot (1-\text{NSE}(Q_d))^2}$
Weighted Euclidean distance	1	$\sqrt{0.50 \cdot (1-\text{NSE}(E_m))^2 + 0.50 \cdot (1-\text{NSE}(Q_d))^2}$
Weighted Euclidean distance	1	$\sqrt{0.25 \cdot (1-\text{NSE}(E_m))^2 + 0.75 \cdot (1-\text{NSE}(Q_d))^2}$
Streamflow-only calibration	1	$1-\text{NSE}(Q_d)$

6.5 Results

For the sake of clarity, the results for station R-5055 have been chosen as the guiding thread facilitating the presentation of this section. Figures corresponding to stations R-5005 and R-2011 are collected in Supporting Information, and will be appropriately referred to throughout the section.

6.5.1 Metrics selection and DELSA sensitivity measures

First, a Monte Carlo experiment was conducted to explore parameter space for the LH sample (Fig. 6.3). The metrics selection was carried out based on the relative performance of VIC at daily and monthly time scales both for streamflow (Fig. 6.3a) and evaporation (Fig. 6.3b). The model yielded daily NSE(Q) (NSE(Q_d) hereinafter) values above 0.6 (Fig. 6.3a), suggesting that NSE(Q_d) was an appropriate metric to evaluate the streamflow performance. On the contrary, NSE(E) reflected poorer results at daily scale, with monthly estimates (NSE(E_m) hereinafter) always above their daily counterparts (Fig. 6.3b). In this case, NSE(E_m) was chosen as the metric of interest, which is also in line with GLEAM not being an observational dataset. NSE(Q_d) and NSE(E_m) were also selected for the other two catchments (Fig. S6.1, S6.7).

The DELSA sensitivity analysis method was subsequently applied to study how model parameters affected the selected metrics. The DELSA sample was obtained from the joint representation of both metrics using a radial filter over the LH sample in order to extract the 500 points closest to the top-right corner of the scatterplot (Fig. 6.3c). This was done to investigate sensitivity only in the region representing the highest model performance. As shown in Fig. 6.3d, the 1% perturbations for the DELSA sample led to NSE scores close to those corresponding to the original 500 points, thus ensuring a suitable coverage of this area for the parameter screening procedure.

The DELSA first-order sensitivity measures for streamflow and evaporation are depicted in Fig. 6.4. Eleven out of the 20 parameters evaluated with DELSA were identified as important with respect to at least one of the performance metrics. This included the five soil parameters (b_i, D_S, W_S, D_m, d_2), the two routing parameters ($rout_1$, $rout_2$) and four vegetation parameters ($depth_{1\mathrm{f}}$, $depth_{2\mathrm{f}}$, $rmin_\mathrm{f}$, LAI_f). Among soil parameters, D_S and D_m evoked the strongest model response in terms of NSE(Q_d) and NSE(E_m) (Fig. 6.4a,b). The two routing

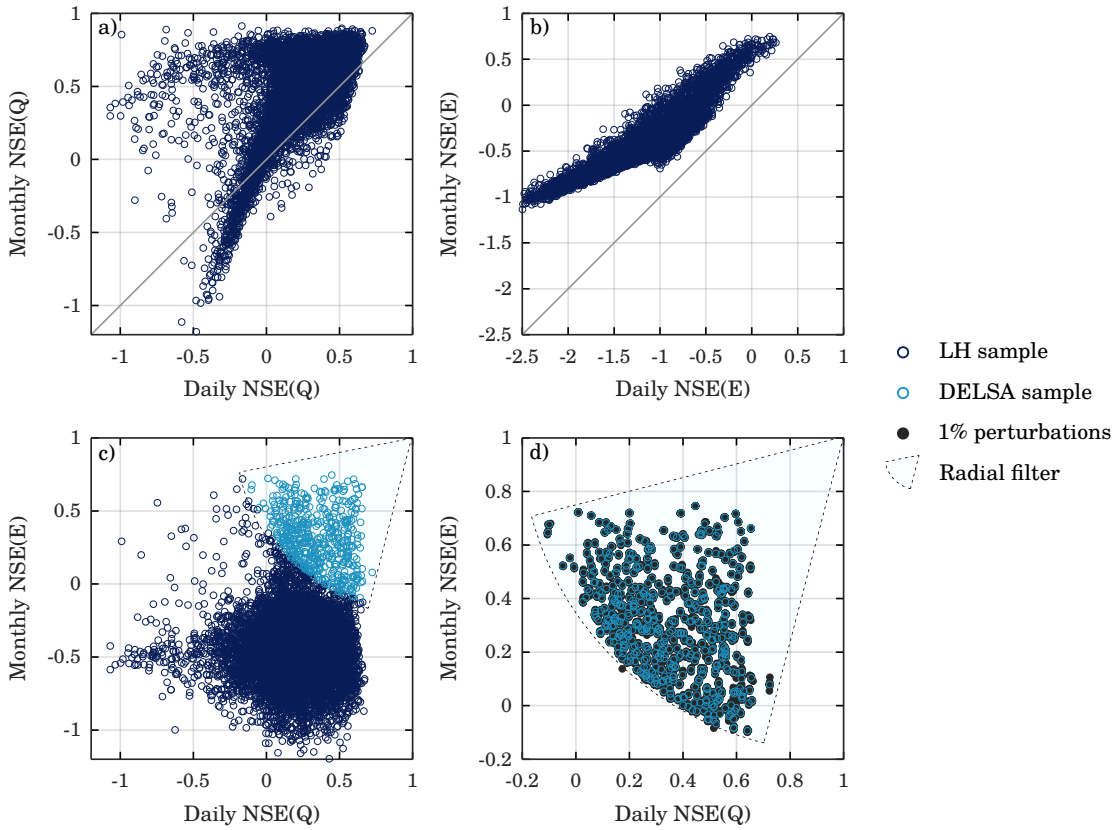


Fig. 6.3: Results of the Monte Carlo experiment for station R-5055 to study parameter space. (a) Daily and monthly estimates of NSE(Q) for the Latin Hybercube sample (LH sample). (b) Daily and monthly estimates of NSE(E) for the LH sample. (c) Joint representation of the selected performance metrics, NSE(Q_d) and NSE(E_m), and radial filter to extract the closest 500 points to the ideal solution (DELSA sample). (d) DELSA sample and the 1% perturbations where DELSA sensitivity analysis is to be performed.

parameters were the most important parameters to NSE(Q_d), while the absence of a feedback mechanism between the VIC model structure and the routing scheme implied a null effect on NSE(E_m). The four vegetation parameters only affected the sensitivity measures for NSE(E_m), although their effects were substantially lower compared to D_S and D_m. As for the other catchments, the soil and routing parameters produced similar sensitivities to that obtained for R-5055, while the number of important vegetation parameters varied from two for R-5005 (Fig. S6.2) to five for R-2011 (Fig. S6.8).

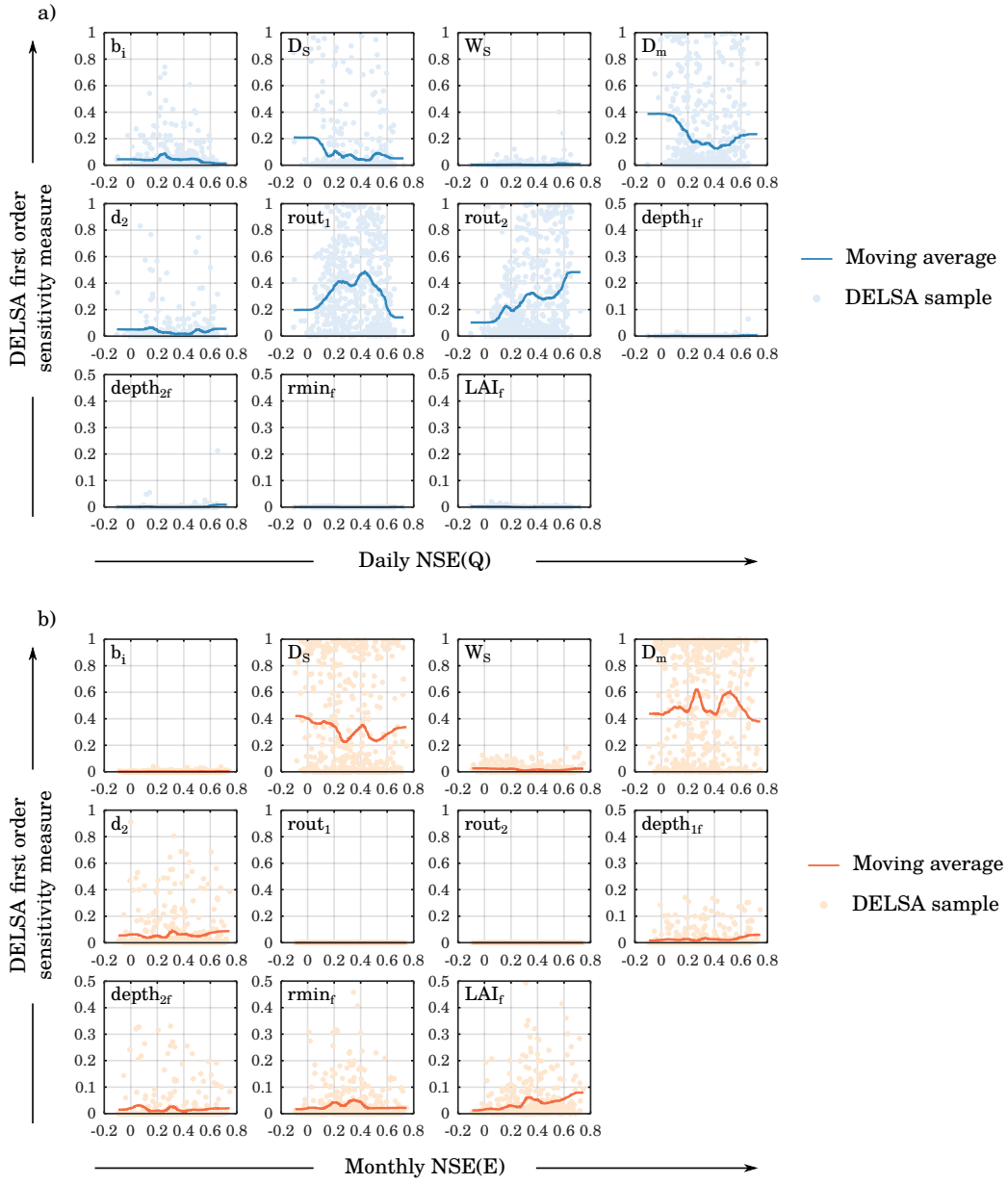


Fig. 6.4: DELSA first order sensitivities against model performance for station R-5055 computed for streamflow (a) and evaporation (b). Only the most influential parameters for at least one of the selected metrics are shown. The moving average has been depicted on top of the DELSA sample in order to facilitate the screening procedure.

6.5.2 Pareto-based sensitivity analysis

As shown in Fig. 6.5a, a total of six different parameterizations were evaluated during this stage by progressively increasing the degrees of freedom of VIC when optimized in a Pareto setting: firstly, a reference parameterization was defined as the combination of the five soil parameters and the two routing parameters (SR parameterization); secondly, the SR parameterization was further extended by including each of the four vegetation parameters one at a time (SR+$depth_{1f}$, $SR + depth_{2f}$, SR+$rmin_f$ and SR+LAI_f); finally, the full parameterization encompassed all the 11 parameters identified with DELSA. An analogous approach was followed during the Pareto-based sensitivity analysis for the other two catchments (Fig. S6.3a, S6.9a).

The Pareto fronts obtained for the SR parameterization and the full parameter-

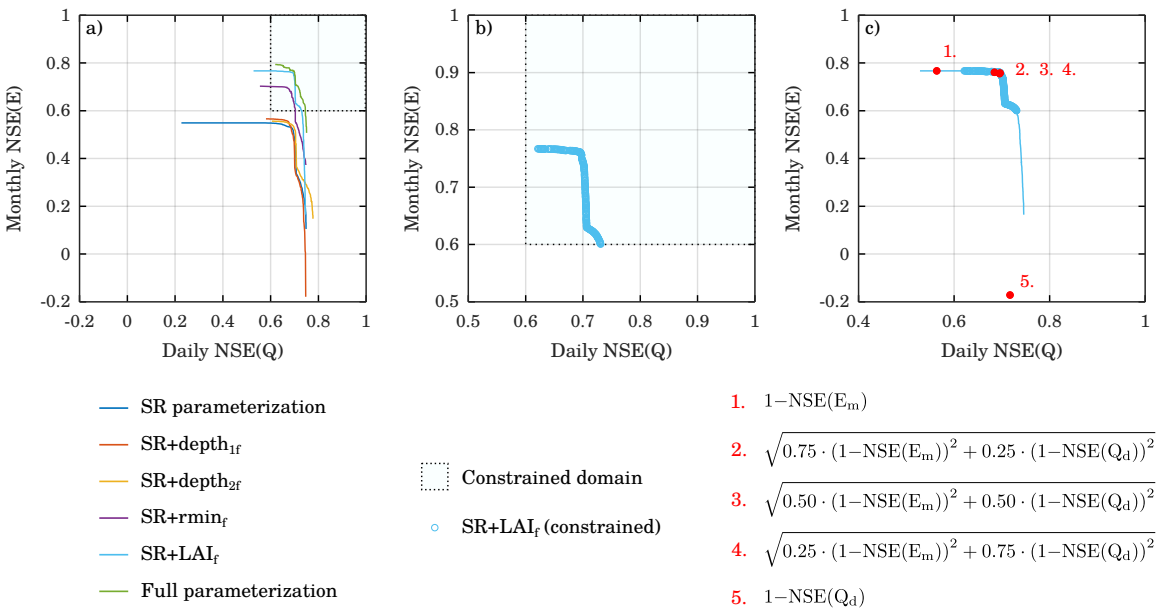


Fig. 6.5: Results of the Pareto-based sensitivity analysis and the multi-objective calibration for station R-5055. (a) Pareto fronts corresponding to the six parameterizations evaluated during the Pareto-based sensitivity analysis. (b) Pareto front resulting from the constrained multi-objective calibration for the selected parameterization. (c) Comparative view of the optimal solutions obtained with the Pareto optimization exercises and the five single-objective experiments for the objectives indicated below the panels (see also Table 6.2).

ization represent the low-end and top-end performances that can be attained by VIC for each individual implementation. The greatest increase in model performance when including all the parameters was achieved for NSE(E_m) in all the catchments, while the NSE(Q_d) improvement was less pronounced. This is supported by the results of DELSA, as the SR parameterization already contains the most important parameters to NSE(Q_d) (Fig. 6.4a, S2a, S8a).

Other Pareto fronts demonstrate the individual effects of the vegetation parameters added to the SR parameterization. It is visible for R-5055 that the addition of the LAI_f parameter produced a substantial increase in performance for NSE(E_m) close to the full parameterization. The SR+LAI_f parameterization was consequently adopted during the multi-objective calibration exercise (Fig. 6.5b) for it exhibited the biggest performance enhancement and only required one extra degree of freedom with respect to the SR parameterization. Similarly, both LAI_f and $rmin_f$ were added to the SR parameterization for R-5005 (Fig. S6.3b), and $depth_{1f}$ and $depth_{2f}$ in the case of R-2011 (Fig. S6.9b).

6.5.3 Multi-objective calibration and predictive uncertainty

The calibration of the VIC model was carried out through a constrained Pareto optimization based on the SR+LAI_f parameterization (Fig. 6.5b). The limits of acceptability for both NSE(Q_d) and NSE(E_m) were set at 0.6 based on the unconstrained optimization results. The predictive intervals for the constrained Pareto optimization showed a good coverage of the streamflow observations and GLEAM data (Fig. 6.6), highlighting the capability of VIC to reproduce both variables simultaneously. The constrained Pareto optimization for R-5005 was done subject to NSE values greater than 0.5 (Fig S6.3b), and the solutions also reflect a good adjustment to streamflow and evaporation data (Fig. S6.4). In the case of R-2011, the constraints were set at 0.6, resulting in a less pronounced Q-E trade-off (Fig. S6.9b) and a narrower predictive interval (Fig. S6.10).

To provide a more complete picture of the multi-objective calibration results, the Pareto front obtained was divided into three clusters representing different model behaviours for which the parameters distributions were compared (Fig. 6.7): cluster C1 represents the region of the front were NSE(E_m) is maximum and NSE(Q_d) improves; cluster C2 is characterized by varying NSE(E_m) values while NSE(Q_d) remains almost constant; finally, cluster C3

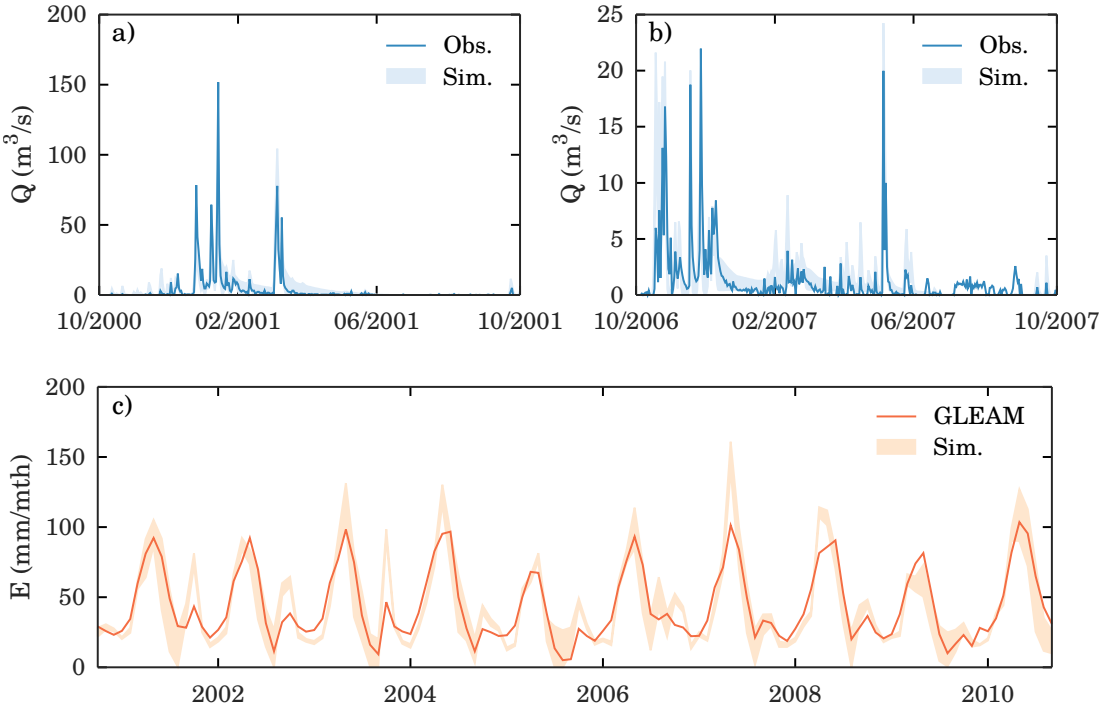


Fig. 6.6: Predictive intervals corresponding to the multi-objective calibration for station R-5055. (a and b) Daily streamflow simulations and observations for two different subperiods. (c) Monthly evaporation simulations and GLEAM data for the complete study period.

includes points where NSE(Q_d) is maximum and both NSE(Q_d) and NSE(E_m) are slightly changing.

The widest boxplots in C1 correspond to D_m and $rout_1$, and therefore are the main drivers of the NSE(Q_d) improvement. In C2, LAI_f is the only parameter responsible for the NSE(E_m) changes, suggesting that small changes of LAI_f lead to great improvements in terms of evaporation performance, and corroborating prior results from the Pareto-based sensitivity analysis. Results for C1 and C2 also confirm the sensitivities estimated by DELSA for D_m and $rout_1$ regarding NSE(Q_d) (Fig. 6.4a), and for LAI_f in relation to NSE(E_m) (Fig. 6.4b). The changes of NSE(Q_d) and NSE(E_m) in C3 are, in turn, less evident when seen through the lens of parameter distributions. These three behaviours are also evinced by the abrupt changes of parameter values between the clusters, particularly for b_i, W_S, D_m, d_2 and $rout_1$.

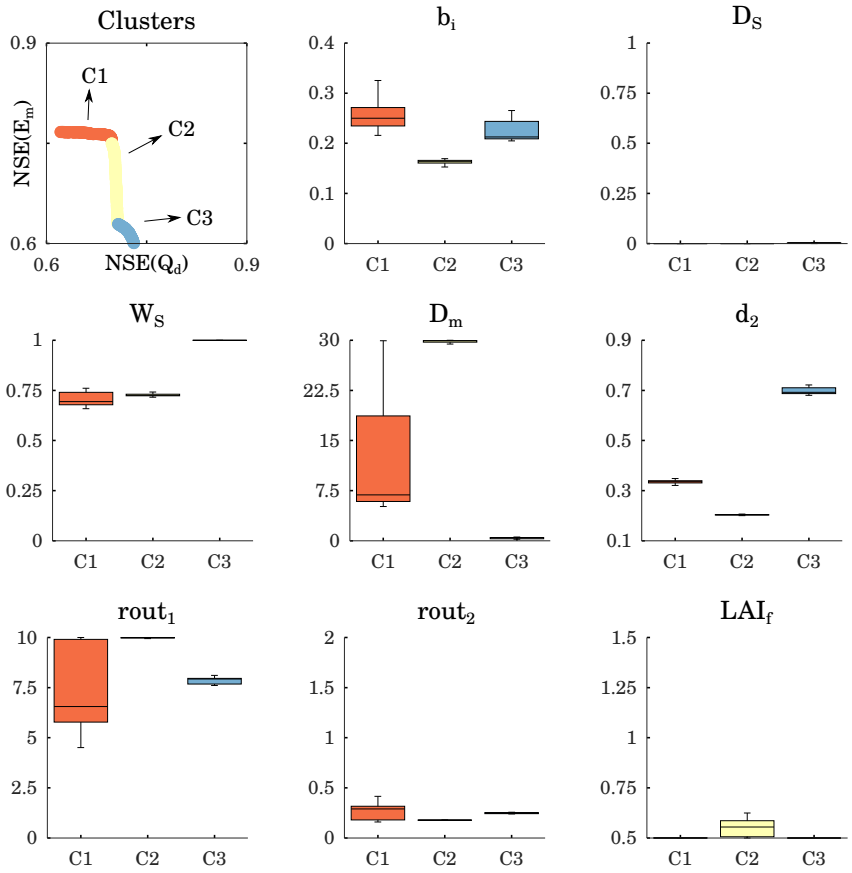


Fig. 6.7: Parameter distributions from the multi-objective calibration exercise for station R-5055. The Pareto front was divided intro three clusters representing well-differentiated regions of model performance: C1 corresponds to the highest NSE(E_m) estimates; C2 contains solutions where NSE(Q_d) remains constant and high and NSE(E_m) improves; C3 is in the region with the highest NSE(Q_d) values and lowest NSE(E_m) estimates. The parameter distributions were evaluated for the three clusters.

Regarding station R-5005, three clusters were defined with similar model behaviours to those described above (Fig. S6.5). In this case, b_i, d_2 and $rout_1$ represent the widest boxplots in the region of NSE(Q_d) enhancement, while $rmin_f$ dominates the NSE(E_m) increase. Finally, only one cluster was defined for station R-2011 (Fig. S6.11), where the small parameter ranges are due to the marginal NSE improvement found in the constrained objective space (Fig. S6.9b).

6.5.4 NSE decomposition and single-objective experiments

Based on Eq. 6.5, the three components of NSE, r, α and β, were calculated for both NSE(Q_d) and NSE(E_m) and each of the studied catchments, and the results are shown in Fig. 6.8, S6 and S12. The Q-E trade-off is evident for all the components in the case of stations R-5055 and R-5005, and r is the

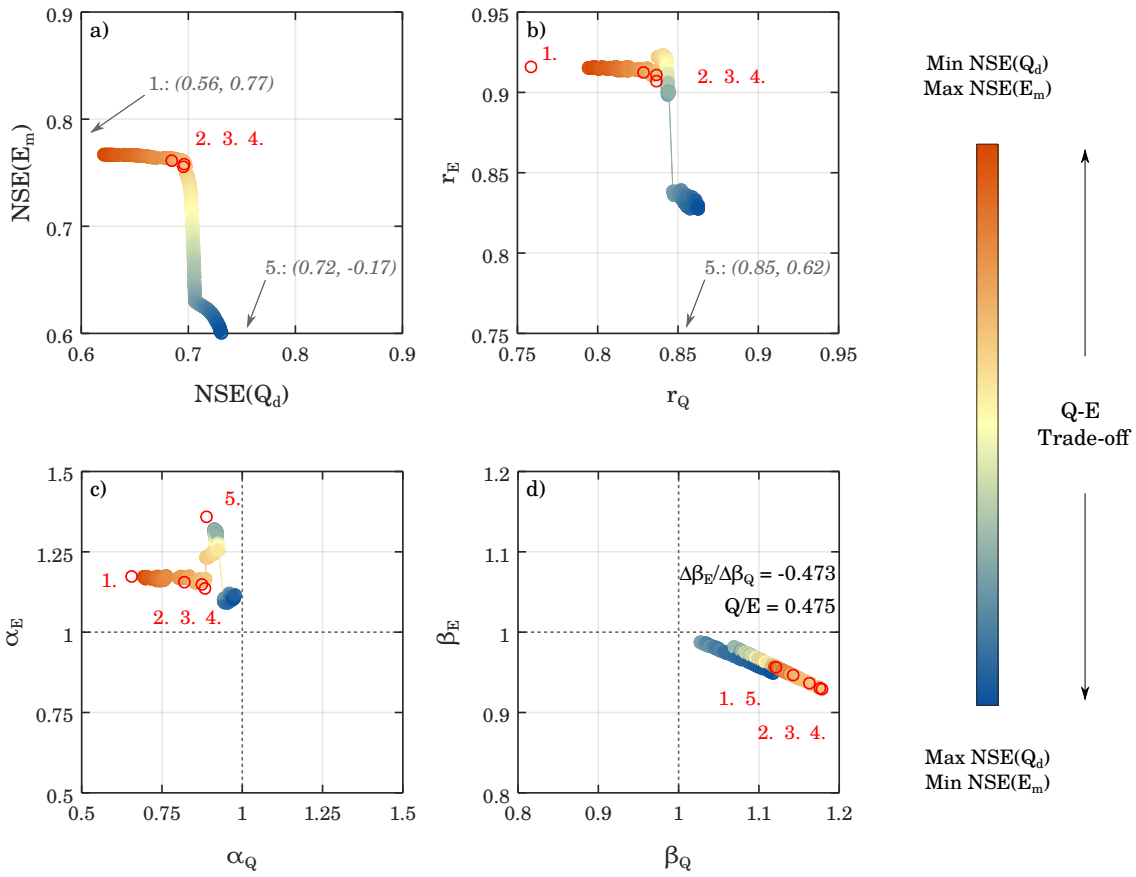


Fig. 6.8: NSE decomposition into r, α and β and trade-off evaluation for station R-5055. The NSE ordering is first determined for the Pareto front corresponding to the multi-objective calibration using a divergent color scale (a). The NSE ordering is then superimposed on the decomposition into r, α and β to evaluate the extent of the Q-E trade-off (b to d). Eq. 6.6 was checked for the β components of NSE(Q_d) and NSE(E_m), and the resulting values are indicated in panel d. The results of the NSE decomposition for the solutions of the five single-objective experiments are depicted in every panel in red, or indicated in grey if outside of the plotting region.

most contributing factor to it as it manifests an identical NSE-ordering to that presented in the Pareto front (see Fig. 6.8b and S6b and compare to Fig. 6.8a and S6a, respectively). Remarkably, cluster C3 (Fig. 6.7) is also noticeable in the NSE decomposition for R-5055 (Fig. 6.8), providing evidence of a complex and non-convex objective space. The small NSE improvement evinced for R-2011 (Fig. S6.12) is, however, a symptom of a weak trade-off that is also reflected in the short ranges of r, α and β.

Lastly, the single-objective calibration experiments for R-5055 tested the ability to attain Pareto-optimal solutions using SCE-UA (Fig. 6.5c, 6.8a). Although SCE-UA solutions mostly reach the Pareto front and resemble its shape, some important features of the objective space are not well reproduced. Notably, the point corresponding to the streamflow-only calibration yielded a sub-optimal solution in comparison with the multi-objective experiment (Fig. 6.5c, 6.8a). Both calibration strategies will be discussed and further analyzed in next section.

α values are concentrated in the top-left quadrant of Fig. 6.8c, S6c and S12c, revealing a generalized underestimation of the variability for the streamflow ($\alpha_{\mathrm{Q}} < 1$) and an overestimation of the variability for evaporation ($\alpha_{\mathrm{E}} > 1$). As expected, β values follow a straight line with negative slope (Eq. 6.6), and its magnitude is close to the value of $Q_{\mathrm{ref}}/E_{\mathrm{ref}}$ computed for each catchment (Fig. 6.8d, S6d and S12d). Lines are mostly located in the bottom-right quadrant, suggesting an overall overestimation of the mean streamflow ($\beta_{\mathrm{Q}} > 1$) and an underestimation of the mean evaporation ($\beta_{\mathrm{E}} < 1$)

6.6 Discussion

6.6.1 Integrating streamflow and evaporation data into sensitivity analysis

The profusion of satellite-based and model evaporation data has proven to be a promising step towards a better understanding of the hydrologic system. Many studies have included evaporation as an evaluation variable to gain further insight and strengthen the results for the streamflow-only calibration (e.g., Bouaziz et al., 2021; Rakovec et al., 2019; Yeste et al., 2020).

Calibration efforts including evaporation have been normally carried out for multiple hydrologic variables either through aggregating functions (Dembélé et al., 2020a,b; Demirel et al., 2018; Széles et al., 2020) or by following pure Pareto approaches (Koppa et al., 2019; Nijzink et al., 2018), and outline the potential for enhancing the representation of hydrologic processes within the modelling framework.

Despite the increasing number of works including evaporation as part of the analysis, there is still no consensus on how evaporation data can be used to improve hydrologic models (Dembélé et al., 2020a). Moreover, the evaporation-related parameters are rarely considered during the sensitivity analysis stage, where streamflow has traditionally played a major role. In the case of the VIC model, this has resulted in numerous parameters with unknown sensitivities (Sepúlveda et al., 2022).

In this respect, and depending on the studied catchment, up to 5 vegetation parameters out of the 13 vegetation parameters initially selected were identified as important to NSE(E_{m}) with DELSA, although their effect was almost unnoticeable for NSE(Q_{d}) (Fig. 6.4, S2, S8). Particularly, LAI_{f} and $rmin_{\mathrm{f}}$ were among the most important parameters in terms of evaporation performance for all the catchments, which is consistent with the findings of Sepúlveda et al. (2022). Likewise, their effect was found to be negligible for NSE(Q_{d}), though it is expected to become important at annual scale according to Melsen and Guse (2019). For their part, the soil and routing parameters reflected similar sensitivities to those reported in previous studies (e.g., Demaria et al., 2007; Gou et al., 2020; Lilhare et al., 2020; Melsen et al., 2016a; Melsen and Guse, 2019; Mendoza et al., 2015; Rakovec et al., 2014).

Although the results of the Pareto-based sensitivity analysis are largely in line with the DELSA first-order sensitivity measures for the studied catchments, the relative importance of each parameterization in terms of joint model performance for both NSE(Q_{d}) and NSE(E_{m}) could not be uncovered by only using DELSA. In this sense, the full parameterization experiments were still subject to overparameterization as they included all the parameters identified with DELSA. The Pareto-based sensitivity analysis successfully overcame this limitation, and its strength lied in its ability to directly pinpoint the most parsimonious implementation leading to a similar performance to that corresponding to the full parameterization.

Thus, no more than two vegetation parameters were required in all cases:

LAI_{f} for station R-5055 (Fig. 6.5), LAI_{f} and $rmin_{\mathrm{f}}$ for R-5005 (Fig. S6.3), and $depth_{1\mathrm{f}}$ and $depth_{2\mathrm{f}}$ for R-2011 (Fig. S6.9). As suggested in Sepúlveda et al. (2022), these differences are likely connected to the existing hydroclimatic gradient between the stations: stations R-5055 and R-5005 are located in the Guadalquivir River Basin, the southern-most and one of the most semi-arid basins in Spain (Yeste et al., 2018), while R-2011 belongs to a more humid climate characteristic of the Duero River Basin (Yeste et al., 2020, 2021).

6.6.2 Benchmark comparison of calibration strategies

Multi-objective calibration and single-objective experiments

The optimization approach developed in this chapter serves the purpose of both a sensitivity analysis and a calibration method, the only difference being the use of constraints during optimization. Whereas the unconstrained implementation included poor estimates for NSE(Q_{d}) and NSE(E_{m}) that are of little interest from a calibration perspective (Fig. 6.5a, S3a, S9a), the use of constraints during the multi-objective calibration produced solutions within the predefined limits of acceptability (Fig. 6.5b, S3b, S9b).

The results for the single-objective experiments mostly reproduced the shape of the Pareto front and yielded optimal solutions (Fig. 6.5c), although they point out important technical issues concerning local optimality. In particular, the streamflow-only calibration demonstrates that the optimization algorithm became stuck in a local minimum (Fig. 6.5c), probably due to the higher density of points revealed during the Monte Carlo experiment for this area (Fig. 6.3c). Additionally, the calibration results for the weighted distances could barely be differentiated from each other as the solutions tended to cluster together in the region closest to the ideal solution (i.e., top-right corner of the scatterplot, Fig. 6.5c).

A similar approach was carried out in Fowler et al. (2018) in an assessment of the Differential Split-Sample Test (DSST, Klemeš, 1986) for two objectives representing the calibration-evaluation trade-off for streamfow through both multi-objective and single-objective experiments. They concluded that none of the single-objective methods could satisfy the limits of acceptability for both objectives at the same time if only the objective for the calibration period was optimized.

Nonetheless, single-objective optimization is still a valid alternative to a full Pareto optimization procedure when working with multiple objectives if the objectives are combined in a single aggregating metric such as the weighted Euclidean distance. As shown in Fig. 6.5c, the performance gain for evaporation with the weighted metric experiments was substantial in comparison with the slight performance loss detected for streamflow. Future efforts will be directed at exploring this issue in more detail for a large sample of Spanish catchments.

NSE diagnosis

An important drawback of using NSE as an objective during optimization is that the calibration process is ultimately bounded to the relation of r, α and β that stems from its very definition. As shown in Gupta et al. (2009), the first derivative of Eq. 6.5 with respect to α indicates that NSE is optimum when $\alpha = r$. This will favour (but not strictly impose) parameter combinations that underestimate α as r is always less than 1, leading to an underestimation of high-flows and an overestimation of low-flows (Gupta et al., 2009; Mizukami et al., 2018). While this is visible for α_{Q}, α_{E} is always greater than 1 for all the catchments (Fig. 6.8c, S6c and S12c), confirming previous research using VIC and NSE as a performance metric for streamflow and evaporation in Rakovec et al. (2019) and Yeste et al. (2020). A possible explanation for this is that r values are generally higher for NSE(E_{m}) than for NSE(Q_{d}). On the other hand, Eq. 6.6 reveals that β values are subject to a linear trade-off when the steady-state is achieved that is exclusively dependent on reference data for streamflow and evaporation. β lines in Fig. 6.8d, S6d and S12d indicate that the steady-state is always reached by overestimating Q and underestimating E in all cases, which is also in agreement with the findings of Rakovec et al. (2019) and Yeste et al. (2020).

It has to be recognized that similar limitations are inherent to any aggregating metric one may choose. For example, the Kling-Gupta Efficiency (KGE, Gupta et al., 2009), which is together with NSE the most used performance metric in hydrologic studies (Clark et al., 2021; Knoben et al., 2019), is based on a Euclidean distance formulation for r, α and β. Limitations for β shown in Eq. 6.6 apply to KGE as well by definition. It is also known that β values tend to be overestimated with KGE if the mean value of reference data is small (Clark et al., 2021; Santos et al., 2018). This constitutes a potential pitfall for regions

undergoing arid conditions such as the Iberian Peninsula (García-Valdecasas Ojeda et al., 2020a, 2021a; Páscoa et al., 2017; Vicente-Serrano et al., 2014), where Rakovec et al. (2016b) observed poor estimates of KGE in a large-sample evaluation of streamflow performance for 400 European river basins.

Besides its intrinsic limitations, the decomposition of NSE into r, α and β provides valuable information as to why NSE values are as they are, and can help elucidate possible causes of model failure as well as potential avenues for improving model performance. The simultaneous consideration of NSE(Q_{d}) and NSE(E_{m}) is also a necessary step in order to explore model performance and alleviate the shortcomings of using a single system-scale metric (Clark et al., 2021), which has been an ever-present goal in this chapter.

6.7 Conclusions

This chapter presents a framework to incorporate streamflow and evaporation data into the sensitivity analysis and calibration stages of the hydrological modelling exercise. The framework is articulated into four consecutive steps inspired by the definition of the Multi-Objective Optimization Problem in order to study the streamflow and evaporation trade-offs, giving rise to what we have referred to as "Pareto-based sensitivity analysis" and "multi-objective calibration". The proposed approach was then applied to three Spanish catchments using VIC.

The overall VIC performance for streamflow and evaporation was firstly explored in a Monte Carlo experiment using daily and monthly estimates of NSE. As a result, the daily NSE(Q) and the monthly NSE(E) were selected as the metrics of reference. A parameter screening procedure based on the DELSA method was then conducted to study parameter sensitivities, and the most influential parameters were chosen in each case. This selection was polished during the Pareto-based sensitivity analysis, and permitted assessing the gains/losses in performance and identifying the most parsimonious parameterization that produced a proper approximation to both streamflow and evaporation data. The multi-objective calibration was performed for the selected parameterization using constraints applied to the objective space to avoid poor performance estimates, and the solutions were compared to five single-objective optimization experiments for one of the catchments. The multi-objective calibration outperformed the single-objective experiments

and produced a better representation of the streamflow and evaporation trade-offs, suggesting the possibility to obtain satisfactory estimates for both variables with hardly any performance loss for streamflow compared to the VIC calibration only against streamflow data.

The use of multiple-criteria methods as in this chapter is a step forward in the direction of "What is my model good for?" rather than "How good is my model?" (Clark et al., 2021), and has guided the development of the approach and the experimental design since their early inception. Answers to the former question started coming during sensitivity analysis, both with DELSA and Pareto-based, and became clearly visible after calibrating VIC. We demonstrate that Pareto-based sensitivity analysis preceding multi-objective calibration is necessary to reach the full potential of multi-objective optimization. Our results will contribute to gain further insight into the nature of the trade-off resulting from the simultaneous adjustment to streamflow and evaporation data.

7 Hydrologic simulations for the Spanish catchments under present and future climate

This chapter is based on a manuscript *in preparation* for submission

Abstract

The joint integration of streamflow and evaporation data into sensitivity analysis and calibration approaches has a great potential to improve the representation of hydrologic processes in modelling frameworks. This chapter aims to investigate the capabilities of the Variable Infiltration Capacity (VIC) model to simulate water storages and fluxes under present and future climatic conditions in a large-sample application comprising 189 headwater catchments located in Spain. The study has been articulated into three parts: (1) a regional sensitivity analysis for a total of 20 soil, routing and vegetation parameters to select the most important parameters conducive to an adequate representation of the streamflow and evaporation dynamics; (2) a two-fold calibration approach against daily streamflow data and monthly evaporation based on the previous parameter selection for VIC, and (3) an analysis of the impacts of climate change in the studied catchments using regional climate simulations carried out with the Weather Research and Forecasting (WRF) model for the Iberian Peninsula. The regional sensitivity analysis revealed that only two vegetation parameters were sufficient to improve the VIC performance for evaporation, and were added to the soil and routing parameter during the calibration stage. Results from the two calibration experiments suggested that, while the performance for streamflow remained close in both cases, the performance for evaporation was highly improved if the objectives for streamflow and evaporation were combined into a single composite function

during optimization. A final assessment of the impacts of climate change for the studied catchments demonstrated that most of the catchments will likely undergo a more evaporative regime and will experience a shift towards soil drying conditions. Results from this investigation will help gain a better understanding of the hydrology of semi-arid regions and will help prepare the ground for a fully gridded implementation of the VIC model in the Spanish domain.

7.1 Introduction

The feasibility of incorporating streamflow and evaporation data into the different stages of the hydrological modelling exercise has been assessed in previous chapters. In Chapter 4 (Yeste et al., 2020), the VIC model was calibrated for 31 headwater catchments belonging to the Duero River Basin against monthly streamflow data, and its performance was evaluated for the monthly simulations of streamflow and evaporation. Results indicated that it was possible to attain good adjustments for both hydrologic variables simultaneously despite the model was calibrated only for one objective related to streamflow. The adjustment to evaporation data, however, was highly dependent on each catchment as the single-objective optimization approach did not include evaporation information.

The Duero River Basin was further investigated in Chapter 5 (Yeste et al., 2021) to determine the projected hydrologic changes under future climate using an ensemble of 18 Euro-CORDEX RCM experiments and the calibrated parameters from Chapter 4. In this case, even though the VIC performance during the calibration period for monthly streamflow was considerably lower using the RCM data compared to the meteorological observations, the ensemble improved the performance for monthly evaporation, which could be related to the model-nature of the evaporation dataset used to evaluate the simulations.

The integration of streamflow and evaporation data was extended to the sensitivity analysis and calibration stages in Chapter 6 by following a multi-objective optimization approach to find Pareto optimal solutions for two objectives. These objectives were specifically selected to characterize the goodness-of-fit of the VIC model to daily streamflow observations and monthly evaporation data, respectively. The VIC model was applied in a case study to examine the trade-offs in model performance during sensitivity analysis and calibration for three headwater catchments: two located in the Guadalquivir River Basin and one in Duero. The approach developed in Chapter 6 effectively overcame the limitations of the so-called streamflow-only calibration and produced satisfactory adjustments to the reference data, highlighting the benefits of Pareto optimization for the bi-objective problem under study.

This chapter builds into all the experience gathered from previous chapters to conduct a large-sample study for 189 Spanish headwater catchments using VIC. Sensitivity analysis and model calibration will be carried out in tandem for two performance metrics representing the model fits to daily streamflow

observations and monthly evaporation. These metrics will be combined into a single weighted metric to explore the effect of different weights on the performance for the streamflow and evaporation simulations during calibration. Catchment-wide simulations of climate change will be finally carried out for the Spanish domain using the calibrated parameters to quantify the long-term and short-term changes in water storages and fluxes.

7.2 Data

7.2.1 Streamflow data

In this work, the catchments selection was based on the complete set of 441 Spanish headwater catchments (Section 2.1 in Chapter 2) and a data quality control for the daily streamflow observations gathered from SAIH-ROEA. A baseline period was defined for the 20 hydrologic years spanning from Oct 1990 to Sep 2010 according to two criteria: 1) the overlapping period for daily precipitation and temperature observations (see next section) is Jan 1950 - Dec 2012, implying that the selected baseline period covers the two most recent decades of the overlapping period; 2) these two decades are a common choice in previous studies using VIC over the Contiguous United States (CONUS) domain (Mizukami et al., 2017; Rakovec et al., 2019), providing fertile ground for the evaluation of the performance of the VIC model across climates.

A total of 189 headwater catchments (Fig. 7.1) comprising 94 reservoirs and 95 gauging stations were selected by setting a maximum percentage of missing values of 10% during the baseline period as the limit of acceptability for inclusion in the modelling framework. Among those catchments, 171 (84 reservoirs + 87 gauging stations) presented less than 5% of missing values, and 144 (80 reservoirs + 64 gauging stations) were below 1%. Table 7.1 collects the number of catchments for every River Basin District included in the streamflow dataset.

An exploratory data analysis of negative values during the baseline period (Fig. 7.2) was subsequently conducted and revealed that 47 reservoirs presented up to a 5% of negative estimates of daily streamflow and 46 reservoirs more than 5%, whereas the 95 gauging stations and only 1 reservoir did not presented values below 0 (Fig. 7.2a). In addition, the percentage ratio of negative to positive values was calculated for each reservoir to quantify their relative

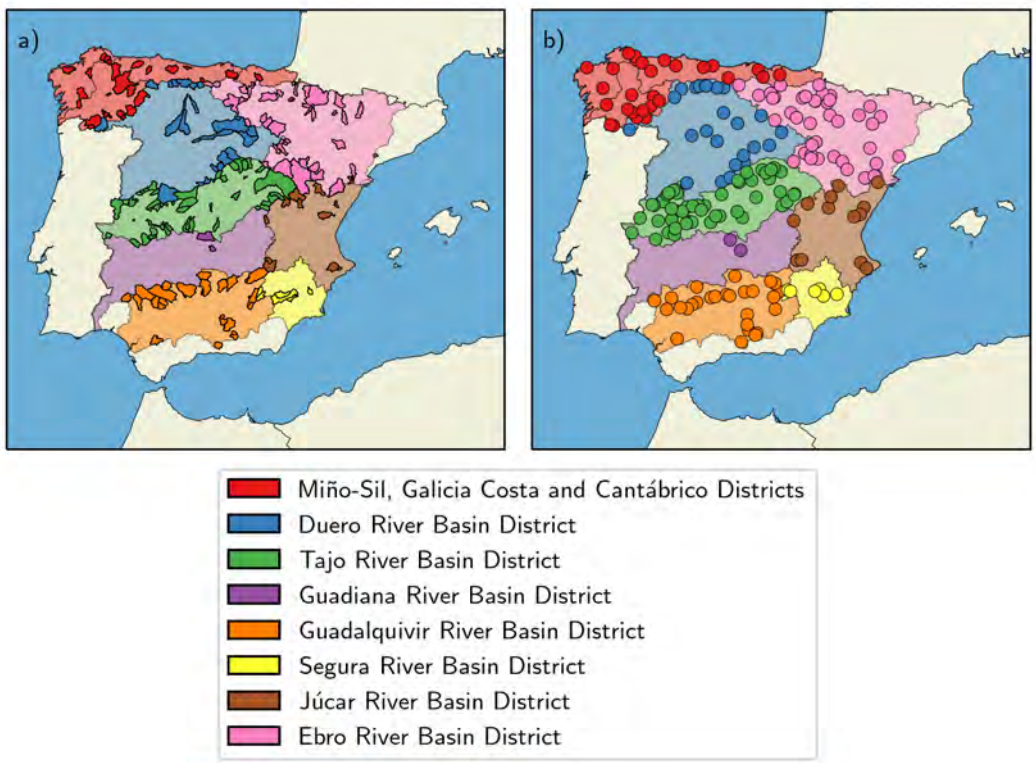


Fig. 7.1: River Basin Districts in Spain and the 189 headwater catchments under study. (a) Topographic boundaries of the 189 headwater catchments. (b) Point-representation of the studied catchments choosing their outlets for the presentation of results.

importance, suggesting that negative values are close to 0 for reservoirs with less than 5% of negative records and become more visible above 5% (the median percentage ratio of negative to positive values for reservoirs with more than 5% of negative records is 2.9%).

One feasible explanation for the presence of negative values in the daily time series of streamflow for reservoirs is that inflow data are calculated applying a daily water balance to water storages and releases from the reservoir (Section 2.1 in Chapter 2). As opposed to the gauging stations, where streamflow is derived from rating curves, the daily water balance in reservoirs can produce negative estimates of streamflow (i.e. inflow) when the variation in the storage is negative and its magnitude is greater than the water releases for any given day. This is likely to happen in those days of the hydrologic year when streamflow is minimum or null and the reduction in the storage is close to the water releases

Table 7.1: Number of headwater reservoirs and gauging stations per River Basin District included in this study.

River Basin District	Reservoirs	Gauging stations	Total
Miño-Sil, Galicia Costa and Cantábrico	13	15	28
Duero	11	12	23
Tajo	14	35	49
Guadiana	2	0	2
Guadalquivir	27	0	27
Segura	6	0	6
Júcar	8	5	13
Ebro	13	28	41
Total	**94**	**95**	**189**

(if any), such as at the beginning of the hydrologic year and in summer. To further test this hypothesis, the monthly distribution of negative values for all the catchments was calculated and compared against the average hydrologic year for streamflow (Fig. 7.2b). Results confirm that most negative values are concentrated in summer months, emerging in late spring and extending to the beginning of the hydrologic year.

On the basis of this initial exploratory analysis, negative values were considered as null for the purpose of this chapter, and the 189 headwater catchments were included in the study. This assumption was taken in order not to compromise the 10% of missing values criterion, as a more straightforward solution would have been considering negative records as gaps. Future efforts will be placed on further analyzing the effect that this assumption can potentially entail for model calibration and evaluation.

7.2.2 Hydroclimatic data

Four study periods were selected to carry out the VIC model simulations: the already defined baseline period was used for model calibration and evaluation purposes; and three periods based on climatic years (that is, from December to November) were selected to perform climate change simulations, the historical period 1980-2015, the short-term future period 2021-2050 and the long-term

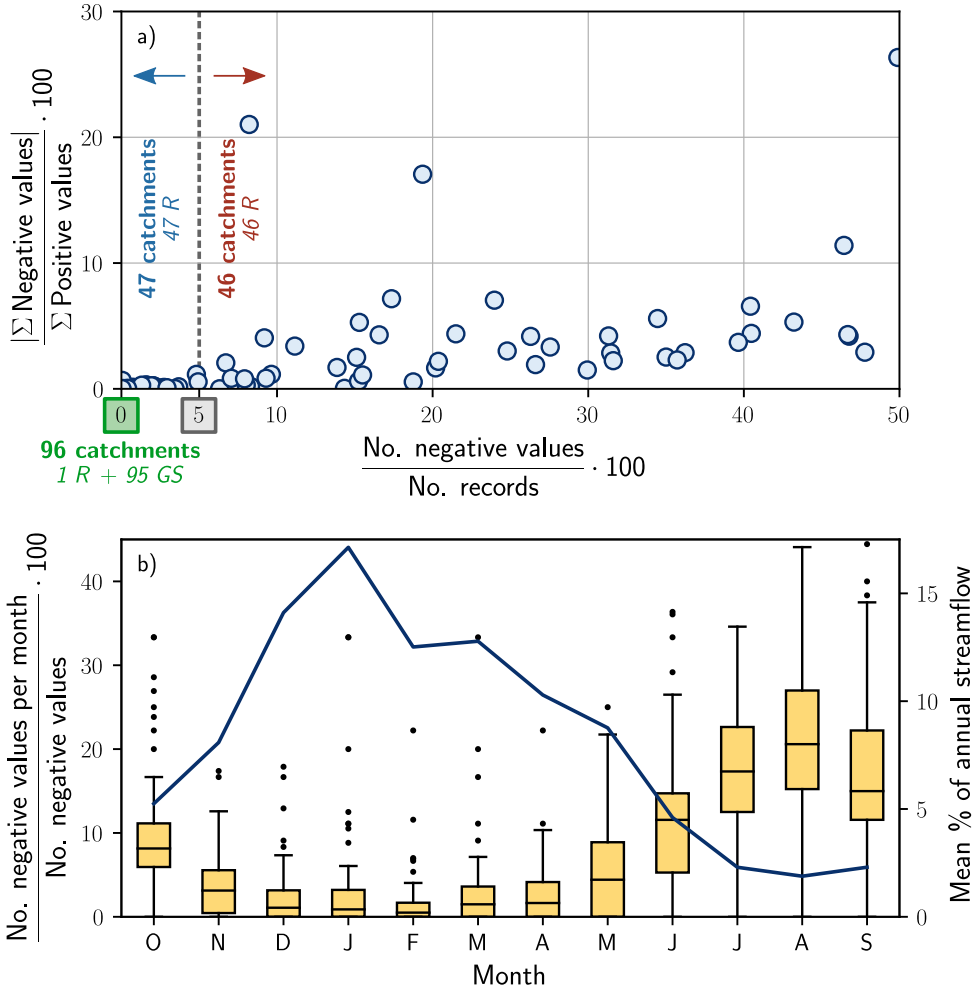


Fig. 7.2: Exploratory data analysis for the negative values in the daily time series of streamflow gathered from CEDEX for the 189 studied catchments during the baseline period. (a) Scatter plot representation of the percentage ratio of negative to positive values (y-axis) against the percentage ratio of the number of negative values to the number of records (x-axis) for each catchment. "*R*" denotes "Reservoir" and "*GS*" denotes "Gauging Station". (b) Monthly distribution of the percentage ratio of the number of negative values per month to the total number of negative values. Blue line corresponds to the average hydrologic year of streamflow expressed as the monthly mean percentage of annual streamflow calculated over all the catchments.

future period 2071-2100. The hydrologic projections were carried out under the RCPs 4.5 and 8.5.

Daily precipitation, maximum temperature and minimum temperature data were gathered for the different study periods from the following sources (see Chapter 2 for a more detailed description): data for the baseline period were collected from SPREAD (Serrano-Notivoli et al., 2017) and STEAD (Serrano-Notivoli et al., 2019). Meteorological forcings for the historical and future periods were taken from the WRFCCSM and WRFMPI simulations in García-Valdecasas Ojeda et al. (2020a,b) for the Iberian Peninsula. Finally, evaporation data were collected from GLEAM version 3.5a (Martens et al., 2017; Miralles et al., 2011). As in previous implementations of VIC for the Duero River Basin (Yeste et al., 2020, 2021), all gridded datasets were downscaled to 0.05º spatial resolution based on a nearest neighbour assignment to match the spatial configuration of the VIC model.

Fig. 7.3 shows the computed values of the runoff ratio (Q/P) and the sum of the runoff and evaporation ratio to precipitation ($(Q+E)/P$) for the baseline period. Two thirds of the catchments manifest Q/P estimates below 0.4 (Fig. 7.3c) and are predominantly located in the southeastern sector of the country (Fig. 7.3a). Approximately 50% of the catchments produce $(Q+E)/P$ estimates between 0.9 and 1.1 (Fig. 7.3d), with negative imbalances ($(Q+E)/P < 1$) mostly corresponding to catchments with a low runoff ratio and positive imbalances ($(Q+E)/P > 1$) towards the northwest (Fig. 7.3b). The effect of these imbalances will be thoroughly examined during the model evaluation stage.

7.3 Methods

7.3.1 Regional sensitivity analysis

Parameter sensitivities were analyzed using the implementation of the Regional Sensitivity Analysis (RSA) method of Hornberger and Spear (1981) in the SAFE Toolbox (Pianosi et al., 2015). RSA is based on a classification of model simulations into behavioural and non-behavioural according to one or more performance metrics, and evaluates differences between parameter distributions corresponding to both classes. The RSA sensitivity index for a given parameter represents the maximum vertical distance between the Cumulative Distribution Functions (CDFs) corresponding to the behavioural and non-behavioural classes, which is equivalent to the Kolmogorov-Smirnov

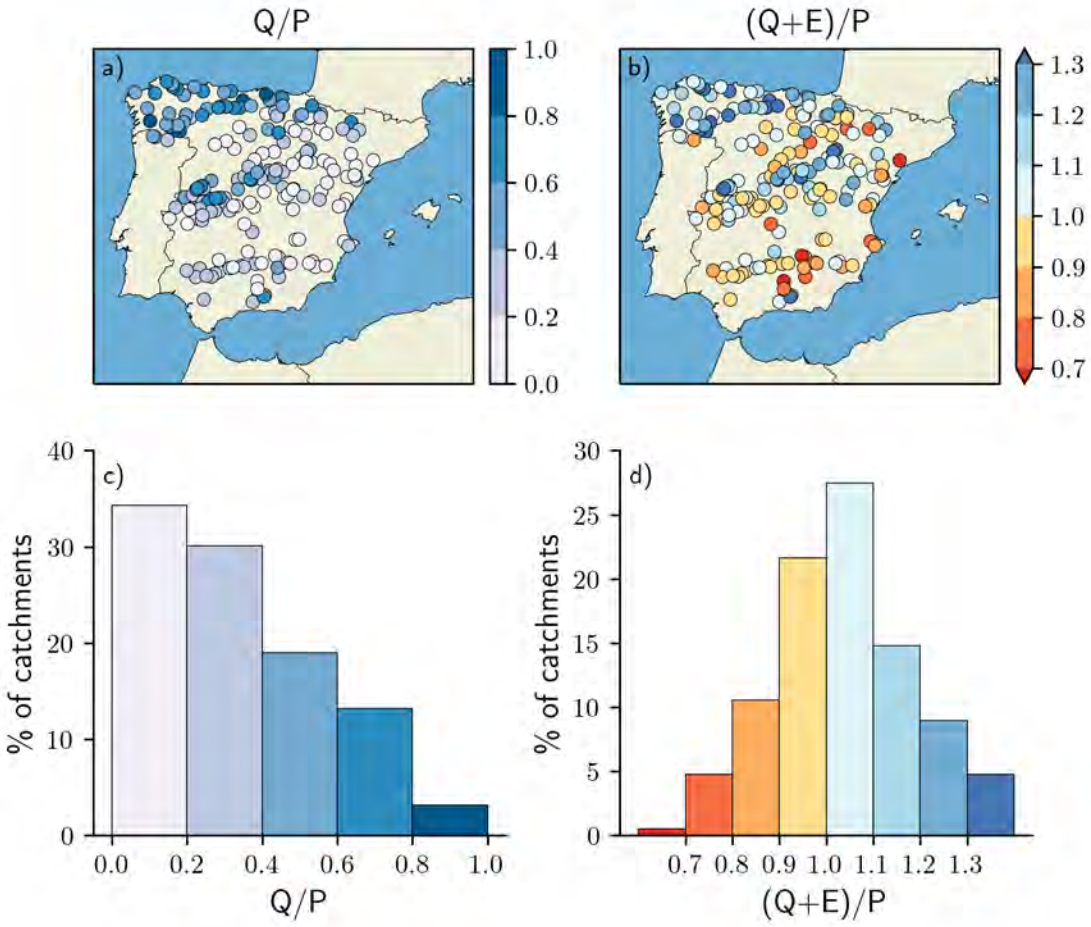


Fig. 7.3: Spatial distribution of the runoff ratio (Q/P) and the sum of the runoff and evaporation ratio to precipitation ($(Q+E)/P$) for the studied catchments during the baseline period. (a) Spatial distribution of Q/P. (b) Spatial distribution of $(Q + E)/P$. (c) Histogram of Q/P estimates. (d) Histogram of $(Q + E)/P$ estimates.

(KS) distance statistic computed in the KS test. Hence, the RSA sensitivity index ranges from 0 to 1, with values closer to 1 indicating a greater parameter sensitivity.

For each of the 189 catchments, the parametric space was explored conducting a Monte Carlo simulation for 10,000 Latin Hypercube samples (Iman and Conover, 1982) extracted from the parameter ranges of the 20 soil, vegetation and routing parameters indicated in Table 6.1 in Chapter 6. RSA was applied to the Nash-Sutcliffe Efficiencies of the Monte Carlo simulations of daily

streamflow (NSE(Q_d)) and monthly evaporation (NSE(E_m)) calculated for the baseline period, and choosing the median NSE(Q_d) and median NSE(E_m) to classify behavioural and non-behavioural simulations.

The two most influential vegetation parameters to any of the two performance metrics for each catchment were selected and incorporated together with the five soil parameters and the two routing parameters (SR parameterization, see also Section 6.5.2 in Chapter 6) into the calibration stage. Adding two extra VIC vegetation parameters to the SR parameterization is sufficient to improve the joint performance against streamflow and evaporation data according to previous research in Chapter 6.

7.3.2 Calibration and evaluation approach

A Split-Sample Test (SST, Klemeš, 1986) was applied to hydroclimatic data and time series of model simulations to calculate the performance metrics for two independent periods of equal duration belonging to the baseline period: a calibration period from Oct 1990 to Sep 2000 and a evaluation period from Oct 2000 to Sep 2010. A spin-up simulation of 10 hydrologic years preceding both the calibration and evaluation periods was conducted to provide initial states of model storages free from the effect of initial conditions (this strategy was also applied to the Monte Carlo simulation described in the previous section). The performance of the VIC model was evaluated through NSE(Q_d) and NSE(E_m) and their decomposition into correlation r, variability α and bias β (see Section 6.4.3 in Chapter 6 for a full description of these statistics and their relationships).

The VIC model was calibrated using the Shuffled-Complex-Evolution Algorithm (SCE-UA) of Duan et al. (1994) following a single-objective optimization approach for the nine selected parameters (five soil parameters, two routing parameters and two vegetation parameters) to minimize a composite function that aggregates the performance metrics for streamflow and evaporation:

$$\text{Minimize} \quad \sqrt{w_Q \cdot (1 - \text{NSE}(Q_d))^2 + w_E \cdot (1 - \text{NSE}(E_m))^2} \tag{7.1}$$

This problem minimizes the two-dimensional weighted Euclidean distance between points (NSE(Q_d),NSE(E_m)) and (1,1), and belongs to the more general weighted-metric method to minimize distances (see Eq. 6.2 in Chapter 6). In

this work, two calibration strategies for different weights combinations in Eq. 7.1 were applied to the VIC model: firstly, a streamflow-only calibration (Q-only calibration hereafter) was performed by choosing $w_{\mathrm{Q}} = 1$ and $w_{\mathrm{E}} = 0$. Secondly, the model was calibrated for a weighted Euclidean distance (Q-E calibration hereafter) selecting two equal weights $w_{\mathrm{Q}} = w_{\mathrm{E}} = 0.5$. The case of two equal weights is equivalent to minimizing the pure Euclidean distance, that is, $w_{\mathrm{Q}} = w_{\mathrm{E}} = 1$, as equal weights do not affect the optimization problem in Eq. 7.1.

7.3.3 Projected hydrologic changes

The calibration strategy yielding a better model performance during the baseline period against streamflow and evaporation simultaneously was selected to study the impacts of climate change in the Spanish catchments. This selection was carried out based on a comparative analysis of model performance using meteorological observations from SPREAD and STEAD and climate simulations from WRFCCSM and WRFMPI.

The projected hydrologic changes were determined following a delta change approach (Hay et al., 2000) to evaluate the differences in mean values between the future and historical periods at annual and seasonal time scales. The delta changes were calculated for precipitation (P), streamflow (Q), actual and potential evaporation (E and E_{p}, respectively) and terrestrial water storage (TWS). The latter was calculated as the sum of the soil moisture estimates for the three soil layers implemented in the VIC model. The statistical significance of the annual and seasonal delta changes was evaluated applying the Mann-Whitney U test at 90% confidence level.

7.4 Results

7.4.1 RSA sensitivity analysis

The RSA sensitivity indices for NSE(Q_{d}) and NSE(E_{m}) are depicted in Fig. 7.4 and Fig. 7.5, respectively. The NSE(Q_{d}) behaviour was mostly controlled by the five soil parameters and the two routing parameters (i.e. SR parameterization), with little or no influence from the vegetation parameters and no clear spatial

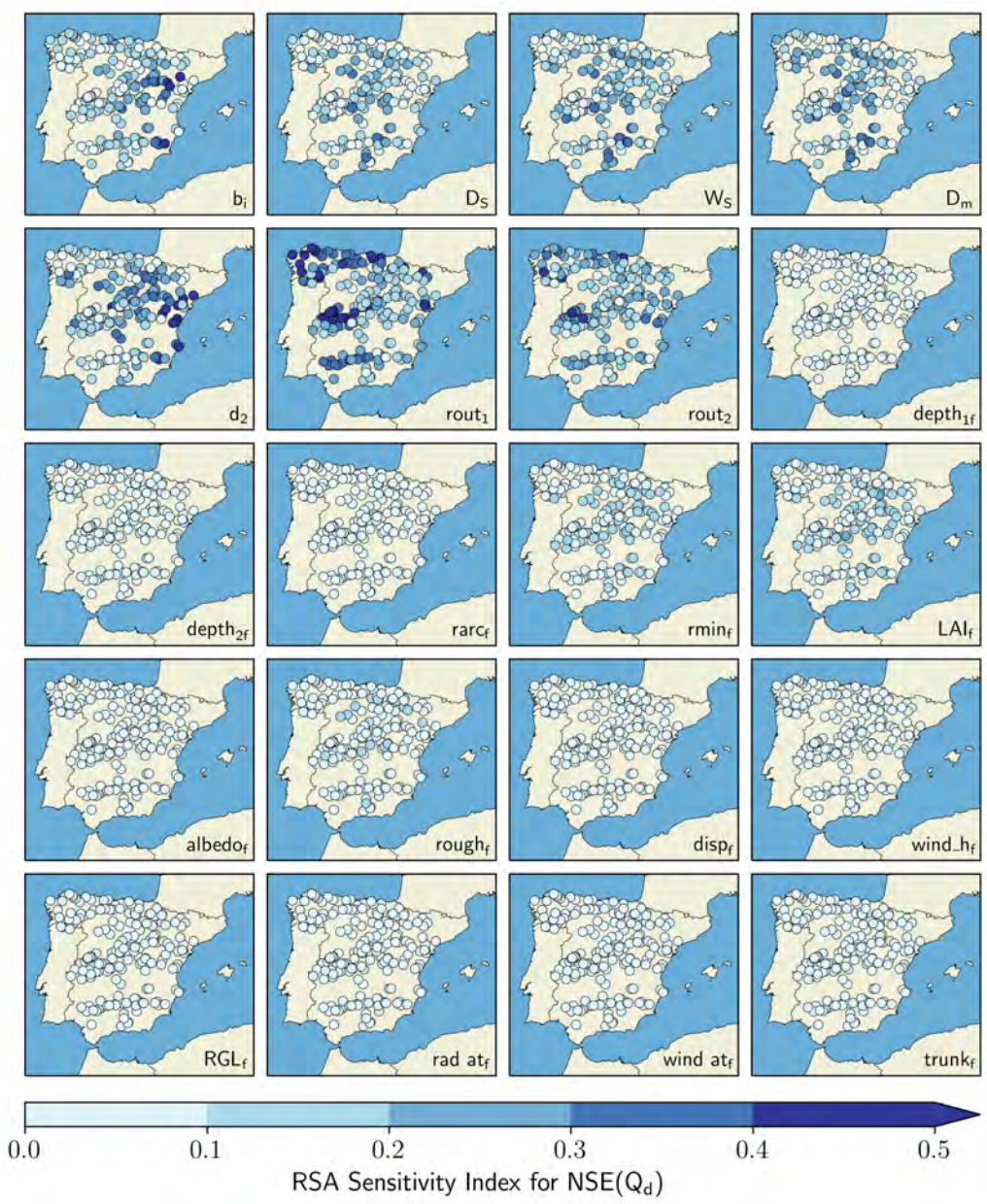


Fig. 7.4: Spatial distribution of the RSA sensitivity indices calculated for NSE(Q_{d}).

pattern for the RSA indices (Fig. 7.4). Among these parameters, the highest sensitivities correspond to d_2, $rout_1$ and $rout_2$, although b_{i}, D_{S}, W_{S} and D_{m} were also influential to the streamflow metric according to several local estimates.

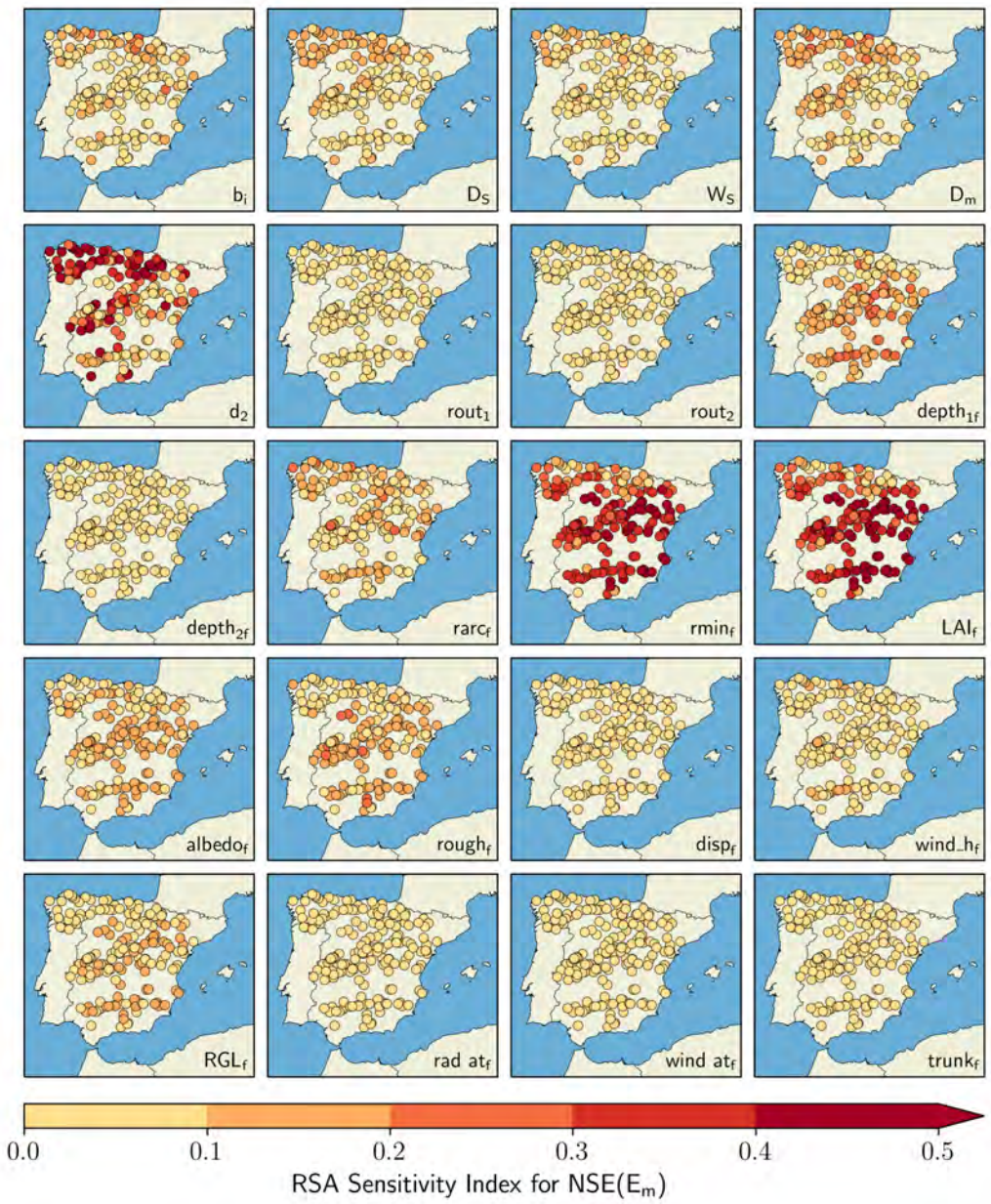


Fig. 7.5: Spatial distribution of the RSA sensitivity indices calculated for NSE(E_{m}).

Contrarily to NSE(Q_{d}), NSE(E_{m}) scores were greatly influenced by the vegetation parameters (Fig. 7.5). The highest sensitivities correspond to $rmin_{\mathrm{f}}$ and LAI_{f}, and manifest a spatial pattern in the northwest-southeast direction with minimum sensitivities occurring for the northern catchments. This pattern

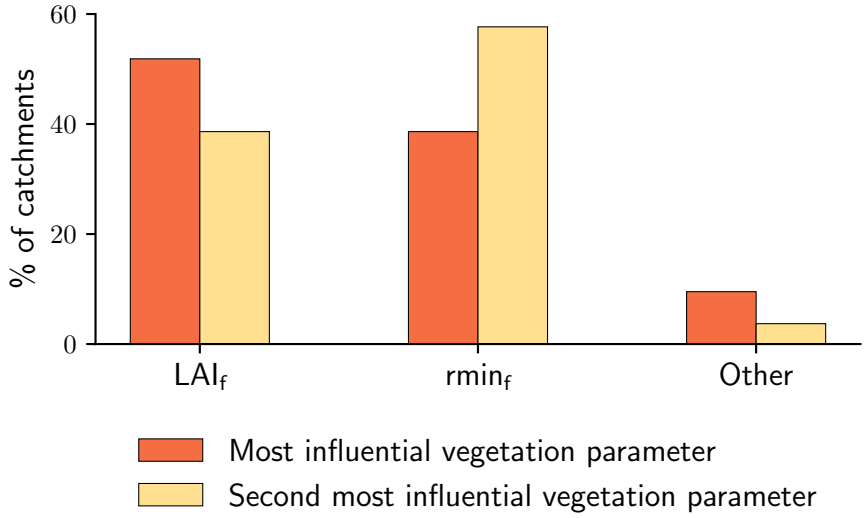


Fig. 7.6: Frequency of the (first) most influential and second most influential vegetation parameters according to the RSA sensitivity indices calculated for NSE(E_m).

is also noticeable, but to a lesser extent, for $depth_{1f}$, $rarc_f$, $albedo_f$, $rough_f$ and RGL_f. d_2 was revealed as the soil parameter with the largest sensitivity towards NSE(E_m), and together with the rest of the soil parameters it reflects an opposite spatial distribution of the RSA index to that observed for $rmin_f$ and LAI_f. As expected from the VIC model implementation, the routing parameters had a null effect on NSE(E_m) due to the routing scheme being exclusively applied to the runoff simulations during a post-process phase.

The two most influential vegetation parameters to any of the two performance metrics analysed were selected according to the values of the RSA index. Fig. 7.6 indicates that LAI_f and $rmin_f$ were the two most influential parameters for the vast majority of the catchments with little presence of other vegetation parameters. For each particular catchment, the two selected vegetation parameters were added to the SR parameterization during the subsequent calibration exercise.

7.4.2 Calibration and evaluation

Fig. 7.7 shows the spatial distributions of the NSE(Q_d) and NSE(E_m) values calculated for the Q-only (Fig. 7.7a,c) and Q-E (Fig. 7.7b,d) calibration strategies during the calibration period. The relative gain/loss in model performance suggests that, while NSE(Q_d) values remained similar for both calibrations (Fig. 7.7a,b), the NSE(E_m) performance was highly improved when the VIC model was calibrated against streamflow and evaporation data simultaneously (Fig. 7.7c,d).

This can also be appreciated in the Cumulative Distribution Functions (CDFs) of NSE(Q_d) and NSE(E_m) and their decomposition depicted in Fig. 7.8 and Fig. 7.9,

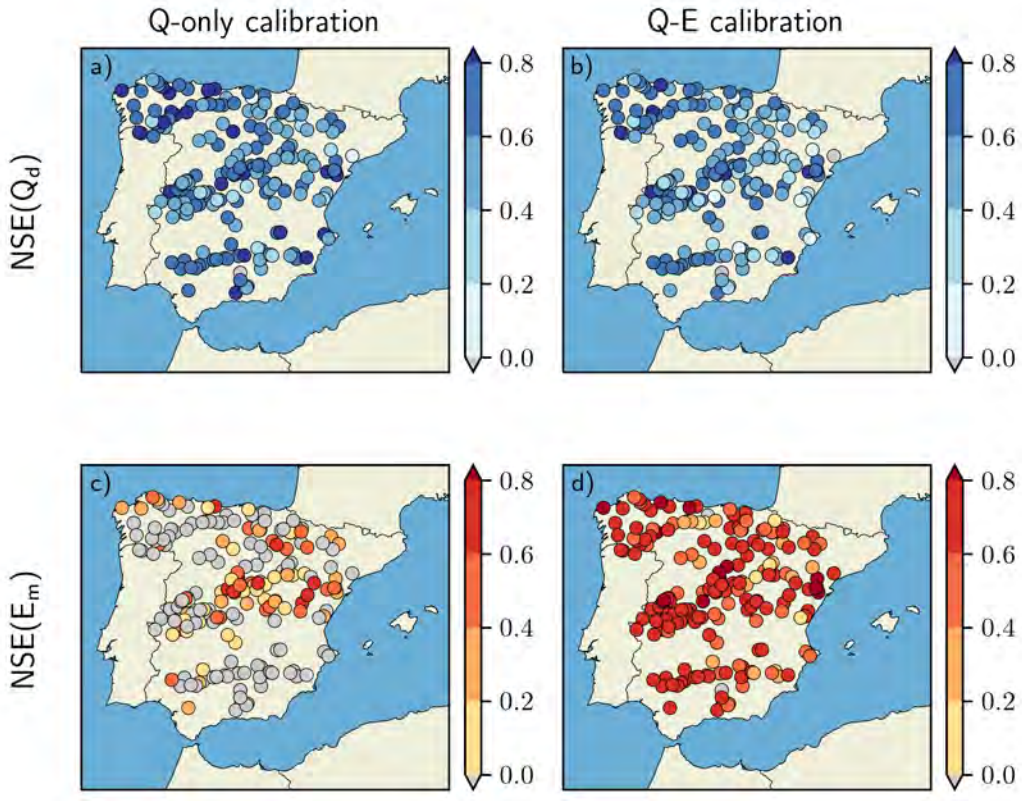


Fig. 7.7: Spatial distribution of NSE(Q_d) and NSE(E_m) for the Q-only calibration (a,c) and the Q-E calibration (b,d) during the calibration period.

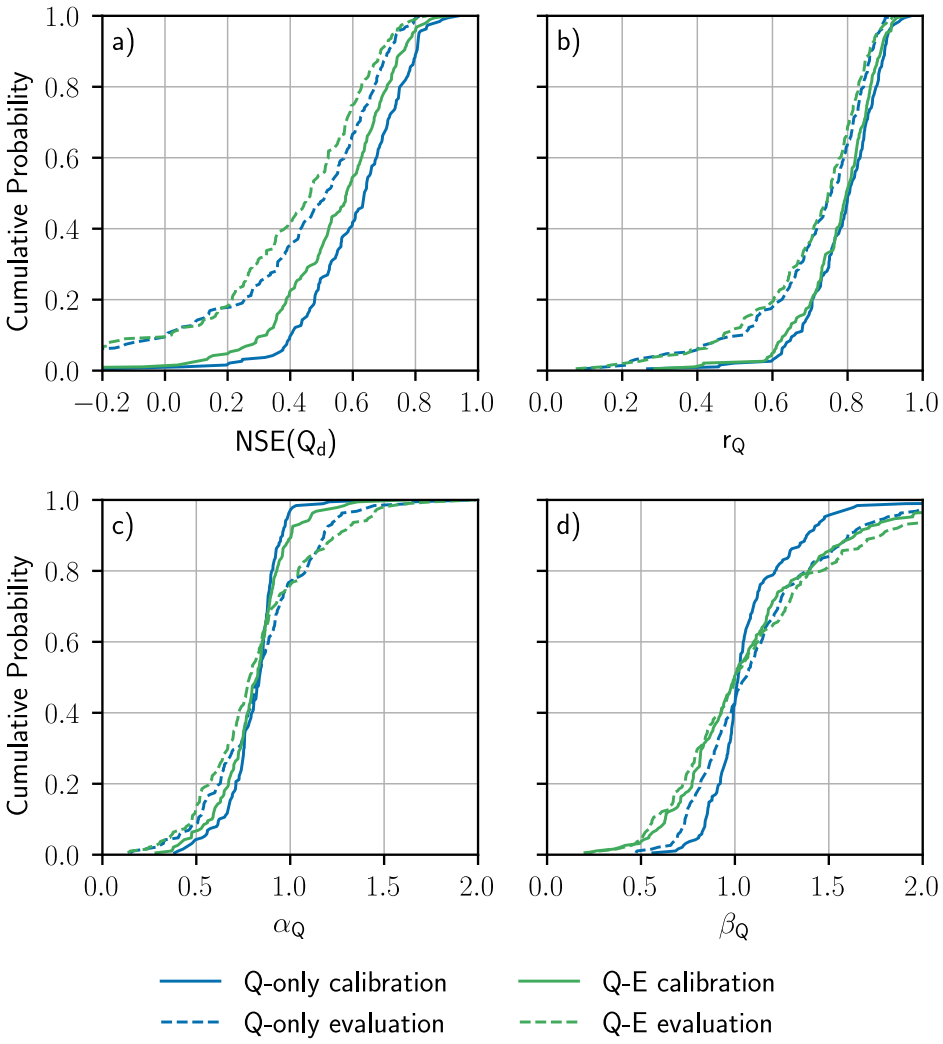


Fig. 7.8: CDFs of (a) NSE(Q_d) and its decomposition into (b) $\mathrm{r_Q}$, (c) α_Q and (d) β_Q for both calibration experiments during the calibration and evaluation periods.

respectively. The median NSE(Q_d) is close to 0.6 for both strategies during the calibration period (Fig. 7.8a), although the streamflow performance is slightly deteriorated for the Q-E calibration as NSE(Q_d) and NSE(E_m) were combined and equally weighted into a composite single-objective function (Eq. 7.1). On the other hand, the median NSE(E_m) during the calibration period calculated for the Q-E calibration is 0.67, while the median NSE(E_m) for the Q-only calibration does not exceed 0. The NSE of monthly streamflow (NSE(Q_m)) was

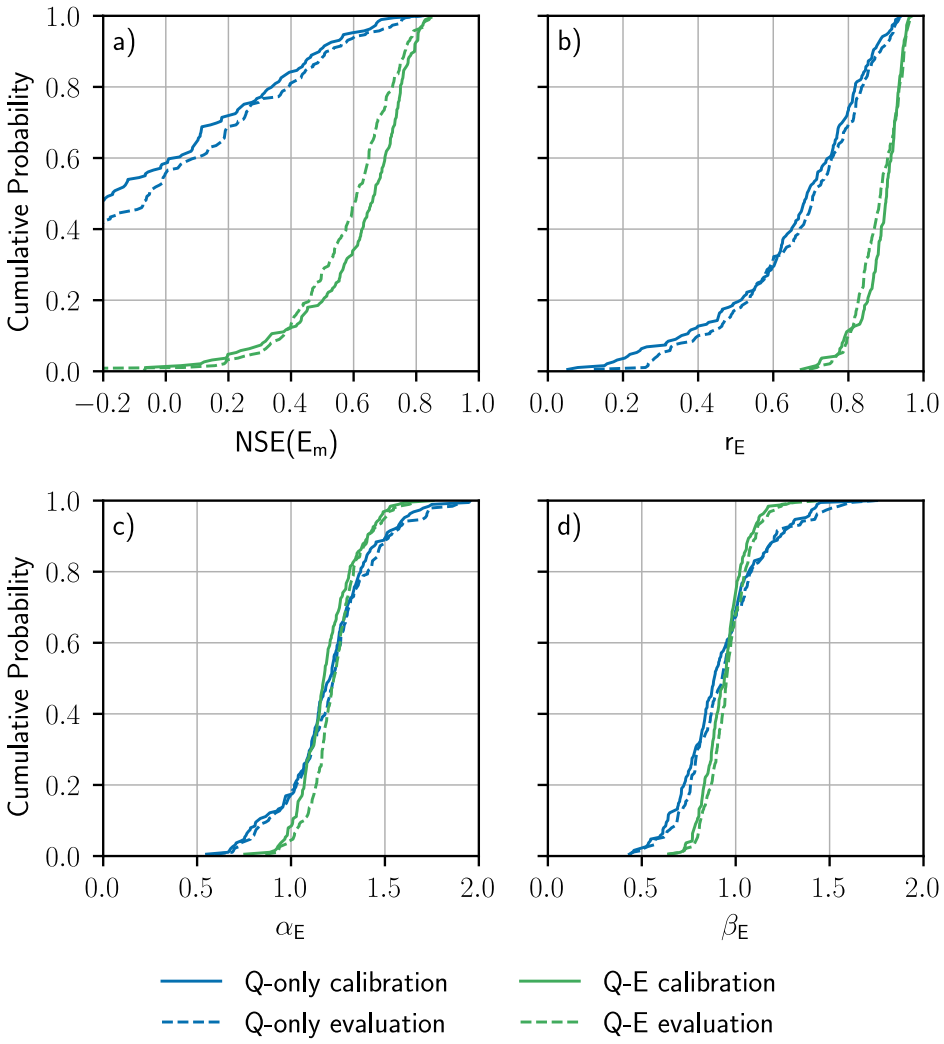


Fig. 7.9: Same as Fig.7.8, but for (a) NSE(E_m) and its decomposition into (b) r_E, (c) α_E and (d) β_E.

also calculated for the studied catchments (Fig. S7.1 in Supporting Information) to provide a more complete picture of the VIC model performance, and results indicate a marked increase in the NSE estimates compared to NSE(Q_d) with a median NSE(Q_m) above 0.7 for both calibration strategies.

The decomposition of NSE(Q_d) and NSE(Q_m) reveals similar r_Q values (Fig. 7.8b, S7.1b) and α_Q estimates generally below 1 for both calibrations (Fig. 7.8c, S7.1c). The main difference between them lies in the β_Q distribution,

which is approximately symmetric around the median for both calibrations but reflects a steeper CDF closer to 1 for the Q-only calibration (Fig. 7.8d, S7.1d). The NSE(E_m) improvement for the Q-E calibration is also evinced in the statistics used to compute it, r_E being the component subject to the greatest enhancement (Fig. 7.9b). α_E and β_E estimates are comparable for both calibration strategies with values slightly closer to 1 corresponding to the Q-E calibration (Fig. 7.9c,d), and point to a generalized overestimation of the variability and a slight underestimation of the bias, respectively.

Finally, the slight to moderate loss in model performance during the evaluation period for NSE(Q_d) (Fig. 7.8a), NSE(Q_m) (Fig. S7.1a) and NSE(E_m) (Fig. 7.9a) is indicative of an acceptable implementation and an adequate predictive capability. A detailed characterization of the VIC model performance including the values of NSE(Q_d), NSE(Q_m) and NSE(E_m) during the complete baseline period is provided in Tables S7.1 to S7.8 in Supporting Information for every River Basin District and each individual catchment. The VIC model skill will be further verified and tested for the baseline period in next section using different meteorological datasets.

7.4.3 Evaluation using multiple meteorological datasets

Fig. 7.10 shows the biases in the Q/P ratio with respect to the observations during the baseline period that stem from both calibration approaches when using SPREAD/STEAD, WRFCCSM and WRFMPI data. Q/P biases corresponding to the Q-only calibrated parameters are broadly in the range ±0.1 for all the datasets (Fig. 7.10a,b,c), while the Q-E calibration increases these deviations as a consequence of calibrating VIC against streamflow and evaporation data simultaneously (Fig. 7.10d,e,f).

Results for SPREAD/STEAD and the Q-only calibration suggest that negative biases tend to be associated with higher Q/P values and vice versa (see Fig. 7.10a and compare to Fig. 7.3a), while the Q/P biases corresponding to the Q-E calibration display an opposite spatial distribution to that observed for the $(Q + E)/P$ values (see Fig. 7.10d and compare to Fig. 7.3b) and exhibit a high negative correlation ($r = -0.91$). For the WRF datasets, there is a predominance of negative Q/P biases for WRFCCSM and positive biases in the case of WRFMPI.

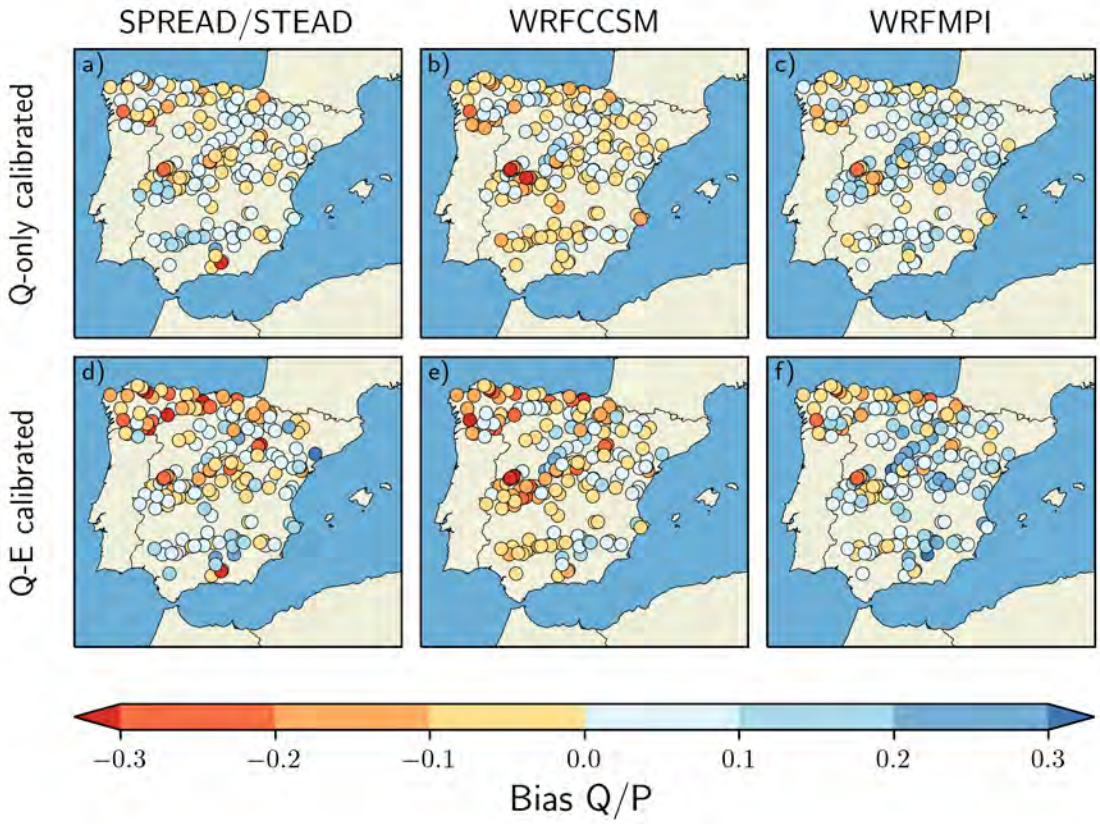


Fig. 7.10: Spatial distribution of the Q/P bias calculated as the difference between the values simulated with the VIC model using SPREAD/STEAD, WRFCCSM and WRFMPI and the observed values during the baseline period (see also Fig. 7.3). (a to c) Q/P bias corresponding to the Q-only calibrated parameters. (d to f) Q/P bias corresponding to the Q-E calibrated parameters.

The distributions of NSE(Q_{d}), NSE(E_{m}) and their decomposition for each meteorological datasets during the complete baseline period are depicted in Fig. 7.11. It is visible that the VIC performance attained for NSE(Q_{d}) using SPREAD/STEAD (SPST) is not achievable with WRFCCSM and WRFMPI (Fig. 7.11a), presumably due to the low r_{Q} values obtained with them (Fig. 7.11b). α_{Q} and β_{Q} estimates for WRFCCSM and WRFMPI also exhibit marked differences with respect to SPREAD/STEAD (Fig. 7.11c,d), although WRFCCSM values are closer to the VIC performance using observations.

The VIC performance for NSE(E_{m}) is, in turn, similar for all the datasets

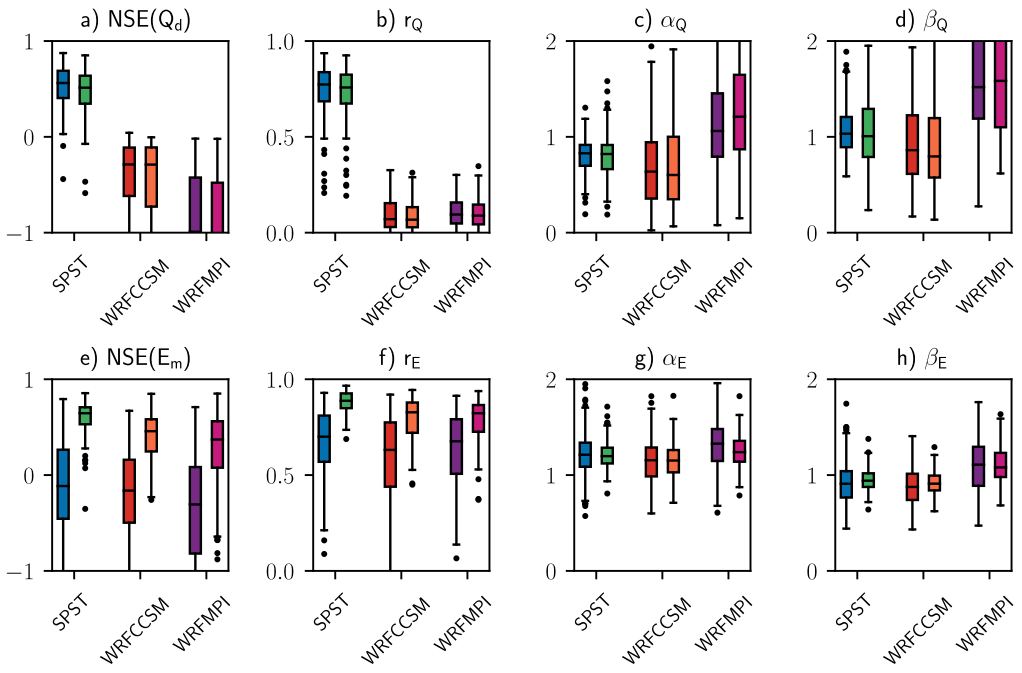


Fig. 7.11: Distributions of (a to d) NSE(Q_d) and its decomposition into r_Q, α_Q and β_Q and (e to h) NSE(E_m) and its decomposition into r_E, α_E and β_E during the baseline period using SPREAD/STEAD (SPST), WRFCCSM and WRFMPI. Blue, red and purple boxplots (i.e., boxplots to the left in each pair group) correspond to Q-only calibrated parameters. Green, orange and pink boxplots (i.e., boxplots to the right in each pair group) correspond to the Q-E calibrated parameters.

and clearly demonstrate the effect of each calibration experiment even for WRFCCSM and WRFMPI (Fig. 7.11e). In the case of evaporation, the r_E values corresponding to the WRF datasets are high (Fig. 7.11f), particularly when using the Q-E calibrated parameters, and α_E and β_E are close to 1 (Fig. 7.11g,h). Again, the performance for WRFCCSM is slightly better than for WRFMPI.

7.4.4 Projected annual and seasonal hydrologic changes

Fig. 7.12 and Fig. 7.13 portray the annual and seasonal delta changes of precipitation (P), streamflow (Q), evaporation (E) and terrestrial water storage (TWS) for the period 2071-2100 under the RCP8.5 scenario using WRFCCSM and WRFMPI, respectively. Results for that period under RCP4.5 and for

the period 2021-2050 under both RCPs are gathered in Fig. S7.2 to S7.7 in Supporting Information. The delta changes of potential evaporation (E_p) are collected as well in Supporting Information in Fig. S7.8 and Fig. S7.9. What follows is a detailed description of the delta changes for the different hydrologic variables under study, with the focus placed on the long-term future period under RCP8.5 given the greater severity of the changes then detected. The delta changes for the rest of the future periods and RCP scenarios will be appropriately referred to throughout the section to provide further insights and relevant information.

Annual *P* is expected to decline throughout the Spanish domain by the end of the century, with significant decreases ranging from ca. 20% in the northern catchments to up to 40% in the south (Fig. 7.12, 7.13). These changes are also evident, but less pronounced, under the RCP4.5 scenario (Fig. S7.6, S7.7), whereas there is a prevalence of non-significant changes for the short-term future period in all cases (Fig. S7.2, S7.3, S7.4, S7.5). At seasonal scale, the *P* delta changes for the long-term future period are negative and significant, exceeding -40% in spring, summer and autumn. Delta changes in winter, by contrast, are mostly non-significant for all the future periods and RCPs, and sometimes present positive values (e.g., Fig. 7.12).

The delta changes of *Q* indicate a generalized decrease at annual and seasonal scales in most future scenarios, showing values broadly below -80% in the long-term future period under RCP8.5 for many southern catchments (these values cannot be observed in Figs. 7.12 and 7.13 as the colorbar for the delta changes has been chosen to adequately represent the projected hydrologic changes considering all the hydrologic variables and all the future study periods at once). The *Q* detriments are in agreement with the *P* decreases since the latter constitute a limiting factor for the runoff generation process. There are some important differences between WRFCCSM and WRFMPI concerning the *Q* delta changes in all the future scenarios as the detriments are more acute for WRFMPI. This is particularly noticeable for the period 2021-2050 under RCP8.5: while WRFCCSM manifests a widespread presence of non-significant *Q* delta changes (Fig. S7.4), WRFMPI is characterized by significant decreases between 20% and 40% at annual scale and above 40% in summer and autumn (Fig. S7.5).

The annual *E* changes reflect a latitudinal gradient for the period 2071-2100 under both RCPs (Fig. 7.12, 7.13, S7.6, S7.7) consisting of: 1) significant diminutions between 20% and 30% for the southern catchments, and 2)

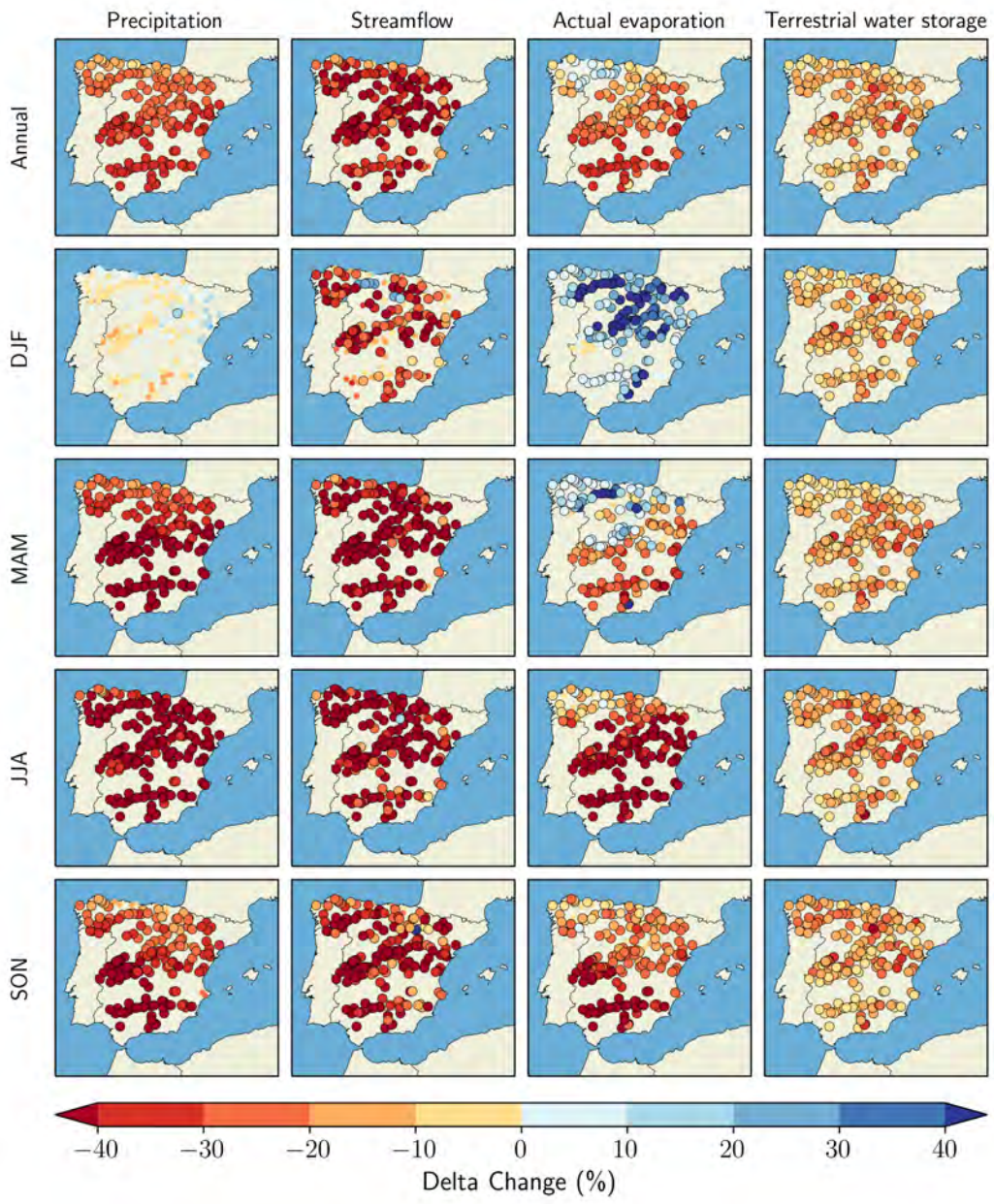


Fig. 7.12: Delta changes of annual and seasonal P, Q, E and TWS for the period 2071-2100 under the RCP8.5 scenario using WRFCCSM. Significant changes at the 90% confidence level have been marked with solid borders.

significant increments of up to 20% for a small portion of the catchments located in the northwestern sector of the country. This gradient is also appreciable for

Fig. 7.13: Delta changes of annual and seasonal P, Q, E and TWS for the period 2071-2100 under the RCP8.5 scenario using WRFMPI. Significant changes at the 90% confidence level have been marked with solid borders.

the short-term future period even though both the increases and decreases are less severe and statistical significant changes are less abundant (Fig. S7.2

to S7.5). The seasonal E changes are subject to a strong intra-annual variability, with significant increases in the northern catchments in winter and spring, and significant decreases emerging in spring in the southern catchments. The summer and autumn E changes, in turn, are mainly positive and in some cases exceed 40% (Fig. 7.12, 7.13). The E behaviour is upper limited by E_{p}, which is expected to raise in all the future scenarios (Fig. S7.8, S7.9).

Lastly, the delta changes of TWS points towards soil drying conditions in all the studied catchments, with most significant decreases occurring for the long-term future period under both RCPs (Fig. 7.12, 7.13, S7.6, S7.7). The annual TWS changes and the seasonal TWS changes in winter and spring reflect a similar spatial pattern with decreases slightly superior corresponding to the southern catchments. In summer and autumn, those changes spread to the northern catchments and become maximum there, causing a flip in the sign of the gradient that finally disappears at annual scale.

7.5 Discussion

7.5.1 Parameter sensitivities

This study expands on previous investigations for the Duero River Basin in Yeste et al. (2020, 2021) by involving the main River Basin Districts in Spain. This permits to obtain more robust information on model realism and the hydrological functioning for a wide range catchments representative of the hydroclimatic variability within the studied region. The RSA sensitivity analysis method allowed for quantifying parameter sensitivities in 189 headwater catchments distributed across the Spanish domain for the various soil, routing and vegetation parameters considered in this work. As in Chapter 6, parameter sensitivities were calculated with respect to NSE(Q_{d}) and NSE(E_{m}), which were the performance metric selected to evaluate the goodness-of-fit of the VIC model simulations for streamflow and evaporation, respectively.

d_2 and the two routing parameters governing the gamma distribution function (i.e., $rout_1$ and $rout_2$) were identified as the most important parameters to NSE(Q_{d}) (Fig. 7.4), highlighting the importance of applying a routing procedure to attain adequate adjustments of the daily streamflow simulations to the observations. The use of a gamma function is a common choice in many studies using VIC to post-process the runoff simulations given its parsimonious

representation of the unit hydrograph (Mizukami et al., 2017; Rakovec et al., 2014, 2019), and constitutes an important leap forward in comparison with the monthly aggregation of runoff carried out in Yeste et al. (2020, 2021). The rest of the soil parameters were also identified as important to NSE(Q_{d}), and yielded comparable sensitivities to those uncovered in many previous works (e.g., Gou et al., 2020; Lilhare et al., 2020; Melsen and Guse, 2019; Mendoza et al., 2015; Yeste et al., 2020). The influence from the vegetation parameters on NSE(Q_{d}), however, was negligible, corroborating the findings of Sepúlveda et al. (2022) in a large-sample application of VIC in Chile.

On the contrary, NSE(E_{m}) was found to be most sensitive to the vegetation parameters, LAI_{f} and $rmin_{\mathrm{f}}$ being the most important vegetation parameters according to the RSA sensitivity indices (Fig. 7.5, 7.6). This is in line with the VIC parameter sensitivities reported in Chapter 6 and in Sepúlveda et al. (2022), suggesting that VIC vegetation parameters have a significant potential to improve the representation of evaporative processes if included in calibration. From among soil parameters, d_2 was the most important parameter to NSE(E_{m}), which could be related to the water uptake by vegetation in the root zone as it is directly affected by the thickness of the VIC soil layers.

7.5.2 Model performance during the baseline period

The large-sample application of the VIC model provided valuable insights into the VIC performance for streamflow and evaporation in the 189 studied catchments. The capability of the VIC model to produce acceptable estimates of NSE(Q_{d}) and NSE(E_{m}) simultaneously was tested by following a two-fold calibration approach based on the weighted Euclidean distance definition for two objectives (Eq. 7.1), namely Q-only and Q-E calibration. While the Q-only calibration led to satisfactory estimates of NSE(Q_{d}) and NSE(E_{m}) scores below 0 for more than half of the catchments (Fig. 7.7), the Q-E calibration substantially improved the NSE(E_{m}) estimates and still showed NSE(Q_{d}) values to those corresponding to the Q-only calibration. Similar conclusions were reached in Chapter 6 according to the five single-objective calibration experiments carried out for one test catchment located in the Guadalquivir River Basin.

The decomposition of NSE into r, α and β and their representation through cumulative distribution functions is a convenient way to facilitate the inter-comparison of multiple model implementations in large-sample studies (e.g.,

Mizukami et al., 2017; Rakovec et al., 2019; Yeste et al., 2020). The same decomposition can also be applied to other performance metrics based on these statistics such as the Kling-Gupta Efficiency (KGE, Gupta et al., 2009) to further understand the causes behind model performance (e.g., Aerts et al., 2022; Rakovec et al., 2016a,b). The VIC model results for NSE(Q_{d}), NSE(E_{m}) and their decomposition into r, α and β (Fig. 7.8, 7.9) revealed a similar performance for daily streamflow and monthly evaporation to that reflected in large-sample applications over the CONUS domain in Mizukami et al. (2017) and Rakovec et al. (2019). The VIC model performance for streamflow was also comparable to previous results for the Duero River Basin (Morán-Tejeda et al., 2014; Yeste et al., 2020), Tajo (Pellicer-Martínez and Martínez-Paz, 2018; Pellicer-Martínez et al., 2021), Guadalquivir (Yeste et al., 2018), Segura (Pellicer-Martinez and Martínez-Paz, 2015a; Pellicer-Martínez et al., 2015b) and Júcar (Marcos-Garcia et al., 2017; Suárez-Almiñana et al., 2020).

As described in Chapter 6, the integration of streamflow and evaporation data into model calibration is ultimately subject to the law of conservation of mass and the magnitude of the imbalance resulting from merging three independent datasets of precipitation, streamflow and evaporation. If the imbalance is significant and the period is long-enough, this will unavoidably result in a data-driven trade-off for the streamflow and evaporation simulations as the model will produce a steady-state solution that cannot match the mean values of streamflow and evaporation data simultaneously. Such condition was thoroughly checked for the two VIC calibrations using meteorological observations and WRF model simulations as model forcings (Fig. 7.10), and results indicated that the bias in the runoff ratio (i.e., Q/P) was higher for the Q-E calibration while its magnitude remained similar for all the datasets.

This limitation will be further explored in future implementations of the VIC model for the Spanish catchments to produce seamless distributed parameters maps and Spanish-wide simulations based on a fully gridded implementation. For the purpose of this study, the Q-E calibration was selected as the reference implementation to carry out the hydrologic projections according to an additional evaluation of model performance using the WRF model simulations (Fig. 7.11). Similarly to the monthly simulations in Yeste et al. (2021) using an ensemble of 18 Euro-CORDEX model experiments, NSE was found to be too demanding for evaluating the model performance for daily streamflow because the daily streamflow dynamics (i.e., r_{Q}) could be hardly captured. In the case of monthly evaporation, the computed NSE(E_{m})

scores were much higher and always comparable to the values corresponding to the meteorological observations.

7.5.3 Future projections

The spatial patterns of the hydrologic delta changes calculated for WRFCCSM and WRFMPI are in agreement in most of the Spanish domain, particularly for the long-term future period under both RCPs (Fig. 7.12, 7.13, S7.6, S7.7). The magnitude of the projected changes, however, differs for both WRF datasets depending on the future scenario, which is in line with findings in García-Valdecasas Ojeda et al. (2020a,b) for the Iberian Peninsula using the same datasets. The inter-model variability can be related to the characteristics inherited from the lateral boundary conditions driving the WRF model simulations (García-Valdecasas Ojeda et al., 2020a,b), and is expected to become diluted by applying a bias correction method to historical data (Hakala et al., 2018; Papadimitriou et al., 2016; Pastén-Zapata et al., 2020; Yeste et al., 2021).

The projected decreases of P are in accordance with results in Argüeso et al. (2012) and García-Valdecasas Ojeda et al. (2020a) based on WRF model simulations for Spain and the Iberian Peninsula, respectively, revealing a more pronounced effect of climate change for the southern catchments. Contrarily to the rest of the seasons, P is expected to slightly increase in some northern and central catchments by the end of the century, yet this compensatory effect is generally non-significant and remains unnoticeable at annual scale. This confirms previous research in Argüeso et al. (2012) and García-Valdecasas Ojeda et al. (2020a) for the northern half of the country.

The partitioning of P into Q and E is expected to shift towards an enhanced evaporative regime as a consequence of climate change in the Spanish catchments. This important shift may pose a threat for the future water security and the availability of water resources in the region, where the occurrence of hydroclimatic extremes in the form of more frequent and longer-lasting droughts is envisaged to increase (García-Valdecasas Ojeda et al., 2021a,b). At a smaller scale, comparable results were reported in Yeste et al. (2021) and Fonseca and Santos (2019) for the Spanish and Portuguese parts of the Duero River Basin, respectively, where Q is projected to dramatically decrease at annual and seasonal time scales.

The delta changes of E are contingent upon the future water availability and the upper limit imposed by E_p, which is projected to uniformly increase for all the future scenarios (Fig. S7.8, S7.9). As a consequence, E is the hydrologic variable subject to the highest intra-annual variability, with increases approximating the E_p limit in winter and decreases reflecting a progressive shortening of water availability for the rest of the seasons. These results partially confirm the findings of García-Valdecasas Ojeda et al. (2020a) for E over the Spanish domain, the main differences being concentrated in winter as significant increases were not detected for the long-term future period under both RCPs in this study. Similarly to Yeste et al. (2021), the disparity in the projected changes of E with respect to García-Valdecasas Ojeda et al. (2020a) can be attributed to the calibration approach followed to attain a more accurate representation of the water balance in the region (see also Yeste et al., 2020).

Finally, TWS changes provide key information of the subsurface water dynamics as an aggregate estimate of soil moisture changes. TWS decreases evidenced a steady movement towards soil dryness for all the future scenarios, resulting in less availability of water for the runoff generation and the evaporative processes. The future behaviour projected for TWS is in agreement with the soil moisture characterization reported in García-Valdecasas Ojeda et al. (2020a) based on the Soil Moisture Index (SMI, Seneviratne et al., 2010). The main difference between TWS and SMI lies in the thickness of the soil profile considered to calculate them: while SMI was calculated using the upper 1 m of soil implemented for the WRF model simulations (García-Valdecasas Ojeda et al., 2020a), TWS aggregates the whole soil profile, which in this work was fixed at 2 m depth according to the VIC model implementation. It is expected that TWS provides a better representation of the soil state for the purpose of this study as all the three soil layers included in the VIC model parameterization are hydrologically active. In any case, both TWS and SMI are soil-state variables that appropriately capture the subsurface water dynamics and have a strong potential to quantify and explain many hydroclimatic processes (García-Valdecasas Ojeda et al., 2020a; Güntner et al., 2017; Heistermann et al., 2022; Ruosteenoja et al., 2018).

7.6 Conclusions

The main findings of this study can be summarized as follows:

- The regional sensitivity analysis helped identify the parameter sensitivities with respect to the selected metrics to evaluate the performance of the VIC model against daily streamflow and monthly evaporation data in all the studied catchments. The soil and routing parameters were revealed as the most important parameters to the streamflow performance, while the influence from the vegetation parameters was negligible. The VIC performance for evaporation was mostly controlled by the soil parameters and two of the vegetation parameters.

- The VIC model was calibrated following a single-objective optimization approach applied to the streamflow simulations exclusively and to the streamflow and evaporation simulations simultaneously. The performance of the VIC model was evaluated for two independent periods, suggesting that it is possible to achieve satisfactory adjustment to both hydrologic variables at the same time if their performance metrics are combined into a composite function that aggregates the individual performance for streamflow and evaporation.

- An additional evaluation of the performance of the VIC model for the streamflow and evaporation simulations was carried out using two regional climate model datasets as meteorological forcings. While the VIC performance for streamflow was poor in comparison to the simulations conducted for the meteorological observations due to the inability of the climate datasets to reproduce the daily dynamics of streamflow, the performance for monthly evaporation was high and remained close to that corresponding to the meteorological observations.

- The impacts of climate change in the Spanish catchments were assessed for the two regional climate model datasets, and the projected hydrologic changes were analyzed for the VIC model simulations of water storages and fluxes for several future scenarios. An enhanced evaporative regime was projected for the future partitioning of precipitation into streamflow and evaporation, and the diminutions of terrestrial water storage indicated a shift towards soil drying conditions for all the studied catchments.

8 General conclusions

The modelling framework developed in this thesis has been successfully applied in a wide range of Spanish catchments and has allowed for the examination of the VIC model performance and its predictive capabilities concerning multiple hydrologic variables. What follows is an overview of the main conclusions reached in this work. These are organized into three different categories addressing the three stages of the hydrological modelling framework: **sensitivity analysis**, **calibration and evaluation**, and **hydrologic projections**. Specific conclusions for each of the four study cases analyzed in this thesis can be found in Chapters 4 to 7.

Sensitivity analysis

Four sensitivity analysis methods were applied to study parameter sensitivities in relation to the different components of the water balance and the performance of the VIC model against streamflow and evaporation data. The main findings for the sensitivity analysis stage can be summarized as follows:

- A *global sensitivity analysis* was applied to quantify the effect of the five soil calibration parameters on the water balance variables for 31 headwater catchments located in the Duero River Basin (Chapter 4). The sensitivity indices for the total runoff resulted from the compounding effect of the surface runoff and baseflow component, and the sensitivity indices for actual evaporation reflected an identical magnitude to the indices for total runoff but with opposite sign due to the model reaching the steady state during the study period. The five soil parameters were important to the soil moisture simulations as they govern the equations describing the water storages and fluxes implemented in VIC for the three soil layers.
- A *hybrid local-global sensitivity analysis* method was applied to an initial selection of 20 soil, vegetation and routing parameters as the parameter screening procedure in three studied catchments located in the

Guadalquivir and Duero River Basins (Chapter 6). The most influential parameters to the performance of the VIC model for daily streamflow and monthly evaporation were chosen in each case, and the selection was further refined during the *Pareto-based sensitivity analysis* following a multi-objective optimization approach without constraints. This strategy facilitated the identification of the most parsimonious parameterization leading to an adequate approximation to streamflow and evaporation data simultaneously, and only required adding two extra vegetation parameters to the soil and routing parameterization.

- A *regional sensitivity analysis* was applied to the same initial selection of 20 soil, vegetation and routing parameters to identify the most influential parameters to the VIC performance of daily streamflow and monthly evaporation for a total of 189 headwater catchments distributed across the Spanish territory (Chapter 7). The distributed maps of sensitivity indices revealed the spatial variability in parameter sensitivities for the Spanish domain. The streamflow performancThe modelling framework developed in this thesis has been successfully applied in a wide range of Spanish catchments and has allowed for the examination of the VIC model performance and its predictive capabilities concerning multiple hydrologic variables. What follows is an overview of the main conclusions reached in this work. These are organized into three different categories addressing the three stages of the hydrological modelling framework: **sensitivity analysis**, **calibration and evaluation**, and **hydrologic projections**. Specific conclusions for each of the four study cases analyzed in this thesis can be found in Chapters 4 to 7.e was mostly driven by the soil and routing parameters, with little influence from the vegetation parameters. The soil parameters and two vegetation parameters were identified as important to the performance for evaporation. This in line with previous results in this thesis for the hybrid local-global and the Pareto-based sensitivity analyses.

Calibration and evaluation

The calibration of the VIC model was carried out following three optimization approaches focused on streamflow and evaporation, and the model performance for the streamflow and evaporation simulations was evaluated at

different temporal resolutions. The following conclusions were drawn from the model calibration and evaluation stage:

- A *single-objective calibration* approach against monthly streamflow observations was applied to the 31 studied catchments in Duero, and the performance of the VIC model for monthly streamflow and monthly evaporation was evaluated for two independent periods (Chapter 4). The VIC model reflected satisfactory adjustments to monthly streamflow and evaporation data and improved the benchmark performance. A high degree of parameter transferability in space was detected in relation to the VIC performance for streamflow, and equifinality in model parameters was observed for many catchments where a considerable number of parameter combinations produced a similar performance estimate for streamflow to that calculated during calibration.

- A *multi-objective calibration* approach was undertaken to adjust the most influential parameters according to the hybrid local-global and the Pareto-based sensitivity estimates for three studied catchments (Chapter 6) and optimize the joint model representation of daily streamflow observations and monthly evaporation data. Constraints were applied to the objective space to optimize model performance within the limits of acceptability established for streamflow and evaporation. The trade-offs in model performance for both variables were well captured in the resulting Pareto front, and indicated that a significant improvement for the evaporation representation can be attained at the expense of a very reduced loss in model performance for streamflow.

- A *composite-single-objective calibration* approach consisting of two strategies was adopted in 189 catchments across Spain to optimize 1) the model representation of daily streamflow , and 2) the joint model performance for daily streamflow and monthly evaporation (Chapter 7). The calibration parameters were selected according to the sensitivity estimates from the regional sensitivity analysis, and the performance of the VIC model for the streamflow and evaporation simulations was evaluated for two independent periods. The combination of the performance metrics of streamflow and evaporation into a composite-single-objective function does not lead to a significant deterioration of the streamflow representation and makes it possible to adequately reproduce both hydrologic variables at the same time.

Hydrologic projections

Hydrologic projections for multiple water balance components were carried out using the calibrated VIC and meteorological forcings gathered from different regional climate change simulations to study the impacts of climate change in the Spanish catchments. The main findings for the hydrologic projections stage are as follows:

- *Hydrologic projections using Euro-CORDEX data* were performed for a selection of 24 headwater catchments located in Duero where the VIC model produced higher performance estimates for monthly streamflow during calibration and evaluation (Chapter 5). The ability of the Euro-CORDEX ensemble to reproduce monthly streamflow and evaporation data was evaluated and compared against the model performance using meteorological observations, showing that, while the performance for streamflow was deteriorated with Euro-CORDEX, the evaporation performance was remarkably improved. This could be potentially due to the model-nature of the satellite-based evaporation dataset[1].

- The projected hydrologic changes in the Duero River Basin evidenced a shift towards an enhanced evaporative regime regarding the partitioning of precipitation into streamflow and evaporation for all the studied catchments. The streamflow diminutions in the northern catchments could be explained by the precipitation decreases and the evaporation increases projected for them, while the streamflow reductions in the southern catchments were driven by a lesser availability of water resources. The snowmelt contribution to the runoff generation was projected to highly diminished in future scenarios and become almost non-existent in catchments where the snowmelt contribution to runoff was relevant under present climate.

- *Hydrologic projections using WRF data* were carried out for the complete set of 189 Spanish headwater catchments to study the impacts of climate change at national scale (Chapter 7). These simulations were implemented separately for the two calibration experiments integrating the composite-single-objective calibration approach. The performance of the VIC model against daily streamflow and monthly evaporation was evaluated for

[1]This, however, has not been observed for the complete set of 189 Spanish catchments after a calibration focused on daily streamflow alone, as will be stated later.

the WRF datasets and was compared to the performance estimates attained using meteorological observations. The poor performance for daily streamflow simulations using WRF data was associated to their inability to reproduce the daily dynamics of the streamflow observations. The performance for monthly evaporation, in turn, was highly affected by the calibration strategy and was comparable to that corresponding to the meteorological observations.

- Similarly to the Duero River Basin, a tendency towards a more evaporative regime was also appreciated for the 189 Spanish catchments, and the latitudinal gradient of the annual and seasonal changes projected for streamflow and evaporation could be satisfactorily captured with the VIC model simulations performed for the complete set of catchments. An additional analysis of the changes in terrestrial water storage revealed increasing soil drying conditions towards the end of the 21th century for the Spanish catchments.

This thesis represents an important contribution to the field of large-sample hydrology and a major advance in the application of multi-criteria approaches aimed at integrating streamflow observations and satellite-based evaporation data. This investigation will also help steer future developments to improve the calibration and evaluation of hydrologic models and provide more reliable hydrologic projections for the future climate. In particular, the hydroclimatic community and water practitioners in Spain can strongly benefit from these results as they provide a baseline reference for the evaluation of different modelling approaches as well as other hydrologic models.

Future research

The contributions of this thesis will be further extended in the near future to explore multiple research lines on the basis of the large-scale implementation of the VIC model for the Spanish catchments. The great amount of hydrologic information will soon become supported with complementary data analysis and modelling approaches that will help strengthen the key results and conclusions from this thesis and will help achieve a more reliable characterization of the water balance for the Spanish domain.

In particular, the research is planned to be extended to other subcatchments located within the topographic area of the studied catchments where there

exists additional streamflow information from upstream gauges. This is an adequate step to assess the parameter transferability in space, and an early answer to the feasibility of undergoing a posterior parameter regionalization approach to estimate seamless parameter fields and develop a Spanish-wide gridded implementation of the VIC model.

The Spanish-wide implementation of VIC will facilitate the comparison against the monthly SIMPA simulations available for streamflow, evaporation and additional hydrologic variables that are annually updated. The calibration of VIC against daily streamflow observations and monthly evaporation data is expected to produce a more reliable quantification of the water balance in Spain as the VIC model has been calibrated to adequately reproduce the daily streamflow dynamics and the predominance of evaporative fluxes in the region has been appropriately weighted.

The use of remote sensing data of subsurface water storages and other sources of evaporation data will be key to advance the evaluation of the VIC simulations. For instance, GLEAM is regularly updated and currently delivers an evaporation dataset exclusively based on satellite data in addition to the evaporation dataset used in this thesis. Moreover, the VIC model simulations of terrestrial water storage will be evaluated against other databases such as GRACE to analyze model consistency for other relevant processes not considered during calibration.

Finally, the emerging field of decadal predictions is intended to provide a more reliable characterization of the future climate for a planning horizon of 1 to 10 years through climate simulations where the soil states have been initialized with observational data. Decadal simulations can be used to forced VIC and assess the hindcast and forecast skills of decadal predictions systems from a hydrological perspecive. These simulations will provide a solid foundation for the future water planning and the development of adaptation strategies that can help mitigate the impacts of climate change in the Spanish catchments.

Conclusiones generales

El ejercicio de modelización desarrollado en esta tesis ha sido aplicado exitosamente en un amplio rango de cuencas españolas y ha permitido examinar el ajuste del modelo VIC y sus capacidades predictivas en relación a múltiples variables hidrológicas. Lo que sigue es un resumen de las principales conclusiones de este trabajo. Éstas se han organizado en tres categorías correspondientes a las tres etapas de la modelización hidrológica: **análisis de sensibilidad**, **calibración y evaluación**, y **proyecciones hidrológicas**. Las conclusiones específicas para cada uno de los casos de estudio analizados en esta tesis pueden encontrarse en los Capítulos 4 al 7.

Análisis de sensibilidad

Se aplicaron cuatro métodos de análisis de sensibilidad para estudiar la sensibilidad de los parámetros en relación a los distintos componentes del balance de agua y al ajuste del modelo VIC frente a datos de caudal y evaporación. Los principales hallazgos para la etapa de análisis de sensibilidad pueden resumirse como sigue:

- Se aplicó un *análisis de sensibilidad global* para cuantificar el efecto de los cinco parámetros del suelo sobre las variables del balance de agua en 31 cuencas de cabecera localizadas en la Demarcación Hidrográfica del Duero (Capítulo 4). Los índices de sensibilidad de la escorrentía total resultaron de un efecto acumulado de las componentes superficial y subterránea de la escorrentía, y los índices de sensibilidad para la evaporación real reflejaron una magnitud idéntica a los de la escorrentía total pero con signo opuesto debido a que el modelo alcanzó el estado estacionario durante el período de estudio. Los cinco parámetros del suelo fueron importantes para las simulaciones de humedad del suelo, dado que gobiernan las ecuaciones de los almacenamientos y los flujos de agua implementados en el modelo VIC para las tres capas de suelo.

- Se aplicó un *análisis de sensibilidad híbrido local-global* sobre un conjunto de 20 parámetros del suelo, la vegetación y la propagación del caudal para realizar una preselección de parámetros en tres cuencas de estudio localizadas en las Demarcaciones Hidrográficas del Guadalquivir y del Duero (Capítulo 6). Los parámetros más influyentes sobre el ajuste del modelo para el caudal diario y la evaporación mensual fueron seleccionados en cada caso, y la selección se refinó durante el *análisis de sensibilidad basado en Pareto* siguiendo un proceso de optimización multi-objetivo sin restricciones. Esta estrategia facilitó la identificación de la parametrización más parsimoniosa conducente a una aproximación adecuada a los datos de caudal y evaporación simultáneamente, y solamente requirió añadir dos parámetros extra de la vegetación a la parametrización del suelo y la propagación del caudal.
- Se aplicó un *análisis de sensibilidad regional* sobre el mismo conjunto inicial de 20 parámetros del suelo, la vegetación y la propagación del caudal para identificar los parámetros más influyentes sobre el ajuste del modelo para el caudal diario y la evaporación mensual en 189 cuencas de cabecera distribuidas sobre el territorio español (Capítulo 7). Los mapas distribuidos de los índices de sensibilidad revelaron la variabilidad espacial de las sensibilidades de los parámetros para el dominio español. El ajuste del caudal estaba principalmente ligado a los parámetros del suelo y de la propagación del caudal, con poca influencia de los parámetros de la vegetación. Se identificaron como importantes para el ajuste a los datos de evaporación los cinco parámetros del suelo y dos parámetros de la vegetación. Esto se encuentra en línea con los resultados obtenidos durante el análisis de sensibilidad híbrido y al análisis de sensibilidad basado en Pareto.

Calibración y evaluación

La calibración del modelo VIC se llevó a cabo siguiendo tres estrategias de optimización centradas en el caudal y la evaporación, y el ajuste del modelo para las simulaciones de caudal y evaporación fue evaluado a distintas resoluciones temporales. Estas son las conclusiones para la etapa de calibración y evaluación:

- Se aplicó una estrategia de *calibración de un solo objetivo* frente a observaciones mensuales de caudal en 31 cuencas de la Demarcación

Hidrográfica del Duero, y se evaluó el ajuste del modelo para el caudal mensual y la evaporación mensual en dos períodos independientes (Capítulo 4). El modelo VIC reflejó ajustes satisfactorios al caudal mensual y la evaporación mensual, mejorando el ajuste seleccionado como referencia comparativa. Se detectó un alto grado de transferabilidad espacial de los parámetros calibrados en relación al ajuste del caudal, y se observó equifinalidad en los parámetros del modelo para muchas cuencas en las que un número considerable de combinaciones de parámetros produjo un ajuste similar para el caudal al calculado durante la calibración.

- Los parámetros más influyentes de acuerdo con el análisis de sensibilidad híbrido y el análisis de sensibilidad basado en Pareto fueron ajustados mediante *calibración multi-objetivo* en tres cuencas de estudio para optimizar el ajuste simultáneo frente a datos diarios de caudal y datos mensuales de evaporación (Capítulo 6). Se aplicaron restricciones sobre el espacio objetivo para optimizar el ajuste del modelo dentro de los límites de aceptabilidad establecidos para el caudal y la evaporación. La frontera de Pareto capturó adecuadamente el efecto compensatorio en el ajuste a ambas variables, sugiriendo que es posible obtener una mejora significativa del ajuste a la evaporación a consta de una pérdida muy reducida del ajuste para el caudal.

- Se adoptaron dos estrategias de *calibración compuesta de un solo objetivo* en 189 cuencas en España para optimizar 1) la representación en el modelo del caudal diario, y 2) el ajuste simultáneo al caudal diario y la evaporación mensual (Capítulo 7). Los parámetros de la calibración fueron seleccionados de acuerdo con las estimaciones de sensibilidad efectuadas con el análisis de sensibilidad regional, y el ajuste del modelo para las simulaciones de caudal y evaporación se evaluó en dos períodos independientes. La combinación de las métricas del caudal y la evaporación en una sola función objetivo no condujo a un deterioro del ajuste del caudal e hizo posible reproducir adecuadamente ambas variables hidrológicas a la vez.

Proyecciones hidrológicas

Se obtuvieron proyecciones hidrológicas para los múltiples componentes del balance de agua usando el modelo VIC calibrado y empleando forzamientos meteorológicos extraídos de diferentes simulaciones regionales de cambio

climático para estudiar los impactos del cambio climático en las cuencas españolas. Las principales conclusiones para la etapa de las proyecciones hidrológicas son las siguientes:

- Se realizaron *proyecciones hidrológicas usando datos de Euro-CORDEX* en una selección de 24 cuencas de cabecera localizadas en la Demarcación Hidrográfica del Duero para las que el modelo VIC produjo un mejor ajuste del caudal mensual durante la calibración y la evaluación (Capítulo 5). Se evaluó la habilidad del ensamblado de Euro-CORDEX para reproducir el caudal y la evaporación mensual y se comparó frente al ajuste usando observaciones meteorológicas, demostrando que, mientras que el ajuste para el caudal empeoró con Euro-CORDEX, el ajuste de la evaporación mejoró notablemente. Esto pudo deberse a la naturaleza de modelo de la base de datos de evaporación empleada[1].

- Los cambios hidrológicos proyectados en la Demarcación Hidrográfica del Duero evidenciaron un cambio hacia un régimen evaporativo más pronunciado en relación con el particionamiento de la precipitación en caudal y evaporación para todas las cuencas de estudio. Las disminuciones del caudal en las cuencas del norte quedaron explicadas por los descensos de precipitación y los aumentos de evaporación proyectados, mientras que los descensos del caudal en las cuencas del sur estaban ligados a una menor disponibilidad de recursos hídricos. Se proyectó un descenso muy alto de la contribución de la fusión nival a la generación de escorrentía, llegando prácticamente a desaparecer en cuencas donde la contribución de la fusión nival a la escorrentía era relevante bajo clima presente.

- Se llevaron a cabo *proyecciones hidrológicas usando datos del modelo WRF* para el conjunto de 189 cuencas españolas de cabecera de cara a estudiar los impactos del cambio climático a escala nacional (Capítulo 7). Estas simulaciones fueron implementadas por separado para los dos ejercicios de calibración que integraron el proceso de calibración compuesta de un solo objetivo. Se evaluó el ajuste del modelo VIC para el caudal diario y la evaporación mensual usando los datos del modelo WRF, y se comparó con los ajustes estimados empleando observaciones meteorológicas. El bajo ajuste al caudal diario empleando datos del modelo WRF estuvo

[1]Esto, sin embargo, no ha sido observado para el conjunto completo de 189 cuencas españolas después de una calibración centrada exclusivamente en el caudal diario, como se destacará después.

asociado a su incapacidad para reproducir las dinámicas diarias de las observaciones de caudal. El ajuste para la evaporación mensual, en cambio, se vio muy afectado por la estrategia de calibración, y fue comparable al ajuste obtenido con las observaciones meteorológicas.

- De forma similar a lo sucedido en el Duero, se apreció una tendencia hacia un régimen evaporativo más acentuado para las 189 cuencas españolas, y se pudo capturar satisfactoriamente el gradiente latitudinal de los cambios anuales y estacionales proyectados con el modelo VIC. Se efectuó un análisis adicional de los cambios en el almacenamiento terrestre de agua, indicando un incremento de las condiciones de aridez en las cuencas españolas hacia el final del siglo 21.

Esta tesis supone una contribución importante al campo de la hidrología macroescala y un avance en la aplicación de enfoques multicriterio centrados en la integración de observaciones de caudal y datos satelitales de evaporación. Esta investagación ayudará a conducir futuros desarrollos para mejorar la calibración y la validación de modelos hidrológicos y ayudará a obtener proyecciones hidrológicas más fiables bajo clima futuro. En particular, la comunidad hidroclimática en España puede beneficiarse de estos resultados dado que proporcionan una referencia base para la evaluación de distintas estrategias de modelización así como de distintos modelos hidrológicos.

Estudios futuros

Las contribuciones de esta tesis se extenderán próximamente para explorar múltiples líneas de investigación tomando como base la implementación macroescala del modelo VIC en las cuencas españolas. La gran cantidad de información hidrológica será pronto complementada con análisis de datos y estrategias de modelización adicionales que ayudarán a reforzar los resultados y la conclusiones clave de esta tesis y a conseguir una caracterización más fidedigna del balance de agua para el dominio español.

En particular, se pretende extender el estudio a otras subcuencas localizadas en el área topográfica de aquellas cuencas de estudio donde existen datos adicionales de caudal en estaciones de aforo localizadas aguas arriba. Esto es un paso adecuado para estudiar la transferabilidad espacial de parámetros, y una respuesta temprana a la posibilidad de llevar a cabo una estrategia de regionalización paramétrica para estimar mapas distribuidos de parámetros

y llevar a cabo una implementación del VIC para una malla cubriendo el dominio español.

La implementación mallada del VIC facilitará la comparación frente a las simulaciones mensuales del model SIMPA para el caudal, la evaporación y otras variables hidrológicas que se actualizan anualmente. Se espera que la calibración del modelo VIC frente a observaciones diarias de caudal y datos mensuales de evaporación sea capaz de producir un cuantifiación más realista del balance de agua en España dado que ha el modelo ha sido calibrado para reproducir adecuadamente las dinámicas diarias de caudal y la predominancia de flujos evaporativos ha sido ponderada adecuadamente.

El uso de datos de teledetección de los almacenamientos subsuperficiales de agua y otras fuentes de datos de evaporación serán clave para avanzar la evaluación de las simulaciones con el modelo VIC. Por ejemplo, GLEAM es actualizado regularmente y en la actualidad ofrece una base de datos de evaporación basada exclusivamente en datos de satélite junto con la base de datos usada en esta tesis. Además, las simulaciones del almacenamiento de agua terrestre serán evaluadas frente a otras bases de datos como GRACE para analizar la consistencia del modelo en relación a otros procesos hidrológicos relevantes que no se han considerado durante la calibración.

Finalmente, el campo emergente de las predicciones decenales tiene como objeto proporcionar una caracterización más fiable del clima futuro para un horizonte de planifación que abarca de 1 a 10 años por medio de simulaciones climáticas donde las condiciones iniciales de humedad del suelo se han fijado con datos observacionales. Las simulaciones decenales pueden ser empleadas para forzar el modelo VIC y determinar la capacidad predictiva de los sistemas de predicciones decenales desde un punto de vista hidrológico. Estas simulaciones constituirán una base sólida para la planificación hidrológica en el futuro y el desarrollo de estrategias de adaptación que puedan ayudar a mitigar los impactos del cambio climático en las cuencas de España.

Supporting Information

Supporting Information for Chapter 5

- 2 Supporting Tables
- 10 Supporting Figures

Table S5.1: Main characteristics of the selected catchments: area (km^2), mean altitude (m) and calibrated NSE value obtained in Yeste et al. (2020).

Code	Name	Area (km^2)	Mean altitude (m)	NSE
R-2001	Cuerda del Pozo	546.7	1319	0.852
R-2011	Arlanzón	106.7	1440	0.9237
R-2013	La Requejada	220.7	1378	0.8263
R-2014	Camporredondo	229.6	1673	0.8777
R-2026	Barrios de Luna	482.9	1496	0.8865
R-2027	Villameca	45.8	1180	0.8254
R-2028	Moncabril (Sistema)	62.9	1712	0.7475
R-2032	Riaño	592.3	1451	0.9409
R-2036	Linares del Arroyo	761.3	1111	0.6662
R-2037	Burgomillodo	803.1	1097	0.8222
R-2038	Santa Teresa	1845.4	1326	0.9145
R-2039	Águeda	788.4	895	0.9521
R-2042	Castro de las Cogotas	853.3	1279	0.8903
R-2043	Pontón Alto	150.4	1582	0.8322
GS-2005	Osma	893.1	1090	0.8307
GS-2016	Pajares de Pedraza	281.3	1298	0.835
GS-2041	Villalcázar de Sirga	307.7	929	0.7106
GS-2049	Cabañes de Esgueva	270.2	995	0.7333
GS-2051	El Espinar	36.7	1610	0.7743
GS-2057	Villovela de Pirón	202	1183	0.723
GS-2089	Morla de la Valdería	281.1	1369	0.9116
GS-2104	Villaverde de Arcayos	371.1	1146	0.8137
GS-2150	Pardavé	223.8	1448	0.7909
GS-2818	Rabal	597.9	678	0.8724

Table S5.2: Linear regression between the fractional changes of C_Y and the delta changes of Q at annual and seasonal time scales. All the r values are statistically significant at a 95% confidence level using the Student's t-test.

Linear regresion	Annual		Autumn		Winter		Spring	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
r	0.837	0.924	0.889	0.899	0.986	0.997	0.963	0.97
Slope	0.741	0.91	0.616	0.785	1.034	0.995	1.014	1.037
Intercept	-6.051	-7.127	-7.637	-9.52	-0.822	-0.924	-1.678	-2.467

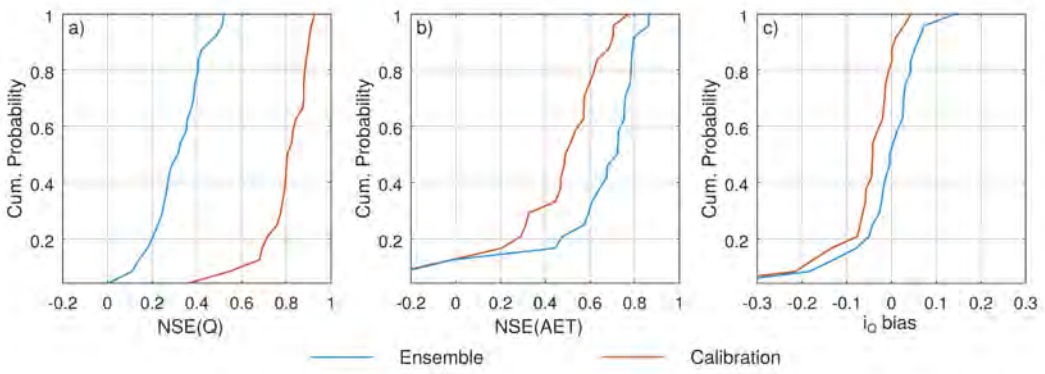


Fig. S5.1: CDFs of the VIC model performance for the period Oct 2000 - Sep 2011 corresponding to the combination historical+RCP8.5: a) NSE for the streamflow simulations against streamflow observations; b) NSE for the AET simulations against GLEAM; c) i_Q bias with respect to the ratio of SPREAD precipitation to streamflow observations. Blue lines represent the ensemble simulation, and orange lines correspond to the calibration results from Yeste et al. (2020).

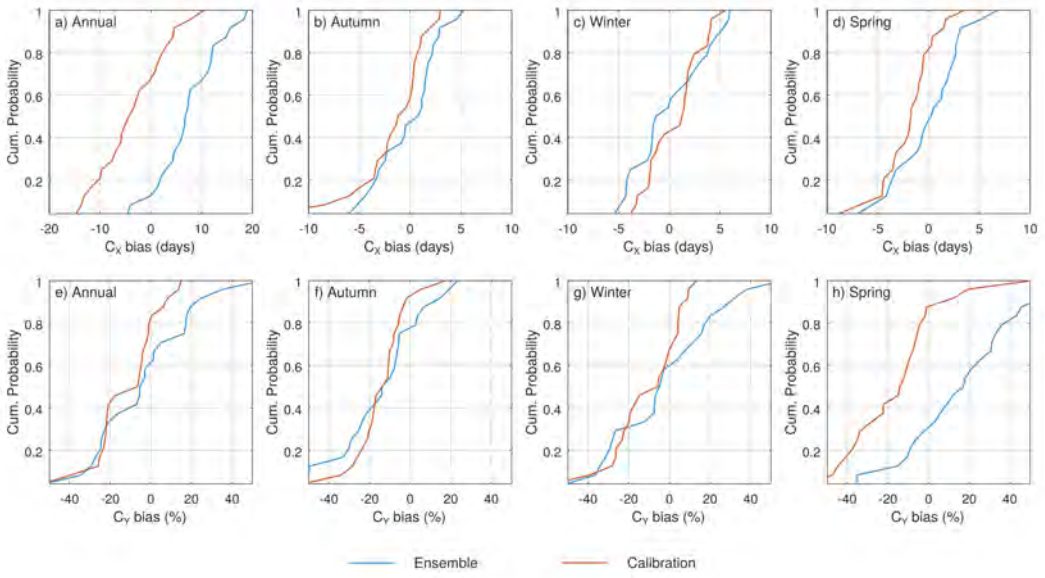


Fig. S5.2: CDFs of the VIC model performance for the annual and seasonal values of C_X (a to d) and C_Y (e to h) corresponding to the period Oct 2000-Sep 2011 and the combination historical+RCP4.5. C_X biases are calculated as the difference between simulated and observed values. C_Y biases represent fractional changes calculated as [(simulations - observations)/observations · 100]. Blue lines represent the ensemble simulation, and orange lines correspond to the calibration results from Yeste et al. (2020).

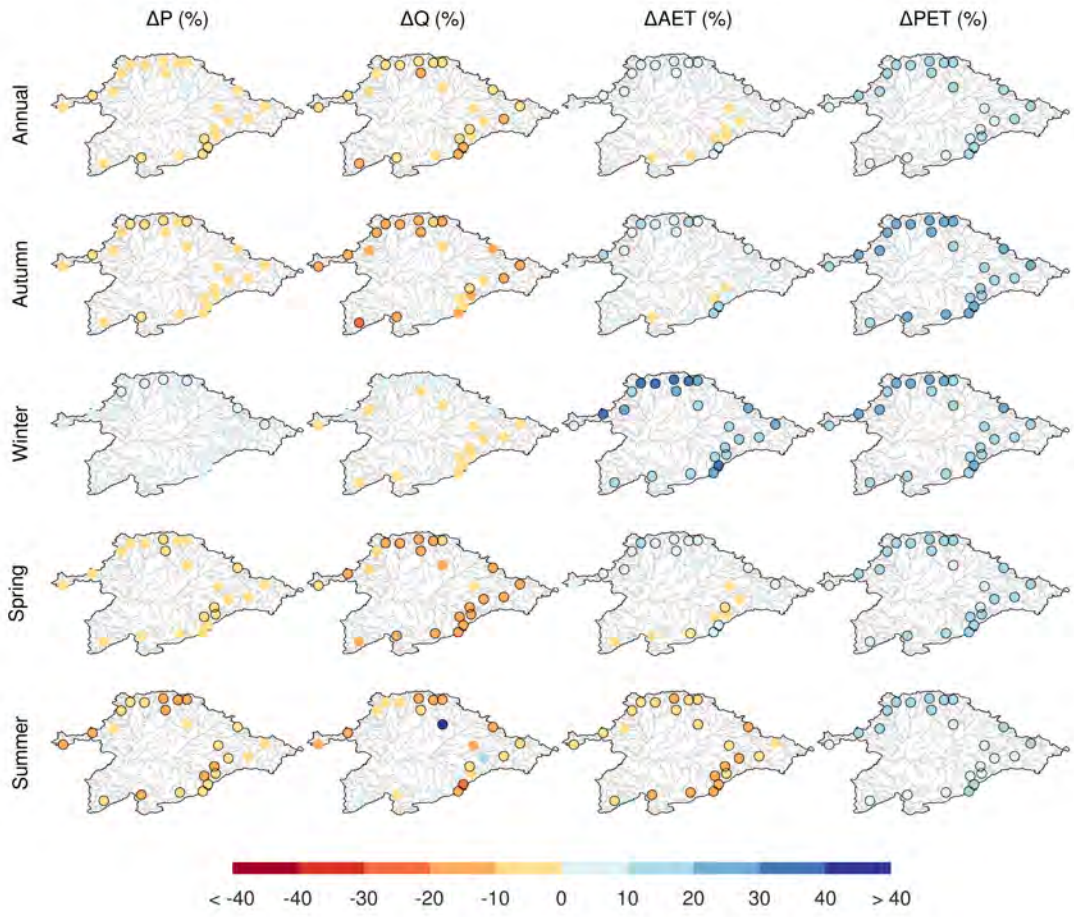


Fig. S5.3: Delta changes of annual and seasonal P, Q, AET and PET for the period 2021-2050 under the RCP4.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

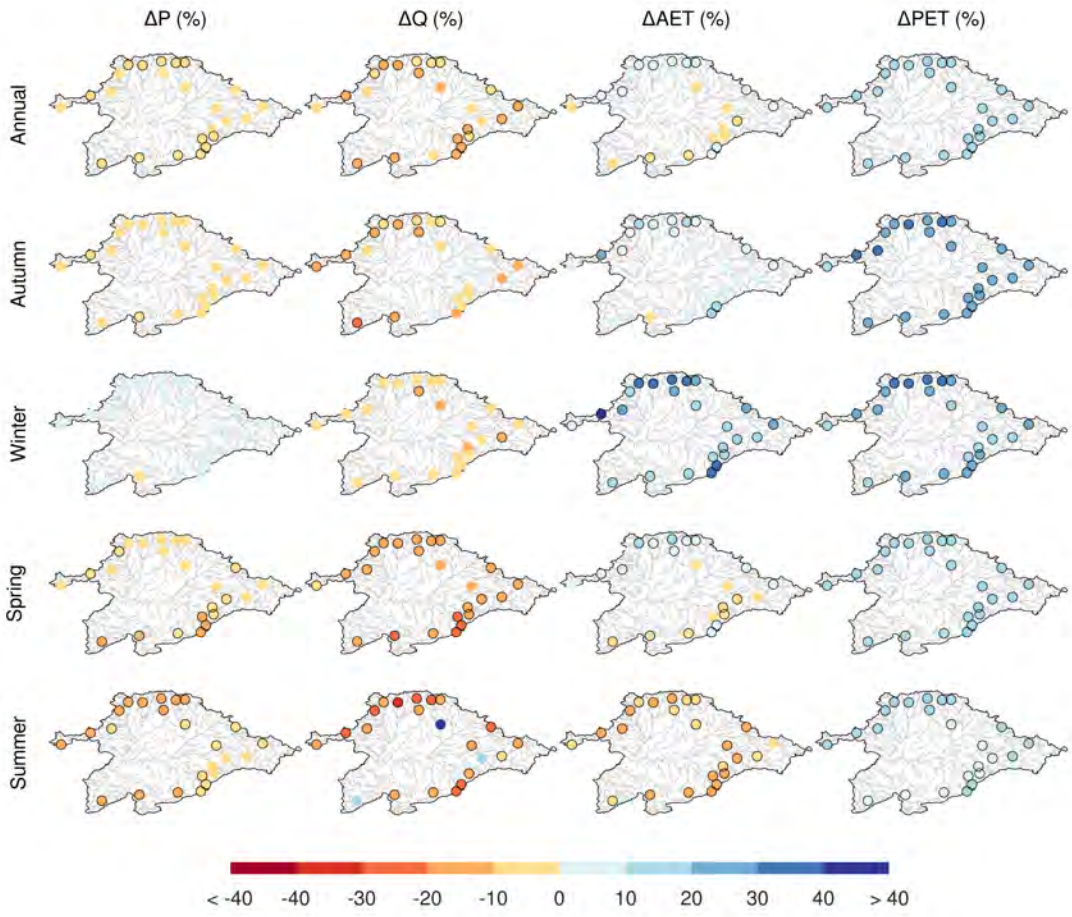


Fig. S5.4: Delta changes of annual and seasonal P, Q, AET and PET for the period 2021-2050 under the RCP8.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

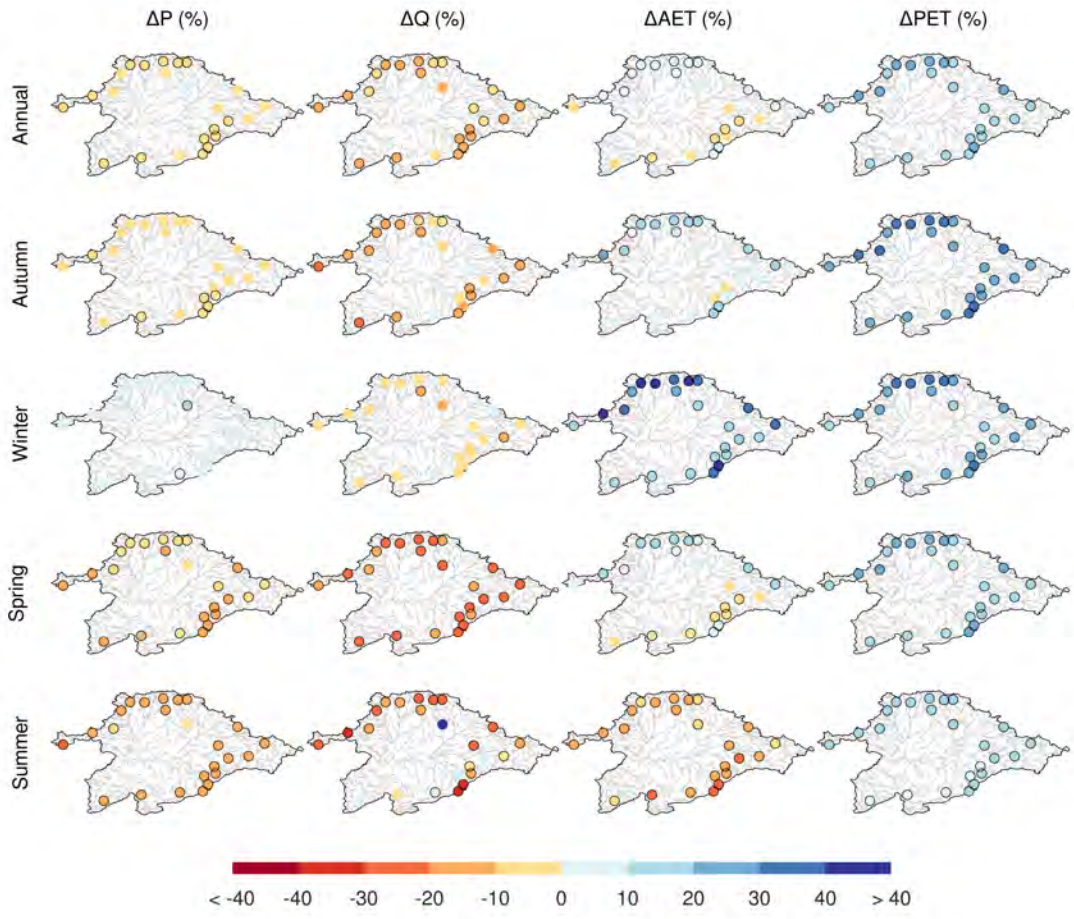


Fig. S5.5: Delta changes of annual and seasonal P, Q, AET and PET for the period 2041-2070 under the RCP4.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

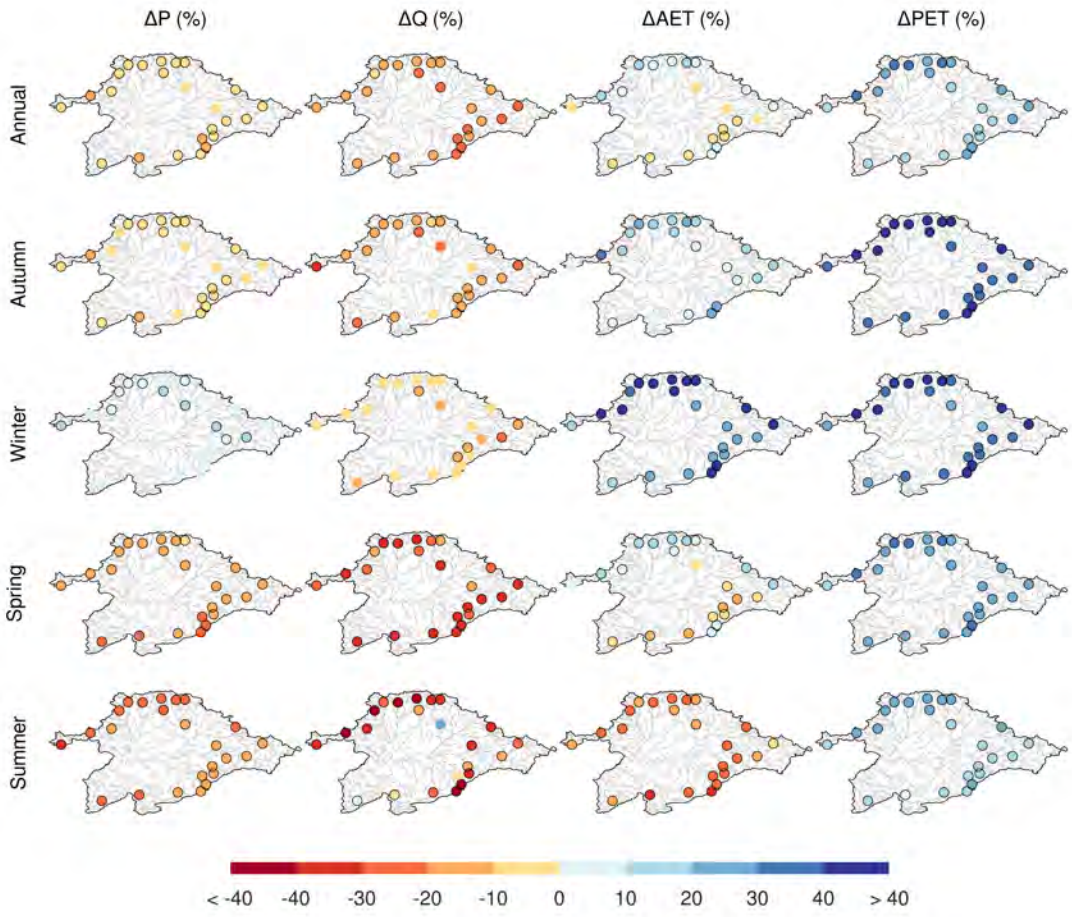


Fig. S5.6: Delta changes of annual and seasonal P, Q, AET and PET for the period 2041-2070 under the RCP8.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

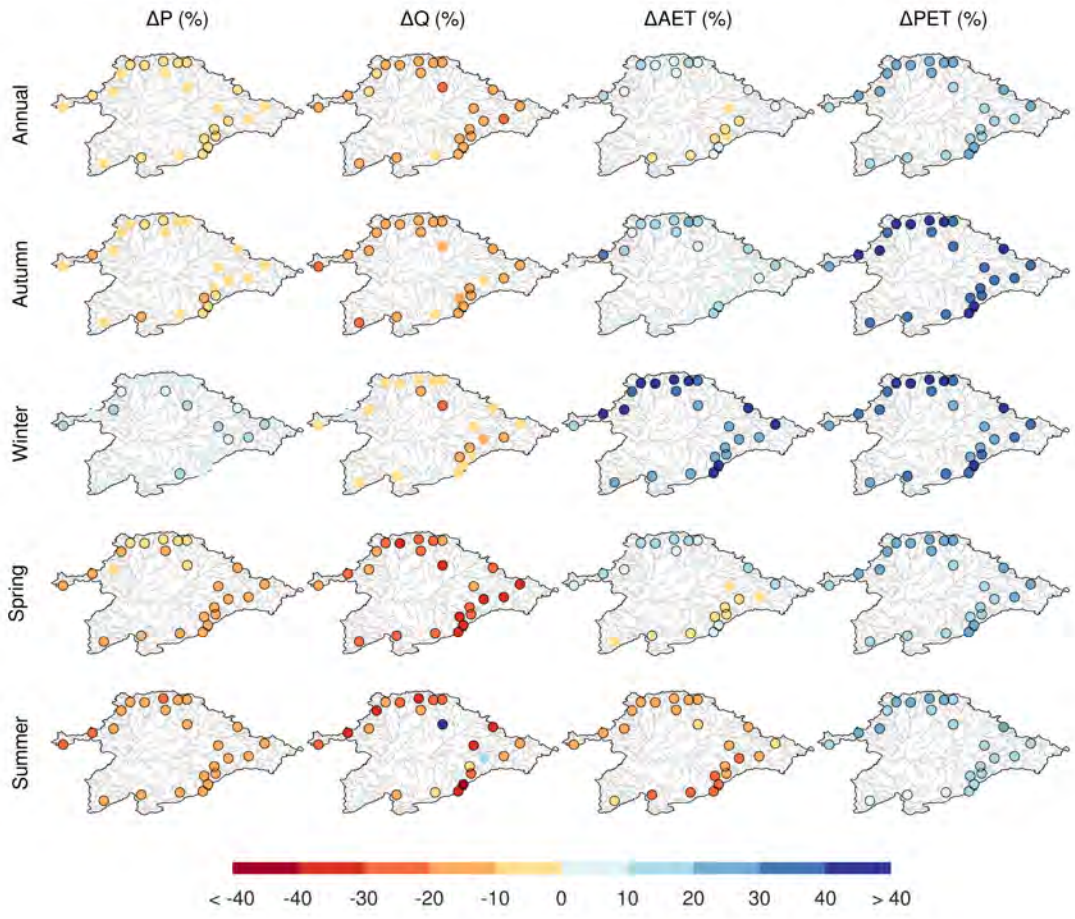


Fig. S5.7: Delta changes of annual and seasonal P, Q, AET and PET for the period 2071-2100 under the RCP4.5 scenario in the studied catchments. Significant changes at the 95% confidence level have been marked with solid borders.

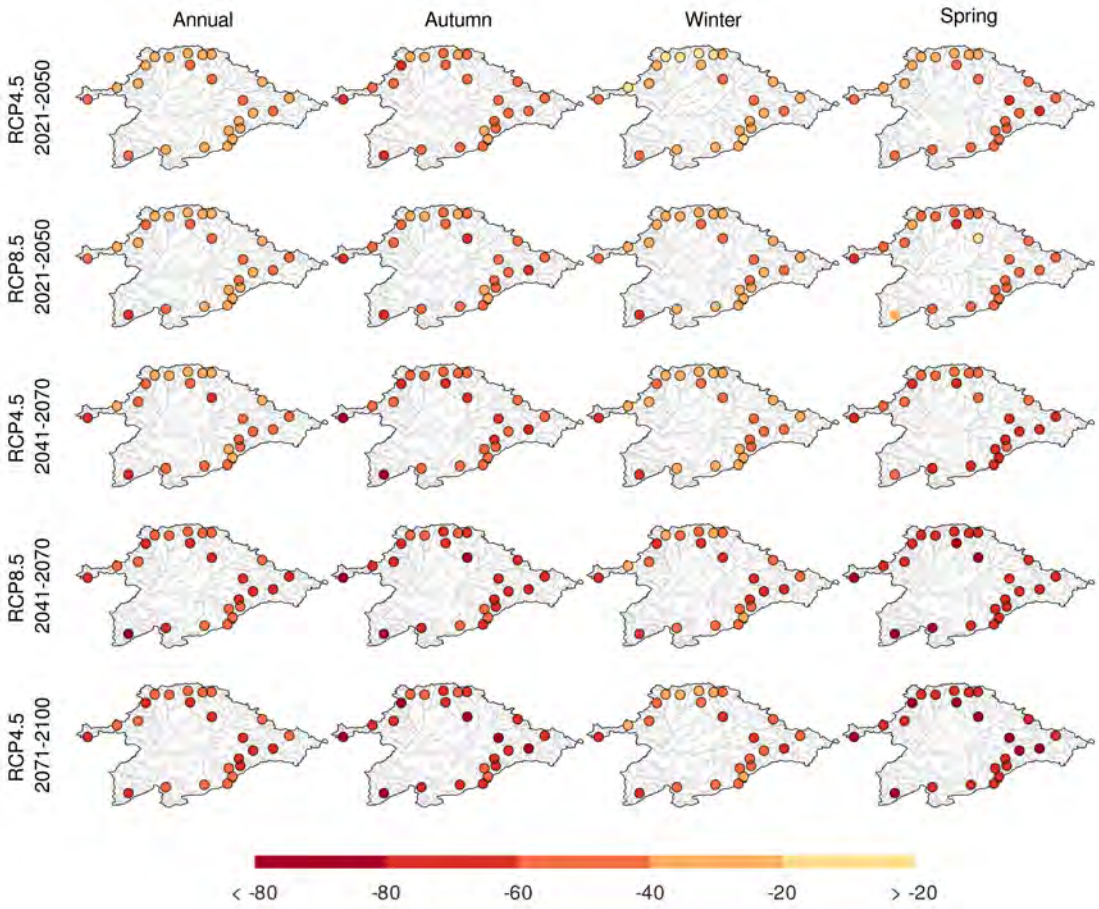


Fig. S5.8: Delta changes (excluding summer) of annual and seasonal Q_{snow} for the indicated periods and RCP scenarios in the studied catchments. Significant changes at a 95% confidence level have been marked with solid borders.

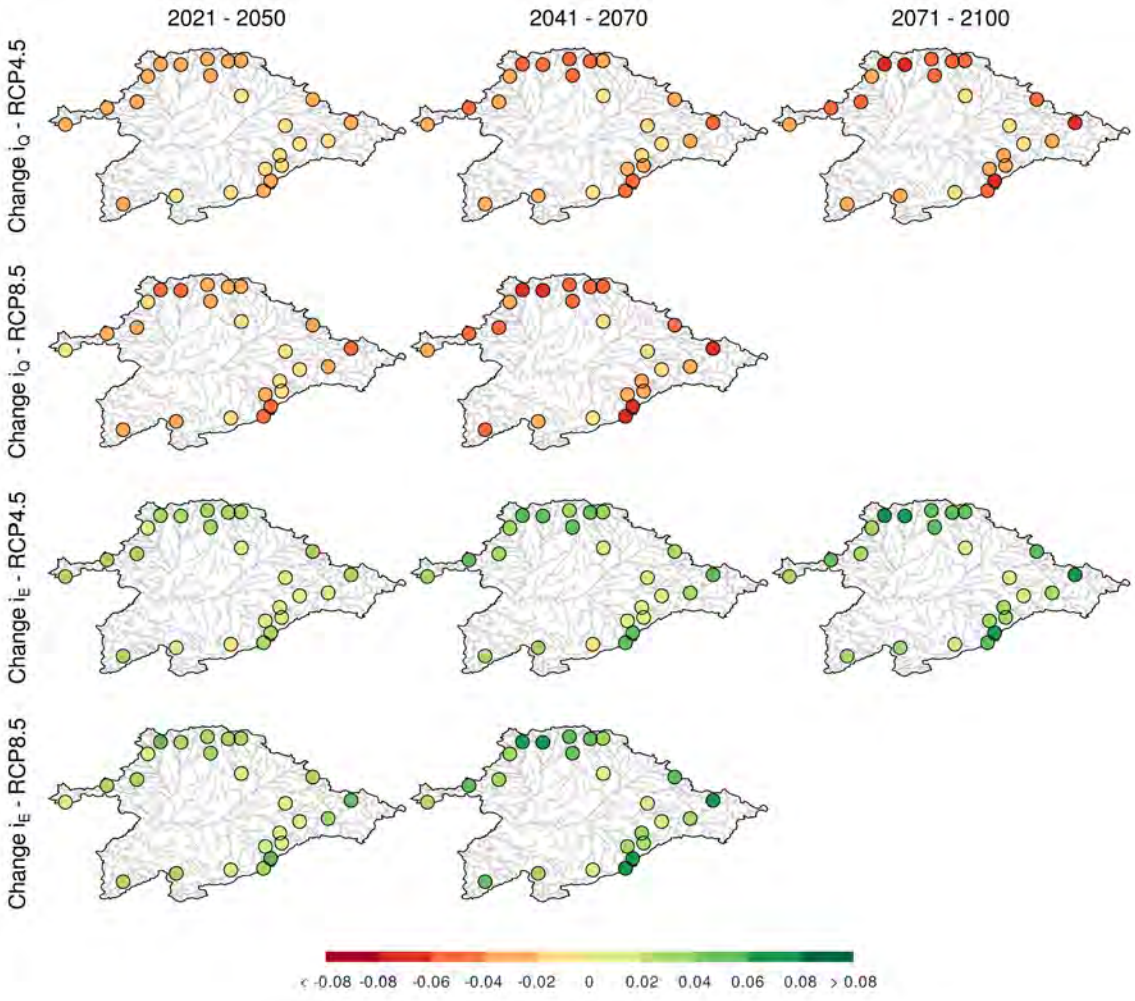


Fig. S5.9: Projected changes of i_Q and i_E as the difference between future and historical values for the indicated periods and RCP scenarios.

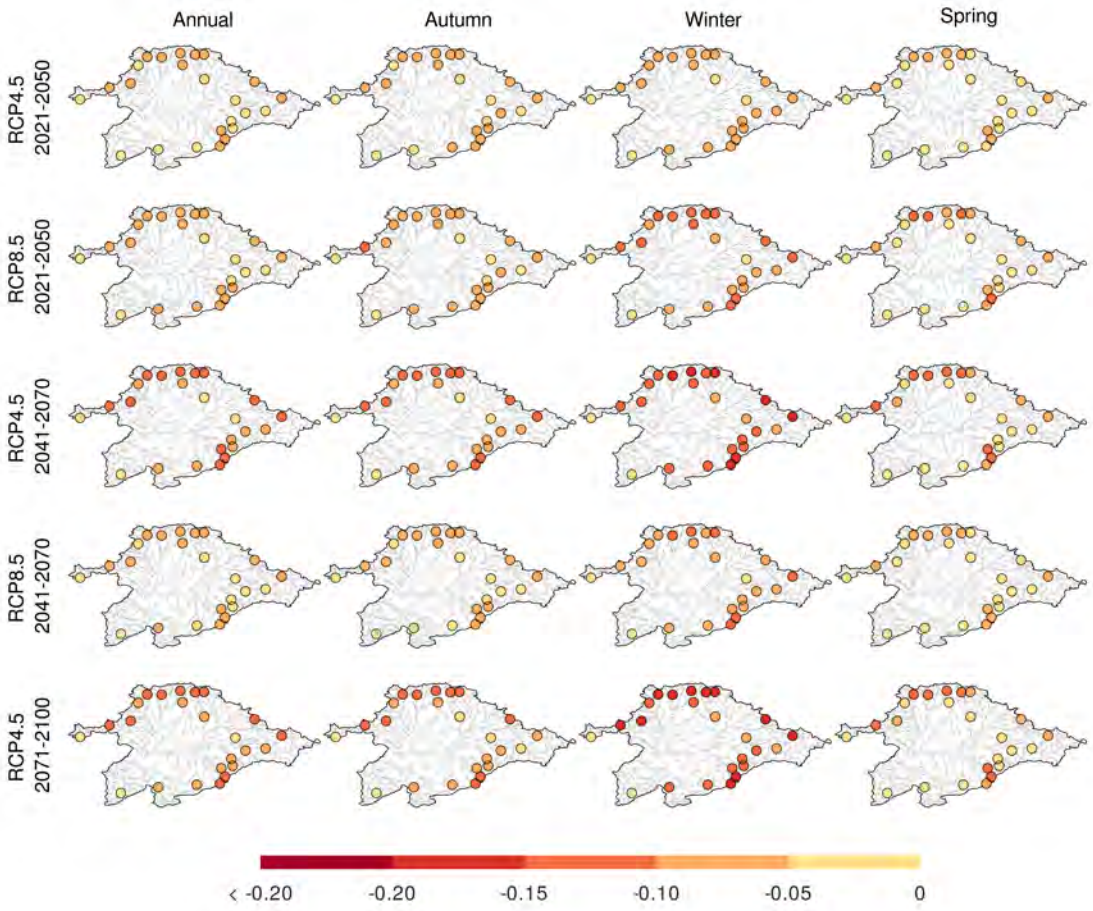


Fig. S5.10: Projected changes of annual and seasonal $Q_{snow}ratio$ (excluding summer) calculated as the difference between future and historical values for the indicated periods and RCP scenarios.

Supporting Information for Chapter 6

- 1 Supporting Text
- 12 Supporting Figures

Text S6.1 Implementation of the optimization algorithms

The implementation details of the algorithms used in the different optimization experiments are given below.

- Pareto-based sensitivity analysis: NSGA-II and MOFPA were each run twice 100 population points randomly extracted from the DELSA sample (i.e., biased initialization) and 25 generations. The four populations were combined to extract the best 100 solutions, and the most contributing algorithm to this selection was run once again, this time for 50 generations, to further tune the results.
- Multi-objective calibration: the most contributing algorithm was finally run for 200 points and 25 generations. A biased initialization strategy was implemented again choosing the final population from the Pareto-based sensitivity analysis and another random selection from the DELSA sample.
- Single-objective experiments: SCEUA was implemented to tackle five objective single-objectives experiments using different weights for streamflow and evaporation. In order to ensure convergence to the Pareto front, the termination criterion was set for a maximum number of iterations of 10000, instead of prespecifying a small size for the parametric space or a threshold for the objective improvement.

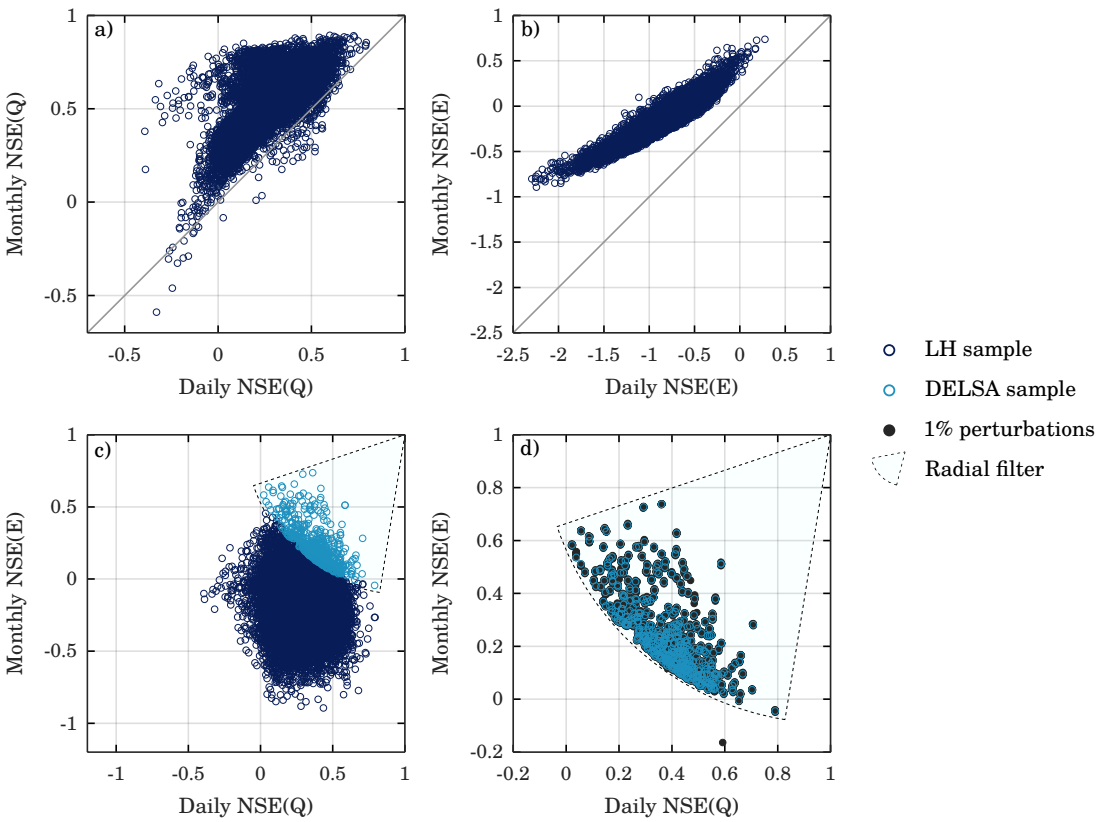


Fig. S6.1: Results of the Monte Carlo experiment for station R-5005 to study parameter space. (a) Daily and monthly estimates of NSE(Q) for the Latin Hybercube sample (LH sample). (b) Daily and monthly estimates of NSE(E) for the LH sample. (c) Joint representation of the selected performance metrics, NSE(Q_{d}) and NSE(E_{m}), and radial filter to extract the closest 500 points to the ideal solution (DELSA sample). (d) DELSA sample and the 1% perturbations where DELSA sensitivity analysis is to be performed.

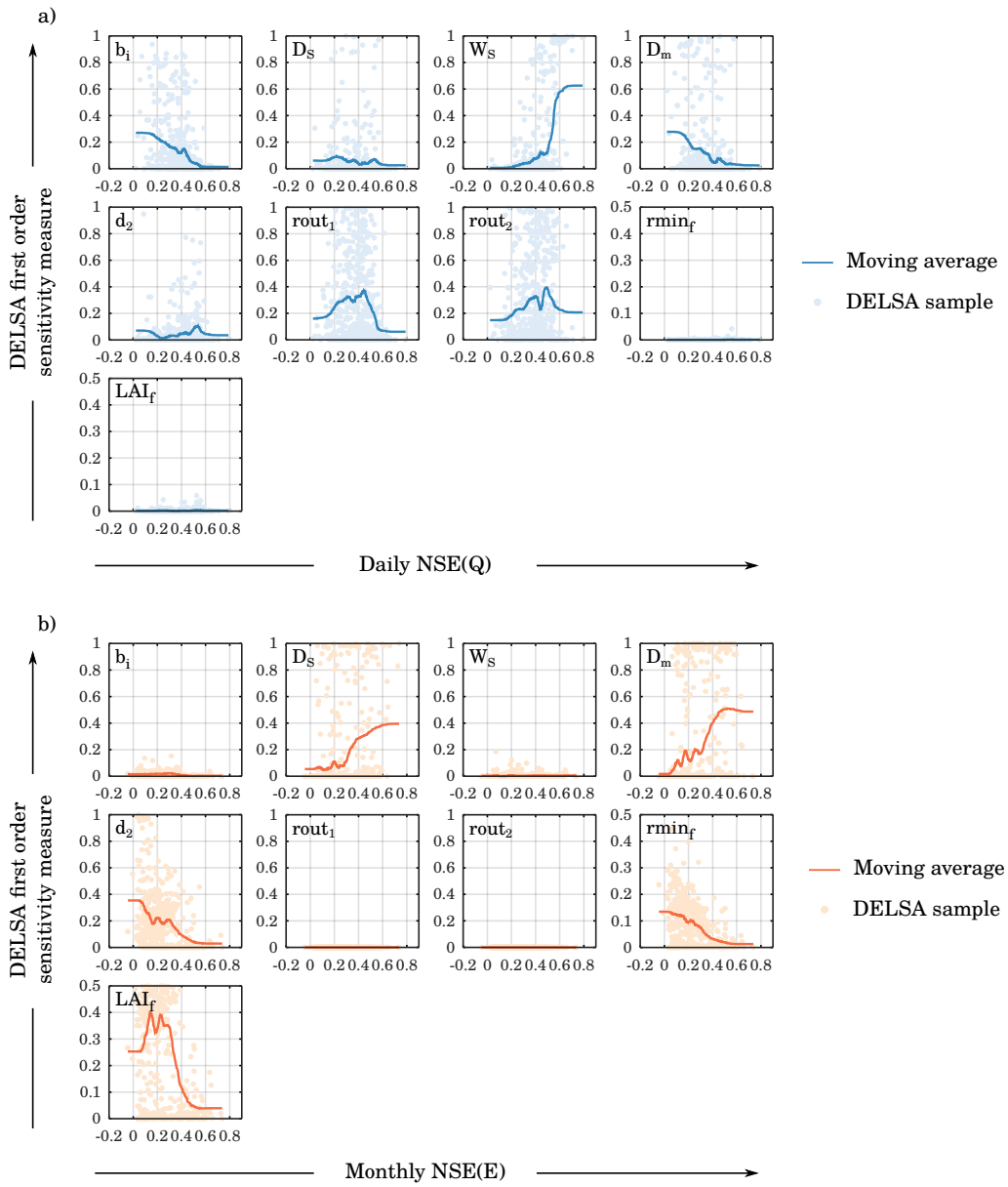


Fig. S6.2: DELSA first order sensitivities against model performance for station R-5005 computed for streamflow (a) and evaporation (b). Only the most influential parameters for the selected metrics are shown.

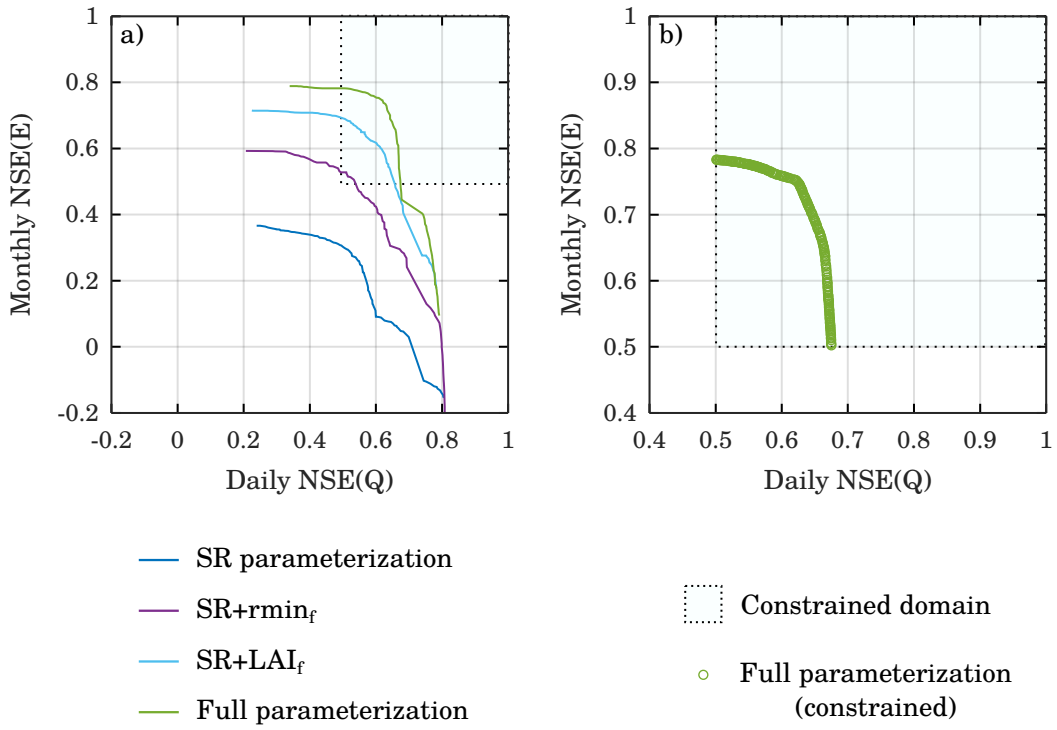


Fig. S6.3: Results of the Pareto-based sensitivity analysis and the multi-objective calibration for station R-5005. (a) Pareto fronts corresponding to the parameterizations evaluated during the Pareto-based sensitivity analysis. (b) Pareto front resulting from the constrained multi-objective calibration for the selected parameterization.

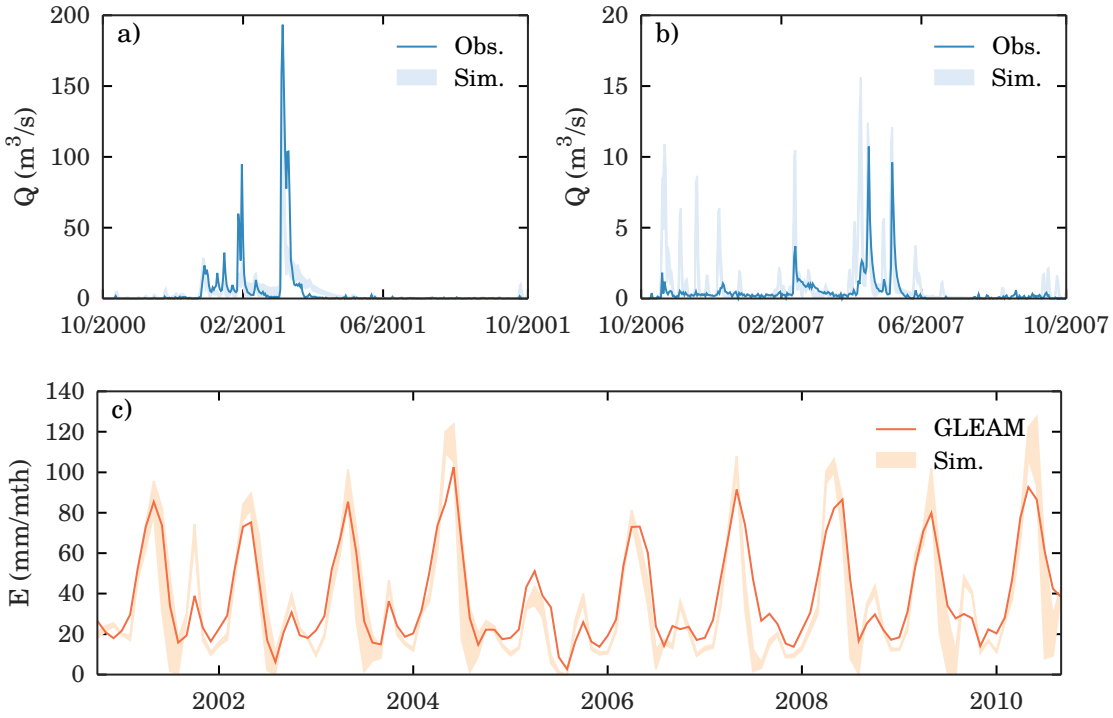


Fig. S6.4: Predictive intervals corresponding to the multi-objective calibration for station R-5005. (a and b) Daily streamflow simulations and observations for two different subperiods. (c) Monthly evaporation simulations and GLEAM data for the complete study period.

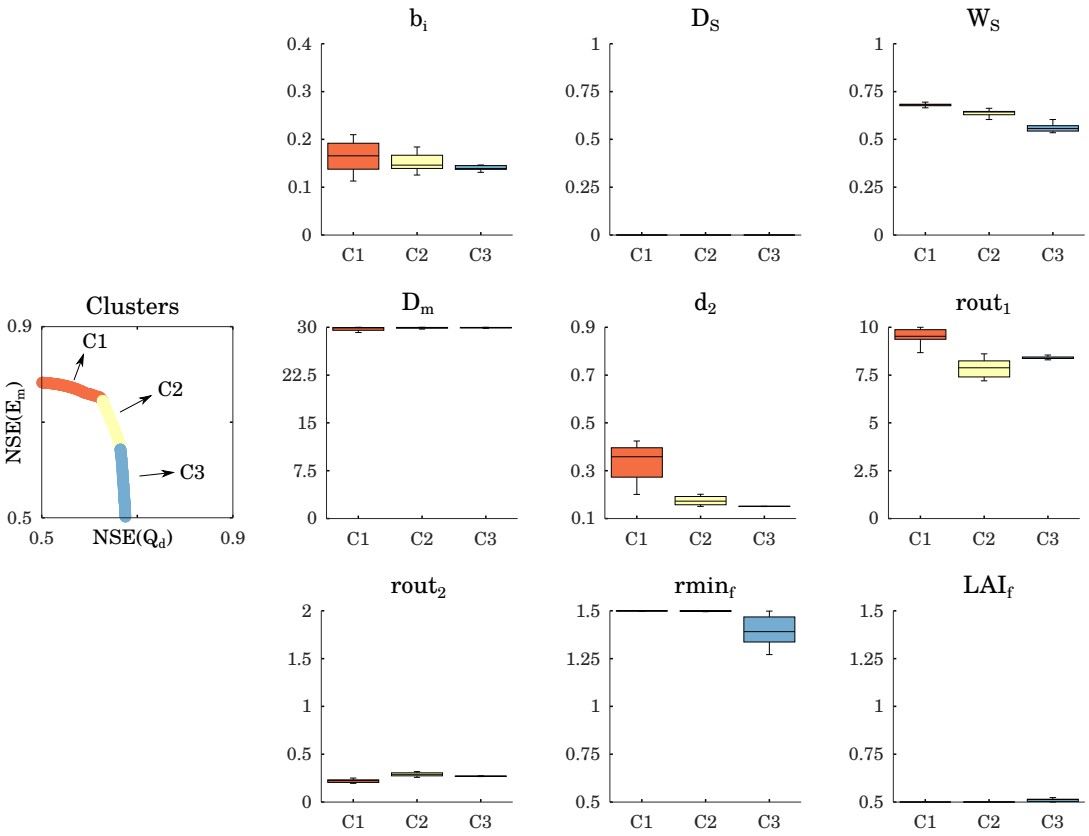


Fig. S6.5: Parameter distributions from the multi-objective calibration exercise for station R-5005. The Pareto front was divided intro three clusters representing well-differentiated regions of model performance.

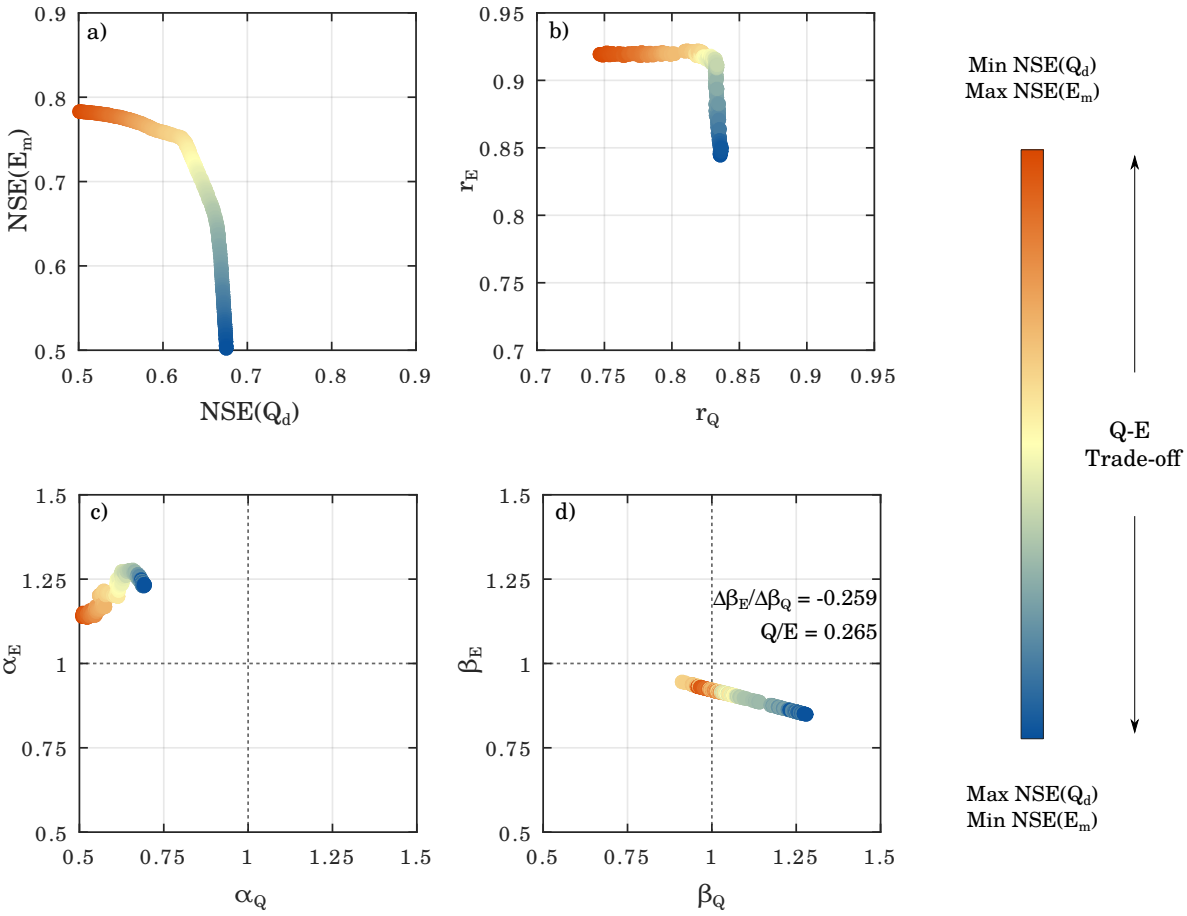


Fig. S6.6: NSE decomposition into r, α and β and trade-off evaluation for station R-5005.

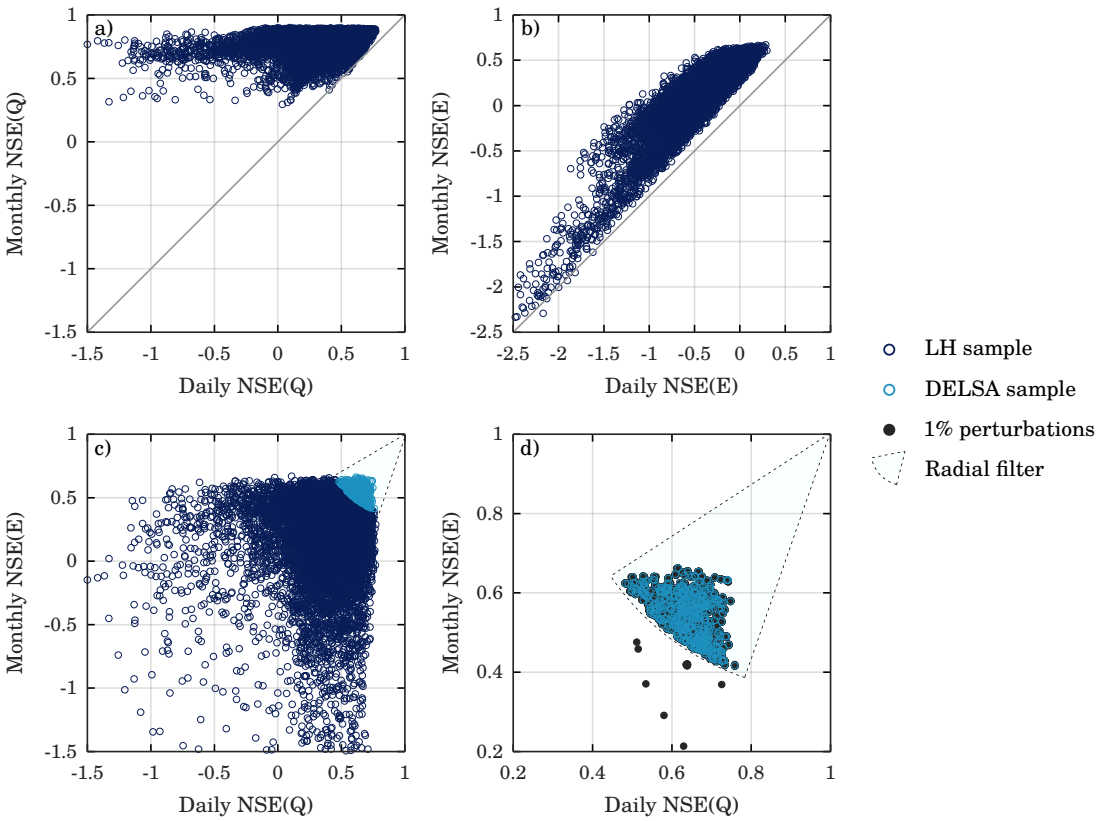


Fig. S6.7: Same as Fig. S6.1, but for station R-2011.

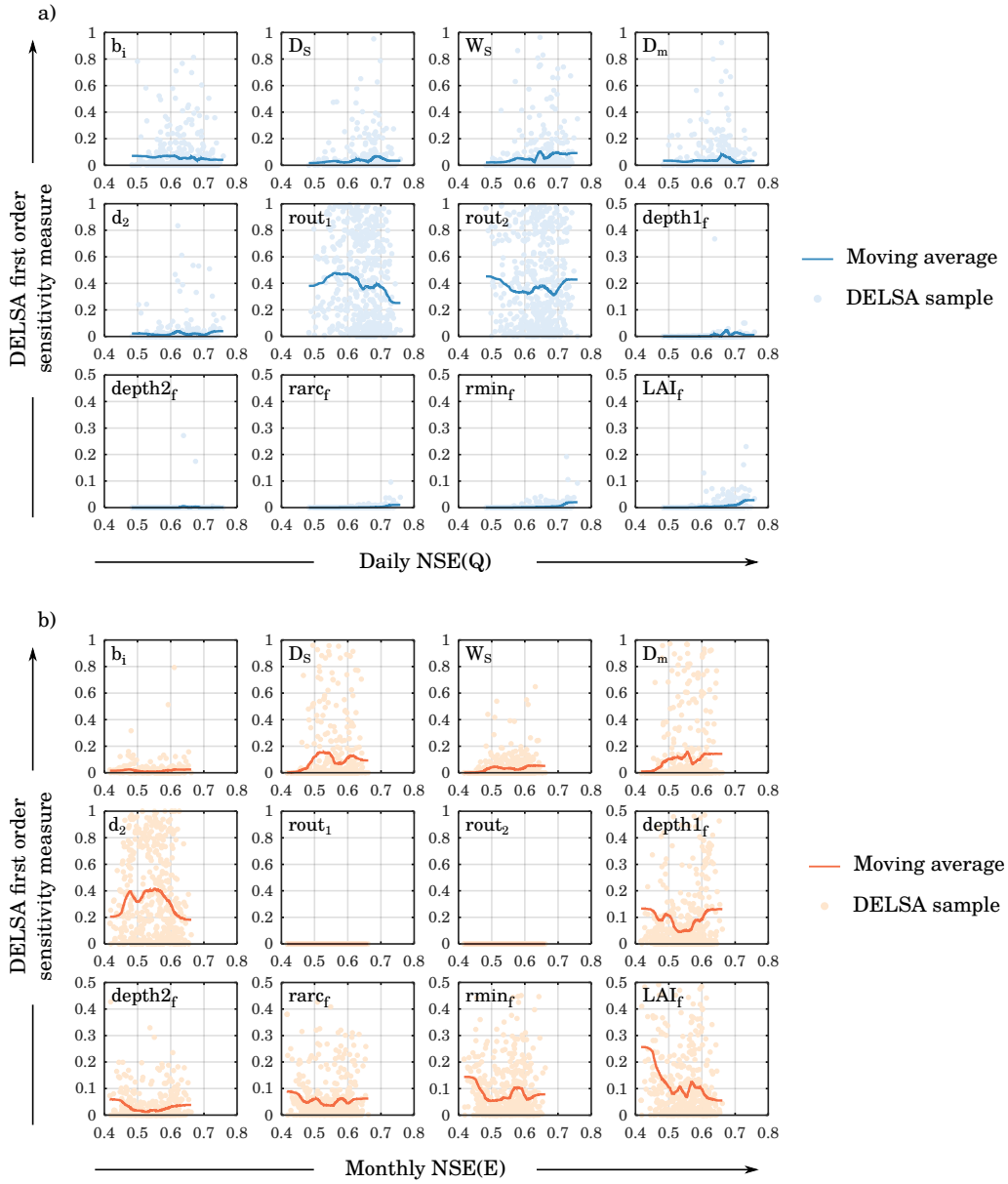


Fig. S6.8: Same as Fig. S6.2, but for station R-2011.

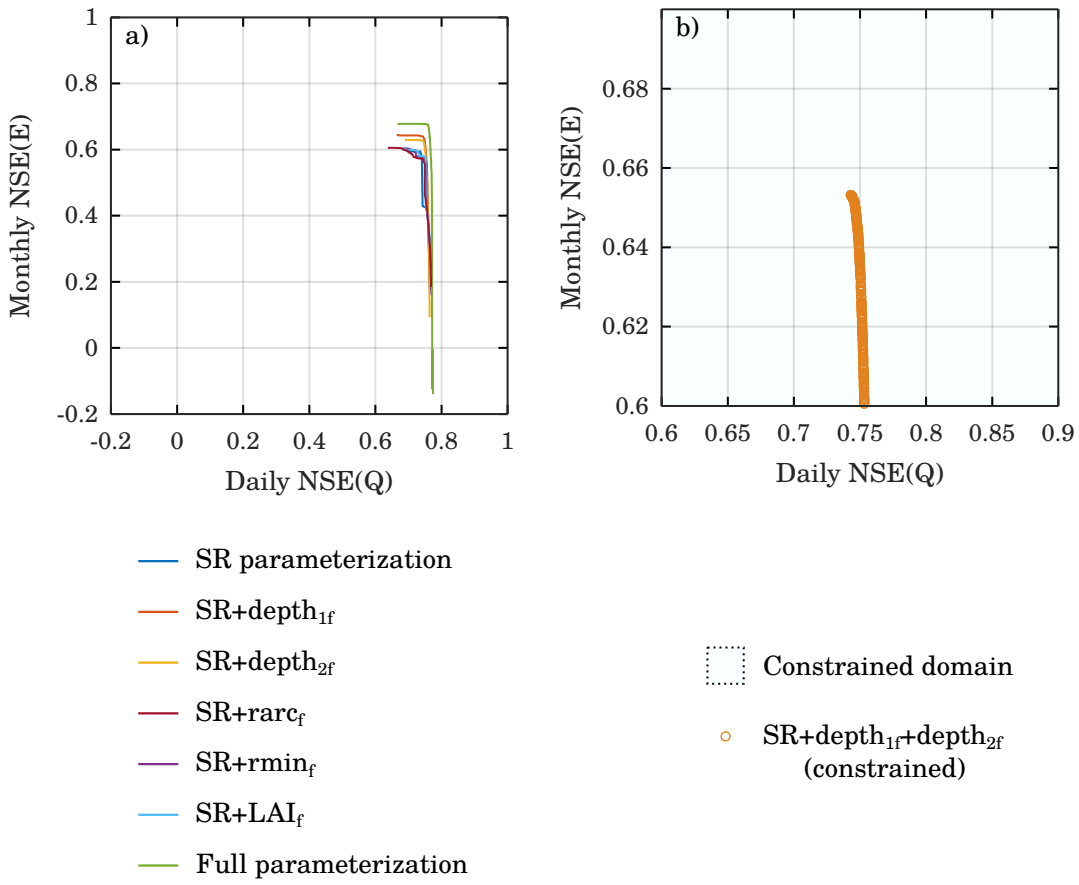


Fig. S6.9: Same as Fig. S6.3, but for station R-2011.

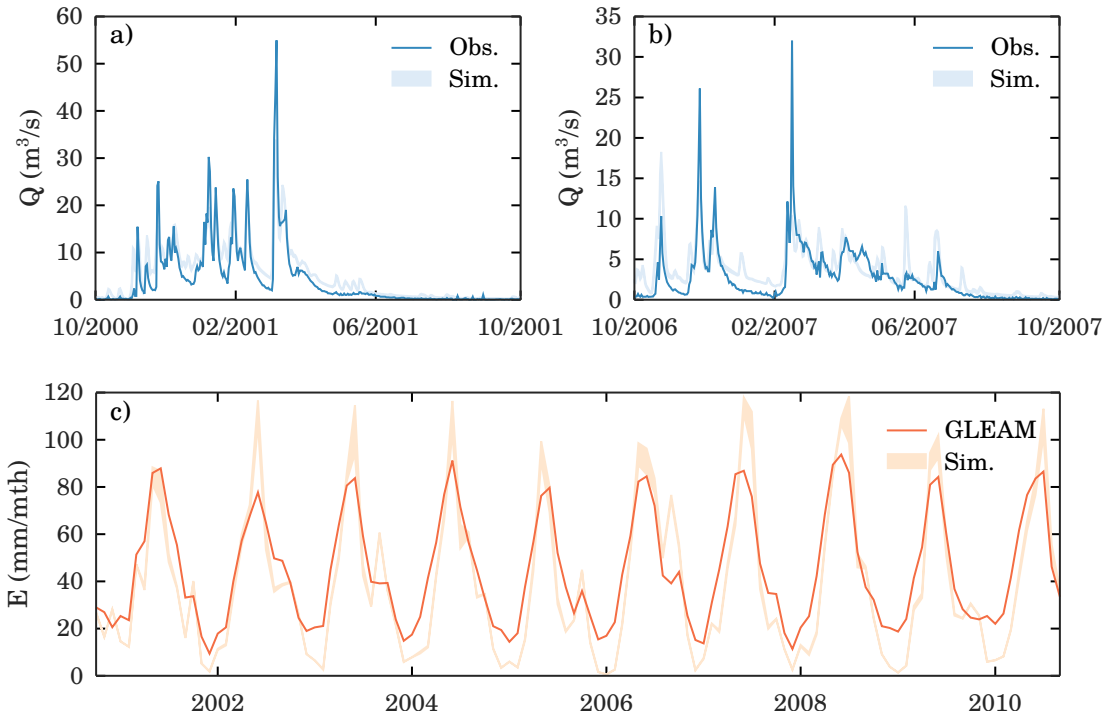


Fig. S6.10: Same as Fig. S6.4, but for station R-2011.

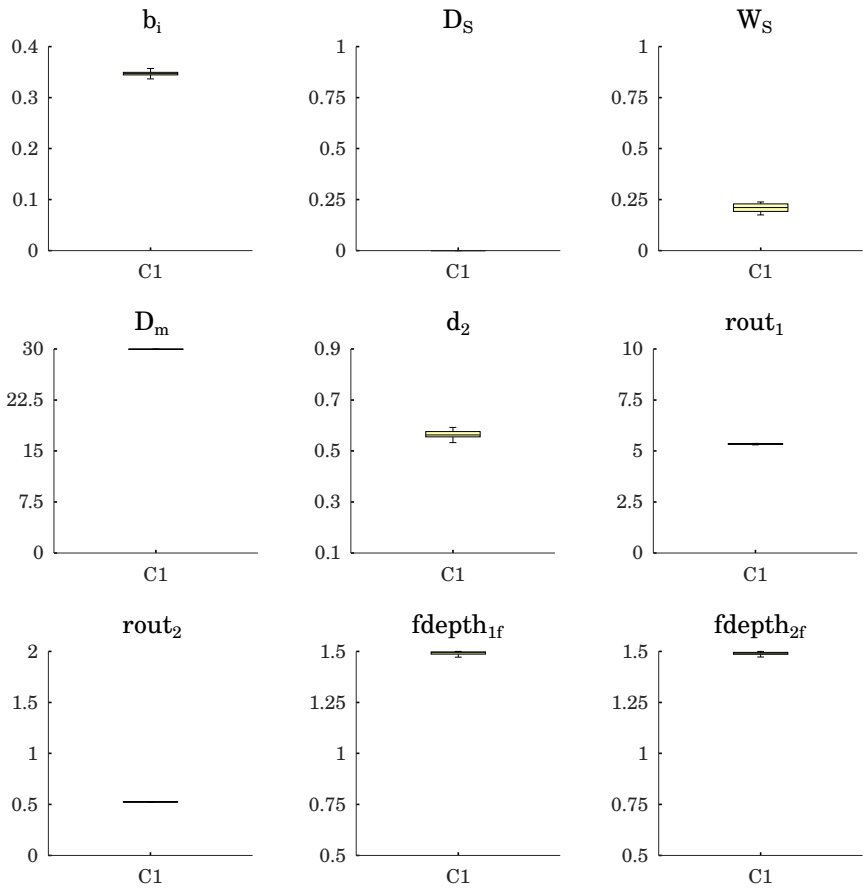


Fig. S6.11: Same as Fig. S6.5, but for station R-2011. In this case, only one cluster representing model performance was defined.

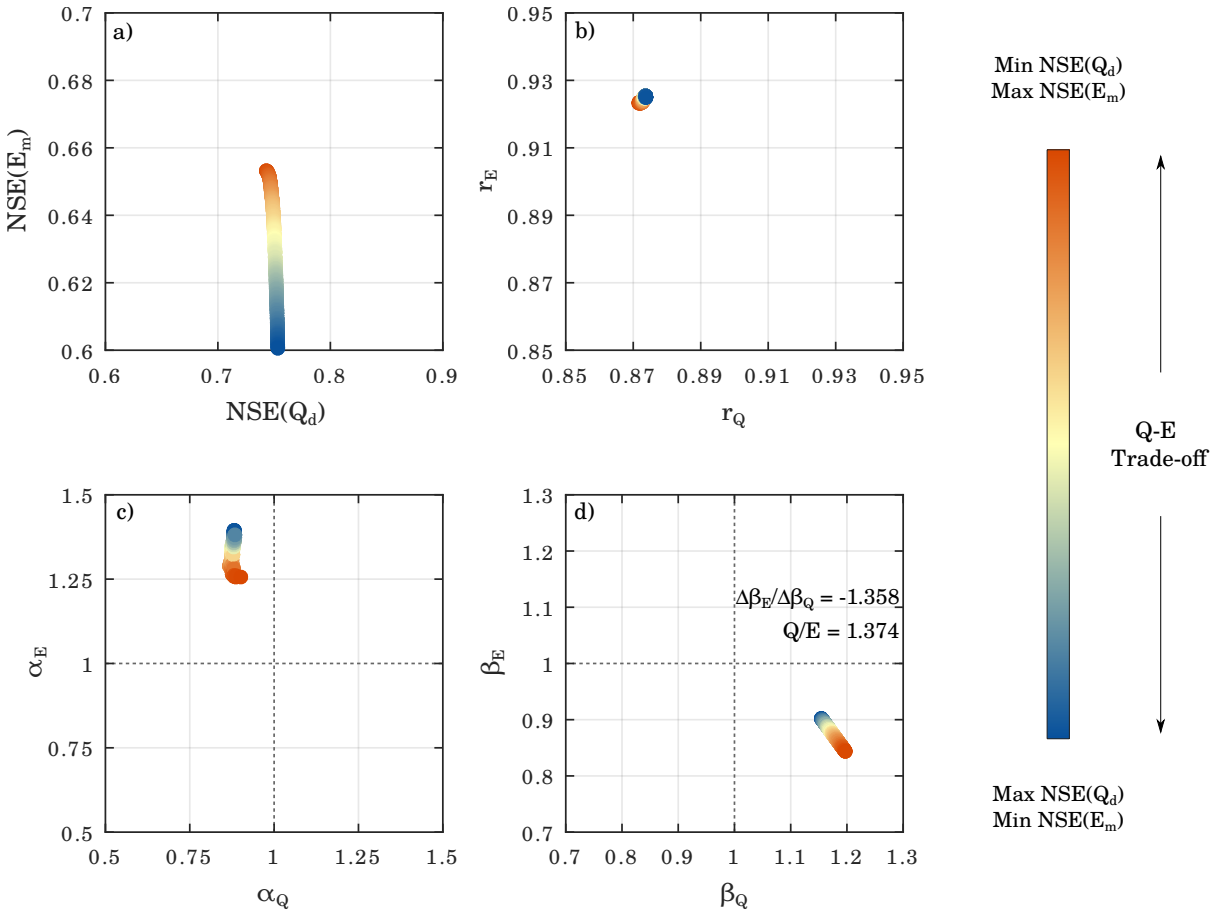


Fig. S6.12: Same as Fig. S6.6, but for station R-2011.

Supporting Information for Chapter 7

- 8 Supporting Tables
- 9 Supporting Figures

Table S7.1: NSE values during the complete baseline period for the Miño-Sil, Galicia Costa and Cantábrico Districts.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-1177	Ordunte	0.29	0.43	-0.09	0.22	0.31	0.65
R-1253	La Cohilla	0.50	0.69	-0.83	0.35	0.38	0.07
R-1406	Salime	0.84	0.91	-0.30	0.80	0.84	0.67
R-1637	Albarellos	0.55	0.78	-0.27	0.57	0.78	0.61
R-1718	Montearenas	0.74	0.79	-0.27	0.74	0.81	0.83
R-1741	San Sebastián	0.49	0.70	-0.90	0.42	0.51	0.54
R-1781	Edrada Conso	0.67	0.85	0.50	0.67	0.83	0.63
R-1790	Chandreja	0.75	0.86	0.05	0.71	0.82	0.64
R-1791	Prada	0.76	0.86	-0.34	0.72	0.79	0.73
R-1795	Leboreiro Mao	0.47	0.50	-0.56	0.46	0.50	0.47
R-1796	Vilasouto	0.60	0.70	-0.23	0.54	0.66	0.72
R-1807	Las Salas	0.50	0.78	-0.46	0.40	0.75	0.71
R-1808	Las Conchas	0.77	0.87	-0.75	0.74	0.87	0.75
GS-1163	Lemona	0.51	0.68	-0.27	0.40	0.43	0.71
GS-1196	Coterillo	0.73	0.80	0.34	0.71	0.80	0.67
GS-1285	Rales	0.80	0.85	0.72	0.74	0.83	0.77
GS-1302	Ozanes	0.75	0.77	0.24	0.75	0.79	0.79
GS-1378	Grado	0.67	0.76	-0.36	0.50	0.39	0.73
GS-1395	Trevias	0.76	0.58	0.27	0.79	0.61	0.82
GS-1427	San Tirso de Abres	0.76	0.80	-0.63	0.59	0.54	0.70
GS-1431	Masma	0.51	0.72	-1.93	0.26	0.25	0.65
GS-1433	San Acisclo	0.51	0.72	0.38	0.50	0.69	0.74
GS-1438	Chavín	0.24	0.61	0.52	0.24	0.61	0.65
GS-1464	Muñiferal	0.64	0.82	0.46	0.63	0.79	0.76
GS-1485	Anllóns	0.76	0.83	0.36	0.64	0.77	0.86
GS-1552	Puente Cira	0.65	0.84	-0.61	0.63	0.83	0.76
GS-1724	Quilós	0.79	0.78	-0.69	0.74	0.73	0.60
GS-1734	Puente de Domingo Flórez	0.76	0.86	-0.59	0.64	0.67	0.68

Table S7.2: NSE values during the complete baseline period for the Duero River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-2001	Cuerda del Pozo	0.62	0.70	-0.75	0.44	0.33	0.67
R-2011	Arlanzón	0.72	0.83	-0.06	0.70	0.81	0.66
R-2013	La Requejada	0.62	0.76	-0.31	0.54	0.61	0.12
R-2014	Camporredondo	0.69	0.83	-0.12	0.63	0.78	0.16
R-2026	Barrios de Luna	0.66	0.77	-0.15	0.58	0.67	0.53
R-2027	Villameca	0.77	0.84	-0.10	0.60	0.63	0.71
R-2030	Porma Juan Benet	0.59	0.67	-0.95	0.48	0.44	0.33
R-2032	Riaño	0.75	0.88	-0.06	0.69	0.84	0.35
R-2036	Linares del Arroyo	0.48	0.64	-0.12	0.63	0.77	0.79
R-2037	Burgomillodo	0.56	0.77	0.51	0.54	0.76	0.79
R-2038	Santa Teresa	0.68	0.84	-0.45	0.60	0.73	0.47
GS-2005	Osma	0.70	0.69	-0.39	0.65	0.57	0.63
GS-2016	Pajares de Pedraza	0.50	0.41	0.79	0.47	0.37	0.80
GS-2031	Peral de Arlanza	0.75	0.84	0.24	0.69	0.79	0.76
GS-2041	Villalcázar de Sirga	0.46	0.50	0.31	0.49	0.58	0.75
GS-2044	Valladolid-Esgueva	0.63	0.63	-0.07	0.46	0.52	0.44
GS-2047	Mediana de Voltoya	0.41	0.51	0.17	0.33	0.36	0.46
GS-2052	Guijas Albas	0.60	0.72	0.64	0.59	0.70	0.69
GS-2057	Villovela de Pirón	0.33	0.46	0.16	0.16	0.23	0.78
GS-2068	Caldas de Nocedo	0.59	0.68	-0.22	0.52	0.59	0.34
GS-2124	Medina de Rioseco	0.29	0.51	-0.54	0.27	0.34	0.53
GS-2126	Villárdiga	0.39	0.08	-0.72	0.57	0.44	0.61
GS-2818	Rabal	0.70	0.78	-0.55	0.58	0.59	0.74

Table S7.3: NSE values during the complete baseline period for the Tajo River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-3006	Entrepeñas	0.62	0.67	0.24	0.74	0.82	0.76
R-3050	El Vado	0.73	0.86	0.49	0.72	0.84	0.68
R-3065	Pálmaces	0.61	0.86	0.11	0.64	0.86	0.55
R-3068	Beleña	0.65	0.84	0.24	0.61	0.77	0.66
R-3111	El Burguillo	0.78	0.91	0.40	0.73	0.85	0.78
R-3145	Jerte Plasencia	0.74	0.87	-0.62	0.73	0.85	0.75
R-3148	Borbollón	0.75	0.90	-0.93	0.68	0.77	0.71
R-3157	El Vellón	0.58	0.82	0.22	0.55	0.80	0.61
R-3160	Rivera de Gata	0.75	0.88	-1.16	0.75	0.88	0.72
R-3189	Navacerrada	0.64	0.79	0.51	0.61	0.74	0.68
R-3191	Navalmedio	0.68	0.81	0.71	0.68	0.80	0.71
R-3196	Pinilla	0.66	0.81	0.52	0.64	0.80	0.65
R-3199	Navalcán	0.71	0.84	-0.36	0.75	0.83	0.53
R-3287	Alcorlo	0.58	0.84	0.69	0.60	0.83	0.72
GS-3045	Priego Escabas	0.64	0.75	0.25	0.59	0.74	0.58
GS-3161	Arenas de San Pedro	0.79	0.93	-0.96	0.85	0.94	0.57
GS-3163	Piedras Albas	0.52	0.59	-0.24	0.48	0.55	0.68
GS-3172	Huete	0.26	0.01	0.56	0.15	0.17	0.67
GS-3173	La Peraleja	0.25	0.56	0.53	0.23	0.42	0.62
GS-3182	Garcibuey	0.68	0.89	-0.58	0.67	0.78	0.69
GS-3185	Santibáñez el Bajo	0.65	0.72	-1.56	0.61	0.70	0.66
GS-3186	Priego Trabaque	0.51	0.89	0.38	0.30	0.83	0.75
GS-3212	Malpica	0.55	0.70	0.37	0.57	0.74	0.73
GS-3213	Alcaudete de la Jara	0.33	0.62	-0.31	0.30	0.59	0.50
GS-3217	Horcajo de Montemayor	0.47	0.59	0.07	0.57	0.70	0.72
GS-3218	Miranda del Castañar	0.51	0.68	0.44	0.49	0.62	0.73
GS-3220	Santa Marta de Magasca	0.43	0.57	0.31	0.51	0.71	0.50
GS-3221	Bohonal de Ibor	0.68	0.84	-0.44	0.64	0.77	0.69
GS-3222	Peraleda de San Román	0.48	0.61	-0.12	0.46	0.67	0.66
GS-3224	Gargüera	0.69	0.88	0.13	0.69	0.87	0.51
GS-3226	Candeleda	0.38	0.54	0.05	0.38	0.53	0.69
GS-3229	Losar de la Vera	0.38	0.59	-0.99	0.38	0.60	0.68
GS-3233	Valdecaba Baja	0.27	0.39	-0.48	0.35	0.56	0.47
GS-3234	Jajaíz de la Vera	0.60	0.83	-0.39	0.66	0.84	0.70
GS-3235	Nuñomoral	0.62	0.81	-0.60	0.64	0.79	0.69
GS-3236	El Ladrillar	0.40	0.52	-0.52	0.29	0.34	0.75
GS-3244	Campillo de Deleitosa	0.63	0.85	-0.69	0.55	0.73	0.65
GS-3246	Ayuela	0.53	0.60	-0.06	0.51	0.58	0.64

Table S7.3 (cont.): NSE values during the complete baseline period for the Tajo River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
GS-3250	Trujillo Magasca	0.41	0.63	-0.11	0.25	0.57	0.49
GS-3251	La Pueblanueva	0.42	0.59	0.32	0.39	0.56	0.60
GS-3254	Aragosa	0.32	0.38	0.12	0.29	0.37	0.33
GS-3255	Torre del Burgo	0.50	0.53	0.17	0.38	0.37	0.46
GS-3260	Villanueva de la Vera	0.37	0.60	-1.48	0.37	0.64	0.66
GS-3266	Camarenilla	0.22	0.46	-1.13	0.19	0.39	0.42
GS-3273	Velilla de San Antonio	0.36	0.39	0.20	0.35	0.37	0.52
GS-3276	Torrejón	0.50	0.66	-0.05	0.52	0.75	0.61
GS-3278	Valencia de Alcántara	0.27	0.45	-0.47	0.31	0.58	0.64
GS-3279	Trujillo Tozo	0.48	0.51	-0.17	0.43	0.58	0.54
GS-3284	Cañaveral	0.34	0.44	-0.55	0.35	0.58	0.67

Table S7.4: NSE values during the complete baseline period for the Guadiana River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-4005	Gasset	0.72	0.87	0.13	0.66	0.81	0.58
R-4007	Torre de Abraham	0.71	0.82	-0.09	0.63	0.70	0.56

Table S7.5: NSE values during the complete baseline period for the Guadalquivir River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-5001	Tranco de Beas	0.74	0.83	-0.32	0.60	0.57	0.69
R-5005	Rumblar	0.55	0.80	-0.29	0.36	0.40	0.37
R-5006	Montoro III	0.60	0.64	-0.24	0.34	0.53	0.43
R-5011	El Pintado	0.69	0.78	-0.85	0.75	0.91	0.64
R-5012	Bembézar	0.56	0.71	-0.75	0.58	0.86	0.65
R-5014	Cala	0.66	0.86	-0.40	0.61	0.80	0.63
R-5016	Aracena	0.73	0.88	0.35	0.67	0.78	0.68
R-5017	Retortillo	0.75	0.92	0.26	0.73	0.88	0.61
R-5018	Guadalmena	0.71	0.74	0.26	0.72	0.78	0.66
R-5020	Cubillas	0.52	0.74	-0.31	0.39	0.58	0.51
R-5021	Los Bermejales	0.80	0.85	0.02	0.80	0.87	0.76
R-5022	Torre del Águila	0.79	0.91	0.18	0.56	0.83	0.56
R-5029	La Bolera	0.46	0.64	-0.53	0.37	0.40	0.47
R-5037	Sierra Boyera	0.50	0.78	-0.14	0.46	0.70	0.46
R-5038	Quéntar	0.56	0.67	-0.05	0.29	0.39	0.49
R-5039	Quiebrajano	-0.10	-0.47	-1.34	-0.47	-1.11	-0.35
R-5044	Dañador	0.38	0.46	-1.12	0.29	0.26	0.32
R-5045	Aguascebas	0.61	0.79	-0.46	0.58	0.71	0.67
R-5046	Guadanuño Cerro Muriano	0.64	0.83	-0.72	0.55	0.73	0.64
R-5047	Martín Gonzalo	0.60	0.83	-0.18	0.47	0.63	0.60
R-5048	Canales	0.35	0.38	-0.08	0.16	0.15	0.46
R-5049	Yeguas	0.71	0.82	-0.72	0.64	0.80	0.69
R-5050	Colomera	0.71	0.79	-0.94	0.64	0.71	0.50
R-5052	La Fernandina	0.73	0.86	-0.30	0.53	0.73	0.63
R-5055	José Torán	0.69	0.82	-0.98	0.67	0.87	0.65
R-5056	Huesna	0.75	0.84	-1.03	0.74	0.88	0.62
R-5062	Guadalmellato	0.63	0.79	-0.89	0.60	0.73	0.68

Table S7.6: NSE values during the complete baseline period for the Segura River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-7001	Anchuricas	0.57	0.75	-0.39	0.53	0.70	0.64
R-7004	La Novia	0.42	0.67	-0.32	0.23	0.05	0.35
R-7005	Alfonso XIII	0.37	0.36	-0.24	0.26	0.33	0.15
R-7006	La Cierva	0.42	0.30	-0.76	0.21	0.05	0.31
R-7010	Santomera	0.85	0.84	-0.04	0.83	0.75	0.62
R-7011	Argos	0.34	0.48	-0.36	0.20	0.37	0.57

Table S7.7: NSE values during the complete baseline period for the Júcar River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-8002	Alcora	0.37	0.28	-0.43	0.42	0.34	0.67
R-8006	Arquillo de San Blas	0.09	0.57	0.71	0.08	0.51	0.81
R-8007	Beniarrés	0.73	0.84	-0.28	0.49	0.19	0.54
R-8015	Guadalest	0.14	0.04	-0.84	-0.07	-0.38	0.40
R-8018	Onda	0.16	-0.36	0.26	0.21	-0.29	0.66
R-8019	Regajo	0.38	0.22	-0.51	0.05	-0.58	0.33
R-8023	La Toba	0.65	0.72	0.15	0.58	0.60	0.63
R-8025	Ulldecona	0.07	0.40	0.34	0.06	0.33	0.76
GS-8028	Villalba Alta	0.35	0.38	0.64	0.35	0.41	0.73
GS-8087	Belmontejo	0.10	0.10	-0.34	-1.50	-0.47	0.51
GS-8104	Los Santos	0.30	0.24	0.48	0.25	0.20	0.52
GS-8137	Tiriez	0.47	0.48	-0.11	0.49	0.55	0.64
GS-8138	Balazote	0.81	0.82	-0.29	0.65	0.70	0.62

Table S7.8: NSE values during the complete baseline period for the Ebro River Basin District.

Code	Name	Q-only calibrated			Q-E calibrated		
		Q_d	Q_m	E_m	Q_d	Q_m	E_m
R-9801	Ebro	0.61	0.81	-0.36	0.49	0.52	0.42
R-9808	Maidevera	0.70	0.79	-0.05	0.28	0.10	0.30
R-9809	Mansilla	0.68	0.65	-0.12	0.53	0.48	0.70
R-9812	La Tranquera	0.04	0.40	0.53	0.04	0.33	0.72
R-9814	Las Torcas	0.23	0.33	0.29	0.21	0.22	0.59
R-9815	Moneva	0.23	0.66	-0.13	0.38	0.50	0.45
R-9817	Cueva Foradada	0.05	0.17	0.73	0.22	0.55	0.74
R-9821	Pena	0.51	0.75	0.45	0.51	0.65	0.65
R-9825	Eugui	0.70	0.85	0.05	0.63	0.84	0.20
R-9829	Yesa	0.55	0.54	0.50	0.54	0.51	0.71
R-9830	Alloz	0.67	0.73	-0.09	0.52	0.47	0.64
R-9841	Vadiello	0.46	0.00	-0.59	0.28	-0.93	0.52
R-9868	Ciurana	-0.44	-1.98	-1.66	-0.59	-5.30	0.57
GS-9031	Zorita	0.87	0.85	0.46	0.80	0.70	0.72
GS-9040	Boltaña	0.40	0.32	0.40	0.38	0.33	0.77
GS-9047	Capella	0.59	0.70	0.21	0.52	0.64	0.58
GS-9052	Beceite	0.03	0.39	-0.25	-0.01	-0.29	0.55
GS-9055	Morata de Jiloca	0.26	0.30	0.35	0.30	0.31	0.74
GS-9057	Embid de Ariza	0.12	0.16	-0.06	0.41	0.53	0.53
GS-9058	Jubera	0.30	0.27	0.24	0.40	0.42	0.32
GS-9064	Aspurz	0.58	0.61	0.04	0.50	0.48	0.65
GS-9071	Estella	0.65	0.47	-0.20	0.64	0.56	0.67
GS-9073	Sangüesa	0.40	0.53	0.14	0.38	0.51	0.74
GS-9078	Garínoain	0.36	0.64	0.42	0.51	0.59	0.77
GS-9079	Urroz	0.63	0.77	0.40	0.68	0.85	0.57
GS-9086	Barásoain	0.24	0.39	0.12	0.26	0.24	0.60
GS-9093	Oña	0.52	0.61	0.52	0.53	0.60	0.66
GS-9095	Barbastro	0.43	0.30	-0.07	0.26	0.33	0.70
GS-9155	Biota	0.41	0.59	-0.12	0.21	0.28	0.28
GS-9165	Miranda de Ebro	0.68	0.81	-0.73	0.45	0.32	0.62
GS-9177	Batea	0.49	0.59	0.29	0.34	0.52	0.78
GS-9184	Ateca	0.30	0.34	-0.25	0.21	0.21	0.48
GS-9185	Cintruénigo	0.42	0.46	0.60	0.33	0.31	0.71
GS-9187	Erla	0.49	0.65	0.42	0.52	0.71	0.66
GS-9188	Berguenda	0.59	0.62	-0.34	0.64	0.76	0.72
GS-9189	Orón	0.14	0.13	0.39	0.07	0.06	0.53
GS-9197	Leza de Río Leza	0.40	0.40	0.60	0.43	0.44	0.71
GS-9221	Larrinoa	0.51	0.53	0.18	0.49	0.51	0.51
GS-9231	Candasnos	0.14	-0.02	-0.12	-0.07	-0.92	0.40
GS-9253	Arnedillo	0.54	0.59	0.69	0.55	0.60	0.76
GS-9261	Trasobares	0.48	0.43	0.06	0.36	0.30	0.51

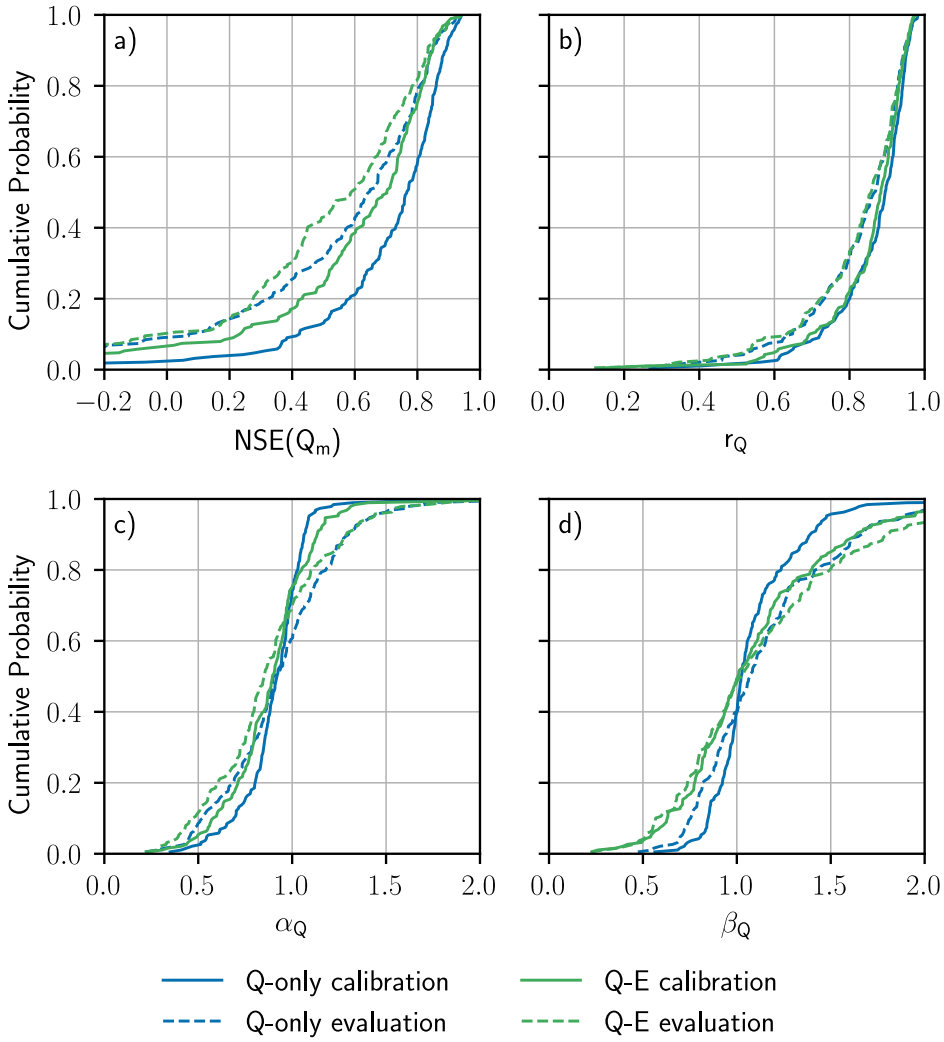


Fig. S7.1: CDFs of (a) NSE(Q_m) and its decomposition into (b) r_Q, (c) α_Q and (d) β_Q for both calibration experiments during the calibration and evaluation periods.

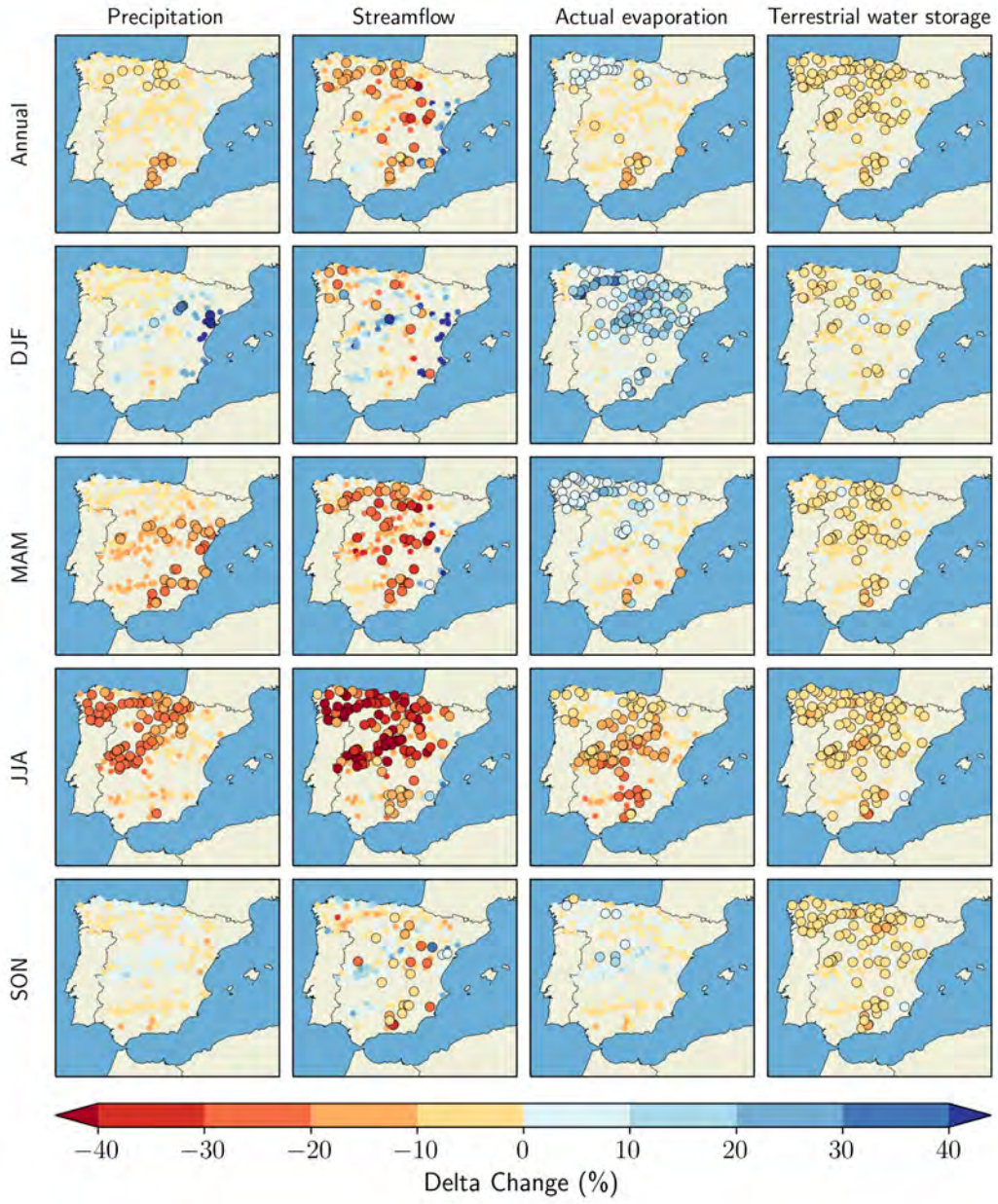


Fig. S7.2: Delta changes of annual and seasonal P, Q, E and TWS for the period 2021-2050 under the RCP4.5 scenario using WRFCCSM. Significant changes at the 90% confidence level have been marked with solid borders.

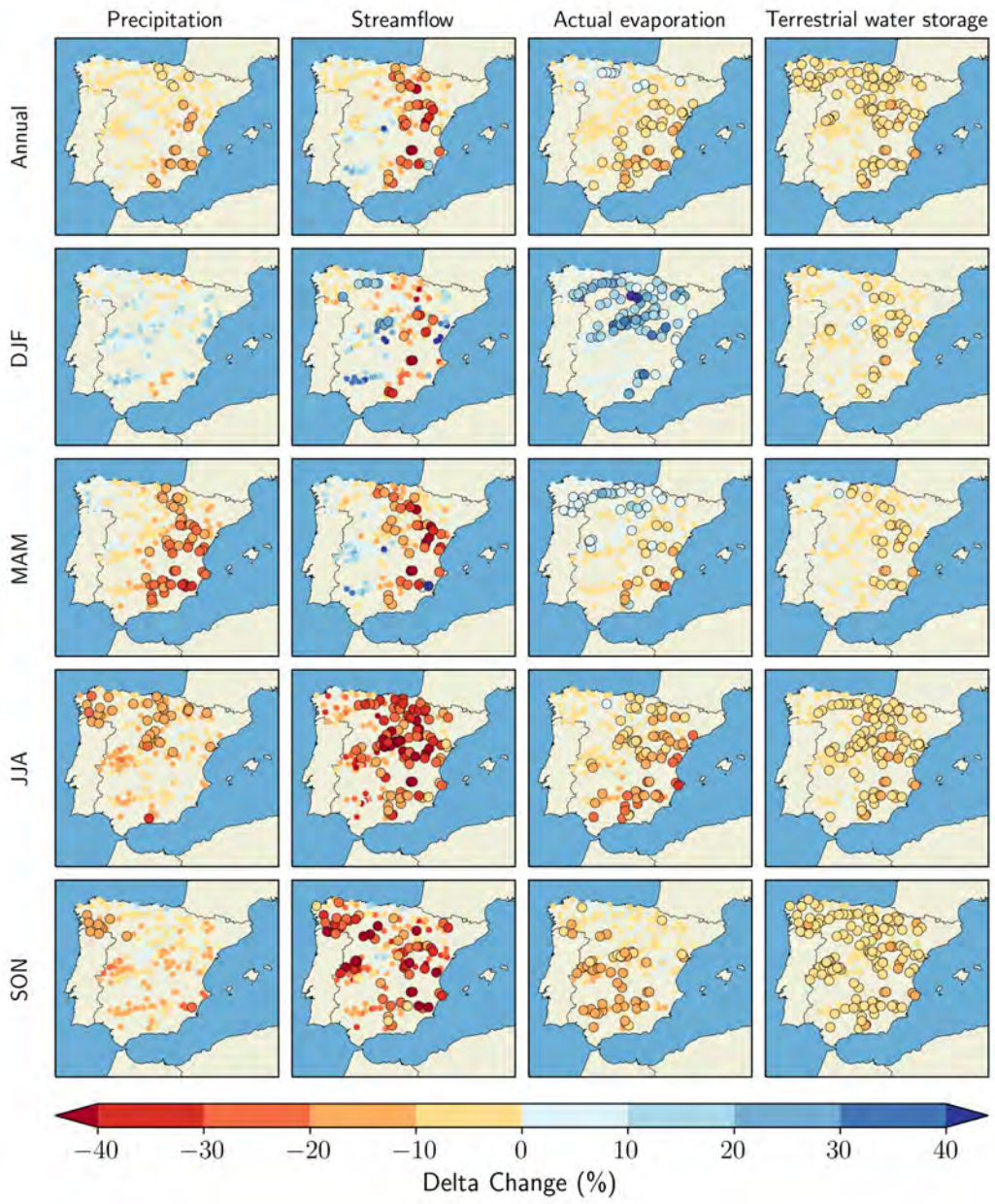


Fig. S7.3: Delta changes of annual and seasonal P, Q, E and TWS for the period 2021-2050 under the RCP4.5 scenario using WRFMPI. Significant changes at the 90% confidence level have been marked with solid borders.

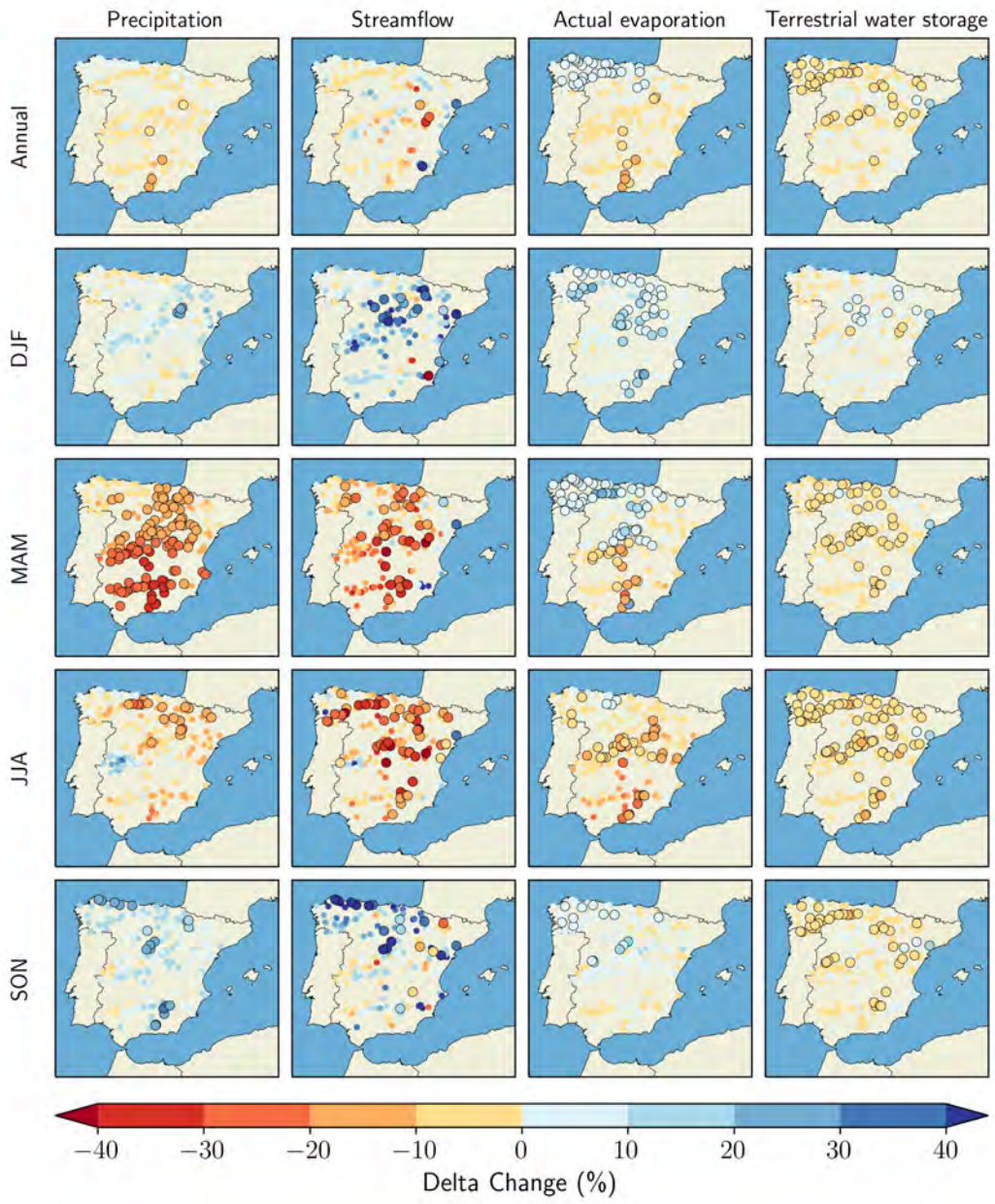


Fig. S7.4: Delta changes of annual and seasonal P, Q, E and TWS for the period 2021-2050 under the RCP8.5 scenario using WRFCCSM. Significant changes at the 90% confidence level have been marked with solid borders.

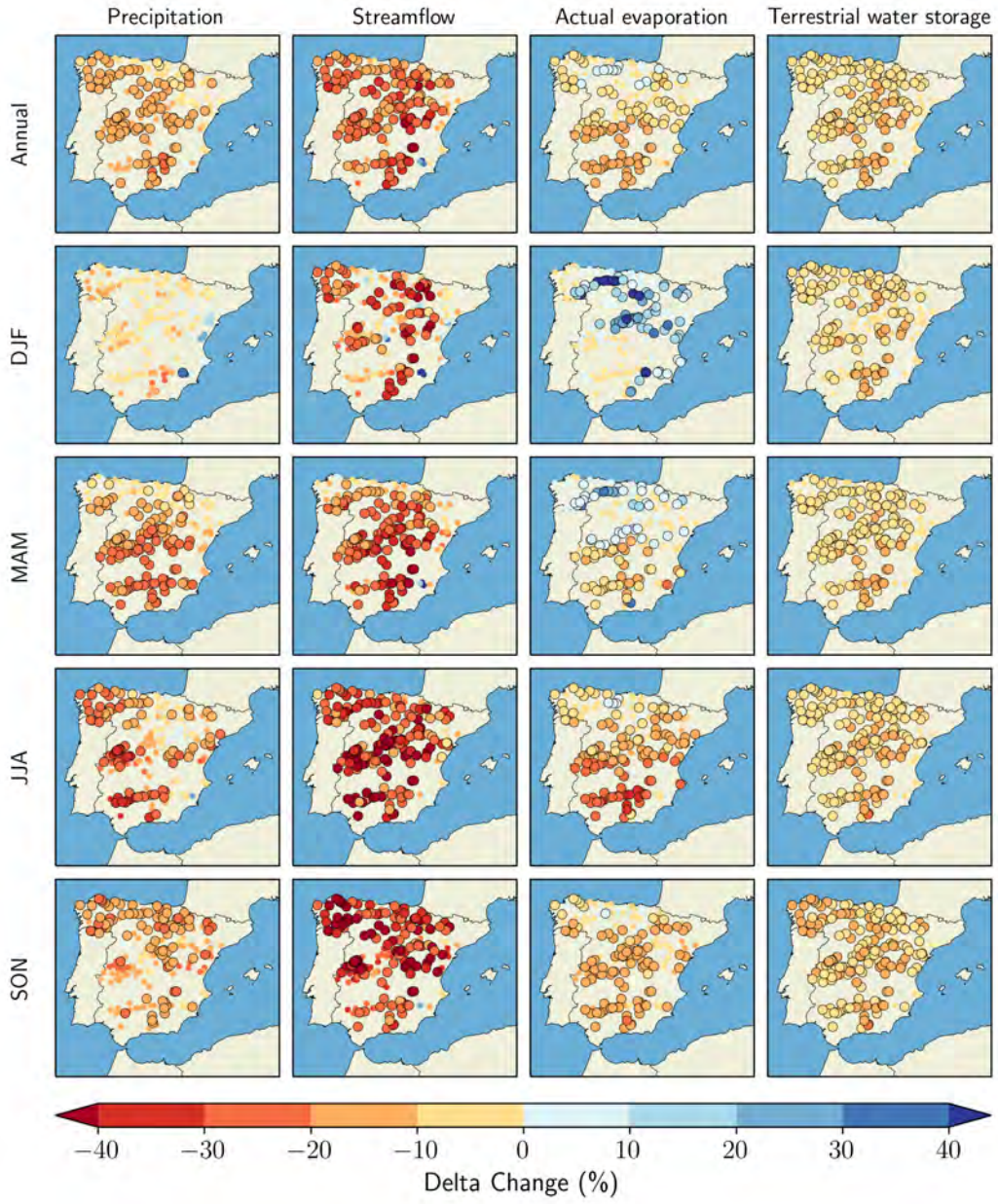


Fig. S7.5: Delta changes of annual and seasonal P, Q, E and TWS for the period 2021-2050 under the RCP8.5 scenario using WRFMPI. Significant changes at the 90% confidence level have been marked with solid borders.

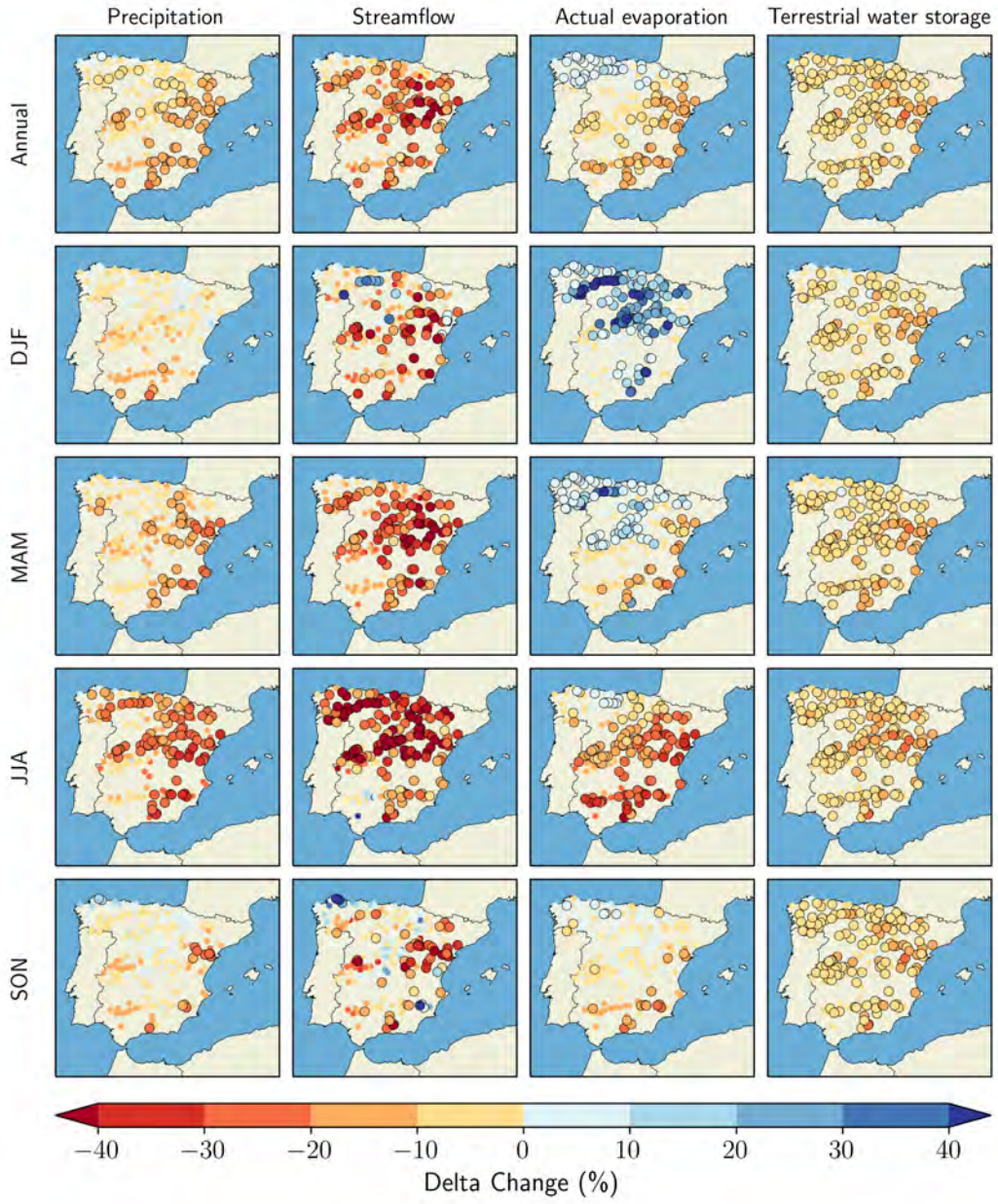


Fig. S7.6: Delta changes of annual and seasonal P, Q, E and TWS for the period 2071-2100 under the RCP4.5 scenario using WRFCCSM. Significant changes at the 90% confidence level have been marked with solid borders.

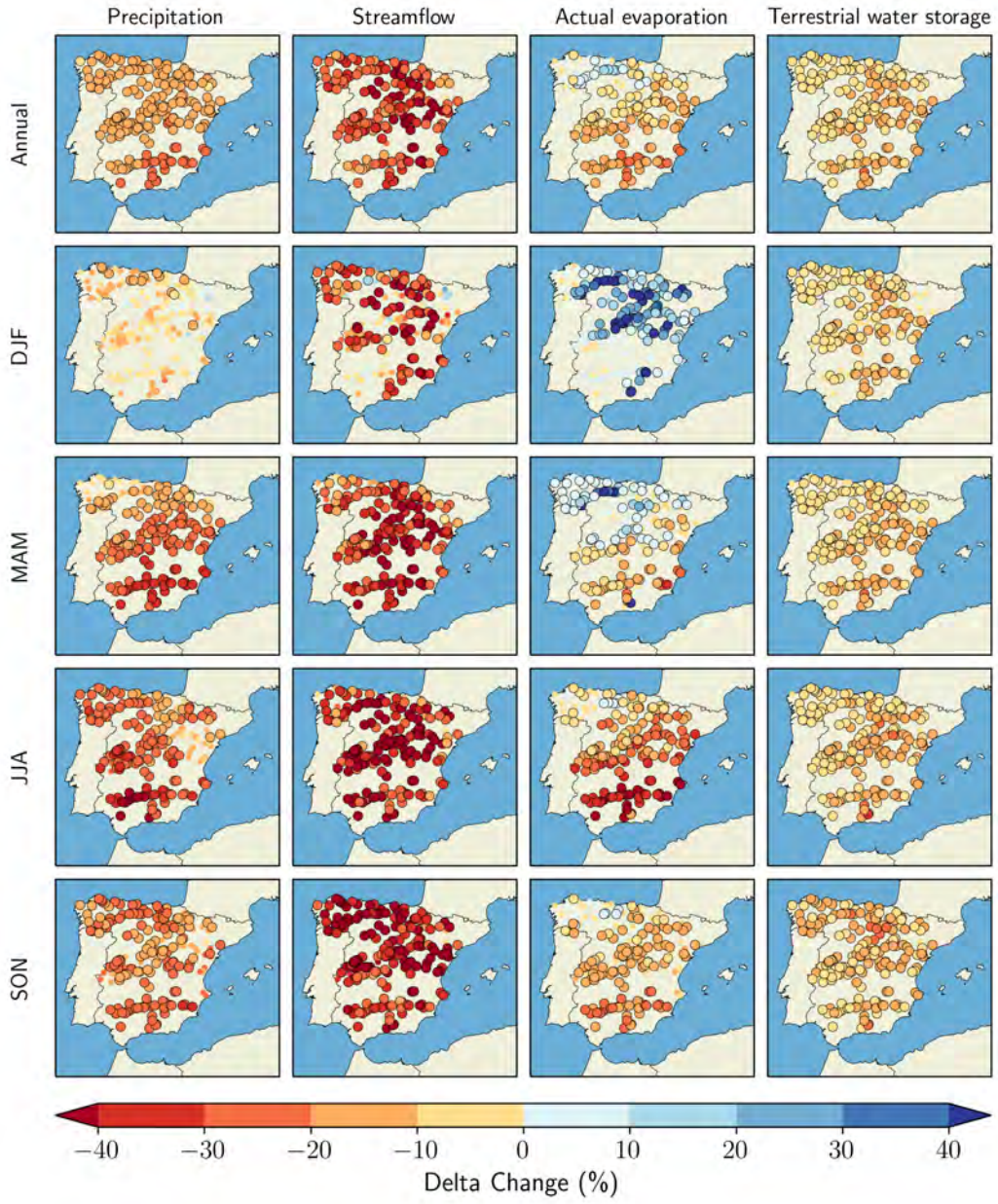


Fig. S7.7: Delta changes of annual and seasonal P, Q, E and TWS for the period 2071-2100 under the RCP4.5 scenario using WRFMPI. Significant changes at the 90% confidence level have been marked with solid borders.

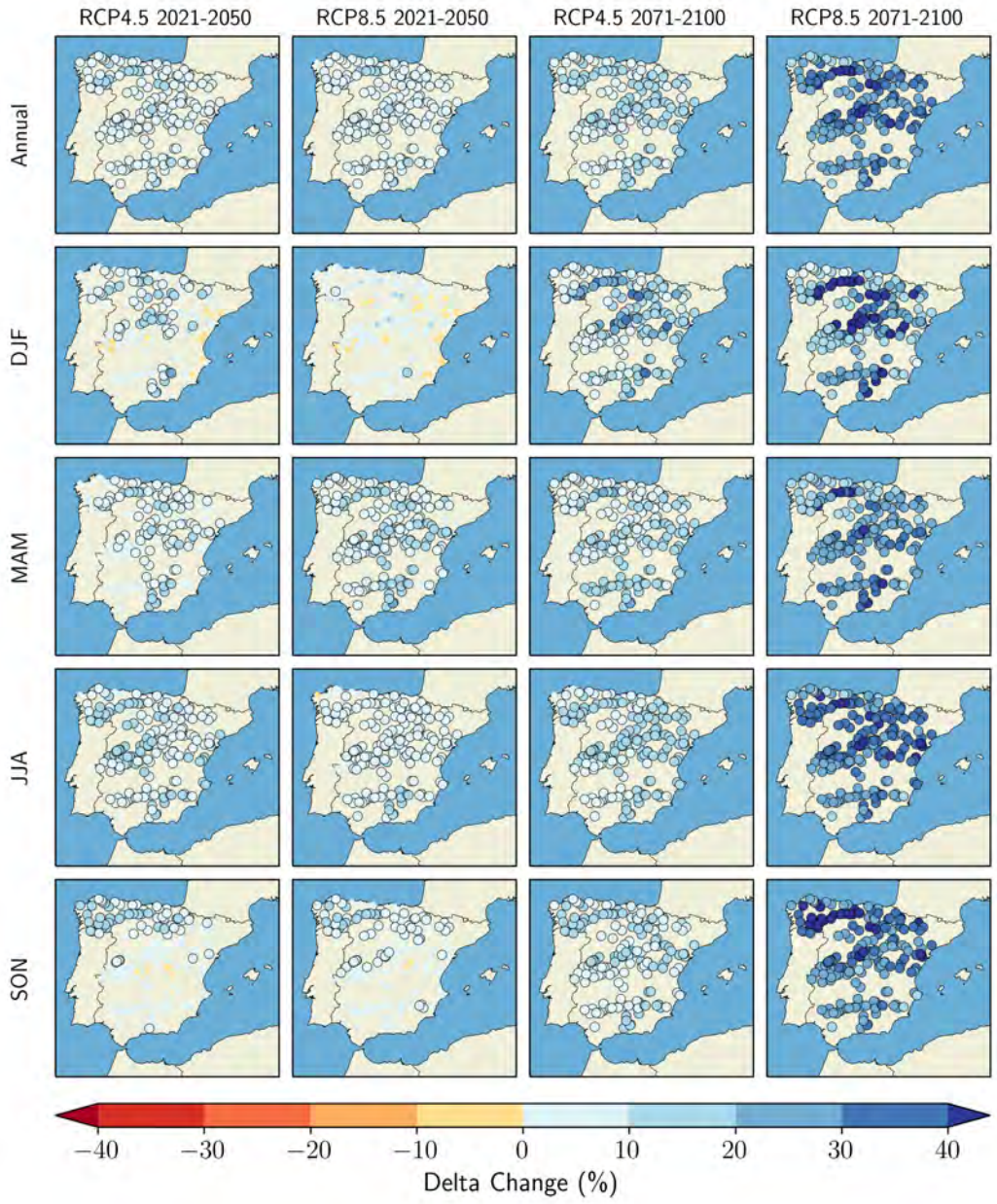


Fig. S7.8: Delta changes of E_p for all the future scenarios corresponding to WRFCCSM. Significant changes at the 90% confidence level have been marked with solid borders.

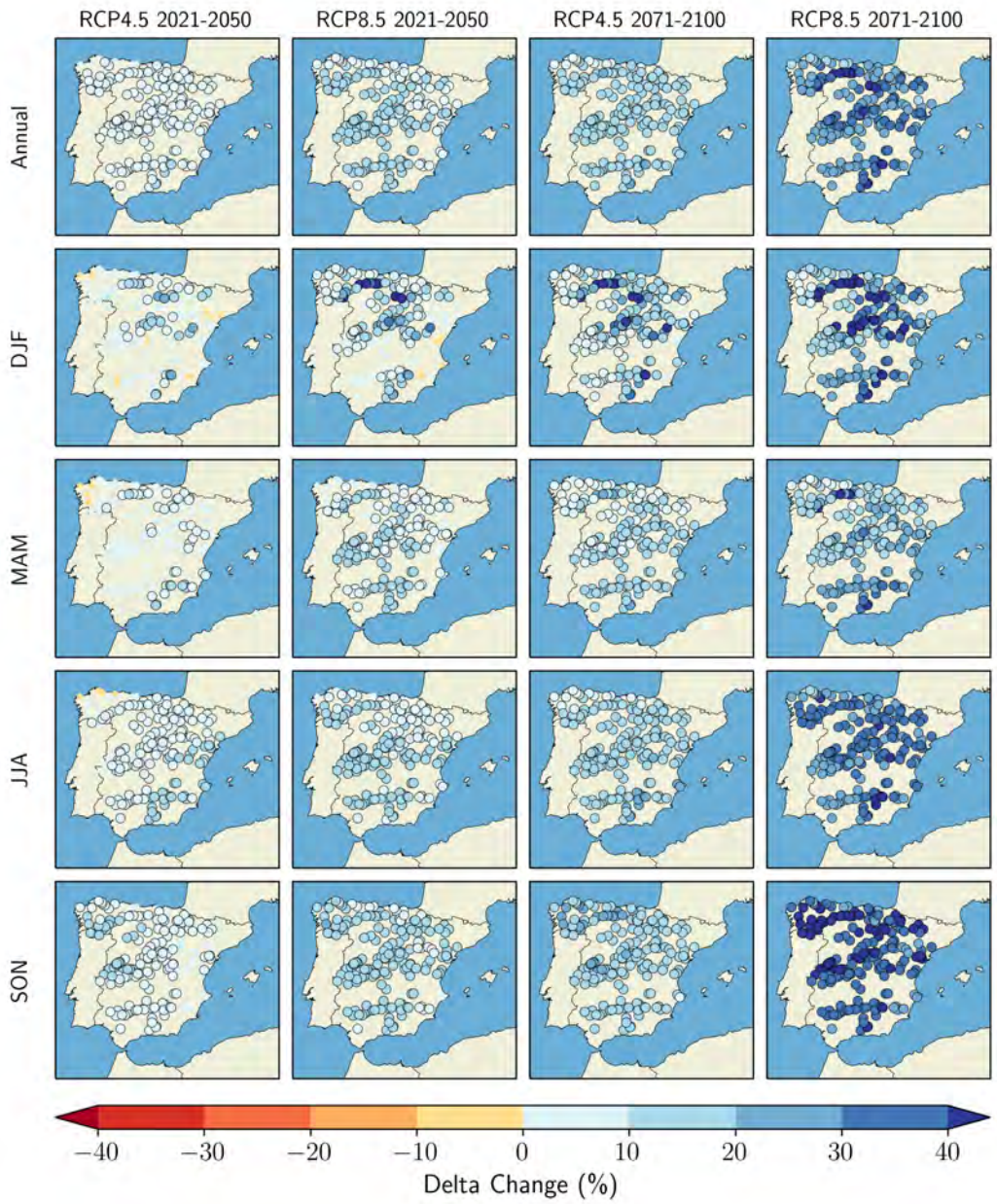


Fig. S7.9: Delta changes of E_p for all the future scenarios corresponding to WRFMPI. Significant changes at the 90% confidence level have been marked with solid borders.

Bibliography

Abraham, A. and Jain, L.: Evolutionary Multiobjective Optimization, in: Evolutionary Multiobjective Optimization: Theoretical Advances and Applications, edited by Abraham, A., Jain, L., and Goldberg, R., pp. 1–6, Springer London, London, doi:10.1007/1-84628-137-7_1, 2005.

Addor, N. and Melsen, L. A.: Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models, Water Resources Research, 55, 378–390, doi:10.1029/2018WR022958, 2019.

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, Hydrological Sciences Journal, 65, 712–725, doi:10.1080/02626667.2019.1683182, 2020.

Aerts, J. P., Hut, R. W., Van De Giesen, N. C., Drost, N., Van Verseveld, W. J., Weerts, A. H., and Hazenberg, P.: Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model, Hydrology and Earth System Sciences, 26, 4407–4430, doi:10.5194/hess-26-4407-2022, 2022.

Allen, G. H. and Pavelsky, T. M.: Global extent of rivers and streams, Science, 361, 585–588, doi:10.1126/science.aat0636, 2018.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al.: Crop evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and drainage paper 56, FAO, Rome, 1998.

Alvarez, J., Sánchez, A., and Quintas, L.: SIMPA, a GRASS based Tool for Hydrological Studies, International Journal of Geoinformatics, 1, 2005.

Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, doi:10.5194/hess-22-5817-2018, 2018.

Amponsah, W., Ayral, P. A., Boudevillain, B., Bouvier, C., Braud, I., Brunet, P., Delrieu, G., DIdon-Lescot, J. F., Gaume, E., Lebouc, L., Marchi, L., Marra, F., Morin, E., Nord, G., Payrastre, O., Zoccatelli, D., and Borga, M.: Integrated high-resolution dataset of high-intensity European and Mediterranean flash floods, Earth System Science Data, 10, 1783–1794, doi:10.5194/essd-10-1783-2018, 2018.

Argüeso, D., Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Esteban-Parra, M. J., and Castro-Díez, Y.: High-resolution projections of mean and extreme precipitation over Spain using the WRF model (2070-2099 versus 1970-1999), Journal of Geophysical Research Atmospheres, 117, doi:10.1029/2011JD017399, 2012.

Arsenault, R., Bazile, R., Ouellet Dallaire, C., and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrological Processes, 30, 2734–2736, doi:10.1002/hyp.10880, 2016.

Bandaragoda, C., Tarboton, D. G., and Woods, R.: Application of TOPNET in the distributed model intercomparison project, Journal of Hydrology, 298, 178–201, doi:10.1016/j.jhydrol.2004.03.038, 2004.

Beck, H. E., van Dijk, A. I., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resources Research, 52, 3599–3622, doi:10.1002/2015WR018247, 2016.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrology and Earth System Sciences, 21, 6201–6217, doi:10.5194/hess-21-6201-2017, 2017.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., and Wood, E. F.: Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments, Journal of Geophysical Research: Atmospheres, 125, doi:10.1029/2019JD031485, 2020.

Beck, H. E., Pan, M., Miralles, D. G., Reichle, R. H., Dorigo, W. A., Hahn, S., Sheffield, J., Karthikeyan, L., Balsamo, G., Parinussa, R. M., van Dijk, A. I., Du, J., Kimball, J. S., Vergopolan, N., and Wood, E. F.: Evaluation of 18 satellite- And model-based soil moisture products using in situ measurements from 826 sensors, Hydrology and Earth System Sciences, 25, 17–40, doi:10.5194/hess-25-17-2021, 2021.

Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., and Schüth, C.: Spatially distributed model calibration of a highly managed hydrological

system using remote sensing-derived ET data, Journal of Hydrology, 577, 123 944, doi:10.1016/j.jhydrol.2019.123944, 2019.

Bennett, K. E., Urrego Blanco, J. R., Jonko, A., Bohn, T. J., Atchley, A. L., Urban, N. M., and Middleton, R. S.: Global Sensitivity of Simulated Water Balance Indicators Under Future Climate Change in the Colorado Basin, Water Resources Research, 54, 132–149, doi:10.1002/2017WR020471, 2018.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K.: Rainfall-Runoff Modelling: The Primer: Second Edition, Wiley, doi:10.1002/9781119951001, 2012.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, doi:10.1002/hyp.3360060305, 1992.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11–29, doi:10.1016/S0022-1694(01)00421-8, 2001.

Bhatti, A. M., Koike, T., and Shrestha, M.: Climate change impact assessment on mountain snow hydrology by water and energy budget-based distributed hydrological model, Journal of Hydrology, 543, 523–541, doi:10.1016/j.jhydrol.2016.10.025, 2016.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resources Research, 51, 4923–4947, doi:10.1002/2015WR017173, 2015.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrological Processes, 29, 310–320, doi:10.1002/hyp.10391, 2015.

Blank, J. and Deb, K.: Pymoo: Multi-Objective Optimization in Python, IEEE Access, 8, 89 497–89 509, doi:10.1109/ACCESS.2020.2990567, 2020.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M.,

Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, Science, 357, 588–590, doi:10.1126/science.aan2506, 2017.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y. Y. Y. Y. Y. Y., Chen, Y. Y. Y. Y. Y. Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Prieto Sierra, C., Ramos, M. H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Van Loon, A. F., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J. P., von Freyberg, J., Vorogushyn, S., Wachniew, P.,

Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K., and Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH)–a community perspective, Hydrological Sciences Journal, 64, 1141–1158, doi:10.1080/02626667.2019.1620507, 2019a.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, doi:10.1038/s41586-019-1495-6, 2019b.

Blöschl, G., Kiss, A., Viglione, A., Barriendos, M., Böhm, O., Brázdil, R., Coeur, D., Demarée, G., Llasat, M. C., Macdonald, N., Retsö, D., Roald, L., Schmocker-Fackel, P., Amorim, I., Bělínová, M., Benito, G., Bertolin, C., Camuffo, D., Cornel, D., Doktor, R., Elleder, L., Enzi, S., Garcia, J. C., Glaser, R., Hall, J., Haslinger, K., Hofstätter, M., Komma, J., Limanówka, D., Lun, D., Panin, A., Parajka, J., Petrić, H., Rodrigo, F. S., Rohr, C., Schönbein, J., Schulte, L., Silva, L. P., Toonen, W. H., Valent, P., Waser, J., and Wetter, O.: Current European flood-rich period exceptional compared with past 500 years, Nature, 583, 560–566, doi:10.1038/s41586-020-2478-3, 2020.

Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., De Niel, J., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., and Hrachowitz, M.: Behind the scenes of streamflow model performance, Hydrology and Earth System Sciences, 25, 1069–1095, doi:10.5194/hess-25-1069-2021, 2021.

Ceballos-Barbancho, A., Morán-Tejeda, E., Luengo-Ugidos, M. Á., and Llorente-Pinto, J. M.: Water resources and environmental change in a Mediterranean environment: The south-west sector of the Duero river basin (Spain), Journal of Hydrology, 351, 126–138, doi:10.1016/j.jhydrol.2007.12.004, 2008.

Chaney, N. W., Herman, J. D., Reed, P. M., and Wood, E. F.: Flood and drought hydrologic monitoring: The role of model parameter uncertainty, Hydrology and Earth System Sciences, 19, 3239–3251, doi:10.5194/hess-19-3239-2015, 2015.

Chavez-Jimenez, A., Lama, B., Garrote, L., Martin-Carrasco, F., Sordo-Ward, A., and Mediero, L.: Characterisation of the Sensitivity of Water Resources Systems to Climate Change, Water Resources Management, 27, 4237–4258, doi:10.1007/s11269-013-0404-2, 2013.

Chawla, I. and Mujumdar, P. P.: Isolating the impacts of land use and climate change on streamflow, Hydrology and Earth System Sciences, 19, 3633–3651, doi:10.5194/hess-19-3633-2015, 2015.

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., and Palmer, R. N.: The effects of climate change on the hydrology and water resources of the Colorado River basin, Climatic Change, 62, 337–363, doi:10.1023/B:CLIM.0000013684.13621.1f, 2004.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models, Water Resources Research, 51, 5929–5956, doi:10.1002/2015WR017096, 2015a.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resources Research, 51, 2498–2514, doi:10.1002/2015WR017198, 2015b.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R., and Brekke, L. D.: Characterizing Uncertainty of the Hydrologic Impacts of Climate Change, Current Climate Change Reports, 2, 55–64, doi:10.1007/s40641-016-0034-x, 2016.

Clark, M. P., Bierkens, M. F., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences, 21, 3427–3440, doi:10.5194/hess-21-3427-2017, 2017.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The Abuse of Popular Performance Metrics in Hydrologic Modeling, Water Resources Research, 57, 1–16, doi:10.1029/2020WR029001, 2021.

Coello Coello, C. A.: Recent Trends in Evolutionary Multiobjective Optimization, in: Evolutionary Multiobjective Optimization, edited by Abraham, A.,

Jain, L., and Goldberg, R., pp. 7–32, Springer-Verlag, London, doi:10.1007/1-84628-137-7_2, 2005.

Collados-Lara, A. J., Pardo-Igúzquiza, E., and Pulido-Velazquez, D.: A distributed cellular automata model to simulate potential future impacts of climate change on snow cover area, Advances in Water Resources, 124, 106–119, doi:10.1016/j.advwatres.2018.12.010, 2019.

Coppola, E., Raffaele, F., and Giorgi, F.: Impact of climate change on snow melt driven runoff timing over the Alpine region, Climate Dynamics, 51, 1259–1273, doi:10.1007/s00382-016-3331-0, 2018.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), Journal of Geophysical Research: Atmospheres, 108, doi:10.1029/2002jd003316, 2003.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resources Research, 51, 5531–5546, doi:10.1002/2014WR016532, 2015.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth System Science Data, 12, 2459–2483, doi:10.5194/essd-12-2459-2020, 2020.

Cui, X., Guo, X., Wang, Y., Wang, X., Zhu, W., Shi, J., Lin, C., and Gao, X.: Application of remote sensing to water environmental processes under a changing climate, Journal of Hydrology, 574, 892–902, doi:10.1016/j.jhydrol.2019.04.078, 2019.

Deb, K.: An efficient constraint handling method for genetic algorithms, Computer Methods in Applied Mechanics and Engineering, 186, 311–338, doi:10.1016/S0045-7825(99)00389-8, 2000.

Deb, K.: Multi-objective optimization using evolutionary algorithms, Wiley, 2001.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 182–197, doi:10.1109/4235.996017, 2002.

Demaria, E. M., Nijssen, B., and Wagener, T.: Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, Jour-

nal of Geophysical Research Atmospheres, 112, doi:10.1029/2006JD007534, 2007.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resources Research, 56, 1–26, doi:10.1029/2019WR026085, 2020a.

Dembélé, M., Ceperley, N., Zwart, S. J., Salvadore, E., Mariethoz, G., and Schaefli, B.: Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies, Advances in Water Resources, 143, 103 667, doi:10.1016/j.advwatres.2020.103667, 2020b.

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., and Stisen, S.: Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, Hydrology and Earth System Sciences, 22, 1299–1315, doi:10.5194/hess-22-1299-2018, 2018.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global and Planetary Change, 57, 16–26, doi:10.1016/j.gloplacha.2006.11.030, 2007.

Diffenbaugh, N. S. and Giorgi, F.: Climate change hotspots in the CMIP5 global climate model ensemble, Climatic Change, 114, 813–822, doi:10.1007/s10584-012-0570-x, 2012.

Do, H. X., Westra, S., and Leonard, M.: A global-scale investigation of trends in annual maximum streamflow, Journal of Hydrology, 552, 28–43, doi:10.1016/j.jhydrol.2017.06.015, 2017.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata, Earth System Science Data, 10, 765–785, doi:10.5194/essd-10-765-2018, 2018.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, Journal of Hydrology, 158, 265–284, doi:10.1016/0022-1694(94)90057-4, 1994.

Efstratiadis, A. and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, Hydrological Sciences Journal, 55, 58–78, doi:10.1080/02626660903526292, 2010.

Ercan, M. B. and Goodall, J. L.: Design and implementation of a general software library for using NSGA-II with SWAT for multi-objective

model calibration, Environmental Modelling and Software, 84, 112–120, doi:10.1016/j.envsoft.2016.06.017, 2016.

Estrela, T. and Quintas, L.: El sistema integrado de modelización precipitación-aportación SIMPA, Ingeniería Civil, 0, 43–52, 1996.

Falcón-Cardona, J. G., Hernández Gómez, R., Coello Coello, C. A., and Castillo Tapia, M. G.: Parallel Multi-Objective Evolutionary Algorithms: A Comprehensive Survey, Swarm and Evolutionary Computation, 67, 100 960, doi:10.1016/j.swevo.2021.100960, 2021.

Faris, H., Aljarah, I., Mirjalili, S., Castillo, P. A., and Merelo, J. J.: EvoloPy: An open-source nature-inspired optimization framework in python, IJCCI 2016 - Proceedings of the 8th International Joint Conference on Computational Intelligence, 1, 171–177, doi:10.5220/0006048201710177, 2016.

Fersch, B., Senatore, A., Adler, B., Arnault, J., Mauder, M., Schneider, K., Völksch, I., and Kunstmann, H.: High-resolution fully coupled atmospheric-hydrological modeling: A cross-compartment regional water and energy cycle evaluation, Hydrology and Earth System Sciences, 24, 2457–2481, doi:10.5194/hess-24-2457-2020, 2020.

Fonseca, A. R. and Santos, J. A.: Predicting hydrologic flows under climate change: The Tâmega Basin as an analog for the Mediterranean region, Science of the Total Environment, 668, 1013–1024, doi:10.1016/j.scitotenv.2019.01.435, 2019.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement, Water Resources Research, 54, 9812–9832, doi:10.1029/2018WR023989, 2018.

Fowler, K. J., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth System Science Data, 13, 3847–3867, doi:10.5194/essd-13-3847-2021, 2021.

Francés, F., Vélez, J. I., and Vélez, J. J.: Split-parameter structure for the automatic calibration of distributed hydrological models, Journal of Hydrology, 332, 226–240, doi:10.1016/j.jhydrol.2006.06.032, 2007.

Franchini, M. and Pacciani, M.: Comparative analysis of several conceptual rainfall-runoff models, Journal of Hydrology, 122, 161–219, doi:10.1016/0022-1694(91)90178-K, 1991.

Gampe, D., Nikulin, G., and Ludwig, R.: Using an ensemble of regional climate models to assess climate change impacts on water scarcity in

European river basins, Science of the Total Environment, 573, 1503–1518, doi:10.1016/j.scitotenv.2016.08.053, 2016.

Gao, H., Birkett, C., and Lettenmaier, D. P.: Global monitoring of large reservoir storage from satellite remote sensing, Water Resources Research, 48, 2012WR012 063, doi:10.1029/2012WR012063, 2012.

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

García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Castro-Díez, Y., and Esteban-Parra, M. J.: Evaluation of WRF capability to detect dry and wet periods in Spain using drought indices, Journal of Geophysical Research, 122, 1569–1594, doi:10.1002/2016JD025683, 2017.

García-Valdecasas Ojeda, M., Yeste, P., Gámiz-Fortis, S. R., Castro-Díez, Y., and Esteban-Parra, M. J.: Future changes in land and atmospheric variables: An analysis of their couplings in the Iberian Peninsula, Science of the Total Environment, 722, doi:10.1016/j.scitotenv.2020.137902, 2020a.

García-Valdecasas Ojeda, M., Rosa-Cánovas, J. J., Romero-Jiménez, E., Yeste, P., Gámiz-Fortis, S. R., Castro-Díez, Y., and Esteban-Parra, M. J.: The role of the surface evapotranspiration in regional climate modelling: Evaluation and near-term future changes, Atmospheric Research, 237, doi:10.1016/j.atmosres.2020.104867, 2020b.

García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Romero-Jiménez, E., Rosa-Cánovas, J. J., Yeste, P., Castro-Díez, Y., and Esteban-Parra, M. J.: Projected changes in the Iberian Peninsula drought characteristics, Science of the Total Environment, 757, 143 702, doi:10.1016/j.scitotenv.2020.143702, 2021.

García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Romero-Jiménez, E., Rosa-Cánovas, J. J., Yeste, P., Castro-Díez, Y., and Esteban-Parra, M. J.: Projected changes in the Iberian Peninsula drought characteristics, Science of the Total Environment, 757, 143 702, doi:10.1016/j.scitotenv.2020.143702, 2021a.

García-Valdecasas Ojeda, M., Romero-Jiménez, E., Rosa-Cánovas, J. J., Yeste, P., Castro-Díez, Y., Esteban-Parra, M. J., Vicente-Serrano, S. M., and Gámiz-Fortis, S. R.: Assessing Future Drought Conditions over the Iberian Peninsula: The Impact of Using Different Periods to Compute the SPEI, Atmosphere, 12, 980, doi:10.3390/atmos12080980, 2021b.

Garrote, L., Granados, A., and Iglesias, A.: Strategies to reduce water stress in Euro-Mediterranean river basins, Science of the Total Environment, 543, 997–1009, doi:10.1016/j.scitotenv.2015.04.106, 2016.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, Hydrology and Earth System Sciences, 17, 149–161, doi:10.5194/hess-17-149-2013, 2013.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: An observation-based global gridded runoff dataset from 1902 to 2014, Earth System Science Data, 11, 1655–1674, doi:10.5194/essd-11-1655-2019, 2019.

Good, S. P., Noone, D., and Bowen, G.: Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, Science, 349, 175–177, doi:10.1126/science.aaa5931, 2015.

Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J., and Zhou, R.: Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China, Water Resources Research, 56, 1–19, doi:10.1029/2019WR025968, 2020.

Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), Earth System Science Data, 8, 279–295, doi:10.5194/essd-8-279-2016, 2016.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - A comparison of methods, Hydrology and Earth System Sciences, 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part 2: Quality control, time-series indices and homogeneity assessment, Earth System Science Data, 10, 787–804, doi:10.5194/essd-10-787-2018, 2018.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., and Seneviratne, S. I.: Observed Trends in Global Indicators of Mean and Extreme Streamflow, Geophysical Research Letters, 46, 756–766, doi:10.1029/2018GL079725, 2019.

Güntner, A., Reich, M., Mikolaj, M., Creutzfeldt, B., Schroeder, S., and Wziontek, H.: Landscape-scale water balance monitoring with an iGrav superconducting gravimeter in a field enclosure, Hydrology and Earth System Sciences, 21, 3167–3182, doi:10.5194/hess-21-3167-2017, 2017.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Haddeland, I., Matheussen, B. V., and Lettenmaier, D. P.: Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model, Water Resources Research, 38, 29–1–29–10, doi:10.1029/2001wr000854, 2002.

Hakala, K., Addor, N., and Seibert, J.: Hydrological modeling to evaluate climate model simulations and their bias correction, Journal of Hydrometeorology, 19, 1321–1337, doi:10.1175/JHM-D-17-0189.1, 2018.

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., and Mao, Y.: The variable infiltration capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility, Geoscientific Model Development, 11, 3481–3496, doi:10.5194/gmd-11-3481-2018, 2018.

Han, J., Yang, Y., Roderick, M. L., McVicar, T. R., Yang, D., Zhang, S., and Beck, H. E.: Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments, Water Resources Research, 56, 1–16, doi:10.1029/2020WR027392, 2020.

Hansen, M. C., Sohlberg, R., Defries, R. S., and Townshend, J. R.: Global land cover classification at 1 km spatial resolution using a classification tree approach, International Journal of Remote Sensing, 21, 1331–1364, doi:10.1080/014311600210209, 2000.

Hanzer, F., Förster, K., Nemec, J., and Strasser, U.: Projected cryospheric and hydrological impacts of 21st century climate change in the Ötztal Alps (Austria) simulated using a physically based approach, Hydrology and Earth System Sciences, 22, 1593–1614, doi:10.5194/hess-22-1593-2018, 2018.

Hasan, E., Tarhule, A., Kirstetter, P. E., Clark, R., and Hong, Y.: Runoff sensitivity to climate change in the Nile River Basin, Journal of Hydrology, 561, 312–321, doi:10.1016/j.jhydrol.2018.04.004, 2018.

Hay, L. E., Wilby, R. L., and Leavesley, G. H.: A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States, Journal of the American Water Resources Association, 36, 387–397, doi:10.1111/j.1752-1688.2000.tb04276.x, 2000.

Heistermann, M., Bogena, H., Francke, T., Guntner, A., Jakobi, J., Rasche, D., Schron, M., Dopper, V., Fersch, B., Groh, J., Patil, A., Putz, T., Reich, M., Zacharias, S., Zengerle, C., and Oswald, S.: Soil moisture observation in a forested headwater catchment: Combining a dense cosmic-ray neutron sensor network with roving and hydrogravimetry at the TERENO site Wüstebach, Earth System Science Data, 14, 2501–2519, doi:10.5194/essd-14-2501-2022, 2022.

Hengl, T., De Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G., Walsh, M. G., and

Gonzalez, M. R.: SoilGrids1km - Global soil information based on automated mapping, PLoS ONE, 9, doi:10.1371/journal.pone.0105992, 2014.

Her, Y., Yoo, S. H., Cho, J., Hwang, S., Jeong, J., and Seong, C.: Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions, Scientific Reports, 9, doi:10.1038/s41598-019-41334-7, 2019.

Hernandez-Suarez, J. S., Nejadhashemi, A. P., and Deb, K.: A novel multi-objective model calibration method for ecohydrological applications, Environmental Modelling & Software, 144, 105 161, doi:10.1016/j.envsoft.2021.105161, 2021.

Herrera, S., Soares, P. M., Cardoso, R. M., and Gutiérrez, J. M.: Evaluation of the EURO-CORDEX Regional Climate Models Over the Iberian Peninsula: Observational Uncertainty Analysis, Journal of Geophysical Research: Atmospheres, 125, doi:10.1029/2020JD032880, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, doi:10.1002/qj.3803, 2020.

Hornberger, G. and Spear, R.: An approach to the preliminary analysis of environmental systems, Journal of Environmental Management, 12, 7–18, 1981.

Iglesias, A., Garrote, L., Flores, F., and Moneo, M.: Challenges to manage the risk of water scarcity and climate change in the Mediterranean, Water Resources Management, 21, 775–788, doi:10.1007/s11269-006-9111-6, 2007.

Iman, R. L. and Conover, W. J.: A distribution-free approach to inducing rank correlation among input variables, Communications in Statistics - Simulation and Computation, 11, 311–334, doi:10.1080/03610918208812265, 1982.

IPCC: Climate Change 2014: Impacts, Adaptation and Vulnerability: Part B: Regional Aspects: Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, doi:10.1017/CBO9781107415386, 2014.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, vol. In Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/9781009157896, 2021.

Ishida, K., Ercan, A., Trinh, T., Kavvas, M. L., Ohara, N., Carr, K., and Anderson, M. L.: Analysis of future climate change impacts on snow distribution over mountainous watersheds in Northern California by means of a physically-based snow distribution model, Science of the Total Environment, 645, 1065–1082, doi:10.1016/j.scitotenv.2018.07.250, 2018.

Ishida, K., Ohara, N., Ercan, A., Jang, S., Trinh, T., Kavvas, M. L., Carr, K., and Anderson, M. L.: Impacts of climate change on snow accumulation and melting processes over mountainous regions in Northern California during the 21st century, Science of the Total Environment, 685, 104–115, doi:10.1016/j.scitotenv.2019.05.255, 2019.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J. F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: New high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, doi:10.1007/s10113-013-0499-2, 2014.

Jerez, S., Montavez, J. P., Gomez-Navarro, J. J., Jimenez, P. A., Jimenez-Guerrero, P., Lorente, R., and Gonzalez-Rouco, J. F.: The role of the land-surface model for climate change projections over the Iberian Peninsula, Journal of Geophysical Research Atmospheres, 117, doi:10.1029/2011JD016576, 2012.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, doi:10.1080/02626668609491024, 1986.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, doi:10.5194/hess-23-4323-2019, 2019.

Konapala, G., Mishra, A. K., Wada, Y., and Mann, M. E.: Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation, Nature Communications, 11, 1–10, doi:10.1038/s41467-020-16757-w, 2020.

Koppa, A., Gebremichael, M., and Yeh, W. W.: Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach, Advances in Water Resources, 130, 129–146, doi:10.1016/j.advwatres.2019.06.005, 2019.

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., and Miralles, D. G.: A deep learning-based hybrid model of global terrestrial evaporation, Nature Communications, 13, 1–11, doi:10.1038/s41467-022-29543-7, 2022.

Lamontagne, J. R., Barber, C. A., and Vogel, R. M.: Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data, Water Resources Research, 56, 1–25, doi:10.1029/2020WR027101, 2020.

LaTorre, A., Molina, D., Osaba, E., Poyatos, J., Del Ser, J., and Herrera, F.: A prescription of methodological guidelines for comparing bio-inspired optimization algorithms, Swarm and Evolutionary Computation, 67, 100 973, doi:10.1016/j.swevo.2021.100973, 2021.

Laumanns, M.: Self-adaptation and Convergence of Multiobjective Evolutionary Algorithms in Continuous Search Spaces, in: Evolutionary Multiobjective Optimization, edited by Abraham, A., Jain, L., and Goldberg, R., pp. 33–53, Springer-Verlag, London, doi:10.1007/1-84628-137-7_3, 2005.

Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., and Mankin, J. S.: The potential to reduce uncertainty in regional runoff projections from climate models, Nature Climate Change, 9, 926–933, doi:10.1038/s41558-019-0639-x, 2019.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, Water Resources Research, 51, 7309–7342, doi:10.1002/2015WR017616, 2015.

Li, C., Wang, L., Wanrui, W., Qi, J., Linshan, Y., Zhang, Y., Lei, W., Cui, X., and Wang, P.: An analytical approach to separate climate and human contributions to basin streamflow variability, Journal of Hydrology, 559, 30–42, doi:10.1016/j.jhydrol.2018.02.019, 2018.

Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, Journal of Geophysical Research Atmospheres, 114, doi:10.1029/2008JD010969, 2009.

Li, H., Li, X., Yang, D., Wang, J., Gao, B., Pan, X., Zhang, Y., and Hao, X.: Tracing Snowmelt Paths in an Integrated Hydrological Model for Understanding Seasonal Snowmelt Contribution at Basin Scale, Journal of Geophysical Research: Atmospheres, 124, 8874–8895, doi:10.1029/2019JD030760, 2019.

Liang, W., Bai, D., Wang, F., Fu, B., Yan, J., Wang, S., Yang, Y., Long, D., and Feng, M.: Quantifying the impacts of climate change and ecological restoration on streamflow changes based on a Budyko hydrological model in China's Loess Plateau, Water Resources Research, 51, 6500–6519, doi:10.1002/2014WR016589, 2015.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research, 99, doi:10.1029/94jd00483, 1994.

Liang, X., Wood, E. F., and Lettenmaier, D. P.: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, Global and Planetary Change, 13, 195–206, doi:10.1016/0921-8181(95)00046-1, 1996.

Liang, X., Guo, J., and Leung, L. R.: Assessment of the effects of spatial resolutions on daily water flux simulations, Journal of Hydrology, 298, 287–310, doi:10.1016/j.jhydrol.2003.07.007, 2004.

Lilhare, R., Pokorny, S., Déry, S. J., Stadnyk, T. A., and Koenig, K. A.: Sensitivity analysis and uncertainty assessment in water budgets simulated by the variable infiltration capacity model for Canadian subarctic watersheds, Hydrological Processes, 34, 2057–2075, doi:10.1002/hyp.13711, 2020.

Liu, W., Wang, L., Sun, F., Li, Z., Wang, H., Liu, J., Yang, T., Zhou, J., and Qi, J.: Snow Hydrology in the Upper Yellow River Basin Under Climate Change: A Land Surface Modeling Perspective, Journal of Geophysical Research: Atmospheres, 123, 12,676–12,691, doi:10.1029/2018JD028984, 2018.

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., and Bierkens, M. F. P.: Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products, Hydrology and Earth System Sciences, 21, 3125–3144, doi:10.5194/hess-21-3125-2017, 2017.

López-Moreno, J. I., Beguería, S., and García-Ruiz, J. M.: The management of a large Mediterranean reservoir: Storage regimens of the Yesa Reservoir, Upper Aragon River basin, Central Spanish Pyrenees, Environmental Management, 34, 508–515, doi:10.1007/s00267-003-0249-1, 2004.

López-Moreno, J. I., Goyette, S., and Beniston, M.: Impact of climate change on snowpack in the Pyrenees: Horizontal spatial variability and vertical gradients, Journal of Hydrology, 374, 384–396, doi:10.1016/j.jhydrol.2009.06.049, 2009.

Lorenzo-Lacruz, J., Vicente-Serrano, S. M., López-Moreno, J. I., Morán-Tejeda, E., and Zabalza, J.: Recent trends in Iberian streamflows (1945-2005), Journal of Hydrology, 414-415, 463–475, doi:10.1016/j.jhydrol.2011.11.023, 2012.

Lorenzo-Lacruz, J., Moŕan-Tejeda, E., Vicente-Serrano, S. M., and Ĺopez Moreno, J. I.: Streamflow droughts in the Iberian Peninsula between 1945 and 2005: Spatial and temporal patterns, Hydrology and Earth System Sciences, 17, 119–134, doi:10.5194/hess-17-119-2013, 2013.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. A., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., and Reed, P. M.: Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions, Environmental Modelling and Software, 62, 271–299, doi:10.1016/j.envsoft.2014.09.013, 2014.

Majone, B., Villa, F., Deidda, R., and Bellin, A.: Impact of climate change and water use policies on hydropower potential in the south-eastern Alpine region, Science of the Total Environment, 543, 965–980, doi:10.1016/j.scitotenv.2015.05.009, 2016.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environmental Research Letters, 10, doi:10.1088/1748-9326/10/11/114016, 2015.

Marcos-Garcia, P., Lopez-Nicolas, A., and Pulido-Velazquez, M.: Combined use of relative drought indices to analyze climate change impact on meteorological and hydrological droughts in a Mediterranean basin, Journal of Hydrology, 554, 292–305, doi:10.1016/j.jhydrol.2017.09.028, 2017.

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903–1925, doi:10.5194/gmd-10-1903-2017, 2017.

Martinez, G. F. and Gupta, H. V.: Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous United States, Water Resources Research, 46, doi:10.1029/2009WR008294, 2010.

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E. F., Pan, M., Sheffield, J., and Samaniego, L.: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 ºC, Hydrology and Earth System Sciences, 22, 1017–1032, doi:10.5194/hess-22-1017-2018, 2018.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-

domain hydrological models: Case study for a mesoscale pre-Alpine basin, Hydrology and Earth System Sciences, 20, 2207–2226, doi:10.5194/hess-20-2207-2016, 2016b.

Melsen, L. A. and Guse, B.: Hydrological Drought Simulations: How Climate and Model Structure Control Parameter Sensitivity, Water Resources Research, 55, 10 527–10 547, doi:10.1029/2019WR025230, 2019.

Melsen, L. A. and Guse, B.: Climate change impacts model parameter sensitivity-implications for calibration strategy and model diagnostic evaluation, Hydrology and Earth System Sciences, 25, 1307–1332, doi:10.5194/hess-25-1307-2021, 2021.

Melsen, L. A., Teuling, J. A., Torfs, J. J., Uijlenhoet, R., Mizukami, N., and Clark, P. M.: HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models, Hydrology and Earth System Sciences, 20, 1069–1079, doi:10.5194/hess-20-1069-2016, 2016a.

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., Uijlenhoet, R., and Teuling, A. J.: Mapping (dis)agreement in hydrologic projections, Hydrology and Earth System Sciences, 22, 1775–1791, doi:10.5194/hess-22-1775-2018, 2018.

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., and Arnold, J. R.: Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts, Journal of Hydrometeorology, 16, 762–780, doi:10.1175/JHM-D-14-0104.1, 2015.

Mendoza, P. A., Mizukami, N., Ikeda, K., Clark, M. P., Gutmann, E. D., Arnold, J. R., Brekke, L. D., and Rajagopalan, B.: Effects of different regional climate model resolution and forcing scales on projected hydrologic changes, Journal of Hydrology, 541, 1003–1019, doi:10.1016/j.jhydrol.2016.08.010, 2016.

Meresa, H. K. and Romanowicz, R. J.: The critical role of uncertainty in projections of hydrological extremes, Hydrology and Earth System Sciences, 21, 4245–4258, doi:10.5194/hess-21-4245-2017, 2017.

Miettinen, K.: Nonlinear Multiobjective Optimization, vol. 12 of *International Series in Operations Research & Management Science*, Springer US, Boston, MA, doi:10.1007/978-1-4615-5563-6, 1998.

Minville, M., Cartier, D., Guay, C., Leclaire, L.-A., Audet, C., Le Digabel, S., and Merleau, J.: Improving process representation in conceptual hydrological model calibration using climate simulations, Water Resources Research, 50, 5044–5073, doi:10.1002/2013WR013857, 2014.

Miralles, D. G., Holmes, T. R., De Jeu, R. A., Gash, J. H., Meesters, A. G., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, doi:10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Annals of the New York Academy of Sciences, 1436, 19–35, doi:10.1111/nyas.13912, 2019.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, Water Resources Research, 53, 8020–8040, doi:10.1002/2017WR020401, 2017.

Mizukami, N., Rakovec, O., Newman, A., Clark, M., Wood, A., Gupta, H., and Kumar, R.: On the choice of calibration metrics for high flow estimation using hydrologic models, Hydrology and Earth System Sciences Discussions, pp. 1–16, doi:10.5194/hess-2018-391, 2018.

Morán-Tejeda, E., Ceballos-Barbancho, A., and Llorente-Pinto, J. M.: Hydrological response of Mediterranean headwaters to climate oscillations and land-cover changes: The mountains of Duero River basin (Central Spain), Global and Planetary Change, 72, 39–49, doi:10.1016/j.gloplacha.2010.03.003, 2010.

Morán-Tejeda, E., Ignacio, L. M., Antonio, C. B., and Sergio M., V. S.: Evaluating Duero's basin (Spain) response to the NAO phases: Spatial and seasonal variability, Hydrological Processes, 25, 1313–1326, doi:10.1002/hyp.7907, 2011a.

Morán-Tejeda, E., López-Moreno, J. I., Ceballos-Barbancho, A., and Vicente-Serrano, S. M.: River regimes and recent hydrological changes in the Duero basin (Spain), Journal of Hydrology, 404, 241–258, doi:10.1016/j.jhydrol.2011.04.034, 2011b.

Morán-Tejeda, E., Ceballos-Barbancho, A., Llorente-Pinto, J. M., and López-Moreno, J. I.: Land-cover changes and recent hydrological evolution in the Duero Basin (Spain), Regional Environmental Change, 12, 17–33, doi:10.1007/s10113-011-0236-7, 2012a.

Morán-Tejeda, E., Lorenzo-Lacruz, J., López-Moreno, J. I., Ceballos-Barbancho, A., Zabalza, J., and Vicente-Serrano, S. M.: Reservoir Management in the Duero Basin (Spain): Impact on River Regimes and the Response to Environmental Change, Water Resources Management, 26, 2125–2146, doi:10.1007/s11269-012-0004-6, 2012b.

Morán-Tejeda, E., Lorenzo-Lacruz, J., López-Moreno, J. I., Rahman, K., and Beniston, M.: Streamflow timing of mountain rivers in Spain: Recent changes and future projections, Journal of Hydrology, 517, 1114–1127, doi:10.1016/j.jhydrol.2014.06.053, 2014.

Morán-Tejeda, E., López-Moreno, J. I., and Sanmiguel-Vallelado, A.: Changes in Climate, Snow and Water Resources in the Spanish Pyrenees: Observations and Projections in a Warming Climate, in: Advances in Global Change Research, vol. 62, pp. 305–323, Springer International Publishing, doi:10.1007/978-3-319-55982-7_13, 2017.

Moratiel, R., Snyder, R. L., Durán, J. M., and Tarquis, A. M.: Trends in climatic variables and future reference evapotranspiration in Duero Valley (Spain), Natural Hazards and Earth System Science, 11, 1795–1805, doi:10.5194/nhess-11-1795-2011, 2011.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Nathan, R. J., McMahon, T. A., Peel, M. C., and Horne, A.: Assessing the degree of hydrologic stress due to climate change, Climatic Change, 156, 87–104, doi:10.1007/s10584-019-02497-4, 2019.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, doi:10.5194/hess-19-209-2015, 2015.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., and Nearing, G.: Benchmarking of a physically based hydrologic model, Journal of Hydrometeorology, 18, 2215–2225, doi:10.1175/JHM-D-16-0284.1, 2017.

Nijssen, B., O'donnell, G. M., Hamlet, A. F., and Lettenmaier, D. P.: Hydrologic sensitivity of global rivers to climate change, Climatic Change, 50, 143–175, doi:10.1023/A:1010616428763, 2001.

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Nooijen, R. R. P., Savenije, H. H. G., and Hrachowitz, M.: Constraining Conceptual Hydrological Models With Multiple Information Sources, Water Resources Research, 54, 8332–8362, doi:10.1029/2017WR021895, 2018.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R.,

Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, doi:10.5194/gmd-9-3461-2016, 2016.

Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S., and Tootle, G.: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US, Hydrology and Earth System Sciences, 18, 67–84, doi:10.5194/hess-18-67-2014, 2014.

Papadimitriou, L. V., Koutroulis, A. G., Grillakis, M. G., and Tsanis, I. K.: High-end climate change impact on European runoff and low flows - Exploring the effects of forcing biases, Hydrology and Earth System Sciences, 20, 1785–1808, doi:10.5194/hess-20-1785-2016, 2016.

Páscoa, P., Gouveia, C. M., Russo, A., and Trigo, R. M.: Drought trends in the Iberian Peninsula over the last 112 years, Advances in Meteorology, 2017, doi:10.1155/2017/4653126, 2017.

Pastén-Zapata, E., Jones, J. M., Moggridge, H., and Widmann, M.: Evaluation of the performance of Euro-CORDEX Regional Climate Models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods, Journal of Hydrology, 584, doi:10.1016/j.jhydrol.2020.124653, 2020.

Pavlyukevich, I.: Lévy flights, non-local search and simulated annealing, Journal of Computational Physics, 226, 1830–1844, doi:10.1016/j.jcp.2007.06.008, 2007.

Pellicer-Martinez, F. and Martínez-Paz, J. M.: Contrast and transferability of parameters of lumped water balance models in the Segura River Basin (Spain), Water and Environment Journal, 29, 43–50, doi:10.1111/wej.12091, 2015a.

Pellicer-Martínez, F. and Martínez-Paz, J. M.: Climate change effects on the hydrology of the headwaters of the Tagus River: Implications for the management of the Tagus-Segura transfer, Hydrology and Earth System Sciences, 22, 6473–6491, doi:10.5194/hess-22-6473-2018, 2018.

Pellicer-Martínez, F., González-Soto, I., and Martínez-Paz, J. M.: Analysis of incorporating groundwater exchanges in hydrological models, Hydrological Processes, 29, 4361–4366, doi:10.1002/hyp.10586, 2015b.

Pellicer-Martínez, F., Gomariz-Castillo, F., Portela, M. M., Martínez-Alcalá, I. M., and Martínez-Paz, J. M.: Modulation of the goodness of fit in hydrological modelling based on inner balance errors, PloS one, 16, e0260 117, doi:10.1371/journal.pone.0260117, 2021.

Peterson, T. J., Saft, M., Peel, M. C., and John, A.: Watersheds may not recover from drought, Science, 372, 745–749, doi:10.1126/science.abd5085, 2021.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis, Environmental Modelling and Software, 70, 80–85, doi:10.1016/j.envsoft.2015.04.009, 2015.

Pokhrel, P. and Gupta, H. V.: On the use of spatial regularization strategies to improve calibration of distributed watershed models, Water Resources Research, 46, 1–17, doi:10.1029/2009WR008066, 2010.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C. E., Thiery, W., Veldkamp, T., Zhao, F., and Wada, Y.: Global terrestrial water storage and drought severity under climate change, Nature Climate Change, 11, 226–233, doi:10.1038/s41558-020-00972-w, 2021.

Pool, S., Vis, M., and Seibert, J.: Regionalization for Ungauged Catchments — Lessons Learned From a Comparative Large-Sample Study, Water Resources Research, 57, 1–25, doi:10.1029/2021WR030437, 2021.

Puertes, C., Lidón, A., Echeverría, C., Bautista, I., González-Sanchis, M., del Campo, A. D., and Francés, F.: Explaining the hydrological behaviour of facultative phreatophytes using a multi-variable and multi-objective modelling approach, Journal of Hydrology, 575, 395–407, doi:10.1016/j.jhydrol.2019.05.041, 2019.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resources Research, 50, 409–426, doi:10.1002/2013WR014063, 2014.

Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resources Research, 52, 7779–7792, doi:10.1002/2016WR019430, 2016a.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287–307, doi:10.1175/JHM-D-15-0054.1, 2016b.

Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., Clark, M. P., and Samaniego, L.: Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United

States, Journal of Geophysical Research: Atmospheres, 124, 13 991–14 007, doi:10.1029/2019JD030767, 2019.

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., Gutmann, E., Dudhia, J., Chen, F., Barlage, M., Yates, D., and Zhang, G.: Climate Change Impacts on the Water Balance of the Colorado Headwaters: High-Resolution Regional Climate Model Simulations, Journal of Hydrometeorology, 15, 1091–1116, doi:10.1175/JHM-D-13-0118.1, 2014.

Ritter, A. and Muñoz-Carpena, R.: Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, Journal of Hydrology, 480, 33–45, doi:10.1016/j.jhydrol.2012.12.004, 2013.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004.

Rummukainen, M.: State-of-the-art with regional climate models, Wiley Interdisciplinary Reviews: Climate Change, 1, 82–96, doi:10.1002/wcc.8, 2010.

Ruosteenoja, K., Markkanen, T., Venäläinen, A., Räisänen, P., and Peltola, H.: Seasonal soil moisture and drought occurrence in Europe in CMIP5 projections for the 21st century, Climate Dynamics, 50, 1177–1192, doi:10.1007/s00382-017-3671-4, 2018.

Saha, A., Joseph, J., and Ghosh, S.: Climate controls on the terrestrial water balance: Influence of aridity on the basin characteristics parameter in the Budyko framework, Science of the Total Environment, 739, doi:10.1016/j.scitotenv.2020.139863, 2020.

Salmoral, G., Willaarts, B. A., Troch, P. A., and Garrido, A.: Drivers influencing streamflow changes in the Upper Turia basin, Spain, Science of the Total Environment, 503-504, 258–268, doi:10.1016/j.scitotenv.2014.07.041, 2015.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis practices: Strategies for model-based inference, Reliability Engineering and System Safety, 91, 1109–1125, doi:10.1016/j.ress.2005.11.014, 2006.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: The primer, John Wiley & Sons, Ltd, doi:10.1002/9780470725184, 2008.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, doi:10.1029/2008WR007327, 2010.

Santos, L., Thirel, G., and Perrin, C.: Technical note: Pitfalls in using log-transformed flows within the KGE criterion, Hydrology and Earth System Sciences, 22, 4583–4591, doi:10.5194/hess-22-4583-2018, 2018.

Savenije, H. H.: HESS opinions: "The art of hydrology", Hydrology and Earth System Sciences, 13, 157–161, doi:10.5194/hess-13-157-2009, 2009.

Schulzweida, U.: CDO User Guide (Version 1.9.8), doi:10.5281/zenodo.3539275, 2019.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Sepúlveda, U. M., Mendoza, P. A., Mizukami, N., and Newman, A. J.: Revisiting parameter sensitivities in the variable infiltration capacity model across a hydroclimatic gradient, Hydrology and Earth System Sciences, 26, 3419–3445, doi:10.5194/hess-26-3419-2022, 2022.

Serrano-Notivoli, R., Beguería, S., Saz, M. Á., Longares, L. A., and De Luis, M.: SPREAD: A high-resolution daily gridded precipitation dataset for Spain - An extreme events frequency and intensity overview, Earth System Science Data, 9, 721–738, doi:10.5194/essd-9-721-2017, 2017.

Serrano-Notivoli, R., Beguería, S., and De Luis, M.: STEAD: A high-resolution daily gridded temperature dataset for Spain, Earth System Science Data, 11, 1171–1188, doi:10.5194/essd-11-1171-2019, 2019.

Siderius, C., Biemans, H., Wiltshire, A., Rao, S., Franssen, W. H., Kumar, P., Gosain, A. K., van Vliet, M. T., and Collins, D. N.: Snowmelt contributions to discharge of the Ganges, Science of the Total Environment, 468-469, doi:10.1016/j.scitotenv.2013.05.084, 2013.

Soares, P. M., Cardoso, R. M., Lima, D. C., and Miranda, P. M.: Future precipitation in Portugal: high-resolution projections using WRF model and EURO-CORDEX multi-model ensembles, Climate Dynamics, 49, 2503–2530, doi:10.1007/s00382-016-3455-2, 2017.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, Journal of Hydrology, 523, 739–757, doi:10.1016/j.jhydrol.2015.02.013, 2015.

Stewart, I. T., Cayan, D. R., and Dettinger, M. D.: Changes toward earlier streamflow timing across western North America, Journal of Climate, 18, 1136–1155, doi:10.1175/JCLI3321.1, 2005.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrology and Earth System Sciences, 6, 85–100, doi:10.5194/hess-6-85-2002, 2002.

Suárez-Almiñana, S., Solera, A., Madrigal, J., Andreu, J., and Paredes-Arquiola, J.: Risk assessment in water resources planning under climate change at the Júcar River basin, Hydrology and Earth System Sciences, 24, 5297–5315, doi:10.5194/hess-24-5297-2020, 2020.

Sun, R., Hernández, F., Liang, X., and Yuan, H.: A Calibration Framework for High-Resolution Hydrological Models Using a Multiresolution and Heterogeneous Strategy, Water Resources Research, 56, 1–26, doi:10.1029/2019WR026541, 2020.

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., and Blöschl, G.: The Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a Small Catchment, Water Resources Research, 56, doi:10.1029/2019WR026153, 2020.

Tapiador, F. J., Navarro, A., Moreno, R., Sánchez, J. L., and García-Ortega, E.: Regional climate models: 30 years of dynamical downscaling, Atmospheric Research, 235, doi:10.1016/j.atmosres.2019.104785, 2020.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Te Chow, V., Maidment, D., and Mays, L.: Applied Hydrology, Civil Engineering, McGraw-Hill, 1988.

Teutschbein, C. and Seibert, J.: Regional climate models for hydrological impact studies at the catchment scale: A review of recent modeling strategies, Geography Compass, 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, 2010.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Thackeray, C. W., Derksen, C., Fletcher, C. G., and Hall, A.: Snow and Climate: Feedbacks, Drivers, and Indices of Change, Current Climate Change Reports, 5, 322–333, doi:10.1007/s40641-019-00143-w, 2019.

Themeßl, M., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31, 1530–1544, doi:10.1002/joc.2168, 2011.

Thrasher, B., Maurer, E. P., McKellar, C., and Duffy, P. B.: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrology and Earth System Sciences, 16, 3309–3314, doi:10.5194/hess-16-3309-2012, 2012.

Tomas-Burguera, M., Vicente-Serrano, S. M., Peña-Angulo, D., Domínguez-Castro, F., Noguera, I., and El Kenawy, A.: Global Characterization of the Varying Responses of the Standardized Precipitation Evapotranspiration Index to Atmospheric Evaporative Demand, Journal of Geophysical Research: Atmospheres, 125, doi:10.1029/2020JD033017, 2020.

Tóth, B., Weynants, M., Pásztor, L., and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrological Processes, 31, 2662–2666, doi:10.1002/hyp.11203, 2017.

Tramblay, Y., Koutroulis, A., Samaniego, L., Vicente-Serrano, S. M., Volaire, F., Boone, A., Le Page, M., Llasat, M. C., Albergel, C., Burak, S., Cailleret, M., Kalin, K. C., Davi, H., Dupuy, J. L., Greve, P., Grillakis, M., Hanich, L., Jarlan, L., Martin-StPaul, N., Martínez-Vilalta, J., Mouillot, F., Pulido-Velazquez, D., Quintana-Seguí, P., Renard, D., Turco, M., Türkeş, M., Trigo, R., Vidal, J. P., Vilagrosa, A., Zribi, M., and Polcher, J.: Challenges for drought assessment in the Mediterranean region under future climate scenarios, Earth-Science Reviews, 210, doi:10.1016/j.earscirev.2020.103348, 2020.

Vano, J. A., Nijssen, B., and Lettenmaier, D. P.: Seasonal hydrologic responses to climate change in the Pacific Northwest, Water Resources Research, 51, 1959–1976, doi:10.1002/2014WR015909, 2015.

VanShaar, J. R., Haddeland, I., and Lettenmaier, D. P.: Effects of land-cover changes on the hydrological response of interior Columbia River basin forested catchments, Hydrological Processes, 16, 2499–2520, doi:10.1002/hyp.1017, 2002.

Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., and Espejo, F.: Evidence of increasing drought severity caused by temperature rise in southern Europe, Environmental Research Letters, 9, doi:10.1088/1748-9326/9/4/044001, 2014.

Vieira, J., Cunha, M. C., and Luís, R.: Integrated Assessment of Water Reservoir Systems Performance with the Implementation of Ecological Flows under Varying Climatic Conditions, Water Resources Management, 32, 5183–5205, doi:10.1007/s11269-018-2153-8, 2018.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, Remote Sensing of Environment, 115, 801–823, doi:10.1016/j.rse.2010.11.006, 2011.

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., Huang, Y., Koboltschnig, G., Litaor, M. I., López-Moreno, J. I., Lorentz, S., Schädler, B., Schreier, H., Schwaiger, K., Vuille, M., and Woods, R.: Climate change and mountain water resources: Overview and recommendations for research, management and policy, Hydrology and Earth System Sciences, 15, 471–504, doi:10.5194/hess-15-471-2011, 2011.

Vormoor, K., Lawrence, D., Heistermann, M., and Bronstert, A.: Climate change impacts on the seasonality and generation processes of floods - Projections and uncertainties for catchments with mixed snowmelt/rainfall regimes, Hydrology and Earth System Sciences, 19, 913–931, doi:10.5194/hess-19-913-2015, 2015.

Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, Proceedings of the National Academy of Sciences of the United States of America, 104, 708–711, doi:10.1073/pnas.0610471104, 2007.

Vrugt, J. A., Robinson, B. A., and Hyman, J. M.: Self-adaptive multimethod search for global optimization in real-parameter spaces, IEEE Transactions on Evolutionary Computation, 13, 243–259, doi:10.1109/TEVC.2008.924428, 2009.

Wolpert, D. and Macready, W.: No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation, 1, 67–82, doi:10.1109/4235.585893, 1997.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophysical Research Letters, 35, doi:10.1029/2008GL034648, 2008.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, Journal of Geophysical Research: Atmospheres, 107, ACL 6–1–ACL 6–15, doi:10.1029/2001JD000659, 2002.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge

for monitoring Earth's terrestrial water, Water Resources Research, 47, 1–10, doi:10.1029/2010WR010090, 2011.

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., and Wang, J.: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, Water Resources Research, 50, 2693–2717, doi:10.1002/2013WR014710, 2014.

Xu, X., Yang, D., Yang, H., and Lei, H.: Attribution analysis based on the Budyko hypothesis for detecting the dominant cause of runoff decline in Haihe basin, Journal of Hydrology, 510, 530–540, doi:10.1016/j.jhydrol.2013.12.052, 2014.

Yang, X., Yu, X., Wang, Y., Liu, Y., Zhang, M., Ren, L., Yuan, F., and Jiang, S.: Estimating the response of hydrological regimes to future projections of precipitation and temperature over the upper Yangtze River, Atmospheric Research, 230, doi:10.1016/j.atmosres.2019.104627, 2019.

Yang, X.-S.: Flower pollination algorithm for global optimization, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7445 LNCS, pp. 240–249, Springer Berlin Heidelberg, doi:10.1007/978-3-642-32894-7_27, 2012.

Yang, X.-S. and Deb, S.: Eagle strategy using Lévy walk and firefly algorithms for stochastic optimization, vol. 284, Springer Berlin Heidelberg, doi:10.1007/978-3-642-12538-6_9, 2010.

Yang, X.-S., Karamanoglu, M., and He, X.: Flower pollination algorithm: A novel approach for multiobjective optimization, Engineering Optimization, 46, 1222–1237, doi:10.1080/0305215X.2013.832237, 2014.

Yang, Y., McVicar, T. R., Donohue, R. J., Zhang, Y., Roderick, M. L., Chiew, F. H., Zhang, L., and Zhang, J.: Lags in hydrologic recovery following an extreme drought: Assessing the roles of climate and catchment characteristics, Water Resources Research, 53, 4821–4837, doi:10.1002/2017WR020683, 2017.

Yeste, P., Dorador, J., Martin-Rosales, W., Molero, E., Esteban-Parra, M. J., and Rueda, F. J.: Climate-driven trends in the streamflow records of a reference hydrologic network in Southern Spain, Journal of Hydrology, 566, 55–72, doi:10.1016/j.jhydrol.2018.08.063, 2018.

Yeste, P., García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Castro-Díez, Y., and Esteban-Parra, M. J.: Integrated sensitivity analysis of a macroscale hydrologic model in the north of the Iberian Peninsula, Journal of Hydrology, 590, 125 230, doi:10.1016/j.jhydrol.2020.125230, 2020.

Yeste, P., Rosa-Cánovas, J. J., Romero-Jiménez, E., García-Valdecasas Ojeda, M., Gámiz-Fortis, S. R., Castro-Díez, Y., and Esteban-Parra, M. J.: Projected

hydrologic changes over the north of the Iberian Peninsula using a Euro-CORDEX multi-model ensemble, Science of The Total Environment, 777, 146 126, doi:10.1016/j.scitotenv.2021.146126, 2021.

Zhang, B., Xia, Y., Long, B., Hobbins, M., Zhao, X., Hain, C., Li, Y., and Anderson, M. C.: Evaluation and comparison of multiple evapotranspiration data models over the contiguous United States: Implications for the next phase of NLDAS (NLDAS-Testbed) development, Agricultural and Forest Meteorology, 280, 107 810, doi:10.1016/j.agrformet.2019.107810, 2020.

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., and Thomas, M.: Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, Hydrology and Earth System Sciences, 21, 821–837, doi:10.5194/hess-21-821-2017, 2017.

Zhao, R.-J., Zuang, Y.-L., Fang, L.-R., Liu, X.-R., and Zhang, Q.-S.: The Xinanjiang Model, IAHS-AISH Publication, pp. 351–356, 1980.