Journal of Hydrology 590 (2020) 125230

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

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

Patricio Yeste[a], Matilde García-Valdecasas Ojeda[b], Sonia R. Gámiz-Fortis[a], Yolanda Castro-Díez[a], María Jesús Esteban-Parra[a,*]

[a] Dept. Applied Physics, University of Granada, Spain
[b] Instituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Italy

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Massimiliano Zappa, Associate Editor

Keywords:
Duero River Basin
VIC model
Calibration
Model evaluation
Sensitivity analysis
Equifinality

ABSTRACT

Process-based hydrologic models allow to identify the behavior of a basin providing a mathematical description of the hydrologic processes underlying the runoff mechanisms that govern the streamflow generation. This study focuses on a macroscale application of the Variable Infiltration Capacity (VIC) model over 31 headwater subwatersheds 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 subwatersheds; (2) an integrated sensitivity analysis concerning the soil parameters chosen 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 subwatersheds that largely improved the benchmark performance. The spatial performance reflected a high transferability for most parameter combinations, and the least transferable were related to subwatersheds 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 subwatershed. 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 subwatersheds 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.

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). As a part of the Mediterranean region, the Iberian Peninsula conforms a vulnerable area that has been identified as a hotspot (Diffenbaugh and Giorgi, 2012) where the streamflows have shown a marked reduction during the last half century (Lorenzo-Lacruz et al., 2012, 2013).

Being able to identify the hydrologic behavior 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

* Corresponding author.
E-mail address: esteban@ugr.es (M.J. Esteban-Parra).

https://doi.org/10.1016/j.jhydrol.2020.125230
Received 4 September 2019; Received in revised form 10 June 2020; Accepted 22 June 2020
Available online 26 June 2020

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 modeling 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., 2019).

In the context of climate modeling, these models are usually called Land-Surface Models (LSMs) and are coupled to General Circulation Models (GCMs) and Regional Climate Models (RCMs) as the land scheme that allows to simulate the biophysical processes involved in the land-atmosphere interaction (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., 2015). In this respect, the Variable Infiltration Capacity (VIC) macroscale hydrologic model (Liang et al., 1994, 1996) has played the role of both LSM and hydrologic model in many previous studies. Melsen et al. (2016b) provided sufficient evidence in a meta-analysis of 192 peer-reviewed studies where the VIC model was calibrated and validated. Current research has aimed at calibrating 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) and 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 subwatersheds of an important basin located in the north of the Iberian Peninsula, the Duero River Basin. For this end, the hydrologic modeling exercise here developed has been divided into three interrelated parts:

- The calibration of the VIC model for the selected subwatersheds 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 subwatersheds 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

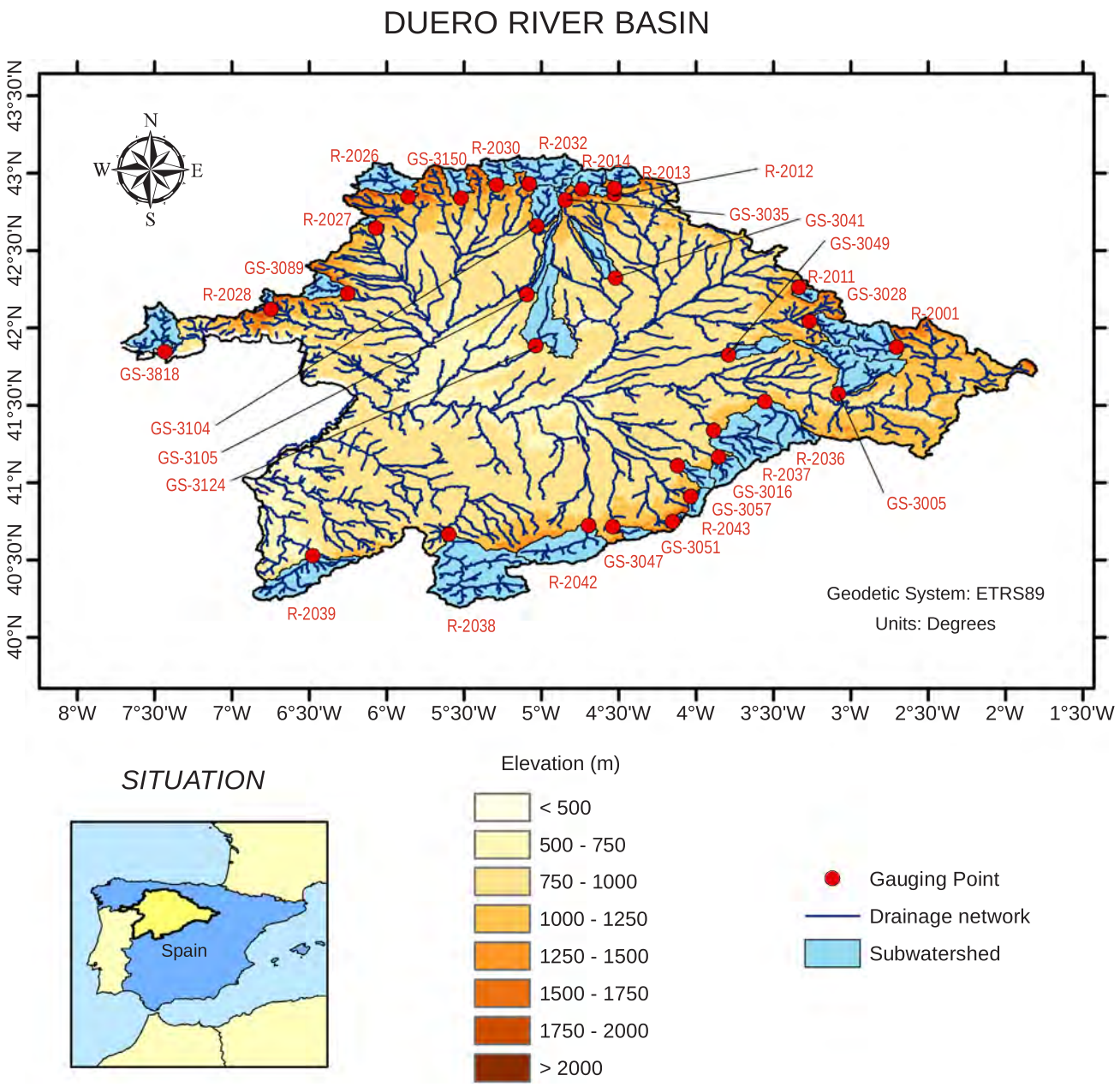

Fig. 1. Duero River Basin and the 31 studied subwatersheds. The prefix "R-" denotes "Reservoir" and the prefix "GS-" denotes "Gauging Station".

hydroclimatic and land-surface variables and contributed to a better understanding of the hydrologic behavior of the Duero River Basin. The hydrologic modeling analysis carried out in this work 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 Sect. 2 and 3 the study area and the methods are described. Sect. 4 gathers the results of the three-part approach and Sect. 5 corresponds to the discussion of the key results. Finally, the main conclusions of this study are provided in Sect. 6.

2. Study area

The Duero River Basin constitutes the largest basin of the Iberian Peninsula with a surface of 98,073 km^2. It is a shared territory between Spain and Portugal, characterized by a high water contribution (~15,000 hm^3/year). The study is focused on the Spanish part of the basin (Fig. 1), which represents the 80% of the area (78,859 km^2). Most of this territory constitutes a plain 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 50,000 hm^3 which is mostly lost into the atmosphere through evaporative fluxes (~35,000 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 2 °C in January, while summer is soft with maximum temperatures occurring in July (~20.5 °C).

The Duero River Basin is regulated by a total of 31 reservoirs where the reservoir inflow records are estimated through a water balance of the daily storages and water releases. Additional streamflow monitoring is also carried out in a large network of ~200 gauging stations where the streamflow records are calculated through the rating curves.

3. Methods

3.1. Hydrologic dataset

The streamflow records were gathered in a monthly basis from the Spanish Centre for Public Work Experimentation and Study (CEDEX, *Centro de Estudios y Experimentación de Obras Públicas*) database for all the reservoirs and gauging stations of the Duero Basin. An analysis of the percentage of gaps in the time series revealed that the period from October 2000 to September 2011 presents less than 5% of missing values for all the time series. Therefore, it was chosen as the study period for this work. A reference hydrologic network was then defined applying the criterion of the absence of upstream hydraulic structures (Whitfield et al., 2012). This selection reduced the number of reservoirs and gauging stations for the analysis to 16 reservoirs and 15 gauging stations covering the headwaters of the Duero River Basin (Fig. 1). Table 1 collects the main characteristics of these subwatersheds: 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), for the study period.

Precipitation, maximum temperature and minimum temperature data were extracted from two high-resolution (~5 km × 5 km) daily gridded datasets: SPREAD (Serrano-Notivoli et al., 2017) for precipitation data and STEAD (available at https://doi.org/10.20350/digitalCSIC/8622) for temperature data. Both datasets cover Peninsular Spain and the Balearic and Canary Islands and were built with information from observed precipitation and maximum and minimum temperature at a varying number of meteorological stations provided by several administrations including the Spanish Meteorological Agency (AEMET) and some hydrologic confederations. Atmospheric pressure, incoming shortwave radiation, incoming longwave radiation, vapor pressure and wind speed data were taken from daily outputs of high resolution (0.088°, ~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).

3.2. Hydrologic modeling

3.2.1. The VIC model

The VIC model (Liang et al., 1994, 1996) is a semi-distributed macroscale hydrologic model that computes both the water and the energy balance within the grid cell. The sub-grid variability in land cover classes is evaluated statistically, and a spatially heterogeneous structure for the infiltration capacity is assumed using the formulation described in the Xinanjiang model (Zhao et al., 1980). This approach takes into account the sub-grid variability in the soil moisture storage capacity.

The water balance in the VIC model considers three types of evaporation: evaporation from bare soil, evaporation from the canopy layer for each vegetation class and transpiration from the different types of vegetation. Potential evapotranspiration is calculated from the Penman-Monteith equation, and it represents the atmospheric demand for water vapor. Actual evapotranspiration in a grid cell is obtained as the sum of the three evaporation types weighted by the fraction of the area corresponding to each land cover class.

Different algorithms for the runoff generation process must be contemplated depending on the number of soil layers defined. The three-layer VIC model (VIC-3L) is the most common application and has

Table 1
Main characteristics of the 31 subwatersheds.

Code	Name	Area (km^2)	Mean elevation (m)	P_{an} (mm/year)	PET_{an} (mm/year)	Q_{an} (hm^3/year)
R-2001	CUERDA DEL POZO	546.7	1319	1122	900	188.7
R-2011	ARLANZON	106.7	1440	1293	798	77.6
R-2012	CERVERA DE RUESGA	54.3	1284	1005	859	86.1
R-2013	REQUEJADA, LA	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	AGUEDA	788.4	895	1189	1333	392.7
R-2042	CASTRO DE LAS COGOTAS	853.3	1279	527	1231	89.0
R-2043	PONTON ALTO	150.4	1582	990	993	78.3
GS-3005	OSMA	893.1	1090	728	1084	121.0
GS-3016	PAJARES DE PEDRAZA	281.3	1298	634	1196	61.6
GS-3028	SALAS DE LOS INFANTES	353.5	1257	988	926	104.7
GS-3035	OTERO DE GUARDO	69.2	1492	1702	838	29.3
GS-3041	VILLALCAZAR DE SIRGA	307.7	929	664	1011	35.5
GS-3047	MEDIANA DE VOLTOYA	130.4	1347	518	1128	15.4
GS-3049	CABAÑES DE ESGUEVA	270.2	995	658	1046	24.7
GS-3051	ESPINAR, EL	36.7	1610	905	1032	18.8
GS-3057	VILLOVELA DE PIRON	202.0	1183	615	1212	33.0
GS-3089	MORLA DE LA VALDERIA	281.1	1369	1001	945	146.1
GS-3104	VILLAVERDE DE ARCAYOS	371.1	1146	1041	942	135.3
GS-3105	SANTERVAS DE CAMPOS	277.1	900	591	1107	28.8
GS-3124	MEDINA DE RIOSECO	908.4	790	450	1315	37.2
GS-3150	PARDAVE	223.8	1448	1356	787	220.8
GS-3818	RABAL	597.9	678	1328	1146	305.7

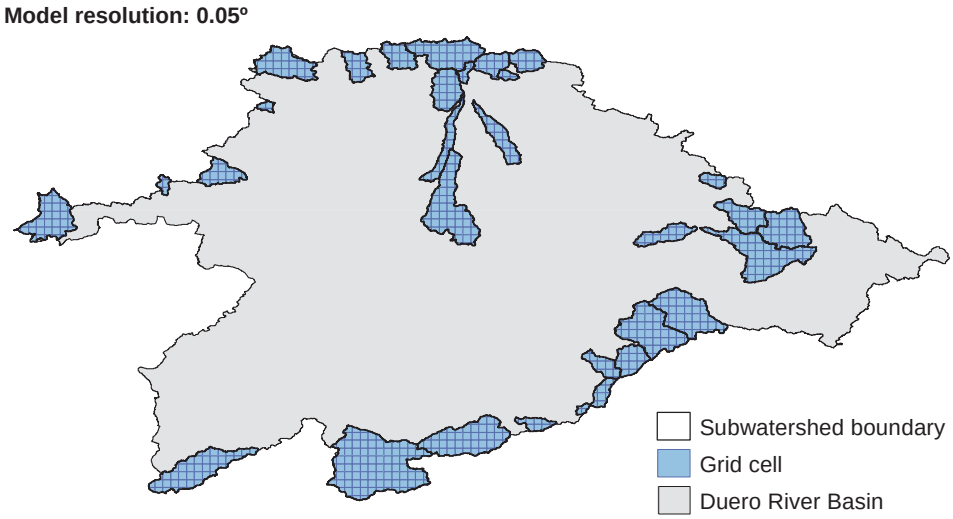

Fig. 2. VIC model implementation.

been chosen for this work since it is a modification of the two-layer VIC model (VIC-2L) originally developed to better represent the runoff generation process (Liang et al., 1996). Surface runoff is generated through an infiltration excess applying the Xinanjiang formulation (Zhao et al., 1980) to the upper two soil layers:

$$Q_d = \begin{cases} P - z_2 \cdot (\theta_S - \theta_2) + z_2 \cdot \theta_S \cdot \left(1 - \frac{i_0 + P}{i_m}\right)^{1+b_i}, & P + i_0 \le i_m \\ P - z_2 \cdot (\theta_S - \theta_2), & P + i_0 \ge i_m \end{cases} \quad (1)$$

For each time step Q_d [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:

$$Q_b = \begin{cases} \frac{D_S \cdot D_m}{W_S \cdot \theta_S} \cdot \theta_3, & 0 \le \theta_3 \le W_S \cdot \theta_S \\ \frac{D_S \cdot D_m}{W_S \cdot \theta_S} \cdot \theta_3 + \left(D_m - \frac{D_S \cdot D_m}{W_S}\right) \cdot \left(\frac{\theta_3 - W_S \cdot \theta_S}{\theta_S - W_S \cdot \theta_S}\right)^2, & \theta_3 \ge W_S \cdot \theta_S \end{cases} \quad (2)$$

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 characteristic large grid size for macroscale models makes that the VIC snow model conceptualizes the snow processes partitioning each grid cell of the spatial domain 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 class and the outputs consist of the snow depth and the snow water equivalent for each grid cell. 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, making the model suitable for applications in any part in the world (Liang et al., 1994, 1996).

3.2.2. Soil and vegetation parameters

The required soil parameters for the application of the VIC model were obtained from SoilGrids1km (Hengl et al., 2014) and EU-SoilHydroGrids ver1.0 (Tóth et al., 2017), all of them with a spatial resolution of 1 km × 1 km. 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 VIC model handles land uses information as a set of vegetation parameters for the different vegetation classes specified in a vegetation library. Here the UMD Global land cover classification (Hansen et al., 2000) was chosen with a spatial resolution of 1 km × 1 km, and the vegetation parameters (i.e. Leaf-Area Index, roots depth and roots coverage) were fixed for each vegetation class following the recommendations of the GLDAS Project (https://ldas.gsfc.nasa.gov/gldas/vegetation-parameters).

3.2.3. Aggregation method

The water balance mode of the VIC model allows to simulate the surface runoff and the baseflow for each grid cell of the spatial domain where the model is implemented. Since the conceptualization of the VIC model does not include the horizontal transport processes between contiguous grid cells, the runoff generated needs to be transported to a certain outlet in order to determine simulated values of streamflow that can be compared with existing observations. An approach consisting of a monthly aggregation of the total runoff (the sum of surface runoff and baseflow) for all the grid cells within a given subwatershed, weighted by the fractional area of each cell inside the subwatershed, was selected for this work. This approach has the advantage of working with the real boundaries of the subwatersheds and is not dependent on the resolution of the grid cells.

3.2.4. Model implementation

The version of the model used in this work was VIC 4.2.d (https://vic.readthedocs.io/en/vic.4.2.d/). The VIC model was implemented in the water balance mode at a daily time step and with a spatial resolution of 0.05° (~5 km × 5 km) for the 31 studied subwatersheds (Fig. 2). The meteorological forcings were interpolated to the grid cells of the spatial domain following a nearest neighbor assignment. The soil parameters and the elevation were averaged for each grid cell, and the vegetation parameters were kept at the original resolution of 1 km × 1 km because the VIC model allows to consider the sub-grid variability of the land uses for each grid cell. As in the soil database, the depth of each grid cell was set at 2 m, fixing the thickness of the first soil layer (d_1) at 0.1 m and varying the thicknesses of the second (d_2) and third (d_3) layers during the calibration and the sensitivity analysis.

The outputs of the VIC model were finally aggregated into monthly

Table 2
Selected parameters for the calibration.

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

values for each subwatershed in order to accomplish the calibration, model evaluation and sensitivity analysis.

3.3. Parameter calibration

The model was calibrated for the period from October 2000 to September 2009 choosing the Nash-Sutcliffe Efficiency (*NSE*, 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 model. Table 2 indicates the selected parameters for the calibration process together with their upper and lower bounds. These parameters were chosen following the recommendations for the calibration of the VIC model (https://vic.readthedocs.io/en/vic.4.2.d/Documentation/Calibration/). This selection is also in agreement with previous studies (e.g. Chawla and Mujumdar, 2015; Liang et al., 1996, 2004; Melsen et al., 2016b; Oubeidillah et al., 2014). The calibration was carried out 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.

3.4. 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 $alpha_d$ and $alpha_m$. *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. $alpha_d$ is the ratio between the standard deviation of the simulations and the standard deviation of the observations. $alpha_m$ 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. The model performance was compared to a benchmark performance based on the streamflow simulations carried out with the Integrated System for Rainfall-Runoff Modeling (SIMPA) model (Estrela and Quintas, 1996; Álvarez et al., 2004) for the domain of Spain. This model has been implemented and calibrated by CEDEX, and it is periodically updated and used among the different water districts of Spain as a tool for water planning and water resources management purposes (see https://www.miteco.gob.es/en/agua/temas/evaluacion-de-los-recursos-hidricos/evaluacion-recursos-hidricos-regimen-natural/ for more detail).

The spatial assessment of the VIC model streamflow simulations was done through a cross-evaluation exercise that aimed at analyzing the spatial transferability of the five soil parameters chosen for the calibration. The VIC model was run for each subwatershed using the calibrated parameters of the remaining 30 subwatersheds, 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 subwatershed with two *AET* products: the Global Land Evaporation Amsterdam Model (GLEAM) version 3.3a (Martens et al., 2017; Miralles et al., 2011), and the WRFERA simulation for the Iberian Peninsula provided in García-Valdecasas Ojeda et al. (2017). GLEAM comprises a set of algorithms for the estimation of land evaporation using satellite information and reanalysis data. GLEAM v3.3a presents several improvements with respect to its predecessors, such as the use of ERA5 data instead of ERA-Interim, and provides a global dataset spanning the period 1980–2014 with a spatial-resolution of 0.25° × 0.25°. The WRFERA simulation for the Iberian Peninsula constitutes a suitable tool for studying the *AET* behavior, and its performance was recently evaluated showing a good ability to represent the land-surface processes in this region (García-Valdecasas Ojeda et al., 2020). 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.

3.5. 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 behavior 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 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 subwatershed 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.

3.5.1. 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 choosing the five calibration parameters and considering the upper a lower bounds specified in Table 2. 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 10,000 random samples. This process allowed to define a sampling matrix, $\mathbf{\Theta}$, of order $m \times n$, where m represents the number of samples ($m = 10{,}000$) 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.

3.5.2. Multiple linear regression

The outputs of interest from the VIC model were those components included in the water balance: surface runoff (Q_d), 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:

$$\boldsymbol{y} = a_0 \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \sum_{i=1}^{n} a_i \cdot \boldsymbol{\Theta}_i \tag{3}$$

where $\boldsymbol{y}$ 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 $\boldsymbol{\Theta}_i$ is the column of the sampling matrix corresponding to the parameter i. The standardized regression coefficients, β_i, are then calculated for each parameter:

$$\beta_i = \frac{\sigma_{\boldsymbol{\Theta}_i}}{\sigma_{y_p}} \cdot a_i \tag{4}$$

Here, $\sigma_{\boldsymbol{\Theta}i}$ and σ_{yp} are the standard deviations of $\boldsymbol{\Theta}_i$ and the predicted values of $\boldsymbol{y}$, respectively. β_i^2 represents the relative contribution of the parameter i to the variance of the model output of interest, being $\sum_{i=1}^{n} (\beta_i^2) \leq 1$ 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 behavior, 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. Results

4.1. Calibration results

The values of the four skill metrics for the calibration and validation periods are shown in Table 3. Fig. 3 depicts the simulated streamflows during both periods together with the observed streamflows for six selected subwatersheds (R-2011, R-2037, R-2038, GS-3005, GS-3089 and GS-3150) located in different parts of the basin. The *NSE* for the calibration period presents values above 0.75 in 19 out of the 31 subwatersheds and reaches values above 0.85 in 10 subwatersheds, 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 corresponding ones for the calibration process, although minimum values of *NSE* below 0 were attained for 3 subwatersheds. 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 3, Fig. 3). For instance, stations R-2012 and GS-3035 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 $alpha_d$ and $alpha_m$ in such cases. The values of $alpha_d$ and $alpha_m$ are below 0.4 for station R-2012, and this evidences a clear underestimation of both the variability of the streamflow (i.e. low $alpha_d$) and its total volume during the period (i.e. low $alpha_m$). In the case of station GS-3035, the variability of the simulated streamflow is almost identical to the observations ($alpha_d \sim 1$), while the streamflow volume is greatly overestimated ($alpha_m \sim 2$). It is interesting to note that high *NSE* values were obtained for subwatersheds with varying sizes, with good fits for both small-sized (e.g. R-2011 and GS-3089) and medium-sized (e.g. R-2038 and GS-3005) subwatersheds, emphasizing the ability of the VIC model to provide accurate predictions of the streamflow across different spatial scales.

Fig. 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. 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 $alpha_d$ and $alpha_m$ (Fig. 4c, d) show that most of the subwatersheds 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 $alpha_d$ values above 1.5. The overestimation is also evident for the streamflow volume, and in this case the presence of values of $alpha_m$ above 1.5 is noticeable both for VIC and SIMPA.

4.2. Cross-evaluation of the calibrated parameters

Fig. 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 subwatersheds and values between 0.6 and 0.8 in 12 of them. There is a spatial pattern for the optimal *NSE* distribution, and most of the lowest values were obtained for the subwatersheds located in the headwater areas of the central depression of the basin, being mainly associated to gauging stations (e.g. GS-3041 and GS-3105).

Fig. 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 subwatershed (i.e. *NSE* bias). The results can be interpreted in two different directions: firstly, each row reflects the degree of receptivity of a given subwatershed to the calibrated parameters obtained for every subwatershed including itself. The presence of darker blue tones indicates a greater parameter receptivity for the target subwatershed (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 subwatersheds 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 subwatershed 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 subwatersheds 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 subwatersheds with both high optimal *NSE* estimations (e.g. R-2011 and GS-3089) and low optimal *NSE* values (e.g. R-2012 and GS-3057). The lowest degrees of receptivity followed a similar behavior but were much less abundant (e.g. R-2001 and GS-3105), 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. 1).

Table 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-3005	0.8307	0.9156	1.0006	1.0305	0.5003	0.8161	0.6139	0.6316
GS-3016	0.8350	0.9173	0.8664	1.0787	0.7798	0.9126	0.7777	0.8281
GS-3028	0.6290	0.8320	1.0762	1.0830	0.6980	0.8465	0.7543	0.8810
GS-3035	0.5047	0.8850	0.9824	2.1179	−2.1931	0.7974	1.4392	3.1461
GS-3041	0.7106	0.8497	0.9556	1.0304	0.1784	0.8344	0.3184	0.3503
GS-3047	0.6277	0.8148	0.9591	1.2317	0.4813	0.7479	1.0270	1.0153
GS-3049	0.7333	0.8738	1.0208	1.1350	−0.0513	0.8926	1.5098	1.5572
GS-3051	0.7743	0.8909	0.7924	0.8814	0.7591	0.9226	0.6616	0.8019
GS-3057	0.7230	0.8864	1.1062	1.1746	0.6515	0.9176	0.7842	0.6004
GS-3089	0.9116	0.9564	1.0038	1.0377	0.9026	0.9561	0.8707	0.9415
GS-3104	0.8137	0.9208	1.0955	1.0734	0.6694	0.8985	0.5954	0.7873
GS-3105	0.6204	0.8032	0.9603	1.0057	0.3723	0.8010	0.5803	0.6466
GS-3124	0.7397	0.8604	0.8403	0.9855	0.3129	0.8246	1.3097	0.8383
GS-3150	0.7909	0.9118	0.8582	0.7828	0.8507	0.9436	1.1388	0.9702
GS-3818	0.8724	0.9403	1.0319	1.0953	0.8975	0.9720	0.8782	0.7920

4.3. Model performance for the AET simulations

Fig. 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. 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.

A comparable behavior 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. 6b). The SIMPA model produces *r* values slightly higher when compared against GLEAM and hardly reaches a value of 0.8 for very few subwatersheds.

Lastly, there is a generalized underestimation of the *AET* variability and the total volume for both model structures, with $alpha_d$ and $alpha_m$ values predominantly below 1 and markedly slower for SIMPA (Fig. 6c, d). The wide distance between the CDFs corresponding to $alpha_d$ 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 $alpha_m$, 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. 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. 7 (a to g) for all the subwatersheds. 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. 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_d: the values of r^2 are above 0.7 for all the subwatersheds (Fig. 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. (1), where a relation of exponential type between Q_d 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_d, and this effect is related to an increase in *AET*.
- Q_b: the values of r^2 mostly range from 0.5 to 0.7, with some values close to 0.8 (Fig. 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

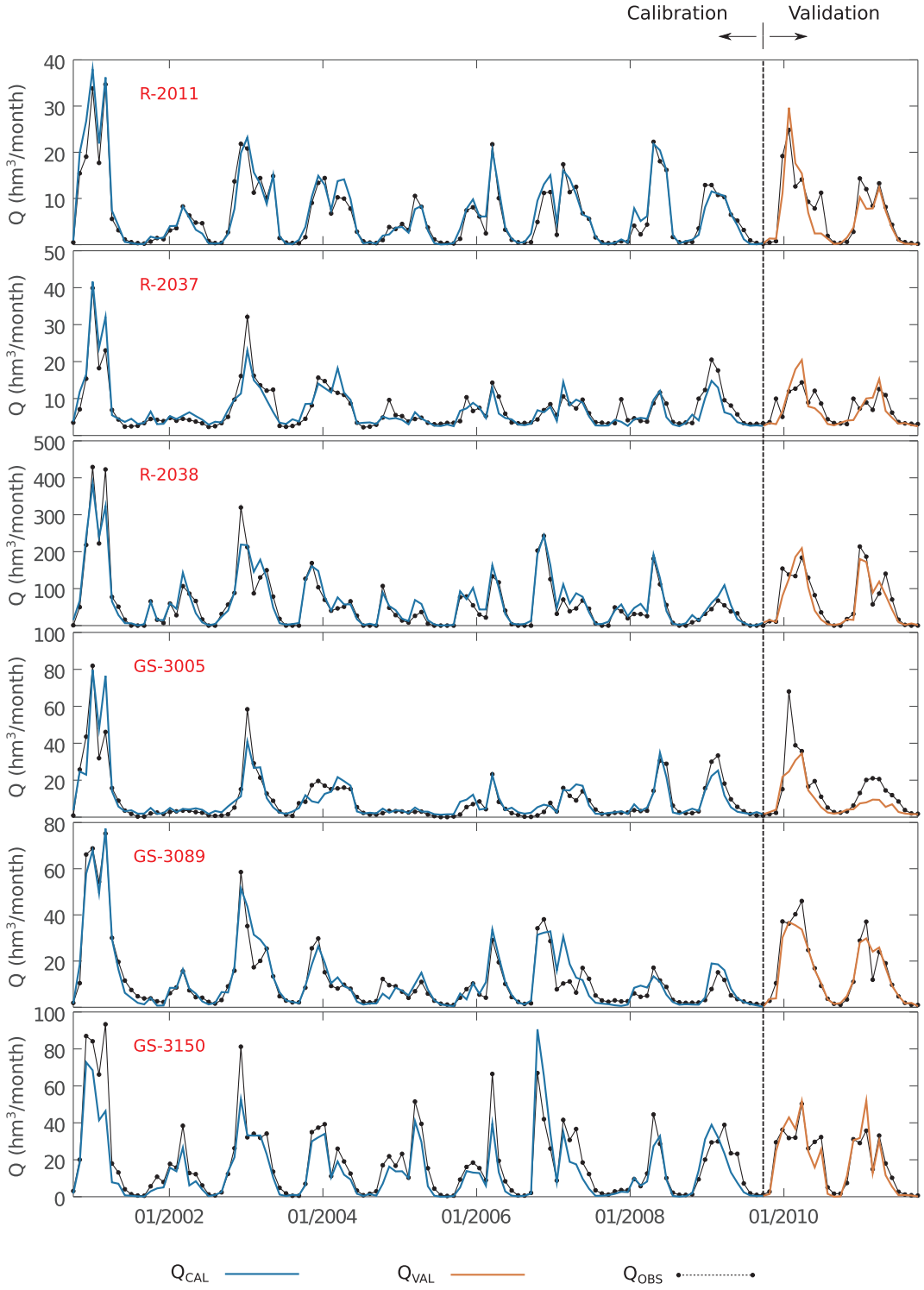

Fig. 3. Time series of the observed streamflows along with the simulated ones for the calibration and de validation periods for six example subwatersheds.

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. (2)). The amplitude of the β coefficients for d_2 is broader than for the rest of the parameters but always negative except for two subwatersheds.

- 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. 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 behavior the spaghetti plots of the Monte Carlo simulation for an example subwatershed (R-2038) are depicted in Fig. 8. As shown there, the time series of the observed streamflow (Fig. 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

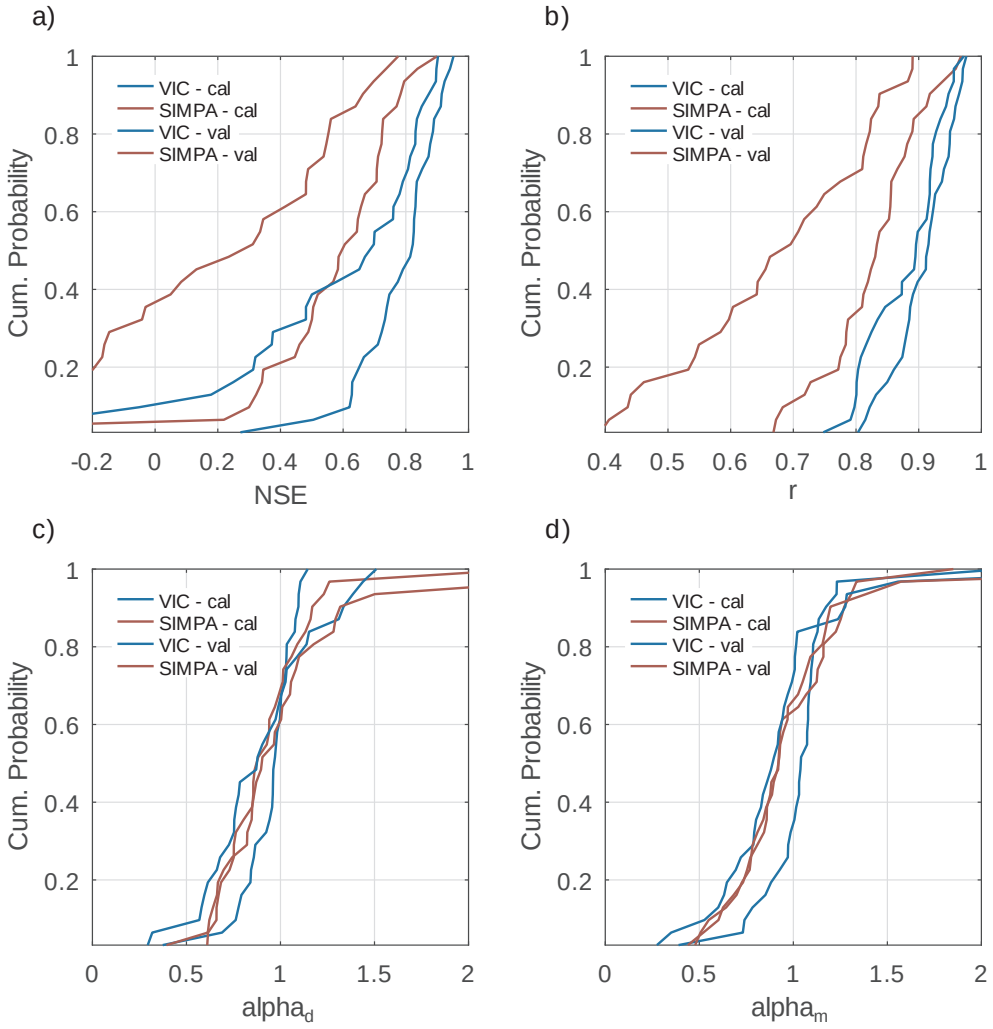

Fig. 4. CDFs of the four skill metrics for the streamflow simulations: a) *NSE*; b) *r*; c) $alpha_d$, and d) $alpha_m$. Blue lines represent the VIC model performance and red lines represent the SIMPA model (benchmark) performance. Straight lines correspond to the calibration period and dotted lines correspond to the validation period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subwatershed, 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 subwatershed 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. 7). Fig. 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. 8d corresponds to the current vegetation parameters.

- SM_1: the values of r^2 are widely scattered and range from 0.35 to values above 0.9 (Fig. 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. 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 subwatersheds (Fig. 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. 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. 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. 8g).

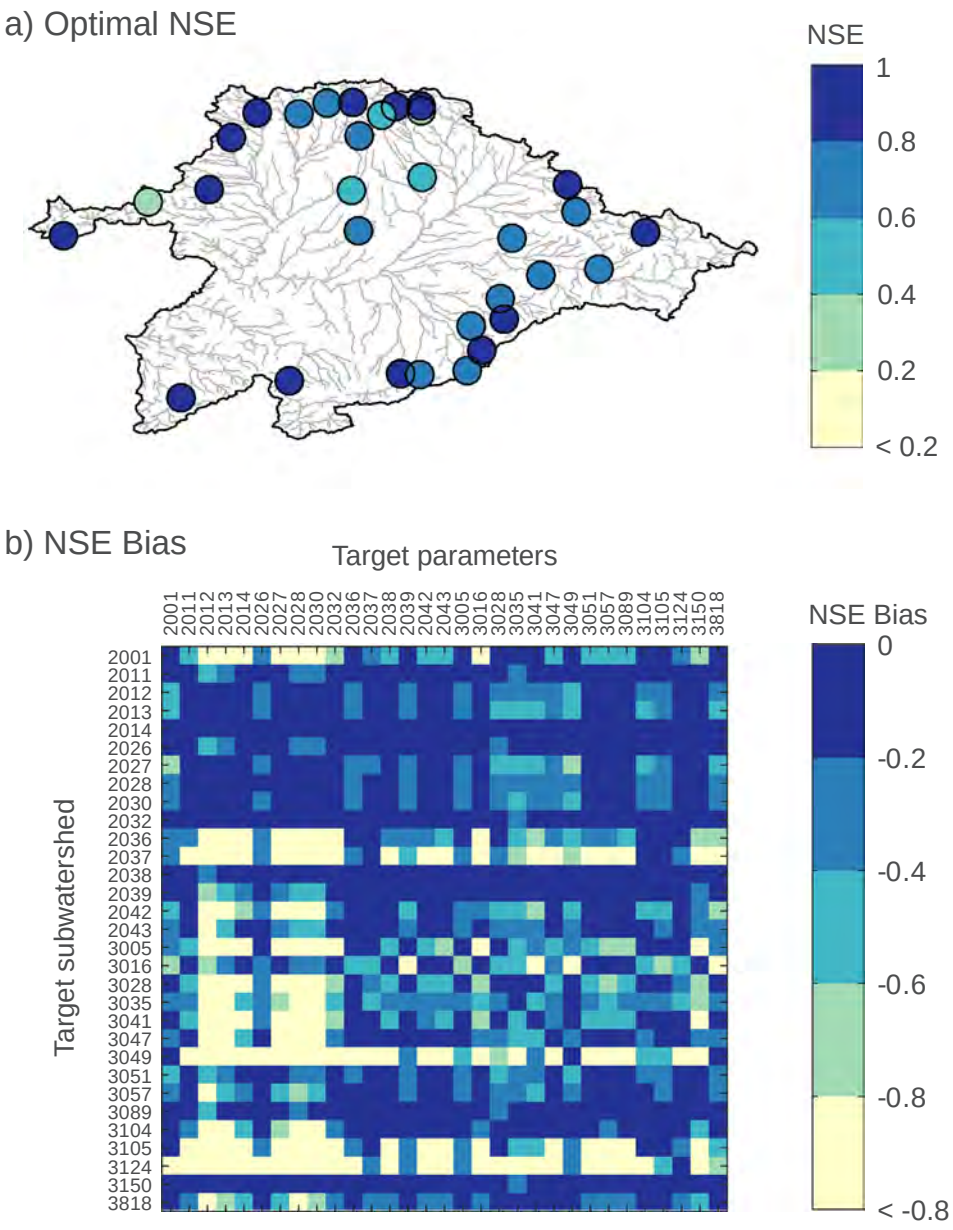

Fig. 5. Cross-evaluation of the calibrated parameters for the 31 subwatersheds: a) spatial distribution of the optimal *NSE* values, and b) cross-performance of the different parameter combinations for each subwatershed.

4.5. Assessing equifinality and the 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 subwatersheds 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 subwatershed 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.

It is clear that for the majority of the subwatersheds 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. 9 shows the dotty plots of the five parameters of the calibration for two subwatersheds (R-2038 and GS-3089) as an example of this analysis. For both subwatersheds 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). The shape of the dotty plots supplies useful information about the behavior of the parameter samples as a set. For example, the dotty plot of d_2 for the subwatershed 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.

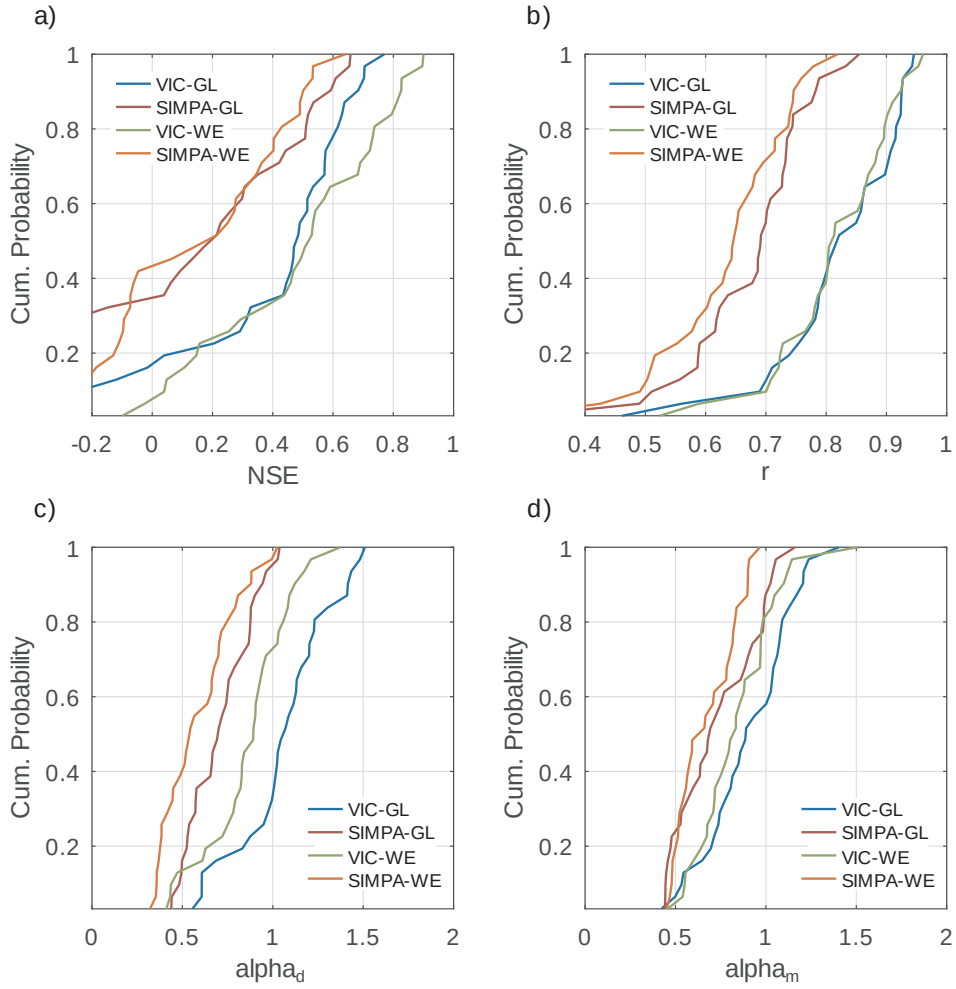

Fig. 6. CDFs of the four skill metrics for the *AET* simulations: a) *NSE*; b) *r*; c) $alpha_d$, and d) $alpha_m$. Blue and green lines represent the VIC model performance against GLEAM and WRFERA, respectively. Red and orange lines represent the SIMPA model (benchmark) performance against GLEAM and WRFERA, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Discussion

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 study performs well for a large number of subwatersheds 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 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 calibrating the VIC model and evaluating its 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 calculated in this work (Fig. 4) showed higher *NSE* and *r* values than those obtained for the subwatersheds of the CONUS domain, as well as $alpha_d$ and $alpha_m$ 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 (see Table 3 there in). Moreover, while the CDFs of our work were calculated for 31 subwatersheds, 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

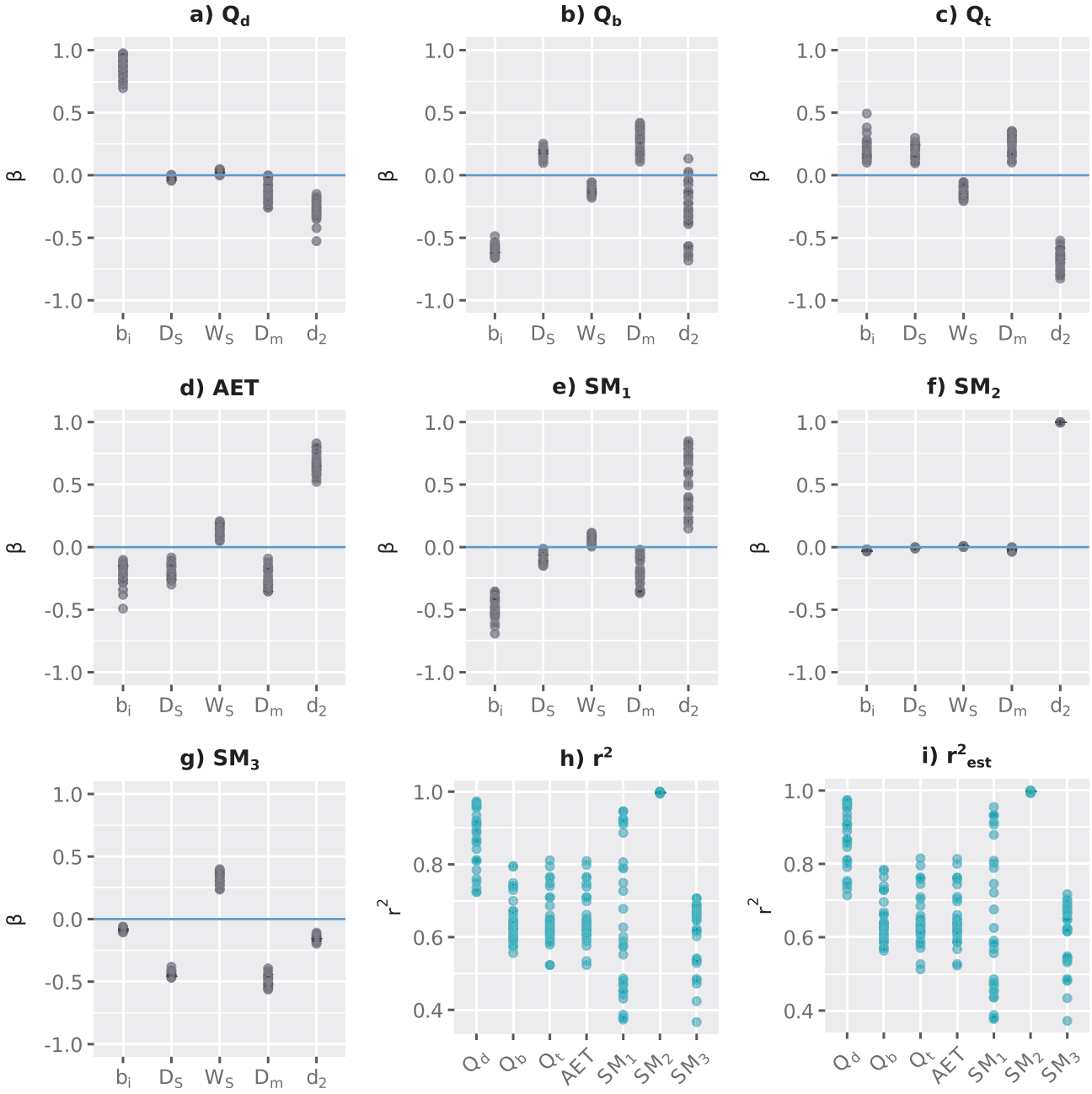

Fig. 7. (a to g) β coefficients for the 31 subwatersheds. h) r^2 value from the multiple linear regression. i) r^2 estimated from the β coefficients.

streamflow with the observations for all the studied subwatersheds. 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 subwatersheds.

Finally, the aggregation method has proven to be accurate and efficient in this work, 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 subwatershed size increases, but it is expected that for subwatersheds 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).

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 subwatershed with each parameter combination. Many subwatersheds 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 subwatersheds where the Monte Carlo analysis showed a lesser number of simulations close to the optimum (see Table 4 and compare with Fig. 5b). Meanwhile, the parameter combinations reflecting the highest *NSE* biases were found for various subwatersheds located in the north of the basin. This latitudinal gradient requires further research in order to study the main drivers of its behavior (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

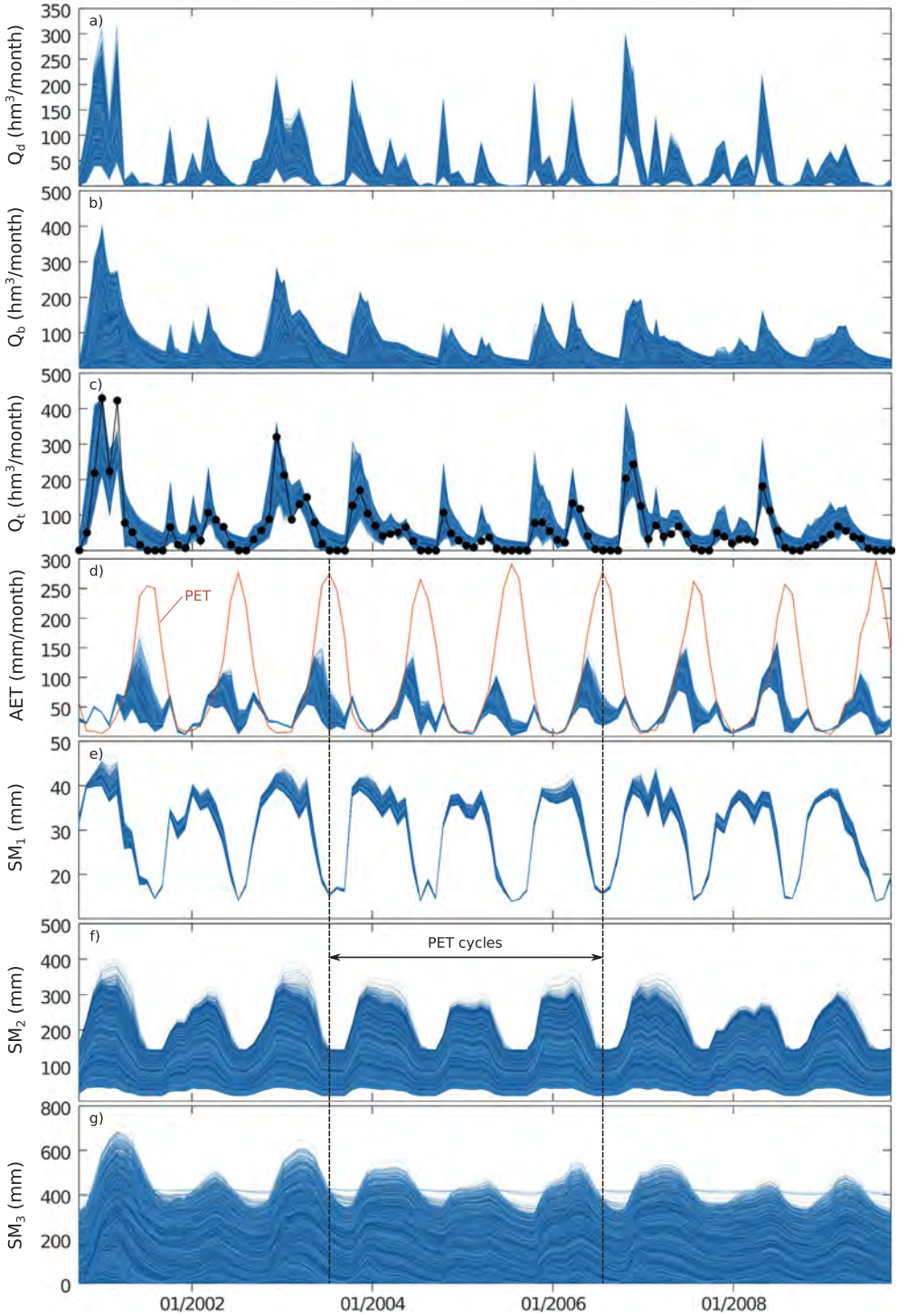

Fig. 8. Spaghetti plots of the water balance components resulting from the Monte Carlo simulation for the subwatershed R-2038.

streamflow. Since the streamflow was the only variable subject to calibration in this work, 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 $alpha_m$ 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., 2020).

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

Table 4
Behavior of *NSE* in the Monte Carlo simulations for assessing equifinality and the efficiency of the calibration algorithm.

Code	NSE_{cal}	Number of simulations with $NSE > NSE_{cal} - 0.05$	Number of simulations with $NSE > NSE_{cal}$
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-3005	0.8307	32	5
GS-3016	0.8350	1758	0
GS-3028	0.6290	103	0
GS-3035	0.5047	373	0
GS-3041	0.7106	0	0
GS-3047	0.6277	267	0
GS-3049	0.7333	0	0
GS-3051	0.7743	2446	0
GS-3057	0.7230	518	3
GS-3089	0.9116	3578	3
GS-3104	0.8137	93	0
GS-3105	0.6204	8	0
GS-3124	0.7397	1	0
GS-3150	0.7909	3484	0
GS-3818	0.8724	26	0

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_d 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 subwatersheds.

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 subwatersheds, 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.

6. Conclusions

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

[1] The VIC performance for the streamflow simulations reflected good results for most of the studied subwatersheds 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 modeling 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.

[2] 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 subwatersheds 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.

[3] 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 behavior 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.

[4] 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 subwatersheds. 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.

CRediT authorship contribution statement

Patricio Yeste: Conceptualization, Methodology, Software, Validation, Writing - original draft. **Matilde García-Valdecasas Ojeda:** Investigation, Data curation. **Sonia R. Gámiz-Fortis:** Writing - review & editing, Supervision. **Yolanda Castro-Díez:** Writing - review & editing, Supervision. **María Jesús Esteban-Parra:** Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

All the simulations were conducted in the ALHAMBRA cluster

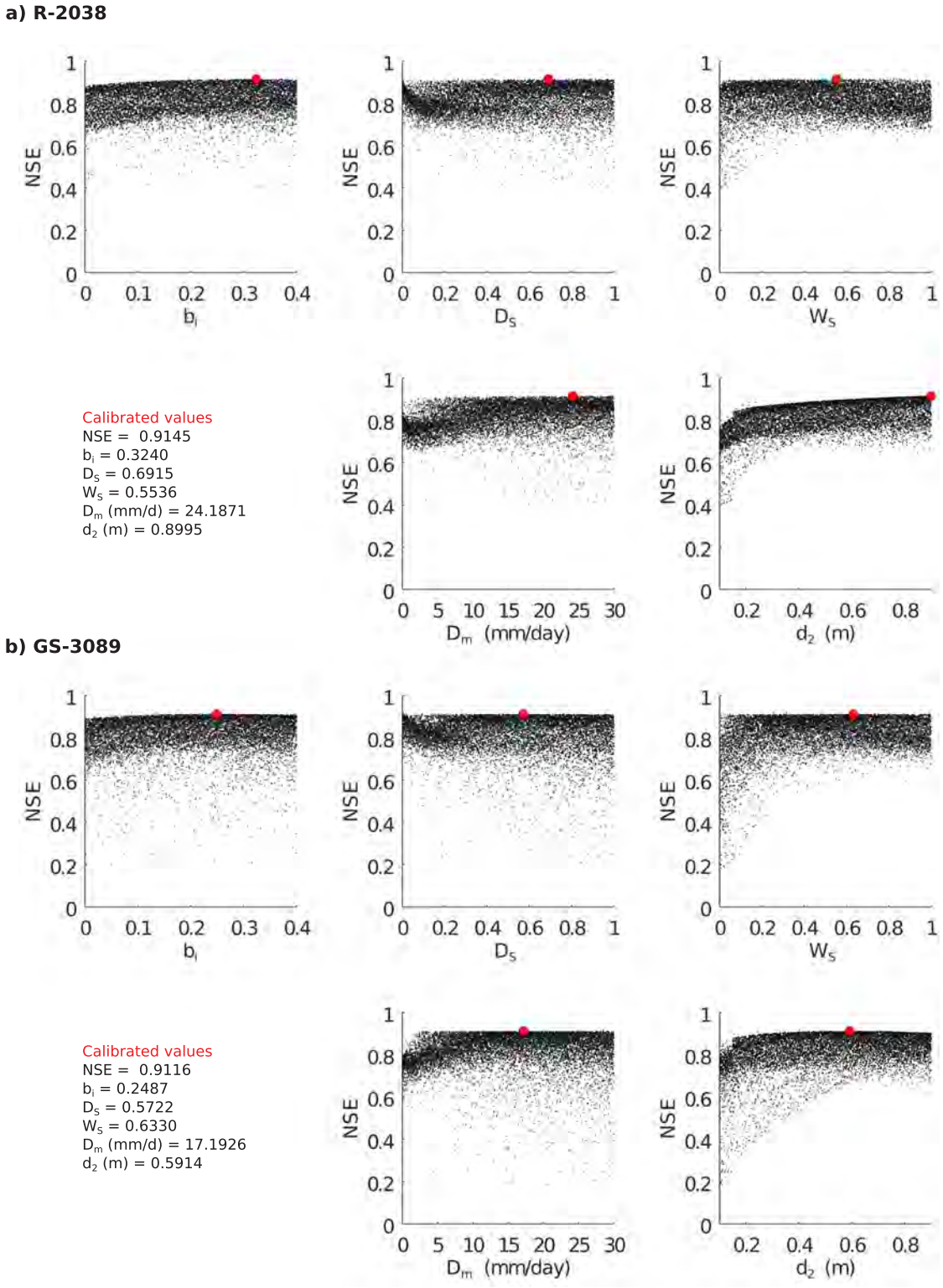

Fig. 9. Dotty plots for two subwatersheds: a) R-2038 and b) GS-3089. Red dot corresponds to the calibrated value for the corresponding parameter using the SCE-UA algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(http://alhambra.ugr.es/) of the University of Granada. This work was partially funded by the Spanish Ministry of Economy and Competitiveness projects CGL2013-48539-R and CGL2017-89836-R, with additional support from the European Community Funds (FEDER) and by FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento/B-RNM-336-UGR18 project. The first author was supported by the Ministry of Education, Culture and Sport of Spain (FPU grant FPU17/02098). We thank two anonymous referees whose comments and indications improved the paper significantly.

References

Álvarez, J., Sánchez, A., Quinta, L., 2004. SIMPA, a GRASS based tool for Hydrological Studies. Proc. FOSS/GRASS Users Conf.

Bennett, K.E., Urrego Blanco, J.R., Jonko, A., Bohn, T.J., Atchley, A.L., Urban, N.M., Middleton, R.S., 2018. Global sensitivity of simulated water balance indicators under future climate change in the Colorado Basin. Water Resour. Res. 54, 132–149. https://doi.org/10.1002/2017WR020471.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Beven, K., 2012. Rainfall-Runoff Modelling: The Primer: Second Edition. Rainfall-Runoff Model. Prim. Second Ed. https://doi.org/10.1002/9781119951001.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

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

Blöschl, G., Bierkens, M.F.P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H.G., 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., Chen, Y., Chifflard, P., Claps, P., Clark, M.P., Collins, A.L., Croke, B., Dathe, A., David, P.C., de Barros, F.P.J., 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.R., 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.M., 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., Zhang, Y., 2019. Twenty-three unsolved problems in hydrology (UPH)–a community perspective. Hydrol. Sci. J. 64, 1141–1158. https://doi.org/10.1080/02626667.2019.1620507.

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

Chaney, N.W., Herman, J.D., Reed, P.M., Wood, E.F., 2015. Flood and drought hydrologic monitoring: The role of model parameter uncertainty. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-19-3239-2015.

Chavez-Jimenez, A., Lama, B., Garrote, L., Martin-Carrasco, F., Sordo-Ward, A., Mediero, L., 2013. Characterisation of the sensitivity of water resources systems to climate change. Water Resour. Manage. 27, 4237–4258. https://doi.org/10.1007/s11269-013-0404-2.

Chawla, I., Mujumdar, P.P., 2015. Isolating the impacts of land use and climate change on streamflow. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-19-3633-2015.

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

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., Rasmussen, R.M., 2015. A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resour. Res. https://doi.org/10.1002/2015WR017198.

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., Tarpley, J.D., 2003. Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). J. Geophys. Res. Atmos. https://doi.org/10.1029/2002jd003316.

Demaria, E.M., Nijssen, B., Wagener, T., 2007. Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model. J. Geophys. Res. Atmos. https://doi.org/10.1029/2006JD007534.

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

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. https://doi.org/10.1016/0022-1694(94)90057-4.

Estrela, T., Quintas, L., 1996. El sistema integrado de modelización precipitación – aportación SIMPA. Ing. Civ. 104, 43–52.

Franchini, M., Pacciani, M., 1991. Comparative analysis of several conceptual rainfall-runoff models. J. Hydrol. https://doi.org/10.1016/0022-1694(91)90178-K.

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

García-Valdecasas Ojeda, M., Gámiz-Fortis, S.R., Castro-Díez, Y., Esteban-Parra, M.J., 2017. Evaluation of WRF capability to detect dry and wet periods in Spain using drought indices. J. Geophys. Res. https://doi.org/10.1002/2016JD025683.

García-Valdecasas Ojeda, M., Rosa-Cánovas, J.J., Romero-Jiménez, E., Yeste, P., Gámiz-Fortis, S.R., Castro-Díez, Y., Esteban-Parra, M.J., 2020. The role of the surface evapotranspiration in regional climate modelling: Evaluation and near-term future changes. Atmos. Res. 237, 104867. https://doi.org/10.1016/j.atmosres.2020.104867.

Garrote, L., Granados, A., Iglesias, A., 2016. Strategies to reduce water stress in Euro-Mediterranean river basins. Sci. Total Environ. 543, 997–1009. https://doi.org/10.1016/j.scitotenv.2015.04.106.

Haddeland, I., Matheussen, B.V., Lettenmaier, D.P., 2002. Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model. Water Resour. Res. https://doi.org/10.1029/2001wr000854.

Hansen, M.C., Sohlberg, R., Defries, R.S., Townshend, J.R.G., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens. https://doi.org/10.1080/014311600210209.

Hengl, T., De Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km - global soil information based on automated mapping. PLoS One. https://doi.org/10.1371/journal.pone.0105992.

Iman, R., Conover, W.J., 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. Commun. Stat. - Simul. Comput. https://doi.org/10.1080/03610918208812265.

IPCC: Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part B: Regional Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report, vol. 2, Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9781107415386.

Li, H., Luo, L., Wood, E.F., Schaake, J., 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. J. Geophys. Res. Atmos. https://doi.org/10.1029/2008JD010969.

Liang, X., Guo, J., Leung, L.R., 2004. Assessment of the effects of spatial resolutions on daily water flux simulations. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2003.07.007.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. https://doi.org/10.1029/94JD00483.

Liang, X., Wood, E.F., Lettenmaier, D.P., 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. Glob. Planet. Change. https://doi.org/10.1016/0921-8181(95)00046-1.

Lorenzo-Lacruz, J., Morán-Tejeda, E., Vicente-Serrano, S.M., López-Moreno, J.I., 2013. Streamflow droughts in the Iberian Peninsula between 1945 and 2005: Spatial and temporal patterns. Hydrol. Earth Syst. Sci. 17, 119–134. https://doi.org/10.5194/hess-17-119-2013.

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

Martens, B., Miralles, D.G., Lievens, H., Van Der Schalie, R., De Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.

Melsen, L.A., Teuling, J.A., Torfs, J.J.F.P., Uijlenhoet, R., Mizukami, N., Clark, P.M., 2016a. HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-20-1069-2016.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., Uijlenhoet, R., 2016b. Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-20-2207-2016.

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15, 453–469. https://doi.org/10.5194/hess-15-453-2011.

Mizukami, N., Clark, M.P., Newman, A.J., Wood, A.W., Gutmann, E.D., Nijssen, B., Rakovec, O., Samaniego, L., 2017. Towards seamless large-domain parameter estimation for hydrologic models. Water Resour. Res. 53, 8020–8040. https://doi.org/10.1002/2017WR020401.

Morán-Tejeda, E., Ceballos-Barbancho, A., Llorente-Pinto, J.M., 2010. Hydrological response of Mediterranean headwaters to climate oscillations and land-cover changes: the mountains of Duero River basin (Central Spain). Glob. Planet. Change 72, 39–49. https://doi.org/10.1016/j.gloplacha.2010.03.003.

Morán-Tejeda, E., Ceballos-Barbancho, A., Llorente-Pinto, J.M., López-Moreno, J.I., 2012a. Land-cover changes and recent hydrological evolution in the Duero Basin (Spain). Reg. Environ. Change. 12, 17–33. https://doi.org/10.1007/s10113-011-0236-7.

Morán-Tejeda, E., Ignacio, L.M., Antonio, C.B., Sergio, M., 2011a. Evaluating Duero's basin (Spain) response to the NAO phases: spatial and seasonal variability. Hydrol. Process. 25, 1313–1326. https://doi.org/10.1002/hyp.7907.

Morán-Tejeda, E., López-Moreno, J.I., Ceballos-Barbancho, A., Vicente-Serrano, S.M., 2011b. River regimes and recent hydrological changes in the Duero basin (Spain). J. Hydrol. 404, 241–258. https://doi.org/10.1016/j.jhydrol.2011.04.034.

Morán-Tejeda, E., Lorenzo-Lacruz, J., López-Moreno, J.I., Ceballos-Barbancho, A., Zabalza, J., Vicente-Serrano, S.M., 2012b. Reservoir management in the Duero Basin

(Spain): impact on river regimes and the response to environmental change. Water Resour. Manage. 26, 2125–2146. https://doi.org/10.1007/s11269-012-0004-6.
Morán-Tejeda, E., Lorenzo-Lacruz, J., López-Moreno, J.I., Rahman, K., Beniston, M., 2014. Streamflow timing of mountain rivers in Spain: recent changes and future projections. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2014.06.053.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models Part I: a discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Nijssen, B., O'donnell, G.M., Hamlet, A.F., Lettenmaier, D.P., 2001. Hydrologic sensitivity of global rivers to climate change. Clim. Change 50, 143–175. https://doi.org/10.1023/A:1010616428763.
Oubeidillah, A.A., Kao, S.C., Ashfaq, M., Naz, B.S., Tootle, G., 2014. A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-18-67-2014.
Pellicer-Martínez, F., Martínez-Paz, J.M., 2018. Climate change effects on the hydrology of the headwaters of the Tagus River: Implications for the management of the Tagus-Segura transfer. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-22-6473-2018.
Rakovec, O., Mizukami, N., Kumar, R., Newman, A.J., Thober, S., Wood, A.W., Clark, M.P., Samaniego, L., 2019. Diagnostic evaluation of large-domain hydrologic models calibrated across the contiguous United States. J. Geophys. Res. Atmos. 124, 13991–14007. https://doi.org/10.1029/2019JD030767.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis. The Primer. Glob. Sensit. Anal. Prim. https://doi.org/10.1002/9780470725184.
Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2006. Sensitivity analysis practices: Strategies for model-based inference. Reliab. Eng. Syst. Saf. https://doi.org/10.1016/j.ress.2005.11.014.
Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46, 1–25. https://doi.org/10.1029/2008WR007327.
Savenije, H.H.G., 2009. HESS opinions: "The art of hydrology". Hydrol. Earth Syst. Sci. 13, 157–161. https://doi.org/10.5194/hess-13-157-2009.
Serrano-Notivoli, R., Beguería, S., Saz, M.Á., Longares, L.A., De Luis, M., 2017. SPREAD: A high-resolution daily gridded precipitation dataset for Spain - An extreme events frequency and intensity overview. Earth Syst. Sci. Data. https://doi.org/10.5194/essd-9-721-2017.
Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. J. Hydrol. 523, 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013.
Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. Hydrol. Process. https://doi.org/10.1002/hyp.11203.
Vano, J.A., Nijssen, B., Lettenmaier, D.P., 2015. Seasonal hydrologic responses to climate change in the Pacific Northwest. Water Resour. Res. 51, 1959–1976. https://doi.org/10.1002/2014WR015909.
VanShaar, J.R., Haddeland, I., Lettenmaier, D.P., 2002. Effects of land-cover changes on the hydrological response of interior Columbia River basin forested catchments. Hydrol. Process. https://doi.org/10.1002/hyp.1017.
Whitfield, P.H., Burn, D.H., Hannaford, J., Higgins, H., Hodgkins, G.A., Marsh, T., Looser, U., 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. Hydrol. Sci. J. https://doi.org/10.1080/02626667.2012.728706.
Wood, A.W., Lettenmaier, D.P., 2008. An ensemble approach for attribution of hydrologic prediction uncertainty. Geophys. Res. Lett. https://doi.org/10.1029/2008GL034648.
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., Whitehead, P., 2011. Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res. 47. https://doi.org/10.1029/2010WR010090.
Yang, Y., Pan, M., Beck, H.E., Fisher, C.K., Beighley, R.E., Kao, S.C., Hong, Y., Wood, E.F., 2019. In quest of calibration density and consistency in hydrologic modeling: distributed parameter calibration against streamflow characteristics. Water Resour. Res. 55, 7784–7803. https://doi.org/10.1029/2018WR024178.
Yeste, P., Dorador, J., Martin-Rosales, W., Molero, E., Esteban-Parra, M.J., Rueda, F.J., 2018. Climate-driven trends in the streamflow records of a reference hydrologic network in Southern Spain. J. Hydrol. 566, 55–72. https://doi.org/10.1016/j.jhydrol.2018.08.063.
Zhao, R.J., Zhuang, Y.L., Fang, L.R., Liu, X.R., Zhang, Q.S., 1980. The Xinanjiang Model, Hydrological Forecasting. Oxford Symp. IAHS129.