Journal of Hydrology 610 (2022) 127816

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Hydrograph separation for tackling equifinality in conceptual hydrological models

Jesús Casado-Rodríguez, Manuel del Jesus *

IHCantabria – Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain

ARTICLE INFO

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

Keywords:
Hydrological modeling
Calibration
Equifinality
Hydrograph separation

ABSTRACT

Attributing a physical meaning to the calibration of a conceptual hydrological model is a risk due to equifinality, i.e., the existence of multiple optimal parameterizations that might or might not represent the actual behavior of a catchment; a risk that propagates to posterior studies that use the outputs of the hydrological model as an input. This study proposes and analyses sequential procedures for calibrating conceptual hydrological models aimed at reducing equifinality. These procedures force the model to reproduce a flow separation of the observed streamflow into quick flow and base flow, which we assumed representative of the run-off generating processes in the catchment. The sequential calibration of the model parameters that control quick flow and base flow forces the model to reproduce the flow separation, and introduces additional constraints in the calibration process that reduce equifinality and improve the overall calibration procedure. We applied this procedure to two mesoscale catchments in the "Picos de Europa" National Park (northern Spain). We compared the performance of the different calibration procedures both in the real scenario and in hypothetical scenarios of land use and soil permeability, to provide a sounder assessment of the ability of the procedures under diverse conditions. Results show that a calibration method applying hydrograph separation ensures models with a better discharge partition, whereas methods that do not apply separation failed in a considerable number of cases. In terms of performance (NSE and bias), the method applying hydrograph separation outperformed the reference method (without separation) for the real scenario, even for total streamflow; in the hypothetical scenarios though, the improvement in process representativeness came at the expense of a slight loss in performance. The sequential methods here developed were more computationally efficient; since they explore the parameter space in subsets, the number of iterations until convergence was a third of that needed with the reference method. In summary, we have developed a simple calibration procedure that ensures a better model behavior (more in line with the underlying conceptualization) with a similar, and even better, performance and a shorter calibration time than the reference method.

1. Introduction

Reproducing the real functioning of a catchment by calibration of a conceptual hydrological model is an extremely complicated task due to the multidimensionality of the optimisation problem (Shokri et al., 2018). The equifinality of a model is the impossibility to select a single best model parameterization from the set of parameterizations with optimal performance (Beven and Freer, 2001). These multiple optimal parameterizations might represent models that behave in distinct ways, i.e., in which the processes generating discharge (overland flow, interflow and base flow) are differently partitioned (Beven and Binely, 1992). The usual calibration procedure, either using manual or automatic calibration, would choose the model parameterization that maximizes a specific performance metric. However, the parameter set with the highest performance might not reproduce the actual catchment functioning; instead, sub-optimal parameterizations might be more realistic (Kavetski, 2018).

Equifinality is partly caused by the large number of parameters present in conceptual hydrological models (even though physical processes-based models may contain even more parameters) and the lack of observations against which to compare the model outputs (Shokri et al., 2018), which in a regular case in hydrology is the observed discharge in one (or several) gauging station. It may even be exacerbated -in conceptual models- due to the limited physical base of some of the

* Corresponding author.
E-mail addresses: jesus.casado@alumnos.unican.es (J. Casado-Rodríguez), manuel.deljesus@unican.es (M. del Jesus).

https://doi.org/10.1016/j.jhydrol.2022.127816
Received 15 November 2020; Received in revised form 1 March 2022; Accepted 8 April 2022
Available online 18 April 2022

processes considered (for instance the separation between overland flow, interflow and baseflow). The uncertainty in the representativeness of the optimized model prevents granting it a physical meaning, and thus its application as a starting point for further studies, such as erosion, vegetation dynamics or nutrient transport, is compromised (Acero Triana et al., 2019; Cain et al., 2019).

A way to tackle equifinality is to increase the number of target variables during calibration, forcing the model to properly reproduce specific additional processes in the water cycle (Beven and Binely, 1992). Several authors have applied satellite products linked to vegetation (Pasquato et al., 2015; Ruiz Pérez, 2016), such as NDVI (normalized difference vegetation index) (Rouse et al., 1974) or LAI (leaf area index), evapotranspiration (Koch et al., 2015; Demirel et al., 2017; Mendiguren et al., 2017; Koch et al., 2018; Stisen et al., 2018; Wambura et al., 2018; Rajib et al., 2018; Jiang et al., 2020), snow (Yassin et al., 2017; Bai et al., 2018; Tuo et al., 2018; Nemri and Kinnard, 2020) or soil moisture (Li et al., 2018), to calibrate spatially distributed hydrological models in order to reduce the uncertainty in the optimization by forcing surface hydrology.

Subsurface hydrology is one of the sources of uncertainty in rainfall-runoff models. Conceptual models often represent subsurface hydrology as a series of linear reservoirs corresponding to soil layers (topsoil, unsaturated and saturated zone, for example). Each reservoir generates a horizontal flow (overland flow, interflow and baseflow, following the previous example) that aggregates in the stream network. parameterizing this set of linear reservoirs is a complicated task mainly for two reasons: the general scarcity of data about soil properties, and the fact that the variable usually targeted in calibration is the total streamflow, i.e., an aggregated value of the outflow from the linear reservoirs. The result is a marked equifinality, in the sense that totally different allocations to each of the horizontal flows may result in a similar streamflow, without any certainty that any of these allocations reproduces the actual partition between surface and subsurface runoff in the catchment. As an example of this, Zhang et al. (2020) compared the base flow simulated by two hydrological models against the base flow obtained from the average of four common hydrograph separation methods, to find out that, even though the models performed properly in terms of total streamflow, the simulated base flow suffered from a strong bias.

Total streamflow can be decomposed into several types of flow representing different sources and delays, which in the most common case are two: quick flow and base flow (Stoelzle et al., 2020). Quick flow (also referred to as direct or storm flow) represents the component directly caused by precipitation events, both by surface runoff or by preferential paths in the soil (Cain et al., 2019). Base flow is the component coming from delayed sources such as groundwater (Hall, 1968; Tallaksen, 1995) and representing the longest and slowest decaying flows (Duncan, 2019); this type of flow is responsible for maintaining streamflow under prolonged dry weather (Stoelzle et al., 2020). Decomposing total streamflow allows us to better understand the catchment processes (Zhang et al., 2017), a knowledge that can be used for several purposes such as improving water resources management in both drought and flood events, characterize aquifers, analyze long-term changes in the water balance (Zhang and Schilling, 2006) and introduce new information in the calibration of hydrological models (Tallaksen, 1995; Zhang et al., 2017; Zhang et al., 2020).

Hydrograph separation methods are techniques that decompose the total streamflow in its components; we will focus in two-component (quick and base flow) separation techniques, even though some authors have explored multi-component separation methods (Stoelzle et al., 2020). Separation methods can be classified in two groups depending on whether they use tracers or not (Stoelzle et al., 2020; Zhang et al., 2020; Zhang et al., 2017). Tracer-based methods use the physico-chemical signature (chloride, electric conductivity, temperature, etc.) of the streamflow to estimate the age of the water and to separate quick flow from base flow. They are the most physically-grounded separation methods but require data which is often not available and whose generation requires time and resources (Tallaksen, 1995; Stoelzle et al., 2020).

Hydrograph-based separation methods only require gauged streamflow (in some cases also precipitation records (Mei et al., apr 2015)). They resort to graphical tools, digital filters and recession curves based on the Boussinesq exponential decay equation (Boussinesq, 1877; Boussinesq, 1904) (Eq. 1, where B_0 represents the baseflow at time 0, B_t represents the baseflow t periods later, and k describes the rate at which baseflow decreases between storm events ((Duncan, 2019))). Graphical methods include the widely used local minimum method (Sloto and Crouse, 1996) developed by the USGS, the United Kingdom Institute of Hydrology (UKIH) method (Gustard et al., 1992), and modifications of it (Piggott et al., Oct 2005). These methods identify turning points in the hydrograph and assume that base flow is the linear interpolation between those points; therefore, their main advantage is their simplicity and absence of parameters to be calibrated, whereas the main drawback is the lack of physical grounds since the recession decay is not applied. Filter methods include the Lyne-Hollick method (Lyne and Hollick, 1979), the Chapman-Maxwell method (Chapman, apr 1999), the Eckhardt method (Eckhardt, 2005) and the Duncan method (Duncan, 2019). They filter the high-frequency variability of total streamflow to generate either a quick flow (Lyne-Hollick method) or a base flow (Chapman-Maxwell, Eckhardt and Duncan methods) time series. These are parametric methods, requiring at least the fitting of k (which is defined as the relation between baseflow at any given time and one period earlier; it is always smaller than 1), the recession constant in the Boussinesq equation (Eq. 1), and in some cases, such as the Eckhardt method, a second parameter named the maximum base flow index (BFI).

$$B_t = B_0 \cdot k^t \tag{1}$$

The objective of this study is to use the additional information about the runoff generating processes that we can extract from hydrograph separation methods to reduce the equifinality in the calibration of a conceptual hydrological model. The idea is to decompose the streamflow records in quick flow and base flow and to sequentially calibrate the horizontal flows in the conceptual hydrological model against these decomposed flows. We hypothesize that such a calibration procedure will constrain the model to mimic the real flow decomposition, hence reducing equifinality by constraining the spectrum of optimal parameterizations. This hypothesis is built upon the idea that introducing internal state values (even though approximated) will help distinguishing among equifinal parameterizations (Beven and Binely, 1992). For the sake of simplicity and to enhance the applicability of our results, we use simple and widely spread methods; in the case of hydrograph separation, we use the local minimum method (Sloto and Crouse, 1996), whereas for the optimization of model parameters we apply the Shuffled Complex Evolution – University of Arizona method (Duan et al., 1992; Duan et al., 1993; Duan et al., 1994). The final goal is to constrain the equifinality of the calibration process and ensure obtaining plausible models without the need for any additional data apart from the observed streamflow.

2. Study area and data

The study area are two mesoscale basins, with an approximate area of 480 km^2 each, in the "Picos de Europa" National Park (northern Spain): the upper Sella and the Cares river basins. Both are middle-high mountain catchments (Fig. 1) with a large elevation range, roughly from 50 to 2600 m.a.s.l., that causes a sharp climate gradient. The climate in this area goes from temperate oceanic (Cfb in Köppen-Geiger's classification) in the valleys and close to the sea, to subarctic/sub-alpine (Dfc) in the mountain tops (Barceló and Nunes, 2009); mean annual precipitation in the area is around 1600 mm and the mean annual temperature is 10.6 °C. Anthropogenic intervention is limited in both catchments; land cover is dominated by shrubland and deciduous forest in lower areas that turns into grassland and rock outcrops in higher altitudes.

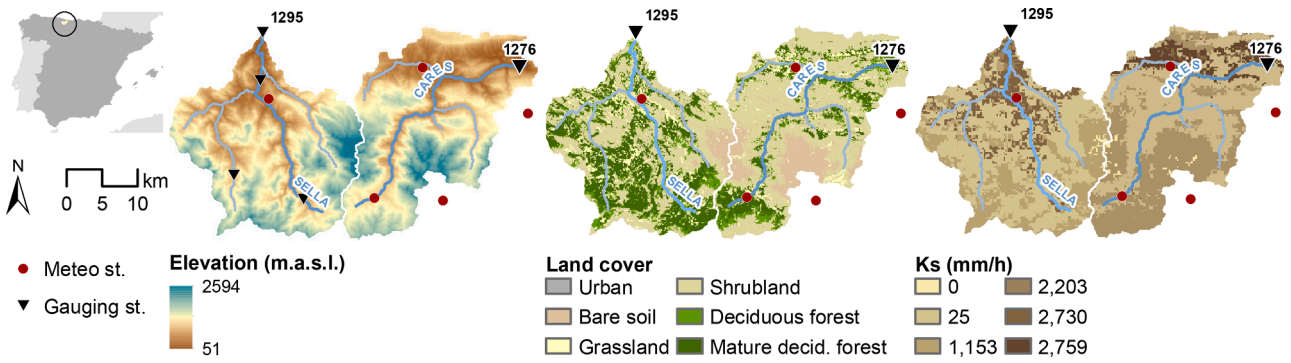

Fig. 1. Elevation, land cover and saturated hydraulic conductivity maps of the study area: upper Sella and Cares river basin (northern Spain). Black triangles represent the gauging stations in the outlets of the basins. Red dots represent the location of the rain gauges.

The hydrometeorologic data required for the study are daily climatic time series as input for the hydrological model and daily streamflow records as the target variable for performance assessment. The Spanish Meteorological Agency (AEMET) provided the climatic data from the stations in the National Park and surrounding area (Fig. 1). From this data, we generated daily maps of precipitation, mean air temperature and diurnal temperature variation by spatial interpolation. We performed a leave-one-out (LOO) cross validation to identify the best spatial interpolation method.

In a LOO procedure, the interpolation model is iteratively fitted excluding data from one station at a time, the variable is interpolated to the excluded location and the correspondence between observed and interpolated values is evaluated; once this has been done for all the stations, an aggregated performance metric (in our case the root mean square error) is computed to compare several spatial interpolation methods. The LOO crossed validation proved that universal kriging, assuming linearity between mean monthly precipitation and elevation, was the highest-performing method to interpolate precipitation. Universal kriging (a.k.a. kriging with an external drift) is a spatial interpolation method that fits a trend to the original data and applies kriging to the residues of that trend (Goovaerts, 2000); in our case, we looked for linear trends with three geographic variables that we deemed relevant: elevation, aspect, and distance to the coast. In the case of temperature (both daily mean temperature and diurnal variation), the best interpolation method was inverse distance weighted over the residuals of the linear regression between monthly mean temperature and elevation. Daily potential evapotranspiration maps were derived by the Hargreaves-Samani method (Hargreaves and Samani, 1985) from the mean air temperature and diurnal temperature variation maps. The Official Gauging Station Network of Spain (ROEA) (CEDEX, 2016) provided the observed streamflow in the gauging stations at the outlet of both catchments (Fig. 1). Since both data sets had a daily temporal resolution, that is the resolution chosen for the hydrological modeling.

Cartographic data consisted of a digital terrain model (DTM), and land cover and soil property maps. The Spanish National Center for Geographic Information (CNIG) provided a 20 m resolution DTM of the study area. The soil property maps were extracted from the database EU-SoilHydroGrids (Tóth et al., 2017), which contains 250 m resolution maps for saturated hydraulic conductivity and available water capacity. Lastly, we used a land cover map created by classification of Landsat images from 2005 (Álvarez-Martínez et al., 2018) with 30 m resolution. We chose a spatial resolution of 100 m as a trade-off between model resolution and computation time. To adapt the data to this resolution, we applied a bi-linear resampling algorithm to continuous variables and a nearest neighbor algorithm to categorical variables.

3. Methods

3.1. Hydrological model

The hydrological model used in this study is TETIS (GIMHA, 2018), a conceptual and spatially distributed model. TETIS has been applied to a broad range of climates, including Mediterranean, alpine, temperate oceanic and tropical climates, specially in Spain (Vélez et al., 2009) but also in other countries such as France (Ruiz-Villanueva et al., Oct 2014), China (Li et al., sep 2017), UK (McGrane et al., feb 2017), Colombia (Peña et al., 2016) and Kenya (Ruiz-Pérez et al., 2017). Its tank-based structure is similar to that of other renowned conceptual models, but unlike most of those, its fully spatially-distributed characteristic grants it a particular interest for analyzing the impacts of land cover change.

TETIS conceptualizes the water cycle as a series of seven reservoirs representing the storages in the water cycle: interception, snow pack, static, surface, gravitational, aquifer and streams (Fig. 2). This study focuses on the last four of these compartments, i.e., those directly affecting runoff generation. TETIS divides the soil column in three layers, each represented by a linear reservoir respectively named surface, gravitational and aquifer storage. The surface and gravitational storage are responsible for the overland flow and interflow, respectively, which discharge into gullies (the lower order streams in the stream network), whereas the aquifer storage produces base flow, which is only discharged into rivers (higher order streams). Streamflow routing along the stream network (gullies and rivers) is simulated using a geomorphological kinematic wave approach. In total, modeling runoff generation requires 25 parameters: 7 corresponding to the three compartments representing soil layers, 9 to gullies and 9 to rivers. To simplify the calibration process, TETIS is able to fit all these parameters by just calibrating 8 dimensionless hyperparameters or correction factors (FC): 2 linked to the surface storage, 3 to the gravitational storage, 2 to the aquifer, and 1 to the streamflow routing in gullies and rivers. The term hyperparameter is used in modeling to denote model parameters that modify other parameters of a lower level. In the case of TETIS, the parameters (in blue in Fig. 2) are maps inferred from physical properties of the soil, the land cover and the topography; they are fixed to keep their spatial structure. Instead of modifying the parameters directly, the hyperparameters (in red in Fig. 2) modify the parameter maps as a whole. These hyperparameters are the values tuned in the model calibration.

As mentioned in Section 2, given the extent of the catchments, the objective of the analysis and the available data, we created models for both catchments with 100 m spatial resolution and daily temporal resolution. The calibration period spans hydrological years (starting the 1st of October) from 2008 to 2013 (both included), and the validation period from 2000 to 2007; in both runs we used the first year as spin-up time.

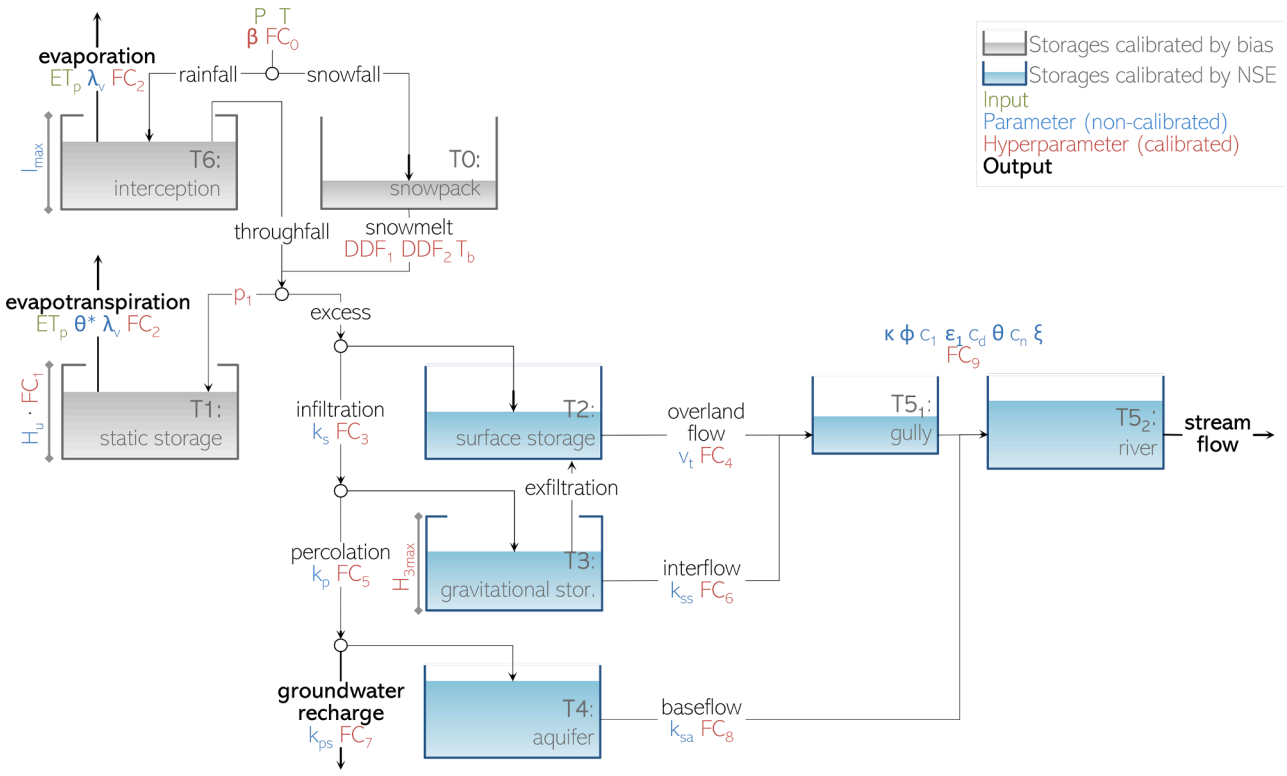

Fig. 2. Conceptualization of the hydrological model TETIS (adapted from (GIMHA, 2018)). The color of the font defines the input variables (green), the parameters (blue), the hyperparameters (red) used to calibrate the model, and the outputs (bold, black). The study focuses in the blue tanks, those affecting runoff generation.

3.2. Hydrograph separation

The main hypothesis behind this study is that the separation of the observed streamflow in two time series (quick flow and base flow) can be attributed to the three horizontal flows in TETIS. We assume that the quick flow represents the aggregation of overland flow and interflow, whereas base flow corresponds to the outflow of the aquifer storage. This assumption is grounded in TETIS' conceptualization (Fig. 2), in which both overland flow and interflow represent streamflow generated by storm events that is discharged into ephemeral streams (named gullies in TETIS), and is further supported by previous publications (Stoelzle et al., 2020; Soto, 2020; Cain et al., 2019; Duncan, 2019; Chow et al., 1988). With this approach, without any additional observation such as soil moisture, NDVI or ET, we have three time series (quick flow, base flow and total streamflow) to calibrate sequentially the horizontal flows generated by the different compartments. In a regular calibration, however, all the compartments are fitted simultaneously against the unique total streamflow.

We adopted as separation method the local-minimum method, taken from the USGS software HYSEP (Sloto and Crouse, 1996). We chose this method based on its simplicity and widespread application (Chen and Teegavarapu, 2020; Soto, 2020; Killian et al., 2019; Eckhardt, apr 2008; Zhang and Schilling, 2006); since the purpose of this study is to develop a calibration procedure, we focused on that front, assuming that if the local-minimum method improves the performance of the calibrated model, more complex and site-specific separation methods will perform better. We analyzed more complex hydrograph separation methods that reproduce the base flow regime more realistically, i.e., complying with the five features of base flow enumerated by Duncan (2019). However, we adopted the local-minimum method for the sake of simplicity, since no separation method reproduces faithfully either the real or the simulated base flow behavior (Tallaksen, 1995). With this approach, we expect to approximate the partition of total streamflow in quick and base flow, even though this method is unable to reproduce a physically-meaningful base flow hydrograph. To prove that, we conducted experiments to check that this simple approach is able to improve the calibration on an array of synthetic cases in TETIS (see the Supplementary Material of this paper).

Fig. 3 shows an extract of the local-minimum method application to the streamflow records in the two gauging stations in the study. The local-minimum method is a graphical method consisting on two steps. The first step is the identification of local minima in the hydrograph (black markers); a local minimum is a day with the lowest streamflow in a centered time window of width $2N^*$ (gray shadings). Eq. 2 defines the overland flow duration in days (N), where A is the area of the catchment in km^2 and $2N^*$ is the closest odd number to the value $2N$. Once the dates of the local minima are identified, the base flow hydrograph is the linear interpolation of the observed flow in those dates (blue line); in case the linear interpolation exceeds the observed hydrograph, the base flow is equal to the total streamflow (mid January in Fig. 3).

$$N = 0.8 \cdot A^{0.2} \ [days] \tag{2}$$

3.3. Model calibration

TETIS includes the automatic optimization algorithm Shuffled Complex Evolution - University of Arizona (SCE-UA) (Duan et al., 1992; Duan et al., 1993; Duan et al., 1994) to calibrate the model. SCE-UA tries to replicate the natural evolution in the optimization process. The modeler defines a n-dimensional parameter space inside which the algorithm will look for the global optimum (n is the number of parameters to be calibrated). The algorithm starts by randomly sampling across this parameter space a population of parameter sets (vectors of length n with a value for each parameter). From this point, it iterates a three-step process (division, evolution and shuffling) until a convergence criterion is fulfilled, and the parameter set with highest performance is selected as the global optimum (Duan et al., 1992).

In the division step, the population of parameter sets is randomly

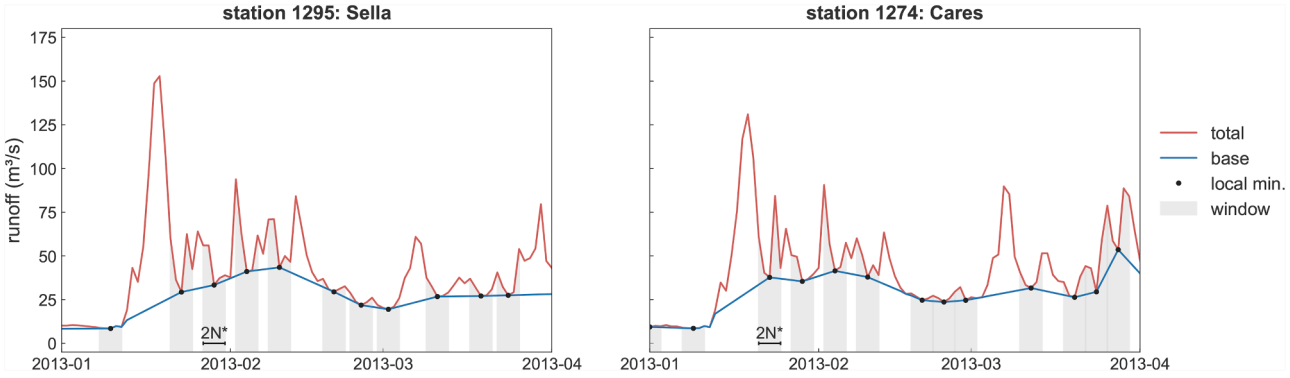

Fig. 3. Base flow separation using the local-minimum method for an extract of the streamflow series in the two gauging stations. The red line is the observed streamflow and the blue line the separated base flow; black markers represent the local minima based on a centered window of width $2N*$ (5 days for both catchments), shown in the figure by a gray shading.

divided into groups (a.k.a. complexes). During evolution, each complex evolves independently towards higher-performing members. The evolution is based on the competitive complex evolution (CCE) algorithm (Duan et al., 1992). CCE selects a couple of members from the complex (parents) and creates a new parameter set (offspring); the offspring replaces in the complex the worst-performing parent. CCE iterates allowing each member in the complex a chance to generate offspring, so there is no loss of information. As in natural evolution, the probability of a member being selected as parent is not equal, those members with higher performance are more likely to be selected. Also following natural evolution, offspring can be created by reflection (using information from the parents) or mutation (to avoid local minima and add new information). At some point, the evolution stops and complexes are shuffled to make up again a unique population; this step prevents the algorithm from finding local instead of global optima.

To grade the performance in SCE-UA, we must define an objective criteria. In this study, we used two objective criteria: the bias (Eq. 3) and the Nash–Sutcliffe efficiency coefficient (NSE, Eq. 4) (Nash and Sutcliffe, 1970).

$$bias = \frac{\sum_{t=1}^{n}\left(Q_{s,t} - Q_{o,t}\right)}{\sum_{t=1}^{n} Q_{o,t}} \cdot 100 \left[\%\right] \tag{3}$$

$$NSE = 1 - \frac{\sum_{t=1}^{n}\left(Q_{o,t} - Q_{s,t}\right)^2}{\sum_{t=1}^{n}\left(Q_{o,t} - \overline{Q}_o\right)^2} \left[-\right] \tag{4}$$

Where $Q_{o,t}$ and $Q_{s,t}$ are respectively the observed and simulated streamflow for day t, and $\overline{Q}_o$ is the mean observed streamflow.

The first step in the calibration procedure is minimizing the bias, so that the model generates the correct amount of streamflow. In this step, we fit the hyperparameters of the three storage compartments in gray shading in Fig. 2 (interception, snowpack and static), i.e., those that control the losses of water in the system by evapotranspiration. According to TETIS' conceptualization (see Fig. 2), water can only exit the interception and static storage compartments through evapotranspiration; therefore, they only affect the water balance, i.e., the bias in the streamflow. The snowmelt exiting the snowpack storage can either feed the static storage (losses through evapotranspiration) or join precipitation to produce streamflow. From the point of view of calibrating the runoff generation processes, it is either a loss or an input; therefore, we also calibrate it against bias. In particular, we adjust six parameters: three parameters of the degree-day snowmelt method (DDF_1, DDF_2, T_b), two parameters controlling soil water sorption (p_1, FC_1), and a factor of the potential evapotranspiration time series (FC_2). Bias fitting is done independently for each catchment; the hyperparameters here fitted are fixed for the rest of the calibration steps, ensuring an unbiased total streamflow.

Once the bias is fitted, the following steps in the calibration try to reproduce the flow regime in the observed streamflow. The objective criteria is the NSE and the calibration focuses on the 8 hyperparameters related to runoff generation. We compare 5 calibration methods (bottom left-hand panel in Fig. 4).

Method 0 represents the reference calibration method in which we do not apply hydrograph separation and we use the full power of the automatic calibration algorithm. We fit simultaneously all the 8 hyperparameters against the total streamflow, allowing the algorithm to explore the parameter space without constraints. The lack of constraint in this type of calibration may lead to non-realistic catchment models, even if the performance is high, which is the essence of equifinality and the reason for this study.

As opposed to this reference method, we analyze four sequential calibration methods. In these methods, we calibrate successive combinations of the reservoirs controlling runoff generation, so that we can calibrate each of these combinations against the most adequate flow (quick, base or total). **Method 1** represents the common practice of many hydrologists, who sequentially calibrate parameters led by expert knowledge and the visual inspection of the simulated streamflow. In this case, we calibrate sequentially the horizontal flows in the order that a raindrop would follow along the conceptualization of the model (Fig. 2): overland flow and interflow (T2 and T3), base flow (T4) and streamflow routing (T5). This method does not apply hydrograph separation, so the target time series in all the three steps is the total streamflow. **Method 2** replicates the three phases in method 1, but applies hydrograph separation; the sum of overland flow and interflow is calibrated against quick flow, the outflow of the aquifer storage against the base flow, and streamflow routing against total streamflow. **Method 3** also applies hydrograph separation, but it inverts the order of the phases in method 2: we first calibrate base flow, secondly quick flow, and finally total streamflow. In order to calibrate tank 4 in the first phase of method 3, we must also fit the two hyperparameters affecting infiltration and percolation (FC_3 and FC_5), which control the inflow to the tank. **Method 4** is a simplification of method 2 in which we fuse the last two phases, i.e., we fit simultaneously base flow and routing against total streamflow.

Tables 1 and 2 define more precisely the setup of the calibration methods. Table 1 defines the phases in each calibration method; for each phase, it indicates the target flow time series, the objective criteria and the parameters tuned. Table 2 defines the search range (minimum and maximum values) and the initial value applied in the calibration of the 8 hyperparameters related to runoff generation processes.

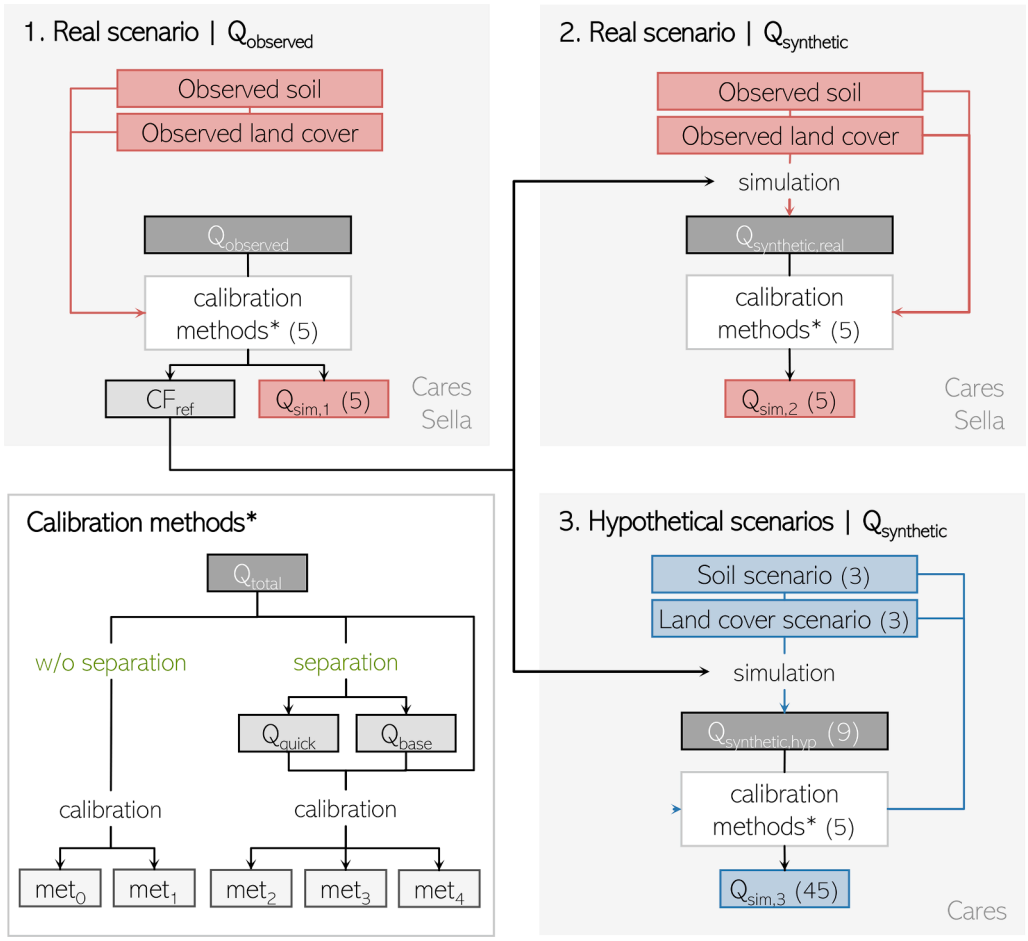

Fig. 4. Scheme of the study. The bottom left-hand panel introduces the 5 calibration methods applied to the three scenarios (hypothetical scenarios, and the real scenario targeting a synthetic or the observed streamflow), which are shown in the other three panels. CF_{ref} stands for the reference parameterization used to generate the synthetic streamflow in cases 2 and 3.

Table 1
Calibration process for each of the methods compared in this study. The first phase, water balance, is common for all the methods. OC stands for objective criteria.

Method	Phase	Flow type	OC	Parameters
All methods	1. Water balance	Total	Bias	$DDF_1, DDF_2, T_b, p_1, FC_1, FC_2$
Method 0	2. T2-T5	Total	NSE	$FC_3, FC_4, FC_5, H_{3,max}, FC_6, FC_7, FC_8, FC_9$
Method 1	2. T2, T3	Total	NSE	$FC_3, FC_4, FC_5, H_{3,max}, FC_6$
	3. T4	”	”	FC_7, FC_8
	4. T5	”	”	FC_9
Method 2	2. T2, T3	Quick	NSE	$FC_3, FC_4, FC_5, H_{3,max}, FC_6$
	3. T4	Base	”	FC_7, FC_8
	4. T5	Total	”	FC_9
Method 3	2. T4	Base	NSE	FC_3, FC_5, FC_7, FC_8
	3. T2, T3	Quick	”	$FC_4, H_{3,max}, FC_6$
	4. T5	Total	”	FC_9
Method 4	2. T2, T3	Quick	NSE	$FC_3, FC_4, FC_5, H_{3,max}, FC_6$
	3. T4, T5	Total	”	FC_7, FC_8, FC_9

Table 2
Search range and initial value applied in the calibration of the parameters related to runoff-generating processes. All the FC parameters are dimensionless. $H_{3,max}$ is measured in mm.

Parameter	Description	Minimun	Initial	Maximum
FC_3	Infiltration	0	0.2	1.5
FC_4	Overland flow	0.001	1	10
FC_5	Percolation	0	0.4	1.5
FC_6	Interflow	0.001	10	5,000
$H_{3,max}$	Max. capacity of tank 3 (mm)	0	100,000	1,000,000
FC_7	Groundwater recharge	0	0.5	1
FC_8	base flow	0.001	200	50,000
FC_9	Streamflow routing	0	1	1.5

3.4. Scenarios

We applied the five calibration methods to three scenarios: the calibration of the real land use and soil scenario using the observed streamflow, and the calibration of synthetic streamflow using either the real land use and permeability maps or hypothetical maps. Fig. 4 summarises the general workflow of the study.

In the first experiment, we calibrated the two catchment models in the regular case, i.e., using the observed streamflow and observed land use and permeability maps. For the subsequent experiments a reference parameterization was required (CF_{ref} in Fig. 4). To make sure that this parameterization represented a feasible set, we used the results from the calibration by the reference method (method 0), selecting from the SCE-

UA iterations those parameterizations we considered behavioral ($|bias| \leq 5\%$ and $NSE \geqslant 0.58$), and calculating for each parameter the median value among those behavioral parameterizations. Since the median of a set of behavioral parameterizations does not need to be behavioral, we checked that the median parameterization did satisfy this requirement.

Experiments two and three are synthetic cases in which we created a fictitious streamflow by running the model with the reference parameterization and various land use and permeability maps. Calibrating synthetic cases has some advantages. Since the target streamflow is generated by the model itself, we are certain that it is possible to reach an ideal performance and we expect to find the original parameterization; problems in reaching any of these two goals must be attributed to a poor calibration procedure or to equifinality. Furthermore, synthetic cases allow us to check if the hydrograph separation method represents the partition between quick flow and base flow in the hydrological model; we compare the synthetic base flow (not used in the calibration procedure) with the base flow obtained from the separation of the synthetic total streamflow.

In experiment two, we simulated a synthetic streamflow using the reference parameterization and the observed land use and permeability maps. We proceeded by calibrating the model with the five methods against this synthetic streamflow. The whole procedure (simulation of a synthetic streamflow and calibration) was repeated for both catchments. In experiment three, with the objective of analysing the applicability of the methodology to diverse conditions without studying a large amount of catchments, we designed a series of hypothetical scenarios of land cover and soil permeability in the Cares basin. We generated synthetic streamflow for the nine hypothetical scenarios and calibrated each of them by the five methods. For the sake of brevity, the methodology of experiment three and the results of the second and third experiments are explained in the Supplementary Material.

3.4.1. Assessment of performance

In all three experiments, the procedure is the same: separation of the target total streamflow and calibration by means of the 5 methods described in Section 3.3. We compared methods in three ways: performance in both the total and separated flows, visual inspection of hydrographs, and values of the optimized parameterization. We must stress that, in the synthetic cases, the flow time series used for calibration and performance assessment are not the same. To calibrate, we feed the algorithm with the time series resulting from the hydrograph separation method, thereby reproducing a real case in which the base flow is unknown. However, we assess performance by comparison with the synthetic base flow, so that we evaluate how the calibrated model reproduces the original partition, instead of the approximation done by the hydrograph separation method.

4. Results

4.1. Observed streamflow in the real scenario

Fig. 5 shows the performance in the calibration and validation periods for the experiment 1, i.e., the real scenario calibrated against the observed streamflow. Left-hand panels show the performance in terms of NSE, central panels for bias, and the top right-hand panel for the number of iterations required in each of the calibration phases. Results are organised by type of flow (quick, base or total) and calibration method (from 0 to 4). Vertical, gray lines indicate the distance to the target value of the objective function.

Regarding the total flow, there are four methods (all but method 3) that performed good in both catchments; total flow bias is close to zero and the NSE can be classified as good (around 0.6, except method 0 with values slightly lower than 0.5). Method 3 has a strong negative bias in both catchments (close to −50%) and a poor NSE (below 0.4). We expected that the reference method (method 0) would be the highest-performing in terms of total flow, since it uses the full power of the automatic algorithm. To our surprise, the reference method is the one with the lowest efficiency among those performing appropriately, whereas methods applying flow separation (methods 2 and 4) performed slightly better than the rest, even for the total flow. Total flow performance in the validation period is basically similar; method 3 misbehaves while the other four methods are still unbiased, but show a certain loss in NSE. This loss is larger in the Sella basin, which causes that the efficiency in the Sella model is lower than that of the Cares model during the validation period, contrary to what happened for the calibration period.

The four well-performing models in terms of total flow behave differently when looking at the partition in quick and base flow. The two methods without hydrograph separation (methods 0 and 1) show a strong bias, even larger than 100% (quick flow in method 0 for the Sella

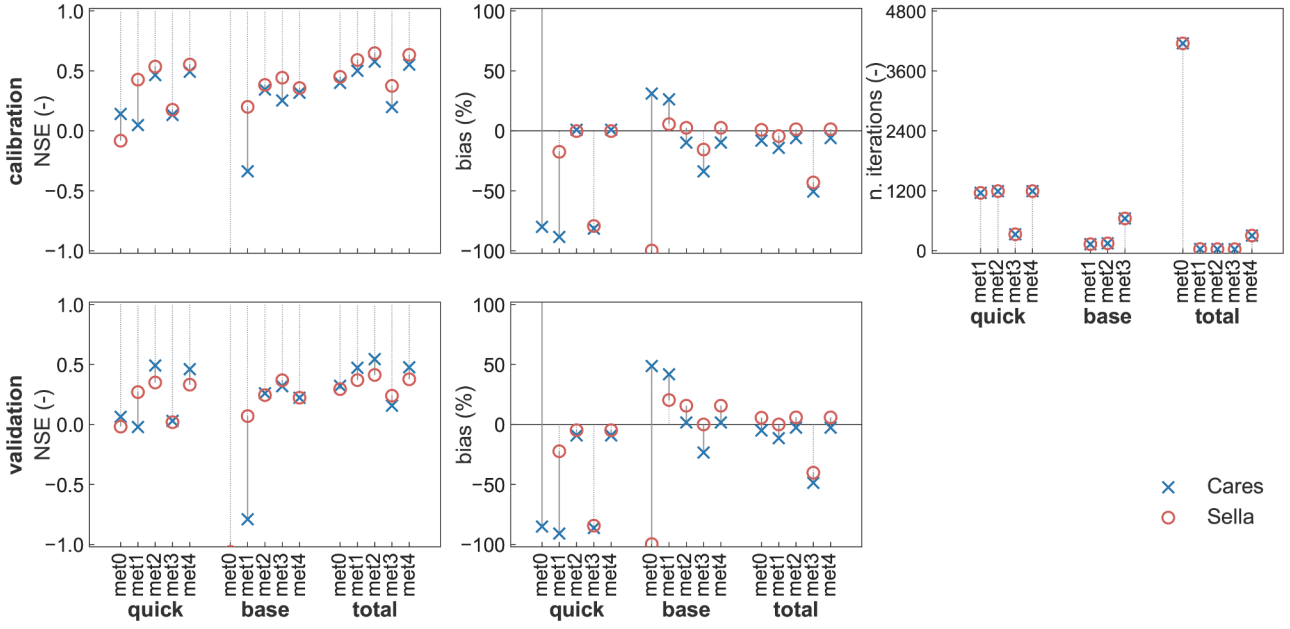

Fig. 5. Performance (NSE and bias) in the calibration and validation for the real scenario with the observed streamflow for the Cares and upper-Sella river basins. The top right-hand panel exhibits the number of iterations required for calibration. The vertical lines represent the spread between catchments.

basin). The NSE for these two methods is poor for both separated flows (close to −1 for the base flow) and distinctly lower than the two methods that apply hydrograph separation (methods 2 and 4). These last two methods are the ones with the highest performance; the separated bias is close to null and the NSE is larger than the rest of methods (with the exception of base flow in the Sella model). The similar performance of these two methods comes from the fact that method 4 is a simplification of method 2 with a common quick flow calibration phase. Method 3, rejected in terms of total flow performance, also misbehaves in the separated flows; it shows negative bias in both quick and base flow (stronger in the former), whereas NSE is poor for the quick flow (0.15 in average), but acceptable for the base flow (0.35 in average), similar to methods 2 and 4. The performance of separated flows in the validation period renders a similar behavior, where methods 2 and 4 are those performing better. As well as in the total flow, the loss of efficiency in the validation is larger for the Sella model; this loss is more marked for the base than the quick flow, both in terms of bias and NSE.

The null variability in the number of iterations among catchment models is remarkable. Regardless of the catchment, the number of iterations that SCE-UA requires to converge is controlled by the number of parameters to be optimized in a given phase. This fact favours sequential calibration methods (all but method 0), that instead of analysing an eight dimensional parameter space (as many as parameters involved in runoff generation), explore subsets of it. The result is that sequential methods require a third of the number of iterations needed by the reference method. This reduction in computation time is gained at the expense of a less exhaustive analysis of the parameter space and, thus, a lower probability of finding the global optimum. The difference between sequential methods applying flow separation (methods 2, 3 and 4) and the common practice (method 1) is that we use our knowledge about how the catchment may work to ensure optimized models with a proper flow partition.

The scatter plots in Fig. 6 provide an insight into the previous performance results. The plots compare the observed flow Q_{obs} with the simulated flow Q_{sim} (generated with the optimized parameterization), each dot represents the pair of values for a day. Colours depict types of flow in order to show how the calibrated model reproduces flow partition. In a perfect fit, all the dots would be placed along the 1:1 line, which means equal observed and simulated flow. We must stress that in this experiment the observed quick and base flows are not actually observed, but created with the hydrograph separation method. This means that the partition has no real meaning. Nevertheless, the results of the synthetic experiments (Supplementary Material) prove that the simplicity of the hydrograph separation method here employed was representative enough of the real behavior of the catchment.

A visual analysis shows that 5 out of the 10 plots represent a correct model, namely, method 1 in the Sella basin and methods 2 and 4 in both basins. Method 0 optimized completely opposite models for the two basins; whereas in the Cares model base flow accounts for almost the entire total flow (therefore quick flow is basically non-existent), in the Sella model there is no base flow and the total flow is equal to the quick flow. Given the similar climate and geology in these two adjacent basins, it is extremely unlikely that the two basins function in such a different manner; most likely, the real catchments do not behave in any of these extreme ways. The Cares model optimized by method 1 also suffers from an incorrect flow partition; similarly to method 0, the model attributes almost the entire total flow to the base flow, so there is no quick flow. Method 3 fails to reproduce quick flow, probably due to a wrong parameterization of infiltration and percolation during the base flow fitting.

The optimized hyperparameter values obtained by each calibration method are shown in Fig. 7. Each plot contains the values for a catchment; hyperparameters are grouped by the type of flow they directly affect. The dotted, gray line spans the extreme optimized values to show the variability among methods. Parameter values are normalized by their search range (Table 2) to be able to show parameters with different orders of magnitude.

The four methods with an optimal total flow performance (methods 0, 1, 2 and 4) optimized diverse parameterizations; an example of the equifinality problem. When comparing those methods that partition flow correctly (methods 2 and 4, and method 1 in the Sella basin), the variability in the parameterizations is notably reduced: FC_3, FC_5, FC_6, FC_7, FC_8 and FC_9 adopt similar values in the Sella basin, and parameterizations in the Cares basin are almost identical for methods 2 and 4. This means that the parameter range that properly addresses flow partition is narrower than that which only reproduces correctly the total flow. Our approach attempts to guide the automatic calibration algorithm towards this limited range, thus avoiding cases where a correct total flow calibration does not represent a realistic catchment functioning.

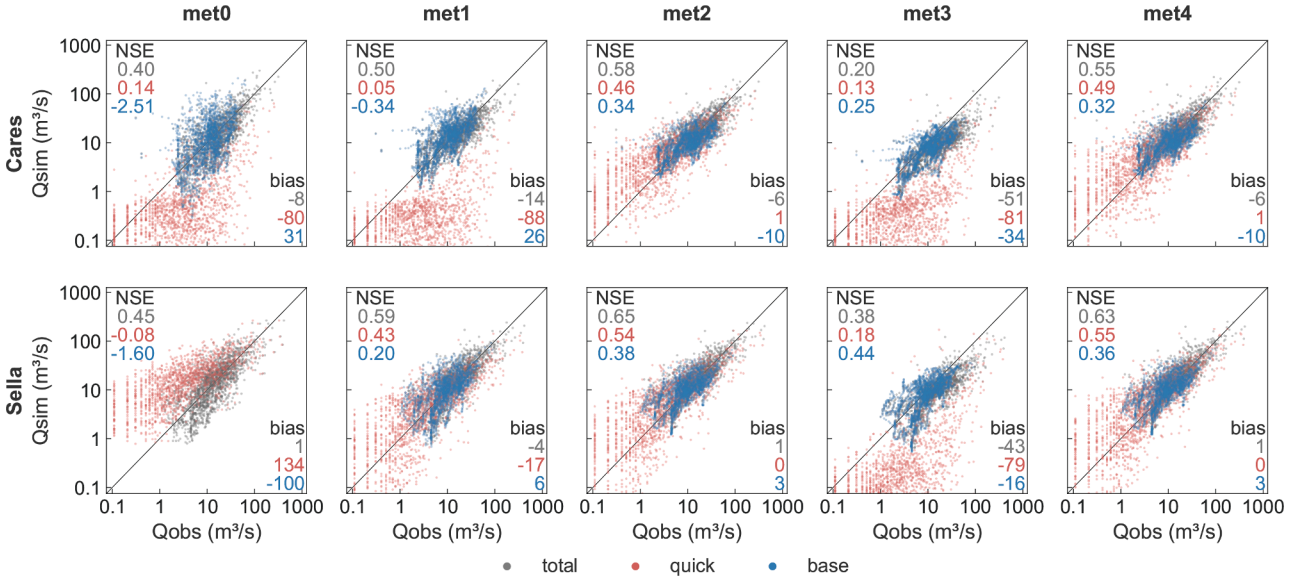

Fig. 6. Scatter plots of the observed vs simulated daily flows in the calibration of the real scenario with the observed streamflow for the Cares and upper-Sella river basins. Rows correspond to each of the catchments; columns correspond to the calibration methods. It includes the performance in terms of NSE and bias for the three types of flow, represented by colours.

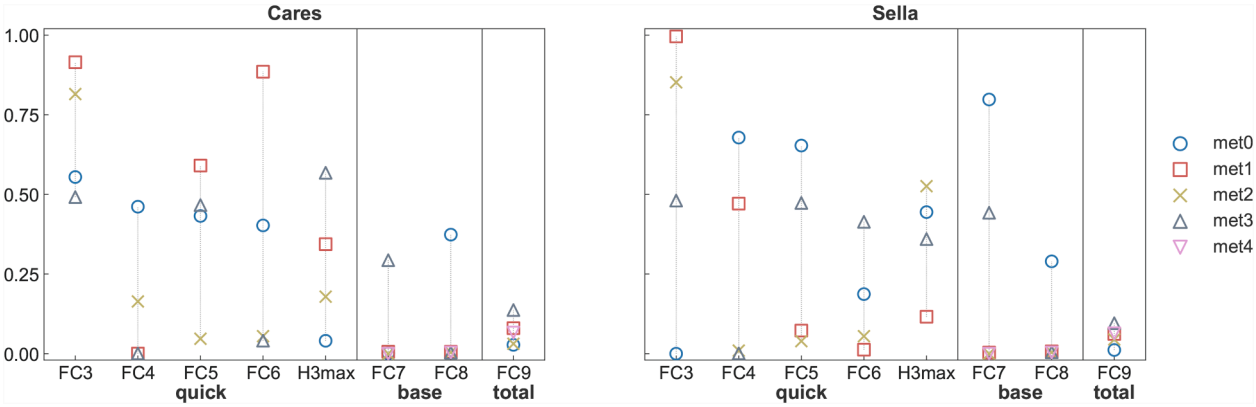

Fig. 7. Optimized parameterization in the calibration of the real scenario with the observed streamflow for the Cares and Sella basins. Hyperparameter values are normalized by their search range.

In the majority of cases the extreme optimized values correspond to one of the methods that could not reproduce the flow partition (methods 0 and 3, and method 1 in the Cares basin). Quick flow parameters show larger variability than the base and total flow parameters, with the exception of FC_7 in the Sella model. The variability in the optimized values might be linked to the sensitivity of the parameter. Highly sensitive parameters have a narrow range of values with optimum performance (Pianosi et al., 2016), so the calibrated parameter value is similar regardless of the calibration method. Therefore, base and total flow parameters would be more sensitive than quick flow parameters.

4.2. Synthetic streamflow in the real scenario

The results of this experiment are thoroughly commented on the Supplementary Material. To summarize, this experiment proved that the proposed correspondence between the quick and base flow time series resulting from hydrograph separation and those simulated with TETIS is not perfect, but approximate enough. In terms of performance in the simulation of total flow, only methods 1 and 2 obtained a high NSE and a negligible bias in both catchments. When analysing separated flows, method 1, which does not apply flow separation, is clearly outperformed by method 2, which does apply flow separation. Regarding optimized hyperparameter values, none of the calibration methods was able to replicate the reference parameterization, proving that the calibration methods here applied do not remove equifinality even in a simplified scenario such as a synthetic case. However, methods that apply flow separation get values closer among them and closer to the reference value.

These results prove that, even though method 2 does not remove equifinality, it can at least constrain it and force the calibration to get parameterizations closer to the reference and with a correct behavior in terms of runoff partition.

4.3. Hypothetical scenarios

The results of this experiment are thoroughly commented on the Supplementary Material. The eighteen calibrations carried out in this experiment are a perfect example of the problem of equifinality and the potential of the sequential methods here exposed.

When analysing the total flow performance, all methods performed successfully both in terms of NSE and bias. As expected, the reference method (method 0) was the highest-performing one. The difference between methods appears when looking at the performance for the quick and base flow, where methods that apply flow separation stand out, specially method 2. In some scenarios, method 0 represents the perfect example of the risk of granting physical meaning to an optimised model without further tests: the simulation of total flow was utterly perfect, but the partition clearly wrong. Methods applying flow separation, instead, reproduce adequately flow partition at the expense of a slight loss in total flow performance.

The automatic optimization algorithm could not find the reference parameterization in any of the hypothetical scenarios. This is the case even for method 0, for which we allowed the algorithm to exhaustively explore the parameter space. However, methods using hydrograph separation (2, 3 and 4) reduce the variability in the optimized hyperparameter values among scenarios, hence constraining the equifinality problem.

5. Discussion

The multiple cases here presented dwell in the difficulty of coping with the equifinality problem and the need to constrain the calibration process in order to tackle it. We show how diverse parameterizations representing different catchment functioning are equivalent in terms of total flow performance. Our approach attempts to distinguish which of those parameterizations also reproduce the actual catchment processes, which we represent by the partition of the total flow.

Results show that the reference calibration method, in which we apply no constraints in the parameter space to be evaluated, outperforms the methods here designed in the hypothetical scenarios, but not in the real scenario, neither with a synthetic nor the observed streamflow as calibration target. Apart from that fact, the reference method has two disadvantages. In the first place, in a considerable number of cases the method optimizes catchment models physically unfeasible, i.e., in which the partition in quick and base flow is incorrect. From our results, this fact occurs in all the four real case scenarios and in three out of nine hypothetical scenarios. In the second place, this method is computationally less efficient than the sequential methods; since the algorithm must explore at once the eight-dimensional parameter space, the number of iterations it requires is approximately 3 times larger than that needed in the sequential methods, which explore subsets of that parameter space.

As a method in between the reference and the partition methods here developed, we included the common practice of many hydrologists, who calibrate models by sequentially fitting parameters always with the total streamflow as target. Results prove that this method outperforms the reference method in the real scenarios, but is the one with the lowest performance in the hypothetical scenarios. In both real and hypothetical scenarios it shows high performance in terms of total flow, but it fails to reproduce flow partition in one out of four real scenarios and all of the nine hypothetical scenarios.

Among the methods that apply flow separation, method 2 stands out as the highest-performing of all. In this method, we sequentially calibrate storages from top to bottom soil layers and lastly the streamflow

routing. The main attribute of this method is that it ensures a correct flow partition in all cases. In the real scenario (both synthetic and observed streamflow) this method outperforms all of the others, even unexpectedly for the total flow; in the hypothetical scenarios, the improvement in flow partition is gained at the expense of a slight loss of total flow performance when compared with the reference method. Because this method explores subsets of the parameter space, it may not find the global optimum (what explains the loss of total flow performance in the hypothetical scenarios), but it reduces the computational effort to a third.

None of the methods is able to resolve the equifinality in the calibration of the conceptual hydrological model. Proof of that fact are the diverse parameterizations that performed correctly in the synthetic cases, in which the calibration algorithm should have found similar parameter values to the reference parameterization. The fact that the automatic calibration algorithm, when applied with total freedom such as in method 0, has not been able to replicate neither the streamflow nor the parameterization in the synthetic cases of the real scenario, raises doubts about its efficiency in real scenarios. Although equifinality is not removed, the results show that constraining the global search in the parameter space, such as in method 2, guides the automatic algorithm towards a reduced region in the parameter space with better performance and, more importantly, realistic catchment models. This study does not attempt to invalidate SCE-UA as a global optimization algorithm, but we draw attention to the importance of constraining the calibration of conceptual hydrological models in order to prevent equifinality from generating unrealistic catchment models. This is of vital importance when the hydrological model is the starting point of a posterior analysis, for instance, erosion, nutrient transport or vegetation dynamics.

We have employed the local-minimum method to separate the total streamflow into its quick and base flow components, which is a simplistic approach. To prove that the simplicity of this method does not affect the outcomes of the study, we evaluated its performance in synthetic cases, both in the real and hypothetical scenarios. The negligible bias between separated and synthetic flows (see Figure S1) proves that the separation method properly represents the flow partition in TETIS. However, there is ground for improvement in this part; applying more realistic separation methods, such as those presented in Duncan (2019); Eckhardt, apr 2008; Chapman, apr 1999; Lyne and Hollick, 1979 or tracer methods if available, will enhance the capabilities of the sequential calibration procedures here developed, specifically in terms of base flow NSE. There is a limitation in this regard, which is the representation of the aquifer's outflow in the hydrological model; no matter how realistic the separation method is, the base flow performance is limited to the ability of the model to simulate base flow, which in the case of TETIS seems to be excessively reactive, giving support to the implementation of a cascade of linear reservoirs instead of a single one (Tallaksen, 1995). Future research might also deal with the idea of dividing streamflow in more than two components (Stoelzle et al., 2020). If we could separate it in three components (surface runoff, interflow and base flow), we could calibrate individually each of the three soil storage reservoirs involved in runoff generation, boosting the advantages of the sequential calibration method here developed. This approach may also improve the observed lack of sensitivity of the quick flow parameters, by calibrating these 5 parameters in two instead of one single phase.

In this study we have not addressed equifinality in crucial processes such as snow melt processes and notably evapotranspiration, represented in TETIS by three specific storage reservoirs (interception, snowpack and static). Current research in our group deals with this shortcoming; we are exploring how to incorporate remotely sensed products in the calibration of snow cover and evapotranspiration, exploiting also the spatially distributed property of TETIS (Koch et al., 2015; Demirel et al., 2017; Mendiguren et al., 2017; Koch et al., 2018; Bai et al., 2018; Tuo et al., 2018; Nemri and Kinnard, 2020).

The results here presented are specific to a type of climate and hydrological regime. We consider that the benefits of sequential calibration and flow partition can be extrapolated to other climates and hydrological regimes, though further analyses must be done to prove it; specially regarding the hydrograph separation method.

6. Conclusions

In this study, we have analyzed the equifinality in the calibration of a conceptual hydrological model and developed a sequential calibration method that limits its consequences. The idea is to calibrate independently the outflows of the different runoff-generating reservoirs in the hydrological model against a time series representative of that process; to create these target time series, we apply hydrograph separation using the local-minimum method. The study focuses on two mesoscale catchments in the northern side of the "Picos de Europa" National Park (Spain); in these two catchments, we carried out three calibration experiments: the real case scenario with both the observed and a synthetic streamflow as targets in the calibration, and hypothetical scenarios of land cover and soil permeability with synthetic streamflows.

Results prove that a sequential calibration method using separated flows leads the automatic calibration algorithm towards model parameterizations that better reproduce the runoff-generating processes occurring in the catchment. In the real scenario, this procedure not only improves the performance for the separated flows, but also for total streamflow. Only when applied to hypothetical scenarios of soil permeability and land cover, the improved performance in the separated flows comes at the expense of a slight loss of total streamflow performance. The constraints imposed in the sequential calibration procedures reduce the range in the parameter space of the behavioral models, hence they reduce equifinality. On top of the previous, sequential calibration methods are computationally more efficient since they explore the parameter space in subspaces; the setup of this study reduced the number of iterations in the optimization algorithm to a third.

This study presents a first, promising attempt to use hydrograph separation to improve the calibration of conceptual hydrological models. Further research should implement more realistic separation methods that better reproduce the nature of base flow, or decompose streamflow in more than two components. We strongly suggest imposing constraints, such as separated flows, in the calibration of hydrological models to induce models that better reproduce catchment processes. This is of grand importance if the purpose of hydrological model is not focused strictly on streamflow, but to analyse catchment processes such as erosion, impacts of land cover change, groundwater recharge or nutrient cycling, for instance.

CRediT authorship contribution statement

Jesús Casado-Rodríguez: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Visualization. **Manuel del Jesus:** Methodology, 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

This research was originally funded by the Spanish National Research Agency and the European Regional Development Fund through the research project GESDIVAH (Ref. BIA2016-78397-P). The first author is sponsored by the Spanish Ministry of Education and Vocational Training through a pre-doctoral scholarship (FPU 17/

05353). Manuel del Jesus acknowledges the financial support from the Government of Cantabria through the Fénix Programme. The authors would like to thank José Manuel Álvarez Martínez for contributing the land cover map.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.jhydrol.2022.127816.

References

Acero Triana, J.S., Chu, M.L., Guzman, J.A., Moriasi, D.N., Steiner, J.L., 2019. Beyond model metrics: the perils of calibrating hydrologic models. J. Hydrol. 578 (August), 124032.

Álvarez-Martínez, J.M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., Juanes, J.A., 2018. Modelling the area of occupancy of habitat types with remote sensing. Methods Ecol. Evol. 9 (3), 580–593.

Bai, P., Liu, X., Liu, C., 2018. Improving hydrological simulations by incorporating GRACE data for model calibration. J. Hydrol. 557, 291–304.

Barceló, A.M., Nunes, L.F., 2009. Iberian Climate Atlas 1971-2000. Agencia Estatal de Meteorologia (Ministerio de Medio Ambiente y Medio Rural y Marino). Instituto de Meteorologia de Portugal.

Beven, K., Binely, A., 1992. The future of distributed model: calibration and uncertainty prediction. Hydrol. Process. 6 (May 1991), 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 (1–4), 11–29.

Boussinesq, J., 1877. Essai sur la theorie des eaux courantes. Memoires presentes par divers savants a l'Academie des Sciences de l'Institut National de France, Tome XXIII 1.

Boussinesq, J., 1904. Recherches théoriques sur l'écoulement des nappes d'eau infiltrées dans le sol et sur le débit des sources. J. de Mathématiques Pures et Appliquées 10, 5–78.

Cain, M.R., Ward, A.S., Hrachowitz, M., 2019. Ecohydrologic separation alters interpreted hydrologic stores and fluxes in a headwater mountain catchment. Hydrol. Process. 33 (20), 2658–2675.

CEDEX, 2016. Anuario de aforos 2013-2014. Tech. rep., Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente.

Chapman, T., apr 1999. A comparison of algorithms for stream flow recession and baseflow separation. Hydrol. Process. 13, 701–714.

Chen, H., Teegavarapu, R.S., 2020. Comparative analysis of four baseflow separation methods in the south atlantic-gulf region of the U.S. Water 12 (120), 1–17.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology. McGraw-Hill, Singapore.

Demirel, M.C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., Stisen, S., 2017. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. Hydrol. Earth Syst. Sci. Discuss. 1–22.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28 (4), 1015–1031.

Duan, Q., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. J. Optim. Theory Appl. 76 (3), 501–521.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284.

Duncan, H.P., 2019. Baseflow separation – a practical approach. J. Hydrol. 575, 308–313.

Eckhardt, K., 2005. How to construct recursive digital filters for baseflow separation. Hydrol. Process. 19 (2), 507–515.

Eckhardt, K., apr 2008. A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. J. Hydrol. 352, 168–173.

GIMHA, 2018. Description of the Distributed Conceptual Hydrological Model Tetis V.9.0.1. Tech. rep.. In: Research Group of Hydrological and Environmental Modelling (GIMHA). Universitat Politècnica de València, Valencia.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129.

Gustard, A., Bullock, A., Dixon, J.M., 1992. Low flow estimation in the United Kingdom. Tech. rep. Institute of Hydrology, Wallingford.

Hall, F.R., 1968. Base-Flow Recessions—a review. Water Resour. Res. 4 (5), 973–983.

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Appl. Eng. Agric. 1 (2), 96–99.

Jiang, L., Wu, H., Tao, J., Kimball, J.S., Alfieri, L., Chen, X., 2020. Satellite-based evapotranspiration in hydrological model calibration. Remote Sens. 12 (3).

Kavetski, D., 2018. Parameter estimation and predictive uncertainty quantification in hydrological modelling. In: Handbook of Hydrometeorological Ensemble Forecasting. Springer, Ch. Parameter, pp. 1–42.

Killian, C.D., Asquith, W.H., Barlow, J.R., Bent, G.C., Kress, W.H., Barlow, P.M., Schmitz, D.W., 2019. Characterizing groundwater and surface-water interaction using hydrograph-separation techniques and groundwater-level data throughout the Mississippi Delta, USA. Hydrogeol. J. 27, 2167–2179.

Koch, J., Jensen, K.H., Stisen, S., 2015. Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. Water Resour. Res. 51, 9127–9140.

Koch, J., Demirel, M.C., Stisen, S., 2018. The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models. Geosci. Model Dev. 11 (5), 1873–1886.

Li, Z., Fang, H., sep 2017. Modeling the impact of climate change on watershed discharge and sediment yield in the black soil region, northeastern China. Geomorphology 293, 255–271.

Li, Y., Grimaldi, S., Pauwels, V.R., Walker, J.P., 2018. Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: the impact on predictions at gauged and ungauged locations. J. Hydrol. 557, 897–909.

Lyne, V., Hollick, M., 1979. Stochastic time-variable rainfall-runoff modelling. In: Institute of Engineers Australia National Conference. pp. 89–92.

McGrane, S.J., Hutchins, M.G., Miller, J.D., Bussi, G., Kjeldsen, T.R., Loewenthal, M., feb 2017. During a winter of storms in a small UK catchment, hydrology and water quality responses follow a clear rural-urban gradient. J. Hydrol. 545, 463–477.

Mei, Y., Anagnostou, E.N., apr 2015. A hydrograph separation method based on information from rainfall and runoff records. J. Hydrol. 523, 636–649.

Mendiguren, G., Koch, J., Stisen, S., 2017. Spatial pattern evaluation of a calibrated national hydrological model - A remote-sensing-based diagnostic approach. Hydrol. Earth Syst. Sci. 21 (12), 5987–6005.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-a discussion of principles*. J. Hydrol. 10, 282–290.

Nemri, S., Kinnard, C., 2020. Comparing calibration strategies of a conceptual snow hydrology model and their impact on model performance and parameter identifiability. J. Hydrol. 582 (May 2019), 124474.

Pasquato, M., Medici, C., Friend, A.D., Francés, F., 2015. Comparing two approaches for parsimonious vegetation modelling in semiarid regions using satellite data. Ecohydrology 8 (6), 1024–1036.

Peña, L.E., Barrios, M., Francés, F., 2016. Flood quantiles scaling with upper soil hydraulic properties for different land uses at catchment scale. J. Hydrol. 541, 1258–1272.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Soft. 79, 214–232.

Piggott, A.R., Moin, S., Southam, C., Oct 2005. A revised approach to the UKIH method for the calculation of baseflow. Hydrol. Sci. J. 50 (5), 911–920.

Rajib, A., Evenson, G.R., Golden, H.E., Lane, C.R., 2018. Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. J. Hydrol. 567 (April), 668–683.

Rouse, J.W., Hass, R.H., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the great plains with ERTS. In: Third Earth Resources Technology Satellite (ERTS) symposium, vol. 1. pp. 309–317.

Ruiz Pérez, G., 2016. On the use of satellite data to calibrate a parsimonious ecohydrological model in ungauged basins. Ph.D. thesis, Universidad Politécnica de Valencia.

Ruiz-Pérez, G., Koch, J., Manfreda, S., Caylor, K., Francés, F., 2017. Calibration of a parsimonious distributed ecohydrological daily model in a data-scarce basin by exclusively using the spatio-temporal variation of NDVI. Hydrol. Earth Syst. Sci. 21, 6235–6251.

Ruiz-Villanueva, V., Stoffel, M., Bussi, G., Francés, F., Bréthaut, C., Oct 2014. Climate change impacts on discharges of the Rhone River in Lyon by the end of the twenty-first century: model results and implications. Reg. Environ. Change 15 (3), 505–515.

Shokri, A., Walker, J.P., van Dijk, A.I., Wright, A.J., Pauwels, V.R., 2018. Application of the patient rule induction method to detect hydrologic model behavioural parameters and quantify uncertainty. Hydrol. Process. 32 (8), 1005–1025.

Sloto, R.A., Crouse, M.Y., 1996. Hysep: a computer program for streamflow hydrograph separation and analysis. U.S. Geological Survey Water-Resources Investigations Report 96-4040, 54.

Soto, B., 2020. Trends of hydrograph components in rivers of North of Iberian Peninsula during 1972–2012. Environ. Earth Sci. 79 (16), 1–15.

Stisen, S., Koch, J., Sonnenborg, T.O., Refsgaard, J.C., Bircher, S., Ringgaard, R., Jensen, K.H., 2018. Moving beyond run-off calibration—multivariable optimization of a surface–subsurface–atmosphere model. Hydrol. Process. 32 (17), 2654–2668.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., Tallaksen, L.M., 2020. Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions. Hydrol. Earth Syst. Sci. 24, 849–867.

Tallaksen, L., 1995. A review of baseflow recession analysis. J. Hydrol. 165, 349–370.

Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. Hydrol. Process. 31 (14), 2662–2666.

Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. A multi-objective approach to improve SWAT model calibration in alpine catchments. J. Hydrol. 559, 347–360.

Vélez, J.J., Puricelli, M., López Unzu, F., Francés, F., 2009. Parameter extrapolation to ungauged basins with a hydrological distributed model in a regional framework. Hydrol. Earth Syst. Sci. 13 (2), 229–246.

Wambura, F.J., Dietrich, O., Lischeid, G., 2018. Improving a distributed hydrological model using evapotranspiration-related boundary conditions as additional constraints in a data-scarce river basin. Hydrol. Process. 32 (6), 759–775.

Yassin, F., Razavi, S., Wheater, H., Sapriza-Azuri, G., Davison, B., Pietroniro, A., 2017. Enhanced identification of a hydrologic model using streamflow and satellite water storage data: a multicriteria sensitivity analysis and optimization approach. Hydrol. Process. 31 (19), 3320–3333.

Zhang, Y.K., Schilling, K.E., 2006. Increasing streamflow and baseflow in Mississippi River since the 1940 s: effect of land use change. J. Hydrol. 324 (1–4), 412–422.
Zhang, J., Zhang, Y., Song, J., Cheng, L., 2017. Evaluating relative merits of four baseflow separation methods in Eastern Australia. J. Hydrol. 549, 252–263.
Zhang, J., Zhang, Y., Song, J., Cheng, L., Kumar Paul, P., Gan, R., Shi, X., Luo, Z., Zhao, P., 2020. Large-scale baseflow index prediction using hydrological modelling, linear and multilevel regression approaches. J. Hydrol. 585, 124780.