Key Points:
- A flexible simulation environment coupled with dynamic identifiability analysis forms a diagnostic tool for process-based model building
- Alternative model configurations in terms of process representation, numerical solver, and parametrization are compiled
- Identifiability of model structures varies in space and time driven by the governing hydro-meteorological conditions

Supporting Information:
- Supporting Information S1

Correspondence to:
T. Pilz,
tobias.pilz@pik-potsdam.de

How to Tailor My Process-Based Hydrological Model? Dynamic Identifiability Analysis of Flexible Model Structures

Tobias Pilz[1,2] , **Till Francke[1]** , **Gabriele Baroni[3]** , and **Axel Bronstert[1]**

[1]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany, [2]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [3]Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy

Abstract In the field of hydrological modeling, many alternative representations of natural processes exist. Choosing specific process formulations when building a hydrological model is therefore associated with a high degree of ambiguity and subjectivity. In addition, the numerical integration of the underlying differential equations and parametrization of model structures influence model performance. Identifiability analysis may provide guidance by constraining the a priori range of alternatives based on observations. In this work, a flexible simulation environment is used to build an ensemble of semidistributed, process-based hydrological model configurations with alternative process representations, numerical integration schemes, and model parametrizations in an integrated manner. The flexible simulation environment is coupled with an approach for dynamic identifiability analysis. The objective is to investigate the applicability of the framework to identify the most adequate model. While an optimal model configuration could not be clearly distinguished, interesting results were obtained when relating model identifiability with hydro-meteorological boundary conditions. For instance, we tested the Penman-Monteith and Shuttleworth & Wallace evapotranspiration models and found that the former performs better under wet and the latter under dry conditions. Parametrization of model structures plays a dominant role as it can compensate for inadequate process representations and poor numerical solvers. Therefore, it was found that numerical solvers of high order of accuracy do often, though not necessarily, lead to better model performance. The proposed coupled framework proved to be a straightforward diagnostic tool for model building and hypotheses testing and shows potential for more in-depth analysis of process implementations and catchment functioning.

1. Introduction

Computer models are imperfect abstractions and simplifications of the real world transferred into computer code. As such, they necessarily impose uncertainties when simulating a certain process or the evolution of a variable of interest. In surface hydrology, one main objective of modeling is to transfer a precipitation signal into a discharge hydrograph at a certain river section. However, the diversity of landscapes, data sets, and specific research objectives led to the development of a large number of different hydrological models. These can vary in their conceptualization, how and to which degree of realism hydrological processes are represented, the discretization of a landscape into spatial model units, model runtime, initialization efforts, the number of parameters, whether parameters should be calibrated or not, and under which environmental conditions they are appropriate simulation tools (e.g., Clark et al., 2011; Fenicia et al., 2016; Weiler & Beven, 2015).

The structure of a model is related to the perceptual model of the real-world system and therefore reflects the system understanding of the model developer, which is in turn based on evidence from observations and experience (Beven, 2009; Wrede et al., 2015). Consequently, structural model uncertainties are associated with the general conception of a model, incorporated mathematical equations, and computer code (Gupta et al., 2012). Furthermore, model developers are usually confronted with a number of ambiguities, such as multiple equally plausible equations for a certain process. In addition, input and output of the model, state variables, calibration parameters, and the scale of operation need to be defined. Eventually, the applicability of the model needs to be proven in case studies. Model development is therefore problem or even catchment specific, where no straightforward solutions exist and compromises need to be made (Fenicia et al., 2016;

Höge et al., 2018; Jakeman et al., 2006). Consequently, structural uncertainties can be attributed to misconceptions of the general system, a lack of process understanding, spatial and temporal scaling issues, subjectivity and ambiguity regarding important decisions during model building, and random programming errors.

Flexible model frameworks have been recognized as important tools to address structural uncertainty (Clark et al., 2008, 2015; Fenicia et al., 2011; Kneis, 2015; Knoben et al., 2019). They offer a common infrastructure for the straightforward exchange of model structures, while the preparation of input data and handling of output files are independent from the specific model configuration. In that way, they provide efficient tools for hypothesis testing and process understanding. These frameworks, however, have not been fully exploited as often single or few model structures have been compared (e.g., Fenicia et al., 2016; Kavetski & Fenicia, 2011). Moreover, this type of analysis has found limitations due to interactions between the input factors and nonlinearities in the model response (Saltelli & Annoni, 2010).

The use of more complex diagnostic and prognostic modeling tools in a fully integrated way has been recognized as a promising approach (Pianosi et al., 2016; Uusitalo et al., 2015). Among others, in the last decades frameworks of global sensitivity analysis (GSA) have been identified as important tools for model assessment and improvement (e.g., Baroni & Tarantola, 2014; Pianosi & Wagener, 2016; Savage et al., 2016). Specifically, by means of GSA, it is possible to investigate the importance of different uncertain model elements and their interaction (Günther et al., 2019; Savage et al., 2016; Stahn et al., 2017). Complementary frameworks of identifiability analyses have been used to identify adequate model configurations and parametrizations (e.g., Ajami et al., 2007; Coxon et al., 2014; Herman et al., 2013; Mustafa et al., 2020; Wagener et al., 2003). Identifiability analysis relates to sensitivity analysis in a way that it tries to reduce the uncertainty of model output by constraining the a priori range of sensitive model elements based on additional information, such as observations (Ghasemizade et al., 2017; Guillaume et al., 2019).

Difficulties to implement the aforementioned approaches are related to the need of a specific sampling design or computational burden. As an alternative, Monte Carlo (MC) filtering approaches, also referred to as regional sensitivity analysis (RSA), can be used to overcome these limitations (Spear & Hornberger, 1980). These approaches have been used for sensitivity analysis (Tang et al., 2007) and identifiability analysis (Wagener et al., 2003) but only focusing on parameters.

In this study, we extend the use of the MC filtering approach as implemented in the DYNIA framework by Wagener et al. (2003) to account also for model structures, and we explore its capability for model building. This shall be achieved by coupling it with a flexible model environment. The aim of this study is to identify the optimal model configuration with respect to parametrization, process representation, and numerical solver of the underlying ordinary differential equations (ODEs). By employing a flexible modeling platform, process equations can be easily exchanged, and the integration of ODEs is separated from the process implementations. Moreover, instead of using simplified conceptual approaches as many other studies, more complex process-based representations are employed. In addition, spatial and temporal variability will be considered by employing dynamic identifiability analysis (DYNIA) over catchments with different hydro-meteorological characteristics. It is hypothesized that this will provide more insights on model functioning and process behavior from a set of equally plausible process-based model structures tailored to the hydro-meteorological conditions of the area under investigation. The following specific research questions are addressed:

1. How well identifiable is a set of equally likely process representations, numerical ODE solvers, and parameter realizations?
2. How does identifiability vary over time with different meteorological conditions?
3. How consistent is the pattern for different parts of a catchment with varying hydrological conditions?

The paper is structured as follows: Section 2 introduces the general framework for flexible model identification. In section 3, the flexible model environment together with a case study are presented to evaluate the proposed framework of model identification. The results of the case study and a discussion of their implications are given in sections 4 and 5, respectively. The final conclusions are presented in section 6.
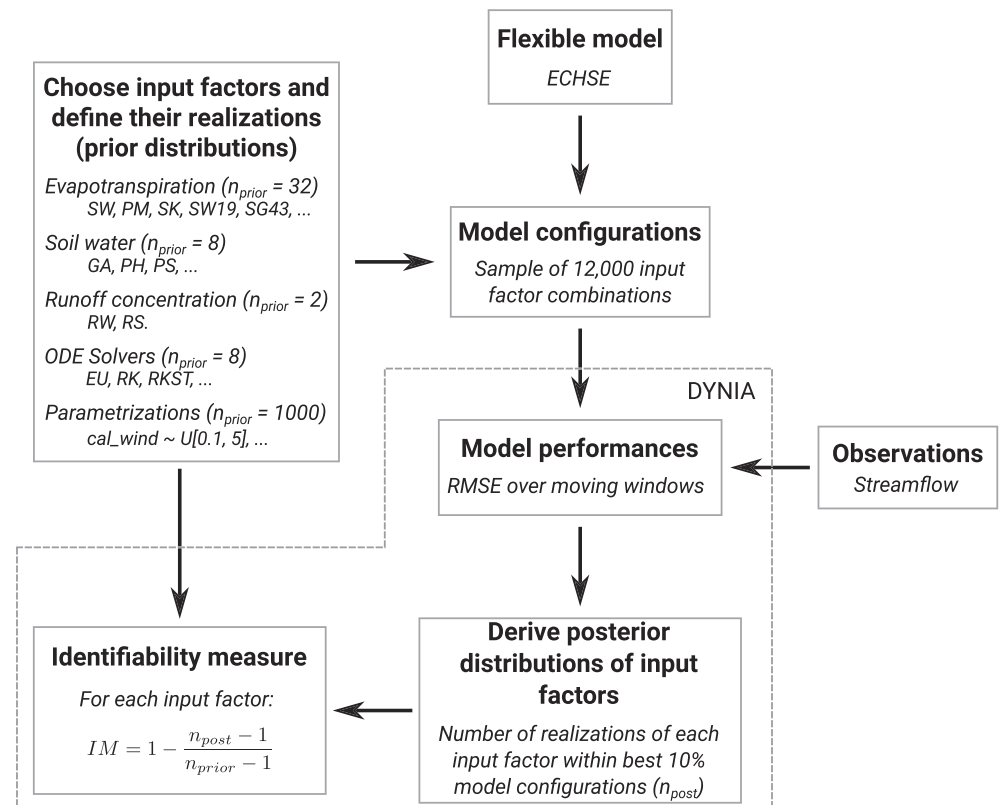
Figure 1. Outline of the framework introduced in this paper. In **bold** the generic steps are presented (section 2), and in *italics* the implementations within this case study are highlighted (section 3).

2. Framework for Flexible Model Identification

Figure 1 outlines the proposed framework for model identification. The generic part is described in the following, while the application within a specific case study is presented in section 3.

In order to determine the most adequate model structure, multiple alternatives need to be tested against each other. This is best achieved by implementing them in a single flexible environment. In the past, several such model platforms have been developed (Clark et al., 2008, 2015; Fenicia et al., 2011; Kneis, 2015; Knoben et al., 2019). They allow for the rapid exchange of model structures independent of specific input data and output requirements. As they differ in their features, such as incorporated process formulations, underlying model conception, programming language, or the degree of flexibility in adding further process representations, the user has to choose a framework according to the requirements of the specific case study and research questions.

Denoted as *input factors*, in this study as in some others (e.g., Pianosi et al., 2016), are elements of a model that can be flexibly adapted before a model simulation. These include, for instance, parameters, input data sets, or process representations. Input factors are composed of *realizations*, such as a specific parameter value, input time series, or process formula. The definition of input factors and their realizations depends on the goal of a study. To define the *prior distribution* of an input factor, assumptions about the underlying distribution are required, for example, normal distribution with mean and standard deviation or uniform distribution within certain limits. Within this framework, input factors should be selected and defined in a way that it is known from experience that variation in the input factors will lead to variation in the output of the model, that is, that the input factors exhibit some degree of sensitivity. Furthermore, each input factor realization should represent a plausible working hypothesis for catchment behavior as otherwise the result of identifiability analysis would be biased.

Studies solely focusing on parameters typically define input factors in a way that they are continuously distributed. To account for discrete nonscalar input factors, as becomes necessary when focusing on, for

example, alternative process equations, each realization of an input factor is associated with a scalar value. In that way, the set of considered realizations forms a discrete uniform distribution. This is a generic strategy for the quantification and ranking of different input factors in the context of environmental model application and allows for the incorporation of both numeric and nonscalar input factors (Baroni & Tarantola, 2014; Lilburne & Tarantola, 2009; Plischke et al., 2013; Savage et al., 2016).

A specific combination of realizations of each input factor determines a *model configuration*. In order to reduce the amount of work, model preprocessing, such as landscape discretization and preparation of input data, should be independent of a specific model configuration. This also avoids side effects where specific preprocessing steps unintentionally influence the result of the subsequent identifiability analysis. On the other hand, certain preprocessing steps can as well be explicitly considered as input factor within the framework.

Following the Bayesian philosophy of statistics, the goal of this framework is to take further observations into account and update the prior distributions of the input factors, that is, derive their *posterior distributions*. In the case of complex, nonlinear hydrological models, no analytical solutions exist for this problem. Instead, MC-based strategies are efficient means to analyze the multitude of competing model configurations in a systematic and computationally feasible manner. For this framework, the DYNIA approach by Wagener et al. (2003) was adopted. It is based on the RSA (also called MC filtering) approach by Spear and Hornberger (1980) by partitioning an ensemble of model configurations into behavioral (acceptable model performance) and nonbehavioral sets. The DYNIA framework uses the same basis of RSA with the aim of dynamically assessing the information content of input factors over moving time windows. In that way, the influence of varying hydro-meteorological conditions, for example, wet and dry periods, on the identifiability of model structures can be assessed. The posterior distribution of each input factor is derived by analyzing the set of behavioral model configurations, that is, determining the frequencies of realizations in the case of discrete-valued input factors.

Identifiability is assessed by comparing the prior and posterior distributions of input factors. This can be done, for instance, by defining some metric. For our framework, we propose the identifiability measure by comparing the number of realizations in the prior and posterior distributions of each input factor (n_{prior} and n_{post}, respectively):

$$IM = 1 - \frac{n_{post} - 1}{n_{prior} - 1} \text{ with } n_{prior}, n_{post} \in \mathbb{N}_{>0}. \tag{1}$$

IM ranges between zero and one. A value of zero is obtained when n_{post} has the same value as n_{prior}, that is, when all realizations of an input factor defined in its prior distribution are still present in the posterior distribution. This means the specific input factor is not identifiable as all its realizations can lead to behavioral model performance. In contrast, IM is one if n_{post} is one, which happens when only one realization of an input factor is left in its posterior distribution. In that case, the input factor is well identifiable as only one specific realization is associated with behavioral model performance. Note that n_{prior} and n_{post} cannot be zero as each model configuration is composed of one realization of each input factor and therefore at least one realization of each input factor will also occur in the set of behavioral model configurations. If there should be no behavioral model configuration distinguishable, this shows that either the performance criteria for the partition of model configurations are too strict or the defined input factor realizations are no meaningful representations of catchment behavior.

More information on identifiability can also be obtained by analyzing the posterior distributions in greater detail. This can be achieved by analyzing the single best model configuration or visual inspection of the posterior distributions, that is, the posterior frequencies of realizations.

In the following section, a case study is presented to illustrate the functionality and capabilities of the proposed framework.

3. Case Study

3.1. Study Area

For the model simulations, the Isábena catchment in northeastern Spain was selected (Figure 2). The watershed comprises an area of about 425 km^2 and is located at the southern edge of the Pyrenees. It is a mesoscale
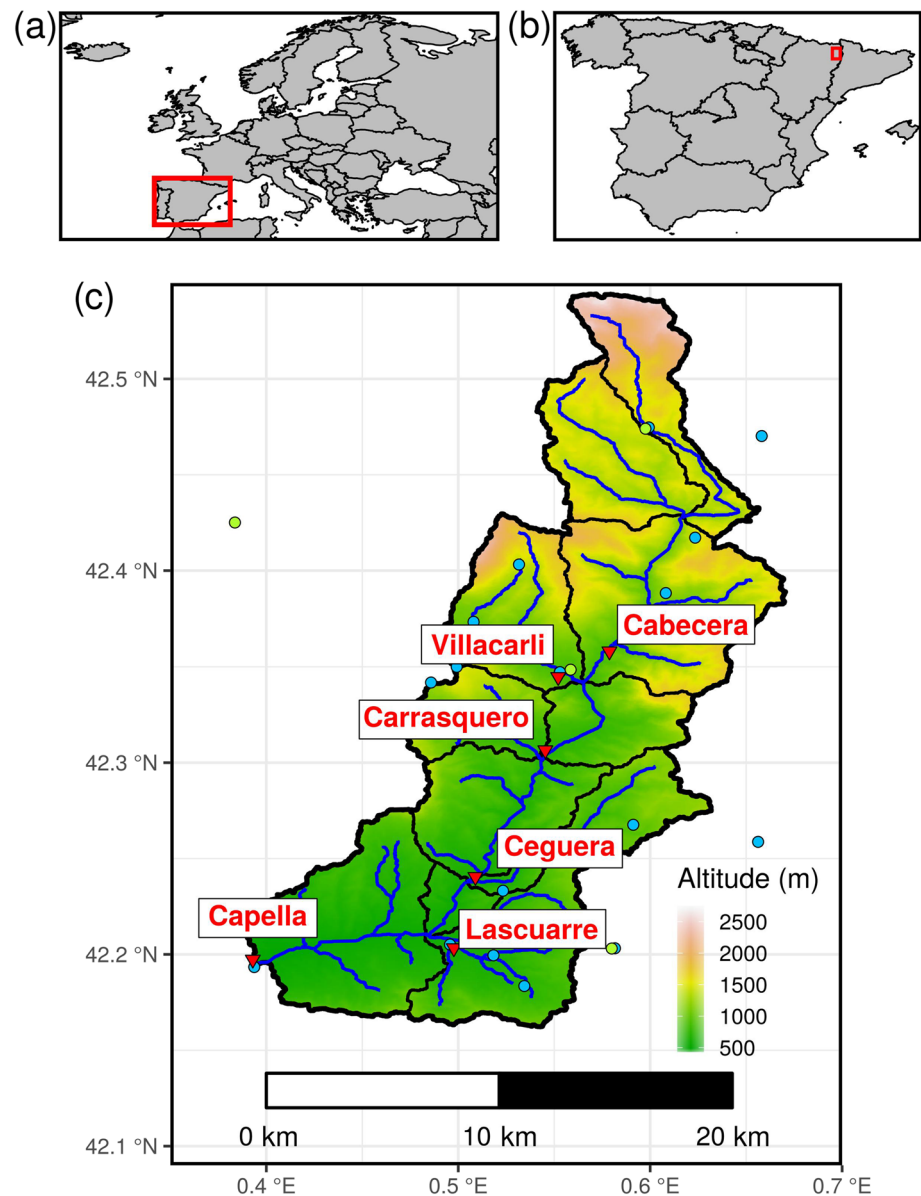
Figure 2. Overview over the Isábena catchment (c) and its location within Spain (b) and Europe (a). In panel (c), background color is based on the DEM used for model initialization, thin black lines within the catchment outline subbasins (delineated with the lumpR software and corresponding to the studied subcatchments except for Capella, which comprises the whole study area, and Cabecera extending over the full northeast), red triangles mark the position of discharge gauges, and blue and green points show gauges of rainfall and other meteorological variables, respectively. Note that not all stations used in this study are visible due to feature overlays and because some stations are located slightly outside the plot.

headwater catchment of the Ebro river basin. Therefore, the mountainous topography is characterized by a heterogeneous relief with altitudes ranging from about 500 to 2,700 m. Consequently, precipitation is spatially heterogeneous with annual sums ranging from 450 mm in the lowlands up to 1,600 mm in the upland parts of the catchment and a spatial average of about 770 mm. The climate is continental Mediterranean (relatively wet and cold), characterized by Atlantic and Mediterranean influences (García-Ruiz et al., 2001).

The hydrological regime is influenced by rain and snow. Floods may occur in spring as a consequence of precipitation events amplified by snowmelt or in late summer and autumn caused by convective thunderstorms. Mean annual discharge at the catchment outlet is 4.1 m³/s, while a maximum instantaneous value of 370 m³/s has been observed (period 1945–2009). Minimum discharge can be less than 1 m³/s, but the

Table 1
Characteristics of the Study Area and Delineated Subcatchments Referring to the Analysis Period 2013 to 2015

Gauge	A (km^2)	Gauge elevation (m a.s.l.)	P (mm/year)	Q (m^3/s)	RC (%)
Villacarli	41	866	673	0.30	27
Cabecera	145	841	901	2.09	46
Carrasquero	25	762	673	0.44	68
Ceguera	29	598	599	0.13	21
Lascuarre	44	565	572	0.06	6
Capella	424	490	743	4.57	40

Note. Capella is located at the outlet of the study area, while all other gauges refer to individual subcatchments; A, catchment area referring to the model setting of this study; P, average annual rainfall (spatially interpolated from station data, see text); Q, average discharge; RC, runoff coefficient.

river never falls completely dry. The study area is not regulated, thus the hydrological regime is determined by natural factors only. It is mainly composed of deciduous woodland, agriculture, pasture, and bushes in the valley bottoms with evergreen oaks and pines. Table 1 shows further hydrological statistics of the area including selected subcatchments.

The area has been investigated in many research projects, including intensive hydro-sedimentological monitoring (see Bronstert et al., 2014, and references therein). Consequently, a rich data set exists with relatively high spatial coverage of meteorological and discharge measurements. In addition, the mountainous catchment consists of several subcatchments with varying hydrological and meteorological conditions, which provide different settings to apply the proposed framework of this study and investigate the influence of hydro-meteorological conditions on the identification of model structures. Recorded time series data have recently been published and described by Francke, Foerster, et al. (2018).

3.2. The Flexible ECHSE

The ecohydrological simulation environment (ECHSE) is a software designed for flexible model building (Kneis, 2015). The environment consists of a generic part and the model engines. The former is the basis of each ECHSE-based model and defines the format and general structure of input and output files, provides data types for model development (state variables, parameters, input, and output variables), and contains methods for the actual simulation, including a number of ODE solvers. The latter is the actual model and consists of code provided by the user, that is, the actual process formulations. Applications of the ECHSE environment can be found in Kneis et al. (2014), Abon et al. (2016), and Kneis et al. (2017).

For this study, a new model engine has been designed for ECHSE, which is oriented at the semidistributed, process-based hydrological and sedimentological model WASA-SED (Güntner & Bronstert, 2004; Mueller et al., 2010). The key feature is an efficient hierarchical landscape disaggregation scheme over multiple scales. These spatial scales include subbasins, which contain landscape units (LUs) as representative hillslopes, further characterized by different terrain components (TCs) that constitute segments of the hillslope. The latter in turn consist of several soil-vegetation components (SVCs) described by a characteristic soil profile with specific vegetation cover. Each spatial unit is associated with areal shares of certain spatial units of the next lower scale, that is, except for the subbasins, the spatial units have no explicit spatial reference. Explicitly represented processes include evapotranspiration, infiltration, both infiltration-excess and saturation-excess runoff, soil water movement, exfiltration, as well as lateral runoff redistribution. Groundwater is represented by a simplified linear storage approach. River flow is described by a simple unit hydrograph routing with evaporation as only source for transmission losses.

The spatial discretization scheme allows for spatially semidistributed hydrological simulations over large scales up to about 100,000 km^2 with acceptable computational burden, while taking small-scale processes along hillslopes explicitly into account. This multiscale feature puts particular emphasis on dryland hydrological peculiarities, such as evaporation from patchy vegetated surfaces, channel transmission losses, infiltration excess in case of crusted soil, and limited connectivity of runoff phenomena (see Bronstert et al., 2014 for more details). The model has been successfully applied in the present study area as well as several other dryland regions in Spain (Bronstert et al., 2014; Francke, Baroni, et al., 2018; Mueller et al., 2009, 2010), Brazil (de Araújo & Medeiros, 2013; Krol et al., 2011; Medeiros et al., 2010, 2014; Pilz, Delgado, et al., 2019),

and India (Jackisch et al., 2014). Therefore, it has been chosen as highly plausible model hypothesis for the selected study area and enhanced by further process representations and ODE solvers within ECHSE.

3.3. Data and Model Initialization

All model simulations are based on the same initialization procedure, including the discretization of the study area into model units, parametrization of soil and vegetation, and the preparation of meteorological inputs. Landscape discretization and preparation of model input files was performed using the lumpR software, a package for the free and open-source environment R (Pilz et al., 2017). The software has been designed for the initialization of models incorporating hillslope-based discretizations schemes, such as WASA-SED, and has the option to generate input files for ECHSE. The basis of the algorithms were a $15 \, m \times 15 \, m$ DEM processed from ASTER raw data, a soil type, and a land-use map including parametrizations of the soil and vegetation types. The initialization procedure comprised the delineation of the catchment and subbasins (outlined in Figure 2), the derivation of further model units (the LUs, TCs, and SVCs described in section 3.2), calculation and checking of parameters, and the generation of model input files. The spatial setup consisted of 11 subbasins, 24 LUs, 72 TCs, and 93 SVCs.

Meteorological and discharge data were obtained from the data set of Francke, Foerster, et al. (2018). This includes rainfall data from 18 stations, temperature from nine stations, and solar radiation and air humidity data from two stations within or in close vicinity to the study area. Gaps in the time series were interpolated with data from neighboring stations. Furthermore, station data needed to be interpolated to the centroids of the subbasins by employing inverse distance interpolation (IDW), which was realized using the geostat R package of the ECHSE toolbox (Kneis, 2012). For model evaluation, spatially distributed discharge measurements from the Isábena river at the catchment outlet (Capella) and from five subcatchments were used (Table 1).

All experiments were conducted under the following experimental design. The simulation and analysis period covered 3 years from 1 January 2013 to 31 December 2015 with a temporal resolution of 1 day. This decision is based on a compromise between data availability and computational feasibility. On the one hand, the period should be sufficiently long to cover the hydrological catchment dynamics under different conditions. On the other hand, although subdaily measurements are available and numerical ODE solvers would be expected to be more reliable under hourly resolution, model runtimes with hourly resolution would be too long to achieve results in acceptable time, when applied in the presented framework. To bring model states into equilibrium and avoid artificial effects on outputs, a warm-up was conducted prior to any simulation run. This warm-up run consisted of iterations over 1 year (1 January 2012 to 31 December 2012) until convergence was achieved, that is, until the sum of hydrological storages at the end of a warm-up iteration deviated by less than 0.1 % from the sum of storages at the end of the previous iteration.

3.4. Input Factor Definition

In line with the proposed generic framework, the following five input factors were defined for this case study: (Ia) structural uncertainty with respect to evapotranspiration subprocesses; (Ib) uncertainty in the representation of soil water processes; (Ic) runoff concentration approaches; (II) numerical integration of underlying ODEs; and (III) parametrization. We selected the input factors as they either reflect key landscape features that have a large impact on simulation results but are hard to quantify (parametrization), reflect ambiguities in the representation of key processes (process representation), or are known to be neglected in most other studies while still exhibiting quite a large impact (ODE solvers). Table 2 provides an overview of the input factors, which are further described in the following.

3.4.1. Process Representations

For nine hydrological (sub-)processes, alternative formulas were implemented in ECHSE. For the demonstration of the proposed approach, it was decided to provide two alternatives for each process component in order to keep the computational burden within feasible limits and allow for a straightforward interpretation of results. These alternatives consist of the respective representations copied from the original WASA-SED model and an alternative. In most of the cases, approaches are similar but may vary in degree of detail. For instance, the evapotranspiration approach by Shuttleworth and Wallace (1985) is deviated from the Penman-Monteith formula but consists of a more detailed conception of resistances to account for patches of bare soil instead of assuming a homogeneous vegetation cover. While an overview of implemented processes is given by Table 2, Tables S1 and S2 in the supporting information provide the associated formulas.

Table 2
Description of the Input Factors

Symbol	Process/group	Implementation
(Ia) Evapotranspiration		
SW	Evapotranspiration	Shuttleworth and Wallace (1985)
PM	Evapotranspiration	Penman-Monteith
SK	Stomatal resistance	Saugier and Katerji (1991), equation 4
SW19	Stomatal resistance	Shuttleworth and Wallace (1985), equation 19
SG43	Roughness length	Shuttleworth and Gurney (1990), equation 43
BT	Roughness length	Brutsaert (1975)
SG42	Displacement height	Shuttleworth and Gurney (1990), equation 42
SG41	Displacement height	Shuttleworth and Gurney (1990), equation 41
AN	Clear sky radiation	Ångström formula
AL	Clear sky radiation	Allen et al. (2005), equation 19
(Ib) Soil water		
GA	Infiltration	Modified Green-Ampt (Güntner, 2002)
PH	Infiltration	Philip (1957)
PS	Percolation	SWAT approach (described in Güntner, 2002)
PR	Percolation	Simplified Richards' equation
VG	Soil water retention	van Genuchten (described in Maidment, 1993)
CB	Soil water retention	Campbell (described in Maidment, 1993)
(Ic) Runoff concentration		
RW	Runoff concentration	Physically based: lateral redistribution along a hillslope (Güntner, 2002)
RS	Runoff concentration	Conceptual (requires calibration): delay of runoff concentration by linear storage approach
(II) ODE solvers *(each can be applied with [CONS] or without [FREE] solution constraints)*		
EU	First-order accurate	Explicit Euler, fixed time step lengths
RK	Fourth-order accurate	Explicit Runge-Kutta Cash-Karp, fixed time step lengths
RKST	Fourth-order accurate	Explicit Runge-Kutta Cash-Karp, adaptive time step lengths
BDF	Dynamic accuracy	Backward differentiation formula in Nordsieck form (predictor-corrector), multistep
(III) Parametrization		
cal_wind	Evapotranspiration	Windspeed correction factor, $\mathcal{U}(0.1, 5)$
cal_ks	Soil water movement	Correction factor for saturated hydraulic conductivity, $\mathcal{U}(0.01, 100)$
cal_kfbed	Groundwater recharge	Correction factor for bedrock conductivity, $\mathcal{U}(0.01, 100)$
Phil_cal	Infiltration	Parameter for Philip's equation (m/s), $\mathcal{U}(0.2, 1)$
str_surf	Runoff concentration	Surface runoff retention factor, $\mathcal{U}(0.001, 0.1)$
str_inter	Runoff concentration	Subsurface runoff (interflow) retention factor, $\mathcal{U}(0.01, 0.5)$
str_base	Runoff concentration	Groundwater runoff (baseflow) retention factor, $\mathcal{U}(0.1, 10)$

Note. Equations of process formulas are shown in Tables S1 and S2.

3.4.2. ODE Solvers

Hydrological model structures commonly comprise a set of ODEs to describe the evolution of state variables. These ODEs need to be integrated over discrete time steps along a model application. However, complex hydrological models typically contain nonlinear ODEs, which are analytically intractable. Consequently, numerical approximation methods, also referred to as *ODE solvers*, need to be employed, which raise a number of mathematical issues that need to be considered, such as convergence (the solver needs to converge to a solution), order of accuracy (how well solutions are approximated), and stability (the solution needs to be stable and must not oscillate). Even though it has been shown that the use of oversimplified ODE solvers can induce high uncertainties and lead to wrong conclusions (Gupta et al., 2012; Kavetski & Clark, 2011; Schoups et al., 2010), this topic still receives little attention in the field of surface hydrology. Consequently,

an aspect of this study was to explicitly consider uncertainty related to the ODE solver as part of the identifiability analysis. To do so, several ODE solvers were implemented in ECHSE using the external *GNU Scientific Library* (Galassi et al., 2017). Four different solvers were considered in this study, varying in the order of accuracy and stability (Table 2). More detailed information about ODE solvers and the selected implementations are provided in the supporting information.

3.4.3. Parametrization

To reflect uncertainty in the parametrization, seven parameters were considered whose general features are hard or not measurable at the targeted scale. The associated value ranges and distributions were selected based on experience from previous applications of the WASA-SED model in the study area and similar environments (Bronstert et al., 2014; Francke, Baroni, et al., 2018; Mueller et al., 2009, 2010). These parameters influence different components of the model including evapotranspiration, soil water movement, groundwater recharge, infiltration, and runoff concentration. All parameters are globally effective parameters (e.g., as factors; see Table 2) and independent from the spatial model setup, that is, the parameters are used to scale the values for the entire catchment. For instance, parameter cal_ks affects the saturated hydraulic conductivity of all soil units.

It should be noted that four parameters (Phil_cal, str_surf, str_inter, and str_base) are associated with specific process implementations while the remaining three parameters are effective for all process implementations. This exhibits correlation among the input factors, which needs to be considered when choosing a specific algorithm for identifiability or sensitivity analysis.

3.5. Prior Distribution and Model Configurations

The prior distributions of the input factors are defined by their realizations and assumed underlying distribution. Table 3 provides an overview where each line represents a specific input factor realization. Each input factor realization was associated with a number in order to obtain numerical input factors with discrete distributions (column ID in Table 3). With respect to the prior distribution it was assumed that the input factors are uniformly distributed, that is, each realization of a specific input factor was assumed to be an equally likely representation of catchment characteristics or processes and was given equal weight.

Regarding the input factors reflecting process representations (input factors Ia to Ic), the realizations were defined by all possible combinations of the two alternative representations for each subprocess. This resulted in $n_{prior} = 2^n$ realizations in the prior distribution for each input factor, where n is the number of considered subprocesses: that is, 5, 3, and 1 subprocesses for evapotranspiration, soil, and runoff concentration, respectively, resulting in 32, 8, and 2 realizations (n_{prior}) for input factors Ia, Ib, and Ic, respectively. The ODE solvers (input factor II) were combined with the possibility of constrained or freely evolving solutions, which resulted in eight realizations for this input factor (four solver variants times two variants of constraints). To represent the parameter space (input factor III), 1,000 realizations were taken by Latin hypercube sampling from the log-transformed and uniformly distributed parameter spaces (except for Phil_cal, which was not transformed).

The resulting model configurations are derived by the combination of input factor realizations, that is, each model configuration consists of one specific realization of each input factor. Because of computational constraints, of the 4,096,000 possible model configurations, 12,000 samples were randomly drawn. The model was subsequently evaluated for each sample. It should be noted that this framework involves no explicit parameter calibration of the model. The resulting ensemble of model runs was then analyzed following the DYNIA framework as follows.

3.6. Posterior Distribution and Dynamic Identifiability Measure

The root mean square error (RMSE) was selected as performance metric. For the dynamic analysis, a performance value was calculated for each simulation day d over a moving window of the width $2w + 1$ as follows:

$$RMSE(d) = \sqrt{\frac{1}{2w + 1} \sum_{i=d-w}^{d+w} \left(q_s(i) - q_o(i)\right)^2}, \tag{2}$$

where q_s is the simulated and q_o the observed discharge and w a parameter defining the window size. For this study, w was set to a value of 15 days, which resulted in a total moving window size of 31 days.

Table 3
Input Factor Realizations

ID	Implementations				
Ia: Evapotranspiration					
1	PM	SW19	BT	SG41	AN
2	SW	SW19	BT	SG41	AN
3	PM	SK	BT	SG41	AN
4	SW	SK	BT	SG41	AN
5	PM	SW19	SG43	SG41	AN
6	SW	SW19	SG43	SG41	AN
7	PM	SK	SG43	SG41	AN
8	SW	SK	SG43	SG41	AN
9	PM	SW19	BT	SG42	AN
10	SW	SW19	BT	SG42	AN
11	PM	SK	BT	SG42	AN
12	SW	SK	BT	SG42	AN
13	PM	SW19	SG43	SG42	AN
14	SW	SW19	SG43	SG42	AN
15	PM	SK	SG43	SG42	AN
16	*SW*	*SK*	*SG43*	*SG42*	*AN*
17	PM	SW19	BT	SG41	AL
18	SW	SW19	BT	SG41	AL
19	PM	SK	BT	SG41	AL
20	SW	SK	BT	SG41	AL
21	PM	SW19	SG43	SG41	AL
22	SW	SW19	SG43	SG41	AL
23	PM	SK	SG43	SG41	AL
24	SW	SK	SG43	SG41	AL
25	PM	SW19	BT	SG42	AL
26	SW	SW19	BT	SG42	AL
27	PM	SK	BT	SG42	AL
28	SW	SK	BT	SG42	AL
29	PM	SW19	SG43	SG42	AL
30	SW	SW19	SG43	SG42	AL
31	PM	SK	SG43	SG42	AL
32	SW	SK	SG43	SG42	AL
Ib: Soil					
1	PH	PS	VG		
2	*GA*	*PS*	*VG*		
3	PH	PR	VG		
4	GA	PR	VG		
5	PH	PS	CB		
6	GA	PS	CB		
7	PH	PR	CB		
8	GA	PR	CB		
Ic: Runoff concentration					
1	*RW*				
2	RS				

Table 3 *(continued)*

ID							
			Implementations				
II: ODE solvers							
1	EU	FREE					
2	*EU*	*CONS*					
3	RK	FREE					
4	RK	CONS					
5	RKST	FREE					
6	RKST	CONS					
7	BDF	FREE					
8	BDF	CONS					
III: Parametrization							
ID	cal_wind	cal_ks	cal_kfbed	Phil_cal	str_surf	str_inter	str_base
1	2.130	1.980	0.091	0.455	0.003	0.147	5.720
2	0.635	15.911	3.374	0.796	0.025	0.176	0.980
3	1.844	0.029	0.022	0.290	0.022	0.065	0.161
…	…	…	…	…	…	…	…
998	0.185	0.021	3.823	0.226	0.093	0.290	2.514
999	0.400	1.451	0.136	0.433	0.011	0.030	4.773
1000	4.455	0.458	0.147	0.702	0.045	0.076	2.340

Note. Highlighted realizations mark implementations in the original WASA-SED model. For symbols of process and ODE solver representations, see Table 2.

Based on the model performances, the posterior distributions of the input factors have to be determined. There are different strategies how this can be achieved. In this study, as in the original DYNIA paper of Wagener et al. (2003), an informal Bayesian approach as in the GLUE framework (Beven & Binley, 1992) was employed by grouping the model performances into a behavioral group by arbitrarily selecting the best 10% performing model configurations with respect to RMSE measures. Consequently, the group of behavioral (i.e., acceptable) model runs generically consisted of 1,200 values out of the 12,000 samples.

The frequencies (i.e., the number of occurrences) of input factor realizations within the set of behavioral model runs constitute the posterior distributions. The number of realizations with frequencies greater than zero gives a value of n_{post} for each input factor. Together with n_{prior}, the identifiability measure IM was calculated for each input factor following Equation 1. Note that, when comparing the IM values of the input factors, the different characteristics and number of realizations have to be kept in mind, which define each input factor. The values are therefore not directly comparable.

These processing steps were applied over both the full simulation period (static identifiability analysis) in order to identify the optimal model structure and a moving window over the simulation period (DYNIA) to study the influence of varying meteorological conditions. Spatial variability, and hence the influence of hydrological characteristics, was investigated by analyzing the full study area and individual subcatchments with distinct characteristics.

In addition to the analysis of IM values, the posterior distributions of the input factors were examined in greater detail by visual inspection. This was done by counting the occurrences (also denoted as frequencies) of individual input factor realizations in the posterior distribution. The higher the frequency, the more often a realization contributes to adequate model performance. In that way, even for input factors with zero identifiability (all realizations occur at least once in the set of behavioral configurations), information can be derived how well certain realizations are performing.

3.7. Convergence and Robustness

Credibility of the applied sampling of model configurations was assessed by means of (i) analysis of convergence (is a sample size of $N = 12,000$ configurations sufficient?) and (ii) analysis of robustness (is the sampled posterior distribution independent of specific samples?). Convergence is typically analyzed by
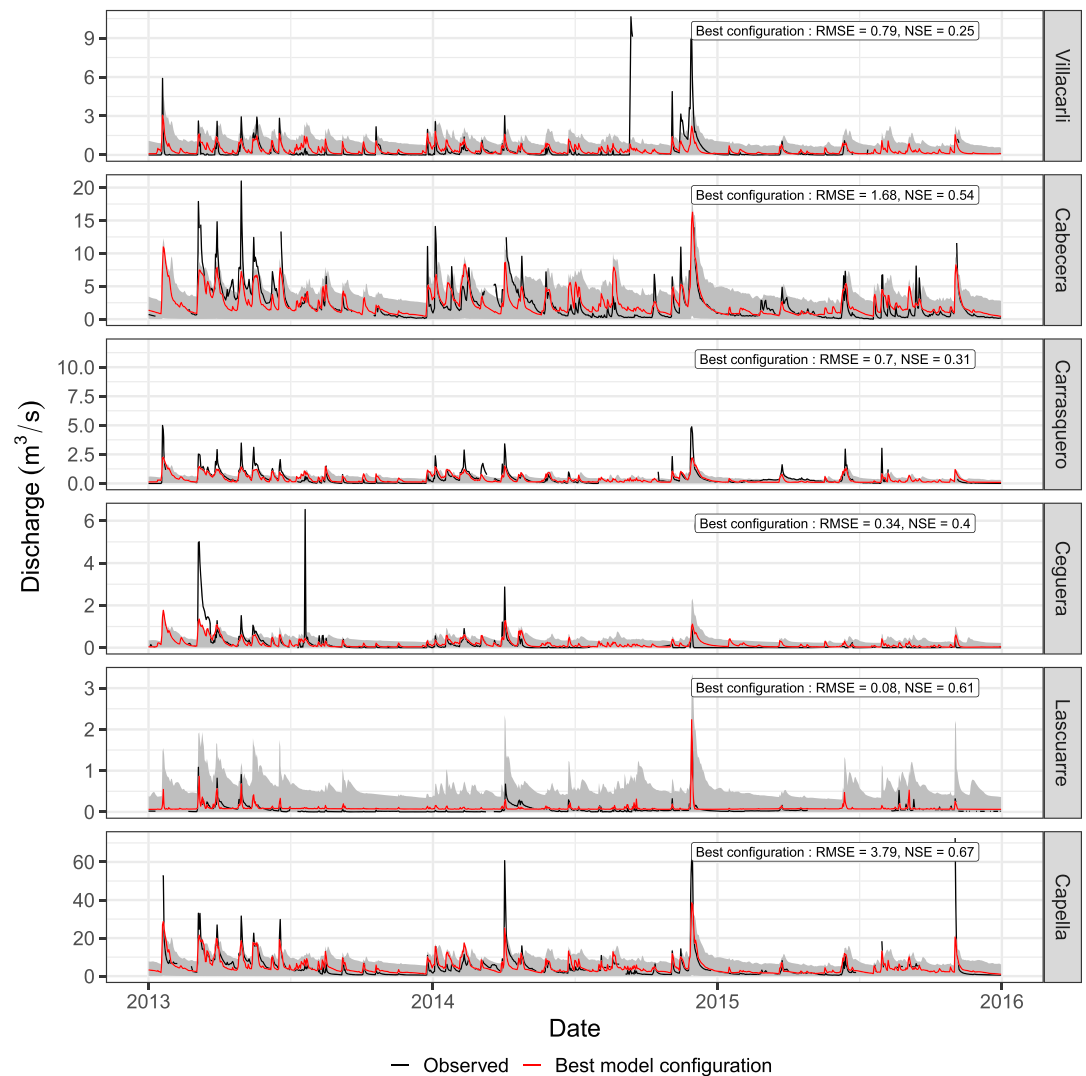
Figure 3. Discharge simulations in comparison to observations for each studied subcatchment and the whole study area (Capella). The gray area illustrates the 90% probability range of all 11,968 successful model configurations reflecting the prior distribution of input factors.

subsampling, that is, the computation of the target variable for increasing N from the original sample. This was coupled with a bootstrapping approach to analyze robustness, which consists of sampling from the N model results N_b times with replacement and subsequent recalculation of the target variable. The range over the values of the target variable from the N_b bootstrapping procedures is a measure of robustness. In that way, no additional model evaluations were required (Pianosi et al., 2016).

In this study, N_b was set to a value of 1,000. The target variable, for which convergence and robustness were assessed, was the posterior distribution of input factors (namely, n_{post}) and the identifiability measure (IM).

4. Results

4.1. Model Simulations
4.1.1. Model Errors
To carry out the experiments with the different model structures and parametrizations, the ECHSE environment was run 12,000 times according to the sampled model configurations. Individual simulation runtimes varied between less than a minute and more than 1 hr, mainly depending on the ODE solver and workload of the high-performance cluster, where the simulations were carried out. During model evaluation, 32 runtime errors occurred, that is, for the further analyses, 11,968 model configurations have been considered. The
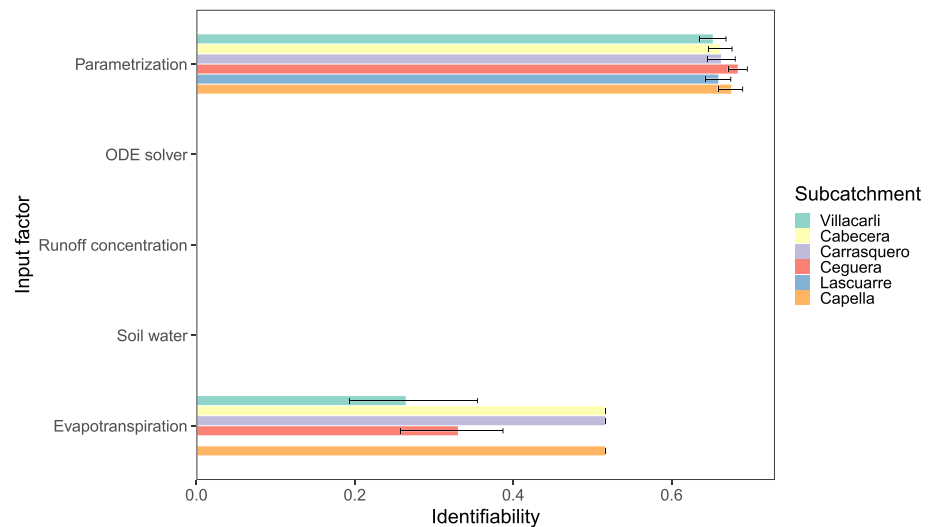
Figure 4. Identifiability measures for the different input factors and subcatchments. Black errorbars represent the 95% confidence interval estimated via bootstrapping.

runtime errors occurred exclusively when employing the backward differentiation formula (BDF) without solution constraints as ODE solver. This is not surprising as it is the only implicit solver considered in this analysis, which are known to be limited by the given constraints of accuracy and maximum number of iterations. However, the number of errors is low, even in comparison to all runs with the BDF solver ($\sim$3,000), and thus no impact on the general conclusions of this study are expected.

4.1.2. Discharge Simulation

Observed discharge values for most time steps fall into the 90% probability range of the 11,968 prior model configurations (Figure 3). This shows that the range of prior model configurations is able to capture the observed streamflow response. However, especially large discharge peaks are often underestimated by the model ensemble as well as by the single best model runs. Nash-Sutcliffe values range from 0.25 for the Villacarli headwater catchment to 0.67 for the basin outlet at Capella. The RMSE for the best performing model configuration generally lies in the order of magnitude of average discharge observations (compare with Table 1). However, especially for the smaller subcatchments (Villacarli, Carrasquero, and Ceguera) RMSE is relatively high. In general, the model is performing better for the larger (sub-)catchments (Cabecera and Capella).

4.2. Static Identifiability Analysis

After determining the posterior model configurations and associated input factor distributions, it was possible to determine the identifiability measure, first for the whole analysis period. Parametrization and evapotranspiration turned out to be the only input factors with a certain degree of identifiability (Figure 4). For the latter, however, spatial differences exist. At gauge Lascuarre, with its comparatively low runoff coefficient (see Table 1), no superior model structures with respect to evapotranspiration processes could be distinguished from the set of a priori structures. For the other input factors, all realizations of the prior distribution also occur in the posterior distribution resulting in identifiability measures of zero for all gauges.

More detailed information can be obtained by analyzing the posterior distributions of input factors (Figure 5). The Penman-Monteith approach (PM, odd-numbered IDs; see Table 3) has the highest frequency values in the posterior distribution of that input factor. For the other evapotranspiration subprocesses, no obvious patterns can be seen. In contrast to the other gauges, at Lascuarre, the Shuttleworth-Wallace approach (SW, even-numbered realizations) achieves high posterior frequencies. That means that the PM approach (almost exclusively constituting the best 10% of configurations) is clearly the superior evapotranspiration model for most parts of the study area, regardless of the choices for the other evapotranspiration subprocesses, while for Lascuarre, the SW approach excels.

Regarding the soil water processes, the approach by van Genuchten (VG, IDs 1 to 4) obtained the highest posterior frequencies at most gauges. Again, there is a different picture for gauge Lascuarre, where no
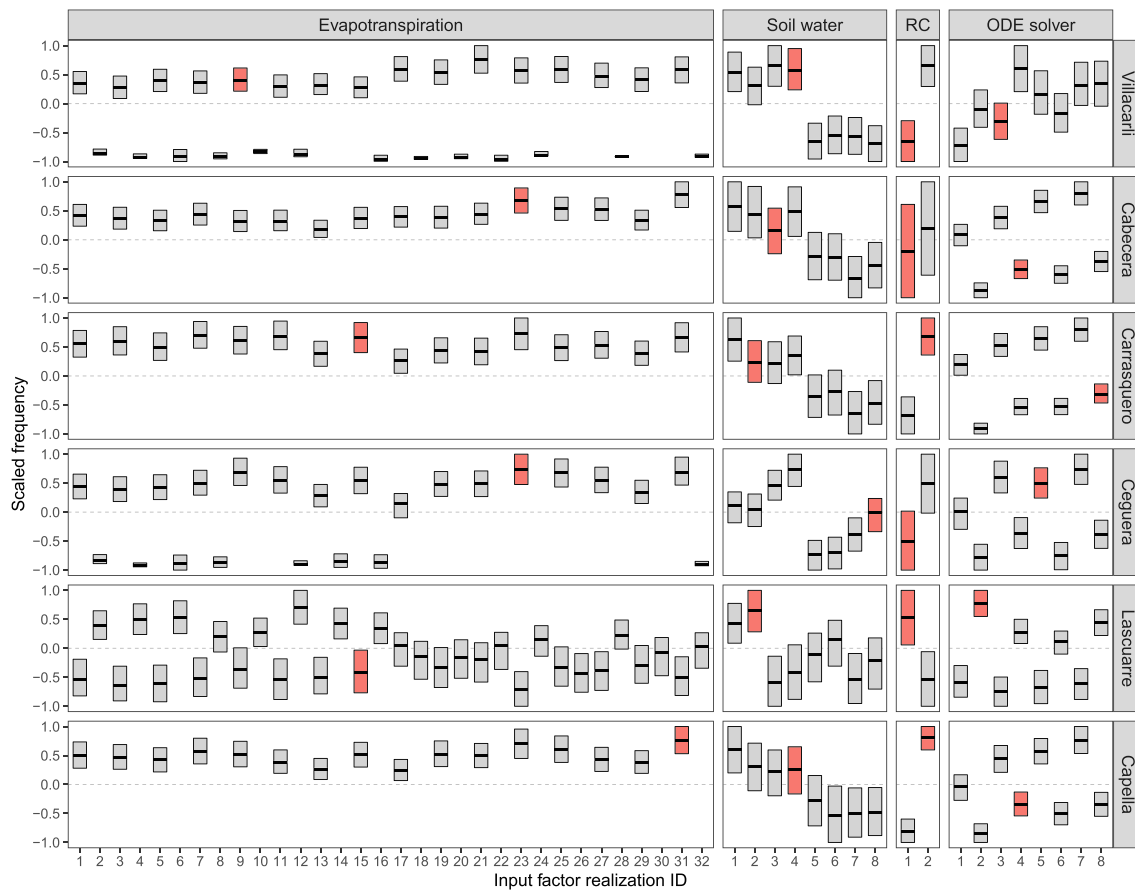
Figure 5. Posterior distributions of input factors for each subcatchment. Note that the distributions are discrete and shown as, for better interpretability, frequency values scaled to [−1, 1] by dividing each frequency value by the lowest or highest value of the respective input factor depending on whether the posterior value is less or greater than the prior value. Consequently, negative (positive) values indicate posterior frequencies less (larger) than prior frequency values. The boxes represent the 95% confidence intervals estimated via bootstrapping, the black horizontal line is the bootstrap ensemble mean. Red boxes mark the realizations related to the best model configuration (red line in Figure 3). For the meaning of realization IDs, consult Table 3.

clear pattern can be distinguished. For runoff concentration, ID 2 (conceptual calibration-based approach) achieved slightly higher frequencies for all gauges but for Lascuarre.

The ODE solvers without solution constraints (odd-numbered IDs) exhibit the highest posterior frequencies for all gauges except Villacarli and Lascuarre. Moreover, solvers with higher order of accuracy obtained slightly higher values. This, however, is opposed to the findings for Lascuarre, where solvers with solution constraints (even-numbered IDs) dominate and the explicit Euler approach even achieved the highest posterior frequencies.

It should be noted that the best performing model configuration is often not associated with input factor realizations with the highest posterior frequencies. That means that, in some occasions such as ODE solver 4 for catchments Cabecera and Capella, an input factor realization is associated with the very best model performance, but, overall, configurations associated with that realization only rarely belong to the set of behavioral configurations.

Despite the high identifiability measure, the posterior distributions of input factor parametrization show distinct peaks only for three of the seven parameters (cal_kfbed [groundwater recharge], cal_ks [soil water movement], and cal_wind [evapotranspiration]; Figure 6). The other parameters are relatively equally distributed and show little distinction from the prior distribution. Differences among gauges are small. The only exception is parameter cal_wind, for which at gauge Villacarli a peak in the posterior distribution for values greater than zero (resulting in increasing values of evapotranspiration) can be seen, while for the other gauges peaks are at smaller values (translating into less evapotranspiration). Parameters cal_kfbed
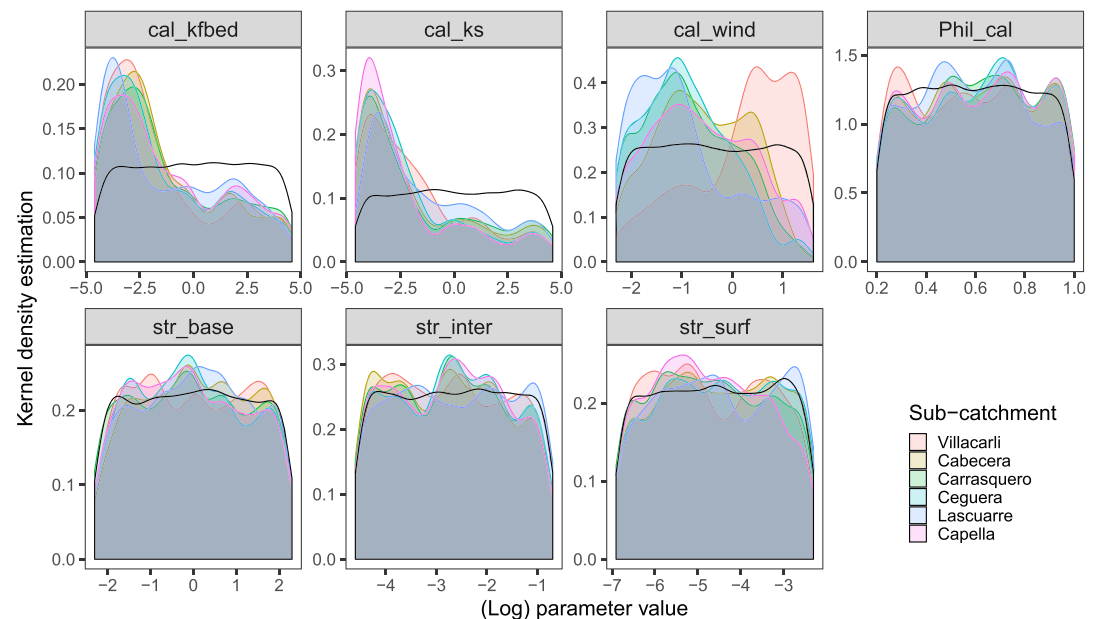
Figure 6. Kernel density plots of the posterior (colored) and prior (black line) distributions of parameters. Shown are log-transformed parameter spaces except for Phil_cal. For parameter symbols, see Table 2.

and cal_ks show a tendency toward smaller values, which results in less groundwater recharge and delayed subsurface runoff generation.

4.3. DYNIA

Identifiability may change over time depending on the current boundary conditions (Figure 7). This holds especially true for input factor evapotranspiration, which is varying considerably, while for the other input factors, identifiability remains more ore less constant. There is some evidence that identifiability of evapotranspiration is enhanced during wet periods. However, differences exist among subcatchments, as for Lascuarre the pattern seems to be reversed and for Ceguera there are periods of enhanced identifiability during wet and dry periods. For the other input factors, no relationships become visible.

The dynamic analysis of posterior frequencies, however, provides more detailed insights and reveals patterns even for input factors other than evapotranspiration (Figure 8). It shows that the clear pattern of high posterior values of odd-valued realizations of evapotranspiration (i.e., preference of the PM formula) is blurred or even reversed during periods of low flows, where even-valued realizations (the SW formula) dominate the posterior distribution (e.g., end of 2013 or beginning of 2015). That means that the SW approach is more suitable for dry conditions. In contrast, for soil water, such low flow periods lead to high posterior frequencies of IDs 7 and 8 (soil water retention model after Campbell and percolation modeled by simplified Richards' approach), while during high flows, ID 1 with completely different equations is favored (see Table 3). A model structure for runoff concentration is best identifiable during peak flows (ID 2, conceptual approach), but for most of the simulation period, posterior frequencies were close to prior frequencies. The pattern for ODE solvers is mostly blurred but shows a small tendency toward even-numbered realizations (constrained solvers) during low flows. During high flows, however, although the pattern is still blurred, odd-numbered (unconstrained) solvers with higher order of accuracy are more frequent in the posterior. However, it should be kept in mind that absolute identifiability measures for soil water, runoff concentration, and ODE solver are low and small absolute values of posterior frequencies are amplified due to the scaling in Figure 8.

For input factor parametrization (Figure 9), clear signals can only be seen for parameters cal_kfbed (groundwater recharge) and cal_ks (soil water movement). Parameter cal_kfbed appears to be best identifiable during dry periods with a tendency toward smaller values (less groundwater recharge, more soil moisture and interflow). In contrast, cal_ks is best identifiable during wet periods with a tendency toward smaller values (reduced conductivity of soil, more surface runoff). For cal_wind (evapotranspiration) and str_base (baseflow), patterns can be distinguished as well but are more difficult to generalize. Both during peak discharge and low flow periods, cal_wind is well identifiable with tendency toward smaller values (less
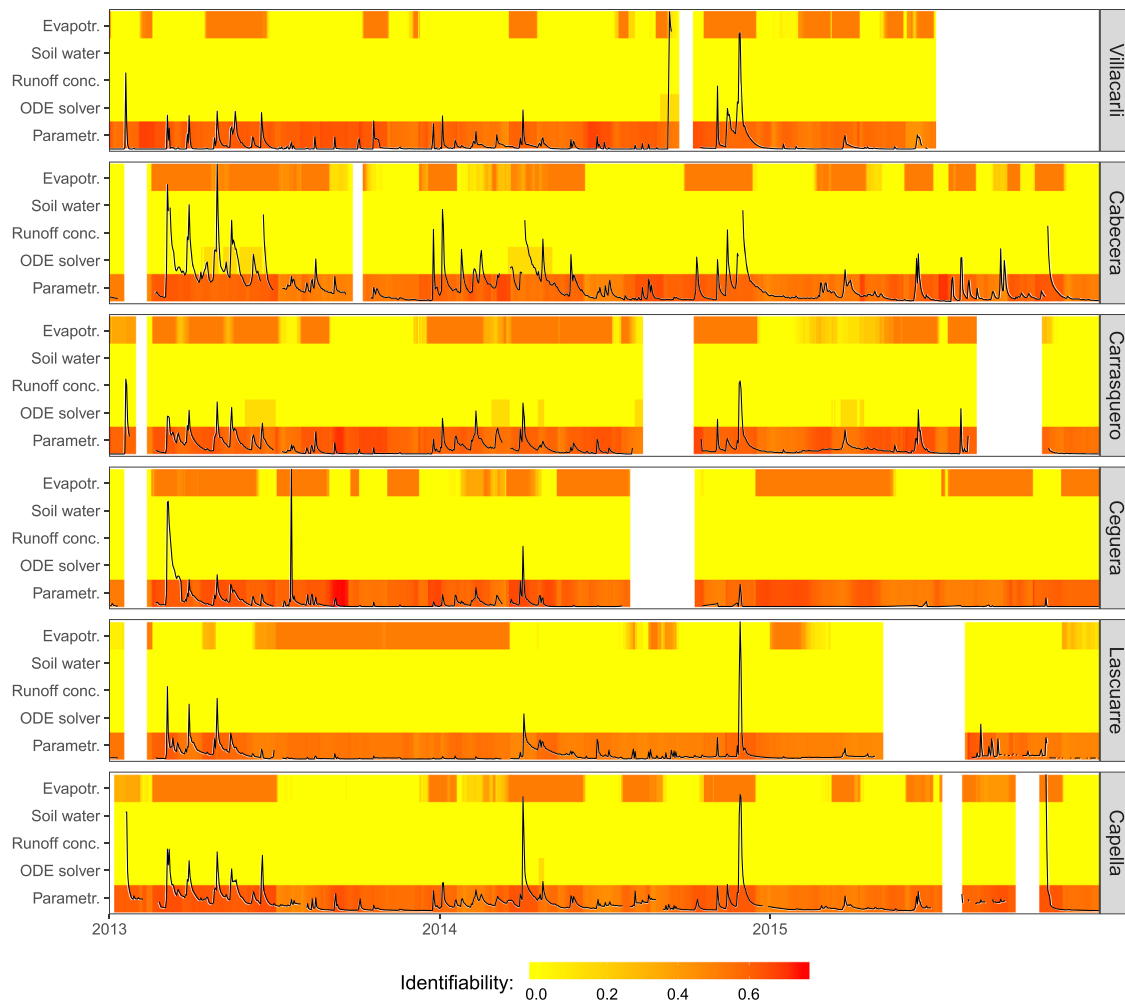
Figure 7. Identifiability measures over the simulation period. Black lines represent the observed discharge hydrograph at each gauge.

evapotranspiration), while in times of intermediate flow identifiability tends toward larger values (more evapotranspiration). Parameter str_base shows a tendency toward small values (small retention, quick release of flows) at the beginning of flow events with values increasing over the discharge event (increasing retention and prolonged release of flows; e.g., at the end of year 2013 or several times in 2015).

5. Discussion

5.1. Methodology and Identifiability Measure

The methodology presented in this paper can be easily integrated in any flexible model framework. However, some specific settings need to be selected based on the specific application. These aspects together with the implications of the decisions made in this specific case study are discussed in the following.

For this study, the ECHSE environment was chosen as it is relatively flexible in terms of model conception, specific process representations, and ODE solvers. Other such frameworks, to the authors' knowledge, focus on a specific conception and offer less freedom to the user, for example, with respect to alternative ODE solvers. Yet the proposed framework should also work with other model environments or even model ensembles.

The model conception in this case study was oriented at the complex hillslope-based discretization scheme of the process-based model WASA-SED. The main intention of this decision was that former studies analyzing hydrological model structures focused on conceptual models. In contrast, this study presents a first in-depth analysis of alternative process-based model structures.
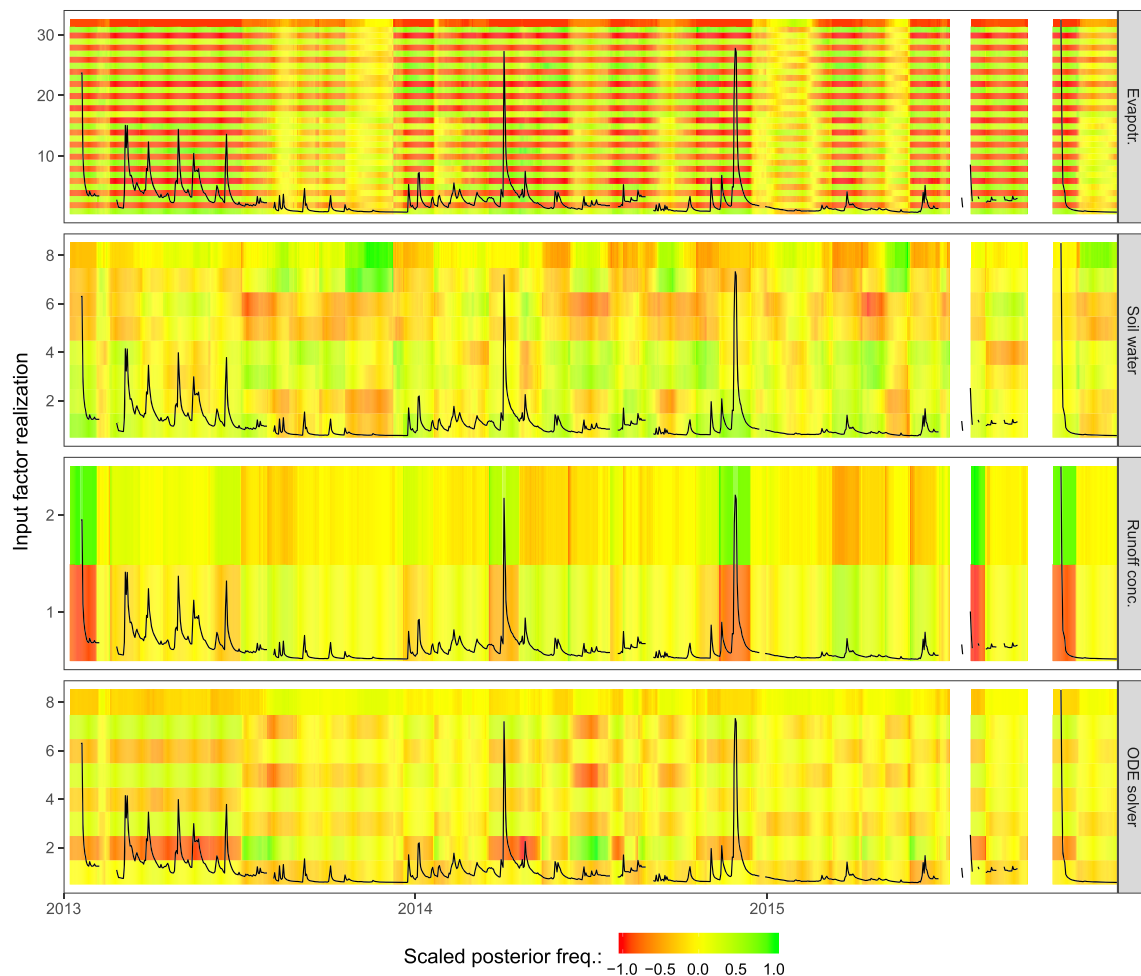
Figure 8. Posterior frequencies over the simulation period for gauge Capella (the catchment outlet) scaled to [− 1, 1] as in Figure 5. Black lines represent the observed discharge hydrograph.

The definition of input factors strongly depends on the aim of a study. For this case study, the objectives included the impact of different ODE solvers on simulation performance, which has rarely been investigated before, and alternative representations of considered processes. Yet the choice of specific alternatives is subjective and has to be constrained to limit the (considerable) amount of programming work and computational burden and to facilitate the interpretation of results. In future studies, it would be interesting to incorporate more or different process representations and/or include different sources of uncertainty as input factors, such as landscape discretization or input data (e.g., different rainfall data sets).

To account for the influence of parametrization, 1,000 parameter realizations were sampled. Even though this number is rather low, more realizations would have increased the number of model configurations and consequently called for a higher number of samples and, therefore, computational burden. Yet a separate bootstrap analysis shows that the identifiability measure converges well with increasing number of parameter realizations (Figure S4).

To limit the number of model simulations, out of the more than four million configurations, 12,000 samples were drawn reflecting the prior distribution. The confidence intervals shown in Figures 4 and 5 are narrow enough to indicate robust results, that is, that posterior distributions and associated identifiability measures are independent from a specific sample. Also the results are converging with increasing sample size (Figure S3; note that the confidence intervals might be underestimated at small sample sizes as explained in the figure caption and addressed by Isaksson et al., 2008). Yet it can be seen that for evapotranspiration, the confidence intervals for subcatchments Ceguera and Villacarli are still large (Figure 4). This can happen when some realizations of an input factor achieved only very few occurrences in the posterior distribution
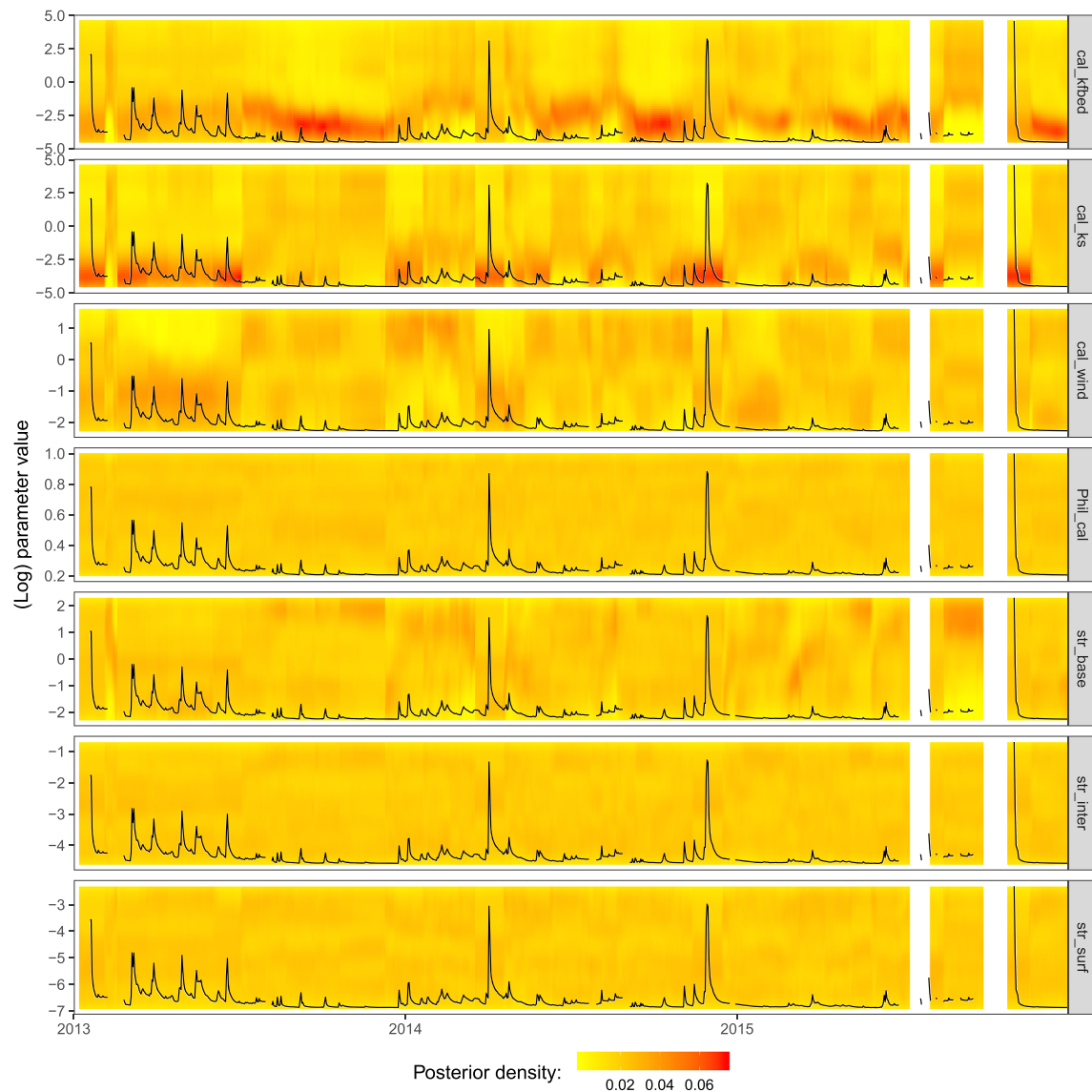
Figure 9. Posterior densities of model parameters derived via kernel density estimation and scaled to sum up to one at each time step for each parameter. Parameter values are log-transformed, except for Phil_cal. Shown is only gauge Capella, the catchment outlet. Black lines represent the observed discharge hydrograph.

(as further supported by Figure 5) and therefore sometimes are included and sometimes excluded from the posterior distribution when resampling the simulation results during bootstrapping.

The number of required samples of the prior distribution also depends on the number of realizations of the input factors. When the number of samples is low and each realization of an input factor appears only a few times or just once in the prior distribution, it is likely to achieve a higher identifiability as the posterior and prior distributions are more likely to differ from each other. However, a clear recommendation of the minimum number of samples with respect to the input factor with the highest number of realizations cannot be given. In the end, a compromise between sample size and computational demand needs to be found.

The results of parameter identification and sensitivity analyses are influenced by the employed performance metric (Francke, Baroni, et al., 2018; Guse et al., 2017). For this study, the RMSE was selected as in the studies of Wagener et al. (2003) and Pianosi and Wagener (2016). Experiments with the Nash-Sutcliffe index gave very similar results (not shown). Yet, RMSE, as well as the Nash-Sutcliffe index and other metrics using squared residuals, is biased toward higher values and the timing of the hydrograph while deficiencies of a

model in reproducing low flows are less strictly penalized. This aspect is relevant when aggregating over long periods and may therefore influence the findings of the static identifiability analysis of this work.

In addition to the performance metric, different variables of interest can be chosen for the comparison of simulations with observations. For the study area, a comprehensive data set of streamflow observations from different gauge locations is available. Besides, streamflow is the most widely used variable of interest in hydrological model applications and best reflects the hydrological cycle, that is, the transformation of rainfall into river discharge, of a catchment. These are the main reasons why this variable has been chosen for this case study. However, for instance, to derive more detailed information about evapotranspiration for soil water processes, also other variables, such as evapotranspiration and soil moisture measurements, can be taken into account. Yet such measurements are often, as in this case, not available and associated with high uncertainties.

Another important decision is the determination of the posterior distribution. Herein, the best 10% model configurations in terms of RMSE determine the posterior distributions of input factors. However, RMSE values are hard to interpret and at first sight it is not clear whether the best 10% simulation produced plausible results at all. For our case study, it turned out that all configurations within the posterior distribution achieved acceptable performances with NSE values greater than zero over the full analysis period (see Figure S5). The advantage of the informal MC filtering approach is that it is relatively straightforward to implement and easy to understand. The disadvantage is that it is somewhat subjective and lacking mathematical rigor. Besides, the identifiability measure was defined in a way that each input factor has to be analyzed separately and the identifiability measures of input factors are not directly comparable. This resulted from the discrete distribution of input factors and the highly different numbers of realizations. In general, different frameworks are able to present different insights and have different advantages and deficits. For instance, approaches of GSA have been used for the comparison of time-varying sensitivity indices of discrete input factors (Pianosi & Wagener, 2016) and can be further used to identify noninfluential input factors, which MC filtering does not allow (Pianosi et al., 2016). In that way, they account for some of the shortcomings of MC filtering but require larger sample sizes and impose a higher computational burden.

The dynamic analysis is influenced by the decision on the length of the moving window. This in particular is the case if sensitivity of the considered input factors varies in time, for example, because they represent processes occurring within larger or smaller characteristic time scales than the chosen window width (Massmann et al., 2014). However, objective selection criteria are missing, and it was decided that a window of 1 month be used. In that way, random measurement uncertainties are less likely to dominate the results (as would be the case for a small window size), but the window is still small enough to identify relevant process realizations at varying hydrological conditions. Experiments with varying window size (not shown) revealed that a smaller window size leads to a more blurred pattern, while with larger windows patterns can be more easily detected. However, the conclusions which can be derived from Figures 8 and 9 (i.e., the dependency of posterior values from wetness conditions) do not change. On the other hand, a small window might aid in more detailed analyses of process dynamics. For instance, identifiability patterns changed slightly for evapotranspiration when using a smaller window, especially for gauge Villacarli, where more periods of high identifiability appeared. For the other gauges, patterns became more diverse as well but not relevant regarding the conclusions of this study.

5.2. Spatiotemporal Patterns of Identifiability

Static analysis over the whole simulation period as well as the dynamic approach using a moving window found evapotranspiration and parametrization to be the only input factors showing a certain degree of identifiability. Yet the temporal analysis provides much more insights into model functioning. It allows for a more detailed comparison of prior and posterior distributions and reveals patterns also for the other input factors, which are overall poorly identifiable. In addition, the analysis shows that the time-varying dominance of certain input factors is to a large degree driven by meteorological conditions. This conclusion is well in line with other studies emphasizing the added value of a temporal analysis in the context of sensitivity or identifiability analysis (e.g., Ghasemizade et al., 2017; Guse et al., 2014; Herman et al., 2013; Pianosi & Wagener, 2016; Reusser & Zehe, 2011; Savage et al., 2016).

It was found that during wet periods, the PM approach clearly dominated the posterior distribution and the parameter cal_wind was directed toward reduced evapotranspiration amounts. During dry periods, the SW approach was dominant with a less clear pattern for cal_wind. This is supported by the fact that the

SW approach was most dominant for subcatchment Lascuarre, which is the driest part of the study area. The SW formula relaxes the big-leaf assumption of the PM approach in a way that it accounts for bare soil and is therefore a more sophisticated approach for sparse crops and patchy vegetation (Shuttleworth & Wallace, 1985). While subcatchment Lascuarre has the largest fraction of cropland, it is not clear whether this really translates to a more patchy vegetation and thus no better suitability of either approach can be inferred a priori. It rather seems that moisture condition is the most influential factor for the selection of an evapotranspiration model in comparison to landscape characteristics (homogeneous vs. patchy vegetation cover).

An interesting finding is that unconstrained ODE solvers with high order of accuracy perform better during wet periods. A possible explanation is that implausible model states, which are likely caused by unconstrained solvers under rainfall conditions, can still produce a more realistic streamflow dynamic. This can be attributed to the faster soil water fluxes with characteristic time scales less than the model's temporal resolution. Under such circumstances, ODE solvers without solution constraints could serve as compensation for the rather coarse daily resolution of the model runs. For instance, consider a high amount of daily precipitation, which occurs in fact just within a few hours rather than equally over a full day. A large signal of rainfall input within a single model time step would likely cause a high amount of surface runoff, while if the signal would be distributed over several sub steps, it would rather result in less overland flow and a less sharp runoff signal. In that way, the temporary storage of water in the soil, although exceeding the physical boundaries, is delaying the runoff signal and could therefore still result in a more realistic system behavior. In contrast, under dry conditions, constrained ODE solvers are favored as they keep the model states within physical limits, which results in a more realistic streamflow dynamic.

In addition to the temporal patterns, some differences among the subcatchments were found. This is especially true for gauge Lascuarre in comparison to the rest of the catchment. Apart from the issue of the evapotranspiration model, which already has been discussed, constrained ODE solvers are more clearly favored at this than at the other gauges. This can be attributed to the distinct hydrological and meteorological conditions in this subcatchment (see Table 1). In contrast to the other subcatchments, Lascuarre is characterized by a very low runoff coefficient, sharper discharge peaks, less precipitation, less steep topography, and more agricultural areas. This supports the findings of van Werkhoven et al. (2008), who found distinct patterns of parametric controls in dry and wet catchments. Their findings for parametric controls can therefore be extended to certain process realizations and even ODE solvers.

5.3. Is There an Optimal Model Structure?

The most straightforward approach to address the question of the optimal model structure would be to select the best performing realization. However, it was shown that the best performing model run does not necessarily refer to the highest posterior frequencies of the analyzed input factors (Figure 5). This suggests a high degree of interaction among input factors, that is, only very specific model configurations result in a good performance of a certain model structure, while changes in the other input factors may significantly deteriorate model behavior.

It was found that spatial variability, even in catchments of the lower mesoscale such as investigated in this work, can be substantial and lead to contrasting conclusions in neighboring subcatchments. The same contrasting conclusions are derived when examining temporal patterns. For instance, it was consistently found that under dry conditions, the SW model is a more plausible evapotranspiration model, while under wet conditions, the PM approach was favored. In future studies, more advanced methods, such as machine learning techniques, could be employed to derive relationships between catchment characteristics and meteorological conditions as predictors and certain model process formulations as response variable. This would allow the designing of the most likely model configuration prior to an application based on the characteristics of the catchment to be investigated. Flexible simulation environments such as ECHSE (among others) enable such a task and serve as a toolbox for the modeler (Clark et al., 2008, 2015; Fenicia et al., 2011; Kneis, 2015; Knoben et al., 2019).

Finally, the identifiability analysis sheds some lights into model functioning and complex interactions between the different model components. Quite surprisingly, even ODE solvers of low order of accuracy, at least in some occasions, achieved rather high rankings in the posterior distribution. This suggests that model deficiencies can be compensated by, albeit unrealistic, parametrizations or process formulations. Consequently, there is a chance to obtain the right answers for the wrong reasons, a phenomenon resulting from,

for example, overparametrization of a model, which has been acknowledged in numerous studies in the field of hydrological modeling (e.g., Fenicia et al., 2016; Kavetski & Clark, 2010; Kirchner, 2006; Samaniego et al., 2010; Schoups et al., 2008). In order to avoid corrupt model parametrization, in line with the findings of other studies (e.g., Kavetski & Clark, 2011), robust ODE solvers of high order of accuracy should be used. In addition, the use of ODE solvers with solution constraints appears unsatisfactorily from a mathematical point of view as the model is forcibly pushed into plausible states. However, this as well calls for finer discretizations (in space and time) than they were used in this study as in many others. For instance, separate analysis of the soil water module in ECHSE revealed that simulations with ODE solvers of higher order of accuracy produce soil moisture states within plausible ranges also without solution constraints when running the model with shorter temporal resolution (1 hr) and more soil horizons with gradually varying soil parameters (not shown in this paper). On the other hand, model runtimes increase dramatically and prevent detailed analyses of ensemble-based approaches as presented in this study. Therefore, solution constraints can still be a practical means, though unsatisfactorily, to keep model states within plausible limits and achieve good model performances under specific boundary conditions (in this case study, dry conditions; see Figure 8).

6. Conclusions

This study investigated the spatiotemporal identifiability of multiple model structures. The experiments were conducted in a lower mesoscale mountainous catchment in northeastern Spain. To carry out simulations, the flexible simulation environment ECHSE was used, which enabled the rapid implementation of different alternatives for process representation (with respect to subprocesses of evapotranspiration, soil water movement, and runoff concentration) and ODE solvers. This flexible model environment was coupled with DYNIA. The approach is generic and flexible in a way that a user can freely choose a flexible simulation environment, the input factors and realizations (process formulations, parametrization, etc.), how to derive the posterior distribution, and how to investigate the results.

While the framework was designed primarily for model structure identification, it also allowed obtaining information about process behavior under changing boundary conditions. The main findings with respect to the initially stated research questions shall be briefly summarized as follows.

1. Parametrization and subprocesses of evapotranspiration turned out to be the only identifiable input factors. Yet more patterns were identified by inspecting the posterior distributions.
2. Identifiability patterns vary over time. The Penman-Monteith approach appeared to be superior during wet periods while during dry periods, the Shuttleworth-Wallace approach led to better model performances. Moreover, unconstrained ODE solvers with high order of accuracy performed better during wet periods, while solvers with solution constraints obtained better model performance during dry periods.
3. The results of model identification are clearly influenced by hydrological characteristics. While identifiability patterns are relatively consistent over areas with similar hydrological characteristics, identified model structures are most distinct for the subcatchment with the most diverging characteristics with respect to land use, topography, rainfall sum, and runoff coefficient.

Identifiability patterns might be influenced by interactions among input factors and compensation effects. This could explain why in some rare cases even ODE solvers of low order of accuracy achieved good model results and therefore high rankings in the posterior distribution. Therefore, it is difficult to decide for a specific model configuration, as the model obtaining the best performance metrics might be influenced by such compensation effects. Consequently, the results of this study impose the following further research questions:

1. How can the compensation effects be eliminated and model identification made more robust? How do temporal resolution and ODE solvers dictate these issues and how do they interact?
2. If temporal resolution and ODE solvers are crucial, how can they be addressed while maintaining feasible model runtimes?
3. Could data science (e.g., machine learning) be combined with process knowledge to determine the most adequate model structure for a study area before conducting time consuming model evaluations?

While answering these questions will be the focus of new studies, we believe that coupling a flexible model framework with identifiability analysis will play an important role to address these issues. In our case study,

we have shown that DYNIA based on MC filtering together with the introduced identifiability measure is one simple and practical approach that can be easily employed with any flexible model environment.

Data Availability Statement

The data, R scripts, and ECHSE model, on which this study is based, were published with GFZ Data Services. The data and R scripts to reproduce the identifiability analysis are accessible from https://doi.org/10/5880/PIK.2019.016 (Pilz, Francke, et al., 2019). The ECHSE environment with the new WASA-SED engine is available from https://doi.org/10.5880/pik.2019.017 (Pilz, 2019). The resources are freely accessible and contain detailed license information.

References

Abon, C. C., Kneis, D., Crisologo, I., Bronstert, A., David, C. P. C., & Heistermann, M. (2016). Evaluating the potential of radar-based rainfall estimates for streamflow and flood simulations in the Philippines. *Geomatics, Natural Hazards and Risk*, *7*(4), 1390–1405. https://doi.org/10.1080/19475705.2015.1058862

Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, *43*, W01403. https://doi.org/10.1029/2005WR004745

Allen, R. G., Walter, I. A., Elliott, R., Terry Howell, D. I., & Jensen, M. (2005). The ASCE standardized reference evapotranspiration equation. Reston, Virginia, USA: Environmental and Water Resources Institute of the American Society of Civil Engineers (ASCE).

Baroni, G., & Tarantola, S. (2014). A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling & Software*, *51*, 26–34. https://doi.org/10.1016/j.envsoft.2013.09.022

Beven, K. (2009). *Environmental modelling: An uncertain future?* London, UK: Routledge.

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Bronstert, A., de Araújo, J.-C., Batalla, R. J., Costa, A. C., Delgado, J. M., Francke, T., et al. (2014). Process-based modelling of erosion, sediment transport and reservoir siltation in mesoscale semi-arid catchments. *Journal of Soils and Sediments*, *14*(12), 2001–2018. https://doi.org/10.1007/s11368-014-0994-1

Brutsaert, W. (1975). Comments on surface roughness parameters and the height of dense vegetation. *Journal of the Meteorological Society of Japan*, *53*(1), 96–98. https://doi.org/10.2151/jmsj1965.53.1_96

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*, W09301. https://doi.org/10.1029/2010WR009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*, 2498–2514. https://doi.org/10.1002/2015WR017198

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*, W00B02. https://doi.org/10.1029/2007WR006735

Coxon, G., Freer, J., Wagener, T., Odoni, N. A., & Clark, M. (2014). Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. *Hydrological Processes*, *28*(25), 6135–6150. https://doi.org/10.1002/hyp.10096

de Araújo, J. C., & Medeiros, P. H. A. (2013). Impact of dense reservoir networks on water resources in semiarid environments. *Australian Journal of Water Resources*, *17*(1), 87–100. https://doi.org/10.7158/13241583.2013.11465422

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, *47*, W11510. https://doi.org/10.1029/2010WR010174

Fenicia, F., Kavetski, D., Savenije, H. H., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, *52*, 954–989. https://doi.org/10.1002/2015WR017398

Francke, T., Baroni, G., Brosinsky, A., Foerster, S., López-Tarazón, J. A., Sommerer, E., & Bronstert, A. (2018). What did really improve our meso-scale hydrological model? A multi-dimensional analysis based on real observations. *Water Resources Research*, *54*, 8594–8612. https://doi.org/10.1029/2018WR022813

Francke, T., Foerster, S., Brosinsky, A., Sommerer, E., Lopez-Tarazon, J. A., Güntner, A., et al. (2018). Water and sediment fluxes in mediterranean mountainous regions: Comprehensive dataset for hydro-sedimentological analyses and modelling in a mesoscale catchment (river Isábena, NE Spain). *Earth System Science Data*, *10*(2), 1063–1075. https://doi.org/10.5194/essd-10-1063-2018

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., et al. (2017). GNU scientific library release 2.4. https://www.gnu.org/software/gsl/.

García-Ruiz, J. M., Beguería, S., López-Moreno, J. I., Lorente, A., & Seeger, M. (2001). *Los recursos hídricos superficiales del pirineo aragonés y su evolución reciente.* Logroño, Spain: Geoforma.

Ghasemizade, M., Baroni, G., Abbaspour, K., & Schirmer, M. (2017). Combined analysis of time-varying sensitivity and identifiability indices to diagnose the response of a complex environmental model. *Environmental Modelling & Software*, *88*, 22–34. https://doi.org/10.1016/j.envsoft.2016.10.011

Guillaume, J. H. A., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., et al. (2019). Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, *119*, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007

Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in snowpack simulations assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance. *Water Resources Research*, *55*, 2779–2800. https://doi.org/10.1029/2018WR023403

Güntner, A. (2002). Large-scale hydrological modelling in the semi-arid north-east of Brazil. *PIK Report* (Vol. *77*). Potsdam, Germany: Potsdam Institute for Climate Impact Research.

Güntner, A., & Bronstert, A. (2004). Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *Journal of Hydrology*, *297*(1–4), 136–161. https://doi.org/10.1016/j.jhydrol.2004.04.008

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*, W08301. https://doi.org/10.1029/2011WR011044

Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., & Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, *21*(11), 5663–5679. https://doi.org/10.5194/hess-21-5663-2017

Guse, B., Reusser, D. E., & Fohrer, N. (2014). How to improve the representation of hydrological processes in SWAT for a lowland catchment temporal analysis of parameter sensitivity and model performance. *Hydrological Processes*, *28*(4), 2651–2670. https://doi.org/10.1002/hyp.9777

Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, *49*, 1400–1414. https://doi.org/10.1002/wrcr.20124

Höge, M., Wöhling, T., & Nowak, W. (2018). A primer for model selection: The decisive role of model complexity. *Water Resources Research*, *54*, 1688–1715. https://doi.org/10.1002/2017WR021902

Isaksson, A., Wallman, M., Gransson, H., & Gustafsson, M. G. (2008). Cross-validation and bootstrapping are unreliable in small sample classification. *Pattern Recognition Letters*, *29*(14), 1960–1965. https://doi.org/10.1016/j.patrec.2008.06.018

Jackisch, C., Zehe, E., Samaniego, L., & Singh, A. K. (2014). An experiment to gauge an ungauged catchment: Rapid data assessment and eco-hydrological modelling in a data-scarce rural catchment. *Hydrological Sciences Journal*, *59*(12), 2103–2125. https://doi.org/10.1080/02626667.2013.870662

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resources Research*, *46*, W10511. https://doi.org/10.1029/2009WR008896

Kavetski, D., & Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes*, *25*(4), 661–670. https://doi.org/10.1002/hyp.7899

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*, W11511. https://doi.org/10.1029/2011WR010748

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*, W03S04. https://doi.org/10.1029/2005WR004362

Kneis, D. (2012). geostat: Utilities for spatial interpolation. https://github.com/echse/echse_tools/tree/master/R/packages/geostat.

Kneis, D. (2015). A lightweight framework for rapid development of object-based hydrological model engines. *Environmental Modelling & Software*, *68*, 110–121. https://doi.org/10.1016/j.envsoft.2015.02.009

Kneis, D., Abon, C., Bronstert, A., & Heistermann, M. (2017). Verification of short-term runoff forecasts for a small Philippine basin (Marikina). *Hydrological Sciences Journal*, *62*(2), 205–216. https://doi.org/10.1080/02626667.2016.1183773

Kneis, D., Chatterjee, C., & Singh, R. (2014). Evaluation of TRMM rainfall estimates over a large Indian river basin (Mahanadi). *Hydrology and Earth System Sciences*, *18*(7), 2493–2502. https://doi.org/10.5194/hessd-11-1169-2014

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular assessment of rainfall-runoff models toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, *12*(6), 2463–2480. https://doi.org/10.5194/gmd-12-2463-2019

Krol, M. S., de Vries, M. J., Oel, P. R., & de Araújo, J. C. (2011). Sustainability of small reservoirs and large scale water availability under current conditions and climate change. *Water Resources Management*, *25*(12), 3017–3026. https://doi.org/10.1007/s11269-011-9787-0

Lilburne, L., & Tarantola, S. (2009). Sensitivity analysis of spatial models. *International Journal of Geographical Information Science*, *23*(2), 151–168. https://doi.org/10.1080/13658810802094995

Maidment, D. R. (Ed.) (1993). *Handbook of hydrology*. New York, USA: McGraw-Hill, Inc.

Massmann, C., Wagener, T., & Holzmann, H. (2014). A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales. *Environmental Modelling & Software*, *51*, 190–194. https://doi.org/10.1016/j.envsoft.2013.09.033

Medeiros, P. H. A., de Araújo, J. C., Mamede, G. L., Creutzfeldt, B., Güntner, A., & Bronstert, A. (2014). Connectivity of sediment transport in a semiarid environment: A synthesis for the upper Jaguaribe basin, Brazil. *Journal of Soils and Sediments*, *14*(12), 1938–1948. https://doi.org/10.1007/s11368-014-0988-z

Medeiros, P. H. A., Güntner, A., Francke, T., Mamede, G. L., & de Arajo, J. C. (2010). Modelling spatio-temporal patterns of sediment yield and connectivity in a semi-arid catchment with the WASA-SED model. *Hydrological Sciences Journal*, *55*(4), 636–648. https://doi.org/10.1080/02626661003780409

Mueller, E. N., Francke, T., Batalla, R. J., & Bronstert, A. (2009). Modelling the effects of land-use change on runoff and sediment yield for a meso-scale catchment in the southern Pyrenees. *Catena*, *79*(3), 288–296. https://doi.org/10.1016/j.catena.2009.06.007

Mueller, E. N., Güntner, A., Francke, T., & Mamede, G. (2010). Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model. *Geoscientific Model Development*, *3*(1), 275–291. https://doi.org/10.5194/gmd-3-275-2010

Mustafa, S. M. T., Nossent, J., Ghysels, G., & Huysmans, M. (2020). Integrated Bayesian multi-model approach to quantify input, parameter and conceptual model structure uncertainty in groundwater modeling. *Environmental Modelling & Software*, *126*, 104654. https://doi.org/10.1016/j.envsoft.2020.104654

Philip, J. R. (1957). The theory of infiltration: 4. Sorptivity and algebraic infiltration equations. *Soil Science*, *84*(3), 257–264.

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. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*, *30*(22), 3991–4003. https://doi.org/10.1002/hyp.10968

Pilz, T. (2019). A new process-based model engine for the Eco-hydrological Simulation Environment (ECHSE). GFZ Data Services. https://doi.org/10.5880/pik.2019.017

Pilz, T., Delgado, J. M., Voss, S., Vormoor, K., Francke, T., Costa, A. C., et al. (2019). Seasonal drought prediction for semiarid northeast Brazil: What is the added value of a process-based hydrological model? *Hydrology and Earth System Sciences*, *23*(4), 1951–1971. https://doi.org/10.5194/hess-23-1951-2019

Pilz, T., Francke, T., Baroni, G., & Bronstert, A. (2019). Data and R-scripts for dynamic identifiability analysis of process-based hydrological model structures. GFZ Data Services. https://doi.org/10.5880/PIK.2019.016

Pilz, T., Francke, T., & Bronstert, A. (2017). lumpr 2.0.0: An R package facilitating landscape discretisation for hillslope-based hydrological models. *Geoscientific Model Development*, *10*(8), 3001–3023. https://doi.org/10.5194/gmd-10-3001-2017

Plischke, E., Borgonovo, E., & Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, *226*(3), 536–550. https://doi.org/10.1016/j.ejor.2012.11.047

Reusser, D. E., & Zehe, E. (2011). Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resources Research*, *47*, W07550. https://doi.org/10.1029/2010WR009946

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, *25*(12), 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Saugier, B., & Katerji, N. (1991). Some plant factors controlling evapotranspiration. *Agricultural and Forest Meteorology*, *54*(2), 263–277. https://doi.org/10.1016/0168-1923(91)90009-F

Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, *52*, 9146–9163. https://doi.org/10.1002/2015WR018198

Schoups, G., van de Giesen, N. C., & Savenije, H. H. G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research*, *44*, W00B03. https://doi.org/10.1029/2008WR006836

Schoups, G., Vrugt, J. A., Fenicia, F., & van de Giesen, N. C. (2010). Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. *Water Resources Research*, *46*, W10530. https://doi.org/10.1029/2009WR008648

Shuttleworth, W. J., & Gurney, R. J. (1990). The theoretical relationship between foliage temperature and canopy resistance in sparse crops. *Quarterly Journal of the Royal Meteorological Society*, *116*(492), 497–519. https://doi.org/10.1002/qj.49711649213

Shuttleworth, W. J., & Wallace, J. S. (1985). Evaporation from sparse crops—An energy combination theory. *Quarterly Journal of the Royal Meteorological Society*, *111*(469), 839–855. https://doi.org/10.1002/qj.49711146910

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, *14*(1), 43–49. https://doi.org/10.1016/0043-1354(80)90040-8

Stahn, P., Busch, S., Salzmann, T., Eichler-Löbermann, B., & Miegel, K. (2017). Combining global sensitivity analysis and multiobjective optimisation to estimate soil hydraulic properties and representations of various sole and mixed crops for the agro-hydrological SWAP model. *Environmental Earth Sciences*, *76*(10), 367. https://doi.org/10.1007/s12665-017-6701-y

Tang, Y., Reed, P., Wagener, T., & van Werkhoven, K. (2007). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences*, *11*(2), 793–817. https://doi.org/10.5194/hess-11-793-2007

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software*, *63*, 24–31. https://doi.org/10.1016/j.envsoft.2014.09.017

van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*, W01429. https://doi.org/10.1029/2007WR006271

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, *17*(2), 455–476. https://doi.org/10.1002/hyp.1135

Weiler, M., & Beven, K. (2015). Do we need a community hydrological model? *Water Resources Research*, *51*, 7777–7784. https://doi.org/10.1002/2014WR016731

Wrede, S., Fenicia, F., Martínez-Carreras, N. A., Juilleret, J., Hissler, C., Krein, A., et al. (2015). Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. *Hydrological Processes*, *29*(12), 2731–2750. https://doi.org/10.1002/hyp.10393