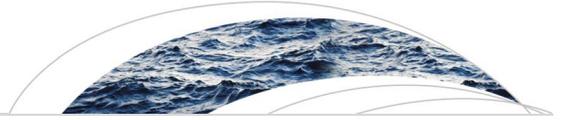




Originally published as:

Francke, T., Baroni, G., Brosinsky, A., Förster, S., López-Tarazón, J. A., Sommerer, E., Bronstert, A. (2018): What Did Really Improve Our Mesoscale Hydrological Model? A Multidimensional Analysis Based on Real Observations. - *Water Resources Research*, 54, 11, pp. 8594—8612.

DOI: <http://doi.org/10.1029/2018WR022813>



Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR022813

Key Points:

- The proposed framework assesses the effect of model enhancements (MEs), e.g., improved data, resolution, model structure, on model performance
- It allows a differentiated evaluation of the effect of these ME according to multiple objectives and their interaction
- For the case study (hydro-sedimentological model in Mediterranean basin), improved rainfall data proved most beneficial

Supporting Information:

- Supporting Information S1

Correspondence to:

T. Francke,
francke@uni-potsdam.de

Citation:

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

Received 23 FEB 2018

Accepted 11 JUL 2018

Accepted article online 18 JUL 2018

Published online 5 NOV 2018

What Did Really Improve Our Mesoscale Hydrological Model? A Multidimensional Analysis Based on Real Observations

Till Francke¹ , Gabriele Baroni^{1,2} , Arlena Brosinsky^{1,3}, Saskia Foerster³ , José A. López-Tarazón^{1,4,5}, Erik Sommerer³, and Axel Bronstert¹

¹Institute of Earth and Environmental Science, Universität Potsdam, Germany, ²UFZ—Helmholtz Centre for Environmental Research, Leipzig, Germany, ³GFZ German Research Centre for Geosciences, Potsdam, Germany, ⁴Mediterranean Ecogeomorphological and Hydrological Connectivity Research Team, Department of Geography, Universitat de las Islas Baleares, Palma, Spain, ⁵Fluvial Dynamics Research Group, Department of Environment and Soil Sciences, Universitat de Lleida, Spain

Abstract Modelers can improve a model by addressing the causes for the model errors (data errors and structural errors). This leads to implementing *model enhancements* (MEs), for example, meteorological data based on more monitoring stations, improved calibration data, and/or modifications in process formulations. However, deciding on which MEs to implement remains a matter of expert knowledge. After implementing multiple MEs, any improvement in model performance is not easily attributed, especially when considering different objectives or aspects of this improvement (e.g., better dynamics vs. reduced bias). We present an approach for comparing the effect of multiple MEs based on real observations and considering multiple objectives (MMEMO). A stepwise selection approach and structured plots help to address the multidimensionality of the problem. Tailored analyses allow a differentiated view on the effect of MEs and their interactions. MMEMO is applied to a case study employing the mesoscale hydro-sedimentological model WASA-SED for the Mediterranean-mountainous Isábena catchment, northeast Spain. The investigated seven MEs show diverse effects: some MEs (e.g., rainfall data) cause improvements for most objectives, while other MEs (e.g., land use data) only affect a few objectives or even decrease model performance. Interaction of MEs was observed for roughly half of the MEs, confirming the need to address them in the analysis. Calibration and increasing the temporal resolution showed by far stronger impact than any of the other MEs. The proposed framework can be adopted in other studies to analyze the effect of MEs and, thus, facilitate the identification and implementation of the most promising MEs for comparable cases.

1. Introduction

Uncertainty is an inevitable phenomenon in modeling the response of water and matter fluxes in complex geosystems (Beven, 2007; Guzman et al., 2015). According to Matott et al. (2009), two groups of uncertainty can be distinguished: input uncertainty and model uncertainty. The former comprises input data, response data, and parameter uncertainty. The latter covers model structure, resolution, spatiotemporal correspondence, and code. The identification and the quantification of the different sources of uncertainty among these groups in a specific model application are crucial steps for improving model performance. This assessment is difficult, however, due to the complexity of the models, the different and often unknown characteristics of the features, and the interactions among them. For these reasons, deciding on how to improve the model performance and which features to implement (e.g., meteorological data based on more monitoring stations, improved calibration data, and/or modifications in process formulations) remains in many cases a matter of expert knowledge. The aim of this study is to present a practical and efficient approach to assess and select between multiple possible changes in the model that could help in supporting model applications and further improvements. Since these features are assembled to enhance the model, they are called *model enhancements* (MEs) henceforward. Arguably, this contributes to the confusion in taxonomies in that regard (Matott et al., 2009), but equivalent terms like *model component* (Guzman et al., 2015), *scenario* (Droogers et al., 2008), and *factor* (Pechlivanidis et al., 2011) struck us as being ambiguous. Here, *ME* does not imply that they are indeed *improving* the model (indeed, some may turn out to be detrimental) nor does it mean they are only related to the model itself (i.e., process formulation): instead, better input data, better resolution, and so forth are also considered MEs.

In this context, a broad range of diagnostic tools for exploring the relationship of the input and output space of a model is already available for different purposes (Matott et al., 2009; Uusitalo et al., 2015). Among these, sensitivity analyses have received much attention in recent decades as a tool to quantify the relative importance of the different sources of uncertainty (Ferretti et al., 2016). In this context, numerous methods have been developed that are presented and discussed in several review papers (Pianosi et al., 2016; Razavi & Gupta, 2015; Saltelli et al., 2007; Song et al., 2015).

The simplest methods rely on scenario-based realizations for which One-at-A-Time (OAT) approaches are implemented. In this approach, one feature is changed while keeping the others constant (e.g., Loosvelt et al., 2013; Paton et al., 2013). Studies within this category usually compare a limited number of MEs (<5), and the MEs are usually represented by real (i.e., practically implementable) enhancements introduced in the model application, such as rainfall measurements or refined land use maps (Chen et al., 2011; Paton et al., 2013). Similarly, different model structures can be compared (e.g., Fenicia et al., 2008), especially when using flexible modeling frameworks (Clark et al., 2008, 2015; Kneis, 2015). Overall, the number of simulations is usually limited (<20 simulations) allowing their use also for computationally demanding models. While OAT is relatively easy to implement and is still used in many studies (Ferretti et al., 2016), it has some limitations due to the nonlinear response of many models as well as interacting effects among the enhancements analyzed (Saltelli & Annoni, 2010). To overcome these limitations, other methods such as global sensitivity analyses methods based on extensive sampling of multiple combinations of features are applied (e.g., Borgonovo et al., 2017; Pianosi & Wagener, 2015; Rakovec et al., 2014; Razavi & Gupta, 2016). Here MEs are usually described by a priori statistical assumptions, and the analysis is based on sampling a defined probabilistic distribution that is supposed to describe the uncertainty of the specific ME. For this reason, this approach originally focused on the effects of different parameters on the model response, that is, parameter uncertainty. More recently, however, these strategies have also been extended to account for other sources of uncertainty (Baroni & Tarantola, 2014; Lilburne & Tarantola, 2009; Savage et al., 2016; Shoaib et al., 2016; Stahn et al., 2017). Thus, different model structures could be also evaluated in the assessment, especially when combined with flexible modeling frameworks (e.g., Clark et al., 2008, 2015; Kneis, 2015). In addition, the tools were integrated in practical code packages (looss et al., 2016; Pianosi et al., 2015) and can now be easily integrated in any modeling applications (Shin et al., 2013). Still, the main drawback of these types of approaches is the large computational requirement (e.g., >500 runs) that could quickly become unfeasible when the number of sources of uncertainty or their probability range is relatively high (Sarrazin et al., 2016).

This discontinuity between OAT and global sensitivity analyses approaches motivated us to propose an *intermediate* approach that could also be applied when the model computational effort is high (i.e., in the order of hours) and/or the number of MEs is relatively large (i.e., >5). This new framework for analyzing the effect of MEs yields quantitative information and ranks the effect of the implemented enhancements. It extends the OAT approach to two (and potentially more) starting points to explore the possible interactions between the features considered. All sources of model uncertainty (input data, response, process representation, etc.) can be integrated in the analysis. Finally, the assessment of the performance is conducted on different objectives of the model response to provide a comprehensive view of the model performance. In general, the framework could also be applied to synthetic perturbations of each ME (defined by a priori statistical distribution). However, the choice of these distributions can strongly affect the outcome and the respective conclusions (Baroni et al., 2017; Haghnegahdar & Razavi, 2017; Zhang et al., 2015). For this reason, we decided to demonstrate the method using practically implementable features as MEs (e.g., better rainfall data). The method is demonstrated on an example of water and sediment modeling in a mesoscale catchment in NE Spain for which several enhancements were derived from long-term monitoring, extensive field campaigns, and remote sensing data. Where applicable, the MEs are also quantified for guiding the definition of a priori ranges in similar studies.

The paper is structured as follows: section 2 explains the general concept of the proposed method. As an example, the method is applied to a specific case study in section 3. In section 4, the results are presented, while potential and limitations are discussed in section 5, both in terms of the method and in relation to the specific application. In section 6, we summarize and present concluding remarks. The paper is complemented by an extensive annex.

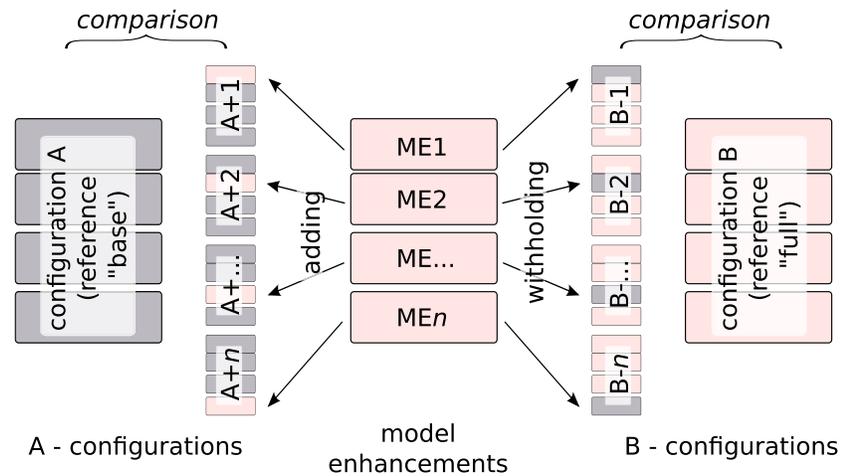


Figure 1. Forward and backward selection of model enhancements to and from configurations A and B, respectively.

2. The Multidimensional Framework

The proposed method relies on the multidimensional analysis of both multiple model enhancements and multiple objectives. Henceforward, it is referred to as MMEMO.

2.1. Forward and Backward Selection

MMEMO focuses on quantifying the effect of explicit changes of features on the model performance. Although the framework could also be applied to synthetical perturbations of these features, we aim to use concrete realizations of them, that is, as resulted from a measurement campaign.

Each ME can change the model performance P . However, these changes may be different whether the ME is added to (withheld from) the base (full) configuration. For this reason, the evaluation of the effect of an ME is done in two ways: a forward selection and a backward elimination. In the former, each ME is added to the base configuration (A, Figure 1, left). Any change in performance is compared with respect to A. In the latter, each ME is withheld from the configuration, including all available MEs (B, Figure 1, right). Change in performance is compared with respect to B. These configurations are depicted in Figure 1 as “A+” and “B–”, respectively. In contrast to the standard OAT approach, this aims at allowing a more differentiated assessment of the role of the individual ME with regard to their interaction (e.g., Campolongo et al., 2007; Morris, 1991). Assuming m_{max} MEs to be assessed and k couples of reference configurations (A and B), a total of $N = (m_{max} + 1) \cdot 2 \cdot k$ configurations is treated. A minimum of two reference configurations (i.e., $k = 1$) needs to be considered, however, for a more detailed analysis, an increased number of couples k can be also included.

2.2. Quantifying the Effect of an ME on Different Objectives

The effect of an ME m is reflected in a change in the performance measure P . It can be expressed as a relative change of P_m when compared to the reference configuration P_{ref} (*scenario impact*; Droogers et al., 2008), normalized by the maximum achieved improvement among all m_{max} MEs:

$$IP_m = s \cdot \frac{P_{ref} - P_m}{\max_{j=1}^{m_{max}} (P_{ref} - P_j)}, \quad (1)$$

with s : sign correction ($s = 1$ for adding an ME as in the A+ configurations; $s = -1$ for withholding an ME as in the B– configurations). Thus, a positive value of IP marks an improvement of the model, while a negative one indicates degradation. The configuration with the largest gain in performance will get the value of 1. The resulting values allow a ranking of the MEs (similar to Chen et al., 2011).

Additionally to forward or backward selection of an ME (A or B configuration), we can consider n_o different objectives (aspects of improvement) by choosing different expressions for the model performance P_i , with

$i \in \{1, \dots, n_o\}$. These objectives depend on the target variable (e.g., water or sediment), the chosen performance measure (e.g., focus on dynamics or bias), the spatial focus (e.g., output at subcatchment or basin outlet), the temporal focus (e.g., daily or seasonal prediction), and so forth. Thus, a given ME may increase the performance of the model for a specific objective i , while its effect may be detrimental when looked at from a different point of view for objective i' . Therefore, each of the configurations depicted in Figure 1 is evaluated with regard to the different objectives selected. Accordingly, corresponding performance measures $P_{m,i}$ and their related indices $IP_{m,i}$ need to be calculated through all n_o objectives.

2.3. Interpretation and Visualization of IP Values

To facilitate the recognition of patterns and improve the detailed interpretation of the results, the IP values can be color-coded and plotted in matrices.

For a less detailed assessment of the effect of an ME m , following reasoning presented in Morris (1991), the resulting IP values can also be aggregated to allow for a more general assessment.

Similarly, for quantifying the degree of interaction and nonlinear effects of an ME, we use the spread of its effects: we computed the absolute standardized difference ASD_m as a measure of relative range of the IP values between the corresponding configurations (e.g., $A + 5$ vs. $B - 5$):

$$ASD_m = \frac{|IP_m(A) - IP_m(B)|}{\text{mean}(IP_m(A), IP_m(B))} \quad (2)$$

3. Case Study

3.1. Materials and Methods

The MMEMO method is applied to a case study using the WASA-SED model on the mesoscale Isábena catchment.

The following sections describe the study area (section 3.1.1), introduce the WASA-SED model (section 3.1.2), the general model setup (section 3.1.3), the specific MEs (section 3.1.4), and objectives (section 3.1.5) considered in the analysis.

3.1.1. Study Area

The study is conducted on the Isábena basin located in the Southern Central Pyrenees and their foothills (NE Iberian Peninsula). The River Isábena is the main tributary of the River Ésera. Both constitute the most important tributaries of the River Cinca, in turn the second largest tributary of the Ebro (Figure 2). The Isábena drains an area of 445 km² and can be divided into six subbasins: Cabecera (146 km², representing 33% of the total catchment area), Villacarli (42 km², 9%), Carrasquero (25 km², 6%), Ceguera (28 km², 6%), Lascuarre (45 km², 10%), and the remaining outlet subbasin Capella.

The Isábena basin is characterized by heterogeneous relief (450–2,720 m above sea level) and pronounced topography. This leads to marked temperature and precipitation gradients (Verdú, 2003) and a high spatial variability in precipitation (annual precipitation: 450–1,600; mean: 770 mm). The area has a Continental Mediterranean climate, wet and cold, with both Atlantic and Mediterranean influences (García-Ruiz et al., 2001). The hydrology of the basin is characterized by a rain-snow-fed regime. Floods typically occur in spring (i.e., frontal precipitation events and snowmelt) and in late summer and autumn (i.e., localized thunderstorms). The mean annual discharge estimated at the outlet of the basin (Capella gauging station, Figure 2) is 4.1 m³/s, with a maximum instantaneous discharge of 370 m³/s recorded in August 1963 (return period of 94 years calculated by the Gumbel method from the series of annual maximum instantaneous discharge for the period 1945–2009).

The Isábena drains together with the Ésera into the Barasona Reservoir, delivering large amounts of suspended sediments which are responsible for its severe siltation problem. Instantaneous suspended sediment concentrations of up to 350 g/L have been measured at the basin outlet (López-Tarazón et al., 2009), generating mean suspended sediment loads above 250,000 t/year (i.e., period 2005–2010; López-Tarazón & Batalla, 2014) and yielding specific values of ~600 t·km⁻²·year⁻¹. These values can be considered very high in comparison with similarly sized catchments in this and other regions (de Vente et al., 2006; López-Tarazón et al., 2012). Sediments mainly originate from the middle part of both basins (viz., in Villacarli and Carrasquero

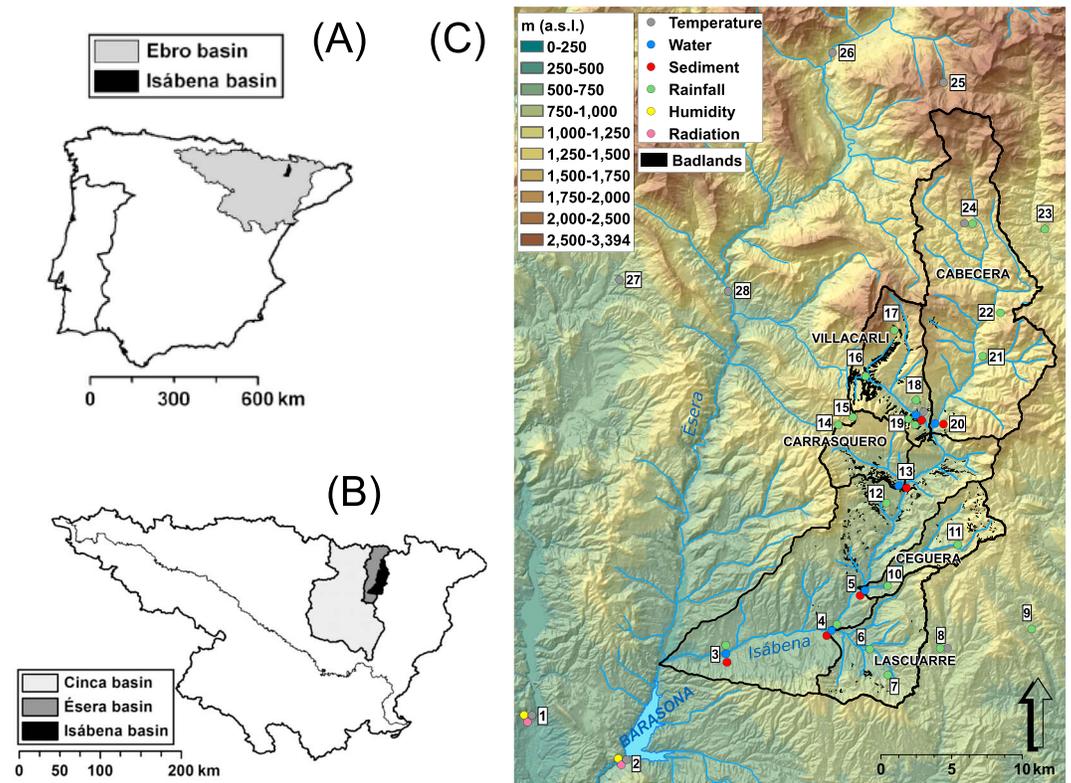


Figure 2. (a) Location of the Isábena catchment in the Ebro basin and in the Iberian Peninsula. (b) Location of the Isábena, the Ésera, and the Cinca Catchments within the Ebro River Basin. (c) Digital elevation model of the Isábena Basin, where the main subbasins, the location of the badland areas, the Barasona Reservoir, and the meteo and gauging stations are shown (see Table A1 of annex for station Ids used in the map).

subbasins in the case of the Isábena; Figure 2), from a corridor consisting of valleys excavated over Eocene marls with sandstones. At the surface, these marls form typical badland structures, which are the main source of sediment yield. Vegetation is mainly composed of deciduous woodland, agriculture, pasture, and bushes in the valley bottoms with evergreen oaks, pines, and bushes on higher ground.

The Isábena catchment has been subject to intensive hydro-sedimentological monitoring over the past decade (e.g., Bronstert et al., 2014), resulting in a favorable instrumentation situation (Figure 2), which enabled testing the multidimensional analysis presented. The time-series data used have been published and described by Francke et al. (2017b).

3.1.2. The Hydro-Sedimentological Model WASA-SED

WASA-SED (Güntner & Bronstert, 2004; Mueller et al., 2010) is a process-oriented hydro-sedimentological model with a multiscale approach to reflect catchment complexity.

The hydrological module calculates interception losses, evaporation, and transpiration using the modified Penman-Monteith approach (Shuttleworth & Wallace, 1985) and infiltration with the Green-Ampt approach (Green & Ampt, 1911) in a multilayer scheme. Snow, infiltration-excess and saturation-excess runoff as well as its lateral redistribution and subsurface runoff between the model's spatial components are considered. WASA-SED represents ground water with a linear storage approach.

The sediment module in WASA-SED provides four erosion equations of sediment generation by using derivatives of the Universal Soil Loss Equation (USLE; Williams, 1995; Wischmeier & Smith, 1981). These estimates of gross erosion can optionally be coupled to a transport-capacity limitation constrained by overland flow or a sediment delivery ratio prespecified on the scale of landscape units.

Subcatchment runoff is routed through the river network considering river exfiltration and evaporation. The water routing is based on the kinematic wave approximation after Muskingum (e.g., as described in Chow

Table 1
Overview of Model Enhancements

ID	Model enhancement	Configuration A (base)	Configuration B (all MEs)	Description	Source of uncertainty ^a
ME 1	Rainfall	Official gauges	+ Research gauges	Additional rain gauges	id, mc
ME 2	Discharge	Outlet only	+ Subbasin gauges	Additional stream gauges	ir, mc
ME 3	Land use	Official data	Remote sensing based	Refined Land-use data	ip, mc
ME 4	LAI/C-factor values	Literature		Refined LAI/C data, no seasonality	ip
ME 5	LAI/C seasonality	None	SDR ^b	Seasonality in LAI/C	ip, ms
ME 6	Connectivity index	None		Refined sediment connectivity	ms
ME 7	Flow concentration	Classic WASA	Fc index	Refined water connectivity	ms

Note. ME = model enhancement; LAI = leaf-area index; C-factor = cover factor.

^aClassification of source of uncertainty (Matott et al., 2009): Input uncertainty: id = input data; ir = response; ip = parameter; model uncertainty: mc = correspondence; mr = resolution; ms = structure. ^bSediment delivery ratio after Vigiak et al. (2012).

et al., 1988). River flow, velocity, and flow depth are calculated for each river stretch and each time step using the Manning equation.

Reservoirs can be accounted for in a lumped or explicit mode. Both modes include the reservoirs' hydrological and sediment balance to a different degree of detail (Mamede, 2008). WASA-SED is tailored for the use at the mesoscale ($\sim 10^0$ – 10^5 km²). To preserve landscape variability in an efficient manner, WASA-SED employs a hierarchical discretization approach: subcatchment, landscape unit, terrain component, and soil-vegetation component. At the subcatchment level, flow routing and water abstractions as well as reservoir management are performed. Each subcatchment contains several landscape units (LUs). These LUs group typical toposequences and similar lithology and land-surface form. They consist of terrain components (TCs) decomposing the LUs into homogenous parts. Each TC is characterized by a different slope gradient and by a specific composition of soil-vegetation-components, which combine soil types and vegetation classes. These different discretization units can be derived with an automatic catena-based discretization method (Francke, Güntner, et al., 2008).

3.1.3. General Model Setup and Experimental Design of MEMO

The modeling period spans January 2009 till April 2016. The model's initial configuration is based on mostly publicly available input data summarized in supporting information Table A2 of the annex. It is referred to as (*base*) configuration A. The model is run in hourly and daily resolution, both in uncalibrated and calibrated mode resulting in a total of $k = 2 \cdot 2 = 4$ reference configurations A (details in section 3.1.4, last paragraph). The calibration was performed for the target variables water and sediment. Each combination of MEs was calibrated separately. Guided by previous studies (Francke, 2009; Güntner, 2002), the most influential parameters for the target variables were selected for calibration (see Table A6). The objective function used in the optimization was based on the root-mean-square error (RMSE). When the configuration included flux data at the subbasin level (i.e., with ME 2), the mean of all subbasin RMSEs was used; otherwise, only the fluxes at the outlet were considered. For water, this concerns measured and modeled discharge; for sediment, monthly sediment yields (cf. Table 4). We employed the heuristic, gradient-free optimization algorithm dynamically dimensioned search (Tolson & Shoemaker, 2007) as implemented in the pps0-package (Francke, 2015; see details in see section 9.5.8 of annex).

Starting from these four configurations A, we evaluate the impact of $m_{max} = 7$ selected MEs. These MEs are listed in Table 1 and briefly summarized in the following. The annex (A10.5) contains further technical details on their derivation. The MEs have been chosen to correspond to all relevant sources of uncertainty as identified by Matott et al. (2009) and comprise the findings and developments of the SESAM-II project (Bronstert et al., 2014). Configuration B refers to the model setup in which all these are implemented, that is, added to the base configuration A (see Figure 1). Intermediate configurations originate from adding an ME to A or withholding it from B. This resulted in a total of $N = (m_{max} + 1) \cdot 2 \cdot k = 64$ configurations.

3.1.4. Description of Implemented MEs

3.1.4.1. ME 1, Rainfall—Additional Gauges

The official rain gauge network was complemented by 12 rain gauges, thus increasing rain gauge density from 1.1 to 3.6 gauges per 100 km². The resulting changes for the modeled subcatchments are summarized in Table 2. Besides changes in the mean, configuration B features higher spatial variability in rainfall

Table 2
Subbasin Statistics for Annual Rainfall and Discharge (1 January 2009 to 6 January 2014) Based on Official (Config. A) and Complemented Rain Gauge Network (Config. B) and Stream Gauges, Respectively (see Figure 2)

Subbasin	Mean annual rainfall			Discharge	
	Config. A (mm)	Config. B (mm)	+/- (%)	Mean (m ³ /s)	Max (m ³ /s)
Cabecera	859	903	5.1	2.07	91.4
Villacarli	765	852	11.4	0.27	14.3
Carrasquero	736	796	8.2	0.24	9.0
Ceguera	671	620	-7.6	0.21	22.4
Lascuarre	575	584	1.6	0.08	13.0
Capella	646	637	-1.4	4.18	123.9
q99 ₁	5	6	22.7		
q99 ₂₄	32	36	11.9		
Mean (ρ_1)	0.87	0.74	-15.5		
Mean (ρ_{24})	0.95	0.90	-4.7		

Note. q99_{xx} = 99 percentile; mean (ρ_{xx}) = mean Pearson correlation among the time series for resolution of xx hours.

and more intense peak rainfall, as attested by the decrease in mean station correlation and increase in the 99% percentile.

3.1.4.2. ME 2, Discharge and Sediment Flux at Subbasin Level

The flux measurements at the outlet were complemented by monitoring data from the five subcatchments. This increased the total length of discharge time series from 1,961 to 5,526 days (+182%).

Sediment fluxes were aggregated to the monthly level. Here adding data from subcatchment monitoring increases coverage from 33 to 156 months.

3.1.4.3. ME 3, Land Use—Improved Map

The use of satellite images, ground-truthing campaigns, and an object-oriented approach for classification allowed for the derivation of an improved land use map. Differences to the base configuration A are especially pronounced in the assignment of the grassland, shrub, and forest classes (Table 3).

3.1.4.4. ME 4, LAI-/C-Factor Values—Derivation From Remote Sensing

Fractional cover of green vegetation was assessed from optical remote sensing imagery and a multiple end-member spectral mixture analysis (MESMA) supported by extensive ground truthing. This allowed the estimation of leaf-area index (LAI). In comparison to the base configuration A, values differed by -2.3 to +2.7 m²/m² (mean: -0.54 m²/m²; see Table A4 of annex).

Using additional field survey data for cover fractions of rocks and dry vegetation, the USLE C-factor (cover factor) was calculated for the land use classes agricultural, shrubland, grassland, and badland, resulting in changes of the updated values of -0.33 to +0.17 (mean: -0.04). Spatially and temporally constant C-factor values from the literature were assigned to the remaining classes and no-data pixels (see Table A5 of the annex).

3.1.4.5. ME 5, LAI-/C-Factor Seasonality—Derivation From Remote Sensing

Based on the maps of LAI and C-factor derived for the entire study area for several dates during the growing season, an average seasonal cycle of LAI and C-factor for each land use class was obtained. This was done by fitting a cyclic linear piecewise regression to mean LAI and C-factor data for all dates and for the four most important land use classes agricultural, shrubland, grassland, and badland. Compared to configuration A, this resulted in changes of -1.6 to 1.8 m²/m² (mean: 0.52 m²/m²) for the LAI and -0.36 to 0.01 (mean: -0.09) for the C-factor. The resulting values as used in the configurations are listed in Tables A4 and A5 in the annex.

3.1.4.6. ME 6, SDR—Enhanced Representation of Sediment Retention and Delivery

The sediment delivery ratio (SDR) was computed according to Vigiak et al. (2012). This computation is based on the index of connectivity (Borselli

Table 3
Land Use Classification Results for Improved Land Use Map

Land use class	Areal fraction (%)		Difference (B - A)
	Config. A	Config. B	
Agricultural	12.9	10.5	-2.5
Badland	0.6	1.9	1.3
Forest dec. + mixed	22.8	32.2	9.4
Forest con.	6.5	18.0	11.4
Shrubland	0.0	1.7	1.7
Open soil	32.7	29.1	-3.6
Grassland	22.2	6.2	-16.0
Rocks	1.6	0.5	-1.1

et al., 2008), which, in turn, relies on DEM analysis and the spatially derived C-factor (see ME 4). For the use in the model, the SDR values were averaged for each landscape unit to match the model input structure. The resulting values range from 0.002 to 0.01 (mean: 0.005).

3.1.4.7. ME 7, Flow Concentration Index—Approximating Flow Convergence

The concentration of overland flow on hillslopes has various implications for hydrological and sediment connectivity.

Its original representation in WASA-SED (configuration A) uses a static empirical estimation (a specific fraction $f_{sheet2rill}$ of the overland flow is considered channelized and not subject to further interaction with the lower TCs, depending on the areal fractions of the TCs; see Güntner, 2002).

We extended the model by refining the estimation of $f_{sheet2rill}$ and also allowing for flow divergence by the factor $f_{rill2sheet}$. Both coefficients were parameterized using topography-driven algorithms. Thus, the respective mean values of $f_{sheet2rill}/f_{rill2sheet}$ are 0.28/0 (configuration A) and 0.82/0.22 (configuration B).

3.1.4.8. Benchmarks Calibration and Temporal Resolution

As explained in the previous section, the analysis includes $k = 4$ couples of reference configurations (hourly and daily resolution and both in uncalibrated and calibrated mode). This allows looking at the different effect of each ME within these settings, which we found of special interest from the application point of view and because modifying them is standard practice in model studies. More details on the calibration in section 10.5.8 of the annex.

By additionally employing comparisons *between* the four couples, we could also isolate the effect of calibrating and increasing the temporal resolution on the IP values. This was done by pairwise comparison of all corresponding uncalibrated and calibrated (daily and hourly) configurations, respectively, and aggregating the results. As this slightly differs from the treatment of the other MEs, calibration and temporal resolution are formally not treated as MEs here, although this also could be done in other settings.

The employed optimization algorithm includes stochastic components. To account for the resulting random effects, we replicated the calibration of the four reference base configurations 10-fold. The scatter in the resulting performance measures must be considered as inherent noise in this case, that is, changes in performance below this threshold should not be interpreted. Consequently, the respective 95 percentile ranges served to denote IP values falling below these thresholds (*insignificant improvements*).

3.1.5. Differentiating Model Performance: Looking at Different Objectives

Following the rationale of MMEMO, the effect of the MEs is measured with regard to different objectives. They include different target variables (water vs. sediment flux), spatial (at the catchment outlet vs. at the subbasins), and temporal (dynamics vs. yield) foci. Dynamics were assessed with the RMSE (equation (3)); agreement in total yield was quantified using the absolute bias P_{Bias} (equation (4)):

$$P_{RMSE} = \sqrt{\frac{\sum_{i=1}^{n_t} ((Q_{obs}(i) - Q_{sim}(i))^2)}{n_t}}, \quad (3)$$

$$P_{Bias} = \Delta t \cdot \left| \sum_{i=1}^{n_t} Q_{obs}(i) - Q_{sim}(i) \right|, \quad (4)$$

$$P_{Bias,rel} = \frac{P_{Bias}}{\Delta t \cdot \sum_{i=1}^{n_t} Q_{obs}(i)}, \quad (5)$$

where n_t is the number of time steps in the simulation period, Q_{sim} simulated flux, and Q_{obs} observed flux of water or sediment. For each objective, the corresponding performance measure P is chosen as shown in Table 4. The relative bias $P_{Bias,rel}$ (equation (5)) solely serves for reporting the model performance in the reference configurations (see section 4.1).

To ensure direct comparability with the metrics from the *daily* configurations, the results of the configurations in *hourly* resolution aggregated to daily resolution before the metrics are computed.

Table 4
Improvement Aspects (Objectives and Associated Change in Performance) Considered in the Case Study and Performance Measures Used

Improvement aspect (category of objective)	States	Number of states	Modifications to performance measure $P = P(Q_{sim}, Q_{obs})$
Target variable	Water Sediment	2	$Q_{sim} = Q_{wat}$ ^a $Q_{sim} = Q_{sed}$ ^a
Spatial focus	Subcatchment Outlet	2	$P = \text{mean}(P(\text{subbasins}))$ $P = P(\text{outlet})$
Temporal focus	Dynamics Yield	2	$P = P_{RMSE}$ $P = P_{Bias}$

Note. The total number of objectives considered is $2^3 = 8$.
 P = performance measure, Q_{wat} = simulated discharge; Q_{sed} = simulated sediment flux; units = water: (V/T); sediment = (M/T).
^aFor sediment, Q_{sim} and Q_{obs} are aggregated to monthly values to reduce the effect of intermediate storage in the riverbed.

4. Results

4.1. Model Performance in the Reference Configurations

The following analysis focuses on the improvements resulting from the different MEs, that is, the changes they caused. For putting these into perspective, the performances of the reference configurations are given in Table 5. The full list of calculated performance measures can be found in the Supporting Information S1.

For all objectives, the performance measures vary greatly. Although calibration used only a single objective function targeting dynamics, the calibrated configurations outperform their uncalibrated counterparts in all other objectives, too. The B configurations usually do not yield a higher performance than the A configurations, despite comprising all considered MEs. This already reveals that at least some of the MEs do not improve model performance. Applying the MEMMO methodology as described in the following sections provides further insights.

4.2. Effect of a Single Model Enhancement on Multiple Objectives

As showcase, we present the effect of *one* ME (ME 1, improved rainfall) on *all* considered objectives of improvement. Figure 3 shows that the denser rain gauge network mostly leads to an improved model performance. This improvement is apparent for all water-related metrics, especially concerning yield. Conversely, the sediment metrics mainly experience insignificant (hashed cells) or negative changes in performance. This degradation is especially apparent for the dynamics at the outlet, for both water and sediment metrics.

Table 5
Performance P of Reference Configurations

Objective	Water				Sediment					
	Sub		Out		Sub		Out			
	Dynamics ^a	Yield ^b	Dynamics ^a	Yield ^b	Dynamics ^c	Yield ^b	Dynamics ^c	Yield ^b		
A	Daily	Uncalibrated	1.19	0.38	3.86	0.40	13,244	10.60	33,213	1.37
		Calibrated	1.02	0.44	2.84	0.29	7,479	4.59	12,755	0.12
	Hourly	Uncalibrated	1.32	0.40	4.11	0.45	7,560	2.05	19,976	0.42
		Calibrated	1.22	0.44	3.29	0.37	7,031	6.27	11,513	0.00
B	Daily	Uncalibrated	1.34	0.47	4.42	0.55	8,048	0.95	26,331	0.99
		Calibrated	1.13	0.45	3.34	0.44	7,907	0.83	25,893	0.96
	Hourly	Uncalibrated	1.46	0.46	4.60	0.58	8,102	0.99	26,523	1.00
		Calibrated	1.31	0.43	3.82	0.49	6,973	2.45	16,916	0.40

Note. Performance measures of dynamics expressed as P_{RMSE} (equation (3)), yield as $P_{Bias,rel}$ (see equation (5) and Table 4).
^a[m³/s]. ^b[-]. ^c[t/month].

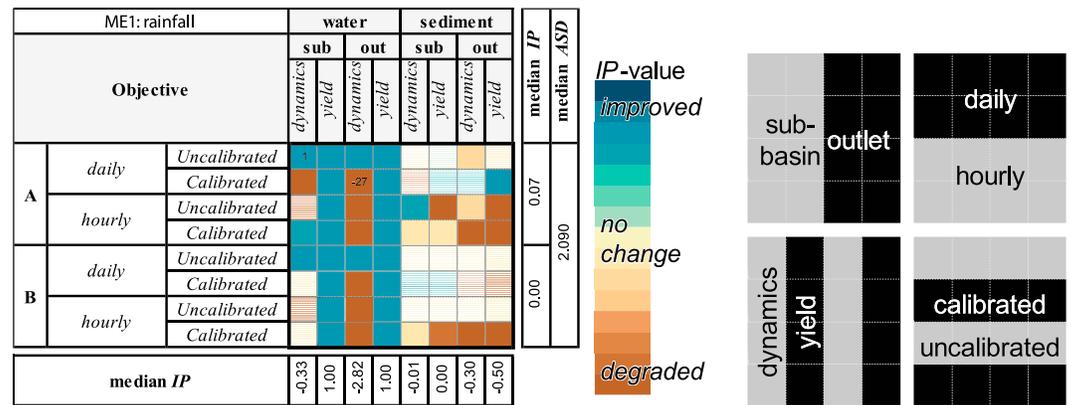


Figure 3. Effect of model enhancement ME 1 (rainfall) expressed as *IP* value. The *IP* values are color-coded, with the maximum and minimum added in black. *IP* values below the noise level are displayed with a line pattern. The matrices are arranged such that the effect of an ME concerning a certain objective is plotted in a contiguous way (see scheme on the right). ME = model enhancement.

The other MEs also show quite diverse effects on the different objectives (see 10.6 in annex). Generalizable patterns are rare among the investigated MEs.

Thus, the effect of any ME must be assessed separately for each objective of interest; a positive effect of an ME for one objective does not necessarily correspond to the improvement of another.

The matrices are arranged such that the effect of an ME concerning a certain objective is plotted in a contiguous way (see scheme on the right).

Thus, while such complex matrices provide detailed information for each single ME and for assessing the effect of interactions (see section 4.3), comparing the effect of multiple MEs requires assessments aggregated by the A/B configurations (see section 4.4) or by the metrics (see section 4.5). Figure 4 illustrates these steps.

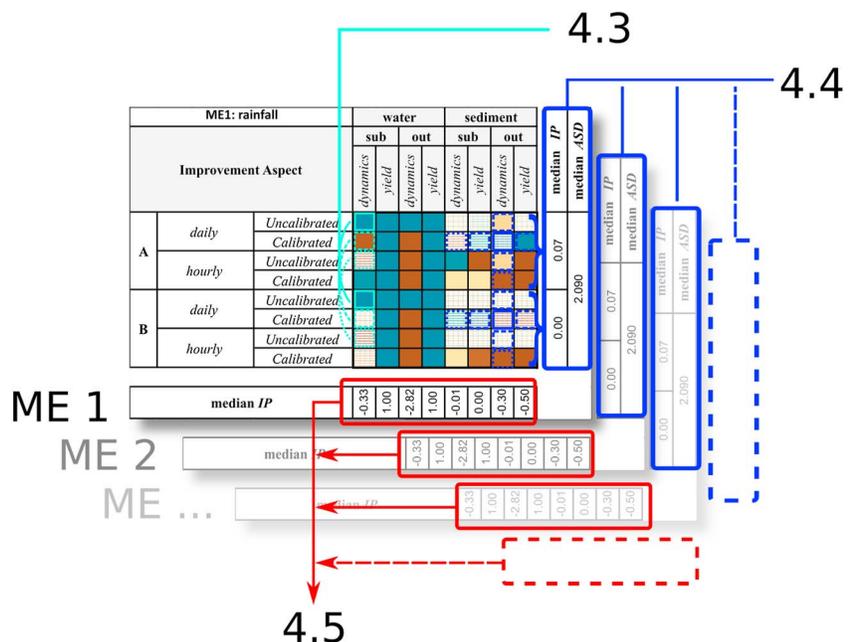


Figure 4. Schematic illustration of aggregating the detailed results. The numbers refer to the following paper sections. ME = model enhancement.

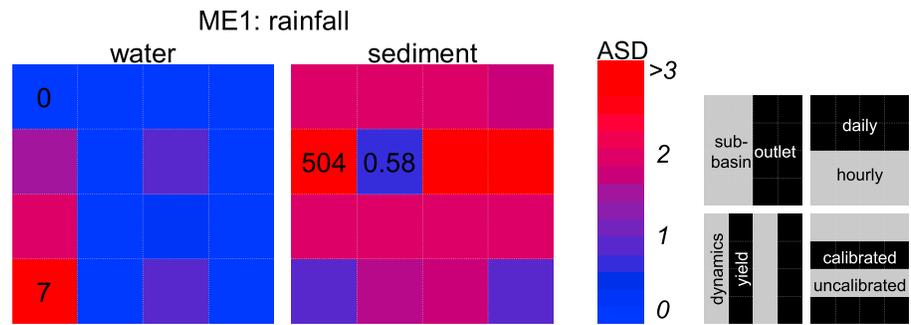


Figure 5. ASD values (relative range of the *IP* values between the corresponding A and B configurations) for all considered metrics as arranged in Figure 5) for ME 1. ME = model enhancement.

4.3. Quantifying Interaction

Differences between the A and B configurations are a result of the interaction of the MEs: the effect of an ME can be different, whether it is applied to a base configuration A or to full configuration B. To quantify the apparent high degree of interaction between the MEs, we employed the ASD, that is, the difference between the *IP* values of corresponding configurations (e.g., “A + 5” vs. “B – 5”, see equation (2)).

Figure 5 suggests that this interaction varies considerably between the different objectives: while water-related objectives seem to be less affected, the calibrated sediment model in daily resolution shows the largest effects with ASD values of over 500.

Likewise, the ASD values for the other MEs show great variability without appreciable patterns (Figure 13, annex). This implies that interactions were observed for some objectives, while they were absent for others.

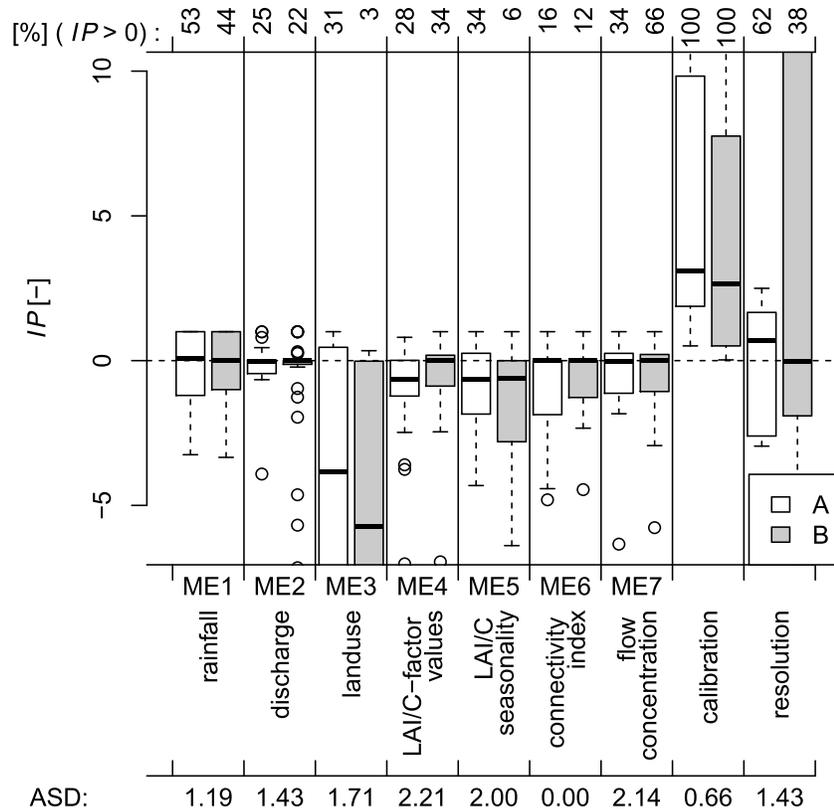


Figure 6. *IP* values of all objectives, grouped by ME and forward/backward selection (A+/B-). The numbers above the panel report the percentage of the positive *IP* value in the respective group. The values below the bottom line show median ASD values for each ME. ME = model enhancement.

However, the values suggest that for roughly half of the MEs, pronounced effects of interaction must be expected, which cannot be inferred from traditional OAT approaches. For this reason, these results point at the difficulties to generalize the effect of a specific ME for other case studies and emphasize the need for identifying such interactions. The following section summarizes these findings to allow the comparison of the MEs.

4.4. Intercomparison of MEs and Their Interaction

Figure 6 summarizes the two previous sections, that is, it aggregates the ME's impacts over *all* objectives. There is large scatter in the *IP* values caused by the different MEs. Most *IP* values fall within the range between -1 and 1 . Especially for MEs 2, 4, 6, and 7, this range is considerably smaller, suggesting a generally limited influence of the respective MEs.

In about 40% of the cases, the effect of an ME is positive (i.e., improving model performance) with A configurations (43%) benefitting more than B configurations (36%). With around 50% of the configurations being improved by rainfall (ME 1), this ME provokes the most beneficial effects on its introduction. However, some MEs (e.g., ME 3) only rarely cause an improvement of the model, such as 31% of the A configurations and only 3% of the B configurations (see top axis of Figure 6). In contrast, calibrating the model or increasing its resolution yields much higher *IP* values. With their value exceeding 1 , this suggests that the associated model improvement is often much stronger than of any ME. They also cause much more consistent improvements, with 100% and 50% improved configurations, respectively.

The median ASD values per MEs (Figure 6, bottom line numbers) are a measure of an ME's interactions. A low value suggests low interaction, for example, ME 6, whereas a high value indicates high interactions. For example, ME 4, 5, and 7 have high ASD values. Thus, their impact depends strongly on the configuration context.

4.5. Ranking of MEs

From the point of model application, it may be of interest to know (I) which ME is most promising to improve a given objective; or, conversely, (II) which objective is most improved by an ME. These questions are addressed in Figure 7, which ranks the MEs according to median *IP* values.

Concerning (I), Figure 7 (colored bars at left of cells) suggests that for many objectives, improved rainfall data (ME 1) and subbasin fluxes (ME 2) are the most improving MEs, while the modeling benefits the least from improved land use data (ME 3). Specifically for sediments, surprisingly connectivity index (ME 6) causes the least improvement. All improvements obtained by the MEs are considerably lower than the effect of calibration. Regarding sediments, also increased resolution yields greater benefits than any of the MEs.

Concerning (II), Figure 7 (colored bars at bottom of cells) shows that ME 1 has its largest positive impact on water yield metrics. Its most detrimental effect concerns water dynamics, which also applies to ME 2–5.

Thus, despite the multidimensionality of the improvements, in the current case, some MEs like rainfall (ME 1) clearly provide the greatest benefit to the model. Figure 8 visualizes this in detail for every objective separately by displaying the two most beneficial MEs. Again, rainfall (ME 1) scores the highest number of objectives, especially regarding water fluxes, namely, water yield. LAI/C-factor values (ME 4) and LAI/C-seasonality (ME 5) also seem to play some role in the B and A configurations, respectively.

For sediment, the picture is less clear: Land use (ME 3), connectivity index (ME 6), and flow concentration (ME 7) seem to predominate but each for specific objectives only (A configurations, uncalibrated daily, and uncalibrated B configurations, respectively).

Many objectives could be improved *only* by as single ME (viz., ME 1 or 2), that is, all other MEs did not yield any improvement at all (cells without circles in the matrix). For nine objectives, none of the MEs had a positive effect (white cells). For these, the performance was only increased by calibration (two objectives) or temporal resolution (three objectives). Four objectives of the calibrated hourly configurations were never improved.

5. Discussion

5.1. Toward Further Improvements of the Model—What Can We Learn?

The presented case study benefitted from a favorable data situation which allowed the comparison of multiple, very diverse MEs with the MMEMO approach. Including both water and sediment in the modeling shed

		Objectives							
		water				sediment			
		sub		out		sub		out	
		dynamics	yield	dynamics	yield	dynamics	yield	dynamics	yield
median IP	ME1	-0.33	1	-2.82	1	-0.01	0	-0.3	-0.5
	ME2	0.5	0	-1.96	0	0.33	-0.02	-0.02	-0.14
	ME3	-12.8	-5.23	-19.6	-5.34	-0.03	-0.04	0.06	-0.02
	ME4	-1.06	-0.53	-3.12	-0.42	0	-0.02	0	0
	ME5	-2.09	-0.09	-3.57	-0.08	-0.4	-0.26	-1.02	-0.31
	ME6	0	0	0	0	-0.37	-1.34	-2.18	-3.25
	ME7	-0.5	-0.01	-0.42	-0.02	0.06	-0.02	0	-0.02
	calibration	18.3	2.29	37.4	2.98	1	0.78	2.8	3.35
resolution	-4.92	-2.85	-11.3	-1.78	0.76	0.85	1.67	1.74	

Figure 7. Comparative effect of model enhancements, expressed by median IP values (cf. bottom line summaries in Figures 3 and 12 in the annex). Bars at the bottom/left cell margin indicate the two highest (green) and lowest (red) IP values of each row/column. Thick bars mark the highest (green)/lowest (red) value; thin bars mark the second highest (green) /second lowest (red) value.

additional light on some of the MEs. Still, some characteristics of the case study must be acknowledged that could affect specific conclusions drawn thereof (see 10.8 of annex).

Applying the MMEMO approach revealed that the resulting effect of adding or withholding these MEs on performance metrics was very diverse and depended on the metric and context. Thus, it confirmed the necessity of this multidimensional approach to identify the different behaviors of the MEs.

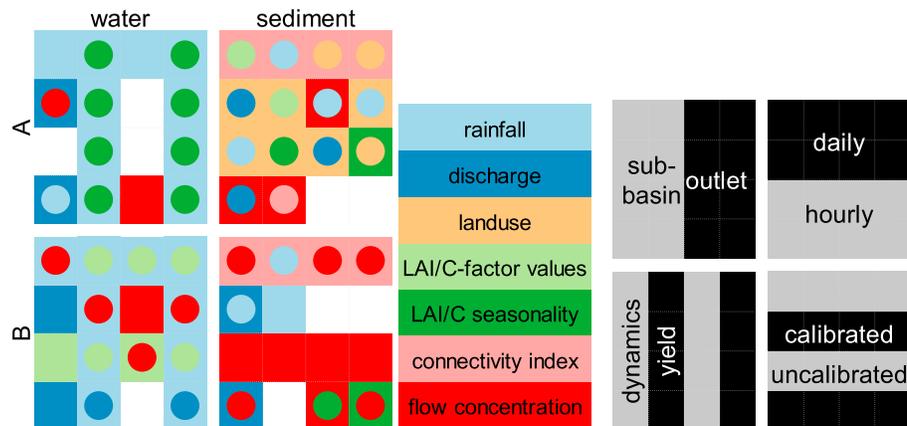


Figure 8. The two most beneficial model enhancements for each objective. Each cell of the colored matrices denotes the most beneficial model enhancement (color of cell) and the second-to-best (colored circle). The layout is the same as in Figure 3. Example: Top-left cell: The performance for modeling of water fluxes in the A configuration at the subbasin level, in daily resolution, in terms of dynamics, for the uncalibrated model was most improved by model enhancement 1 (rainfall, light blue cell color). No other model enhancement caused any improvement for this constellation (no colored circle in this cell). LAI = leaf-area index; C-factor = cover factor.

Ranking first among the MEs, additional *rainfall* (ME 1) proved beneficial for roughly half of the objectives. This ME also showed one of the least tendencies of interaction, suggesting that a gain in model performance can be expected for similar model setups. This corroborates the common assumption that precipitation is generally the biggest driver of uncertainty (Guzman et al., 2015). However, as pointed out by Fencia et al. (2008), improved model performance due to improved rainfall data has mainly been confirmed for *virtual experiments*, while studies using real data are less common.

On average, land use information (ME 3) and improved LAI/C-factor values and their dynamics (MEs 4 and 5) yielded little or no benefit. However, the detailed analysis revealed that these MEs still can be among the most powerful for very specific objectives (see Figure 8).

In virtually all cases, calibration caused considerably larger improvements than any of the MEs; for sediment metrics, this was also true for increasing the temporal resolution. In other words, although the employed model WASA-SED is a semiprocess-based model (Bronstert et al., 2014; Mueller et al., 2010) for data-scarce situations, the additional information contained in the added MEs never produced the same increase in performance as did calibration and changed resolution. A similar conclusion was gained by Butts et al. (2004), who found by comparing 10 different model structures, that model calibration was always essential to yield a *good* model performance. This suggests that for the used model, the additional information content of the tested MEs is still quite low when compared to calibration. Consequently, even for this catchment with relatively good input data situation, calibration is required to achieve high model performance. In addition, a wide range of observations can enable both modelers and field observers to identify dominant hydrological systems dynamics and thus constrain not only parameter but also structural model uncertainty (e.g., Blazkova & Beven, 2009; Graeff et al., 2009). Thus, sufficient observations of the target variable are paramount.

However, prediction (i.e., maximum model performance) is not the sole purpose of modeling (Epstein, 2008). Especially regarding explanation, process understanding, and the verification of theories, the abovementioned phenomena deserve further discussion.

Considering the fidelity of represented processes, we would expect better rainfall data (ME 1) to result in a representation of related processes closer to reality. This is not necessarily the case for calibration, which may overfit a model to observations at the cost of the adequate representations of processes (process fidelity = "right" answer for the "right" reason). In this sense, the apparent improvement of sediment modeling due to the connectivity index (ME 6) for some objectives is an improvement in model performance but not of process fidelity: in the WASA-SED, the SDR concept is an effective formulation of sediment delivery at the scale of the landscape unit that necessarily decreases process fidelity. Without SDR values, the model calculates sediment delivery based on surface runoff, surface roughness, and intrahillslope flow convergence at a smaller model unit scale. Consequently, it constitutes a more process-oriented approach. In this sense, the effect of ME 6 being the strongest for uncalibrated configurations is conclusive.

Conversely, land use (ME 3) and LAI/C-factor seasonality have a very close process-related correspondence in the employed model. Still, their effect on the water metrics is predominantly negative. This could be related to (a) deficits in the presumably improved data or (b) inadequacies in the model concept. Concerning (a), uncertainties associated with the remote sensing data must be expected (see classification accuracies in Table A3). Furthermore, for each land use class, the areal fraction differences between the two configurations must be seen in relation to the parameters associated with the respective land use classes, which can amplify or dampen the effect. Probably even more severe is the class-wise spatial averaging in preprocessing the model input. Thus, the full level of spatial detail as obtained from remote sensing was not fully utilized in the study. Regarding (b), simplifications in the process description come to mind. Specifically, for the study area, the real seasonal dynamics in land cover are driven by temperature, water availability, and agricultural management practices. Thus, they depend on both weather patterns and crop cycles and may vary between years, as opposed to an invariant seasonal cycle based on a limited number of tie points assumed in the employed model. Furthermore, the seasonality of other cycles (rooting depth, macroporosity, soil cohesion, ...) may be predominant but was not included in the data. Thus, this ME especially qualifies for improvements in the next iteration of model refinement.

For the calibration, we chose uni-objective functions for water and sediment. Evidently, this choice will result in models optimized in a fashion that the respective objective function aims at (e.g., good match of dynamics and high flows). Thus, MEs that potentially improve other objectives (i.e., low flows) are not exploited in their

full potential and may therefore have appeared less beneficial. Choosing other objective functions or multi-objective optimization could remedy this shortcoming, if the additional computation times are feasible. Likewise, observational data uncertainty should be considered to prevent sheer numerical overfitting. Several studies have shown that observation uncertainties can be at least in the same range as model uncertainties, be it the errors related with discharge rating curve (Blazkova & Beven, 2009), sediment yield computation (Francke, López-Tarazón, et al., 2008), inaccuracies of soil moisture (Zehe et al., 2010), or uncertain evapotranspiration rates or groundwater levels. Besides using observation values (more or less *hard data*), recently, some authors have proposed to include *soft data* into the analysis and modeling of hydrological systems, such as Seibert and McDonnell (2002), who use field hydrologists' estimates to improve the realism of conceptual models, Hartmann et al. (2015), who use a priori information to constrain modeling uncertainty on a continental scale, and Sebok et al. (2016), who integrated expert knowledge to restrict the uncertainty in the mesoscale catchment water balance. All these refinements could be accommodated in MMEMO and further sharpen its outcomes.

The results of the case study cannot be generalized. They depend on the specific catchment, the chosen target variables, performance measures, degree of detail and quality of the data, and the model used. However, *connectivity index*, *rainfall*, and *calibration* showed the least tendency of interaction. Thus, they are the most promising candidates to transferring the findings of this study to comparable model setups. Therefore, we encourage other researchers to evaluate their MEs using MMEMO and report the results. We hypothesize that a collection of such reports could become a valuable guideline for setting up catchment models and monitoring schemes. Such a guideline could provide a sound basis for estimating benefits of different model enhancements versus incurred efforts in data collection and modeling. If at all, so far, such an advice can only be obtained from subjective expert judgments and the results of only partly transferable studies.

5.2. Merits, Limitations, and Potential Improvements of the Framework

The MMEMO approach aims to compare the potential of different model enhancements in terms of increasing the model performance. The approach explicitly accounts for the high dimensionality of the problem. Some additional features still merit further discussion.

MMEMO is capable of dealing with MEs concerning different sources of uncertainty, for example, input data, parametric, structural (see Table 1). Regarding structural uncertainties, elaborate approaches targeting this issue exist, for example, FUSE (Clark et al., 2008), ECHSE (Kneis, 2015) or SUMMA Alternatives (Clark et al., 2015). MMEMO could be combined with these to allow for more complex assessments in that matter.

The proposed framework is capable of quantifying interactions between MEs. We assume that the existence of such interactions are not specific to our case study but are, in fact, quite common. Thus, the findings of any OAT approach for an ME are unlikely to be generalizable per se. Conversely, identifying MEs showing a low tendency of interaction in different model setups would enable one to generalize the findings of existing OAT studies with greater confidence.

Forward selection and backward elimination were performed only on the two most extreme configurations (i.e., base configuration A and configuration B with all MEs), which effectively resembles an OAT approach with two starting points. This revealed the pronounced interaction of some MEs. Nevertheless, even more intermediate configurations do exist. These could also be sampled, but computational demand increases tremendously (for this case study, it would have increased the computational load 16-fold). Similarly, we only considered MEs in binary states, that is, added or withheld (A+ vs. B−). However, for some MEs, a gradual information could be desirable, for example “until what number of additional rain gauges is model performance still improving?”. A more comprehensive strategy would include all these possible combinations of MEs exhaustively. This would clearly cover all possible interactions of MEs. But while performing such an analysis for a single ME is trivial, it becomes quickly unmanageable to implement in an exhaustive way, as the number of model configurations grows exponentially. At this stage, MMEMO constitutes a compromise and can be considered as a screening approach before conducting extensive sampling design as requested in global sensitivity analysis. Its results can support the need (or not) for further analysis where more specific objectives and model enhancements are required.

For the case study, we evaluated improvement regarding the objectives listed in Table 4.

We assume that the selected objectives already yield a relatively differentiated view of the MEs included. However, this table could be further extended in the number of states considered (through more target variables if measurements exist; intermediate temporal resolutions, etc.). Our choice of performance measures focused on two general categories of objectives, namely, dynamics and yield on two spatial scales. Other more specific measures could be easily implemented, as they are only evaluated on the calibrated models and do not require more simulations. Likewise, computing the performance measures also from a validation data set can additionally reduce the effect of overfitting. However, with these extensions, some of the more detailed visualization schemes and their interpretation would become more difficult or impossible, as the proposed way of using colored matrices can hardly accommodate more measures. Nevertheless, the aggregated assessments (median of the *IP* and *ASD* values) could still work.

We assessed the effect of the MEs on the entire time series. However, various studies have shown that this may change over time or as a function of the condition of the catchment (Ghasemizade et al., 2017; Herman et al., 2013; Pianosi & Wagener, 2016; Reusser et al., 2011). Computing the proposed *IP* and *ASD* values for a moving window would enable related analyses, although the results will likely depend on the choice of the window size. However, the postprocessing of the results could be refined concerning temporal variability of the effects, if of interest.

We considered an ME beneficial, whenever the *IP* value exceeded the noise that was inherent in the randomness of the calibration. Still, as we used *crisp* performance measures, the resulting *IP* values may be affected by sheer numerical changes in the performance measure, which may not always reflect a true and significant improvement of the model, for example, regarding uncertainty of the observations. In this regard, combining uncertainty analysis and sensitivity analysis in a unified framework could be considered to remedy this shortcoming (Baroni & Tarantola, 2014). Methods assessing the significance of difference in model performance (e.g. Craig et al., 2016) could also be included in the approach.

6. Summary and Conclusions

This study proposes a method (MMEMO) to quantify the effect of different model enhancements (MEs) on the performance of an environmental model. This approach aims to fill a niche among the available methods toward a transparent and straightforward evaluation of the effects of multiple MEs on model performance with regard to different evaluation objectives. By extending the OAT approach to the two (extreme) starting positions, it allows the identification of interaction effects. MMEMO specifically considers nonscalar sources of uncertainty expressed as concrete realizations (MEs) of the specific features. In contrast to making numerical exercises based on strong assumptions, MMEMO is tailored for and demonstrated with an example of real observations. The method emphasizes the multiple aspects (i.e., objectives) of model performance and allows ranking the corresponding importance of the MEs. This can prevent wrong conclusions on the importance of MEs that may arise from simple OAT approaches.

The method is demonstrated using a real-world case study of hydro-sedimentological modeling for a meso-scale mountainous catchment in NE Spain. Built on extensive previous studies, seven MEs had been selected as the most relevant enhancements to potentially improve model performance based on an existing base model configuration. Additional *rainfall* (ME 1) proved to yield the most improvement of the model with regard to several objectives, while *calibration* and increased temporal *resolution* mostly outperformed all of the MEs. This underlined the role of sufficient observations and tailored calibration approaches to obtain correct process representations.

While these specific results may not be directly transferrable to other study areas or models, MMEMO may be adopted in other modeling studies to analyze the effect of MEs. Such findings will provide useful guidance in efficiently allocating resources when a model needs to be improved. With its focus on real observations, it could potentially strengthen collaborations between experimentalists (by guiding monitoring) and modelers (by understanding processes and the model predictive capability).

Data and Software Availability and Requirements

The model used data described in Francke et al. (2017b), also available in CUAHSI format via http://hiscentral.cuahsi.org/pub_network.aspx?n=5622 or Francke et al. (2017a).

The scripts to conduct the analysis and visualization are freely available under github.com/TiIf/MMEMO and are described in the Supporting Information, which also contain the central file of the results (Francke_collected_performance_measures.xlsx).

Additional FOSS software used in the study includes the GIS GRASS 6.4.1 (GRASS Development Team, 2012) and the statistical software R 3.3.x (R Core Team, 2016) and its packages. Simulations were performed with WASA-SED (Bronstert et al., 2014; Mueller et al., 2010), revision 250 (<https://github.com/TiIf/WASA-SED>).

Acknowledgments

This research was carried out within the project "Generation, transport and retention of water and suspended sediments in large dryland catchments: Monitoring and integrated modelling of fluxes and connectivity phenomena" (BR 1731/11, GU 987/5, FO 754/1) funded by the "Deutsche Forschungsgemeinschaft" (DFG). During the elaboration of the manuscript, José Andrés López-Tarazón from the SCARCE-CONSOLIDER project (ref. CSD2009-00065) funded by the Spanish Ministry of Economy and Competitiveness and a Marie Curie Intra-European Fellowship (Project "Floodhazards," PIEF-GA-2013-622468, Seventh EU Framework Programme) and second by a Vicenç Mut postdoctoral fellowship (CAIB PD/038/2016). He also acknowledges the Secretariat for Universities and Research of the Department of the Economy and Knowledge of the Autonomous Government of Catalonia for supporting the Consolidated Research Group 2014 SGR 645 (RIUS-Fluvial Dynamics Research Group). G. Baroni received financial support from the Deutsche Forschungsgemeinschaft (DFG) under CI 26/13-1 in the framework of the research unit FOR 2131 "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces". We gratefully acknowledge the fieldwork conducted by numerous colleagues and the analyses done within the MSC theses of Charlotte Wilczok, Anja Kroll, Iris Heine, and Marcus Bauer. We thank the Ebro Water Authorities for permission to install the measuring equipment at the Capella gauging station and for providing hydrological data and Marco Cavalli for providing the index of connectivity script. RapidEye satellite data were provided by the RapidEye Science Archive (RESA) at DLR with resources of the German Federal Ministry of Economic Affairs and Energy. The airborne campaign was conducted by NERC (National Environment Research Council, UK) with funding from EUFAR Transnational Access. This manuscript greatly benefitted from the comments of three anonymous reviewers. Special thanks go to the associate editor Andrew Western for his suggestions helping to improve the manuscript.

References

- Baroni, G., & Tarantola, S. (2014). A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling and Software*, *51*, 26–34. <https://doi.org/10.1016/j.envsoft.2013.09.022>
- Baroni, G., Zink, M., Kumar, R., Samaniego, L., & Attinger, S. (2017). Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. *Hydrology and Earth System Sciences*, *21*(5), 2301–2320. <https://doi.org/10.5194/hess-21-2301-2017>
- Beven, K. (2007). Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process. *Hydrology and Earth System Sciences*, *11*(1), 460–467. <https://doi.org/10.5194/hess-11-460-2007>
- Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, *45*, W00B16. <https://doi.org/10.1029/2007WR006726>
- Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., & Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. *Water Resources Research*, *53*, 7933–7950. <https://doi.org/10.1002/2017WR020767>
- Borselli, L., Cassi, P., & Torri, D. (2008). Prolegomena to sediment and flow connectivity in the landscape: A GIS and field numerical assessment. *Catena*, *75*(3), 268–277. <https://doi.org/10.1016/j.catena.2008.07.006>
- 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>
- Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, *298*(1–4), 242–266. <https://doi.org/10.1016/j.jhydrol.2004.03.042>
- Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, *22*(10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>
- Chen, J., Brissette, F. P. F., Poulin, A., & Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, *47*, W12509. <https://doi.org/10.1029/2011WR010602>
- Chow, V. T., De R. M., & Mays, L. W. (1988). *Applied hydrology*. New York: McGraw-Hill.
- 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>
- Craig, J., Sgro, N., & Tolson, B. (2016). Evaluating prospective hydrological model improvements with consideration of data and model uncertainty. *Geophysical Research Abstracts*, *18*, 10646.
- de Vente, J., Poesen, J., Bazzoffi, P., Van Rompaey, A., & Verstraeten, G. (2006). Predicting catchment sediment yield in Mediterranean environments: the importance of sediment sources and connectivity in Italian drainage basins. *Earth Surface Processes and Landforms*, *31*(8), 1017–1034. <https://doi.org/10.1002/esp.1305>
- Droogers, P., Van Loon, A., & Immerzeel, W. W. (2008). Quantifying the impact of model inaccuracy in climate change impact assessment studies using an agro-hydrological model. *Hydrology and Earth System Sciences*, *12*(2), 669–678. <https://doi.org/10.5194/hess-12-669-2008>
- Epstein, J. M. (2008). Why Model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12.
- Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, *44*, W01402. <https://doi.org/10.1029/2006WR005563>
- Ferretti, F., Saltelli, A., & Tarantola, S. (2016). Trends in sensitivity analysis practice in the last decade. *Science of the Total Environment*, *568*, 666–670. <https://doi.org/10.1016/j.scitotenv.2016.02.133>
- Francke, T. (2009). *Measurement and Modelling of Water and Sediment Fluxes in Meso-Scale Dryland Catchments*. Potsdam: Universität Potsdam. Retrieved from <http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-31525>
- Francke, T. (2015). pps0—R-package pps0: Particle Swarm Optimization and Dynamically Dimensioned Search, optionally using parallel computing based on Rmpi, rev. 0.9–9., R-package. Retrieved from rforge.net/ppso/
- Francke, T., Foerster, S., Brosinsky, A., Sommerer, E., López-Tarazón, J. A., Güntner, A., et al. (2017a). Hydro-sedimentological dataset for the mesoscale mountainous Isábena catchment, NE Spain, GFZ Data Services. <https://doi.org/10.5880/dfgeo.2017.003>
- Francke, T., Foerster, S., Brosinsky, A., Sommerer, E., Lopez-Tarazon, J. A., Güntner, A., et al. (2017b). 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 Discussions*, *1–17*, 1–17. <https://doi.org/10.5194/essd-2017-72>
- Francke, T., Güntner, A., Mamede, G., Müller, E. N., & Bronstert, A. (2008). Automated catena-based discretization of landscapes for the derivation of hydrological modelling units. *International Journal of Geographical Information Science*, *22*(2), 111–132. <https://doi.org/10.1080/13658810701300873>
- Francke, T., López-Tarazón, J. A., Vericat, D., Bronstert, A., & Batalla, R. J. (2008). Flood-based analysis of high-magnitude sediment transport using a non-parametric method. *Earth Surface Processes and Landforms*, *33*(13), 2064–2077. <https://doi.org/10.1002/esp.1654>
- 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: Geofoma.
- Ghasemzade, 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 and Software*, *88*, 22–34. <https://doi.org/10.1016/j.envsoft.2016.10.011>

- Graeff, T., Zehe, E., Reusser, D., Lück, E., Schröder, B., Wenk, G., et al. (2009). Process identification through rejection of model structures in a mid-mountainous rural catchment: Observations of rainfall-runoff response, geophysical conditions and model inter-comparison. *Hydrological Processes*, 23(5), 702–718. <https://doi.org/10.1002/hyp.7171>
- GRASS Development Team (2012). Geographic Resources Analysis Support System (GRASS GIS) Software, Version 6.4.4, Open Source Geospatial Foundation. Retrieved from <http://grass.osgeo.org>
- Green, W. H., & Ampt, G. (1911). Studies on soil physics part I—The flow of air and water through soils. *The Journal of Agricultural Science*, 4(1), 1–24.
- Güntner, A. (2002). Large-scale hydrological modelling in the semi-arid north east of Brazil, (PhD-thesis). University of Potsdam, Potsdam, Germany. Retrieved from <http://opus.kobv.de/ubp/volltexte/2003/62/>
- 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>
- Guzman, J. A., Shirmohammadi, A., Sadeghi, A. M., Wang, X., Chu, M. L., Jha, M. K., et al. (2015). Uncertainty Considerations in Calibration and Validation of Hydrologic and Water Quality Models. *Transactions of the ASABE*, 58(6), 1745–1762. <https://doi.org/10.13031/trans.58.10710>
- Haghnegahdar, A., & Razavi, S. (2017). Insights into sensitivity analysis of Earth and environmental systems models: On the impact of parameter perturbation scale. *Environmental Modelling and Software*, 95, 115–131. <https://doi.org/10.1016/j.envsoft.2017.03.031>
- Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y., & Wagener, T. (2015). A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geoscientific Model Development*, 8(6), 1729–1746. <https://doi.org/10.5194/gmd-8-1729-2015>
- 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>
- Iooss, B., Pujol, G., & Janon, A. (2016). R package: sensitivity: Global Sensitivity Analysis of Model Outputs. Retrieved from <http://cran.r-project.org/package=sensitivity>
- Kneis, D. (2015). A lightweight framework for rapid development of object-based hydrological model engines. *Environmental Modelling and Software*, 68, 110–121. <https://doi.org/10.1016/j.envsoft.2015.02.009>
- 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>
- Loosvelt, L., Vernieuwe, H., Pauwels, V. R. N., De Baets, B., & Verhoest, N. E. C. (2013). Local sensitivity analysis for compositional data with application to soil texture in hydrologic modelling. *Hydrology and Earth System Sciences*, 17(2), 461–478. <https://doi.org/10.5194/hess-17-461-2013>
- López-Tarazón, J. A., & Batalla, R. J. (2014). Dominant discharges for suspended sediment transport in a highly active Pyrenean river. *Journal of Soils and Sediments*, 14(12), 2019–2030. <https://doi.org/10.1007/s11368-014-0961-x>
- López-Tarazón, J. A., Batalla, R. J., Vericat, D., & Francke, T. (2009). Suspended sediment transport in a highly erodible catchment: The River Isábena (Southern Pyrenees). *Geomorphology*, 109(3–4), 210–221. <https://doi.org/10.1016/j.geomorph.2009.03.003>
- López-Tarazón, J. A., Batalla, R. J., Vericat, D., & Francke, T. (2012). The sediment budget of a highly dynamic mesoscale catchment: The river Isábena. *Geomorphology*, 138(1), 15–28. <https://doi.org/10.1016/j.geomorph.2011.08.020>
- Mamede, G. (2008). *Reservoir Sedimentation in Dryland Catchments: Modelling and Management*. Potsdam: Universität Potsdam. Retrieved from <http://opus.kobv.de/ubp/volltexte/2008/1704/>
- Matott, L. S., Babendreier, J. E., & Purucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research*, 45, W06421. <https://doi.org/10.1029/2008WR007301>
- Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161. <https://doi.org/10.2307/1269043>
- 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>
- Paton, F. L., Maier, H. R., & Dandy, G. C. (2013). Relative magnitudes of sources of uncertainty in assessing climate change impacts on water supply security for the southern Adelaide water supply system. *Water Resources Research*, 49, 1643–1667. <https://doi.org/10.1002/wrcr.20153>
- Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., & Wheatley, H. S. (2011). Catchment Scale Hydrological Modelling: A Review Of Model Types, Calibration Approaches And Uncertainty Analysis Methods In The Context Of Recent Developments In Technology And Applications. *Global NEST Journal*, 13(3), 193–214.
- 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 and Software*, 79, 214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for global sensitivity analysis. *Environmental Modelling and Software*, 70, 80–85. <https://doi.org/10.1016/j.envsoft.2015.04.009>
- Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling and Software*, 67, 1–11. <https://doi.org/10.1016/j.envsoft.2015.01.004>
- Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis: Understanding the time-varying importance of uncertainty sources. *Hydrological Processes*, 30(22), 3991–4003. <https://doi.org/10.1002/hyp.10968>
- R Core Team (2016). R: A Language and Environment for Statistical Computing.
- Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., & Uijlenhoet, R. (2014). Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models. *Water Resources Research*, 50, 409–426. <https://doi.org/10.1002/2013WR014063>
- Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and environmental systems models. *Water Resources Research*, 51, 3070–3092. <https://doi.org/10.1002/2014WR016527>
- Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, 52, 423–439. <https://doi.org/10.1002/2015WR017558>
- Reusser, D. E., Buytaert, W., & Zehe, E. (2011). Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. *Water Resources Research*, 47, W07551. <https://doi.org/10.1029/2010WR009947>
- Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, 25(12), 1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2007). *Global Sensitivity Analysis. The Primer*. John Wiley, Chichester, UK. <https://doi.org/10.1002/9780470725184>

- Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, 79, 135–152. <https://doi.org/10.1016/j.envsoft.2016.02.005>
- 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>
- Sebok, E., Refsgaard, J. C., Warmink, J. J., Stisen, S., & Jensen, K. H. (2016). Using expert elicitation to quantify catchment water balances and their uncertainties. *Water Resources Research*, 52, 5111–5131. <https://doi.org/10.1002/2015WR018461>
- Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, 38(11), 1241. <https://doi.org/10.1029/2001WR000978>
- Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. *Journal of Hydrology*, 503, 135–152. <https://doi.org/10.1016/j.jhydrol.2013.08.047>
- Shoaib, S. A., Marshall, L., & Sharma, A. (2016). A metric for attributing variability in modelled streamflows. *Journal of Hydrology*, 541, 1475–1487. <https://doi.org/10.1016/j.jhydrol.2016.08.050>
- 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>
- Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, 523, 739–757. <https://doi.org/10.1016/j.jhydrol.2015.02.013>
- 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. *Environment and Earth Science*, 76(10), 367. <https://doi.org/10.1007/s12665-017-6701-y>
- Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43, W01413. <https://doi.org/10.1029/2005WR004723>
- Uusitalo, L., Lehtikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling and Software*, 63(C), 24–31. <https://doi.org/10.1016/j.envsoft.2014.09.017>
- Verdú, J. M. (2003). *Análisis y modelización de la respuesta hidrológica y fluvial de una extensa cuenca de montaña mediterránea (río Isábena, Pre-Pirineo)*. Lleida, Catalonia, Spain: Universitat de Lleida. Retrieved from <http://www.tdx.cat/TDX-0630107-193135>
- Vigiak, O., Borselli, L., Newham, L. T. H., McInnes, J., & Roberts, A. M. (2012). Comparison of conceptual landscape metrics to define hillslope-scale sediment delivery ratio. *Geomorphology*, 138(1), 74–88. <https://doi.org/10.1016/j.geomorph.2011.08.026>
- Williams, J. R. (1995). The EPIC Model. In V. P. Singh (Ed.), *Computer models of watershed hydrology* (pp. 909–1000). Highlands Ranch, CO: Water Resources Publications.
- Wischmeier, W. H., & Smith, D. D. (1981). Predicting rainfall erosion losses. *Supersedes Agriculture Handbook*, 39(2), 285–269. <https://doi.org/10.1029/TR039i002p00285>
- Zehe, E., Graeff, T., Morgner, M., Bauer, A., & Bronstert, A. (2010). Plot and field scale soil moisture dynamics and subsurface wetness control on runoff generation in a headwater in the Ore Mountains. *Hydrology and Earth System Sciences*, 14(6), 873–889. <https://doi.org/10.5194/hess-14-873-2010>
- Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., & Jensen, K. H. (2015). Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model. *Advances in Water Resources*, 86, 400–413. <https://doi.org/10.1016/j.advwatres.2015.07.018>