Response to reviewers

Referee #1

We would like to thank the Referee for the comments, which are highly appreciated. We will try to improve on the raised issues.

Comment:
Sometimes, I find discussion of the findings contradictory and confusing. For instance, it is mentioned that the introduction of sub-grid heterogeneity leads to improvement in capturing the flow signatures related to peak flows in the low flow period. However, I find the explanation given on pages 13322 – 23 misleading. This explanation applies to only wetlands, but the authors also show a considerable improvement in performance for the urbanized catchment, Orge as well.

Reply:

To clarify on this issue, the urbanized Orge catchment, which showed a relatively poor performance in terms of objective functions, did indeed also show improvements for the low flow signatures. Even though partly urbanized, another large part of the catchment is still classified as wetland (see also Figure 1), where the upward seepage of water may be important. Thus, as we also state in lines 5-9 of page 13322, the more general objective functions over the full validation period showed a decrease in performance, as the relatively high fast flows are not represented sufficiently, but these flows dominate these metrics. The low flow signatures improved at the same time, due to the incorporation of sub-grid variability. In our view, this confirms how misleading objective functions calculated for the full validation period can be, as the peaks in the low flow period are better captured by the topography driven model. Nevertheless, this can go unnoticed when calibrating on more general objective functions.

Comment :
I miss a proper interpretation for the lack of improvement to the models ability to capture the autocorrelations of the flows when the proposed structural changes were introduced. Why does a simpler model respond faster and why should a model that responds faster lead to a better representation of the autocorrelations (page 13323, lines 7-10)?

Reply:

We acknowledge that our discussion here may need further elaboration and, thus, this will be adjusted in the revised manuscript. Basically, the original model showed a quicker response, whereas the adjusted model delayed the signal more. Our explanation for this observation is that the adjusted model has more options, in terms of reservoirs, to store the water. Effectively, this could delay the signal. In case of fast reacting catchments like the Orge or Loisach, this means that the 1-day autocorrelation, which tells us something about the timing of the peaks, is poorly represented.

Besides this model based explanation, the used data may have an influence on the representation of the autocorrelation. The E-OBS data is of rather low resolution (24 km by 24 km), what could lead to a low estimate of the precipitation in locations with a steep topography. When rainfall peaks have averaged out, discharge peaks may not be well represented either, again leading to a low performance for the autocorrelation. Nevertheless, this applies to each of the models.

Comment:

I find the discussion on page 13325, lines 17-25 interesting. Why was it necessary to impose the constraints in Equations 4 and 5 in the first place?

Reply:

In our first hypothesis, more constraints should define the 'plausible' parameter space more, leading to more pronounced differences. The chosen constraints were relatively simple and easy to implement, and, at the same time, there was enough support from literature to confidently apply these constraints. Nevertheless, this appeared not to be true in all cases.

Comment:

Page 13328, last paragraph: Does the considerable improvement in model transferability due to introduction of constraints apply to all parts of the flow regime or only to low flow signatures?

Reply:

The introduction of constraints led to a more general improvement in transferability. For example, in Figure 14d it can be seen that the transferability of mHMtopo with constraints compared to mHMtopo without constraints improves over the full range of signatures. Also Figure 14c shows this for mHM, even though the differences are smaller. We will clarify this in a revised version.

Comment:
Page 13329, lines 14-16, Why is it difficult transferring parameters to this particular catchment from other catchments?

Reply:

Apparently, the derived global relations, like the example in Figure 4, do not hold for this catchment. This could mean that this catchment is significantly different from all other four catchments used in calibration. This may well be, as this catchment is merely gently sloped with agriculture. The Loisach and Broye are more mountainous, whereas the Treene is very flat and wetland dominated. In nature, the Orge catchment should be relatively similar, but this catchment is strongly affected by urbanization. The use of more similar catchments as the Briance in calibration, could maybe improve the transferability to this catchment.

Referee #2 Prof. Tian

We would like to thank prof. Tian for his valuable comments. We highly appreciate the suggestions and would like to improve the issues raised. Herewith, we would like to respond to the comments.

Comment:
I noticed that the improvement of mHM with semi-quantitative constraints improves a little bit when incorporating the additional sub-grid heterogeneity. The authors are often anticipated to improve the performance by comparing to the known 'best' one (not the worst one). More explanations and discussions are preferable here.

Reply:

We are not entirely sure what you mean, but we believe you refer to Figure 10b, which is indeed only discussed briefly at page 13325, lines 1-4. We will add a more elaborate discussion on the improvements of mHMtopo-constrained compared to mHM-constrained in an extra paragraph in section 3.2.3.

Comment:
As the authors test two things, sub-grid heterogeneity representation and expert knowledge based calibration methodology, in this paper, the title of the paper should also reflect the two things.

Reply:

Thank you for this suggestion, we fully agree with it and will change the title: ”The importance of topography controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models”

Comment:

P13303, the last sentence: on the catchment scale â˘A˘T> at the catchment scale.

Reply:

We will correct this.

Comment:

P13305 L7-10: the authors state that the distribution function for maximum unsaturated storage capacities are originally defined in the VIC-model. In my mind this is not true. I suggest the authors refer to Xinanjiang model developed by Zhao (1992, on JoH, 135: 371-381).

Reply:

We agree with this and will add the reference to the Xinanjiang model.

Comment:

The same place with 4): For the Representative Elementary Watershed approach and its closure problem, there is quite a few new publications after Reggiani et al. (1998). I suggest the authors to cite the news as well to reflect the recent advance.

Reply:

We will adjust this and cite the new work as well.

List of changes

-change in title

-additional references to more recent work considering the REW-approach (pages 5-6)

-additional explanation about wetlands with regard to low flow peaks in the largely urbanized Orge catchment. (page 20)

-additional discussion about the representation of the autocorrelation (page 21)

-additional sentence in the discussion of constraints in Eqs. 4 and 5. (page 23)

-addition of discussion of Fig.10b in paragraph 3.2.3. (page 24)

-additional discussion that additional constraints lead to more general improvements in transferability. (page 27)

-additional discussion about why the transferability to the Briance is difficult. (page 27)

-replacement of Figure 10, as 10e was also shown as 10b. The text about Fig. 10b, even though not much, was however correct. (page 57)

The importance of topography controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models

R.C.Nijzink[1], L. Samaniego[2], J.Mai[2], R. Kumar[2], S. Thober[2], M. Zink[2], D. Schäfer[2], H.H.G.Savenije[1], M. Hrachowitz[1]

[1] Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

[2] UFZ - Helmholtz Centre for Environmental Research, Permoserstraße 15, 04318 Leipzig, Germany

Correspondence to: R.C.Nijzink (R.C.Nijzink@tudelft.nl)

Abstract

Heterogeneity of landscape features like terrain, soil, and vegetation properties affect the partitioning of water and energy. However, it remains unclear to which extent an explicit representation of this heterogeneity at the sub-grid scale of distributed hydrological models can improve the hydrological consistency and the robustness of such models. In this study, hydrological process complexity arising from sub-grid topography heterogeneity was incorporated in the distributed mesoscale Hydrologic Model (mHM). Seven study catchments across Europe were used to test whether (1) the incorporation of additional sub-grid variability on the basis of landscape-derived response units improves model internal dynamics; (2) the application of semi-quantitative, expert-knowledge based model constraints reduces model uncertainty; and (3) the combined use of sub-grid response units and model constraints improves the spatial transferability of the model.

Unconstrained and constrained versions of both, the original mHM and mHMtopo, which allows for topography-based sub-grid heterogeneity, were calibrated for each catchment individually following a multi-objective calibration strategy. In addition, four of the study catchments were simultaneously calibrated and their feasible parameter sets were transferred to the remaining three receiver catchments. In a post-calibration evaluation procedure the probabilities of model and transferability improvement, when accounting for sub-grid variability and/or applying expert-knowledge based model constraints, were assessed on the basis of a set of hydrological signatures. In terms of the Euclidian distance to the optimal model, used as overall measure for model performance with respect to the individual signatures, the model improvement achieved by introducing sub-grid heterogeneity to mHM in mHMtopo was on average 13%. The addition of semi-quantitative constraints to mHM and mHMtopo resulted in improvements of 13% and 19% respectively, compared to the base case of the unconstrained mHM. Most significant improvements in signature representations were, in particular, achieved for low flow statistics. The application of prior semi-quantitative constraints further improved the partitioning between runoff and evaporative fluxes. Besides, it was shown that suitable semi-quantitative prior constraints in combination with the transfer function based regularization approach of mHM, can be beneficial for spatial model transferability as the Euclidian distances for the signatures improved on average by 2%. The effect of semi-quantitative prior constraints combined with topography-guided sub-grid

heterogeneity on transferability showed a more variable picture of improvements and deteriorations, but most improvements were observed for low flow statistics.

1 Introduction

A better understanding of the link between landscape heterogeneity and its impact on process dynamics of catchments is urgently required to develop more robust catchment-scale rainfall-runoff models that have the skill to adequately reproduce the observed system response dynamics, even for catchments where no calibration data are available. Besides heterogeneity in the system boundary conditions, including amongst others topography, vegetation or geology (e.g. Knudsen,1986; Rodrìguez-Iturbe, 2006; Tromp-van Meerveld, 2006), climatic variables, i.e. the forcing of models such as precipitation and evaporation, typically exhibit considerable spatial variability (e.g. Obled, 1994; Singh, 1997; Winsemius et al., 2008; Hrachowitz and Weiler, 2011). Together, these factors lead to the concept of the “uniqueness of place” as termed by Beven (2000). Thus, with increasing catchment size it becomes increasingly problematic to treat catchments as lumped entities in models as these are not suitable to accommodate spatial heterogeneity. In other words, this heterogeneity can in reality result in a variety of parallel processes, characterized by considerably different time scales, being simultaneously active. Therefore, lumped representations of catchments frequently fail to adequately represent the dominant features of the observed hydrological response at~~on~~ the catchment scale (e.g. Euser et al., 2015), such as low and high flows at the basin outlet

Experimentally, the importance of intra-catchment process heterogeneity was for example demonstrated by Seibert et al. (2003a). They showed that groundwater table fluctuations can exhibit considerably distinct dynamics between hillslopes and riparian areas near the stream. Similarly, Detty and McGuire (2010) showed that topographically different landscape elements are characterized by different wetting mechanisms, while others, e.g. McGlynn et al. (2004), Jencso et al. (2009) or Spence et al. (2010), systematically documented distinct response patterns in different parts of catchments.

Lumped applications of hydrological models, such as HBV (Bergström, 1992) or GR4J (Perrin et al., 2003) proved valuable in the past under a wide range of environmental conditions and across a range of scales as they appear to capture the core emergent processes of many hydrological systems (e.g. Refsgaard and Knudsen, 1996; Booij, 2005). Nevertheless, in many cases these models may remain serious over-simplifications of the different combinations of the dominant processes underlying the observed response patterns as argued by, among others, Young, (1992), Reichert and Omlin (1997), Perrin et al. (2001), Wagener and Gupta, (2005), Gupta et al. (2012), Zehe et al.,2014, Hrachowitz et al. (2014) and Fovet et al. (2015). In addition, the transferability of these simple models to other (ungauged) basins is limited. In the past, distributed models, such as MIKE-SHE (Refsgaard and Storm, 1995) or DHSVM (Wigmosta et al., 1994), but also (semi-) distributed applications of lumped models were shown to alleviate the issue of over-simplification to a certain extend by accommodating spatial heterogeneity in soil moisture and/or model parameters (e.g. Fenicia et al., 2008; Winsemius et al., 2008; Euser et al., 2015).

However, traditional, conceptual distributed model approaches suffer from several limitations. They are defined by the grid size of the available data or the size of the defined subcatchments, which are in the order of several dozen square kilometers in most applications (e.g. Booij, 2005; Lindström, 2010). Furthermore, although different model parameters allow for some flexibility in accounting for spatial differences, in a large number of cases the defined processes remain the same among individual model units, i.e. the same model architecture is used. This denies the potential for the distinction of different dominant processes belonging to the different parts of the study domain. Even though in some cases triggered by different parameterizations, the importance of this distinction of processes became already apparent in several studies, e.g. Merz and Bárdossy (1998), Zehe et al. (2001), Seibert et al. (2003a) , Das et al (2008).

Thus, as individual model units are often still represented in a lumped way, sub-grid process heterogeneity in these lumped units is merely just reflected by distribution functions, or constitutive relationships. For example, distribution functions for maximum unsaturated storage capacities, such as ~~originally~~ defined in the Xinanjiang model (Zhao, 1992) or the VIC model (Liang et al, 1994), are widely used as a measure of spatial variability of storage capacities on the sub-grid scale. As a second example, the closure problem in the Representative Elementary Watershed approach (Reggiani et al. 1998) addresses the definition of relationships between the spatial variability on the elementary watershed scale and states and fluxes to close the mass and momentum balance equations. Several attempts have been reported to formulate closure relations that allow to accommodate the spatial heterogeneity within the elementary watershed to varying degrees (e.g. Reggiani and Rientjes, 2005; Zhang and Savenije, 2005; Zhang et al.,2006; Tian et al., 2006; Mou et al., 2008; Vannametee et al., 2012), but the search for generally applicable adequate closure relations is still ongoing. ~~Nevertheless, a more explicit definition of processes occurring at different locations within the model unit is not very common.~~

The division of the catchment in several functional units (e.g. Knudsen, 1986; Leavesley and Stannard, 1990; Kite and Kouwen, 1992; Kouwen et al., 1993; Flügel, 1995, Reggiani et al., 1998; Winter, 2001; Seibert et al., 2003b; Uhlenbrook et al., 2004, Schmocker-Fackel et al., 2007, Zehe et al.,2014) may offer a way to address these conceptual short-comings. In spite of the fact that in many cases insufficient data for a detailed delineation of response units are available, it has been recognized (e.g. Beven and Binley, 1979; Knudsen, 1986) that already topographic data can contain important hydrological information. Starting from that premise, Savenije (2010) argued that through the co-evolution of topography, vegetation and hydrology, different landscape features, such as hillslopes, wetlands or plateaus, do have distinct hydrological functions. This implies that topography alone may contain sufficient information to derive dominant hydrological response units. Distinct response units can therefore be identified based on, for example, the height above the nearest drainage, as a proxy for hydraulic head, and local slope (Rennó et al., 2008; Nobre et al, 2011; Gharari et al., 2011). The different dominant processes characterizing these response units can then be combined into a semi-distributed model with landscape elements acting in parallel. This parsimonious approach to account for process heterogeneity at catchment scale proved highly valuable for improving the skill of otherwise lumped models in reproducing observed system response patterns (e.g. Gao et al., 2014; Gharari et al., 2014). They further enhance model transferability without the need for empirical transfer functions (Gao et al., 2015) in widely contrasting environments.

Traditional distributed model applications are characterized by a comparably large parameter space. The typical lack of sufficient model constraints makes it problematic to select meaningful feasible parameter sets. This leads to considerable equifinality (Beven, 1993) and associated problems (cf. Gupta et al., 2008). The need for increased hydrological consistency in models and for more realistic internal model dynamics (i.e. “getting the right answer for the right reasons”; Kirchner, 2006) was recently emphasized as a critical point towards the development of models with higher predictive power (Gupta et al., 2012; Euser et al., 2013; Hrachowitz et al., 2014). This can all be placed in the sense of achieving “the least uncertainty for forecasts” (Kumar, 2011) and needs to be done by more rigorous model testing (e.g.

Andréassian et al., 2009; Coron et al., 2012) to meaningfully constrain the feasible model/parameter space.

An efficient method to constrain the parameter space is model regularization (e.g. Tonkin and Doherty, 2005), for example by the use of transfer functions (e.g. Abdulla and Lettenmaier, 1997; Hundecha et al., 2004; Pokhrel et al., 2008). Being mathematically equivalent to the concept of regionalization, it was also shown that this is a valuable method to improve spatial model transferability (e.g. Götzinger and Bárdossy, 2007; Samaniego et al., 2010; Kumar et al. 2013b). However, regularization frequently relies on empirical relationships between catchment characteristics, such as soils, and individual model parameters with little explicit hydrological meaning. In a different approach it was recently shown that semi-quantitative information on catchment functioning based on expert-knowledge, often referred to as "soft data" (Seibert and McDonnell, 2002; Van Emmerik et al., 2015), can be highly efficient in constraining models (Kapangaziwiri, 2012; Hughes, 2013; Seibert and McDonnell, 2013; Gao et al., 2014; Gharari et al., 2014~~a~~; Hrachowitz et al., 2014).

Considering the potential information embedded in landscapes, the need for simplification and regularization in complex models, and the additional value of expert-based semi-quantitative information, there may be an opportunity to improve distributed hydrological models. To test the value of topography-induced sub-grid process heterogeneity, the principles of landscape driven modelling (Savenije, 2010) were introduced in the distributed, regularized mesoscale Hydrologic Model (mHM; Samaniego et al., 2010; Kumar et al., 2013a). It is hypothesized that, (1) the incorporation of additional sub-grid variability on the basis of topography-derived response units improves model internal dynamics and its predictive power, (2) the application of semi-quantitative, expert-knowledge based model constraints allow the identification of unfeasible parameter sets and thereby reduces model uncertainty, and (3) the combined use of response units and model constraints improves the spatial transferability of the model.

2 Methodology

2.1 Study Areas

Seven catchments were selected in order to cover a variety of climatological, geographical and geological conditions. The geographical locations as well as the classification of topography-based hydrological response units (i.e. hillslopes, wetlands and plateaus) in the study catchments are shown in Figure 1. The set of study sites includes catchments with pronounced relief as well as relatively flat and gently sloped catchments. Therefore, some catchments are almost fully dominated by landscapes classified as hillslopes, whereas others contain higher proportions of wetlands. In addition, the climatic variability is considerable, as indicated by the aridity indices ranging from 0.5 to 1.34. Table 1 summarizes the catchment characteristics.

The North-German Treene catchment is a tributary of the Eider river. It is a lowland catchment characterized by sedimentary soils and peat. The land cover is mostly grassland and low vegetation while only a small percentage is forested or agriculturally used.

The Loisach, Kinzig and Broye catchments are located in mountainous areas, characterized by pronounced relief, steep slopes and the importance of snow. The Loisach and Kinzig catchments are mostly forested, whereas the Broye catchment has mainly open grassland. Sand overlies limestone and other sedimentary bedrock in the Loisach catchment, while the Kinzig catchment is dominated by granite and gneiss series.

The French catchments Orge and Briance are relatively flat with gentle slopes and flat upland areas. Agriculture is the dominant land use, but some forests are also present. The Orge catchment is a tributary of the Seine and contains some of the suburbs of Paris. Thus, it has a significant proportion of urbanized areas (10%). In the Orge, sandy loam soils have formed on limestone geology, while the Briance is characterized by gravel on gneiss bedrock.

The Alzette catchment in Luxembourg is partly covered by forest (33% of the catchment area). The rest of the catchment is more open with grass and shrublands. Limestone, sandstone and schist are the dominant geologic formations with some clay and loam soil in the upper layers.

Daily discharge time series for all study catchments were obtained from the Global Runoff Data Centre (GRDC). The daily meteorological data are the gridded E-OBS precipitation and temperature data from the European Climate Assessment and Dataset (ECA&D). The daily potential evaporation was estimated with the Hargreaves equation (Hargreaves, 1985). A summary of the data sources is given in Table 2.

2.2 Models

2.2.1 Mesoscale Hydrological Model (mHM)

mHM is a distributed, process-based model that uses the cell-wise model architecture shown in Figure 2 in each grid cell of the modelling domain (Samaniego et al., 2010, Kumar et al. 2013a). It contains an interception and snow routine to determine the effective precipitation which enters the soil moisture reservoir. For sealed areas the water is directly routed to a fast reservoir. The water infiltrating into the soil is then partitioned into transpiration and percolation to a fast runoff reservoir, i.e. shallow subsurface flow. In addition, this reservoir recharges a lower reservoir that mimics the base flow component of the runoff. The model has been successfully applied across Germany, Europe and North America (Samaniego et al. 2010, 2013, Kumar et al. 2010, 2013a,b, Livneh et al., 2015, Thober et al., 2015, Rakovec et al., 2015).

2.2.2 Topography driven mHM (mHMtopo)

To test the value of topography variability-induced process heterogeneity in a distributed model, the concepts of FLEXtopo (Savenije, 2010; Gharari et al., 2011) were applied in mHM. Based on the assumption of distinct hydrological functioning of different landscape elements, sub-grid process heterogeneity was accounted for by a model architecture that allowed an explicit representation of landscape classes identified to be dominant in many central European regions: plateaus, hillslopes and wetlands (Savenije, 2010). The landscape classes were defined by the Height Above the Nearest Drainage (Renno et al., 2008; HAND) and local slope. Following Gharari et al. (2011), areas with a low slope (<11%) and high HAND (>5m) were defined as plateaus, areas with high slope (>11%) as hillslopes and areas with low slope and low HAND (<5m) as wetlands. It is acknowledged that these thresholds remain merely assumptions and may need refinement in other regions. Nevertheless, this refinement is out of the scope of this paper and the used threshold values are assumed to give a reasonable delineation of landscape units in the Central European context. The varying proportions of these individual landscape units in each cell in the modelling domain then allow for considerable sub-grid process heterogeneity in the distributed model, as the total outflow of a cell is then the area-weighted average of the outflows from the individual landscape units. The assumptions behind the conceptualizations of the three landscape classes are in the following briefly summarized. For details the reader is referred to Savenije (2010) and Gharari et al. (2014).

The different model structures for these three classes run in parallel, connected by a common groundwater reservoir for each modelling cell, as can be seen in Figure 3. The primary hydrological functions of plateau landscapes are, in the absence of significant topographic gradients, mainly groundwater recharge and evaporation/transpiration, i.e. vertical fluxes. To account for potential agricultural drainage systems a fast reservoir is included in the plateau model structure. Hillslopes are assumed to be the dominant source of storm flow and efficiently contribute to storm runoff through storage excess shallow subsurface flow, e.g. preferential flow, here conceptualized by a fast reservoir. The wetland landscape is assumed to interact stronger with the groundwater. Thus, capillary rise (Cr in Figure 3) is included to interact with the soil moisture reservoir. The wetlands are assumed to have shallow groundwater tables and associated low storage capacities. Therefore, saturation excess overland flow, represented by a fast responding reservoir, and evaporative processes are assumed to be dominant in this landscape unit.

Throughout the rest of this paper, the two models will be referred to as mHM and mHMtopo to distinguish between the original mHM and the topography-guided set-up respectively.

2.3 Model regionalization, regularization and prior constraints

Reducing the feasible model parameter space is strongly associated with a reduction in parameter equifinality and model uncertainty, and can be achieved by imposing constraints on the model, for example by regularization. Only parameter sets that can satisfy these constraints will then be retained as feasible, while others will be discarded. A method that uses empirical transfer functions relating parameter values to physical catchment characteristics, is also a powerful tool to regionalize models.

2.3.1 Multiscale parameter regionalization

The multiscale parameter regionalization (MPR) is the key feature of mHM (Samaniego et al., 2010; Kumar et al. 2013a). The global parameters in mHM are, in contrast to typical models, not hydrologic model parameters (e.g. soil porosity). Instead, the global parameters define the functional relationship between the individual hydrologic model parameters and physical catchment characteristics at the spatial resolution of the data of the latter. A set of global parameters is obtained by simultaneously calibrating on multiple catchments. This set of global parameters can then be transferred to other catchments where the same data of physical catchment characteristics are available without the need for further calibration.

Thus, the functional relationships are used in a first step to estimate model parameters on the spatial resolution of the input data. As depicted in Figure 4, as an example, the leaf area index is linearly linked through global parameters with the hydrologic model parameter of interception capacity (I_{max}). Assuming the relationships are adequate, the use of additional data of preferably multiple, distinct catchments may increase the general validity of these relationships and, thus, the global parameters.

Figure 5 depicts the application of the MPR technique on gridded data. The obtained hydrologic parameters, determined by the functional relationships, still have a resolution equal to the input data. In most cases, this is not equal to the modelling resolution. Therefore, a second step in the MPR is the upscaling of hydrologic parameters to the modelling resolution (in this study 8 km x 8 km). This upscaling can either be achieved by using the harmonic mean, arithmetic mean or maximum value over the cells within the modelled grid cell. The choice of the upscaling method strongly depends on the parameter under consideration. The reader is referred to Samaniego et al. (2010) and Kumar et al. (2013a) for details about the transfer functions and upscaling methods.

The MPR has been adjusted in two ways for use in mHMtopo. The regionalization functions were used for the three individual landscape units, whereby each landscape unit was assigned its own global parameters. In other words, the functional relations between physical catchment characteristics (e.g. soil, slope) and hydrologic parameters were kept the same, but the global parameters of these relations differ between landscape units. For example, the LAI is now individually linked with three global parameters for wetland, hillslopes and plateaus, respectively, to obtain three hydrologic parameters for interception capacity ($I_{max,plateau}$, $I_{max,hillslope}$, $I_{max,wetland}$), see Figure 5.

The second change was in the upscaling. Instead of scaling up over all high resolution cells within a modelling unit, the upscaling was carried out for each landscape class within a modelling unit. The upscale operators for mHMtopo were adopted from similar parameters in mHM. For example, the upscaling of the interception capacities was done by the arithmetic mean, similar to that of the upscaling of interception capacities used in the original mHM (see Figure 5).

2.3.2 Expert knowledge-based prior constraints

In addition to MPR, we tested the value of semi-quantitative, relational prior parameter and process constraints (Gharari et al., 2014; Hrachowitz et al., 2014) for the robustness of process representation and model transferability. In other words, only global parameter sets that satisfied these parameter and process constraints during calibration were accepted as feasible and used in validation and post-calibration evaluation.

Specifically, constraints for the long-term mean annual runoff coefficients were formulated to ensure plausible water partitioning between evaporation and runoff. The limits were chosen as the maximum and minimum annual runoff coefficients C_{Rmax} and C_{Rmin} occurring over the calibration time period. The months May – September were defined as high flow period, whereas low flows were assumed to occur over the months October – April. Only for the Loisach catchment these periods were switched as this catchment has high flows starting in spring due to snowmelt. The following three constraints were used: one taking into account the whole time series (C_R) as well as one for the high flow period (C_{Rhigh}) and one for the low flow period (C_{Rlow}) to improve the seasonal variation of the model response behaviors.

$$C_{Rmin} < C_{Rmodelled} < C_{Rmax} \qquad (1)$$

$$C_{Rhigh,min} < C_{Rhigh,modelled} < C_{Rhigh,max} \qquad (2)$$

$$C_{Rlow,min} < C_{Rlow,modelled} < C_{Rlow,max} \qquad (3)$$

The topography driven model mHMtopo is also constrained on soil moisture storage capacity (S_M). On hillslopes and plateaus the groundwater table can be assumed to be deeper than in wetlands and root systems generate a larger dynamic part of the unsaturated zone (cf. Gao et al., 2014b). Therefore, they are conceptualized to have a higher water storage capacity than wetlands, which are typically characterized by a very shallow groundwater table. This reasoning reflects not only the variable contribution area theory of Dunne (1975) and the concept of topographic wetness index (Beven, 1979), but also results from experimental studies, e.g. Seibert et al. (2003a). Thus, two additional constraints were used for mHMtopo:

$$S_{M,plateau} > S_{M,wetland} \qquad (4)$$

$S_{M,hillslope} > S_{M,wetland}$ (5)

2.4 Experiment set-up

2.4.1 Calibrated model comparison

The two models, i.e. mHM and mHMtopo, were calibrated for each catchment with a random Monte Carlo sampling approach based on 100,000 realizations and a multi-objective strategy using four objective functions: the Nash-Sutcliffe efficiency of flow ($E_{NS,Q}$), the Nash-Sutcliffe efficiency of the logarithm of flow ($E_{NS,logQ}$), the volume error of flow ($E_{V,Q}$) and the Nash-Sutcliffe efficiency of the logarithm of the flow duration curve ($E_{NS,FDC}$). The four objective functions were chosen as they characterize different aspects of the flow response. Therefore, these objective functions are expected to provide hydrologically relatively consistent and robust parameter sets.

This calibration strategy was preferred over other calibration schemes, such as DDS (Tolson and Shoemaker, 2007) or SCE (Duan et al. 1992), to obtain a set of feasible parameter solutions, instead of one optimal solution. As the mathematically optimal solution may not be the hydrologically most adequate solution (cf. Beven, 2006; Kirchner, 2006; Andreassian et al., 2012), this is necessary to make a robust assessment of the model's abilities. Therefore, all parameter sets that satisfy all model constraints and that are contained in the parameter space spanned by the 4-dimensional Pareto front formed by $E_{NS,Q}$, $E_{NS,logQ}$, $E_{V,Q}$ and $E_{NS,FDC}$ were considered as feasible solutions and used for post-calibration evaluation. Considering all feasible solutions as equally likely, the model uncertainty intervals are represented by the envelope of all feasible solutions.

2.4.2 Post-calibration model evaluation

The models' skill to reproduce a variety of observed hydrological signatures, i.e. emergent properties of a system (Eder et al., 2003), was evaluated after calibration to test the hydrological consistency of the models. Hydrological signatures allow evaluating the consistency and reliability of hydrologic simulations by taking more features of the hydrological response into account than only the flow time-series. In a nutshell, the more signatures a model can simultaneously reproduce in addition to the hydrograph, the more plausible it is that a model (and its parameters) adequately reflects the underlying dominant system processes (e.g. Euser et al., 2013). All signatures used in this study were selected based on earlier work (e.g. Sawicz et al., 2011; Euser et al., 2013) and are summarized in Table 3.

Although not fully independent of each other, the signatures, such as the peak flow distribution, the rising limb density and the autocorrelation function of flow, contain information on different aspects of the hydrologic response. The Nash-Sutcliffe efficiency S_{NS} was used as a performance metric to assess the model skill in case of multi-value signatures such as the peak flow distribution or the autocorrelation function. In contrast, the relative error S_{RE} was used for single valued signatures, such as the mean annual runoff. The Euclidian distance D_E to the "perfect model" was used as an overall measure of a model's ability to reproduce all signatures under consideration (e.g. Schoups et al., 2005):

$$D_E = \sqrt{\left(1 - S_{NS,1}\right)^2 + \left(1 - S_{NS,2}\right)^2 \dots + \left(1 - S_{NS,n}\right)^2 + S_{RE,1}{}^2 + S_{RE,2}{}^2 \dots + S_{RE,m}{}^2}, \quad (5)$$

with $S_{NS,i}$ the performance metric of n multi-valued signatures, and $S_{RE,j}$ for the m single valued signatures.

From calibration, a set of feasible parameter sets was obtained for each tested model, which inevitably resulted in varying skills to reproduce the system signatures for the individual parameter sets. The probability that one model outperforms another for a specific signature was computed to objectively quantify the differences between these distributions and to allow an overall assessment which of the tested models exhibit a higher capability to reproduce the individual signatures. As estimates of the empirical performance distributions are available based on all parameter sets retained as feasible, the probability of improvement $P_{I,S}$ can be readily obtained from:

$$P_{I,S} = P(S_1 > S_2) = \sum_{i=1}^{n} P(S_1 > S_2 \mid S_1 = r_i) P(S_1 = r_i) \quad (6)$$

where S_1 and S_2 are the signature performance metrics of the two models, r_i a realization from the S_1 distribution and n the total number of realizations of the S_1 distribution. Thus, a probability of 0.5 indicates that in 50% of the cases model 1 and in 50% of the cases model 2 performs better, i.e. no preference for a model can be identified. In contrast, for $P_{I,S} > 0.5$ it is more likely that model 1 outperforms model 2 with respect to the signature under consideration, and vice versa for $P_{I,S} < 0.5$.

In an additional analysis, the Ranked Probability Score S_{RP} was calculated as a measure for the magnitude of improvement. For details please see the description and Figure S1 in the supplementary material.

2.4.3 Comparison of model transferability

The hydrologic model mHM has been previously shown to have a considerable ability to reproduce the hydrograph when transferring global parameters from calibration catchments to other regions without further recalibration (Samaniego et al., 2010a,b; Kumar et al., 2013a,b; Rakovec et al. 2015). Therefore, it was tested whether the addition of topography-driven sub-grid process heterogeneity and the use of prior constraints in mHM have potential to further improve this transferability. Four catchments were used as donor catchments to obtain one set of global parameters via simultaneous calibration. The Orge, Treene, Broye and Loisach were chosen as donor catchments as they are geographically far from each other, introducing a wide range in climate and catchment characteristics. The receiver catchments are the three remaining catchments of Alzette, Briance and Kinzig.

This was carried out with the same calibration strategy as for the individual catchment calibrations. However, the four objective functions $E_{NS,Q}$, $E_{NS,logQ}$, $E_{V,Q}$ and $E_{NS,FDC}$ were now averaged over the catchments. This lead to global parameters that account for the performance on all donor catchments. These averaged values were then used to determine the pareto space of feasible parameter sets again. The feasible solutions were transferred and used in the three remaining receiver catchments without any further recalibration. We fully acknowledge that this analysis can only give a sense of what is possible and that a full bootstrap procedure and the analysis of more catchments would have allowed a more robust interpretation of the results, but this was unfeasible given the computational demands of the calibration procedure. The calibrations were carried out on the high-performance compute cluster EVE of the UFZ Leipzig which has 84 compute nodes with dual socket Intel Xeon X5650 processors with 64 GB RAM as well as 65 compute nodes with dual socket Intel Xeon E5-2670. Nevertheless, the used calibration strategy needed runtimes of about two weeks per catchment on multiple EVE cores, depending on catchment sizes and lengths of time series.

3 Results and Discussion

3.1 Calibrated model comparison

The two different models mHM and mHMtopo, both with and without additional prior constraints, exhibited adequate and similar calibration performances with respect to all four calibration objective functions (see Figure S2 in the supplementary material). For the validation period it was found that performance generally improved by applying prior constraints and by allowing for topography-guided sub-grid process heterogeneity. This can be seen from Figure 6, where mHM with constraints (darkblue) compared with mHM (lightblue) generally has an increased performance. The same is true for mHMtopo with constraints (orange) compared with unconstrained mHMtopo (grey). At the same time, it can be noted from Figure 6 that the addition of topography-guided sub-grid variability leads to a general moderate improvement in performance. Overall, the introduction of constraints to mHM resulted in an average improvement of 13% with regard to the Euclidian distance D_E for the objective function values in validation. In addition, unconstrained and the constrained mHMtopo exhibited an average increase of 8% and 11%, respectively, for the Euclidian distance D_E compared to the original mHM.

3.1.1 Effect of sub-grid heterogeneity

The incorporation of sub-grid process heterogeneity did not show a clear pattern of improvements or deterioration. Some catchments experienced performance increases in terms of the used objective functions during validation, like the Briance catchment. The predictive performance of others, also in terms of the used objective functions, slightly decreased, such as the Orge catchment. These findings support the results of Orth et al. (2015), who also found that added complexity, here in the sense of an increased number of processes and parameters, not necessarily leads to model improvements. However, these findings are not in line with some other previous work (e.g. Gharari et al., 2013; Gao et al., 2014; Euser et al., 2015), who all concluded that parallel model structures increased model performance. It can be argued that for mHM, whose global parameters are to a certain extent already functions of landscape variability, additional sub-grid process heterogeneity is not warranted by the available data and can thus not be resolved by the model when there are relatively little contrasts in the landscape.

The Treene catchment benefits most from the addition of topography-guided sub-grid heterogeneity (Figure 6). Here, a large area is classified as wetland, where the soil moisture is fed by groundwater through capillary rise. This process is fully absent in the original mHM structure, but an important process in this relatively flat and humid catchment, dominated by peaty soils. These findings also correspond with conclusions by Schmalz (2008a,b), who applied the SWAT model in the same catchment and noticed that shallow groundwater and soil moisture parameters are very sensitive to low flows. It may also be noted that for mHMtopo the bandwidth of the feasible solutions around the observed hydrograph is considerably reduced as compared to mHM, in particular during low flows. Figure 7 shows that in the months April – July the uncertainty range is significantly larger for mHM than for mHMtopo. In addition, it is interesting to note that the lower bound of flow in mHM is reaching towards 0 mm/d in July, whereas mHMtopo still maintains a flow.

In contrast, it can be noticed from Figure 6 that the consideration of sub-grid process heterogeneity causes a decrease in performance compared to the original mHM in the Orge catchment. This catchment has a relatively large urban area of about 10%. In addition, these areas are rather densely populated and the river contains several human made adjustments such as weirs (Le Pape, 2012). Therefore, it is more markedly influenced by anthropogenic disturbances, which are likely not adequately reflected in neither mHM nor mHMtopo. This results in a situation where the more parsimonious mHM is likely to provide a representation of process dynamics that more closely reflects the observed. The higher number of parameters in mHMtopo provides not only more freedom for adequate system representations, but also for misrepresentations. Thus, after an adequate calibration a larger part of the “feasible” mHMtopo parameter sets fails to mimic the observed response patterns in the validation period compared to mHM. In addition, it can also be observed from the hydrographs that the Orge is a fast responding catchment with very spiky flow peaks (Figure 8). The addition of more storage reservoirs in mHMtopo delays the signal more than the simpler model structure, leading to a reduced ability to reproduce this spiky behavior.

3.1.2 Effect of constraints

The applied prior process and parameter constraints, in agreement with Gharari et al. (2014b) and Hrachowitz et al. (2014), helped to increase model performance (Figure 6) and to reduce model uncertainty (Figures 7,8,9) by identifying and discarding a considerable part of model solutions that did not satisfy these constraints. Rather, these discarded solutions violated observed partitioning patterns between runoff and evaporative fluxes and conflicted with our understanding of how the catchments respond. Being merely manifestations of a successful mathematical optimization process, rather than plausible representations of system-internal response dynamics, the discarded solutions underline how deceitful adequate calibration results can be and how a successful identification can result in reduced predictive uncertainty. It must be noted that the effect is strong in the chosen calibration strategy, as a large set containing also less optimal solutions is maintained as feasible, but it has already been shown that other calibration procedures may also benefit from additional constraints (Gharari et al., 2014b). This is true as constraints limit the parameter search space with feasible solutions that the algorithm has to explore. In addition, while traditional calibration procedures may converge to a mathematically optimal fit, additional constraints can test the found solutions for hydrological consistency.

More specifically, the Loisach catchment benefits considerably from the applied constraints. This can be explained by the fact that this is one of the few catchments in this study where snowmelt plays an important role. For this catchment, temperature is in phase with the high flows, which causes difficulties in water partitioning in the unconstrained models, resulting in evaporative fluxes being too high and stream flow being too low. A similar observation for the Loisach was found by Muerth et al. (2013). Even though forced by an ensemble of climate models, the winter flows were too high for an ensemble of hydrological models run for this catchment. Hence, the application of runoff constraints for high and low flow periods lead to a considerable improvement of the model internal dynamics. This is supported by visual inspection of the hydrographs (Figure 9): both, the constraints for mHM and mHMtopo, cause a significant reduction in the uncertainty bandwidth of the modelled hydrograph, particularly during high flow periods. The unconstrained models have a relatively low lower boundary during high flows, whereas the boundaries in the constrained cases stay much closer to the observed values. Nevertheless, it must also be noted that both models tend to slightly underestimate the flows in the high flow period.

3.1.3 Effect of constraints and sub-grid heterogeneity

Comparing the base case of the unconstrained mHM with the most complex constrained mHMtopo (Figure 6) shows that in most cases improvements are observed. As stated before, compared with the unconstrained mHM, the constrained mHMtopo exhibited an average increase of 8% and 11%, respectively, for the Euclidian distance $D_{E.}$ In most cases, a narrowing of the distribution of objective function values can be observed. For example, the Alzette shows a considerable reduction in the bandwidths of the objective function values. Several catchments also show a substantial shift towards more optimal solutions. The Loisach catchment, as an example, is one of the catchments where this can be observed.

The only catchment that does neither show a decrease in bandwidth nor shift upward for any of the four objective function value distributions, is the Orge catchment. Moreover, it shows a strong deterioration in terms of objective functions when constraints and sub-grid heterogeneity are added. The processes included in mHMtopo may not be suitable in this case, as the human influences are strong in this catchment. Thus, as stated before, the more parsimonious mHM better reflects the observed dynamics in this catchment in terms of the objective functions.

3.2 Signature comparison

The two models mHM and mHMtopo, both unconstrained and constrained, were compared for their ability to reproduce a wide range of hydrological signatures (Table 3). This comparison is based on the probabilities of improvement $P_{I,S}$ (Figure 10 and Equation 6), but similar results were obtained with the Ranked Probability Score S_{RP}. The results of S_{RP} can be found in the supplementary material in Figures S3 and S4. Overall, the introduction of constraints to mHM lead to an average improvement of 13% in terms of the Euclidian distance D_E. The introduction of topography had a similar effect with an average improvement of 13% for D_E. The constrained mHMtopo case even experienced an average improvement of 19%.

3.2.1 Effect of sub-grid heterogeneity

Similar to the model performance in the validation periods, no clear pattern emerges for the different models' ability to reproduce the system signatures. The Euclidean distance metric, depicted in the last column of Figure 10a, illustrates that the consideration of sub-grid process heterogeneity in mHMtopo leads to a slight overall improvement compared to mHM. Yet, the effect on individual signatures is diverse with some signatures captured to a better degree, while others could be reproduced less well.

Figure 10a shows that the Treene, Orge and Loisach benefit the most from the addition of sub-grid heterogeneity. Especially the Treene has a rather large probability of improvement for most of the signatures. This supports the previous findings that the wetland related processes, which are added in mHMtopo, are important to consider in this wet, peaty catchment.

It is interesting to note that the Orge and Loisach, which showed a considerable decrease in performance in terms of the four calibration objective functions (Figure 6), now exhibit relatively high probabilities of improvement with respect to the signatures when sub-grid heterogeneity is added (Figure 10a). The signatures with the strongest improvements are related to peaks in the low flow period. Similar to the Treene, the low flow processes are better captured with mHMtopo. The relatively large urban area in the Orge may ~~only~~ merely affect the fast, high flow processes, which leads to low performances for $E_{NS,Q}$ in mHMtopo. Nevertheless, a large area of the Orge catchment is still classified as wetland (see also Figure 1), adding several processes that only become dominant in the dry periods. ~~However~~Thus, the low flow peaks may be more adequately represented in mHMtopo. Besides, the information of low flow peaks is fully masked when looking at, for example, $E_{NS,Q}$ or $E_{NS,logQ}$, as the relative importance of peaks in low flows in these metrics is low. First, these metrics consider the whole period of interest, instead of only the low flow period, and, second, the peaks are relatively small compared to the average high flows. Hence, high performances in terms of $E_{NS,Q}$ or $E_{NS,logQ}$ may be misleading, which is very relevant for automatic calibration schemes that often optimize towards these functions. Improvements in, for example, low flow peaks, may remain unnoticed when calibrating on more general objective functions, such as $E_{NS,Q}$, as they mostly rely on the absolute values of model residuals aggregated over the entire model period. This is the result of the frequent absence of homoscedasticity in the model residuals. Therefore, errors in high flows tend to have a higher weight in the objective function than errors in low flows. For the Loisach, the findings are also in agreement with findings of Velázques et al. (2013) that in particular the performance of low flows depends on the choice of the hydrological model. Apparently, here the low flow processes are not easy to capture as in most hydrological models.

Results for the comparative analysis of the individual signatures instead of catchments indicated a considerable degree of improvement for mHMtopo to represent low flows ($Q_{50,low}$, $Q_{95,low}$, $Q_{5,low}$) and peaks during low flows ($Q_{peak,10}$, $Q_{low,peak,50}$) as can be seen in Figure 10a.

A probability distribution of the performance metric of a signature, so S_{RE} or S_{NS}, may indicate whether the feasible space produces many solutions close to optimal. Ideally, a high peak of the distribution function close to one indicates a strong ability of the model to reproduce a certain signature, whereas a flat and widespread distribution or even negative performance values, indicates a more reduced ability to reproduce the signature. Thus, the improved ability of mHMtopo to reproduce low flow signatures becomes more obvious when looking in detail at the probability distributions of, for example, $Q_{50,low}$ in the Treene catchment (Figure 11). The original model of mHM only allows downward percolation and infiltration, which leads to a larger buffer for soil moisture in dry periods. mHMtopo, on the other hand, sustains a shallow groundwater table in wetlands through an upward flux, which leads to a faster response and thus to a better representation of the peaks during dry periods.

In contrast, the 1-day autocorrelations for the total, low flow and high flow periods, are consistently better represented in the original mHM (Figure 10a,b). This indicates that the timing of the flow peaks is better represented in the original model. Likewise, the rising and declining limb densities (RLD and DLD respectively) are also better captured by the original mHM. Similar to the observation that mHM better captures the fast spikey peaks in the Orge catchment, this suggests again that the more simple model structure (mHM) is able to respond faster, while the more complex model structure (mHMtopo) tends to delay the flow of water. A possible explanation for this observation is that the more complex model has more options, in terms of reservoirs, to store the water. As linear reservoirs keep draining, the use of multiple reservoirs can produce a delayed and flattened signal. In addition, as the flood peaks now consist of contributions of the different reservoirs, more solutions exist to reconstruct these flood peaks. These solutions could also contain flatter, delayed peaks, which affect the 1-day autocorrelation. More specifically, for fast responding catchments like the Orge and Loisach, it means a poor representation of the 1-day autocorrelation in mHMtopo, which offers more storage possibilities and thus more "memory" in the system. However, a closer look at the distributions in detail shows that these differences are small. As an example, Figure 12 shows the 1-day autocorrelation distributions for the Loisach catchment. Here, it is apparent that the distributions of mHM and mHMtopo are in accordance.

The findings presented here are in line with some other comparison studies, such as Reed et al. (2004), Nicolle et al. (2014), Orth et al. (2015) and te Linde (2008) who all found that added complexity can but does not necessarily lead to improvements. However, in contrast with Orth et al. (2015), we found that low flows are better represented by the complex models, whereas they found that low flows were the best represented by a very simple model. Nevertheless, it was stated by Staudinger et al. (2011) that processes in summer low flow periods are more complex due to a stronger interaction between fast storages and evaporation. Therefore, they did not find one particular model structure to represent low flows in summer. In addition, the difficulties to represent low flows have been acknowledged by several authors, such as Smakhtin et al. (2001), Pushpalatha et al. (2011) or Van Esse et al. (2013).

3.2.2 Effect of constraints

Figure 10c shows that the addition of prior constraints to mHM strongly improves the signature representation, in particular for, again, the Treene. Apparently, the seasonal runoff constraints help the model to represent the low flows better, which mHMtopo was able to do through the additional processes included. As the upward flux from the groundwater in mHMtopo is counterbalanced in the constrained mHM by different parameters that most likely influence the fast reservoir coefficient and storage, it remains unclear which of the two conceptualizations, i.e. mHM or mHMtopo, is more adequate in this case. Also the Loisach shows a strong improvement when prior constraints are added to mHM (Figure 10c). The reasoning considering the importance of snow still holds. The seasonal runoff constraints help to identify parameter sets that are better able to reproduce the seasonal flows, which are strongly affected by snowmelt.

The additional constraints imposed to mHM do not significantly affect the performance for the Briance and Orge catchment, as can be seen by the nearly white rows in Figure 10c. Notably, the runoff responses in these catchments are not snow dominated, and as evaporation and rainfall are now out of phase, the original model was already able to capture the seasonality reasonably well.

It can be clearly observed from Figure 10c,d that the applied prior constraints yield a strong improvement, in particular on mHM, and in only about 29% (mHM) and 38% (mHMtopo) of the cases a mostly weak performance reduction is observed. This indicates that, in spite of being constrained by the transfer functions that link parameters to catchment characteristics, additional prior constraints do still contain significant discriminatory information to identify unfeasible model solutions, which is in agreement with findings of Hrachowitz et al. (2014).

The picture is less clear for applying constraints to mHMtopo, but improvements are still observed for the majority of the signatures (Figure 10d; see also the empirical distribution function at the bottom of the figures).

Alzette, Loisach and Orge show some deterioration when constraints are added (Figure 10d), indicating that the topography specific constraints (Equations 4 and 5) may not be fully applicable to these catchments. These catchments show a general decrease in the ability to reproduce several signatures when comparing the unconstrained mHMtopo with the constrained case (Figure 10d). This means that the unconstrained mHMtopo and also the constrained mHM, that does not have these topography specific constraints, will outperform the constrained mHMtopo with respect to these signatures. This is also supported by Figure 10b, that illustrates that for the Alzette, Loisach and Orge the addition of constraints to mHMtopo leads to a reduced ability to represent most signatures compared to the constrained mHM case (see the red pattern in Figure 10b). The rejection of these constraints implies that for these catchments, soil moisture storage capacity in wetlands may be equal or even larger than soil moisture storage capacity in the hillslope and plateau area. This may be true for the Loisach, especially as Kunstmann et al. (2006) found that the karstic nature in these areas even leads to water flowing from the neighboring Ammer catchment to the Loisach. Considering these groundwater leakages, the model may need extra storage to correct for it in the hydrograph.

In Figure 10c,d it may also be noted that the constraints do not add information to mHM and mHMtopo with respect to the autocorrelation functions (AC, AC_{low}, AC_{high}) and rising and declining limb densities (RLD, DLD). This makes sense as the applied constraints here merely affect the seasonal patterns. Therefore, improvements can be observed for signatures addressing low and high flow periods, such as $Q_{low,95}$ and $Q_{high,95}$.

Figure 10d shows that none of the signatures consistently improves or deteriorates. This indicates that care must be taken by including more specific expert knowledge constraints. General constraints, like the runoff constraints, can easily be applied to multiple catchments and lead to improvements as Figure 10c shows, but assumptions about internal model behavior should experimentally be well founded. Even though based on several experimental studies, the topography based parameter constraints applied (Eqs. 4 and 5) were not suitable in all cases, and lead to a random pattern of individual signature improvements/deterioration. Thus, it was expected that additional constraints should narrow down the ‘plausible’ parameter space and would lead to more pronounced differences in performances. Nevertheless, ~~This~~ the results merely support~~s~~ findings of Holländer et al. (2009), where different choices of expert modelers lead to a variety of outcomes.

3.2.3 Combined effect of constraints and sub-grid heterogeneity

Figure 10b shows the effect of additional sub-grid variability on the constrained models. Most of the catchments show a slight overall improvement, indicated by the relatively blue shades for Euclidian distance. In general, the patterns observed in Figure 10b are relatively similar to the patterns observed in Figure 10a. It seems that the applied constraints generally enhance the effects caused by the model structure. This can be seen from more darker colors of red and blue, but also from the flatter distribution function (bottom of Figure 10b). Thus, when the model has already a relatively large probability of improvement for certain signatures, the constraints help to zoom in on the good solutions. When this is not the case, the model drifts further away.

Nevertheless, the Briance and Broye show a more different effect, indicating a positive effect of the constraints for mHMtopo. For the Briance, a red box for the Euclidian distance in Figure 10a turned blue in Figure 10b. The Broye gained higher probabilities of improvement, represented by more dark blue colors in Figure 10b. Apparently, the solutions maintained for the unconstrained mHMtopo case still contained a relatively large number of implausible solutions. Here, the application of constraints helped to narrow the solution space in such a way that mHMtopo showed improvements compared with the original mHM.

However, it must be noted that the Alzette, Loisach and Orge show a relative low probability of improvement again. This due to the rejection of the constraints given in Equations 4 and 5, as discussed before in comparison with Figure 10d.

Figure 10e shows the combined effect of constraints and sub-grid heterogeneity on the signature representation compared with the original, unconstrained mHM. The Euclidian distance in the last column of Figure 10e, shows again that most catchments profit from the addition of constraints and sub-grid heterogeneity to mHM. It was noted before that mHMtopo has an improved ability to represent the low flow statistics, whereas the original mHM better represented fast flows signatures like rising limb density (RLD) or autocorrelation (AC). In Figure 10e, even a further contrast between the fast flow and low flow domain can be observed. More particular, the Treene shows again the most improvements. The rejection of the topography specific constraints in the Alzette, Loisach and Orge introduce also in Figure 10e a more red pattern. Nevertheless, the overall improvements in the low flow domains still lead to a general improvement in the Euclidian distance D_E for the Alzette and Loisach. Only for the Orge catchment, influenced largely by human disturbances, the Euclidian distance D_E shows a clear deterioration in performance.

3.3 Transferability comparison

In a next step, the two models mHM and mHMtopo were calibrated simultaneously on the four catchments Orge, Treene, Broye and Loisach. The parameters were then transferred without further calibration to the three remaining receiver catchments Alzette, Briance and Kinzig. As shown in Figure 13, both models provide a relatively good performance in the validation period with respect to all four calibration objective functions in the receiver catchments as compared to the individual calibration for the same catchments. Compared with the base case of mHM, the Euclidian distances obtained from the calibration objective functions values changed by 2% (mHM with constraints), -4% (mHMtopo) and 1% (mHMtopo with constraints). The Euclidian distances for the signatures improved by 2% for the constrained mHM case. However, mHMtopo had a decrease of 5% and the Euclidian distance almost doubled for the constrained mHMtopo case.

3.3.1 Effect of sub-grid heterogeneity

In general, mHM and mHMtopo showed a considerable ability to reproduce similar objective function values as in the individual calibrations (Figure 13). Both models kept a reasonable performance during validation in terms of the objective function values and did not fail in reproducing the hydrograph with the parameters received from the donor catchments.

For the Alzette, the results obtained with mHM (blue in Figure 13) and mHMtopo (red in Figure 13) are almost identical. For the Briance and Kinzig catchments it is noted that the introduction of sub-grid process heterogeneity, i.e. mHMtopo, leads to a less transferable model. In particular $E_{NS,logQ}$ and $E_{NS,FDC}$ experience a strong decrease in performance (Figure 13). The results also suggest that, in the unconstrained case, the original mHM is better transferable than mHMtopo with respect to catchment signatures (Figure 14a). Most signatures show a low probability of improvement, only some signatures that consider peaks during the low flow periods have a relatively high (blue pattern in Figure 14a) probability of improvement. This indicates again that the more complex mHMtopo mostly affects the low flows.

It should be noted that the transfer functions used in mHMtopo were adopted for similar parameters from the original mHM. However, it may well be that the assumed functional relations are less valid in a more complex setting. The MPR was developed around the simple model structure, and also refined several times (Samaniego et al, 2010; Kumar et al., 2013a). Similar efforts are required for refining the regionalization for a topography driven model in order to make mHMtopo as transferable as the original mHM. In addition, the global parameter ranges, that do not have a real physical meaning, were also derived for the original mHM and may need adjustments for mHMtopo.

3.3.2 Effect of constraints

Imposing prior constraints in mHMtopo leads to a strong increase in performance again in the Kinzig catchment compared to the unconstrained case (Figure 13). This indicates that the applied constraints are very suitable for this catchment, but less for the Briance catchment, where only a minor improvement is observed. The Kinzig catchment is characterized by a rather large elevation difference and relatively high contribution of snow, similar to the Loisach catchment. Hence, the same reasoning for this catchment holds as for the Loisach catchment that the seasonal runoff constraints help in the seasonal flow patterns. Besides, the role of the input data may likely influence the modelling results for this catchment, since the Kinzig catchment has a large difference in elevation.

When comparing the signatures for the constrained mHM and mHMtopo (Figure 14b), it can be observed that the Alzette and Kinzig catchments benefit from additional process heterogeneity and constraints, while the constrained mHM is still better representing the signatures in the Briance catchment. In general, the constraints do not have much influence on the Briance catchment, as indicated by a relatively white row in Figures 14c and 14d. The unconstrained mHM already was better transferable for this catchment compared to mHMtopo (see Figure 14a), this remains the same in the constrained cases. The other two catchments are much more sensitive to the constraints and show now a better transferability, in particular with respect to the low flow signatures.

Furthermore, results shown in Figures 14c,d suggest that prior constraints can add transferability to both models in terms of signatures as highlighted by the probability of improvements for most signatures. For the Kinzig catchment the constrained mHMtopo model is clearly better transferable than the unconstrained mHMtopo as well as mHM with constraints. This was already noted before, when looking at the performances (Figure 13), but it is here confirmed for the signatures.

In general, it can be stated that the addition of topography-guided sub-grid process heterogeneity *per se* does not necessarily lead to a pronounced difference in model transferability in all parts of flow regimes. Some improvements were noticed in low flow signature measures. Significant improvements can rather be observed when applying constraints, as illustrated in ~~particular for the Kinzig~~ in Figures 14c,d. The addition of constraints to mHMtopo shows high probabilities of improvements over the full range of signatures (Figure 14d), in particular for the Kinzig. Also for mHM (Figure 14c), even though more moderate, most of the signatures show a relatively large probability of improvement when applying constraints. This test of model transferability underlines the considerable potential of prior constraints to improve the representation of hydrological signatures.

3.3.3 Effect of constraints and sub-grid heterogeneity

In the transferability test, Alzette and Kinzig have an improved signature representation in terms of the Euclidian distance when constraints and sub-grid heterogeneity both are added to mHM, as can be seen in Figure 14e. For these catchments, the biggest improvements , compared with the base case of the unconstrained mHM, are again observed for the low flow statistics.

The Briance catchment shows a general decrease in the ability to represent the signatures. The constraints did not help here (white rows in Figure 14d) and from Figure 14a it was already observed that the unconstrained mHM was more transferable than mHMtopo. Looking back at Figure 10a, it can also be noted that in the individual calibration mHM slightly outperformed mHMtopo for this catchment with respect to the signatures (lightred Euclidian distance). This indicates that the processes in mHMtopo may not adequately represent the processes in this catchment, which is emphasized when the model receives the parameters derived in other catchments. In addition, the derived global relations may not hold for this catchment. Apparently, this catchment, which is gently sloped with agriculture, is significantly different from the other catchments used in calibration. The calibration catchments of Loisach and Broye are more mountainous catchments, whereas the Treene is very flat and wetland dominated. In nature, the Orge catchment should be relatively similar, but this catchment is strongly affected by urbanization.

3.4 General limitations and outlook

It should be noted that the input data may have a big influence on the experiment. For example, the input resolution of the E-OBS forcing data is 24km by 24km, while the catchments are relatively small. In a few cases, the catchments are just covered by a couple of E-OBS data-cells. In addition, as the E-OBS data is a product derived from the interpolation of station data, peaks in rainfall may have been averaged out. In such cases, the detailed process representation in mHMtopo may thus not be warranted. Due to pronounced topography-induced precipitation heterogeneity (e.g. Hrachowitz and Weiler, 2011) this will be more problematic for catchments with marked relief than for catchments that are characterized by a more subdued topography. For example, the Treene benefits most from mHMtopo and is very flat, whereas the steep Loisach needs additional constraints.

In addition to this, one may wonder what the effect of a different spatial model resolution would be. In the extreme case where one modeling cell could be classified as a certain landscape as a whole, the relative importance of the different processes in mHMtopo will increase. Thus, when the assumed processes in the cell are adequate, the performance will increase. Nevertheless, incorrect functional relations may also become more apparent on finer modelling scales as less upscaling is required.

The assumptions made in the applied functional relationships may also affect the outcomes of this experiment. In future work, these relationship may need refinement for mHMtopo. Besides this, the threshold values to delineate the landscape units were originally derived for one specific catchment. The general validity of these thresholds needs to be tested in future research.

4 Conclusions

In this study the value of incorporating topography-controlled sub-grid process heterogeneity together with semi-quantitative model constraints to increase hydrological consistency and spatial transferability of the distributed, conceptual model mHM was tested. Both, the unconstrained and constrained applications of the original mHM and the topography-based mHMtopo were applied to seven distinct catchments across Europe.

On balance, the addition of topography-based sub-grid ~~variability~~ process heterogeneity moderately improved mHM. Different hydrological signatures indicated that in particular the representation of low flows improved by allowing for increased sub-grid process heterogeneity. This could be contributed mostly to additional processes which were missing in the original mHM. Especially in catchments where the process of capillary rise is likely to be more important, it became clear that low flows signatures were better represented. Nevertheless, the timing of flow peaks was better captured by the original mHM model. In summary, the addition of topography based sub-grid process heterogeneity in the model structure of a distributed model regionalized through soil and land use, was to a moderate degree able to improve the general model performance in the study catchments while more adequately reflecting internal processes.

The use of prior, semi-quantitative constraints proved highly effective in the study catchments as it forces the model to reproduce plausible patterns of partitioning between runoff and evaporative fluxes. Especially in cases where runoff and evaporation are out of phase, the constraints were shown to be valuable. These conclusions were largely drawn from the models' varying ability to reproduce observed catchment signatures.

In addition, it was shown that such an improved hydrological consistency at the sub-grid scale combined with the use of suitable model constraints and functional relationships, can be beneficial for transferring models and predicting flows without further calibration in other catchments.

Concluding, the addition of topography-based sub-grid ~~variability~~ process heterogeneity and the use of prior semi- quantitative constraints were shown to be promising and lead to moderate improvements in terms of process representation and transferability.

Acknowledgements

We would like to acknowledge the European Commission FP7 funded research project "Sharing Water-related Information to Tackle Changes in the Hydrosphere– for Operational Needs" (SWITCH-ON, grant agreement number 603587), as this study was conducted within the context of SWITCH-ON.

We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu)

We acknowledge The Global Runoff Data Centre, D -56002 Koblenz, Germany for providing discharge data.

References

Abdulla, F. A. and Lettenmaier, D. P.: Development of regional parameter estimation equations for a macroscale hydrologic model, Journal of Hydrology, 197, 230 –257, doi:http://dx.doi.org/10.1016/S0022-1694(96)03262-3, 1997.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, Hydrological Processes, 26, 2206–2210, doi:10.1002/hyp.9264, 2012.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrology and Earth System Sciences, 13, 1757–1764, doi:10.5194/hess-13-1757-2009, 2009.

Arino, O., Ramos, J., Kalogirou, V., Defourny, P., and Achard, F.: GlobCover 2009; ESA Living Planet Symposium, 27 June . 2 July 2010, Bergen, Norway.

Bergström, S.: The HBV model: Its structure and applications, Swedish Meteorological and Hydrological Institute, Norrköping,1992.

Beven, K.J.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in Water Resources, 16, 41 –51, doi:http://dx.doi.org/10.1016/0309-1708(93)90028-E, research Perspectives in Hydrology, 1993.

Beven, K.J.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18 – 36,

doi:http://dx.doi.org/10.1016/j.jhydrol.2005.07.007, the model parameter estimation experiment MOPEX MOPEX workshop, 2006.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrology and Earth System Sciences, 4, 203–213, doi:10.5194/hess-4-203-2000, 2000.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43–69, doi:10.1080/02626667909491834, 1979.

Booij, M.: Impact of climate change on river flooding assessed with different spatial model resolutions, Journal of Hydrology, 303, 176 – 198, doi:http://dx.doi.org/10.1016/j.jhydrol.2004.07.013, 2005.

Ciarapica, L. and Todini, E.: TOPKAPI: a model for the representation of the rainfall-runoff process at different scales, Hydrological Processes, 16, 207–229, doi:10.1002/hyp.342, 2002.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, n/a–n/a, doi:10.1029/2011WR011721, w05552, 2012.

Das, T., Bárdossy, A., Zehe, E., and He, Y.: Comparison of conceptual model performance using different representations of spatial variability, Journal of Hydrology, 356, 106 – 118, doi:http://dx.doi.org/10.1016/j.jhydrol.2008.04.008, 2008.

Detty, J. M. and McGuire, K. J.: Topographic controls on shallow groundwater dynamics: implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment, Hydrological Processes, 24, 2222–2236, doi:10.1002/hyp.7656, 2010.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015–1031, doi:10.1029/91WR02985, 1992.

Eder, G., Sivapalan, M., and Nachtnebel, H. P.: Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales, Hydrological Processes, 17, 2125–2149, doi:10.1002/hyp.1325, 2003.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrology and Earth System Sciences, 17, 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.

Euser, T., Hrachowitz, M., Winsemius, H. C., & Savenije, H. H. G.: The effect of forcing and landscape distribution on performance and consistency of model structures. Hydrological Processes, 29, 3727-3743, doi:10.1002/hyp.10445, 2015.

FAO/IIASA/ISRIC/ISSCAS/JRC: Harmonized World Soil Database (version 1.2), Rome. Laxenburg, 2012.

Flügel, W.-A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the

River Bröl, Germany, Hydrological Processes, 9, 423–436, doi:10.1002/hyp.3360090313, 1995.

Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., and Gascuel-Odoux, C.: Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models, Hydrology and Earth System Sciences, 19, 105–123, doi:10.5194/hess-19-105-2015, 2015.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrology and Earth System Sciences, 18, 1895–1915, doi:10.5194/hess-18-1895-2014, 2014a.

Gao, H., Hrachowitz, M., Schymanski, S. J., Fenicia, F., Sriwongsitanon, N., and Savenije, H. H. G.: Climate controls how ecosystems size the root zone storage capacity at catchment scale, Geophysical Research Letters, 41, 7916–7923, doi:10.1002/2014GL061668, http://dx.doi.org/10.1002/2014GL061668, 2014GL061668, 2014b.

Gao, H., Hrachowitz, M., Sriwongsitanon, N., Fenicia, F., Gharari, S. and Savenije, H.H.G. (2015). Accounting for the influence of vegetation and topography improves model transferability. Water Resources Research (in review).

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrology and Earth System Sciences, 15, 3275–3291, doi:10.5194/hess-15-3275-2011, 2011.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrology and Earth System Sciences, 18, 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.

Götzinger, J. and Bárdossy, A.: Comparison of four regionalisation methods for a distributed hydrological model, Journal of Hydrology, 333, 374 – 384, doi:http://dx.doi.org/10.1016/j.jhydrol.2006.09.008, 2007.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, doi:10.1002/hyp.6989, 2008.

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

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from ambient air temperature., American Society of Agricultural Engineers (Microfiche collection)(USA). no. fiche no.85-2517., Chicago,1985.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres (1984–2012), 113, D20119, doi: 10.1029/2008JD010201, 2008.

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., and Flühler, H.: Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data, Hydrology and Earth System Sciences, 13, 2069–2094, doi:10.5194/hess-13-2069-2009, 2009.

Hrachowitz, M. and Weiler, M.: Uncertainty of Precipitation Estimates Caused by Sparse Gauging Networks in a Small, Mountainous Watershed, Journal of Hydrologic Engineering, 16(5), 460–471, doi:10.1061/(ASCE)HE.1943-5584.0000331, 2011.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resources Research, 50, 7445–7469, doi:10.1002/2014WR015484, 2014.

Hughes, D.: A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall runoff model, Journal of Hydrology, 501, 111 – 124, doi:http://dx.doi.org/10.1016/j.jhydrol.2013.07.043, 2013.

Hughes, D. and Sami, K.: A semi-distributed, variable time interval model of catchment hydrology—structure and parameter estimation procedures, Journal of Hydrology, 155, 265 – 291, doi:http://dx.doi.org/10.1016/0022-1694(94)90169-4, 1994.

Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281 – 295, doi:http://dx.doi.org/10.1016/j.jhydrol.2004.01.002, 2004.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: Transferring reach- and plotscale understanding to the catchment scale,Water Resources Research, 45, W04428, doi:10.1029/2008WR007225, 2009.

Jothityangkoon, C., Sivapalan, M., and Farmer, D.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, Journal of Hydrology, 254, 174 – 198, doi:http://dx.doi.org/10.1016/S0022-1694(01)00496-6, 2001.

Kapangaziwiri, E., Hughes, D., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrological Sciences Journal, 57, 1000–1019, doi:10.1080/02626667.2012.690881, 2012.

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

Kite, G. W. and Kouwen, N.: Watershed modeling using land classifications, Water Resources Research, 28, 3193–3200, doi:10.1029/92WR01819, 1992.

Knudsen, J., Thomsen, A., and Refsgaard, J.C..: WATBAL, Nordic Hydrology, 17, 347–362, 1986.

Kouwen, N., Soulis, E., Pietroniro, A., Donald, J., and Harrington, R.: Grouped Response Units for Distributed Hydrologic Modeling, Journal of Water Resources Planning and Management, 119, 289–305, doi:10.1061/(ASCE)0733-9496(1993)119:3(289),1993.

Kumar, P.: Typology of hydrologic predictability, Water Resources Research, 47, W00H05, doi:10.1029/2010WR009769, 2011.

Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, Journal of Hydrology, 392, 54 – 69,doi:http://dx.doi.org/10.1016/j.jhydrol.2010.07.047, 2010.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resources Research, 49, 5700–5714, doi:10.1002/wrcr.20431, 2013a.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, doi:10.1029/2012WR012195, 2013b.

Kunstmann, H., Krause, J., and Mayr, S.: Inverse distributed hydrological modelling of Alpine catchments, Hydrology and Earth System Sciences, 10, 395–412, doi:10.5194/hess-10-395-2006, 2006.

Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, Water Resources Research, 43, W06428, doi:10.1029/2006WR005608, 2007.

Leavesley, G. and Stannard, L.: Application of remotely sensed data in a distributed-parameter watershed model, Application of Remote Sensing in Hydrology, Saskatoon, 47-68, 1990.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, EOS, Transactions American Geophysical Union, 89, 93–94, 2008.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research: Atmospheres, 99, 14 415–14 428,doi:10.1029/94JD00483, 1994.

Lindstrom, G., Pers, C., Rosberg, J., Stromqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, 2010.

Livneh, B., Kumar, R., and Samaniego, L.: Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin, Hydrological Processes, 29, 4638–4655, doi:10.1002/hyp.10601, 2015.

McGlynn, B. L., McDonnell, J. J., Seibert, J., and Kendall, C.: Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations, Water Resources Research, 40, W07504, doi:10.1029/2003WR002494, 2004.

Merz, B. and Bárdossy, A.: Effects of spatial variability on the rainfall runoff process in a small loess catchment, Journal of Hydrology, 212–213, 304 – 317, doi:http://dx.doi.org/10.1016/S0022-1694(98)00213-3, 1998.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resources Research, 43, W05434, doi:10.1029/2006WR005184, 2007.

Mou, L., Tian, F., Hu, H., and Sivapalan, M.: Extension of the Representative Elementary Watershed approach for cold regions: constitutive relationships and an application, Hydrol. Earth Syst. Sci., 12, 565-585, doi:10.5194/hess-12-565-2008, 2008

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrology and Earth System Sciences, 17, 1189–1204, doi:10.5194/hess-17-1189-2013, 2013.

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J.-M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B., and Morice, E.: Benchmarking hydrological models for low-flow simulation and forecasting on French catchments, Hydrology and Earth System Sciences, 18, 2829–2857, doi:10.5194/hess-18-2829-2014, 2014.

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S.: Height Above the Nearest Drainage – a hydrologically relevant new terrain model, Journal of Hydrology, 404, 13 – 29, doi:http://dx.doi.org/10.1016/j.jhydrol.2011.03.051, 2011.

Obled, C., Wendling, J., and Beven, K.: The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data, Journal of Hydrology, 159, 305 – 333,

doi:http://dx.doi.org/10.1016/0022-1694(94)90263-1, 1994.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147 – 159, doi:http://dx.doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Pape, P. L., Ayrault, S., and Quantin, C.: Trace element behaviour and partition versus urbanization gradient in an urban river (Orge River, France), Journal of Hydrology, 472–473, 99 – 110, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.09.042, 2012.

Perrin, C., Michel, C., and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology, 242, 275 – 301, doi:http://dx.doi.org/10.1016/S0022-1694(00)00393-0, 2001.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275 – 289, doi:http://dx.doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Pokhrel, P., Gupta, H. V., and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resources Research, 44, W12419, doi:10.1029/2007WR006615, 2008.

Pushpalatha, R., Perrin, C., Moine, N. L., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve lowflow simulation, Journal of Hydrology, 411, 66 – 76, doi:http://dx.doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and multivariate evaluation of water fluxes and states over European river basins, Journal of Hydrometeorology,17, 287-307, doi: http://dx.doi.org/10.1175/JHM-D-15-0054.1 ~~*in press*~~, 2015.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J.,, and Participants, D.: Overall distributed model intercomparison project results, Journal of Hydrology, 298, 27 – 60, doi:http://dx.doi.org/10.1016/j.jhydrol.2004.03.031, the Distributed Model Intercomparison Project (DMIP), 2004.

Refsgaard, J. C. and Storm, B.: MIKE SHE, in: Computer models of watershed hydrology, edited by: Singh, V. P., Water Resources Publications, Littleton, 809–846, 1995.

Reggiani, P., and T. H. M. Rientjes.: Flux parameterization in the representative elementary watershed approach: Application to a natural basin, Water Resour. Res., 41, *W04013*, *doi:*10.1029/2004WR003693., 2005

Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, Advances in Water Resources, 22, 367 – 398, doi:http://dx.doi.org/10.1016/S0309-1708(98)00012-8, 1998.

Reichert, P. and Omlin, M.: On the usefulness of overparameterized ecological models, Ecological Modelling, 95, 289 – 299, doi:http://dx.doi.org/10.1016/S0304-3800(96)00043-9, 1997.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sensing of Environment, 112, 3469 – 3481, doi:http://dx.doi.org/10.1016/j.rse.2008.03.018, 2008.

Rodríguez-Iturbe, I., Isham, V., Cox, D. R., Manfreda, S., and Porporato,A.: Space-time modeling of soil moisture: Stochastic rainfall forcing with heterogeneous vegetation, Water Resources Research, 42, W06D05, doi:10.1029/2005WR004497, 2006.

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

Samaniego, L., Kumar, R., and Zink, M.: Implications of parameter uncertainty on soil moisture drought analysis in Germany, Journal of Hydrometeorology, 14, 47–68, doi:10.1175/JHM-D-12-075.1, 2013.

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrology and Earth System Sciences, 14, 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

Schmalz, B. and Fohrer, N.: Comparing model sensitivities of different landscapes using the ecohydrological SWAT model, Advances in Geosciences, 21, 91–98, doi:10.5194/adgeo-21-91-2009, 2009.

Schmalz, B., Tavares, F., and Fohrer, N.: Modelling hydrological processes in mesoscale lowland river basins with SWAT—capabilities and challenges, Hydrological Sciences Journal, 53, 989–1000, doi:10.1623/hysj.53.5.989, 2008.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrology and Earth System Sciences, 11, 891–906, doi:10.5194/hess-11-891-2007, 2007.

Seibert, J. and McDonnell, J.: Gauging the Ungauged Basin: Relative Value of Soft and Hard Data, Journal of hydrologic engineering, A4014004, doi:10.1061/(ASCE)HE.1943-5584.0000861, 2013.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resources Research, 38,23–1–23–14, doi:10.1029/2001WR000978,2002.

Seibert, J., Bishop, K., Rodhe, A., and McDonnell, J. J.: Groundwater dynamics along a hillslope: A test of the steady state hypothesis, Water Resources Research, 39, 1014, doi:10.1029/2002WR001404, 2003a.

Seibert, J., Rodhe, A., and Bishop, K.: Simulating interactions between saturated and unsaturated storage in a conceptual runoff model, Hydrological Processes, 17, 379–390, doi:10.1002/hyp.1130, 2003b.

Shamir, E., Imam, B., Morin, E., Gupta, H. V., and Sorooshian, S.: The role of hydrograph indices in parameter estimation of rainfall–runoff models, Hydrological Processes, 19, 2187–2207, doi:10.1002/hyp.5676, 2005.

Singh, V. P.: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, Hydrological Processes, 11, 1649–1669, doi:10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1, 1997.

Smakhtin, V.: Low flow hydrology: a review, Journal of Hydrology, 240, 147 – 186, doi:http://dx.doi.org/10.1016/S0022-1694(00)00340-1, 2001.

Spence, C., Guan, X. J., Phillips, R., Hedstrom, N., Granger, R., and Reid, B.: Storage dynamics and streamflow in a catchment with a variable contributing area, Hydrological Processes, 24, 2209–2221, doi:10.1002/hyp.7492, 2010.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrology and Earth System Sciences, 15, 3447–3459, doi:10.5194/hess-15-3447-2011, 2011.

te Linde, A. H., Aerts, J. C. J. H., Hurkmans, R. T. W. L., and Eberle, M.: Comparing model performance of two rainfall runoff models in the Rhine basin using different atmospheric forcing data sets, Hydrology and Earth System Sciences, 12, 943–957, doi:10.5194/hess-12-943-2008, 2008.

Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D., and Samaniego, L.: Seasonal Soil Moisture Drought Prediction over Europe using the North American Multi-Model Ensemble (NMME), Journal of Hydrometeorology, 2329-2344, doi:10.1175/JHM-D-15-0053.1, 2015.

Tian, F., Hu, H., Lei, Z., and Sivapalan, M.: Extension of the Representative Elementary Watershed approach for cold regions via explicit treatment of energy related processes, Hydrol. Earth Syst. Sci., 10, 619-644, doi:10.5194/hess-10-619-2006, 2006.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resources Research, 43, W01413, doi:10.1029/2005WR004723, 2007.

Tonkin, M. J. and Doherty, J.: A hybrid regularized inversion methodology for highly parameterized environmental models, Water Resources Research, 41, W10412, doi:10.1029/2005WR003995, 2005.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resources Research, 42, W02411, doi:10.1029/2004WR003800, 2006.

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model, Journal of Hydrology, 291, 278 – 296, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.12.038, 2004.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., and Savenije, H.: Predicting the ungauged basin: Model validation and realism assessment, Frontiers in Earth Science, 3, doi:10.3389/feart.2015.00062, 2015

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., and Lobligeois, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrology and Earth System Sciences, 17, 4227–4239, doi:10.5194/hess-17-4227-2013, 2013.

Vannametee, E., Karssenberg, D., Bierkens, M.F.P.: Towards closure relations in the Representative Elementary Watershed (REW) framework containing observable parameters: Relations for Hortonian overland flow, Advances in Water Resources, 43, 52-66, doi:http://dx.doi.org/10.1016/j.advwatres.2012.03.029., 2012

Velázquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., and Turcotte, R.: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources, Hydrology and Earth System Sciences, 17, 565–578, doi:10.5194/hess-17-565-2013, 2013.

Wagener, T. and Gupta, H.: Model identification for hydrological forecasting under uncertainty, Stochastic Environmental Research and Risk Assessment, 19, 378–387, doi:10.1007/s00477-005-0006-5, 2005.

Wigmosta, M. S., Vail, L. W., and Lettenmaier, D. P.: A distributed hydrology-vegetation model for complex terrain, Water Resources Research, 30, 1665–1679, doi:10.1029/94WR00436, 1994.

Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, Oxford, 2011.

Winsemius, H. C., Savenije, H. H. G., and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrology and Earth System Sciences, 12, 1403–1413, doi:10.5194/hess-12-1403-2008, 2008.

Winter, T. C.: THE CONCEPT OF HYDROLOGIC LANDSCAPES1, JAWRA Journal of the American Water Resources Association, 37, 335–349, doi:10.1111/j.1752-1688.2001.tb00973.x, 2001.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, W09417, doi:10.1029/2007WR006716, 2008.

Young, P. C.: Parallel Processes in Hydrology andWater Quality: A Unified Time-Series Approach, Water and Environment Journal, 6, 598–612, doi:10.1111/j.1747-6593.1992.tb00796.x, 1992.

Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modeling water flow and mass transport in a loess catchment, Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 26, 487 – 507, doi:http://dx.doi.org/10.1016/S1464-1909(01)00041-7, 2001.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrology and Earth System Sciences, 18, 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.

Zhang, G. P., Savenije, H. H. G., Fenicia, F., and Pfister, L.: Modelling subsurface storm flow with the Representative Elementary Watershed (REW) approach: application to the Alzette River Basin, Hydrol. Earth Syst. Sci., 10, 937-955, doi:10.5194/hess-10-937-2006, 2006.

Zhang, G. P. and Savenije, H. H. G.: Rainfall-runoff modelling in a catchment with a complex groundwater flow system: application of the Representative Elementary Watershed (REW) approach, Hydrol. Earth Syst. Sci., 9, 243-261, doi:10.5194/hess-9-243-2005, 2005

Zhao, R.-J.: The Xinanjiang model applied in China, Journal of Hydrology, 135, 371-381, http://dx.doi.org/10.1016/0022-1694(92)90096-E, 1992.

Table 1. Overview of the catchments .

Catchment	Country	Area (km^2)	Elevation (mMSL)	Runoff (mm/y)	Aridity Index (EP/P) (-)	Calibration period	Validation period
Alzette	Luxembourg	1172	194-545	286	0.90	01-01-1978 31-12-1980	01-01-1983 31-12-1987
Briance	France	604	211-719	377	0.88	01-01-1982 01-07-1993	02-07-1993 31-12-2004
Broye	Switzerland	396	391-1494	648	0.71	01-01-1995 02-07-1987	03-07-1987 31-12-2009
Kinzig	Germany	955	172-1084	759	0.67	01-01-1951 31-12-1971	01-01-1971 31-12-1990
Loisach	Germany	243	716-2783	960	0.50	01-01-1976 31-12-1988	01-01-1989 31-12-2001
Orge	France	965	38-196	130	1.34	01-01-1968 01-07-1986	02-07-1986 31-12-2004
Treene	Germany	481	-1-80	428	0.75	01-01-1974 01-07-1989	02-07-1989 31-12-2004

Table 2. Overview of the used data.

Data type	Product	Source	Reference
Soil	HWSD	http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1	FAO/IIASA/ISRIC/ISSCAS/JRC (2012)
Topography	SRTM	http://hydrosheds.cr.usgs.gov/index.php	Lehner et al. (2008)
Discharge	GRDC	http://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html	The Global Runoff Data Centre, D -56002 Koblenz, Germany
Precipitation	E-OBS	http://eca.knmi.nl/download/ensembles/ensembles.php	Haylock et al. (2008)
Landcover	Globcover	http://due.esrin.esa.int/page_globcover.php	Arino et al. (2009)

Table 3. Overview of the used signatures.

Signature	Description	Reference
Q_{MA}	Mean annual runoff	
AC	One day autocorrelation coefficient	Montanari and Toth (2007)
AC_{low}	One day autocorrelation low flow period	Euser et al. (2013)
AC_{high}	One day autocorrelation high flow period	Euser et al. (2013)
RLD	Rising limb density	Shamir et al. (2005)
DLD	Declining limb density	Shamir et al. (2005)
Q_{5}	Flow exceeded in 5% of the time	Jothityangkoon et al. (2001)
Q_{50}	Flow exceeded in 50% of the time	Jothityangkoon et al. (2001)
Q_{95}	Flow exceeded in 95% of the time	Jothityangkoon et al. (2001)
$Q_{5,low}$	Flow exceeded in 5% of the low flow time	Yilmaz et al. (2008)
$Q_{50,low}$	Flow exceeded in 50% of the low flow time	Yilmaz et al. (2008)
$Q_{95,low}$	Flow exceeded in 95% of the low flow time	Yilmaz et al. (2008)
$Q_{5,high}$	Flow exceeded in 5% of the high flow time	Yilmaz et al. (2008)
$Q_{50,high}$	Flow exceeded in 50% of the high flow time	Yilmaz et al. (2008)
$Q_{95,high}$	Flow exceeded in 95% of the high flow time	Yilmaz et al. (2008)
Peaks	Peak distribution	Euser et al. (2013)
$Peaks_{low}$	Peak distribution low flow period	Euser et al. (2013)
$Peaks_{high}$	Peak distribution high flow period	Euser et al. (2013)
$Q_{peak,10}$	Flow exceeded in 10% of the peaks	
$Q_{peak,50}$	Flow exceeded in 50% of the peaks	
$Q_{low,peak,10}$	Flow exceeded in 10% of the low flow peaks	
$Q_{low,peak,50}$	Flow exceeded in 10% of the low flow peaks	
$Q_{high,peak,10}$	Flow exceeded in 10% of the high flow peaks	

$Q_{high,peak,50}$	Flow exceeded in 50% of the high flow peaks	
AC_{serie}	Autocorrelation series (200 days lag time)	Montanari and Toth (2007)

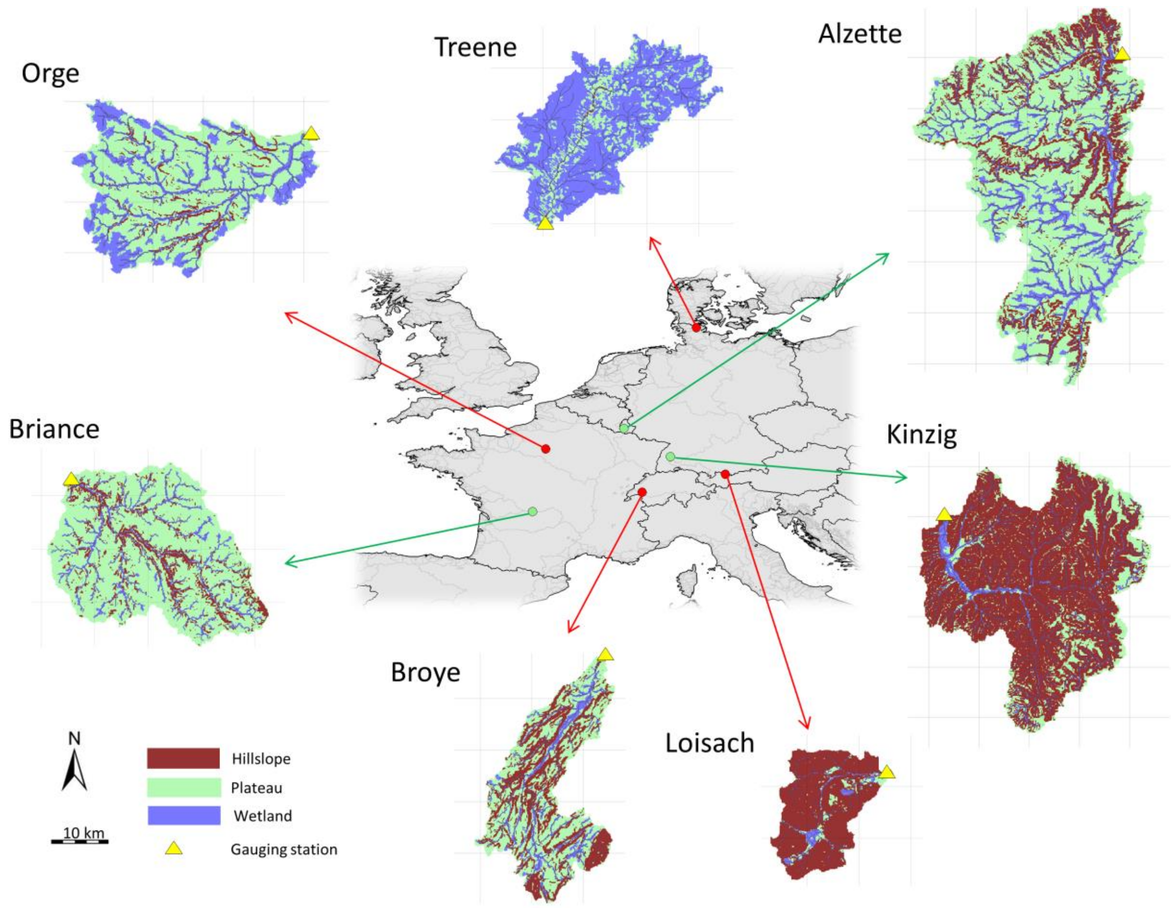


Figure 1. The location of the seven study catchments and their respective landscape classes according to HAND and local slope. Catchments represented by red and green symbols in the context map indicate donor and receiver catchments, respectively, for the transferability analysis. Displayed grids correspond to the modelling grids used in mHM(topo).

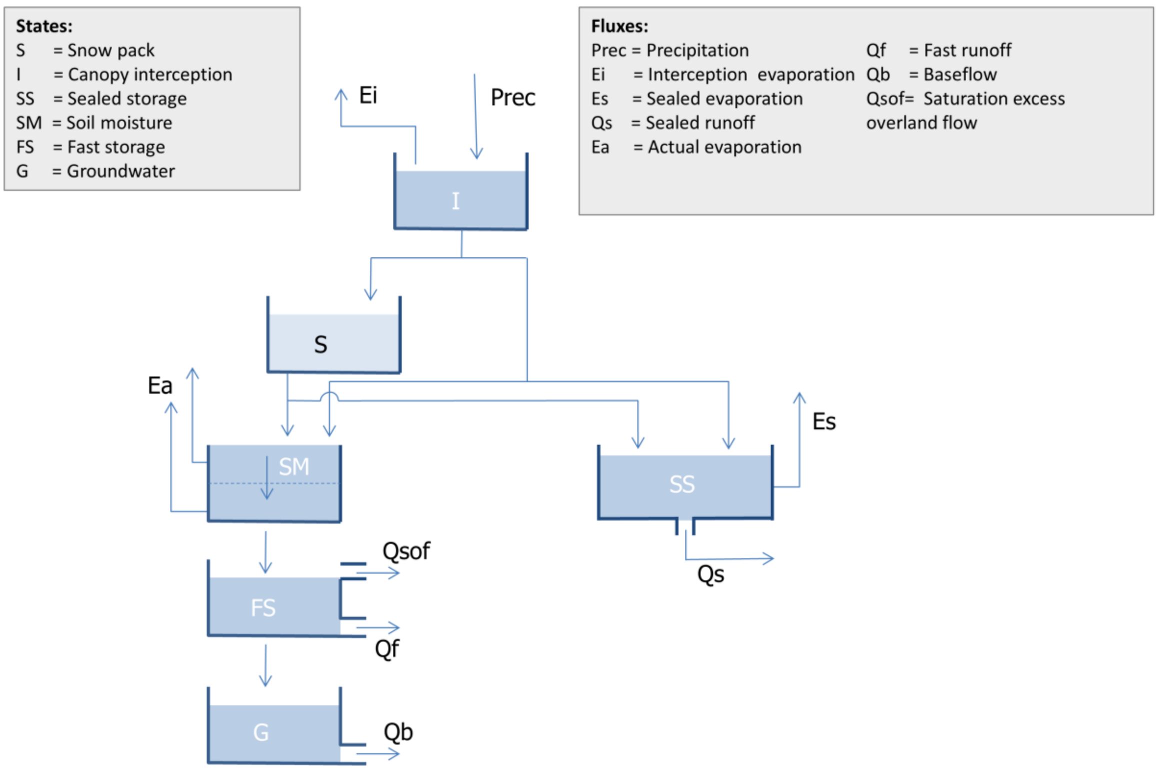


Figure 2. The original mHM model structure. The effective precipitation is determined by an interception (I) and snow routine (S). Afterwards, the effective precipitation enters a soil moisture reservoir (SM) or is directly routed to a fast reservoir that accounts for sealed areas (SS). The water in the soil moisture reservoir either transpires or percolates further down to a fast runoff reservoir (FS), i.e. shallow subsurface flow. Eventually, the base flow component of the runoff is obtained from a slow groundwater reservoir (G).

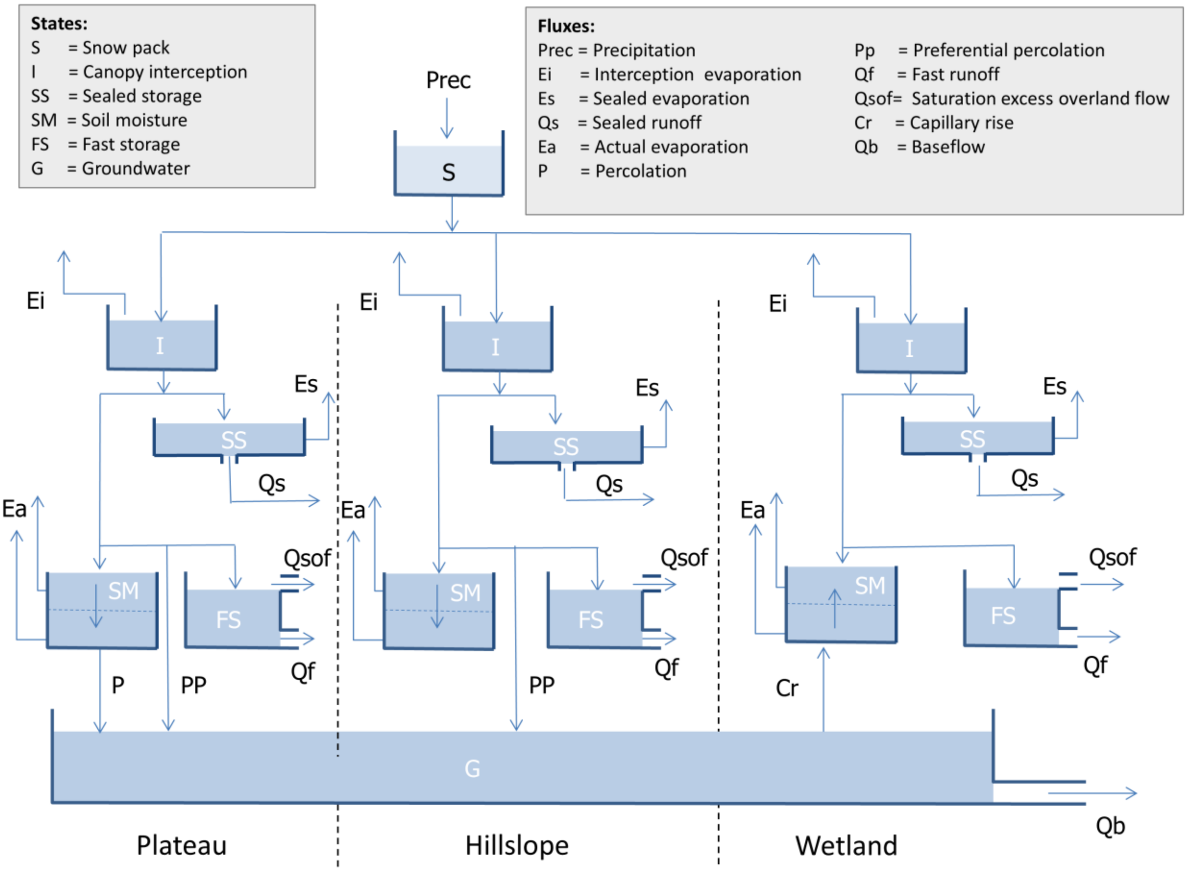


Figure 3. The mHMtopo model structure with different configurations of states and fluxes for the landscape classes plateau, hillslope, and wetland, which are based on topography. First, a shared snow module (S) divides the effective precipitation over the landscape classes. The three classes all have an interception module (I), fast reservoir accounting for sealed areas (SS), soil moisture routine (SM) and fast reservoir (FS). The plateau landscapes are assumed to feed the groundwater through percolation (P) from the soil moisture and preferential percolation (PP). The steeper hillslope areas are assumed to merely feed the groundwater through preferential percolation (PP), whereas the wetlands receive water through capillary rise (Cr). The base flow is determined by a shared groundwater reservoir.

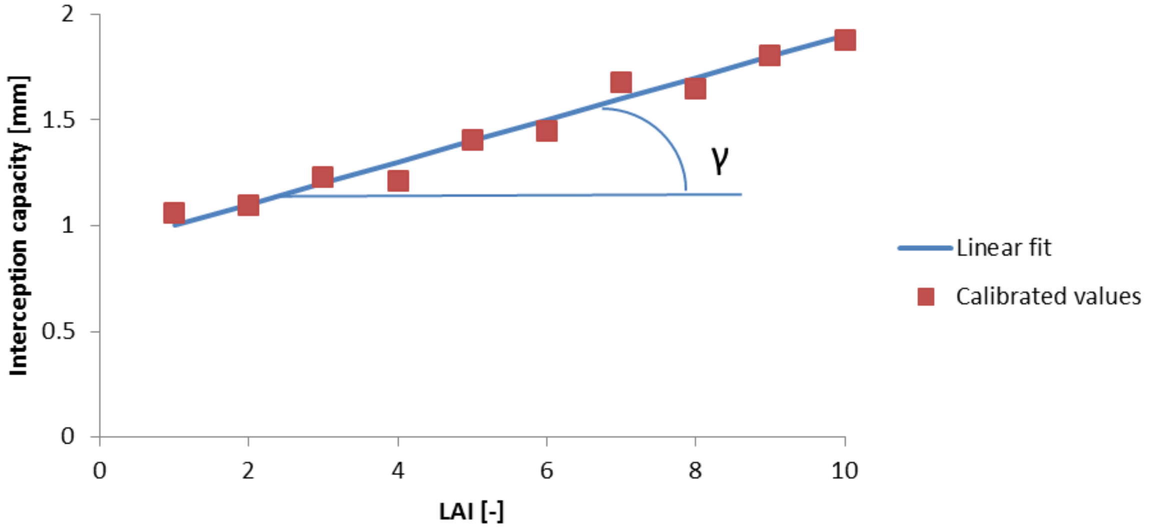


Figure 4. Function relationship between leaf area index (LAI) and the hydrologic parameter interception capacity ($I_{0,max}$) defined by the global parameter γ, based on fictional data for illustration.

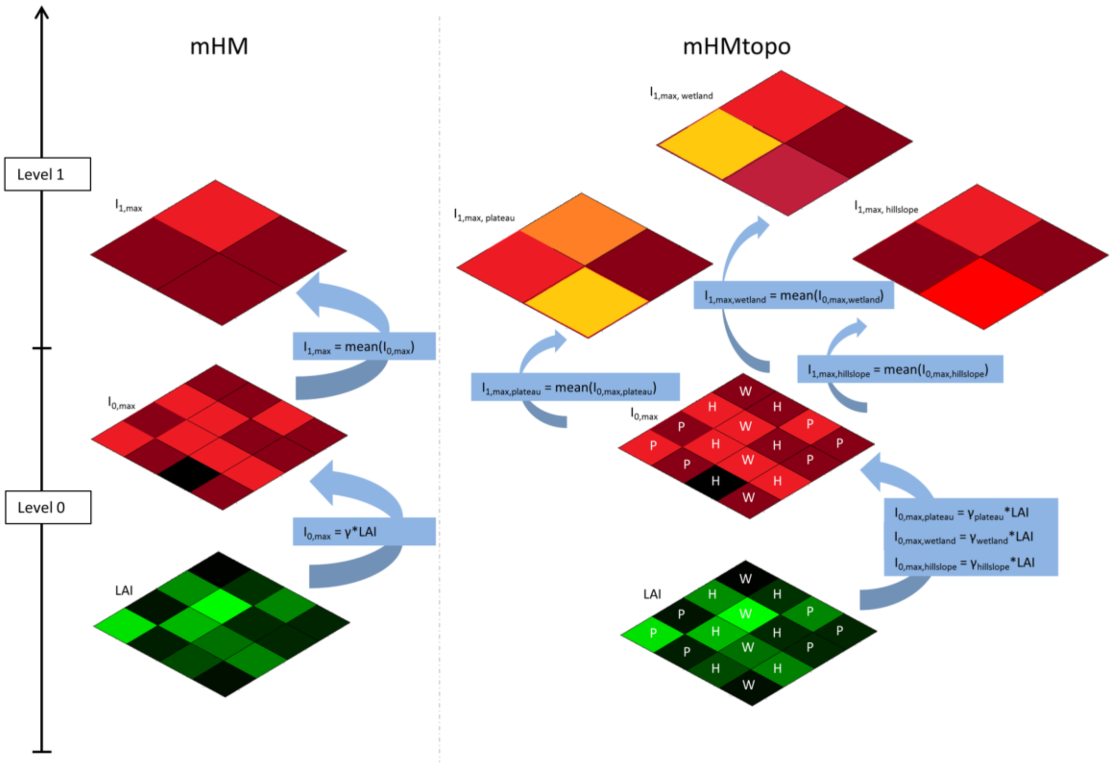


Figure 5. Schematic representation of the original MPR (left) and the adjusted MPR (right) for the maximum interception capacity ($I_{1,max}$). On the input level 0, the leaf area index (LAI) is linked through the global, generally valid, parameter γ with $I_{0,max}$. In a last step, the mean is used for upscaling, yielding $I_{1,max}$ at the modelling resolution. For mHMtopo, the functional relations are the same, but plateau (P), hillslope (H) and wetland (W) have their own global parameters γ. The upscaling is subsequently carried out over each landscape class within each grid cell. This leads to the interception capacities of plateau, hillslope and wetland ($I_{1,max,plateau}$, $I_{1,max,hillslope}$ and $I_{1,max,wetland}$).

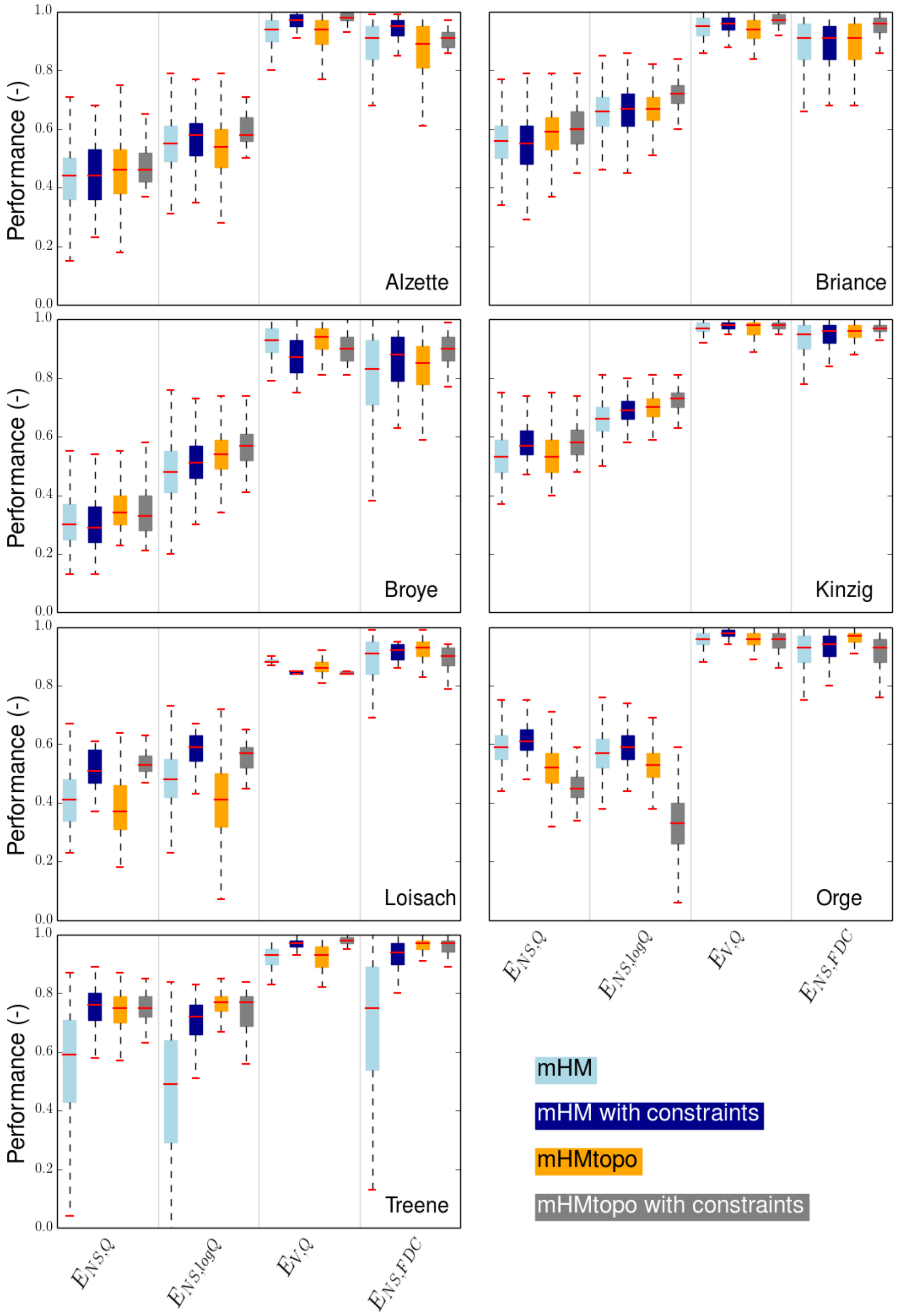


Figure 6. Nash-Sutcliffe efficiency ($E_{NS,Q}$), log Nash-Sutcliffe efficiency ($E_{NS,logQ}$), volume error ($E_{V,Q}$) and log Nash-Sutcliffe efficiency of the flow duration curve ($E_{NS,FDC}$) for the seven catchments in the validation periods. The optimal value for all four criteria is 1, whereas 0 is regarded to have a low performance. The boxplots are formed by the Pareto space spanned by the four objective functions.

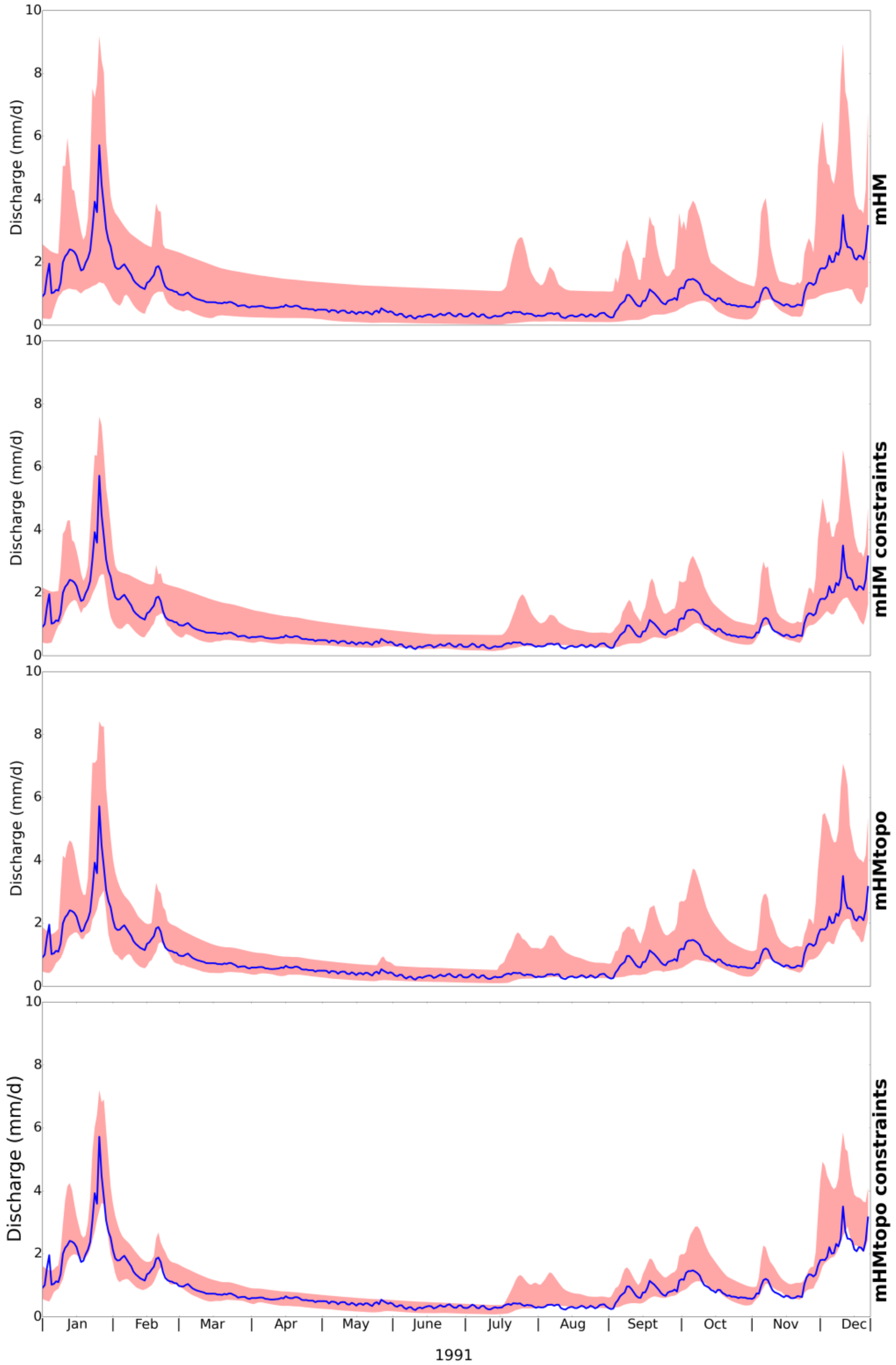


Figure 7. Hydrographs for the Treene catchment, with respectively the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.

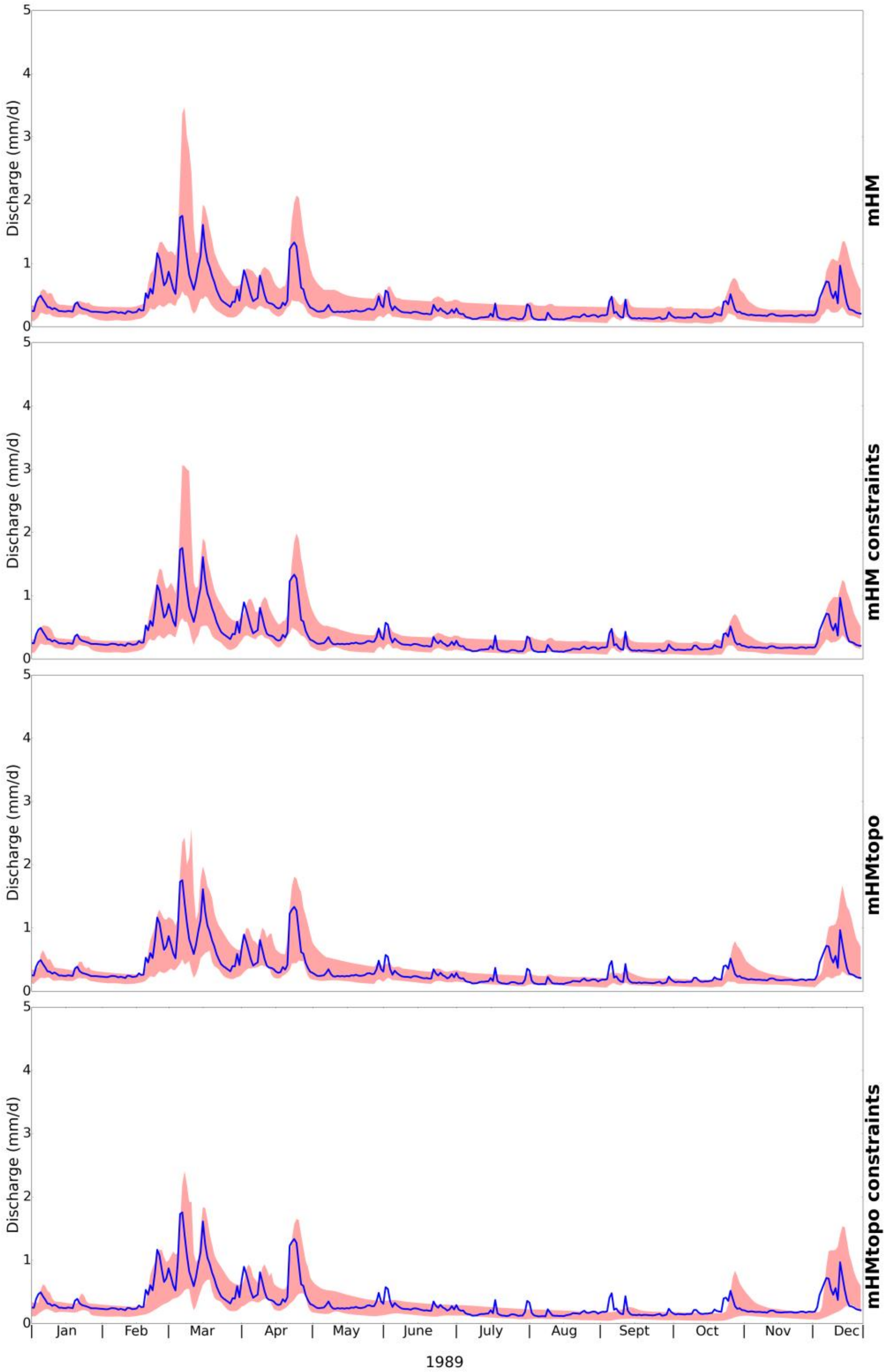


Figure 8. Hydrographs for the Orge catchment, with respectively the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.

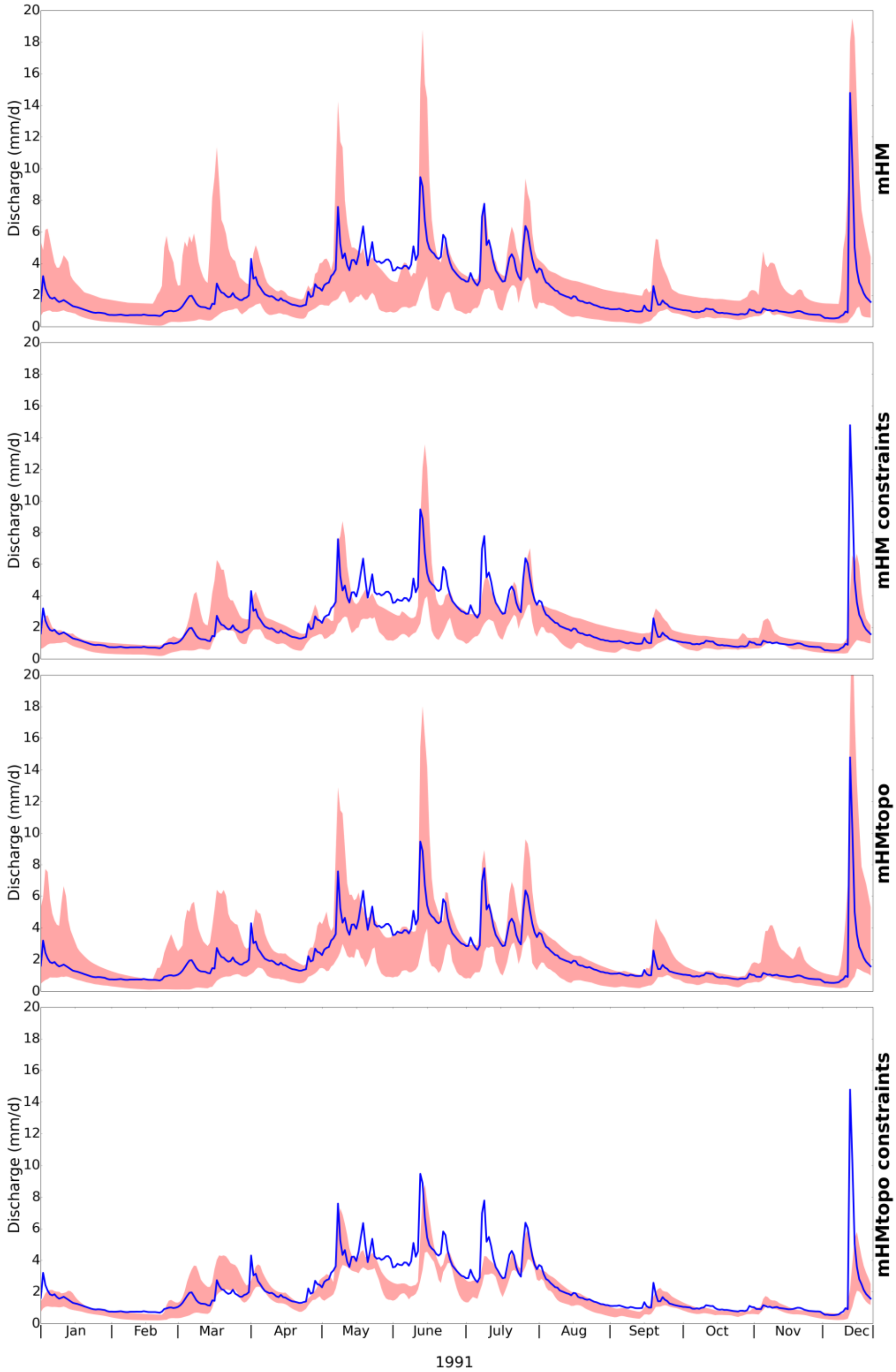
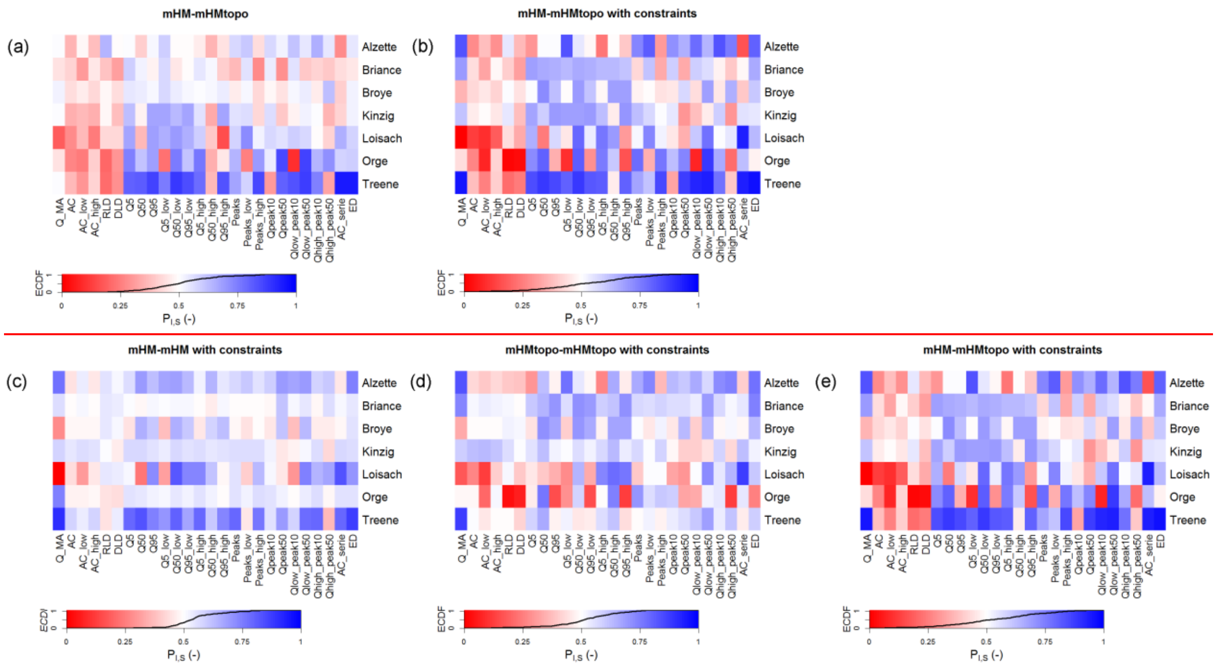


Figure 9. Hydrographs for the Loisach catchment, with respectively the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.



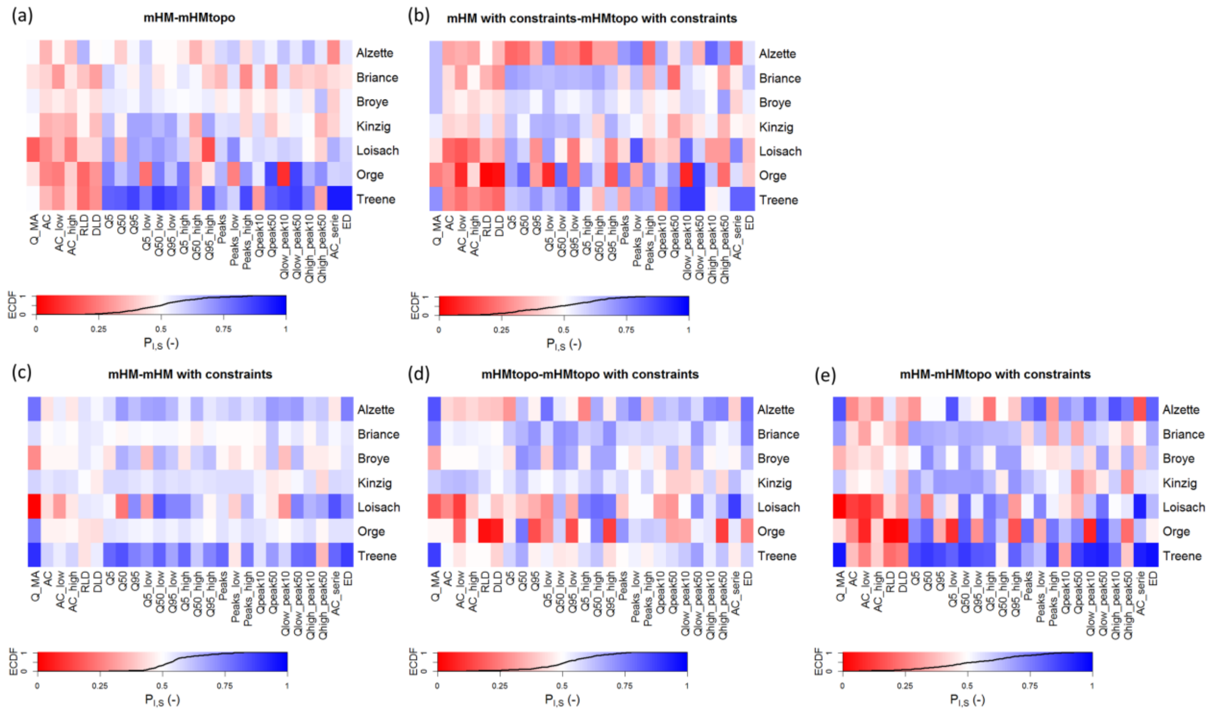


Figure 10. Probabilities of improvements $P_{I,S}$ between (a) mHM and mHMtopo without constraints and (b) with constraints, (c) mHM with and without constraints, (d) mHMtopo with and without constraints and (e) the base case mHM with the constrained mHMtopo case. The colours are linearly related to probability of improvement between 0 (dark red; e.g. probability of mHMtopo outperforming mHM is 0), 0.5 (white; i.e. models are statistically equivalent) and 1 (dark blue; e.g. probability of mHMtopo outperforming mHM is 1). An

empirical cumulative distribution function (ECDF) based on all probabilities of improvement has been added to assess the distribution of these probabilities.

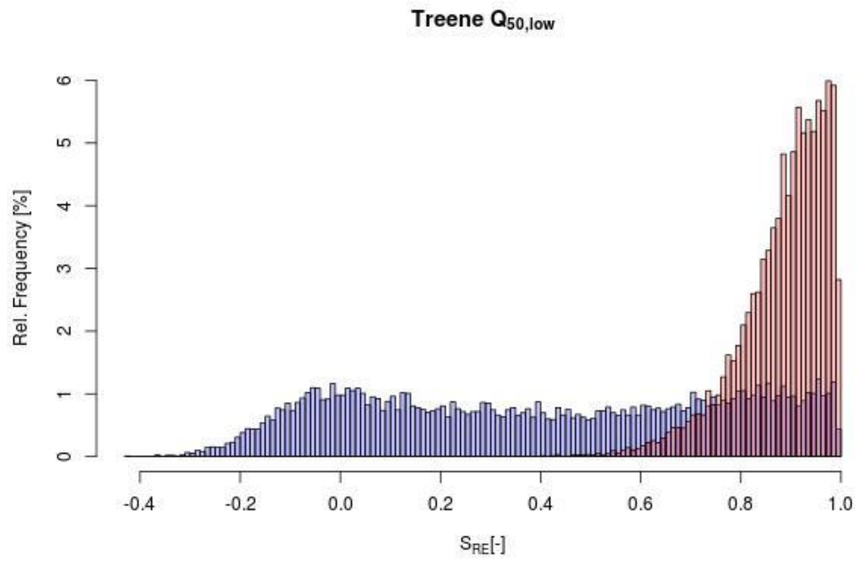


Figure 11. Histograms of the performance distributions for the median of the low flows $Q_{50,low}$ for the Treene catchment on basis of all feasible parameter sets of mHM (blue) and mHMtopo (red). The performance S_{RE} is defined as 1 minus the relative error, leading to an optimal value of 1.

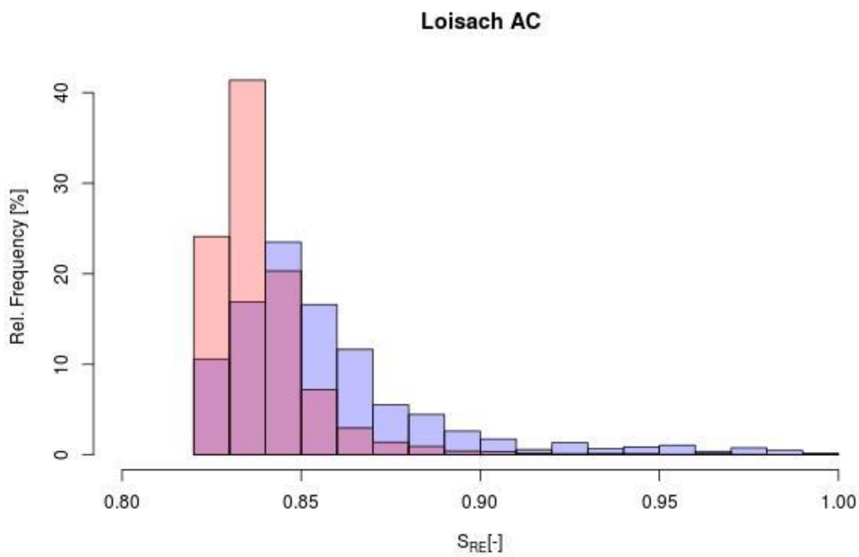


Figure 12. Histograms of the performance distributions for the 1-day autocorrelation of flows for the Loisach catchment on basis of all feasible parameter sets of mHM (blue) and mHMtopo (red). The performance E_{RE} is defined as 1 minus the relative error, leading to an optimal value of 1.

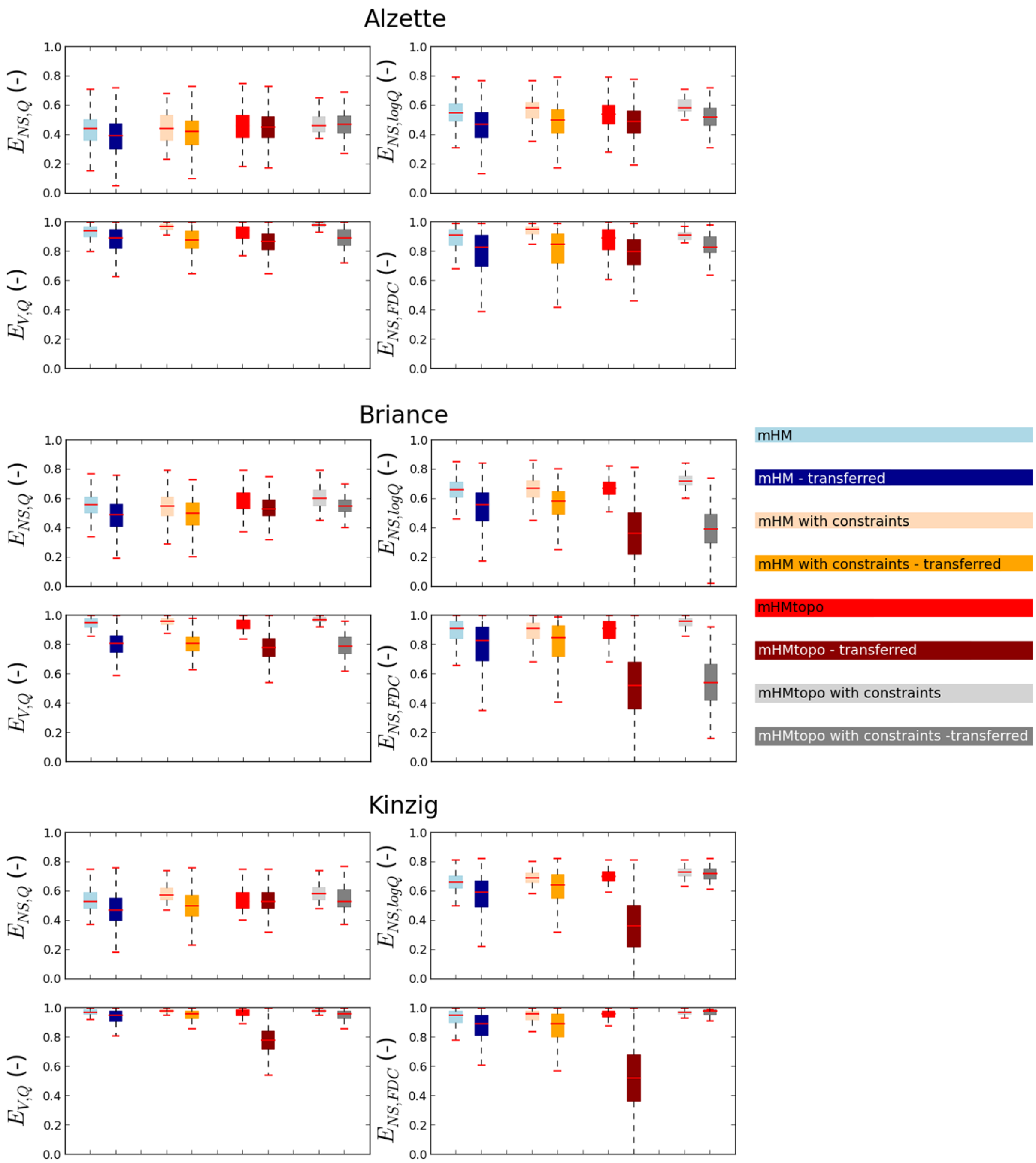


Figure 13. Objective function values of the (a) Alzette, (b) Briance and (c) Kinzig catchments in the validation period for individual calibration (light colours) and when using parameters transferred from the remaining four donor catchments in the multibasin calibration (darker colours).

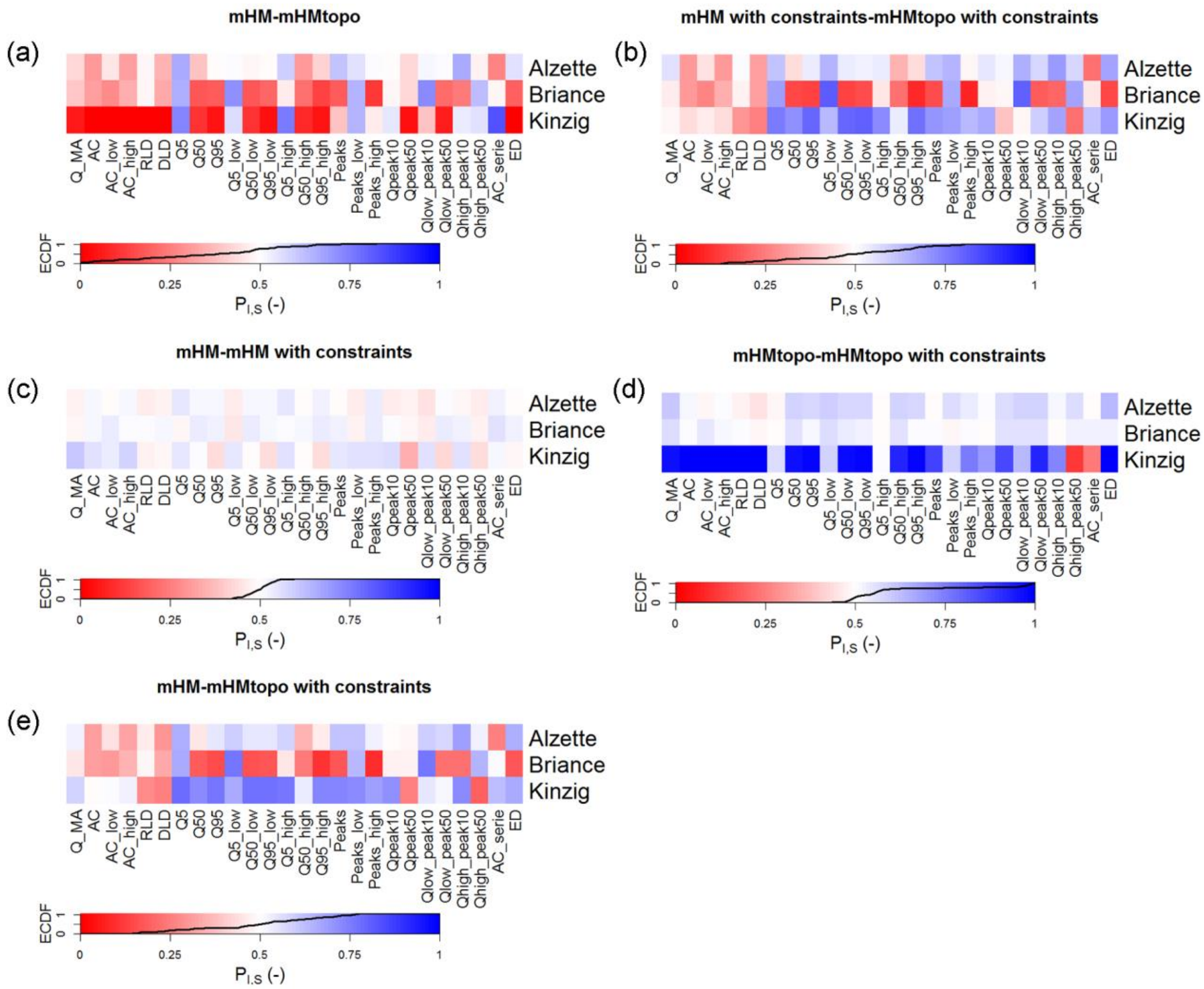


Figure 14. Probabilities of improvements $P_{I,S}$ between (a) mHM and mHMtopo without constraints and (b) with constraints, (c) mHM with and without constraints, (d) mHMtopo with and without constraints and (e) the base case mHM with the constrained mHMtopo case, all after the transfer of global parameters to the three catchments. The colours are linearly related to probability of improvement between 0 (dark red; e.g. probability of mHMtopo outperforming mHM is 0), 0.5 (white; i.e. models are statistically equivalent) and 1 (dark blue; e.g. probability of mHMtopo outperforming mHM is 1). An empirical cumulative distribution function (ECDF) based on all probabilities of improvement has been added to assess the distribution of these probabilities.