TU Delft

Delft University of Technology

Selecting model formulations and parameterizations

Reducing the need for calibration using open data and landscape characteristics

Nijzink, Remko

DOI
10.4233/uuid:8bfae869-25d1-4f64-a24f-90e9514ff7de

Publication date
2018

Document Version
Final published version

Citation (APA)
Nijzink, R. (2018). *Selecting model formulations and parameterizations: Reducing the need for calibration using open data and landscape characteristics*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:8bfae869-25d1-4f64-a24f-90e9514ff7de

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Selecting model formulations and parameterizations

Reducing the need for calibration using open data and landscape characteristics

Selecting model formulations and parameterizations

Reducing the need for calibration using open data and landscape characteristics

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 1 februari 2018 om 15:00 uur

door

Remko Nijzink

civiel ingenieur, Technische Universiteit Delft
geboren te Leidschendam, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. H.H.G. Savenije
copromotor: dr. habil. M. Hrachowitz

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. H.H.G. Savenije,	Technische Universiteit Delft
Dr. habil. M. Hrachowitz,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. ir. N.C. van de Giesen	Technische Universiteit Delft
Prof. dr. ir. S.C. Steele-Dunne	Technische Universiteit Delft
Prof. dr. ir. R. Uijlenhoet	Wageningen University
Assoc. Prof. dr. J. Parajka	TU Vienna
dr. F. Fenicia	Eawag
Prof. dr. ir. M. Bakker	Technische Universiteit Delft, reservelid

Keywords: hydrology, modelling, calibration

Printed by: Ipskamp Printing, Enschede

Front & Back: Designed by Remko Nijzink.

Copyright © 2018 by R.C. Nijzink

ISBN 978-94-028-0930-5

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

Summary

The need for calibration of conceptual hydrological models on river discharge is still large, and the scope of this research is to reduce this need based on new parameter estimation techniques, additional information sources and hydrological understanding.

In a first step, a regularized model and an adjusted regularized model with sub-grid variability based on the landscape, both constrained and unconstrained, were calibrated for seven catchments. Four catchments were also simultaneously calibrated and their feasible parameters transferred to the remaining three receiver catchments. Small improvements were observed by introducing sub-grid variability, whereas the semi-quantitative constraints led to moderate improvements compared to the unconstrained model. Especially low flow statistics improved and suitable prior constraints can aid model transferability.

Subsequently, it was assessed how well the key parameter of root-zone storage capacity could be estimated in three deforested catchments. A recently introduced method based on rainfall and an estimate of transpiration was used to reproduce the temporal evolution of root-zone storage capacities. These values were compared to the values from four hydrological models calibrated for consecutive 2-year windows. Water-balance derived root-zone storage capacities showed a similar signal compared to the calibrated values of the models with a sharp decline in root-zone storage capacity after deforestation, followed by a gradual recovery of 5 to 13 years.

The added value of several combinations of remotely sensed products for parameter estimation was assessed for five different hydrological models in 27 catchments across Europe. A parameter selection process was applied for 1023 possible combinations of ten different data sources, ranging from using 1 to all 10 of these products. High probabilities of improvement, with regard to commonly applied model performance metrics, were obtained when combinations included AMSR-E and ASCAT soil moisture, and GRACE total water storage anomalies. The evaporation products of LSA-SAF and MODIS were less effective for deriving meaningful posterior parameter distributions.

In a last step, a large-scale hydrological model with landscape-derived sub-grid variability was run with 50 random parameterizations for the European continent, with and without semi-quantitative prior parameter constraints. A variable pattern in improvements/deteriorations was observed when evaluated for 397 gauging stations, which shows that the prior parameter constraints were not sufficient to limit the search space effectively.

Concluding, this thesis presents several ways to better optimize different hydrological models without using observed river discharge for varying spatial scales and changing circumstances. In this way, this thesis serves as a stepping stone towards fully predicting in ungauged catchments.

1

Introduction

Hydrological modelling still faces the challenge of ”getting the right answer for the right reasons”, a quote of Kirchner (2006) which has been cited over 300 times, reflecting the ongoing issues in the modelling community. At least, a call has been made by Clark et al. (2016) to unify the theories of our models in a community based effort, but the many unknowns in the system lead to a large number of possible explanations and model hypotheses. Hence, composing a hydrological model can be considered as an art (Savenije, 2009) as the modeler needs to use imagination and creativity to formulate the hydrological theory. This is also complicated by the ”uniqueness of place” (Beven, 2000), overparameterization of our models (Beven, 1993; Gupta et al., 2008) and ongoing reductions in river discharge measurements (Fekete and Vörösmarty, 2002; Hannah et al., 2011; Sivapalan, 2003). These challenges cannot be fully overcome with the current state-of-the-art conceptual hydrological models, and there is still a large need for improvements in the formulation of our models and their parameterizations in order to be able to deal with these challenges.

1.1. Modelling principles

It has been recognized that there is a strong link between climate and vegetation (e.g Donohue et al., 2007; Gentine et al., 2012; Rodríguez-Iturbe, 2000), and also that the landscape is shaped through coevolution by climate and vegetation (e.g. Savenije, 2010; Troch et al., 2015). Especially climatological extremes, such as droughts (e.g. Allen et al., 2010; McDowell et al., 2008; Vose et al., 2016) and floods (e.g. Alila et al., 2009; Gaál et al., 2012) have a large influence on the ecosystem and its equilibrium situation. Taking it one step further, Savenije and Hrachowitz (2017) argued that these linkages lead to self-organization of catchments, as meta-organisms, and patterns will emerge above and below ground, that allow for simple physical equations on the macroscale to sufficiently model the catchment responses. Therefore, simple conceptual hydrological models, which are

parsimonious in the number of parameters to limit the effect of overparameterization, i.e. equifinality (Beven, 1993), can often be applied successfully when the core processes are captured (e.g. Booij, 2005; Refsgaard and Knudsen, 1996). On the other hand, models which are too simple are not able to deal with complex catchment behavior (e.g. Gupta et al., 2012; Hrachowitz et al., 2014; Wagener and Gupta, 2005) and a balance between simplicity and complexity needs to be found. The underlying reason is that hydrological systems are frequently modelled in the realm of organized complexity (Dooge, 1986). As a consequence, they are too complex and "random" to be described in purely mechanistic ways but not complex and "random" enough to be described with purely statistical methods. Many methods have been used to structure the complexity of a catchment, for example by land use (e.g Kite and Kouwen, 1992; Kouwen et al., 1993; Winsemius et al., 2008), soil properties (e.g. McGlynn and McDonnell, 2003; Schmocker-Fackel et al., 2007), geology (Fenicia et al., 2016), the drainage network (e.g. Reggiani et al., 1998) or a combination of several spatial datasets (Flügel, 1995). Several authors, such as Beven and Kirkby (1979), Savenije (2010) or Winter (2001), pointed towards the landscape as a leading concept in defining simple, parsimonious models, that yet reflect the dominant hydrological processes. Eventually, these concepts were translated into models by many researchers, such as Knudsen et al. (1986), Seibert et al. (2003b) and Fenicia et al. (2016), but differences were still found in the way these landscape classes were delineated. As a more specific example, Gharari et al. (2011) translated these concepts into models with landscape based response units for plateaus, hillslopes and wetlands, which were defined by the Height Above the Nearest Drainage Rennó et al. (2008) and terrain slope. Afterwards, Gao et al. (2014a) and Euser et al. (2015) applied similar types of models in the Upper Heihe and the Ourthe catchment in Belgium, showing also for example that model transferability improved.

1.2. Constraining the search space

The hydrological model space, even in the case of rather parsimonious approaches, often remains large, and applying conceptual models without calibration on discharge data is still in its early development. In this respect, the hydrological decade of prediction in ungauged basins (Hrachowitz et al., 2013), even though closed in 2013, still brings challenges forward. Regionalization techniques, as developed by, among others, (e.g. Abdulla and Lettenmaier, 1997; Hundecha and Bárdossy, 2004; Kumar et al., 2013b; Samaniego et al., 2010b), can be used to select feasible model parameters in catchments without discharge data, by using the discharge time series in other gauged catchments nearby. Nevertheless, applications of regionalization techniques have been mostly applied to model formulations that do not incorporate more complex model structures, as this often leads to a larger number of required functional relations for the model parameters. Besides this, the functional relations between model parameters and input data, such as soil data, remain often empirical, and make use of global parameters, which have no clearly defined physical meaning. In addition to the more mathematical regionalization approaches, constraining the parameter space with "soft" data (e.g. Seibert

and McDonnell, 2002; Winsemius et al., 2009) and expert knowledge (e.g. Gharari et al., 2014; Hrachowitz et al., 2014) can help identify suitable model parameters or even model structures. This expert knowledge does however not strongly constrain the parameter space on its own and the availability of more additional data sources, especially remotely sensed information, may provide extra boundaries on this search space for models and their parameterizations. This has been recognized by hydrological modellers and a large number of studies started to appear that deal with this (e.g. Brocca et al., 2010; Rakovec et al., 2016; Werth et al., 2009). However, most of these studies focused on applying a single product for calibration or data assimilation in a model, but more rigorous tests of combinations of products are rare. At most, two or three products are employed to improve hydrological models (e.g. Kunnath-Poovakka et al., 2016; Lopez Lopez et al., 2017), but it can be argued that this is still not enough to efficiently constrain the model search space. A complicating factor arises by the fact that, in most cases, hydrological models reflect a static situation, whereas the system is highly dynamic over time. Often, it is justified to use such as static model formulation as ecosystems tend to organize themselves and establish a static equilibrium situation (e.g. Gao et al., 2014a; Kleidon, 2004), but abrupt disturbances will lead to a system in imbalance. A new equilibrium will need to be established, often enforced by the climatic situation (e.g. Allen et al., 2010; McDowell et al., 2008; Vose et al., 2016) and regenerating vegetation (Li et al., 2007). It may be of considerable value if this time-dynamic behavior of models or model parameters could be incorporated in such a way that the model formulations remain simple and parsimonious. Therefore, it is extremely important to better understand the time-dynamic character of key hydrological parameters, such as the parameter of root zone storage capacity (Milly and Dunne, 1994; Rodríguez-Iturbe et al., 2007), which determines the separation between evaporative fluxes, runoff and moisture storage. An additional challenge in the search for suitable model structures and parameters can be found in assessing to which extent the model represents the right processes. Thorough evaluations of models have been proposed (Andréassian et al., 2009; Clark et al., 2008) and general agreement exists that model evaluation based on the fit of the modelled and observed hydrographs is not sufficient. Therefore, a shift can be observed to more extensive evaluations based on hydrological signatures (e.g. Euser et al., 2013; McMillan et al., 2016; Shafii and Tolson, 2015), such as peak densities and flow duration curves, which allows for a more informative testing of model structures to identify all dominant processes in a catchment. In this way, it was for example shown by Hrachowitz et al. (2014) that more suitable model structures could be selected.

1.3. Science Questions

This thesis focusses on selecting suitable model formulations and parameterizations for a variety of landscapes, climates and vegetation. Following the discussion above, transfer function based modelling has shown a way forward in reducing the need for calibration of models and hence, improved model parameterizations (e.g. Hundecha and Bárdossy, 2004; Kumar et al., 2013a; Samaniego et al., 2010b).

However, model structures remained rather simple for regularized models, whereas previous research suggested that model formulations need to contain more complexity (Hrachowitz et al., 2014; Wagener and Gupta, 2005), structured according to, for example, hydrological response units. Thus, one could wonder whether a combination of a more complex landscape-based model structure as sub-grid variability with a regularization technique is needed to formulate more suitable models:

What is the added value of landscape-variability in a regularized grid-based hydrological model?

In the past, it was shown that model transferability to ungauged sites strongly improved by both model regularization (e.g. Kumar et al., 2013a; Rakovec et al., 2015) and incorporating the landscape in a hydrological model (Gao et al., 2014a). This points in both cases towards an improved hydrological representation in the model, only for different reasons (i.e. model parametrization versus model structure). Hence, it can be expected that combining model regularization with landscape-derived sub-variability should lead to an increased ability of the model to reproduce hydrological responses in catchments not used for calibration (e.g. ungauged catchments). In other words, the hypothesis is tested whether a more complex model structure in a regularized model leads to a more suitable model formulation.

An important part of the model formulation areformed by the model parameters, from which especially the root zone storage capacity plays a key role (Milly and Dunne, 1994; Rodríguez-Iturbe et al., 2007). However, changing conditions may lead to increased model uncertainties (Fenicia et al., 2009; Gharari et al., 2013; Wagener and Gupta, 2005), especially when key parameters cannot be considered constant in time. In this regard, it is of major importance to understand the time-dynamic character of root zone storage capacities:

Can we identify the time-variant behavior of a key parameter of ecosystems, the root zone storage capacity? How does it evolve over time? What is the added value of a time-dynamic formulation of root-zone storage capacity?

By addressing this question, suitable model parameterizations should be found for catchments under change, thus enhancing hydrological understanding about the time-dynamic nature of vegetation dependent parameters. Moreover, a time-dynamic representation of root zone storage capacity as part of the model formulation could be considered as a first step towards more time-variability of parameters in conceptual models, which is strongly needed (Savenije and Hrachowitz, 2017; Wagener and Gupta, 2005). Besides this, it has been recognized that the root-zone storage capacity is a parameter that needs a careful prior estimation in order to arrive at a suitable model formulation (Milly and Dunne, 1994; Rodríguez-Iturbe et al., 2007). Thus, when relatively simple techniques to estimate root-zone storage capacities as formulated by Gao et al. (2014a), de Boer-Euser et al. (2016) and Wang-Erlandsson et al. (2016), can suffice in estimating this parameter and its time-dynamic nature, the need for model calibration on discharge data is again

reduced as model formulations improved.

Besides simple estimation techniques and model regularization for parameter and/or model selection, the large number of new (remotely sensed) data products should provide opportunities for improved model formulations and parameterizations as well. So far, only a limited number of two or three products was used for parameter estimations (e.g. Kunnath-Poovakka et al., 2016; Lopez Lopez et al., 2017; Tian et al., 2017), whereas the combined use of multiple products may lead to more suitable model parameterizations:

What is the added value of different combinations of additional sources of information, e.g. remotely sensed data, in parameter selection procedures?

This question will therefore lead to new insights in how model formulations can be improved by a simultaneous use of different products, relating to different states or fluxes in the models, instead of calibrating on solely river discharge. Besides, assessing combinations of products, instead of assessing the products one by one, will lead to more understanding on the combined strengths of products. Similar as for the case of model regularization with additional sub-grid variability, more information sources in the process of selecting feasible model parameters and/or structures should lead to a more robust model formulation, i.e. a model formulation which can be more easily applied without calibrating on river discharge.

However, for most remotely sensed products, a model needs to be run first to assess its performance, e.g. to assess the correspondence between modelled evaporation and product evaporation, often also referred to as process constraints (e.g. Bulygina et al., 2009; Gharari et al., 2014). These process constraints are generally not possible for large-scale models due to computational demands. Therefore, in a last step, it was tested whether more qualitative expert knowledge prior-parameter constraints, in contrast with a more formal mathematical regularization, suffice already in constraining the parameter search space:

What is the added value of prior parameter constraints for constraining a large-scale landscape-driven model?

It was shown before that the parameter space can be reduced by prior parameter constraints (e.g. Gharari et al., 2014; Hrachowitz et al., 2014), but this was only tested for small scale catchments. With this last step, the power of a relatively simple approach is more extensively explored, in addition to the more formal mathematical model regularization and constraining addressed in the forgoing research questions. These prior parameter constraints are more based on qualitative hydrological understanding, which makes the technique less complicated and model dependent compared to a full mathematical model regularization. For a formal model regularization, functional empirical relations need to be defined first, and the global parameters can only be found after calibration of the model. Besides, the risk remains that the functional relations may not be valid, whereas the more qualitative constraints can be more easily supported with hydrological theory. Therefore,

this last research question should lead to conclusions to which extent simple prior parameter constraints could reduce the parameter search space, whereas at the same time the hydrological understanding is maintained.

Thus, by addressing the research questions above, this thesis explores different parts of the spectrum to improve model formulations and parameterizations, from the different scales of the catchment to the continent, from qualitative, informal prior constraints to more quantitative, formal model regularization, and specific key parameters to full model formulations. The goal of this thesis is therefore to serve as a stepping stone towards conceptual models based on a sound hydrological theory, build upon a rigid understanding of the added value of new data sources and their functional relations with model formulations.

2

Topography-controlled sub-grid heterogeneity

This chapter assesses the suitability of a more complex model formulation based on the landscape (the topography) in a regularized model set-up. Both methods, landscape-driven modelling and model regularization have shown a high potential in order to apply models in catchments which are not used for calibration. Thus, the hypothesis is tested whether the incorporation of additional sub-grid variability on the basis of topography-derived response units leads to more suitable model formulations.

This chapter is based on:
Nijzink, R.C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M.: *The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models*, Hydrol. Earth Syst. Sci., 20, 1151-1176, doi:10.5194/hess-20-1151-2016, 2016.

2.1. 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 et al., 1986; Rodríguez-Iturbe et al., 2006; Tromp-van Meerveld and McDonnell, 2006), climatic variables, i.e. the forcing of models such as precipitation and evaporation, typically exhibit considerable spatial variability (e.g. Hrachowitz and Weiler, 2011; Obled et al., 1994; Singh, 1997; Winsemius et al., 2008). 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 for accommodating spatial heterogeneity. In other words, this heterogeneity can in reality result in a variety of parallel processes, characterized by considerably different timescales being simultaneously active. Therefore, lumped representations of catchments frequently fail to adequately represent the dominant features of the observed hydrological response at 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. Booij, 2005; Refsgaard and Knudsen, 1996). 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), Reicher 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. Euser et al., 2015; Fenicia et al., 2008; Winsemius et al., 2008).

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 of the order of several dozen square kilometres in most applications (e.g. Booij, 2005; Lindström et al., 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; that is, 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 already became apparent in several studies, e.g. Merz and Bárdossy (1998), Zehe et al. (2001), Seibert et al. (2003a), and 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 reflected by distribution functions or constitutive relationships. For example, distribution functions for maximum unsaturated storage capacities, such as 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 the accommodation of the spatial heterogeneity within the elementary watershed to varying degrees (e.g. Mou et al., 2008; Reggiani and Rientjes, 2005; Tian et al., 2006; Vannametee et al., 2012; Zhang and Savenije, 2005; Zhang et al., 2006), but the search for generally applicable adequate closure relations is still ongoing.

The division of the catchment into several functional units (e.g. Flügel, 1995; Kite and Kouwen, 1992; Knudsen et al., 1986; Kouwen et al., 1993; Leavesley and Stannard, 1990; Reggiani et al., 1998; Schmocker-Fackel et al., 2007; Seibert et al., 2003b; Uhlenbrook et al., 2004; Winter, 2001; Zehe et al., 2014) may offer a way to address these conceptual shortcomings. In spite of the fact that in many cases insufficient data for a detailed delineation of response units are available, it has been recognized by several authors (e.g. Beven and Kirkby, 1979; Knudsen et al., 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 (Gharari et al., 2011; Nobre et al., 2011; Rennó et al., 2008). 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., 2014a; Gharari

Table 2.1: Overview of the catchments.

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

et al., 2014). They further enhance model transferability without the need for empirical transfer functions 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 (Euser et al., 2013; Gupta et al., 2012; 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 and Bárdossy, 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 transferabil-

ity (e.g. Götzinger and Bárdossy, 2007; Kumar et al., 2013a; Samaniego et al., 2010b). 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 (Gao et al., 2014a; Gharari et al., 2014; Hrachowitz et al., 2014; Hughes, 2013; Kapangaziwiri et al., 2012; Seibert and McDonnell, 2013).

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; Kumar et al., 2013b; Samaniego et al., 2010b). 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 allows the identification of unfeasible parameter sets and thereby reduces model uncertainty, and that 3) the combined use of response units and model constraints improves the spatial transferability of the model.

2.2. Methodology

2.2.1. 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 Fig. 2.1. 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 2.1 summarizes the catchment characteristics.

The northern 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

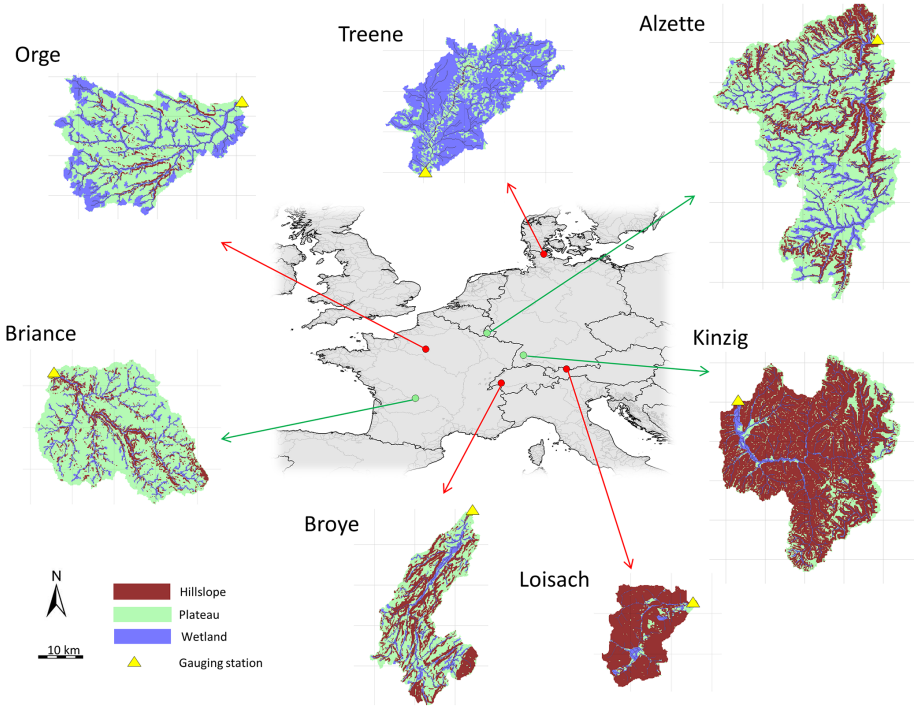

Figure 2.1: The locations of the seven study catchments and their respective landscape classes according to HAND and local slope. Catchments represented by red and green symbols in the context map indicate donor and receiver catchments, respectively, for the transferability analysis. Displayed grids correspond to the modelling grids used in mHM (topo).

Table 2.2: Overview of the used data.

Data type	Product	Source	Reference
Soil	HWSD	http://webarchive.iiasa.ac.at/Research/LUC/	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/	The Global Runoff Data Centre, 56002 Koblenz, Germany
Precipitation	E-OBS	http://eca.knmi.nl/download/ensembles/ensembles.php	Haylock et al. (2008)
Land cover	Globcover	http://due.esrin.esa.int/page_globcover.php	Arino et al. (2010)

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 and Samani, 1985). A summary of the data sources is given in Table 2.2.

2.2.2. Models

mesoscale Hydrological Model (mHM)

mHM is a distributed, process-based model that uses the cell-wise model architecture shown in Fig. 2.2 in each grid cell of the modelling domain (Kumar et al., 2013b; Samaniego et al., 2010b). 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 infil-

trating 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 baseflow component of the runoff. The model has been successfully applied across Germany, Europe and North America (Kumar et al., 2010, 2013a,b; Livneh et al., 2015; Rakovec et al., 2015; Samaniego et al., 2010a,b, 2013; Thober et al., 2015).

Topography-driven mHM (mHMtopo)

To test the value of topography variability-induced process heterogeneity in a distributed model, the concepts of FLEXtopo (Gharari et al., 2011; Savenije, 2010) 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 as dominant in many central European regions: plateaus, hillslopes and wetlands (Savenije, 2010). The landscape classes were defined by the Height Above the Nearest Drainage (Rennó et al., 2008, HAND) and local slope. Following Gharari et al. (2011), areas with a low slope ($< 11\,\%$) and high HAND ($> 5\,m$) were defined as plateaus, areas with high slope ($> 11\,\%$) as hillslopes and areas with low slope and low HAND ($< 5\,m$) 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 study 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 briefly summarized in the following. 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 Fig. 2.3. 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 more strongly with the groundwater. Thus, capillary rise (Cr in Fig. 2.3) 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 thesis, the two models will be referred to as mHM and mHMtopo to distinguish between the original mHM and the topography-guided set-up, respectively.

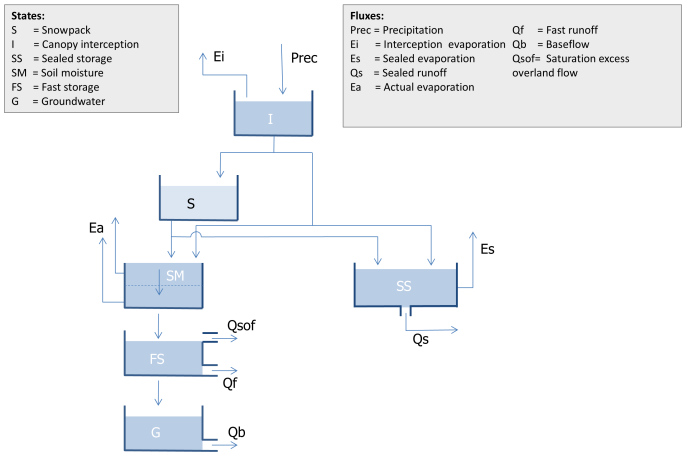

Figure 2.2: The original mHM model structure. The effective precipitation is determined by an interception (I) and a snow routine (S). Afterwards, the effective precipitation enters a soil moisture reservoir (SM) or is directly routed to a fast reservoir that accounts for sealed areas (SS). The water in the soil moisture reservoir either transpires or percolates further down to a fast runoff reservoir (FS), i.e. shallow subsurface flow. Eventually, the baseflow component of the runoff is obtained from a slow groundwater reservoir (G).

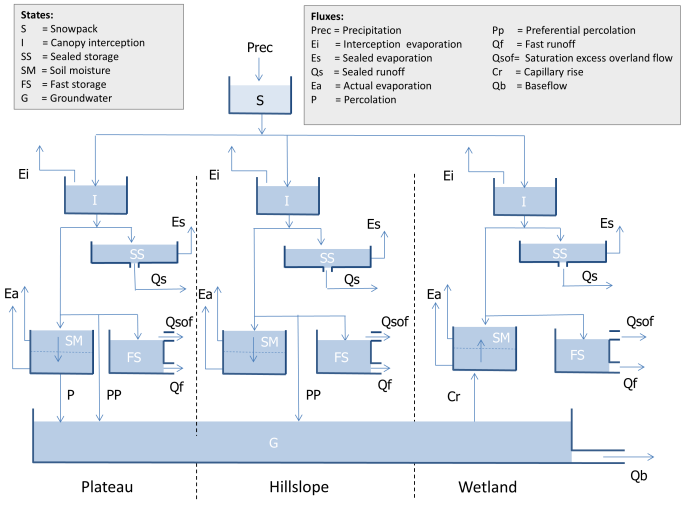

Figure 2.3: The mHMtopo model structure with different configurations of states and fluxes for landscape classes plateau, hillslope, and wetland, which are based on topography. First, a shared snow module (S) divides the effective precipitation over the landscape classes. The three classes all have an interception module (I), a fast reservoir accounting for sealed areas (SS), a soil moisture routine (SM) and a 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 baseflow is determined by a shared groundwater reservoir.

2.2.3. Model 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.

Multiscale parameter regionalization

The multiscale parameter regionalization (MPR) is the key feature of mHM (Kumar et al., 2013b; Samaniego et al., 2010b). 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 Fig. 2.4, 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 2.5 depicts the application of the MPR technique to 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 × 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. (2010b) and Kumar et al. (2013b) for details about the transfer functions and upscaling methods.

The MPR technique 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 Fig. 2.5.

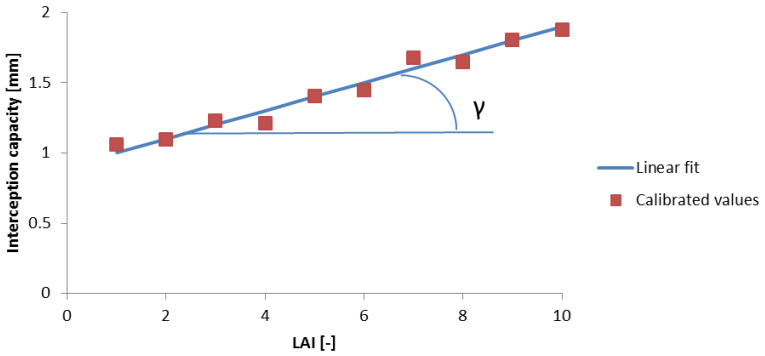

Figure 2.4: Function relationship between leaf area index (LAI) and the hydrologic parameter interception capacity ($I_{0,\max}$) defined by the global parameter γ, based on fictional data for illustration.

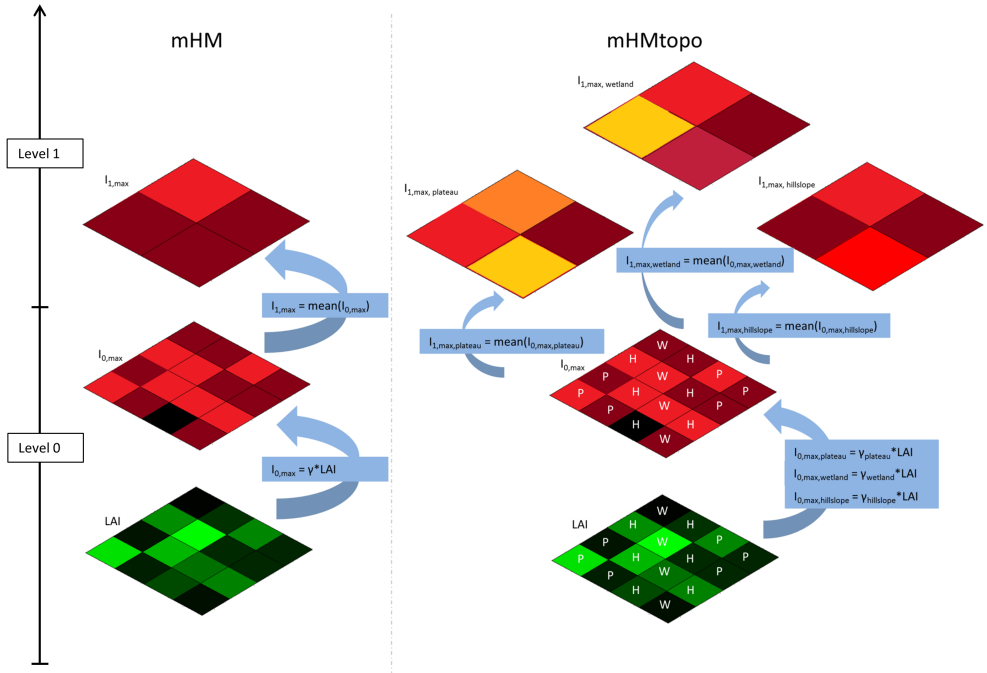

Figure 2.5: Schematic representation of the original MPR (left) and the adjusted MPR (right) for the maximum interception capacity ($I_{1,\max}$). On the input level 0, the leaf area index (LAI) is linked through the global, generally valid, parameter γ with $I_{0,\max}$. In a last step, the mean is 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 ($I_{1,\max,\text{plateau}}$, $I_{1,\max,\text{hillslope}}$ and $I_{1,\max,\text{wetland}}$).

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 Fig. 2.5).

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 a 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 behaviours.

$$C_{\mathrm{Rmin}} < C_{\mathrm{Rmodelled}} < C_{\mathrm{Rmax}} \tag{2.1}$$

$$C_{\mathrm{Rhigh,min}} < C_{\mathrm{Rhigh,modelled}} < C_{\mathrm{Rhigh,max}} \tag{2.2}$$

$$C_{\mathrm{Rlow,min}} < C_{\mathrm{Rlow,modelled}} < C_{\mathrm{Rlow,max}} \tag{2.3}$$

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 et al. (1975) and the concept of a topographic wetness index (Beven and Kirkby, 1979), but also results from experimental studies, e.g. Seibert et al. (2003a). Thus, two additional constraints were used for mHMtopo:

$$S_{\mathrm{M,plateau}} > S_{\mathrm{M,wetland}}, \tag{2.4}$$

$$S_{\mathrm{M,hillslope}} > S_{\mathrm{M,wetland}}. \tag{2.5}$$

2.2.4. Experiment set-up

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 ($E_{\mathrm{NS,Q}}$), the Nash–Sutcliffe efficiency of the logarithm of flow ($E_{\mathrm{NS,logQ}}$), the volume error of flow ($E_{\mathrm{V,Q}}$) and the Nash–Sutcliffe efficiency of the logarithm of the flow duration curve ($E_{\mathrm{NS,FDC}}$). 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 the Dynamically Dimensioned Search algorithm (Tolson and Shoemaker, 2007, DDS) or the Shuffled Complex Evolution method (Duan et al., 1992, SCE), 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. Andréassian et al., 2012; Beven, 2006; Kirchner, 2006), 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 four-dimensional Pareto front formed by $E_{\mathrm{NS,Q}}$, $E_{\mathrm{NS,logQ}}$, $E_{\mathrm{V,Q}}$ and $E_{\mathrm{NS,FDC}}$ were considered to be feasible solutions and used for post-calibration evaluation. Considering all feasible solutions to be equally likely, the model uncertainty intervals are represented by the envelope of all feasible solutions.

Post-calibration model evaluation

The models' skill in reproducing 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 evaluation of 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) will adequately reflect 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. Euser et al., 2013; Sawicz et al., 2011) and are summarized in Table 2.3.

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

Table 2.3: Overview of the used signatures.

Signature	Description	Reference
S_{QMA}	Mean annual runoff	
S_{AC}	One-day autocorrelation coefficient	Montanari and Toth (2007)
$S_{\text{AC,low}}$	One-day autocorrelation low flow period	Euser et al. (2013)
$S_{\text{AC,high}}$	One-day autocorrelation high flow period	Euser et al. (2013)
S_{RLD}	Rising limb density	Shamir et al. (2005)
S_{DLD}	Declining limb density	Shamir et al. (2005)
S_{Q_5}	Flow exceeded in 5 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{50}}$	Flow exceeded in 50 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{95}}$	Flow exceeded in 95 % of the time	Jothityangkoon et al. (2001)
$S_{Q_5,\text{low}}$	Flow exceeded in 5 % of the low flow time	Yilmaz et al. (2008)
$S_{Q_{50},\text{low}}$	Flow exceeded in 50 % of the low flow time	Yilmaz et al. (2008)
$S_{Q_{95},\text{low}}$	Flow exceeded in 95 % of the low flow time	Yilmaz et al. (2008)
$S_{Q_5,\text{high}}$	Flow exceeded in 5 % of the high flow time	Yilmaz et al. (2008)
$S_{Q_{50},\text{high}}$	Flow exceeded in 50 % of the high flow time	Yilmaz et al. (2008)
$S_{Q_{95},\text{high}}$	Flow exceeded in 95 % of the high flow time	Yilmaz et al. (2008)
S_{Peaks}	Peak distribution	Euser et al. (2013)
$S_{\text{Peaks,low}}$	Peak distribution low flow period	Euser et al. (2013)
$S_{\text{Peaks,high}}$	Peak distribution high flow period	Euser et al. (2013)
$S_{Q_{\text{peak},10}}$	Flow exceeded in 10 % of the peaks	
$S_{Q_{\text{peak},50}}$	Flow exceeded in 50 % of the peaks	
$S_{Q_{\text{low,peak},10}}$	Flow exceeded in 10 % of the low flow peaks	
$S_{Q_{\text{low,peak},50}}$	Flow exceeded in 10 % of the low flow peaks	
$S_{Q_{\text{high,peak},10}}$	Flow exceeded in 10 % of the high flow peaks	
$S_{Q_{\text{high,peak},50}}$	Flow exceeded in 50 % of the high flow peaks	
$S_{\text{AC,serie}}$	Autocorrelation series (200-day lag time)	Montanari and Toth (2007)

signatures under consideration (e.g. Schoups et al., 2005):

$$D_{\mathrm{E}} = \sqrt{(1 - S_{\mathrm{NS},1})^2 + (1 - S_{\mathrm{NS},2})^2} \ldots + (1 - S_{\mathrm{NS},n})^2 + S_{\mathrm{RE},1}^2 + S_{\mathrm{RE},2}^2 \ldots + S_{\mathrm{RE},m}^2, \quad (2.6)$$

with $S_{\mathrm{NS},i}$ the performance metric of n multi-valued signatures, and $S_{\mathrm{RE,j}}$ 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 will outperform another for a specific signature was computed to objectively quantify the differences between these distributions and to allow an overall assessment of which of the tested models exhibits a higher ability 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 $P_{\mathrm{I,S}}$ can be readily obtained from:

$$P_{\mathrm{I,S}} = P(S_1 > S_2) = \sum_{i=1}^{n} P(S_1 > S_2 | S_1 = r_i) P(S_1 = r_i), \quad (2.7)$$

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; that is, no preference for a model can be identified. In contrast, for $P_{\mathrm{I,S}} > 0.5$ it is more likely that model 1 outperforms model 2 with respect to the signature under consideration, and vice versa for $P_{\mathrm{I,S}} < 0.5$.

In an additional analysis, the ranked probability score S_{RP} (Wilks, 2011) was calculated as a measure of the magnitude of improvement. For details, please see the description and the Supplement of Nijzink et al. (2016b).

Comparison of model transferability

The mHM hydrologic model has previously been shown to have a considerable ability to reproduce the hydrograph when transferring global parameters from calibration catchments to other regions without further recalibration (Kumar et al., 2013a,b; Rakovec et al., 2015; Samaniego et al., 2010a,b). Therefore, it was tested whether the addition of topography-driven sub-grid process heterogeneity and the use of prior constraints in mHM have the 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 $E_{\mathrm{NS,Q}}$, $E_{\mathrm{NS,logQ}}$, $E_{\mathrm{V,Q}}$

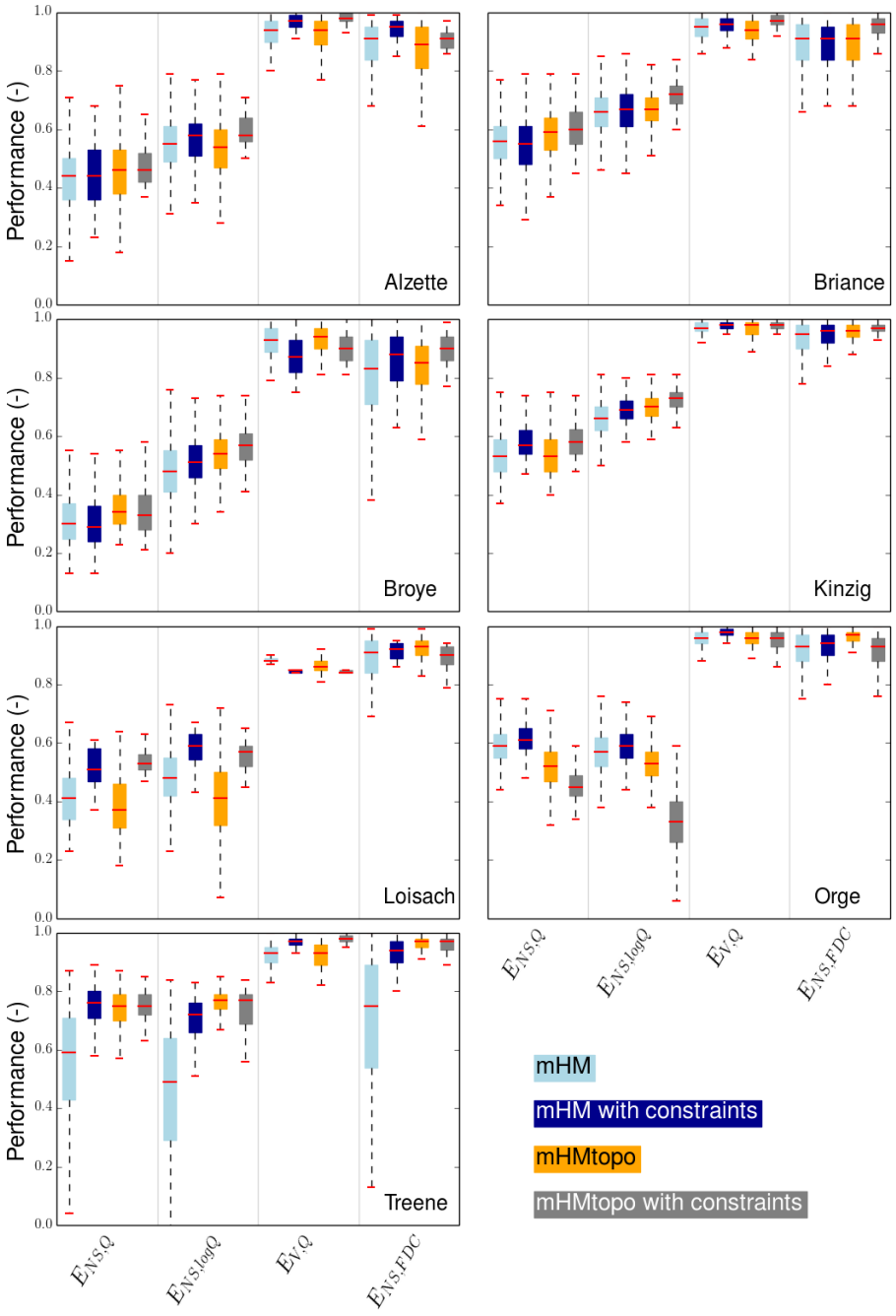

Figure 2.6: Nash–Sutcliffe efficiency ($E_{\mathrm{NS,Q}}$), log Nash–Sutcliffe efficiency ($E_{\mathrm{NS,logQ}}$), volume error ($E_{\mathrm{V,Q}}$) and log Nash–Sutcliffe efficiency of the flow duration curve ($E_{\mathrm{NS,FDC}}$) for the seven catchments in the validation periods. The optimal value for all four criteria is 1, whereas 0 is regarded as having a low performance. The boxplots are formed by the Pareto space spanned by the four objective functions.

and $E_{\mathrm{NS,FDC}}$ were now averaged over the catchments. This led to global parameters that account for the performance in 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 EVE high-performance compute cluster 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 processors. Nevertheless, the used calibration strategy needed run times of about 2 weeks per catchment on multiple EVE cores, depending on catchment sizes and lengths of time series.

2.3. Results and discussion

2.3.1. 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 Fig. S2 in the Supplement of Nijzink et al., 2016b). 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 Fig. 2.6, where mHM with constraints (dark blue) compared with mHM (light blue) 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 Fig. 2.6 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 constrained mHMtopo exhibited an average increase of 8 and 11 %, respectively, for the Euclidian distance D_{E} compared to the original mHM.

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, does not necessarily lead to model improvements. However, these findings are not in line with some other previous work (e.g. Euser et al., 2015; Gao et al., 2014a; Gharari et al., 2014), which 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 few contrasts in the landscape.

The Treene catchment benefits most from the addition of topography-guided sub-grid heterogeneity (Fig. 2.6). 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 is an important process in this relatively flat and humid catchment, dominated by peaty soils. These findings also correspond to conclusions by Schmalz et al. (2008) and Schmalz and Fohrer (2009), 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 2.7 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 reaches towards 0 mm d^{-1} in July, whereas mHMtopo still maintains a flow.

In contrast, it can be noticed from Fig. 2.6 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 et al., 2012). Therefore, it is more markedly influenced by anthropogenic disturbances, which are likely not adequately reflected in either mHM or mHMtopo. This results in a situation where the more parsimonious mHM is likely to provide a representation of process dynamics that more closely reflects those 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 (Fig. 2.8). 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 behaviour.

Effect of constraints

The applied prior process and parameter constraints, in agreement with Gharari et al. (2014) and Hrachowitz et al. (2014), helped to increase model performance (Fig. 2.6) and to reduce model uncertainty (Figs. 2.7, 2.8, 2.9) by identifying and discarding a considerable number 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 deceptive adequate calibration results can be and how a successful identification can result in reduced

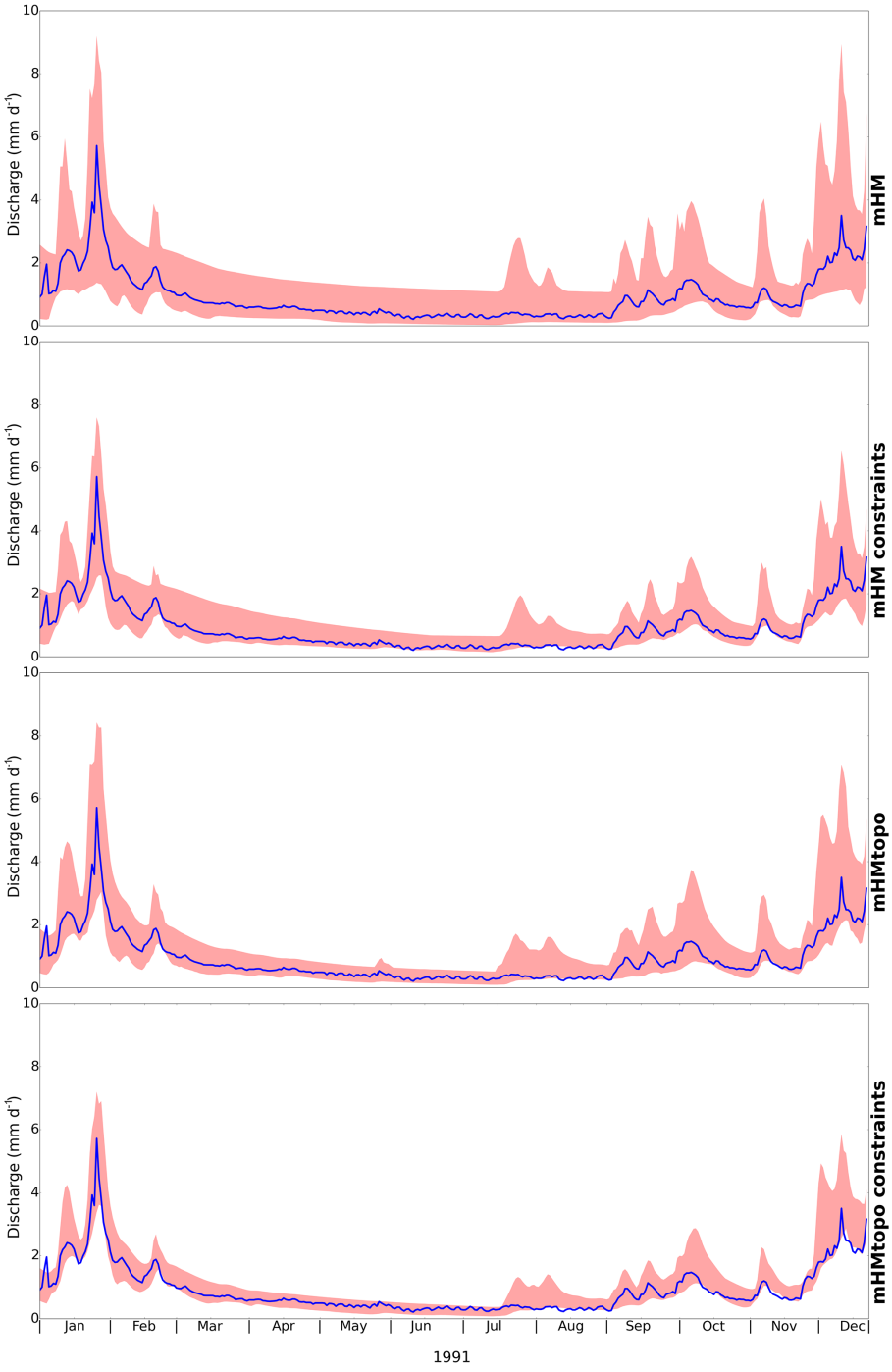

Figure 2.7: Hydrographs for the Treene catchment, with, respectively, the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.

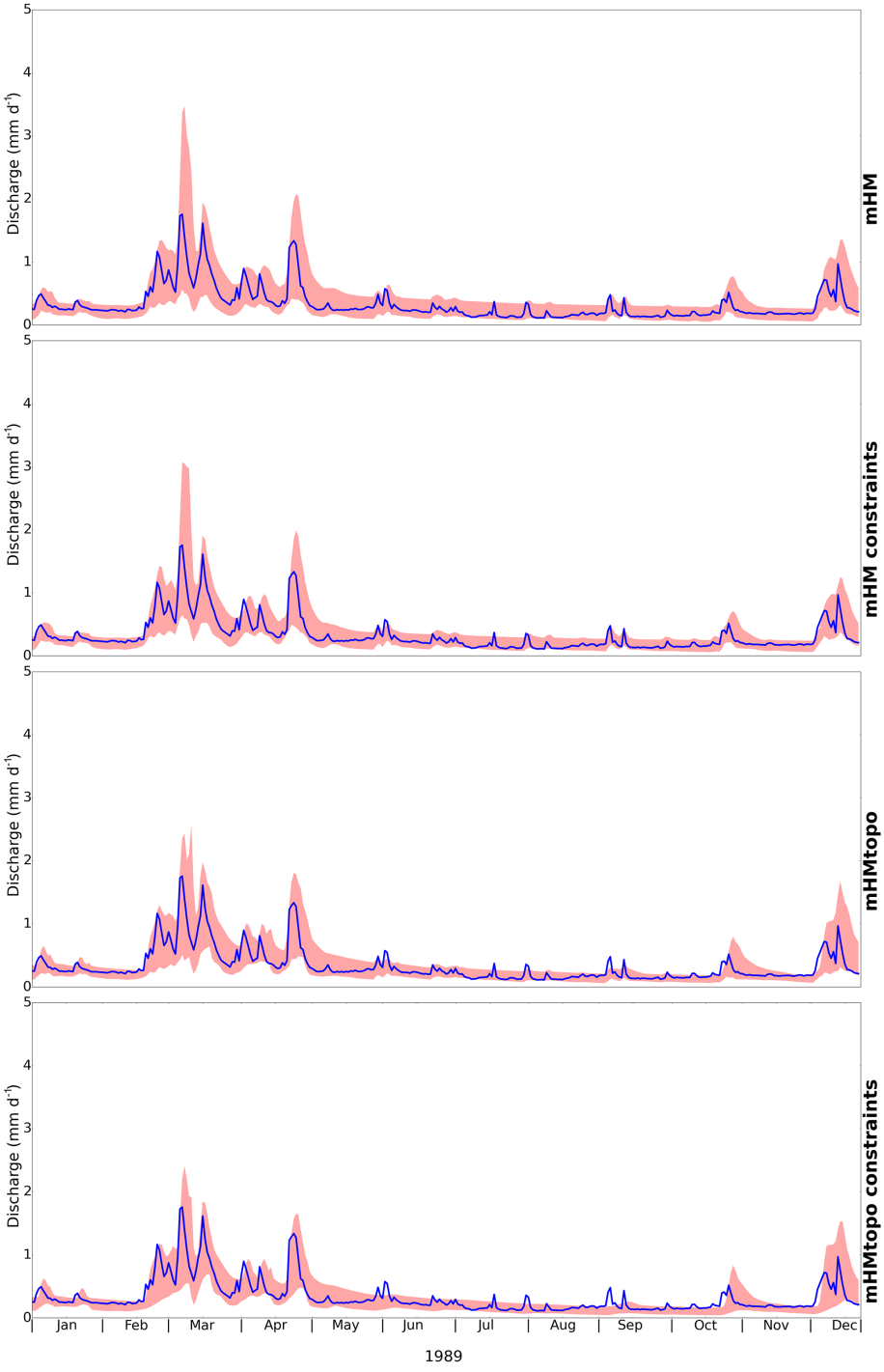

Figure 2.8: Hydrographs for the Orge catchment, with, respectively, the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.

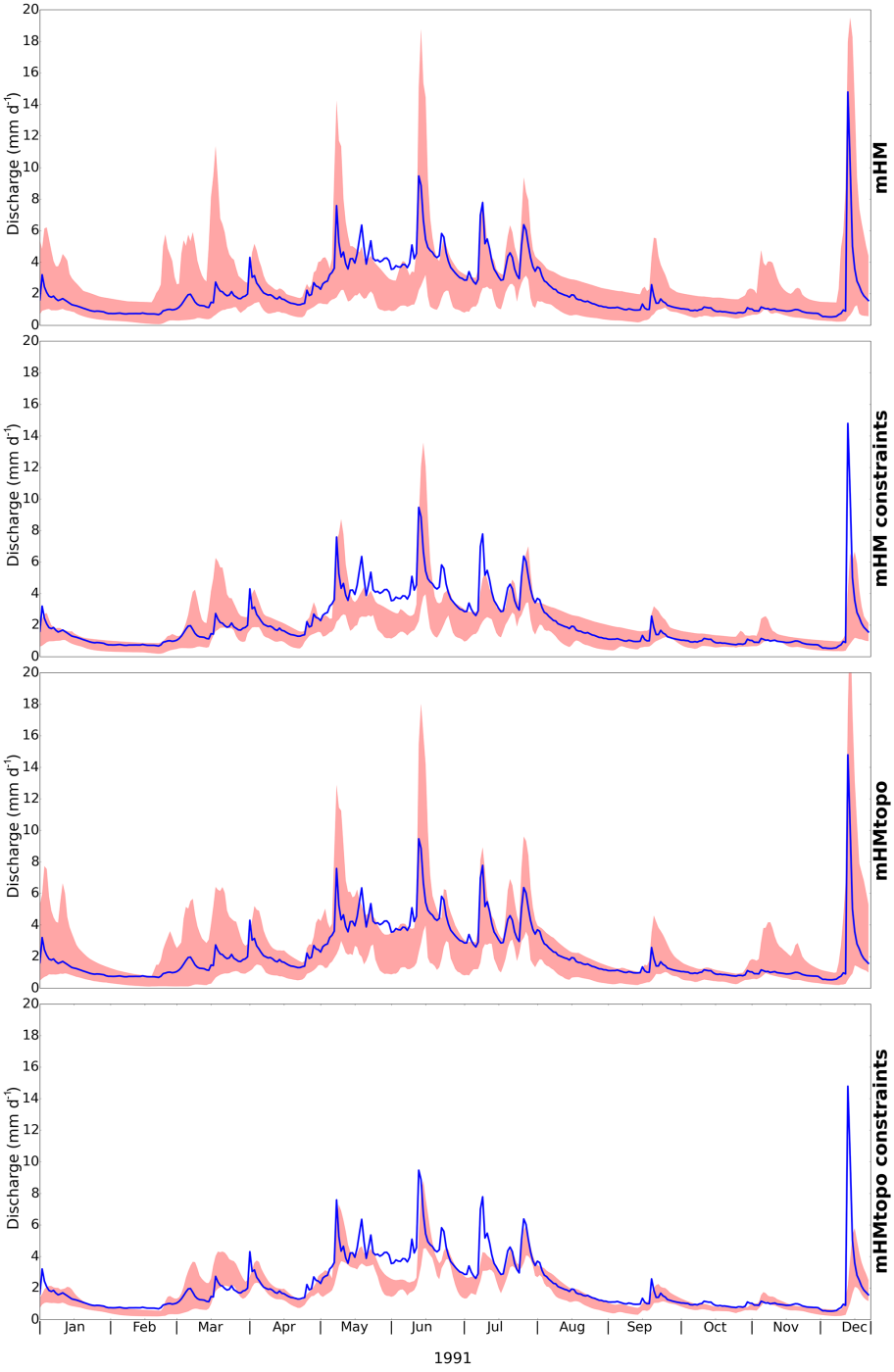

Figure 2.9: Hydrographs for the Loisach catchment, with, respectively, the hydrographs for mHM, mHM with constraints, mHMtopo and mHMtopo with constraints. The red shaded areas represent the envelope spanned by all feasible solutions, whereas the blue line corresponds to observed values.

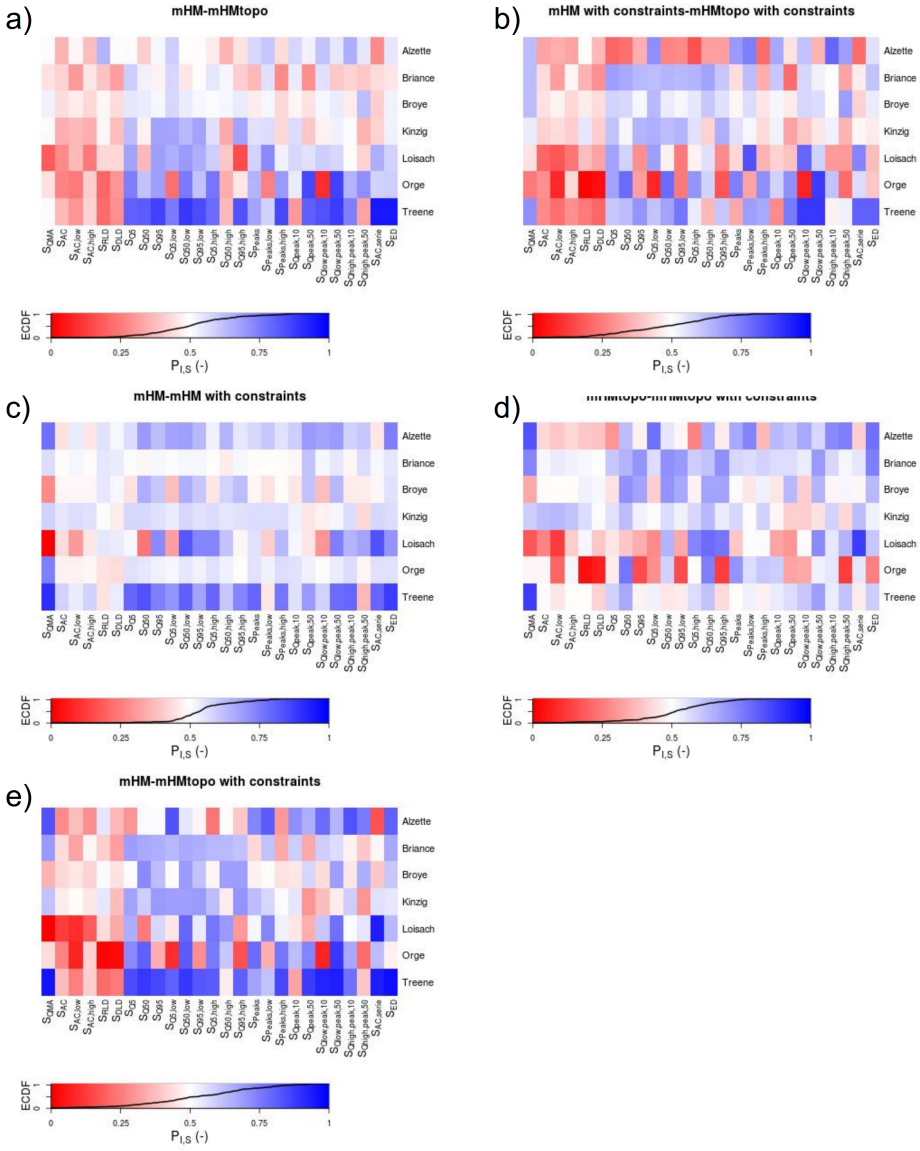

Figure 2.10: Probabilities of improvements $P_{I,S}$ between **(a)** mHM and mHMtopo without constraints and **(b)** with constraints, **(c)** mHM with and without constraints, **(d)** mHMtopo with and without constraints and **(e)** the base case mHM with the constrained mHMtopo case. The colours are linearly related to the probability of improvement between 0 (dark red; e.g. the probability of mHMtopo outperforming mHM is 0), 0.5 (white; i.e. models are statistically equivalent) and 1 (dark blue; e.g. the 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.

predictive uncertainty. It must be noted that the effect is strong in the chosen calibration strategy, as a large set also containing 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., 2014). 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 streamflow 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 led to a considerable improvement in the model's internal dynamics. This is supported by visual inspection of the hydrographs (Fig. 2.9): 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.

Effect of constraints and sub-grid heterogeneity

Comparing the base case of the unconstrained mHM with the most complex constrained mHMtopo (Fig. 2.6) 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 shows neither a decrease in bandwidth nor a 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.

2.3.2. 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 2.3). This comparison is based on the probabilities of improvement $P_{I,S}$ (Fig. 2.10 and Eq. 2.7), but similar results were obtained with the ranked probability score S_{RP}. The results of S_{RP} can be found in the supplementary material of Nijzink et al. (2016b) in Figures S3 and S4. Overall, the introduction of constraints to mHM led 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 %.

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 Fig.2.10a, illustrates that the consideration of sub-grid process heterogeneity in mHMtopo leads to a slight overall improvement compared to mHM. However, the effect on individual signatures is diverse, with some signatures captured to a better degree, while others could be reproduced less well.

Figure 2.10a 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 (Fig. 2.6), now exhibit relatively high probabilities of improvement with respect to the signatures when sub-grid heterogeneity is added (Fig. 2.10a). 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 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 Fig. 2.1), adding several processes that only become dominant in the dry periods. 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.

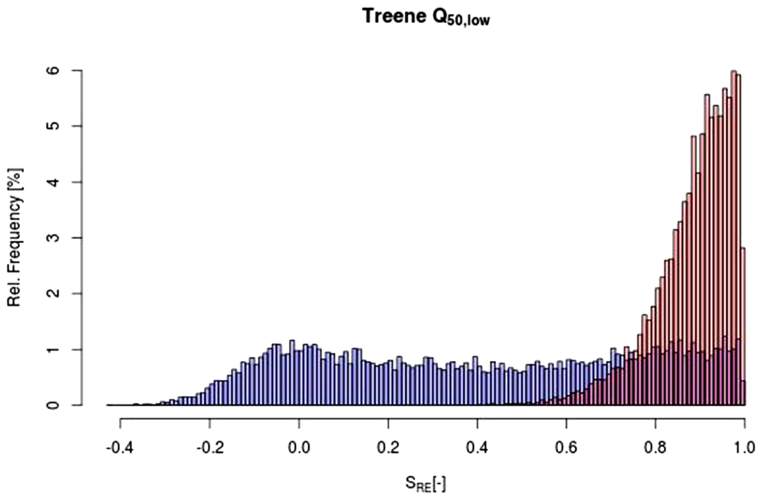

Figure 2.11: Histograms of the performance distributions for the median of the low flows $Q_{50,\text{low}}$ for the Treene catchment on the basis of all feasible parameter sets of mHM (blue) and mHMtopo (red). The performance S_{RE} is defined as 1 minus the relative error, leading to an optimal value of 1.

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 Velázquez 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 ($S_{Q_{50},\text{low}}$, $S_{Q_{95},\text{low}}$, $S_{Q_{5},\text{low}}$) and peaks during low flows ($S_{Q_{\text{peak},10}}$, $S_{Q_{\text{low,peak},50}}$), as can be seen in Fig. 2.10a. 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 1 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, $S_{Q_{50},\text{summer}}$ in the Treene catchment (Fig. 2.11). 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 (Fig. 2.10a, 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 spiky peaks in the Orge catchment, this suggests again that the simpler 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, for storing 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, Fig. 2.12 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 et al. (2008), who all found that added complexity can but does not necessarily lead to improvements. However, in contrast to Orth et al. (2015), we found that low flows are better represented by the complex models, whereas they found that low flows were 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 in representing low flows have been acknowledged by several authors, such as Smakhtin (2001), Pushpalatha et al. (2011) or van Esse et al. (2013).

Effect of constraints

Figure 2.10c 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 (Fig. 2.10c). 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 on mHM do not significantly affect the performance for the Briance and Orge catchment, as can be seen by the nearly white rows in Fig. 2.10c. Notably, the runoff responses in these catchments are not snow

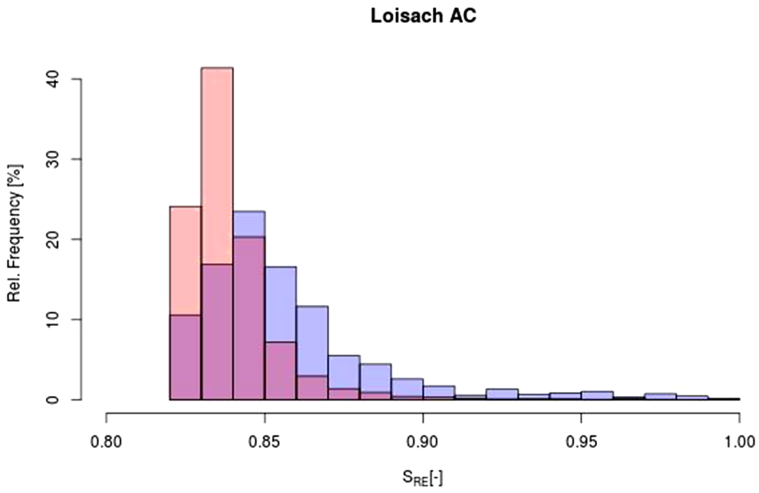

Figure 2.12: Histograms of the performance distributions for the 1-day autocorrelation of flows for the Loisach catchment on the basis of all feasible parameter sets of mHM (blue) and mHMtopo (red). The performance E_{RE} is defined as 1 minus the relative error, leading to an optimal value of 1.

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 Fig. 2.10c, d that the applied prior constraints yield a strong improvement, in particular in mHM, and in only about 29 % (mHM) and 38 % (mHMtopo) of the cases is a mostly weak performance reduction 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 (Fig.2.10d; see also the empirical distribution function at the bottom of the figures).

Alzette, Loisach and Orge show some deterioration when constraints are added (Fig. 2.10d), indicating that the topography specific constraints (Eqs. 2.4 and 2.5) 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 (Fig. 2.10d). This means that the unconstrained mHMtopo and also the constrained mHM, which does not have these topography specific constraints, will outperform the constrained mHMtopo with respect to these signatures. This is also supported by Fig. 2.10b, which 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 Fig.2.10b). The rejection of these constraints implies that for these catchments, soil moisture storage capacity in wetlands may be equal to 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 neighbouring Ammer catchment to the Loisach. Considering these groundwater leakages, the model may need extra storage to correct for it in the hydrograph.

In Fig. 2.10c, d it may also be noted that the constraints do not add information to mHM and mHMtopo with respect to the autocorrelation functions (S_{AC}, $S_{AC,low}$, $S_{AC,high}$) and rising and declining limb densities (S_{RLD}, S_{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 $S_{Q_{95},low}$ and $S_{Q_{95},high}$.

Figure 2.10d 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 Fig. 2.10c shows, but assumptions about internal model behaviour should experimentally be well founded. Even though based on several experimental studies, the topography-based parameter constraints applied (Eqs. 2.4 and 2.5) were not suitable in all cases, and led 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, the results merely support findings of Holländer et al. (2009), where different choices of expert modellers lead to a variety of outcomes.

Combined effect of constraints and sub-grid heterogeneity

Figure 2.10b 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 Fig. 2.10b are relatively similar to the patterns observed in Fig. 2.10a. It seems that the applied constraints generally enhance the effects caused by the model structure. This can be seen from darker colours of red and blue, but also from the flatter distribution function (bottom of Fig. 2.10b). Thus, when the model already has 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 Fig. 2.10a turned blue in Fig. 2.10b. The Broye gained higher probabilities of improvement, represented by more darker blue colours in Fig. 2.10b. 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 relatively low probability of improvement again. This is due to the rejection of the constraints given in Eqs. 2.4 and 2.5, as discussed before in comparison with Fig. 2.10d.

Figure 2.10e 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 Fig. 2.10e 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 (S_{RLD}) or autocorrelation (S_{AC}). In Fig. 2.10e, even a further contrast between the fast flow and low flow domains 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 Fig. 2.10e a redder 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, does the Euclidian distance D_{E} show a clear deterioration in performance.

2.3.3. 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 Fig. 2.13, 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.

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 (Fig. 2.13). 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 Fig. 2.13) and mHMtopo (red in Fig. 2.13) 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, $E_{\mathrm{NS,logQ}}$ and $E_{\mathrm{NS,FDC}}$ experience a strong decrease in performance (Fig. 2.13). The results also suggest that, in the unconstrained case, the original mHM is better transferable than mHMtopo with respect to catchment signatures (Fig. 2.14a). 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 Fig. 2.14a) probability of improvement. This indicates again that the more complex mHMtopo mostly affects

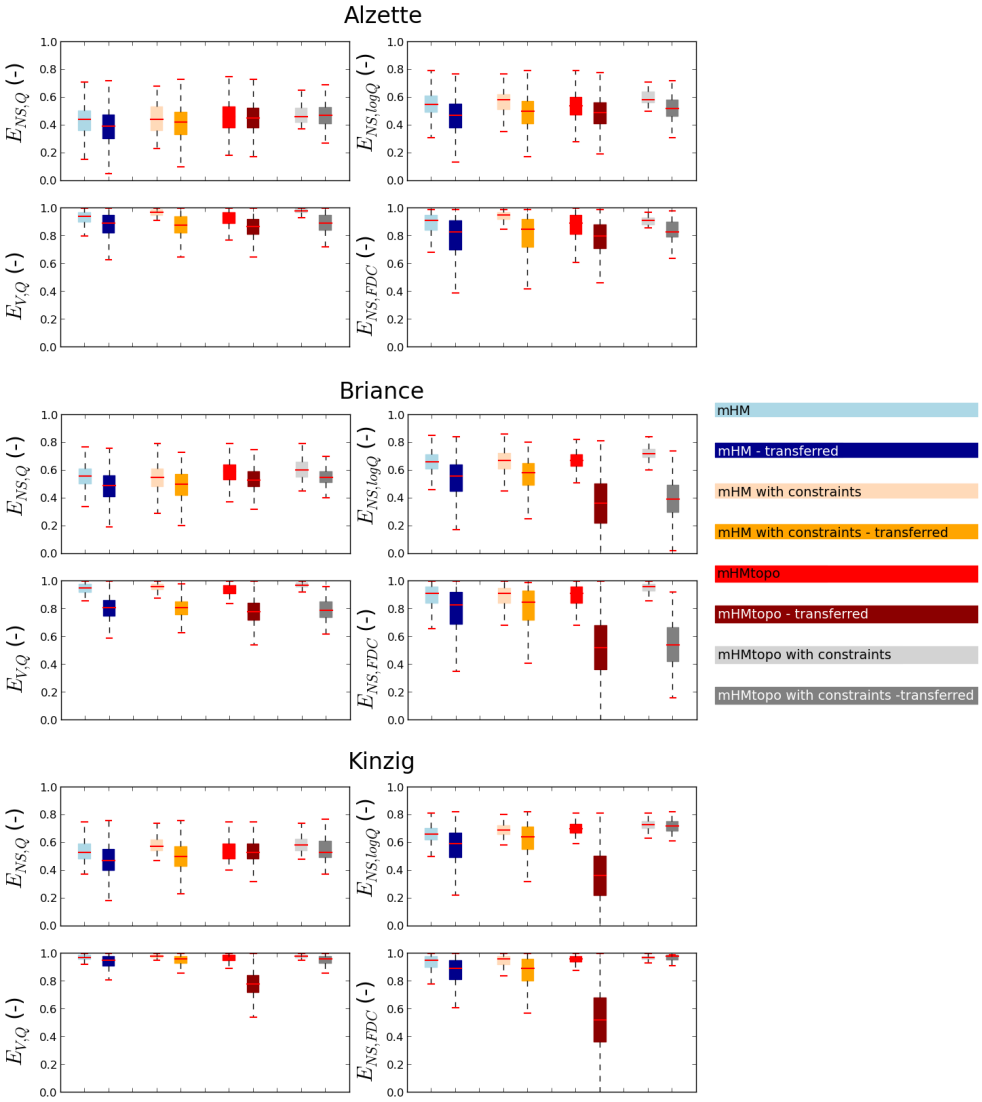

Figure 2.13: Objective function values of the **(a)** Alzette, **(b)** Briance and **(c)** Kinzig catchments in the validation period for individual calibration (light colours) and when using parameters transferred from the remaining four donor catchments in the multibasin calibration (darker colours).

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 (Kumar et al., 2013b; Samaniego et al., 2010b). 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.

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 (Fig. 2.13). This indicates that the applied constraints are very suitable for this catchment, but less so 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 (Fig. 2.14b), it can be observed that the Alzette and Kinzig catchments benefit from additional process heterogeneity and constraints, while the constrained mHM is still better at 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 Fig. 2.14c and d. The unconstrained mHM already was better transferable for this catchment compared to mHMtopo (see Fig. 2.14a); this remains the same in the constrained cases. The other two catchments are much more sensitive to the constraints and now show a better transferability, in particular with respect to the low flow signatures.

Furthermore, results shown in Fig. 2.14c and 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 (Fig. 2.13), 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 Fig. 2.14c, d. The addition of constraints to mHMtopo shows high probabilities of improvements over the full range of signatures (Fig. 2.14d), in particular for the Kinzig. Also, for mHM (Fig. 2.14c), 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.

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 Fig. 2.14e. 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 Fig. 2.14d) and from Fig. 2.14a it was already observed that the unconstrained mHM was more transferable than mHMtopo. Looking back at Fig. 2.10a, it can also be noted that in the individual calibration mHM slightly outperformed mHMtopo for this catchment with respect to the signatures (light-red 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.

2.3.4. 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 24 km by 24 km, 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 are 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.

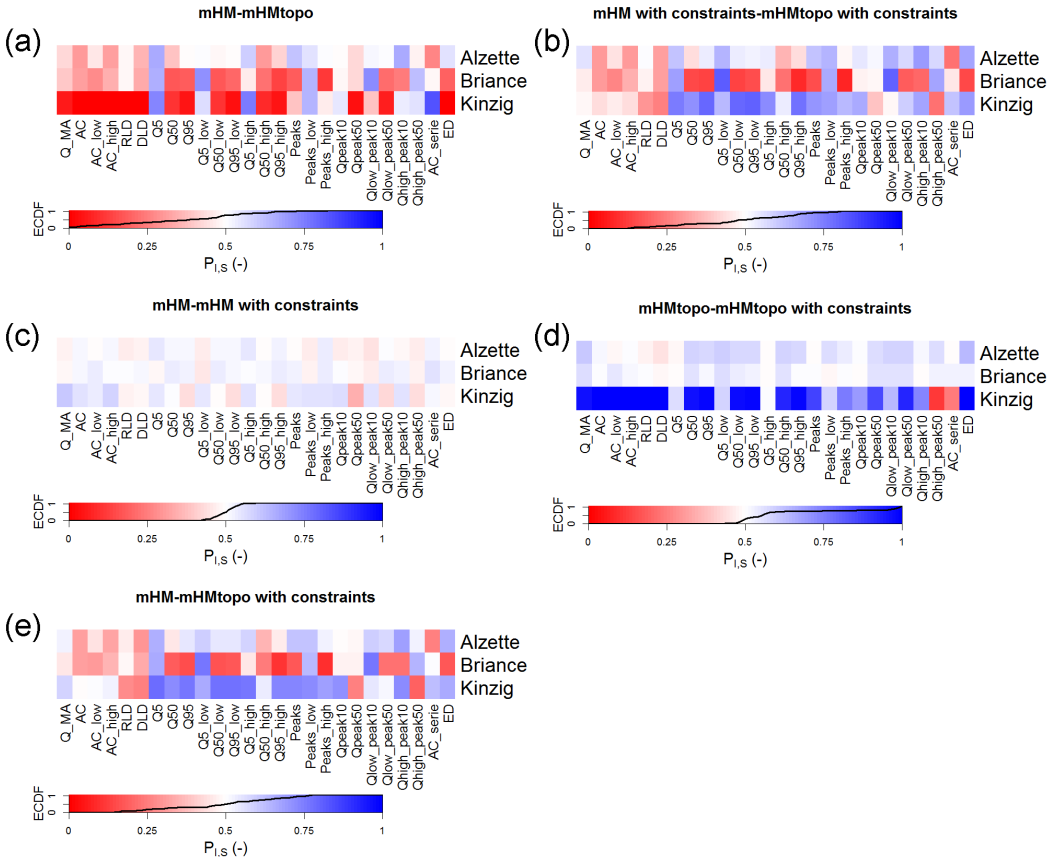

Figure 2.14: Probabilities of improvements $P_{I,S}$ between **(a)** mHM and mHMtopo without constraints and **(b)** with constraints, **(c)** mHM with and without constraints, **(d)** mHMtopo with and without constraints and **(e)** the base case mHM with the constrained mHMtopo case, all after the transfer of global parameters to the three catchments. The colours are linearly related to the probability of improvement between 0 (dark red; e.g. the probability of mHMtopo outperforming mHM is 0), 0.5 (white; i.e. models are statistically equivalent) and 1 (dark blue; e.g. the 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.

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.

2.3.5. 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 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 attributed 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.

In conclusion, the addition of topography-based sub-grid 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.

3

Evolution of root-zone storage capacity

The previous chapter focused on a formal approach to select the full set of model parameters, this chapter focuses on the estimation of a single key parameter, the root-zone storage capacity, in catchments under change. This is a key parameter determining runoff and evaporation, and a correct parameter estimation is therefore crucial in order to obtain a suitable model formulation. Especially in catchments under change, times for recovery towards new equilibriums need to be determined, as the models should be able to adapt to change. Therefore, it is tested whether a simple water-balance estimation of root-zone storage is sufficient for estimating the time-dynamic character of this parameter, in comparison with estimates from five hydrological models.

This chapter is based on:
Nijzink, R.C., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, T., McGuire, K., Savenije, H., and Hrachowitz, M.: *The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change?*, Hydrol. Earth Syst. Sci., 20, 4775-4799, doi:10.5194/hess-20-4775-2016, 2016..

3.1. Introduction

Vegetation, as a core component of the water cycle, shapes the partitioning of water fluxes on the catchment scale into runoff components and evaporation, thereby controlling fundamental processes in ecosystem functioning (Kleidon, 2004; Laio et al., 2001; Rodríguez-Iturbe, 2000), such as flood generation (Donohue et al., 2012), drought dynamics (Seneviratne et al., 2010; Teuling et al., 2013), groundwater recharge (Allison et al., 1990; Jobbágy and Jackson, 2004) and land–atmosphere feedback (Cassiani et al., 2015; Milly and Dunne, 1994; Seneviratne et al., 2013). Besides increasing interception storage available for evaporation (Gerrits et al., 2010), vegetation critically interacts with the hydrological system in a co-evolutionary way by root water uptake for transpiration, towards a dynamic equilibrium with the available soil moisture to avoid water shortage (Donohue et al., 2007; Eagleson, 1978, 1982; Gentine et al., 2012; Liancourt et al., 2012) and related adverse effects on carbon exchange and assimilation rates (Porporato et al., 2004; Seneviratne et al., 2010). Roots create moisture storage volumes within their range of influence, from which they extract water that is stored between field capacity and wilting point. This root-zone storage capacity S_R, sometimes also referred to as plant available water holding capacity, in the unsaturated soil is therefore the key component of many hydrological systems (Milly and Dunne, 1994; Rodríguez-Iturbe et al., 2007).

There is increasing theoretical and experimental evidence that vegetation dynamically adapts its root system, and thus S_R, to environmental conditions, to secure, on the one hand, access to sufficient moisture to meet the canopy water demand and, on the other hand, to minimize the carbon investment for sub-surface growth and maintenance of the root system (Brunner et al., 2015; Schymanski et al., 2008; Tron et al., 2015). In other words, the hydrologically active root-zone is optimized to guarantee productivity and transpiration of vegetation, given the climatic circumstances (Kleidon, 2004). Several studies previously showed the strong influence of climate on this hydrologically active root-zone (e.g. Laio et al., 2001; Reynolds et al., 2000; Schenk and Jackson, 2002). Moreover, droughts are often identified as critical situations that can affect ecosystem functioning evolution (e.g. Allen et al., 2010; McDowell et al., 2008; Vose et al., 2016).

In addition to their general adaption to environmental conditions, vegetation has some potential to adapt roots to such periods of water shortage (Bréda et al., 2006; Mencuccini, 2003; Sperry et al., 2002). In the short term, stomatal closure and reduction of leaf area will lead to reduced transpiration. In several case studies for specific plants, it was also shown that plants may even shrink their roots and reduce soil–root conductivity during droughts, while recovering after re-wetting (Nobel and Cui, 1992; North and Nobel, 1992). In the longer term, and more importantly, trees can improve their internal hydraulic system, for example by recovering damaged xylem or by allocating more biomass for roots (Bréda et al., 2006; Rood et al., 2003; Sperry et al., 2002). Similarly, Tron et al. (2015) argued that roots follow groundwater fluctuations, which may lead to increased rooting depths when water tables drop. Such changing environmental conditions may also provide other plant species with different water demand than the ones present under given conditions,

with an advantage in the competition for resources, as for example shown by Li et al. (2007).

The hydrological functioning of catchments (Black, 1997; Wagener et al., 2007) and thus the partitioning of water into evaporative fluxes and runoff components is not only affected by the continuous adaption of vegetation to changing climatic conditions. Rather, it is well understood that anthropogenic changes to land cover, such as deforestation, can considerably alter hydrological regimes. This has been shown historically through many paired watershed studies (e.g. Alila et al., 2009; Andréassian, 2004; Bosch and Hewlett, 1982; Brown et al., 2005). These studies found that deforestation often leads to generally higher seasonal flows and/or an increased frequency of high flows in streams, while decreasing evaporative fluxes. The timescales of hydrological recovery after such land-cover disturbances were shown to be highly sensitive to climatic conditions and the growth dynamics of the regenerating species (e.g. Brown et al., 2005; Jones and Post, 2004).

Although land-use change effects on hydrological functioning are widely acknowledged, it is less well understood which parts of the hydrological system are affected in which way and over which timescales. As a consequence, most catchment-scale models were originally not developed to deal with such changes in the system, but rather for "stationary" conditions (Ehret et al., 2014). This is true for both top-down hydrological models, such as HBV (Bergström, 1992) or GR4J (Perrin et al., 2003), and bottom-up models, such as MIKE-SHE (Refsgaard and Storm, 1995) or HydroGeoSphere (Brunner and Simmons, 2012). Several modelling studies have in the past incorporated temporal effects of land-use change to some degree (Andersson and Arheimer, 2001; Bathurst et al., 2004; Brath et al., 2006), but they mostly rely on ad hoc assumptions about how hydrological parameters are affected (Fenicia et al., 2009; Legesse et al., 2003; Mahe et al., 2005; Onstad and Jamieson, 1970). Approaches which incorporate the change in the model formulation itself are rare and have only recently gained momentum (e.g. Du et al., 2016; Fatichi et al., 2016; Zhang et al., 2016). This is of critical importance as ongoing changes in land cover and climate dictate the need for a better understanding of their effects on hydrological functioning (Troch et al., 2015) and their explicit consideration in hydrological models for more reliable predictions under change (Hrachowitz et al., 2013; Montanari et al., 2013).

As a step towards such an improved understanding and the development of time-dynamic models, we argue that the root-zone storage capacity, S_R, is a core component determining the hydrological response, and needs to be treated as a dynamically evolving parameter in hydrological modelling as a function of climate and vegetation. Gao et al. (2014b) recently demonstrated that catchment-scale S_R can be robustly estimated exclusively based on long-term water balance considerations. Wang-Erlandsson et al. (2016) derived global estimates of S_R using remote-sensing based precipitation and evaporation products, which demonstrated considerable spatial variability of S_R in response to climatic drivers. In traditional approaches, S_R is typically determined either by the calibration of a hydrological model (e.g. Seibert and McDonnell, 2010; Seibert et al., 2010) or based on soil characteristics and sparse, averaged estimates of root depths, often obtained from

Table 3.1: Overview of the catchments and their sub-catchments (WS).

Catchment	Deforestation period	Treatment	Area [km^2]	Affected Area [%]	Aridity index [–]	Precipitation [$mm\,yr^{-1}$]	Discharge [$mm\,yr^{-1}$]	Potential evaporation [$mm\,yr^{-1}$]	Time series
HJ Andrews WS1	1962–1966	Burned 1966	0.956	100	0.39	2305	1361	902	1962–1990
HJ Andrews WS2	–	–	0.603	–	0.39	2305	1251	902	1962–1990
Hubbard Brook WS2	1965–1968	Herbicides	0.156	100	0.57	1471	1059	784	1961–2009
Hubbard Brook WS3	–	–	0.424	–	0.54	1464	951	787	1961–2009
Hubbard Brook WS5	1983–1984	No treatment	0.219	87	0.51	1518	993	746	1962–2009

literature (e.g. Breuer et al., 2003; Ivanov et al., 2008). This does neither reflect the dynamic nature of the root system nor does it consider to a sufficient extent the actual function of the root-zone: providing plants with continuous and efficient access to water. This leads to the situation where soil porosity often effectively controls the values of S_R used in a model. Consider, as a thought experiment, two plants of the same species growing on different soils. They will, with the same average root depth, then have access to different volumes of water, which will merely reflect the differences in soil porosity. This is in strong contradiction to the expectation that these plants would design root systems that provide access to similar water volumes, given the evidence for efficient carbon investment in root growth (Milly, 1994; Schymanski et al., 2008; Troch et al., 2009a) and posing that plants of the same species have common limits of operation. This argument is supported by a recent study, in which was shown that water-balance-derived estimates of S_R are at least as plausible as soil-derived estimates (de Boer-Euser et al., 2016) in many environments and that the maximum root depth controls evaporative fluxes and drainage (Camporese et al., 2015).

Therefore, using water-balance-based estimates of S_R in several deforested sites as well as in untreated reference sites in two experimental forests, we test the hypotheses that 1) the root-zone storage capacity S_R significantly changes after deforestation, 2) the evolution in S_R can explain post-treatment changes to the hydrological regimes and that 3) a time-dynamic formulation of S_R can improve the

performance of a hydrological model.

3.2. Study sites

The catchments under consideration are part of the HJ Andrews Experimental Forest and the Hubbard Brook Experimental Forest. A summary of the main catchment characteristics can be found in Table 3.1. Daily discharge (Campbell, 2014a; Johnson and Rothacher, 2016), precipitation (Campbell, 2014b; Daly and McKee, 2016) and temperature time series (Campbell, 2014c,d; Daly and McKee, 2016) were obtained from the databases of the Hubbard Brook Experimental Forest and the HJ Andrews Experimental Forest. Potential evaporation was estimated by the Hargreaves equation (Hargreaves and Samani, 1985).

3.2.1. HJ Andrews Experimental Forest

The HJ Andrews Experimental Forest is located in Oregon, USA (44.2^o N, 122.2^o W) and was established in 1948. The catchments at HJ Andrews are described in many studies (e.g. Dyrness, 1969; Harr et al., 1975; Jones and Grant, 1996; Rothacher, 1965; Waichler et al., 2005).

Before vegetation removal and at lower elevations the forest generally consisted of 100- to 500-year old coniferous species, such as Douglas fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*) and western red cedar (*Thuja plicata*), whereas upper elevations were characterized by noble fir (*Abies procera*), Pacific silver fir (*Abies amabilis*), Douglas fir, and western hemlock. Most of the precipitation falls from November to April (about 80 % of the annual precipitation), whereas the summers are generally drier, leading to signals of precipitation and potential evaporation that are out of phase.

Deforestation of HJ Andrews Watershed 1 (WS1) started in August 1962 (Rothacher, 1970). Most of the timber was removed with skyline yarding. After finishing the logging in October 1966, the remaining debris was burned and the site was left for natural regrowth. Watershed 2 (WS2) is the reference catchment, which was not harvested.

3.2.2. Hubbard Brook Experimental Forest

The Hubbard Brook Experimental Forest is a research site established in 1955 and located in New Hampshire, USA (43.9^o N, 71.8^o W). The Hubbard Brook experimental catchments are described in a many publications (e.g. Dahlgren and Driscoll, 1994; Hornbeck, 1973; Hornbeck et al., 1970, 1997; Likens, 2013).

Prior to vegetation removal, the forest was dominated by northern hardwood forest composed of sugar maple (*Acer saccharum*), American beech (*Fagus grandifolia*) and yellow birch (*Betula alleghaniensis*) with conifer species such as red spruce (*Picea rubens*) and balsam fir (*Abies balsamea*) occurring at higher elevations and on steeper slopes with shallow soils. The forest was selectively harvested from 1870 to 1920, damaged by a hurricane in 1938, and is currently not accumulating biomass (Campbell et al., 2013; Likens, 2013). The annual precipitation and runoff is less than in HJ Andrews (Table 3.1). Precipitation is rather uniformly

Table 3.2: Applied parameter ranges for root-zone storage derivation.

Catchment	$I_{max,eq}$ (mm)	$I_{max,change}$ (mm)	T_r (days)
HJ Andrews WS1	1–5	0–5	0–3650
HJ Andrews WS2	1–5	–	–
Hubbard Brook WS2	1–5	5–10	0–3650
Hubbard Brook WS3	1–5	–	–
Hubbard Brook WS5	1–5	0–5	0–3650

spread throughout the year without distinct dry and wet periods, but with snowmelt-dominated peak flows occurring around April and distinct low flows during the summer months due to increased evaporation rates (Federer et al., 1990). Vegetation removal occurred in the catchment of Hubbard Brook Watershed 2 (WS2) between 1965 and 1968 and in Hubbard Brook Watershed 5 (WS5) between 1983 and 1984. Hubbard Brook Watershed 3 (WS3) is the undisturbed reference catchment.

Hubbard Brook WS2 was completely deforested in November and December 1965 (Likens et al., 1970). To minimize disturbance, no roads were constructed and all timber was left in the catchment. On 23 June 1966, herbicides were sprayed from a helicopter to prevent regrowth. Additional herbicides were sprayed in the summers of 1967 and 1968 from the ground.

In Hubbard Brook WS5, all trees were removed between 18 October 1983 and 21 May 1984, except for a 2 ha buffer near an adjacent reference catchment (Hornbeck et al., 1997). WS5 was harvested as a whole-tree mechanical clearcut with removal of 93 % of the above-ground biomass (Hornbeck et al., 1997; Martin et al., 2000), thus including smaller branches and debris. Approximately 12 % of the catchment area was developed as the skid trail network. Afterwards, no treatment was applied and the site was left for regrowth.

3.3. Methodology

To assure reproducibility and repeatability, the executional steps in the experiment were defined in a detailed protocol, following Ceola et al. (2015), which is provided as Supplement of Nijzink et al. (2016a) and online (`http://dl-ng005.xtr.deltares.nl/view/441/`) as part of the SWITCH-ON Virtual Water Science Laboratory.

3.3.1. Water-balance-derived root-zone moisture capacities S_R

The root-zone moisture storage capacities S_R and their change over time were determined according to the methods suggested by Gao et al. (2014b) and subsequently successfully tested by de Boer-Euser et al. (2016) and Wang-Erlandsson et al. (2016). Briefly, the long-term water balance provides information on actual mean transpiration. In a first step, the interception capacity has to be assumed, in order to determine the effective precipitation P_e ($L\,T^{-1}$), following the water balance

equation for interception storage:

$$\frac{\mathrm{d}S_i}{\mathrm{d}t} = P - E_i - P_\mathrm{e} \tag{3.1}$$

with S_i (L) interception storage, P the precipitation $(L\,T^{-1})$, E_i the interception evaporation $(L\,T^{-1})$. This is solved with the constitutive relations:

$$E_i = \begin{cases} E_\mathrm{p} & \text{if } E_\mathrm{p}\mathrm{d}t < S_i \\ \frac{S_i}{\mathrm{d}t} & \text{if } E_\mathrm{p}\mathrm{d}t \geq S_i \end{cases} \tag{3.2}$$

$$P_\mathrm{e} = \begin{cases} 0 & \text{if } S_i \leq I_\mathrm{max} \\ \frac{S_i - I_\mathrm{max}}{\mathrm{d}t} & \text{if } S_i > I_\mathrm{max} \end{cases} \tag{3.3}$$

with, additionally, E_p the potential evaporation $(L\,T^{-1})$ and I_max (L) the interception capacity. As I_max will also be affected by land cover change, this was addressed by introducing the three parameters $I_\mathrm{max,eq}$ (long-term equilibrium interception capacity) (L), $I_\mathrm{max,change}$ (post-treatment interception capacity) (L) and T_r (recovery time) (T), leading to a time-dynamic formulation of I_max:

$$I_\mathrm{max} = \begin{cases} \text{for } t < t_\mathrm{change}\,,\ t > t_\mathrm{change,end} + T_\mathrm{r}: \\ I_\mathrm{max,eq} \\ \\ \text{for } t_\mathrm{change,\,start} < t < t_\mathrm{change,end}: \\ I_\mathrm{max,eq} - \frac{I_\mathrm{max,eq} - I_\mathrm{max,change}}{t_\mathrm{change,end} - t_\mathrm{change,start}} \left(t - t_\mathrm{change,start}\right) \\ \\ \text{for } t_\mathrm{change,end} < t < t_\mathrm{change,end} + T_\mathrm{r}: \\ I_\mathrm{max,change} + \frac{I_\mathrm{max,eq} - I_\mathrm{max,change}}{T_\mathrm{r}} \left(t - t_\mathrm{change,end}\right) \end{cases} \tag{3.4}$$

with $t_\mathrm{change,start}$ the time that deforestation started and $t_\mathrm{start,end}$ the time deforestation finished.

Following a Monte Carlo sampling approach, upper and lower bounds of E_i were then estimated based on 1000 random samples of these parameters, eventually leading to upper and lower bounds for P_e. The interception capacity was assumed to increase after deforestation for Hubbard Brook WS2, as the debris was left at the site. For Hubbard Brook WS5 and HJ Andrews WS1 the interception capacity was assumed to decrease after deforestation, as here the debris was respectively burned and removed. Furthermore, in the absence of more detailed information, it was assumed that the interception capacities changed linearly during deforestation towards $I_\mathrm{max,change}$ and linearly recovered to I_max over the period T_r as well. See Table 3.2 for the applied parameter ranges.

Hereafter, the long-term mean transpiration can be estimated with the remaining components of the long term water balance, assuming no additional gains or losses,

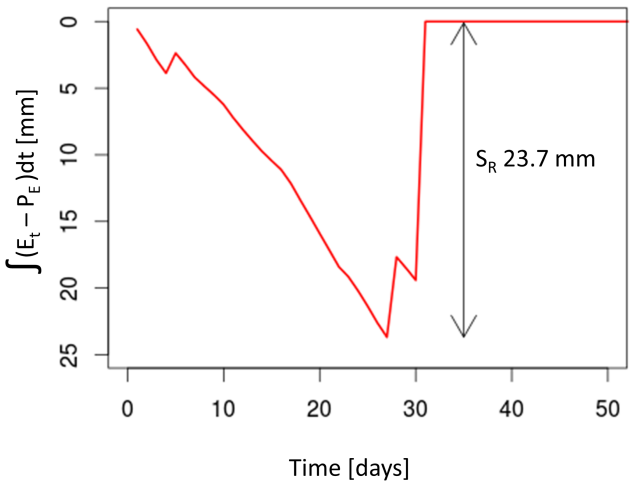

Figure 3.1: Derivation of root-zone storage capacity (S_R) for one specific time period in Hubbard Brook WS2 as difference between the cumulative transpiration (E_t) and the cumulative effective precipitation (P_E).

storage changes and/or data errors:

$$\overline{E}_t = \overline{P}_e - \overline{Q}, \tag{3.5}$$

where $\overline{E_t}$ ($L\,T^{-1}$) is the long-term mean actual transpiration, $\overline{P_e}$ ($L\,T^{-1}$) is the long-term mean effective precipitation and $\overline{Q}$ ($L\,T^{-1}$) is the long-term mean catchment runoff. Taking into account seasonality, the actual mean transpiration is scaled with the ratio of long-term mean daily potential evaporation E_p over the mean annual potential evaporation E_p:

$$\mathrm{E_t}(t) = \frac{\mathrm{E_p}(t)}{\overline{E_p}} \times \overline{E_t}. \tag{3.6}$$

Based on this, the cumulative deficit between actual transpiration and precipitation over time can be estimated by means of an "infinite-reservoir". In other words, the cumulative sum of daily water deficits, i.e. evaporation minus precipitation, is calculated between T_0, which is the time the deficit equals zero, and T_1, which is the time the total deficit returned to zero. The maximum deficit of this period then represents the volume of water that needs to be stored to provide vegetation continuous access to water throughout that time:

$$S_R = \max \int_{T0}^{T1} (E_t - P_e)\,\mathrm{d}t, \tag{3.7}$$

where S_R (L) is the maximum root-zone storage capacity over the time period between T_0 and T_1. See also Fig. 3.1 for a graphical example of the calculation for the

Hubbard Brook catchment for one specific realization of the parameter sampling. The $S_{R,20yr}$ for drought return periods of 20 years was estimated using the Gumbel extreme value distribution (Gumbel, 1941) as previous work suggested that vegetation designs S_R to satisfy deficits caused by dry periods with return periods of approximately 10–20 years (de Boer-Euser et al., 2016; Gao et al., 2014b). Thus, the maximum values of S_R for each year, as obtained by Eq. 3.7, were fitted to the extreme value distribution of Gumbel, and subsequently, the $S_{R,20yr}$ was determined.

For the study catchments that experienced logging and subsequent reforestation, it was assumed that the root system converges towards a dynamic equilibrium approximately 10 years after reforestation. Thus, the equilibrium $S_{R,20yr}$ was estimated using only data over a period that started at least 10 years after the treatment. For the growing root systems during the years after reforesting, the storage capacity does not yet reach its dynamic equilibrium $S_{R,20yr}$. Instead of determining an equilibrium value, the maximum occurring deficit for each year was in that case considered as the maximum demand and thus as the maximum required storage $S_{R,1yr}$ for that year. To make these yearly estimates, the mean transpiration was determined in a similar way as stated by Eq. 3.5. However, the assumption of no storage change may not be valid for 1-year periods. In a trade-off to limit the potential bias introduced by inter-annual storage changes in the catchments, the mean transpiration was determined based on the 2-year water balance, thus assuming negligible storage change over these years.

The deficits in the months October–April are highly affected by snowfall, as estimates of the effective precipitation are estimated without accounting for snow, leading to soil moisture changes that spread out over an unknown longer period due to the melt process. Therefore, to avoid this influence of snow, only deficits as defined by Eq. 3.7, in the period of May–September are taken into consideration, which is also the period where deficits are significantly increasing due to relatively low rainfall and high transpiration rates, thus causing soil moisture depletion and drought stress for the vegetation, which in turn, shapes the root-zone.

3.3.2. Model-derived root-zone storage capacity $S_{u,max}$

The water-balance-derived equilibrium $S_{R,20yr}$ as well as the dynamically changing $S_{R,1yr}$ that reflects regrowth patterns in the years after treatment were compared with estimates of the calibrated parameter $S_{u,max}$, which represents the mean catchment root-zone storage capacity in lumped conceptual hydrological models. Due to the lack of direct observations of the changes in the root-zone storage capacity, this comparison was used to investigate whether the estimates of the root-zone storage capacity $S_{R,1yr}$, their sensitivity to land-cover change and their effect on hydrological functioning, can provide plausible results. Model-based estimates of root-zone storage capacity may be highly influenced by model formulations and parameterizations. Therefore, four different hydrological models were used to derive the parameter $S_{u,max}$ in order to obtain a set of different estimates of the catchment-scale root-zone storage capacity. The major features of the model routines for root-zone moisture tested here are briefly summarized below and detailed descrip-

tions including the relevant equations are provided in the Supplement Sect. S2 of Nijzink et al. (2016a).

FLEX

The FLEX-based model Fenicia et al. (2008) was applied in a lumped way to the catchments. The model has nine parameters, eight of which are free calibration parameters, sampled from relatively wide, uniform prior distributions. In contrast, based on the estimation of a Master Recession Curve (e.g. Fenicia et al., 2006), an informed prior distribution between narrow bounds could be used for determining the slow reservoir coefficient K_s.

The model consists of five storage components. First, a snow routine has to be run, which is a simple degree-day module, similar to that used in, for example, HBV (Bergström, 1976). After the snow routine, the precipitation enters the interception reservoir. Here, water evaporates at potential rates or, when exceeding a threshold, directly reaches the soil moisture reservoir. The soil moisture routine is modelled in a similar way to the Xinanjiang model (Zhao, 1992). Briefly, it contains a distribution function that determines the fraction of the catchment where the storage deficit in the root-zone is satisfied and that is therefore hydrologically connected to the stream and generating storm runoff. From the soil moisture reservoir, water can further vertically percolate down to recharge the groundwater or leave the reservoir through transpiration. Transpiration is a function of maximum root-zone storage $S_{u,\max}$ and the actual root-zone storage, similar to the functions described by Feddes et al. (1978). Water that cannot be stored in the soil moisture storage is then split into preferential percolation to the groundwater and runoff generating fluxes that enter a fast reservoir, which represents fast-responding system components such as shallow subsurface and overland flow.

HYPE

The HYPE model (Lindström et al., 2010) estimates soil moisture for hydrological response units (HRU), which is the finest calculation unit in this catchment model. In the current set-up, 15 parameters were left free for calibration. Each HRU consists of a unique combination of soil and land-use classes with assigned soil depths. Water input is estimated from precipitation after interception and a snow module at the catchment scale, after which the water enters the three defined soil layers in each HRU. Evaporation and transpiration occurs in the first two layers and fast surface runoff is produced when these layers are fully saturated or when rainfall rates exceeds the maximum infiltration capacities. Water can move between the layers through percolation or laterally via fast flow pathways. The groundwater table is fluctuating between the soil layers with the lowest soil layer normally reflecting the base flow component in the hydrograph. The water balance of each HRU is calculated independently and the runoff is then aggregated in a local stream with routing before entering the main stream.

TUW

The TUW model (Parajka et al., 2007) is a conceptual model with a structure similar to that of HBV (Bergström, 1976) and has 15 free calibration parameters. After a

snow module, based on a degree-day approach, water enters a soil moisture routine. From this soil moisture routine, water is partitioned into runoff-generating fluxes and evaporation. Here, transpiration is determined as a function of maximum root-zone storage $S_{u,\max}$ and actual root-zone storage as well. The runoff-generating fluxes percolate into two series of reservoirs. A fast-responding reservoir with overflow outlet represents shallow subsurface and overland flow, while the slower responding reservoir represents the groundwater.

HYMOD

HYMOD (Boyle, 2001) is similar to the applied model structure for FLEX, but only has eight parameters. Besides that, the interception module and percolation from soil moisture to the groundwater are missing. Nevertheless, the model accounts similarly for the partitioning of transpiration and runoff generation in a soil moisture routine. Also for this model, transpiration is a function of maximum storage and actual storage in the root-zone. The runoff-generating fluxes are eventually divided over a slow reservoir, representing groundwater, and a fast reservoir, representing the fast processes.

Model calibration

Each model was calibrated using a Monte-Carlo strategy within consecutive 2-year windows in order to obtain a time series of root-zone moisture capacities $S_{u,\max}$. FLEX, TUW and HYMOD were all run 100 000 times, whereas HYPE was run 10 000 times and 20 000 times for HJ Andrews WS1 and the Hubbard Brook catchments respectively, due to the required runtimes. The Kling–Gupta efficiency for flows (Gupta et al., 2009) and the Kling–Gupta efficiency for the logarithm of the flows were simultaneously used as objective functions in a multi-objective calibration approach to evaluate the model performance for each window. These were selected in order to obtain rather balanced solutions that enable a sufficient representation of peak flows, low flows and the water balance. The unweighted Euclidian distance of the three objective functions served as an informal measure to obtain these balanced solutions (e.g. Hrachowitz et al., 2014; Schoups et al., 2005):

$$L\left(\theta\right) = 1 - \sqrt{\left(1 - E_{\mathrm{KG,Q}}\right)^2 + \left(1 - E_{\mathrm{KG,logQ}}\right)^2}, \tag{3.8}$$

where $L(\theta)$ is the conditional probability for parameter set θ [–], $E_{\mathrm{KG,Q}}$ the Kling–Gupta efficiency [–], $E_{\mathrm{KG,logQ}}$ the Kling–Gupta efficiency for the log of the flows [–].

Eventually, a weighting method based on the GLUE-approach of Freer et al. (1996) was applied. To estimate posterior parameter distributions all solutions with Euclidian distances smaller than 1 were maintained as feasible. The posterior distributions were then determined with the Bayes rule (cf. Freer et al., 1996):

$$L_2\left(\theta\right) = L\left(\theta\right)^n \times L_0\left(\theta\right)/C, \tag{3.9}$$

where $L_0(\theta)$ is the prior parameter distribution [–], $L_2(\theta)$ is the posterior conditional probability [–] , n is a weighting factor (set to 5) [–], and C is a normalizing

constant [–]. 5/95th model uncertainty intervals were then constructed based on the posterior conditional probabilities.

3.3.3. Trend analysis

To test if $S_{R,1yr}$ significantly changes following de- and subsequent reforestation, which would also indicate shifts in distinct hydrological regimes, a trend analysis, as suggested by Allen et al. (1998), was applied to the $S_{R,1yr}$ values obtained from the water-balance-based method. As the sampling of interception capacities (Eq. 3.4) leads to $S_{R,1yr}$ values for each point in time, which are all equally likely in absence of any further knowledge, the mean of this range was assumed as an approximation of the time-dynamic character of $S_{R,1yr}$.

Briefly, a linear regression between the full series of the cumulative sums of $S_{R,1yr}$ in the deforested catchment and the unaffected control catchment is established and the residuals and the cumulative residuals are plotted in time. A 95 %-confidence ellipse is then constructed from the residuals:

$$X = \frac{n}{2}\cos(\alpha), \tag{3.10}$$

$$Y = \frac{n}{\sqrt{n-1}} Z_{p95}\, \sigma_r \sin(\alpha), \tag{3.11}$$

where X presents the x coordinates of the ellipse (T), Y represents the y coordinates of the ellipse (L), n is the length of the time series (T), α is the angle defining the ellipse $(0–2\pi)$ between the diagonal of the ellipse and the x axis (–), Z_{p95} is the value belonging to a probability of 95 % of the standard student t-distribution (–) and σ_r is the standard deviation of the residuals (assuming a normal distribution) (L).

When the cumulative sums of the residuals plot outside the 95 %-confidence interval defined by the ellipse, the null-hypothesis that the time series are homogeneous is rejected. In that case, the residuals from this linear regression where residual values change from either solely increasing to decreasing or vice versa, can then be used to identify different sub-periods in time.

Thus, in a second step, for each identified sub-period a new regression, with new (cumulative) residuals, can be used to check homogeneity for these sub-periods. In a similar way as before, when the cumulative residuals of these sub-periods now plot within the accompanying newly created 95 %-confidence ellipse, the two series are homogeneous for these sub-periods. In other words, the two time series show consistent behaviour over this particular period.

3.3.4. Model with time-dynamic formulation of $S_{u,max}$

In a last step, the FLEX model was reformulated to allow for a time-dynamic representation of the parameter $S_{u,max}$, reflecting the root-zone storage capacity.

As a reference, the long-term water-balance-derived root-zone storage capacity $S_{R,20yr}$ was used as a static formulation of $S_{u,max}$ in the model, and thus kept constant in time. The remaining parameters were calibrated using the calibration strategy

outlined above over a period starting with the treatment in the individual catchments until at least 15 years after the end of the treatment. This was done to focus on the period under change (i.e. vegetation removal and recovery), during which the differences between static and dynamic formulations of $S_{u,\max}$ are assumed to be most pronounced.

To test the effect of a dynamic formulation of $S_{u,\max}$ as a function of forest regrowth, the calibration was run with a temporally evolving series of root-zone storage capacity. The time-dynamic series of $S_{u,\max}$ were obtained from a relatively simple growth function, the Weibull function (Weibull, 1951):

$$S_{u,\max}(t) = S_{\mathrm{R,20yr}} \left(1 - e^{-at^b}\right), \tag{3.12}$$

where $S_{u,\max}(t)$ is the root-zone storage capacity t time steps after reforestation (L), $S_{\mathrm{R,20yr}}$ is the equilibrium value (L), and a (T^{-1}) and b (–) are shape parameters. In the absence of more information, this equation was selected as the first, simple way of incorporating the time-dynamic character of the root-zone storage capacity in a conceptual hydrological model. In this way, root growth is exclusively determined dependent on time, whereas the shape parameters a and b merely implicitly reflect the influence of other factors, such as climatic forcing, in a lumped way. These parameters were estimated based on qualitative judgement so that $S_{u,\max}(t)$ coincides well with the suite of S_{R1yr} values after logging. In other words, the values were chosen by trial and error in such a way that the time-dynamic formulation of $S_{u,\max}(t)$ shows a visually good correspondence with the S_{R1yr} values. This approach was followed to filter out the short-term fluctuations in the S_{R1yr} values, which is not warranted by this equation. Note that this rather simple approach is merely meant as a proof of concept for a dynamic formulation of $S_{u,\max}$.

In addition, the remaining parameter directly related to vegetation, the interception capacity ($I_{\max}$), was also assigned a time-dynamic formulation. Here, the same growth function was applied (Eq. 3.12), but the shape of the growth function was assumed fixed (i.e. growth parameters a and b were fixed to values of 0.001 (day^{-1}) and 1 (–)) loosely based on the posterior ranges of the window calibrations, with qualitative judgement as well. This growth function was used to ensure the degrees of freedom for both the time-variant and the time-invariant models, leaving the equilibrium value of the interception capacity as the only free calibration parameter for this process. Note that the empirically parameterized growth functions can be readily extended and/or replaced by more mechanistic, process-based descriptions of vegetation growth if warranted by the available data, and they were here merely used to test the effect of considering changes in vegetation on the skill of models to reproduce hydrological response dynamics.

To assess the performance of the dynamic model compared to the time-invariant formulation, beyond the calibration objective functions, model skill in reproducing 28 hydrological signatures was evaluated (Sivapalan et al., 2003). Even though the signatures are not always fully independent of each other, this larger set of measures allows a more complete evaluation of the model skill as, ideally, the model should be able to simultaneously reproduce all signatures. An overview of the signatures

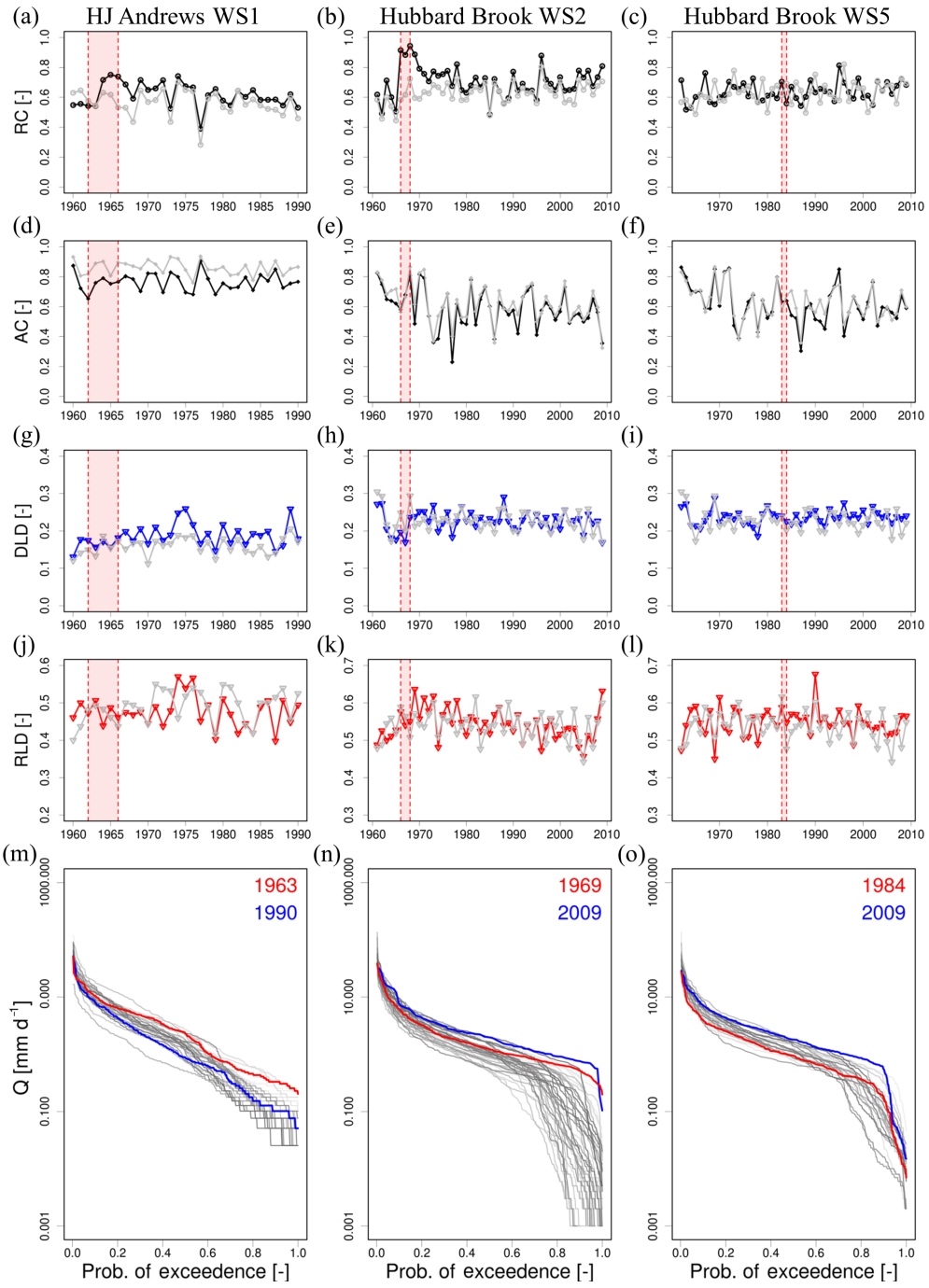

Figure 3.2: Evolution of signatures in time of (**a**–**c**) the runoff coefficient, (**d**–**f**) the 1-day autocorrelation, (**g**–**i**) the declining limb density, (**j**–**l**) the rising limb density with the reference watersheds in grey and periods of deforestation in red shading. The flow duration curves for HJ Andrews WS1, Hubbard Brook WS2 and Hubbard Brook WS5 are shown in **(m)**–**(o)**, where years between the first and last year are coloured from light gray to dark grey as they progress in time.

Table 3.3: Overview of the hydrological signatures.

Signature	Description	Reference
S_{QMA}	Mean annual runoff	
S_{AC}	One day autocorrelation coefficient	Montanari and Toth (2007)
$S_{\text{AC,summer}}$	One day autocorrelation the summer period	Euser et al. (2013)
$S_{\text{AC,winter}}$	One day autocorrelation the winter period	Euser et al. (2013)
S_{RLD}	Rising limb density	Shamir et al. (2005)
S_{DLD}	Declining limb density	Shamir et al. (2005)
S_{Q_5}	Flow exceeded in 5 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{50}}$	Flow exceeded in 50 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{95}}$	Flow exceeded in 95 % of the time	Jothityangkoon et al. (2001)
$S_{Q_5,\text{summer}}$	Flow exceeded in 5 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{50},\text{summer}}$	Flow exceeded in 50 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{95},\text{summer}}$	Flow exceeded in 95 % of the summer time	Yilmaz et al. (2008)
$S_{Q_5,\text{winter}}$	Flow exceeded in 5 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{50},\text{winter}}$	Flow exceeded in 50 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{95},\text{winter}}$	Flow exceeded in 95 % of the winter time	Yilmaz et al. (2008)
S_{Peaks}	Peak distribution	Euser et al. (2013)
$S_{\text{Peaks,summer}}$	Peak distribution summer period	Euser et al. (2013)
$S_{\text{Peaks,winter}}$	Peak distribution winter period	Euser et al. (2013)
$S_{Q_{\text{peak,10}}}$	Flow exceeded in 10 % of the peaks	
$S_{Q_{\text{peak,50}}}$	Flow exceeded in 50 % of the peaks	
$S_{Q_{\text{summer,peak,10}}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{summer,peak,50}}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{winter,peak,10}}}$	Flow exceeded in 10 % of the winter peaks	
$S_{Q_{\text{winter,peak,50}}}$	Flow exceeded in 50 % of the winter peaks	
S_{SFDC}	Slope flow duration curve	Yadav et al. (2007)
S_{LFR}	Low flow ratio (Q_{90} / Q_{50})	
S_{FDC}	Flow duration curve	Westerberg et al. (2011)
$S_{\text{AC,serie}}$	Autocorrelation series (200 days lag time)	Montanari and Toth (2007)

is given in Table 3.3. The results of the comparison were quantified on the basis of the probability of improvement for each signature (Nijzink et al., 2016b):

$$\begin{aligned} P_{I,S} &= P\left(S_{\text{dyn}} > S_{\text{stat}}\right) \\ &= \sum_{i=1}^{n} P\left(S_{\text{dyn}} > S_{\text{stat}} \,|\, S_{\text{dyn}} = r_i\right) P\left(S_{\text{dyn}} = r_i\right), \end{aligned} \tag{3.13}$$

where S_{dyn} and S_{stat} are the distributions of the signature performance metrics of the dynamic and static model, respectively, for the set of all feasible solutions retained from calibration, r_i is a single realization from the distribution of S_{dyn} and n is the total number of realizations of the S_{dyn} distribution. For $P_{I,S} > 0.5$ it is then

more likely that the dynamic model outperforms the static model with respect to the signature under consideration, and vice versa for $P_{I,S} < 0.5$. The signature performance metrics that were used are the relative error (for single-valued signatures) and the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970), for signatures that represent a time series.

In addition, as a more quantitative measure, the ranked probability score, giving information on the magnitude of model improvement or deterioration, was calculated (Wilks, 2011):

$$S_{\mathrm{RP}} = \frac{1}{M-1} \sum_{m=1}^{M} \left[\left(\sum_{k=1}^{m} p_k \right) - \left(\sum_{k=1}^{m} o_k \right) \right]^2, \tag{3.14}$$

where M is the number of feasible solutions, p_k the probability of a certain signature performance to occur and o_k the probability of the observation to occur (either 1 or 0, as there is only a single observation). Briefly, the S_{RP} represents the area enclosed between the cumulative probability distribution obtained by model results and the cumulative probability distribution of the observations. Thus, when modelled and observed cumulative probabilities are identical, the enclosed area goes to zero. Therefore, the difference between the S_{RP} for the feasible set of solutions for the time-variant and time-invariant model formulation was used in the comparison, identifying which model is quantitatively closer to the observation.

3.4. Results

3.4.1. Changes in hydrological response dynamics

We found that the three deforested catchments in the two research forests show on balance similar response dynamics after the logging of the catchments (Fig. 3.2). This supports the findings from previous studies of these catchments (Andréassian, 2004; Bosch and Hewlett, 1982; Hornbeck et al., 1997; Rothacher et al., 1967). More specifically, it was found that the observed annual runoff coefficients for HJ Andrews WS1 and Hubbard Brook WS2 (Fig. 3.2a, b) change after logging of the catchments, also in comparison with the adjacent, undisturbed reference watersheds. Right after deforestation, runoff coefficients increase, followed by a gradual decrease.

The annual autocorrelation coefficients with a 1-day lag time are generally lower after logging than in the years before the change, which can be seen in particular from Fig. 3.2e and f as here a long pre-treatment time series record is available. Nevertheless, the climatic influence cannot be ignored here, as the reference watershed shows a similar pattern. Only for Hubbard Brook WS5 (Fig. 3.2f) the autocorrelation does show reduced values in the first years after logging. Thus, the flows at any time $t + 1$ are less dependent on the flows at t, which points towards less memory and thus less storage in the system (i.e. reduced S_{R}), leading to increased peak flows, similar to the reports of, for example, Patric and Reinhart (1971) for one of the Fernow experiments.

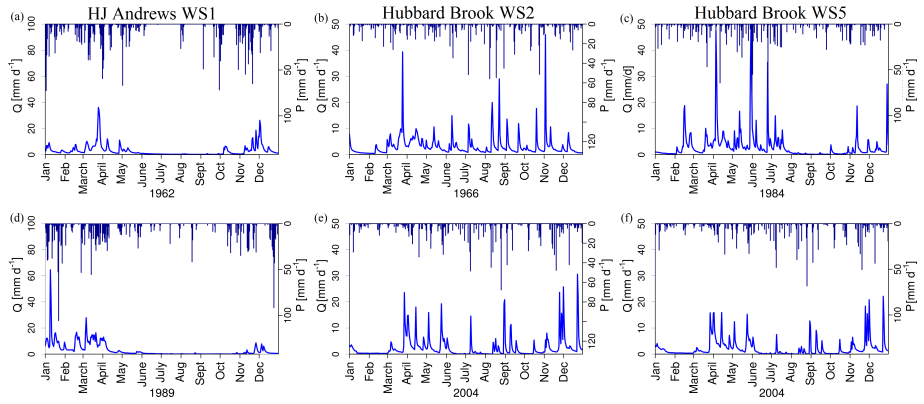

Figure 3.3: Hydrographs for HJ Andrews WS1 in **(a)** 1962 (annual precipitation $P_A = 2018$, $E_{p,A} = 951\,\mathrm{mm\,yr^{-1}}$) and **(b)** 1989 ($P_A = 1752$, $E_{p,A} = 846\,\mathrm{mm\,yr^{-1}}$), Hubbard Brook WS2 in **(c)** 1966 ($P_A = 1222$, $E_{p,A} = 788\,\mathrm{mm\,yr^{-1}}$ and **(d)** 2004 ($P_A = 1296$, annual $E_{p,A} = 761\,\mathrm{mm\,yr^{-1}}$ and Hubbard Brook WS5 in **(e)** 1984 ($P_A = 1480$, annual $E_{p,A} = 721\,\mathrm{mm\,yr^{-1}}$) and **(f)** 2004 ($P_A = 1311$, $E_{p,A} = 731\,\mathrm{mm\,yr^{-1}}$).

The declining limb density for HJ Andrews WS1 (Fig. 3.2g) shows increased values right after deforestation, whereas a longer time after deforestation, the values seem to plot closer to the values obtained from the reference watershed. This indicates that for the same number of peaks, less time was needed for the recession in the hydrograph in the early years after logging. In contrast, the rising limb density shows increased values during and right after deforestation for Hubbard Brook WS2 and WS5 (Fig. 3.2k–l), compared to the reference watershed. Here, less time was needed for the rising part of the hydrograph in the more early years after logging. Thus, the recession seems to be affected in HJ Andrews WS1, whereas the Hubbard Brook watersheds exhibit a quicker rise of the hydrograph.

Eventually, the flow duration curves, as shown in Fig. 3.2m–o, indicate a higher variability of flows, as the years following deforestation plot with an increased steepness of the flow duration curve, i.e. a higher flashiness. This increased flashiness of the catchments after deforestation can also be noted from the hydrographs shown in Fig. 3.3. The peaks in the hydrographs are generally higher, and the flows return faster to the baseflow values in the years right after deforestation than some years later after some forest regrowth, all with similar values for the yearly sums of precipitation and potential evaporation.

3.4.2. Temporal evolution of S_R and $S_{u,\max}$

The observed changes in the hydrological response of the study catchments (as discussed above) were also clearly reflected in the temporal evolution of the root-zone storage capacities as described by the catchment models (Fig. 3.4). The models all exhibited Kling–Gupta efficiencies ranging between 0.5 and 0.8 and Kling–Gupta efficiencies of the log of the flows between 0.2 and 0.8 (see the Supplement of Nijzink et al. (2016b), Figs. S5–S7, with all posterior parameter distributions in

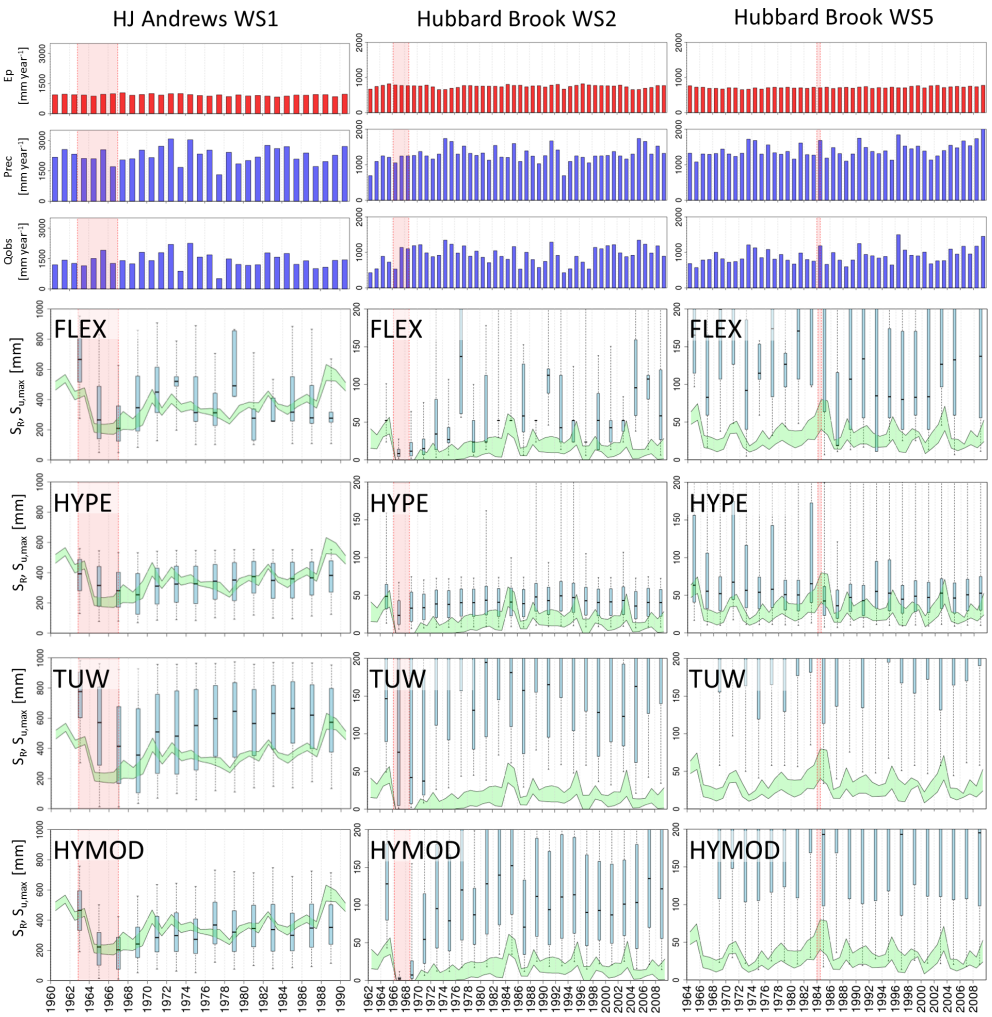

Figure 3.4: Evolution of root-zone storage capacity $S_{R,1yr}$ from water balance-based estimation (green shaded area, a range of solutions due to the sampling of the unknown interception capacity) compared with $S_{u,\max,2yr}$ estimates obtained from the calibration of four models (FLEX, HYPE, TUW, HYMOD; blue box plots) for HJ Andrews WS1, Hubbard Brook WS2 and Hubbard Brook WS5. Red shaded areas are periods of deforestation.

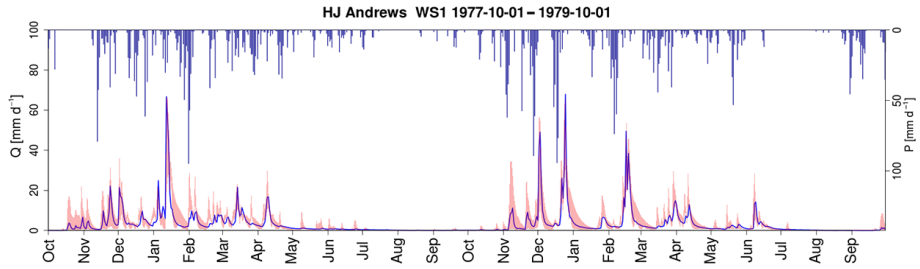

Figure 3.5: Observed and modelled hydrograph for HJ Andrews WS1 for the years of 1978 and 1979, with the red coloured area indicating the 5/95 % uncertainty intervals of the modelled discharge. Blue bars show daily precipitation.

Figs. S10–S27, and the number of feasible solutions in Tables S5–S7). Comparing the water-balance- and model-derived estimates of root-zone storage capacity S_R and $S_{u,max}$, respectively, then showed that they exhibit very similar patterns in the study catchments. Especially for HJ Andrews WS1 and Hubbard Brook WS2, root-zone storage capacities sharply decreased after deforestation and gradually recovered during regrowth towards a dynamic equilibrium of climate and vegetation, whereas the undisturbed reference catchments of HJ Andrews WS2 and Hubbard Brook WS3 showed a rather constant signal over the full period (not shown).

The HJ Andrews WS1 shows the clearest signal when looking at the water-balance-derived S_R, as can be seen by the green shaded area in Fig. 3.4a. Before deforestation, the root-zone storage capacity $S_{R,1yr}$ was found to be around 400 mm. During deforestation, the $S_{R,1yr}$ required to provide the remaining vegetation with sufficient and continuous access to water decreased from around 400 to 200 mm. For the first 4–6 years after deforestation the $S_{R,1yr}$ increased again, reflecting the increased water demand of vegetation with the regrowth of the forest. In addition, it was observed that in the period 1971–1978 $S_{R,1yr}$ slowly decreased again in HJ Andrews.

The four models show a similar pronounced decrease of the calibrated, feasible set of $S_{u,max}$ during deforestation and a subsequent gradual increase over the first years after deforestation. The model concepts, and thus our assumptions about nature, can therefore only account for the changes in hydrological response dynamics of a catchment, when calibrated in a window calibration approach with different parameterizations for each time frame. The absolute values of $S_{u,max}$ obtained from the most parsimonious HYMOD and FLEX models (both with 8 free calibration parameters) show a somewhat higher similarity to $S_{R,1yr}$ and its temporal evolution than the values from the other two models. In spite of similar general patterns in $S_{u,max}$, the higher number of parameters in TUW (i.e. 15) result, due to compensation effects between individual parameters, in wider uncertainty bounds which are less sensitive to change. It was also observed that in particular TUW overestimates $S_{u,max}$ compared to $S_{R,1yr}$, which can be attributed to the absence of an interception reservoir, leading to a root-zone that has to satisfy not only transpiration but all evaporative fluxes.

Hubbard Brook WS2 exhibits a similarly clear decrease in root-zone storage capacity as a response to deforestation, as shown in Fig. 3.4. The water-balance-based $S_{R,1yr}$ estimates approach values of zero during and right after deforestation. In these years the catchment was treated with herbicides, removing effectively any vegetation, thereby minimizing transpiration. In this catchment a more gradual regrowth pattern occurred, which continued after logging started in 1966 until around 1983. Generally, the models applied in Hubbard Brook WS2 show similar behaviour to those in the HJ Andrews catchment. The calibrated $S_{u,max}$ clearly follows the temporal pattern of $S_{R,1yr}$, reflecting the pronounced effects of de- and reforestation. It can, however, also be observed that the absolute values of $S_{u,max}$ exceed the $S_{R,1yr}$ estimates. While FLEX on balance exhibits the closest resemblance between the two values, the TUW model in particular exhibits wide uncertainty bounds with elevated $S_{u,max}$ values. Besides the role of interception evaporation, which is only explicitly accounted for in FLEX, the results are also linked to the fact that the humid climatic conditions with little seasonality reduces the importance of the model parameter $S_{u,max}$, and makes it thereby more difficult to identify by calibration. The parameter is most important for lengthy dry periods when vegetation needs enough storage to ensure continuous access to water. The temporal variation in S_R in Hubbard Brook WS5 does not show such a distinct signal as in the other two study catchments (Fig. 3.4). Moreover, it can be noted that in the summers of 1984 and 1985 the values of $S_{R,1yr}$ are relatively high. Nevertheless, the model-based values of $S_{u,max}$ show again similar dynamics to the water-balance-based $S_{R,1yr}$ values. TUW and HYMOD show again higher model-based values, but FLEX is also now overestimating the root-zone storage capacity.

3.4.3. Trend analysis and change in hydrological regimes

The trend analysis for water-balance-derived values of $S_{R,1yr}$ suggests that for all three study catchments significantly different hydrological regimes in time can be identified before and after deforestation, linked to changes in $S_{R,1yr}$ (Fig. 3.7). For all three catchments, the cumulative residuals plot outside the 95 %-confidence ellipse, indicating that the time series obtained in the control catchments and the deforested catchments are not homogeneous (Fig. 3.7g–i).

Rather obvious break points can be identified in the residual plots for the catchments HJ Andrews WS1 and Hubbard Brook WS2 (Fig. 3.7d–e). Splitting up the $S_{R,1yr}$ time series according to these break points into the periods before deforestation, deforestation and recovery resulted in three individually homogenous time series that are significantly different from each other, indicating switches in the hydrological regimes. The results shown in Fig. 3.4 indicate that these catchments developed a rather stable root-zone storage capacity sometime after the start of deforestation (for HJ Andrews WS1 after 1964, for Hubbard Brook WS2 after 1967). Hence, recovery and deforestation balanced each other, leading to a temporary equilibrium. The recovery signal then becomes more dominant in the years after deforestation. The third homogenous period suggests that the root-zone storage capacity reached a dynamic equilibrium without any further systematic changes. This can be interpreted in the way that in the HJ Andrews WS1, hydrological recov-

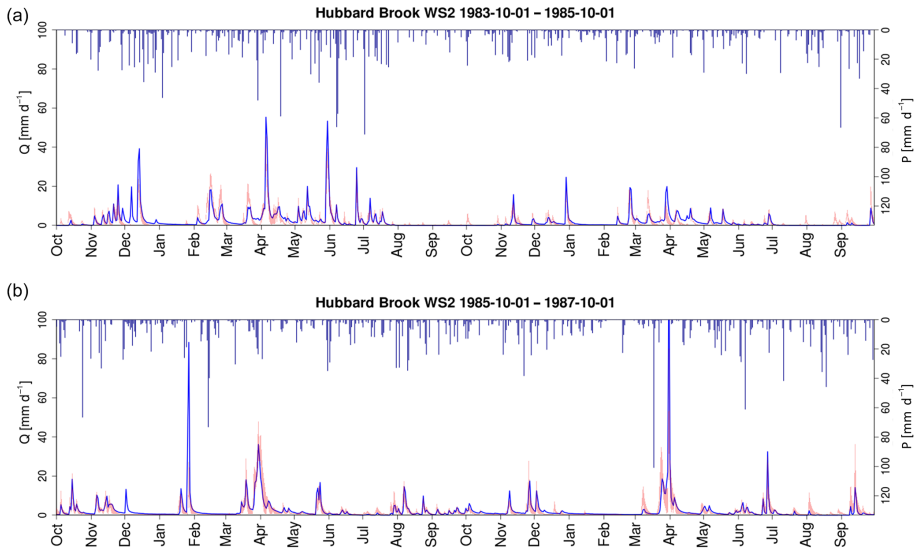

Figure 3.6: Observed and modelled hydrograph for Hubbard Brook WS2 for **(a)** the years of 1984 and 1985 and **(b)** the years of 1986 and 1987, with the red coloured area indicating the 5/95 % uncertainty intervals of the modelled discharge. Blue bars show daily precipitation.

ery after deforestation due to the recovery of the root-zone storage capacity took about 6–9 years (Fig. 3.7p), while Hubbard Brook WS2 required 10–13 years for hydrological recovery (Fig. 3.7q). This strongly supports the results of Hornbeck et al. (2014), who reported changes in water yield for WS2 for up to 12 years after deforestation.

The identification of different periods is less obvious for Hubbard Brook WS5, but the two time series of control catchment and treated catchment are significantly different (see the cumulative residuals in Fig. 3.7i). Nevertheless, the most obvious break point in residuals can be found in 1989 (Fig. 3.7f). In addition, it can be noted that turning points also exist in 1983 and 1985. These years can be used to split the time series into four groups (leading to the periods of 1964–1982, 1983–1985, 1986–1989 and 1990–2009 for further analysis). The cumulative residuals from the new regressions, based on the grouping, plot within the confidence bounds again, and show a period with deforestation (1983–1985) and recovery (1986–1989). Mou et al. (1993) reported similar findings with the highest biomass accumulation in 1986 and 1988, and slower vegetation growth in the early years. Therefore, full recovery took 5–6 years in Hubbard Brook WS5.

3.4.4. Time-variant model formulation

The adjusted model routine for FLEX, which uses a dynamic time series of $S_{u,\max}$, generated with the Weibull growth function (Eq. 3.12), resulted in a rather small impact on the overall model performance in terms of the calibration objective function values (Fig. 3.8b, d, f) compared to the time-invariant formulation of the model.

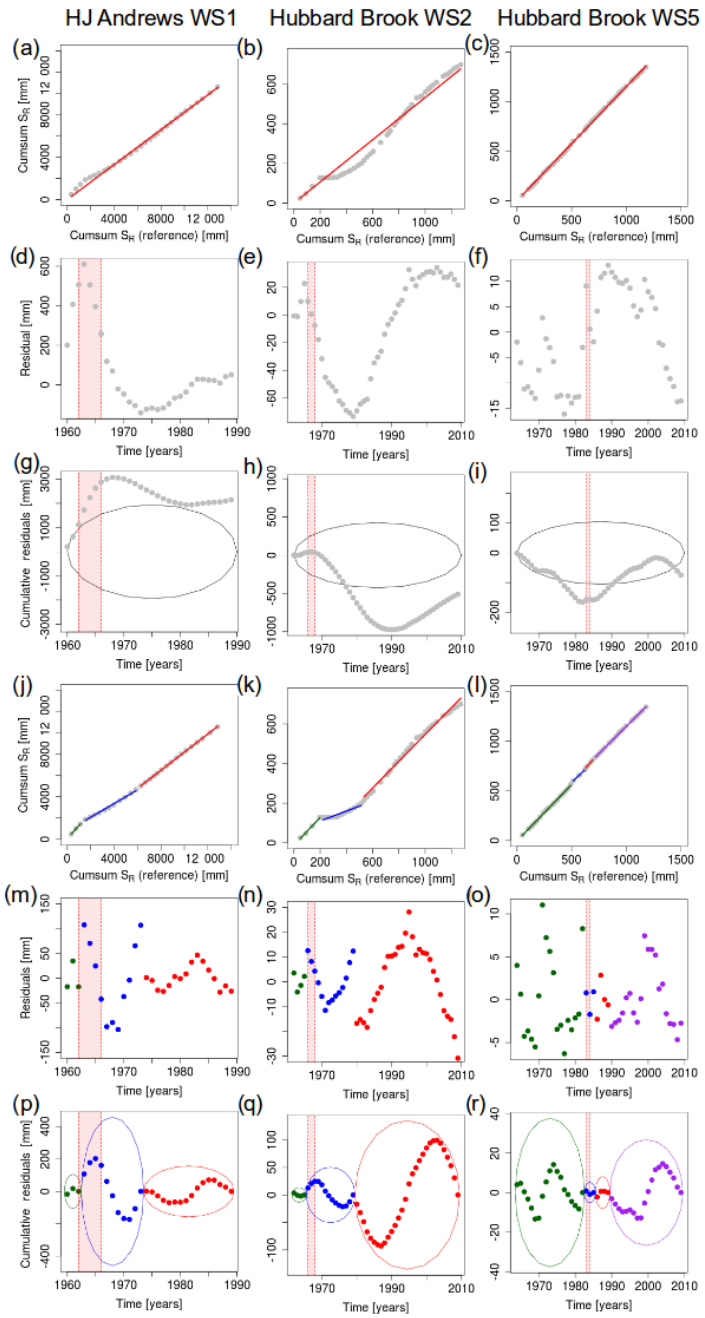

Figure 3.7: Trend analysis for $S_{R,1yr}$ in HJ Andrews WS1, Hubbard Brook WS2 and WS5 based on comparison with the control watersheds with (**a**–**c**) cumulative root-zone storages ($S_{R,1yr}$) with regression, (**d**–**f**) residuals of the regression of cumulative root-zone storages, (**g**–**i**) significance test; the cumulative residuals do not plot within the 95 %-confidence ellipse, rejecting the null-hypothesis that the two time series are homogeneous, (**j**–**l**) piecewise linear regression based on break points in residuals plot, (**m**–**o**) residuals of piecewise linear regression, (**p**–**r**) significance test based on piecewise linear regression with homogeneous time series of $S_{R,1yr}$. The different colours (green, blue, red, violet) indicate individual homogeneous time periods.

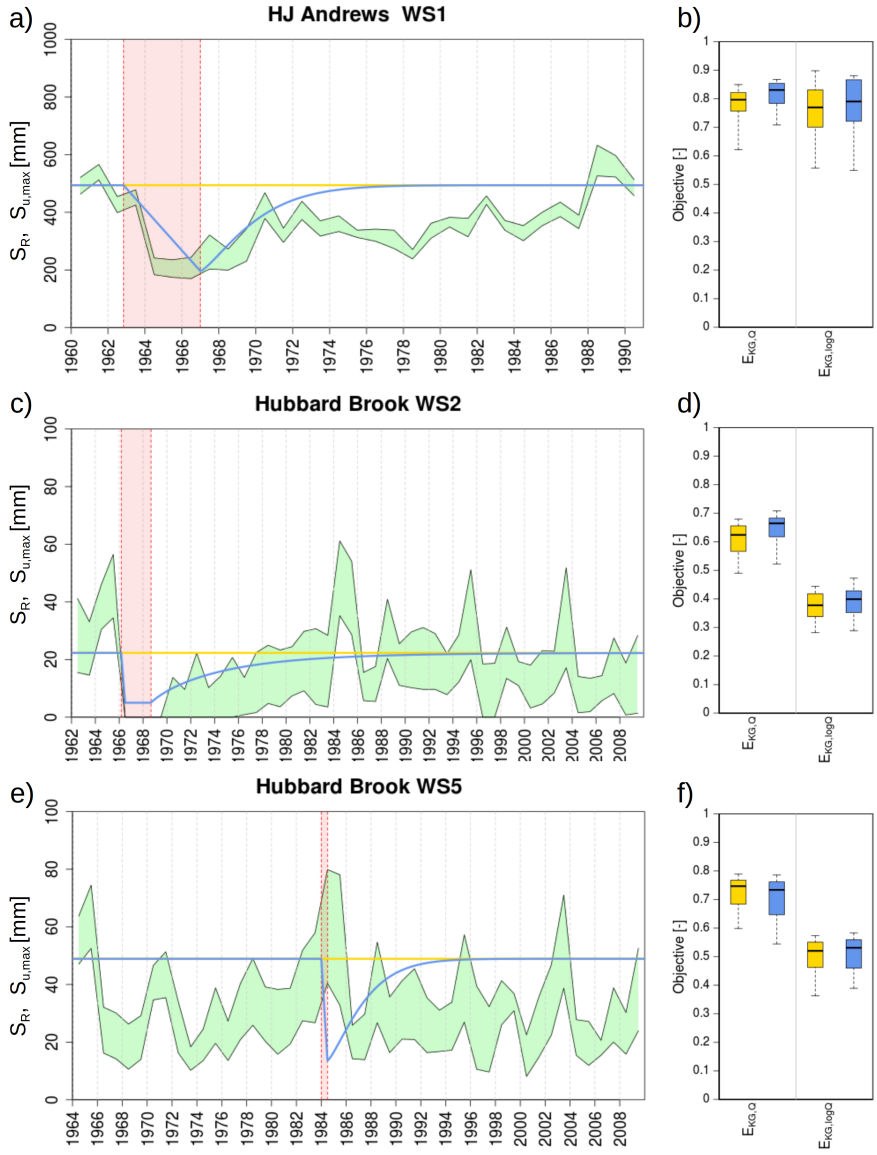

Figure 3.8: The time invariant $S_{u,\max}$ formulation represented by $S_{R,20yr}$ (yellow) and time dynamic $S_{u,\max}$ fitted Weibull growth function (blue) with a linear reduction during deforestation (red shaded area) and mean 20-year return period root-zone storage capacity $S_{R,\ 20yr}$ as equilibrium value for **(a)** HJ Andrews WS1 with $a = 0.0001\,\text{days}^{-1}$, $b = 1.3$ and $S_{R,20yr} = 494\,\text{mm}$ with **(b)** the objective function values, **(c)** Hubbard Brook WS2 with $a = 0.001\ \text{days}^{-1}$, $b = 0.9$ and $S_{R,20yr} = 22\,\text{mm}$ with **(d)** the objective function values, and **(e)** Hubbard Brook WS5 with $a = 0.001\ \text{days}^{-1}$, $b = 0.9$ and $S_{R,20yr} = 49\,\text{mm}$ and with **(f)** the objective function values. The green shaded area represents the maximum and minimum boundaries of $S_{R,1yr}$ from the water balance-based estimation, caused by the sampling of interception capacities.

The strongest improvements for calibration were observed for the dynamic formulation of FLEX for HJ Andrews WS1 and Hubbard Brook WS2 (Fig. 3.8b and d), which reflects the rather clear signal from deforestation in these catchments.

Evaluating a set of hydrological signatures suggests that the dynamic formulation of $S_{u,\max}$ allows the model to have a higher probability of better reproducing most of the signatures tested here (51 % of all signatures in the three catchments) as shown in Fig. 3.9a. A similar pattern is obtained for the more quantitative S_{RP} (Fig. 3.9b), where in 52 % of the cases improvements are observed. Most signatures for HJ Andrews WS1 show a high probability of improvement, with a maximum $P_{I,S} = 0.69$ (for $S_{Q_{95},winter}$) and an average $P_{I,S} = 0.55$. Considering the large difference between the deforested situation and the new equilibrium situation of about 200 mm, this supports the hypothesis that here a time-variant formulation of $S_{u,\max}$ does provide means for an improved process representation and, thus, hydrological signatures. Here, improvements are observed especially in the high flows in summer ($S_{Q_5,\text{summer}}$, $S_{Q_{50},\text{summer}}$) and peak flows (e.g. S_{Peaks}, $S_{\text{Peaks,summer}}$, $S_{\text{Peaks,winter}}$), which illustrates that the root-zone storage affects mostly the fast-responding components of the system.

At Hubbard Brook WS2 a more variable pattern is shown in the ability of the model to reproduce the hydrological signatures. It is interesting to note that the low flows ($S_{Q_{95}}$, $S_{Q_{95},\text{summer}}$, $S_{Q_{50},\text{summer}}$) improve, opposed to the expectation raised by the argumentation for HJ Andrews WS1 that peak flows and high flows should improve. In this case, the peaks are too high for the time-dynamic model.

The probabilities of improvement for the signatures in Hubbard Brook WS5 show an even less clear signal: the model cannot clearly identify a preference for either a dynamic or static formulation of $S_{u,\max}$ (relatively white colours in Fig. 3.9). This absence of a clear preference can be related to the observed patterns in water-balance-derived S_R (Fig. 3.4c), which also does not show a very clear signal after deforestation, indicating that the root-zone storage capacity is of less importance in this humid region characterized by limited seasonality.

3.5. Discussion

3.5.1. Changes in hydrological response dynamics

The changes found in the runoff behaviour of the deforested catchments point towards shifts in the yearly sums of transpiration, which can, except for climatic variation, be linked to the regrowth of vegetation that takes place at a similar pace to the changes in hydrological dynamics. This coincidence of regrowth dynamics and evolution of runoff coefficients was not only noticed by Hornbeck et al. (2014) for the Hubbard Brook, but was also previously acknowledged for example by Swift and Swank (1981) in the Coweeta experiment or Kuczera (1987) for eucalypt regrowth after forest fires.

Therefore, the key role of vegetation in this partitioning between runoff and transpiration (Donohue et al., 2012), or more specifically root-zones (Gentine et al., 2012), necessarily leads to a change in runoff coefficients when vegetation is removed. Similarly, Gao et al. (2014b) found a strong correlation between root-zone

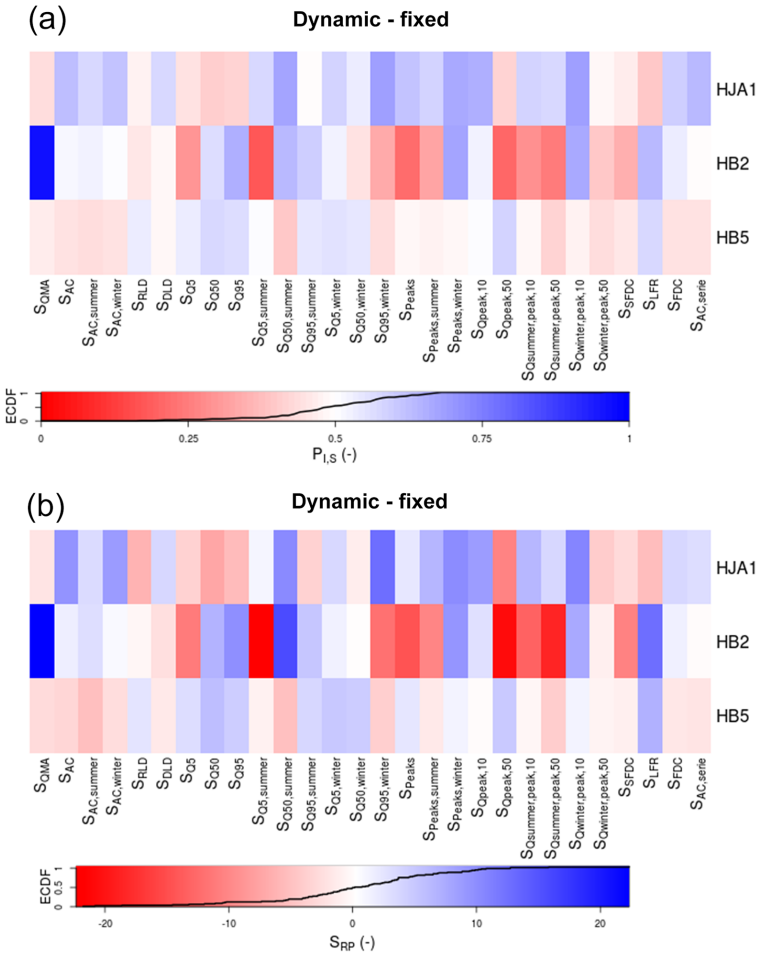

Figure 3.9: Signature comparison between a time-dynamic and time-invariant formulation of root-zone storage capacity in the FLEX model with **(a)** probabilities of improvement and **(b)** Ranked Probability Score for 28 hydrological signatures for HJ Andrews WS1 (HJA1), Hubbard Brook WS2 (HB2) and Hubbard Brook WS5 (HB5). High values are shown in blue, whereas a low values are shown in red.

storage capacities and runoff coefficients in more than 300 US catchments, which lends further support to the hypothesis that root-zone storage capacities may have decreased in deforested catchments right after removal of the vegetation.

3.5.2. Temporal evolution of S_R and $S_{u,max}$

The differences between the Hubbard Brook catchments and HJ Andrews catchments can be related to climatic conditions. In spite of the high annual precipitation volumes, high $S_{R,1yr}$ values are plausible for HJ Andrews WS1 given the marked seasonality of the precipitation in the Mediterranean climate (Köppen–Geiger class Csb) and the approximately 6-month phase shift between precipitation and potential evaporation peaks in the study catchment, which dictates that the storage capacities need to be large enough to store precipitation, which falls mostly during winter, throughout the extended dry periods with higher energy supply throughout the rest of the year (Gao et al., 2014b). At the same time, low $S_{R,1yr}$ values in Hubbard Brook WS2 can be related to the relatively humid climate, and the absence of pronounced rainfall seasonality strongly reduces storage requirements.

It can also be argued that there is a strong influence of the inter-annual climatic variability on the estimated root-zone storage capacities. For example, the marked increase in $S_{R,1yr}$ in Hubbard Brook WS2 in 1985 rather points towards an exceptional year, in terms of climatological factors, than a sudden expansion of the root-zone. It can also be observed from Fig. 3.2 that the runoff coefficient was relatively low for 1985, suggesting either increased evaporation or a storage change. A combination of a relatively long period of low rainfall amounts and high potential evaporation, as can be noted by the relatively high mean annual potential evaporation on top of Fig. 3.4b, may have led to a high demand in 1985. Parts of the vegetation may not have survived these high-demand conditions due to insufficient access to water, explaining the dip in $S_{R,1yr}$ for the following year, which is also in agreement with reduced growth rates of trees after droughts as observed by for example Bréda et al. (2006). The hydrographs of 1984–1985 (Fig. 3.6a) and 1986–1987 (Fig. 3.6b) also show that July–August 1985 was exceptionally dry, whereas the next year in August 1986 the catchment seems to have increased peak flows. This either points towards an actual low storage capacity due to contraction of the roots during the dry summer or a low need of the system to use the existing capacity, for instance to recover other vital aspects of the system.

Nevertheless, Hubbard Brook WS2 does not show a clear signal of reduced root-zone storage, followed by a gradual regrowth. Here, the forest was removed in a whole-tree harvest in winter 1983–1984, followed by natural regrowth. The summers of 1984 and 1985 were very dry summers, as also reflected by the high values of $S_{R,1yr}$. The young system had already developed enough roots before these dry periods to have access to a sufficiently large water volume to survive this summer. This is plausible, as the period of the highest deficit occurred in mid-July and lasted until approximately the end of September, thus long after the beginning of the growing season, allowing enough time for an initial growth and development of young roots from April until mid-July. In addition, the composition of the new forest differed from the old forest, with more pin cherry (*Prunus pensylvanica*) and paper

birch (*Betula papyrifera*). This supports the statements of a quick regeneration as these species have a high growth rate and reach canopy closure in a few years. Furthermore, the forest was not either treated with herbicides (Hubbard Brook WS2) or burned (HJ Andrews WS1), leaving enough low shrubs and herbs to maintain some level of transpiration (Hughes and Fahey, 1991; Martin, 1988). It can thus be argued, similar to Li et al. (2007), that the remaining vegetation experienced less competition and could increase root water uptake efficiency and transpiration per unit leaf area. This is in agreement with Hughes and Fahey (1991), who also stated that several species benefited from the removal of canopies and newly available resources in this catchment. Lastly, several other authors related the absence of a clear change in hydrological dynamics to the severe soil disturbance in this catchment (Hornbeck et al., 1997; Johnson et al., 1991). These disturbances lead to extra compaction, whereas at the same time species were changing, effectively masking any changes in runoff dynamics.

3.5.3. Trend analysis and change in hydrological regimes

The found recovery periods correspond to recovery timescales for forest systems as reported in other studies (e.g. Brown et al., 2005; Elliott et al., 2016; Hornbeck et al., 2014) which found that catchments reach a new equilibrium with a similar timescale as reported here, but in this case with the direct link to the parameter describing the catchment-scale root-zone storage capacity. The timescales are also in agreement with regression models to predict water yield after logging of Douglass (1983), who assumed a duration of water yield increases of 12 years for coniferous catchments.

The timescales found here are around 10 years (5–13 years for the catchments under consideration), but will probably depend on climatic factors and vegetation type. HJ Andrews WS1 has a recovery (6–9 years) slightly shorter compared to Hubbard Brook WS2 (10–13 years), which could depend on the different climatological conditions of the catchments. Nevertheless, it could also be argued that the spraying of herbicides had an especially strong impact on the recovery of vegetation in Hubbard Brook WS2, as the Hubbard Brook WS5 does not show such a distinct recovery signal.

3.5.4. Time-variant model formulation

It was found that a time-dynamic formulation of $S_{u,\max}$ merely improved the high and peak flow signatures for HJ Andrews WS1. Other authors also suggested previously (e.g. de Boer-Euser et al., 2016; Euser et al., 2015; Oudin et al., 2004) that the root-zone storage affects mostly the fast-responding components of the system, by providing a buffer to storm response. Fulfilling its function as a storage reservoir for plant-available water, modelled transpiration is significantly reduced post-deforestation, which in turn results in increased runoff coefficients (cf. Gao et al., 2014b), which have been frequently reported for post-deforestation periods by earlier studies (e.g. Hornbeck et al., 2014; Rothacher, 1970; Swift and Swank, 1981).

Nevertheless, signatures considering the peak flows did not improve for the Hub-

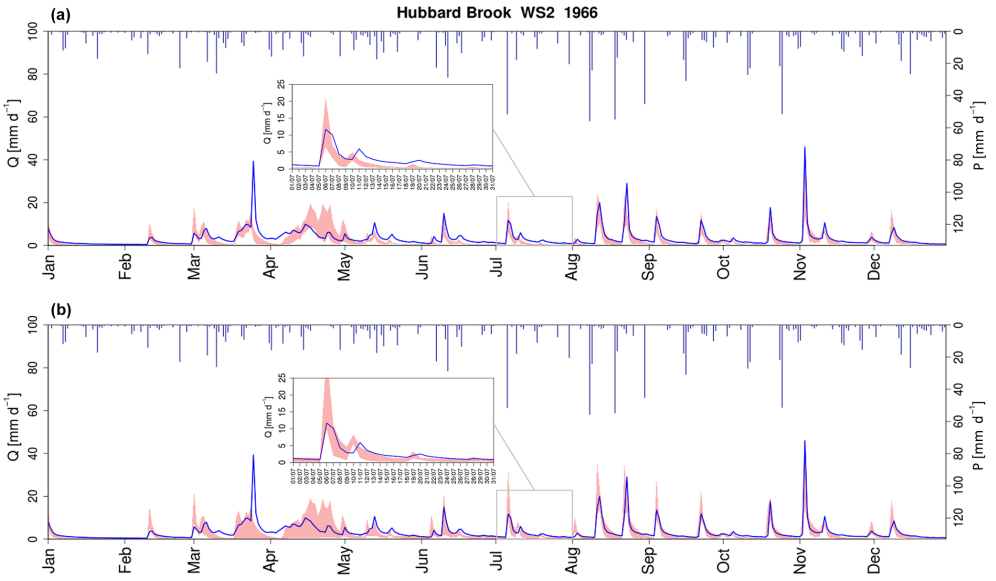

Figure 3.10: Hydrograph of Hubbard Brook WS2 with the observed discharge (blue) and the modelled discharge represented by the 5/95 % uncertainty intervals (red), obtained with **(a)** a constant representation of the root-zone storage capacity $S_{u,\max}$ and **(b)** a time-varying representation of the root-zone storage capacity $S_{u,\max}$. Blue bars indicate precipitation.

bard Brook catchments. Apparently, the model with a constant, and thus higher, $S_{u,\max}$ stored water in the root-zone, reducing recharge to the groundwater reservoir that maintains the lower flows and buffering more water, reducing the peaks. This can also be clearly seen from the hydrographs (Fig. 3.10), where the later part of the recession in the late-summer months is much better captured by the time-dynamic model. Nevertheless, the peaks are too high for the time-dynamic model, which here is linked to an insufficient representation of snow-related processes, as can be seen from the hydrograph (April–May) as well, and possibly by an inadequate interception growth function, both leading to too high amounts of effective precipitation entering the root-zone. An adjustment of these processes would have resulted in less infiltration and a smaller root-zone storage capacity.

It was acknowledged previously by several authors that certain model parameters may need time-dynamic formulations, like Waichler et al. (2005) with time-dynamic formulations of leaf area index and overstore height for the DHSVM model. In addition, Westra et al. (2014) captured long-term dynamics in the storage parameter of the GR4J model with a trend correction, in fact leading to a similar model behaviour to the Weibull growth function in this study. Nevertheless, they only hypothesized about the actual hydrological reasons for this, which aimed at the changing number of farmer dams in the catchment. The results presented here indicate that vegetation, and especially root-zone dynamics, has a strong impact on the long term non-stationarity of model parameters. The simple Weibull equation can be used as an extra equation in conceptual hydrological models to more

closely reflect the dynamics of vegetation. The additional growth parameters may be left for calibration, but can also be estimated from simple water-balance-based estimations of the root-zone storage. In this way, the extra parameters should not add any uncertainty to the model outcomes.

3.6. General limitations

The results presented here depend on the quality of the data and several assumptions made in the calculations. A limiting factor is that the potential evaporation is determined from temperature only, leading to values that may be relatively low and water balances that may not close completely. Generally, this would lead to a discrepancy between the modelled $S_{u,\max}$, where potential evaporation is directly used, and the water-balance estimates of S_R. The models will probably generate higher root-zone storages in order to compensate for the rather low potential evaporation. This can also be noted when looking at Fig. 3.4 for several models.

In addition, the assumption that the water balance closes in the 2-year periods under consideration may often be violated in reality. It can be argued that the estimated transpiration for the calculation of S_R represents an upper boundary, when storage changes are ignored. This would lead to estimates of S_R that may be lower than presented here. Nevertheless, attempts with 5-year water balances to reduce the influence of storage changes (see Fig. S9 in the Supplement of Nijzink et al. (2016b)), showed that similar patterns were obtained. Values here were slightly lower due to more averaging in the estimation of the transpiration by the longer time period used for the water balance. Nevertheless, a strong decrease after deforestation and gradual recovery can still be observed.

The issues raised here can be fully avoided when, instead of a water-balance-based estimation of the transpiration, remote sensing products are used to estimate the transpiration, similar to Wang-Erlandsson et al. (2016). However, water-balance-based estimates may provide a rather quick solution.

The transpiration estimates were also only corrected for interception evaporation, thus assuming a negligible amount of soil evaporation. Making this additional separation is typically not warranted by the available data and would result in additional uncertainty. The transpiration estimates presented here merely represent an upper limit of transpiration and will be lower in reality due to soil evaporation. Thus, the values for $S_{R,1yr}$ may expected to be lower in reality as well.

3.7. Conclusions

In this study, three deforested catchments (HJ Andrews WS1, Hubbard Brook WS2 and WS5) were investigated to assess the dynamic character of root-zone storage capacities using water balance, trend analysis, four different hydrological models and one modified model version. Root-zone storage capacities were estimated based on a simple water balance approach. Results demonstrate a good correspondence between water-balance-derived root-zone storage capacities and values obtained by a 2-year moving window calibration of four distinct hydrological models.

There are significant changes in root-zone storage capacity after deforestation,

which were detected by both a water-balance-based method and the calibration of hydrological models in two of the three catchments. More specifically, root-zone storage capacities showed, for HJ Andrews WS1 and Hubbard Brook WS2, a sharp decrease in root-zone storage capacities immediately after deforestation with a gradual recovery towards a new equilibrium. This could to a large extent explain post-treatment changes to the hydrological regime. These signals were however not clearly observed for Hubbard Brook WS5, probably due to soil disturbance, a new vegetation composition and a climatologically exceptional year. Nevertheless, trend analysis showed significant differences for all three catchments with their corresponding, undisturbed reference watersheds. Based on this, recovery times were estimated to be between 5 and 13 years for the three catchments under consideration.

These findings underline the fact that root-zone storage capacities in hydrological models, which are more often than not treated as constant in time, may need time-dynamic formulations with reductions after logging and gradual regrowth afterwards. Therefore, one of the models was subsequently formulated with a time-dynamic description of root-zone storage capacity. Particularly under climatic conditions with pronounced seasonality and phase shifts between precipitation and evaporation, this resulted in improvements in model performance as evaluated by 28 hydrological signatures.

Even though this more complex system behaviour may lead to extra unknown growth parameters, it has been shown here that a simple equation, reflecting the long-term growth of the system, can already suffice for a time-dynamic estimation of this crucial hydrological parameter. Therefore, this study clearly shows that observed changes in runoff characteristics after land-cover changes can be linked to relatively simple time-dynamic formulations of vegetation-related model parameters.

4

Constraining models with multiple information sources

The increasing amount of data leads to more opportunities to identify suitable model formulations and parameterizations, in addition to model regularization and simple parameter estimation techniques. Especially as the number of available products increase, it becomes highly urgent to exploit multiple data sources simultaneously, and extract the relevant information from it. Therefore, instead of assessing the added value of the products one by one, the added value of the different combinations of products is assessed, ranging from using one to ten products, in order to obtain an increased understanding how the simultaneous use of several data sources can lead to a robust model parameterization.

This chapter is based on:
Nijzink, R.C., Almeida, S., Pechlivanidis, I., Capell, R., Gustaffsons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Savenije, H., and Hrachowitz, M.: *Constraining conceptual hydrological models with multiple information sources*, Water Resources Research, submitted, 2017.

4.1. Introduction

Computational techniques have been advancing and an abundance of new sources of information have become available over the recent years, but selecting meaningful parameters for catchment-scale hydrological models, in particular for predictions in ungauged catchments, remains problematic (Blöschl et al., 2013; Hrachowitz et al., 2013; Sivapalan, 2003), and is further exacerbated by the world-wide ongoing reductions of stream gauging networks (Fekete and Vörösmarty, 2002; Hannah et al., 2011; Sivapalan, 2003).

The dependency on streamflow data for model calibration can, to a certain extent, be reduced by directly estimating individual model parameters (or at least defining non-uniform parameter prior distributions) from exploiting their links with readily available observations of other quantities than streamflow, which are observable at the scale of the model application, such as topographic considerations (Smith et al., 2016) or the long-term water balance (e.g. de Boer-Euser et al., 2016; Gao et al., 2014b; Nijzink et al., 2016a). Similarly, when no streamflow observations are available, traditional regionalization techniques use climatic and physiographic data, to establish transfer functions that allow an indirect estimation of the actual model parameters (Götzinger and Bárdossy, 2007; Hundecha and Bárdossy, 2004; Hundecha et al., 2016; Merz and Blöschl, 2004; Samaniego et al., 2010b; Wagener and Wheater, 2006). Alternatively, catchment signatures can be used to condition the prior parameter space, as an alternative to streamflow calibration (Almeida et al., 2016; Bárdossy, 2007; Bulygina et al., 2009; Castiglioni et al., 2010, 2011; Yadav et al., 2007). In addition, model constraints or limits of acceptability (Beven, 2006), based on qualitative information without clear target values, often referred to as "soft data" (Seibert and McDonnell, 2002; Winsemius et al., 2009) or "expert knowledge" (Gharari et al., 2014; Kelleher et al., 2017), meant to avoid physically implausible representations of the system, were in the past shown to be valuable to limit the feasible model parameter space (i.e. Freer et al., 2004; Hrachowitz et al., 2014). These constraints can be implemented either as *a priori* defined inequality constraints on parameters or on processes (Ambroise et al., 1996). The latter allows to contain the dynamics of individual model components to some degree (cf. Gharari et al., 2014; Wagener and Montanari, 2011), such as limiting long-term evaporation to values expected from the Budyko curve (e.g. Gerrits et al., 2009).

The increasing availability of remotely sensed data may provide ample opportunities to further constrain hydrological models and their parameters. While several recent reviews highlight their potential for applications in hydrology (e.g. AghaKouchak et al., 2015; Hrachowitz and Clark, 2017; Pechlivanidis and Arheimer, 2015; Xu et al., 2014), it can also be argued that remotely sensed high-resolution streamflow data are rather far from becoming reality (Lettenmaier et al., 2015). Although successful attempts of using remotely sensed streamflow for model calibration have been reported (e.g. Sun et al., 2015; Tourian et al., 2017), the specific orbits of the observation satellites lead to spatial and temporal limitations, and only larger rivers can be monitored due to the large resolution. In contrast, products providing estimates of evaporation have in the past been shown to have considerable value

for model applications, as summarized by several studies that point at the different advantages and disadvantages of these products (e.g. Verstraeten et al., 2008; Zhang et al., 2016). Besides evaporation products, the central importance of soil moisture and snow storage for the Earth's water cycle made it a focus of research efforts in the remote sensing community, which developed several satellite missions dedicated to soil moisture and snow cover mapping, such as the Soil Moisture and Ocean Salinity (SMOS; Kerr et al., 2012), Soil Moisture Active and Passive (SMAP; Brown et al., 2013) or NASA's Earth Observing System (Greenstone and King, 1999) missions. Furthermore, the Gravity Recovery and Climate Experiment (GRACE; Tapley et al., 2004) led to new, valuable information on total water storage based on remotely sensed gravity anomalies. These are just a few examples, while more remotely sensed products are currently available and new satellite missions are planned (e.g. GRACE-FO, SWOT), which will further increase the information available for hydrological modelling.

The challenge remains, though, how to select data that are suitable for use in hydrological model applications and to assess how they can support the modelling process in a meaningful and effective way. So far, information from remote sensing has been incorporated in applications of hydrological models in several ways. For example, data assimilation techniques are commonly used to update the states of a model (e.g. Liu and Gupta, 2007; Liu et al., 2012; Reichle, 2008). This can help to improve internal model dynamics and the resulting hydrological predictions (Crow and Ryu, 2009; Tangdamrongsub et al., 2015). Yet, it can be argued that the added value of data assimilation is actually an indicator of inadequate model parameters and/or model formulations (Spaaks and Bouten, 2013). Alternatively and directly addressing this issue, remotely sensed data can be directly used as calibration variables and thus to select feasible model parameters (e.g. Immerzeel and Droogers, 2008; Lopez Lopez et al., 2017; Pechlivanidis and Arheimer, 2015; Sutanudjaja et al., 2014). Although the above strategies are in principle a valid way forward, spatial and temporal mismatches between hydrological models and remotely sensed data (Vereecken et al., 2008; Xu et al., 2014) place some limitations on the value of these data. Similarly, hydrological variables are not directly observed by most remote sensors, but rather inferred from models that link the observed variable with some hydrologically relevant variable, thus introducing an additional source of uncertainty.

A large number of studies previously assessed the added value of different remote sensing products, either for data assimilation or model calibration. These studies generally focused either on a single remote sensing product, for example GRACE (e.g. Lo et al., 2010; Mulder et al., 2015; Rakovec et al., 2015; Werth et al., 2009), SCAT soil moisture (e.g. Parajka et al., 2009) and ASCAT soil moisture (Brocca et al., 2010), or on one single model state or flux with a combination of products such as soil moisture (Wanders et al., 2014). Nevertheless, the combined effects of several products, that deal with multiple model states and fluxes simultaneously, has only recently gained some attention, but this remains rather limited to two different model states or fluxes (Kunnath-Poovakka et al., 2016; Lopez Lopez et al., 2017; Tian et al., 2017). Full bootstrap procedures where multiple combina-

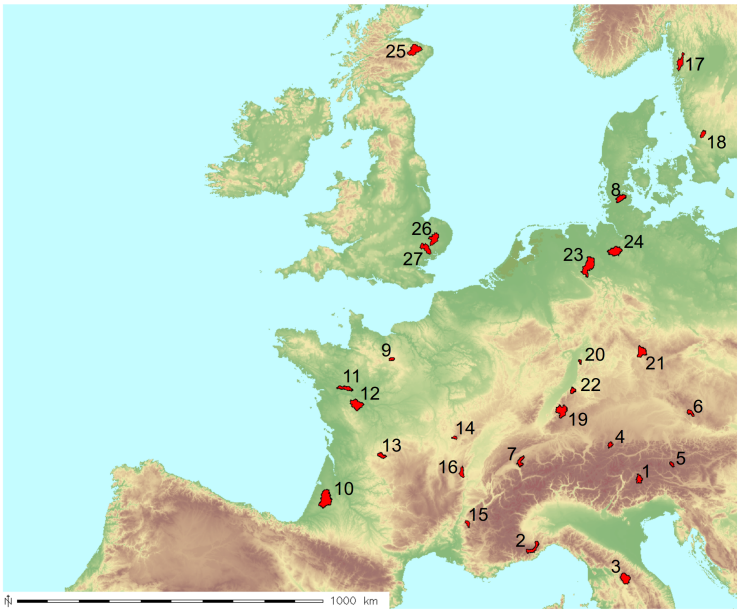

Figure 4.1: The 27 study catchments and their location in Europe. See Table 4.1 for the catchment characteristics

tions of remote sensing products are tested have not been reported so far.

Thus, the objective of this paper is to explore the value of combining several types of remotely sensed data products that reflect different water balance components, to effectively and consistently constrain the parameter space of five different conceptual hydrological models. Specifically, we test the hypotheses that the combined use of different remote sensing products can 1) produce model performances similar to traditional calibration on streamflow and 2) improve the representation of model internal dynamics and hydrological signatures in comparison with traditional calibration on streamflow.

4.2. Methodology

A detailed stepwise description of this experiment, with the model codes and links to the data, can be found in an online experiment protocol (`http://dl-ng005.xtr.deltares.nl/view/66/`) as part of the SWITCH-ON Virtual Water Science Laboratory. This protocol is developed to facilitate full experiment reproducibility and repeatability, according to the requirements suggested by Ceola et al. (2015).

4.2.1. Study Areas

A set of 27 European catchments was selected in order to cover a variety of landscapes, climates and vegetation. The study sites included lowland catchments in the

UK and Germany as well as more mountainous catchments in Austria and France. The study catchments also exhibit considerable climatic differences with aridity indices ranging from 0.5 – 1.1 and mean areal precipitation from 627 – 1593 mm/year. Estimates of daily potential evaporation were derived from data (air temperature, dewpoint temperature, wind speed, long wave radiation, short wave radiation) from ERA-Interim (Dee et al., 2011) according to the Penman formulation as prescribed by FAO (Allen et al., 1998), the air temperature of the ERA-Interim data was also applied as forcing data for the snow modelling. Daily precipitation was derived from the MSWEP dataset (Beck et al., 2017). Time series of streamflow covering recent years and with sufficient length (approx. 10 years of data) were obtained from the Global Runoff Data Centre (GRDC). An overview of the catchments is provided in Table 4.1 and Figure 4.1.

4.2.2. Models

Five different rainfall-runoff models were applied to account for different model structures, which are briefly described here. For more details about model structures, parameters and prior parameter ranges the reader is referred to the Supplementary Material S1 of Nijzink et al. (2017).

FLEX

The FLEX model (Fenicia et al., 2008) is a lumped model that consists of four storage components and a snow module. The snow module, based on a degree-day approach, runs first and determines the effective precipitation consisting of rainfall and snowmelt. After this, the water enters an interception reservoir, from which intercepted water can evaporate and/or leave the reservoir after exceeding a certain threshold. The remaining precipitation after interception is split into runoff and infiltration in the subsequent step. The infiltrated water is stored in the soil moisture reservoir, from which transpiration takes place. A portion of the runoff goes to a fast reservoir, another portion to the groundwater reservoir through preferential percolation. The model uses 8 parameters which are left free for calibration.

FLEXtopo

The FLEXtopo model (Savenije, 2010) uses hydrological response units based on different landscape elements to capture the core processes for different parts in the landscape. In this set-up, the landscape units were defined as plateau, hillslope and wetland, similar to previous applications (de Boer-Euser et al., 2017; Gao et al., 2014a; Gharari et al., 2014). For each model unit, a snow routine is followed by an interception reservoir and unsaturated reservoir. For plateau landscapes, recharge to the groundwater can happen through matrix percolation, as a function of soil moisture, and preferential percolation through macropores or cracks and fissures. In contrast, the hillslope areas are only assumed to contribute to the groundwater through preferential percolation and the wetlands even receive water from the groundwater reservoir through capillary rise. In the original application (Gao et al., 2014a) FLEXtopo uses proportionalities between parameters of different landscape classes (e.g. interception capacity of forest bigger than grass), which

Table 4.1: Overview of the catchments.

Catchment	Area (km^2)	Aridity index (EP/P) (–)	Prec. ($mm\,yr^{-1}$)	Runoff ($mm\,yr^{-1}$)	Calibration period
Gadera	394	0.52	1095	670	01/10/1999-30/09/2009
Tanaro	500	0.81	1059	625	01/10/2003-30/09/2012
Arno	751	0.81	1069	398	01/10/2003-30/09/2013
Vils	198	0.32	1593	1338	01/10/2000-30/09/2010
Grossarler	144	0.36	1508	1202	01/10/2000-30/09/2010
Grosse Muehl	253	0.55	1156	739	01/10/2000-30/09/2010
Broye	396	0.50	1219	610	01/10/1999-30/09/2009
Treene	481	0.71	917	435	01/10/1999-30/09/2004
Risle	147	1.0	751	304	01/10/2001-30/09/2011
Leyre	1587	0.98	913	283	01/10/2001-30/09/2011
Erdre	463	0.98	789	181	01/10/2001-30/09/2011
Layon	928	1.10	694	105	01/10/2006-30/09/2011
Glane	297	0.80	981	401	01/10/2001-30/09/2011
Dragne	117	0.75	996	435	01/10/2002-30/09/2011
Roubion	190	0.87	920	276	01/10/2001-30/09/2011
Azergues	333	0.84	887	317	01/10/2001-30/09/2011
Enning-Dalsaelven	634	0.57	978	646	01/10/2005-30/09/2014
Fyllean	263	0.62	937	749	01/10/2004-30/09/2014
Kinzig	955	0.49	1344	744	01/10/2002-30/09/2012
Modau	91	0.99	705	234	01/10/2002-30/09/2012
Rodach	716	0.76	830	458	01/10/2002-30/09/2012
Pfinz	232	0.74	920	248	01/10/2003-30/09/2013
Hunte	1409	0.80	806	219	01/10/2002-30/09/2012
Wuemme	934	0.77	855	342	01/10/2002-30/09/2012
Deveron	955	0.49	1030	636	01/10/2002-30/09/2012
Little Ouse	757	1.05	627	143	01/10/2004-30/09/2012
Stour	657	1.03	632	153	01/10/2002-30/09/2012

limit the feasible parameter space. Here, in this comparative analytical framework additional conditions were not implemented for FLEXtopo, leaving a relatively wide parameter space. In total, 24 parameters are left free for calibration.

HYMOD

The HYMOD model (Boyle, 2001; Wagener et al., 2001) runs first a snow module from which rainfall and snow melt continue towards the unsaturated zone. Here, evaporation is determined as a function of soil moisture and runoff is generated, based on a distribution function of maximum storage capacities in the catchment (often referred to as beta-function). This runoff is divided over a series of fast reservoirs and one slow reservoir. The contributions of the fast flows and slow flows eventually determine the final streamflow. In total, 8 parameters are free for calibration.

HYPE

The HYPE model (Lindström et al., 2010) runs first a snow module, after which the model structure contains three soil layers with assigned soil depths. Water can evaporate from the first two layers, and runoff is generated when the maximum storage capacity of these layers is reached or when maximum infiltration capacities are exceeded. Water can percolate downwards through matrix flow or preferentially through fast flow paths. The lowest soil layer reflects the groundwater contribution to the streamflow and an additional aquifer routine can be applied. Eventually, a routing function is applied to the total outflows to obtain the final streamflow. A set of 22 model parameters was selected for optimization in the parameter selection procedures.

TUW

The TUW model (Parajka et al., 2007) uses a similar model structure as originally applied in the HBV model (Bergström, 1992). First, a snow routine is run based on a degree-day approach, after which water enters the soil moisture reservoir and becomes available for evaporation. Here, evaporation is determined as a function of soil moisture, and runoff is generated based on a distribution function of maximum storage capacities as well. The runoff continues to a fast reservoir, which has an additional overflow outlet to represent a very fast component. Percolation from the fast reservoir towards the slow, groundwater reservoir takes place subsequently. In a last step, the sums of slow and fast runoff components are routed through the system with a triangular lag functions. The TUWmodel has 15 parameters free for calibration.

4.2.3. Data sources for constraining parameters

Nine different remote sensing products and an analytical framework from four functionally similar groups were tested in this study for their information content to select meaningful model parameters and thus to constrain the feasible parameter space. Each group provides information about a different component of the hydrological system: (1) soil moisture, (2) evaporation, (3) total water storage and

Table 4.2: Details of the remote sensing products.

Product	Version)	Spatial Resolution	Temporal Resolution	Reference	Model state/flux	Performance metric
AMSR-E LPRM	V2	25km	daily	Owe et al. (2008)	Soil moisture	Squared correlation coefficient
ASCAT-SWI	V3	0.1°	daily	Wagner et al. (1999)	Soil moisture	Squared correlation coefficient
SMOS	V620	Ca. 15km	2-3 days	Kerr et al. (2012)	Soil moisture	Squared correlation coefficient
NDII	V6	500m	daily	Sriwongsitanon et al. (2016)	Soil moisture	Squared correlation coefficient
Budyko				Budyko (1974)	Evaporation	Squared correlation coefficient
LSA-SAF		3km	daily	Ghilain et al. (2011)	Evaporation	Squared correlation coefficient
MOD16	V5	500m	8-day	Mu et al. (2011)	Evaporation	Squared correlation coefficient
GRACE		100km	30days	Tapley et al. (2004)	Total water storage	Squared correlation coefficient
MOD10	V5	500m	daily	Hall et al. (2002)	Snow	Squared correlation coefficient
MYD10	V5	500m	daily	Hall et al. (2002)	Snow	Squared correlation coefficient

(4) snow accumulation. A general overview of the used products and the main specifications can be found in Table 4.2.

The first group contains soil moisture estimates from four different remote sensing products. One of the soil moisture products used in this study is derived from the Advanced SCATterometer (ASCAT) on board the Metop satellite, which uses C-band (5.255 GHz) to estimate surface soil moisture. Scatterometer data processed with the algorithm provided by Wagner et al. (1999) was used in this experiment, representing the Soil Water Index (SWI) or the relative soil moisture in the root zone. The second soil moisture product comes from the Land Parameter Retrieval Model (LPRM; Owe et al., 2008) with data from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E, with C and X-band) as input, and represents the top 2-3 cm of soil moisture. The last soil moisture product explored in this study is obtained from the Soil Moisture and Ocean Salinity (SMOS; Kerr et al., 2012) mission, also representing the soil moisture in the upper centimeters of the soil (L-band, 1-2 GHz). In addition, the Normalized Difference Infrared Index (NDII) was calculated based on MODIS images, as recent results suggest a link to root zone soil moisture storage (Sriwongsitanon et al., 2016). Even though most of the products represent only soil moisture in the top soil, the products were directly compared to the soil moisture states of the models, without adjustments or exclusions of specific days (e.g. excluding snow days). Therefore, it was assumed that at least a linear relationship exists between modelled soil moisture state and the observations of the soil moisture products, even though these do not represent exactly the same soil moisture state as the model. Thus, the squared correlation coefficient was used as a performance metric for all soil moisture products.

The second group contains evaporation estimates from two remote sensing products and the Budyko framework. Specifically, the daily product from the EUMETSAT's Land Surface Analysis – Satellite Application Facility (LSA-SAF; Ghilain et al., 2011) was selected as well as the MOD16 product from the MODerate Resolution Imaging Spectroradiometer (MODIS; Mu et al., 2011). The MOD16 8-day evaporation product is based on a Penman-Monteith approach, and the final product consists of soil evaporation, transpiration and interception evaporation from the canopy. The LSA-SAF evaporation product uses a similar Penman-Monteith approach, but the products differ in, amongst other things, formulations of aerodynamic and stomatal resistances, the (absence of) explicit accounting for interception evaporation, the temporal resolution of the satellite as well as the differences in other technical satellite specifications. Besides these remotely sensed daily and 8-daily products, the analytical Budyko framework (Budyko, 1974) was applied to obtain an additional long-term estimate of evaporation as model constraint. Also for this group of products, the squared correlation coefficient was used as a performance metric, only for the Budyko framework the relative error was used as this only comprises a single value instead of a time series. For the MOD16 8-day evaporation product the comparison was made for the modelled 8-day evaporation as well.

The third group provides estimates of changes in total water storage from one remote sensing product. To do that, data on gravity anomalies from the Gravity Re-

covery and Climate Experiment (GRACE; Tapley et al., 2004) were linked to water storage fluctuations. Similar to Rakovec et al. (2016) the GRACE data from the three processing centers of CSR (Center for Space Research, University of Texas, USA), GFZ (Geoforschungs Zentrum Potsdam, Germany) and JPL (Jet Propulsion Laboratory, USA) were merged into one combined product for the analysis. The GRACE water storage anomalies relative to the baseline of 2004-2009 were also corrected to represent the anomalies over the time series under consideration, which differed per study catchment. The squared correlation coefficient between modelled total water storage anomalies and GRACE storage anomalies was used as performance metric again.

The fourth group contains information on snow accumulation and depletion patterns from two remote sensing products. The MODIS Terra Snow Cover (Hall et al., 2002) and MODIS Aqua Snow Cover (Hall et al., 2002) products both provide fractional snow cover on a daily basis and 500m resolution. This snow cover is determined by the Normalized Difference Snow Index (NDSI) and relies on the fact that while clouds have large reflectance in both visible and infrared bands, snow has a larger reflectance only in the visible domain. The average was taken over all non-cloud covered cells over a catchment, but this was only done when the cloud coverage did not exceed 60% of the catchment, similar to Parajka and Blöschl (2008), in order to determine reliable snow coverage values. These catchment averaged snow coverage values were compared with modelled snow water equivalents (FLEX, FLEXtopo, TUW), assuming a linear relation between coverage and snow water equivalent. Only for HYPE, comparing modelled and observed snow coverage was directly possible due to the more extensive snow module. The squared correlation coefficient was again computed as performance metric.

In general, all products were processed in order to be compatible with the model scales, which was in most cases the catchment scale, as the models were applied in a lumped manner. Thus, the average of the cells covering the catchment was determined and used in the analysis. Only for FLEXtopo, the products were averaged over a subarea of the catchment, the landscape units, whereas the other models all used catchment averaged values. For example, the LSA-SAF evaporation was averaged over the catchment area defined as plateau landscape, in order to compare with the modeled evaporation from the plateau model structure.

4.2.4. Parameter selection - remote sensing data

As a first step, random parameter sets were generated for each of the five models. This was done using Latin Hypercube sampling to achieve a somewhat homogenous exploration of the respective parameter spaces. The parameters were sampled from uniform prior distributions with parameter ranges set as wide as possible without becoming unrealistic. FLEX and HYMOD were sampled 80,000 times, whereas FLEXtopo, HYPE and TUW were sampled 100,000 times due to the larger number of free calibration parameters. These parameter sets were then used together with the daily input data to generate either 80,000 or 100,000 model realizations per model for each catchment, covering a time period of approximately 10-years (see Table 4.1).

Table 4.3: Model parameters and the data sources that are related to it for determining posterior parameter bounds.

	Source	FLEX	FLEXtopo	HYPE	HYMOD	TUW
Soil moisture	**AMSR-E**	$Mmelt$ $Tthresh$ $Imax$ $Sumax$ $Beta$ Kf Ks D	$Mmelt$ $Tthresh$ $Imax_p$ $Imax_h$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$ $Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$ $Rrsc2$ $macrate$	Ts $Cfmax$ CRF CWH Sm $Beta$ $Alfa$ Rs Rf	Csf Ddf Tr Ts $Meltt$ FC $Beta$ $lprat$ $K2$ $cperc$
	ASCAT	$Mmelt$ $Tthresh$ $Imax$ $Sumax$ $Beta$ Kf Ks D	$Mmelt$ $Tthresh$ $Imax_p$ $Imax_h$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$ $Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$ $Rrsc2$ $macrate$	Ts $Cfmax$ CRF CWH Sm $Beta$ $Alfa$ Rs Rf	Csf Ddf Tr Ts $Meltt$ FC $Beta$ $lprat$ $K2$ $cperc$
	NDII	$Mmelt$ $Tthresh$ $Imax$ $Sumax$ $Beta$ Kf Ks D	$Mmelt$ $Tthresh$ $Imax_p$ $Imax_h$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$ $Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$ $Rrsc2$ $macrate$	Ts $Cfmax$ CRF CWH Sm $Beta$ $Alfa$ Rs Rf	Csf Ddf Tr Ts $Meltt$ FC $Beta$ $lprat$ $K2$ $cperc$
	SMOS	$Mmelt$ $Tthresh$ $Imax$ $Sumax$ $Beta$ Kf Ks D	$Mmelt$ $Tthresh$ $Imax_p$ $Imax_h$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$ $Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$ $Rrsc2$ $macrate$	Ts $Cfmax$ CRF CWH Sm $Beta$ $Alfa$ Rs Rf	Csf Ddf Tr Ts $Meltt$ FC $Beta$ $lprat$ $K2$ $cperc$

Table 4.3: Model parameters and the data sources that are related to it for determining posterior parameter bounds (continued).

	Source	FLEX	FLEXtopo	HYPE	HYMOD	TUW
Evaporation	**Budyko**	$Imax$ $Sumax$ $Beta$	$Imax_p$ $Imax_h$ $Imax_w$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$	Sm $Beta$	FC $BETA$ $Lprat$
	LSA-SAF	$Imax$ $Sumax$ $Beta$	$Imax_p$ $Imax_h$ $Imax_w$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$	Sm	FC $BETA$ $Lprat$
	MOD16	$Imax$ $Sumax$ $Beta$	$Imax_p$ $Imax_h$ $Imax_w$ $Sumax_p$ $Sumax_h$ $Sumax_w$	$wcfc$ lp $mactrinf$	Sm	FC $BETA$ $Lprat$
Snow	**MOD10**	$Mmelt$ $Tthresh$	$Mmelt$ $Tthresh$	$Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$	Ts $Cfmax$ CFR CWH	Csf Ddf Tr Ts $Meltt$
	MYD10	$Mmelt$ $Tthresh$	$Mmelt$ $Tthresh$	$Ttpd$ $Ttpi$ $Ttmp$ $Cmlt$	Ts $Cfmax$ CFR CWH	Csf Ddf Tr Ts $Meltt$

Subsequently, it was evaluated how well the modelled state and/or flux variables of each model realization were able to reproduce the different data from the group of remote sensing products that corresponds to that specific model state or flux. For example, modelled evaporation for each of the 80,000(100,000) model realizations for each catchment was evaluated against the different evaporation estimates provided by the data from group two (see section 4.2.3). The squared correlation coefficient was used as a performance metric for model evaluation against each remote sensing product, emphasizing the models' ability to reproduce the temporal dynamics of a given variable, but ignoring the magnitude of the variable itself and its fluctuations over time. For evaluation against long-term evaporation from the Budyko curve the relative error was used. See also Table 4.2 for which performance metric was used per product.

In a next step, posterior parameter distributions were determined based on a weighting procedure (e.g. Freer et al., 1996). Initially, this was carried out with a single performance metric related to a specific product for all model parameters, to assess the sensitivity of the parameters for a given product. The vertical distance between the cumulative distribution of the posterior parameter distribution and the prior, uniform parameter distribution (i.e. the Kolmogorov-Smirnoff distance), both averaged for all the catchments, served as an informal indication of sensitivity. This sensitivity was combined with simple reasoning (e.g. snow performance metrics should relate to snow parameters) to relate parameters to relevant performance metrics. Thus, the final weighting was carried out on the basis of one or a combination of m selected performance metrics to model states/fluxes, assumed to be relevant for the parameter under consideration. Table 4.3 gives an overview of the parameters and the remotely sensed data sources that these parameters were eventually linked to. For example, the posterior parameter distributions linked to snow processes were calculated using weights derived from the model's ability to reproduce the satellite snow cover data (the snow products, see section 4.2.3). In the case that multiple products were used for the evaluation of a model component and the construction of the posterior distributions of the associated parameters, a combined performance metric was formulated, based on the Euclidean distance of the model performances with respect to the individual products (Eq. 4.1), thus treating the performance metrics equally important. Hereafter, rescaling was applied (Eq. 4.2) to maintain values between 0 and 1.

$$E_{obj,combined} = 1 - \sqrt{\left(1 - E_{\text{obj,1}}\right)^2 + \ldots + \left(1 - E_{\text{obj,n}}\right)^2}, \tag{4.1}$$

$$L\left(\theta\right) = E_{obj,combined,scaled} = \frac{E_{\text{obj,combined}} - min\left(E_{\text{obj,combined}}\right)}{max\left(E_{\text{obj,combined}}\right) - min\left(E_{\text{obj,combined}}\right)} \tag{4.2}$$

Where $E_{\text{obj, combined}}$ is the combined objective function (or performance metric) using m remote sensing products for evaluation, $E_{\text{obj,m}}$ is the objective for product *m*, $E_{obj,combined,scaled}$ is the scaled objective and $L(\theta)$ is an informal likelihood

weight for parameter set θ. The posterior distributions were then determined as (Freer et al., 1996):

$$L_2(\theta) = L(\theta)^n \times L_0(\theta)/C, \tag{4.3}$$

where $L_0(\theta)$ is the prior parameter distribution [-], $L_2(\theta)$ is the posterior probability distribution [-] , n is a weighting factor (set to 10) [-], and C is a normalizing constant [-], the parameter set θ consists of the relevant parameters of interest for the informal weights, and the other parameters not sensitive for obtaining the formal weights.

This procedure was repeated for all possible combinations of remote sensing products from the four groups (Table 4.3), starting with a single product and ending with the combined use of all 10 products simultaneously. This resulted in a total of 1023 different possible combinations of remote sensing products for the evaluation of the associated model components. After weighting, the 25^{th} and 75^{th} quartiles of the posterior parameter distributions were retained as feasible parameter bounds. This remains a mere subjective choice, but in this way the higher values of the posterior likelihood are always retained and the new parameter ranges are always determined on a sufficient number of samples. The alternative of cutting off the distributions at a certain performance level, as for example done in the GLUE methodology (Beven and Freer, 2001), is equally subjective, with the additional risk of not having any feasible solutions to determine posteriors. It is fully acknowledged that in absence of a clear posterior distribution the new bounds may incorrectly consider some solutions unfeasible (i.e. false negatives), but therefore, the weighting factor n was set relatively high ($n = 10$). In this way, and in combination with the use of the 25^{th} and 75^{th} quartiles as bounds, considerable discriminative power can be obtained, zooming in on the solutions which are considered correct.

4.2.5. Parameter selection – streamflow observations

All five models for all 27 study catchments were calibrated in a traditional way on observed streamflow data to provide a reference benchmark. Thus, all model realizations were evaluated against streamflow with a multi-objective strategy based on the Nash-Sutcliffe efficiency ($E_{\text{NS,Q}}$) of flow and the Nash-Sutcliffe efficiency of the logarithm of the flow ($E_{\text{NS,logQ}}$). Also here, the two objective functions were combined based on the Euclidean distance of these two objective functions:

$$E_{obj,combined} = 1 - \sqrt{\left(1 - E_{\text{NS,Q}}\right)^2 + \left(1 - E_{\text{NS,logQ}}\right)^2}, \tag{4.4}$$

Where $E_{\text{obj, combined}}$ the combined objective. Model runs were retained as feasible when both $E_{\text{NS,Q}}$ and $E_{\text{NS,logQ}}$ were higher than 0. For clarity, hereafter the term *calibration* refers to the traditional parameter selection procedure based on streamflow as outlined here, whereas *constraining* refers to the parameter selection procedure based on the remote sensing data sources as outlined in section 4.2.4.

4.2.6. Added value for streamflow

The added value of the individual remote sensing products was assessed by computing the probability that improvement occurred when including a specific data source (see section 4.2.2) to constrain the feasible parameter space of a given model, compared to not including this specific data source. Improvement was considered here in terms of the Euclidian distance between Nash-Sutcliffe efficiency of the flows and logarithm of the flows. To do so, all 1023 possible combinations of the one to ten potential data sources from the four groups specified in section 4.2.2 were separated in combinations *with* and *without* a specific product, leading to 512 combinations *with*, and 511 combinations *without* this product. Each combination has its own set of feasible solutions (as defined by the posterior parameter distributions) with the associated performance measures (such as $E_{\text{NS,Q}}$ and $E_{\text{NS,logQ}}$). The overall probability of improvement of including one product is then estimated by merging the distribution of performance measures for each combination *with* a specific product with the distributions of performance measures of all other combinations with this product into the marginal probability of improvement when using this product (Figure 4.2). For example, if combinations A, B and C each included the remote sensing product GRACE, with 100, 200, 250 feasible solutions, respectively, the final distribution contained 550 feasible solutions ($100 + 200 + 250$). Following the same approach for all combinations $without$ this product, a second marginal distribution of performances is established. Comparing the overall distribution with this product to the one without this product, the probability of improvement ($P_{I,S}$) can then be calculated according to Nijzink et al. (2016b):

$$P_{\text{I,S}} = P(S_{with} > S_{without}) = \sum_{i=1}^{n} P(S_{with} > S_{without} | S_{with} = r_i) P(S_{with} = r_i), \quad (4.5)$$

where S_{with} and $S_{without}$ are the distributions of the performance metrics *with* and *without* a certain product, respectively, r_i is a single realization from the distribution of S_{with} and n is the total number of realizations of the S_{with} distribution. For $P_{I,S} > 0.5$ it is then more likely that including a product leads to improved objective function values, and vice versa for $P_{I,S} < 0.5$.

4.2.7. Added value for hydrological signatures

To assess the potential of using different remote sensing products to improve the representation of hydrological signatures compared to those obtained in an unconstrained situation, the feasible solutions of the parameter selection procedure as described in section 4.2.4 were evaluated for a set of 27 hydrological signatures (Table 4.4), as previously defined by, among others, Shamir et al. (2005), Yilmaz et al. (2008), Euser et al. (2013), Pechlivanidis and Arheimer (2015) and Kuentz et al. (2017). The probability of improvement in the representation of hydrological signatures when including a specific remote sensing product was determined in comparison with a reference situation, which in this case corresponds to the distribution of signatures obtained from the unconstrained models (Figure 4.2b).

Corresponding to Eq. 4.6, the probability of improvement ($P_{I,S}$) can be calculated for each signature and each catchment (Nijzink et al., 2016b) as:

$$P_{I,S} = P(S_c > S_u) = \sum_{i=1}^{n} P(S_c > S_u | S_C = r_i) P(S_C = r_i), \quad (4.6)$$

where S_C and S_u are the distributions of the signature performance metrics (e.g. squared correlation coefficient) of the constrained model and reference unconstrained model, respectively, for the set of all feasible solutions retained from constraining or calibration, ri is a single realization from the distribution of S and n is the total number of realizations of the S distribution.For $P_{I,S} < 0.5$ it is then more likely that the model outperforms the reference model with respect to the signature under consideration, and vice versa for $P_{I,S} < 0.5$.Even though not all signatures are fully uncorrelated, high probabilities of improvement for all catchments and all signatures simultaneously points towards more consistent model behavior (cf. Euser et al., 2013).

The probabilities of improvement are calculated for all 279 cases (27 catchments with 27 signatures each) for each combination of data sources (i.e. a certain set of constraints), with the improvement relative to the unconstrained reference model. The relative occurrences of certain probabilities of improvement can be inspected by means of cumulative density plots. See also Figure 4.2b for a stepwise clarification of this approach. As the different combinations of products lead to a varying degree of signature reproduction, different cumulative density curves will emerge. Therefore, improvements can be identified by shifts in the cumulative density plots towards higher probabilities of improvement.

4.3. Results and Discussion

4.3.1. Linking parameters and data sources

In a first step, all parameters had to be related to relevant data sources. This was done based on the sensitivity for each parameter to a certain product (Figure 4.3), in terms of the average vertical distance between the empirical prior distribution and the posterior distribution (Kolmogorov-Smirnoff distance; i.e. the higher the distance, the more sensitive a parameter is to the information provided by a given product). For FLEX (Figure 4.3a), it can be noted that the two snow parameters (*Mmelt* and *Tthresh*) react to the two snow products, as expected. More interestingly, also the soil moisture products and GRACE influence the snow parameters, which can be explained by the role of snowmelt entering the unsaturated storage. A similar argumentation holds for the parameter of maximum interception capacity *Imax*, which is, in addition to solely the evaporation products, also affected by the soil moisture products. The soil moisture parameters *Sumax* and *Beta* exhibit a high sensitivity to the group of soil moisture products, but also to the evaporation products and GRACE. The parameters *Kf*, *Ks* and *D* are to some extent linked to GRACE, but also to the soil moisture products. Thus, each parameter of FLEX can be related to certain products, and can hence be constrained.

Table 4.4: Overview of the hydrological signatures.

Signature	Description	Reference
S_{QMA}	Mean annual runoff	
S_{AC}	One day autocorrelation coefficient	Montanari and Toth (2007)
$S_{\text{AC,summer}}$	One day autocorrelation the summer period	Euser et al. (2013)
$S_{\text{AC,winter}}$	One day autocorrelation the winter period	Euser et al. (2013)
S_{RLD}	Rising limb density	Shamir et al. (2005)
S_{DLD}	Declining limb density	Shamir et al. (2005)
S_{Q_5}	Flow exceeded in 5 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{50}}$	Flow exceeded in 50 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{95}}$	Flow exceeded in 95 % of the time	Jothityangkoon et al. (2001)
$S_{Q_5,\text{summer}}$	Flow exceeded in 5 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{50},\text{summer}}$	Flow exceeded in 50 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{95},\text{summer}}$	Flow exceeded in 95 % of the summer time	Yilmaz et al. (2008)
$S_{Q_5,\text{winter}}$	Flow exceeded in 5 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{50},\text{winter}}$	Flow exceeded in 50 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{95},\text{winter}}$	Flow exceeded in 95 % of the winter time	Yilmaz et al. (2008)
S_{Peaks}	Peak distribution	Euser et al. (2013)
$S_{\text{Peaks,summer}}$	Peak distribution summer period	Euser et al. (2013)
$S_{\text{Peaks,winter}}$	Peak distribution winter period	Euser et al. (2013)
$S_{Q_{\text{peak},10}}$	Flow exceeded in 10 % of the peaks	
$S_{Q_{\text{peak},50}}$	Flow exceeded in 50 % of the peaks	
$S_{Q_{\text{summer,peak},10}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{summer,peak},50}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{winter,peak},10}}$	Flow exceeded in 10 % of the winter peaks	
$S_{Q_{\text{winter,peak},50}}$	Flow exceeded in 50 % of the winter peaks	
S_{SFDC}	Slope flow duration curve	Yadav et al. (2007)
S_{LFR}	Low flow ratio (Q_{90} / Q_{50})	
S_{FDC}	Flow duration curve	Westerberg et al. (2011)

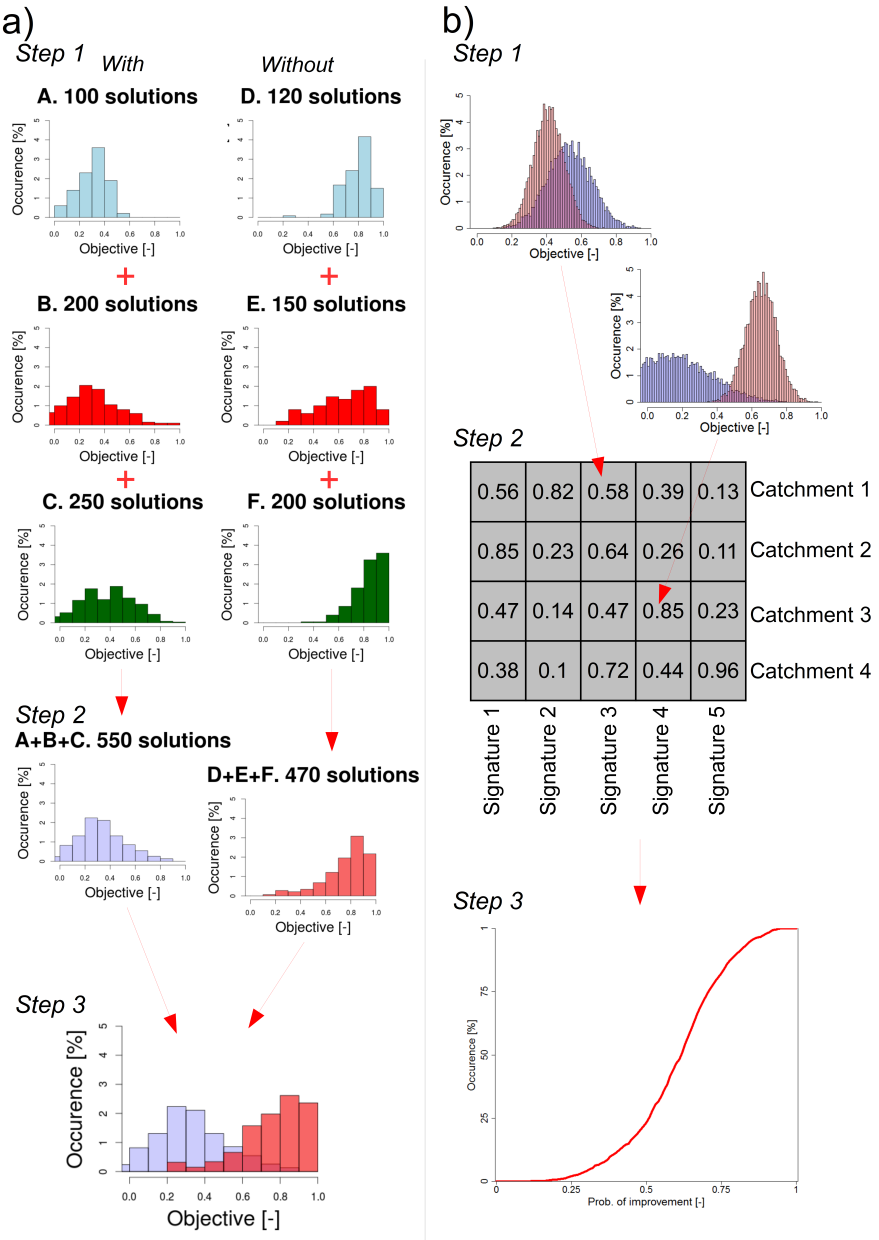

Figure 4.2: Schematized representation of **(a)** the procedure to calculate the probabilities of improvement for each product by adding up all distributions in Step 1 with a certain product, as shown for examples A (100 feasible solutions), B(200), C(250), which leads to a marginal distribution in Step 2 of 550 solutions (100+200+250) and without a certain product, as shown for distributions D, E, F (120+150+200=470 solutions), which can be compared in Step 3 to calculate the probability of improvement. In **(b)** the signature analysis is displayed, with in step 1 the density distributions for the performance metrics for a specific signature in the constrained case (red) and unconstrained case (blue), step 2 the probability of improvement derived from these distributions for each signature and each catchment and step 3 the resulting cumulative occurrences for the probabilities of improvement.

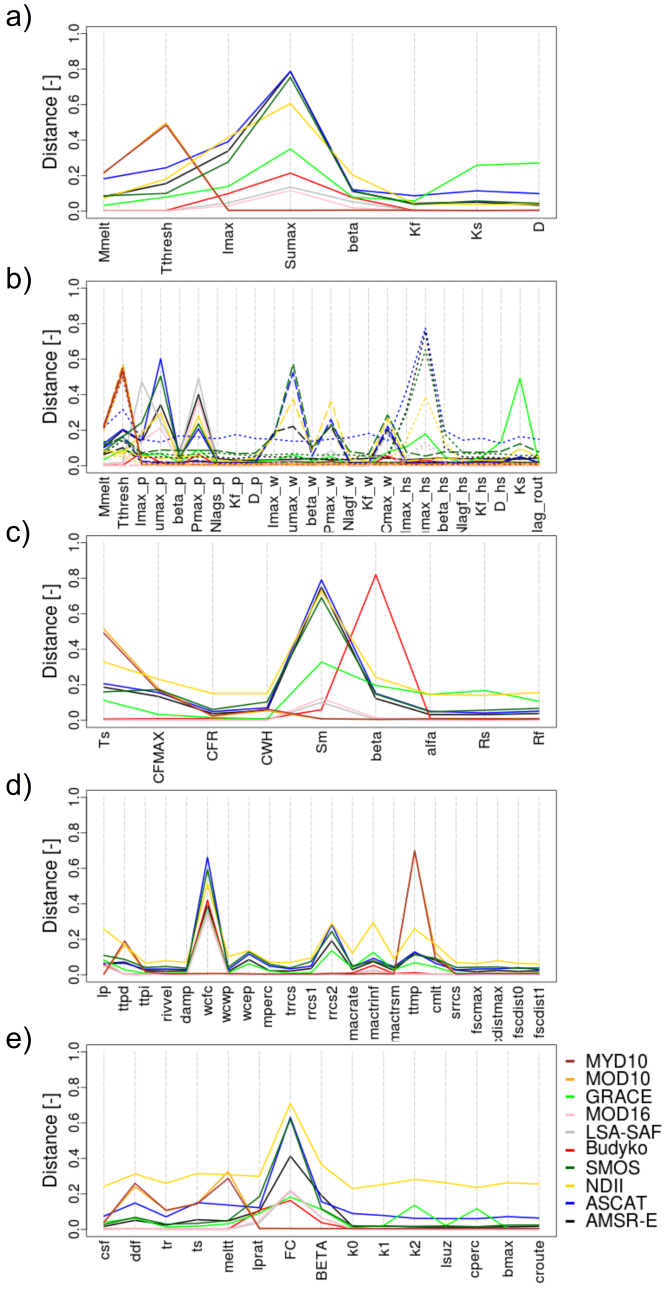

Figure 4.3: Sensitivity in average vertical distance between the empirical distribution curves of posterior and prior (uniform) parameter distributions for **(a)** FLEX, **(b)** FLEXtopo, **(c)** HYMOD, **(d)** HYPE, **(e)** TUW. Different colors indicate different products, with for FLEXtopo a distinction per landscape class with plateau, wetland (dashed) and hillslope (dotted).

The sensitivity plot for FLEXtopo (Figure 4.3b) shows similar behavior, as similar model parameters are used as in FLEX, but now only applied in different landscape classes. Even though no complete overlap could be found, corresponding parameters in FLEXtopo were constrained with the same products as for FLEX to maintain consistency. However, as the number of parameters is higher, while the number of sampled random parameter sets is just slightly higher, the analysis of this model is based on a lower sampling density. Thus, to avoid the situation that no solutions remain due to too many parameter constraints, parameters that did not show a consistent sensitivity for a specific group of products were mainly left unconstrained, such as *Pmax*, *Cmax*, *Kf*, *Beta* and *D* for the three landscape classes (Figure 4.3b).

The relations between products and parameters for HYMOD (Figure 4.3c) show a rather consistent pattern compared to FLEX and FLEXtopo. Also in this case, the snow parameters (*Ts*, *CFMAX*, *CFR*, *CWH*) are not just sensitive to the snow products, but also the soil moisture products. For *Ts* and *CFMAX* also GRACE has an influence, which happened as well for the snow parameters of FLEX. *CFR* and *CWH* do not show this, indicating that this refreezing factor (*CFR*) and water holding capacity of snow (*CWH*) have, apparently, a minor influence in the storage anomalies. The maximum soil moisture *Sm* can be constrained with the soil moisture products, GRACE and the evaporation products. For the second soil moisture parameter *beta*, the soil moisture products matter, but especially a high sensitivity to the Budyko framework can be observed. The last parameters of *Rs*, *Rf* and *alfa* (relating to the reservoir coefficients and runoff generation) can be linked to the soil moisture products as well as GRACE.

It can be noted for HYPE (Figure 4.3d) that the general sensitivity is lower compared to the other models. Nevertheless, the snow parameters (*ttpd*, *ttpi*, *ttmp*, *cmlt*) can be constrained in a similar fashion as for FLEX and HYMOD, thus with the snow products, soil moisture products and GRACE. In addition, *rrcs2* and *macrate*, both related to groundwater dynamics, show considerable sensitivity to GRACE and the soil moisture products. The parameters controlling soil moisture and, thus transpiration *lp*, *wcfc* and *mactrinf*, relate to GRACE, the soil moisture products and the evaporation products. The other parameters are left unconstrained as the sensitivities here are generally low or do not show a clear sensitivity to one of the groups of products.

The snow parameters *csf*, *ddf*, *tr*, *ts* and *meltt* of the TUW model (Figure 4.3e) can also be constrained with the snow products, GRACE and the soil moisture products. The soil moisture parameters of *lprat*, *FC* and *BETA* show a high sensitivity to the soil moisture products, but also again GRACE and the evaporation products. The groundwater parameters *k2* and *cperc* relate to GRACE and the soil moisture products, the remainder of the parameters is left unconstrained as no clear preference for certain groups of products can be identified.

4.3.2. Streamflow calibration performances versus constrained performances

In a next step, the performances of the five models when constrained with all different combinations of data sources (section 4.2.4) are compared with the per-

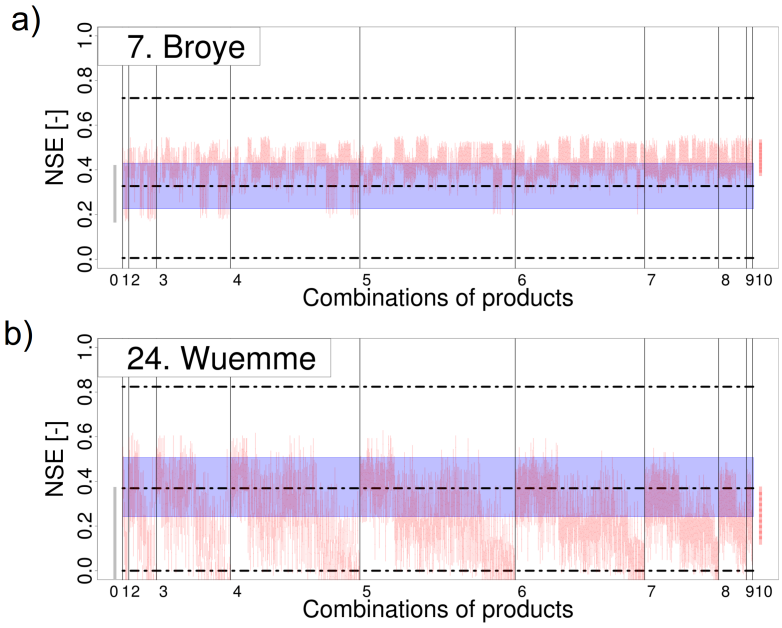

Figure 4.4: Comparison of Nash-Sutcliffe performances for the calibrated models (25^{th} and 75^{th} quartile correspond to the blue band, dashed lines represent the 5th and 95th quartiles) and the constrained model applications (red line for each combination of constraints, representing the 25^{th} and 75^{th} quartiles) for **(a)** FLEX and the Broye catchment, **(b)** TUW and the Wuemme catchment The unconstrained model performances (corresponding to the origin of the plot, i.e. combinations of products equal to 0) are displayed in grey.

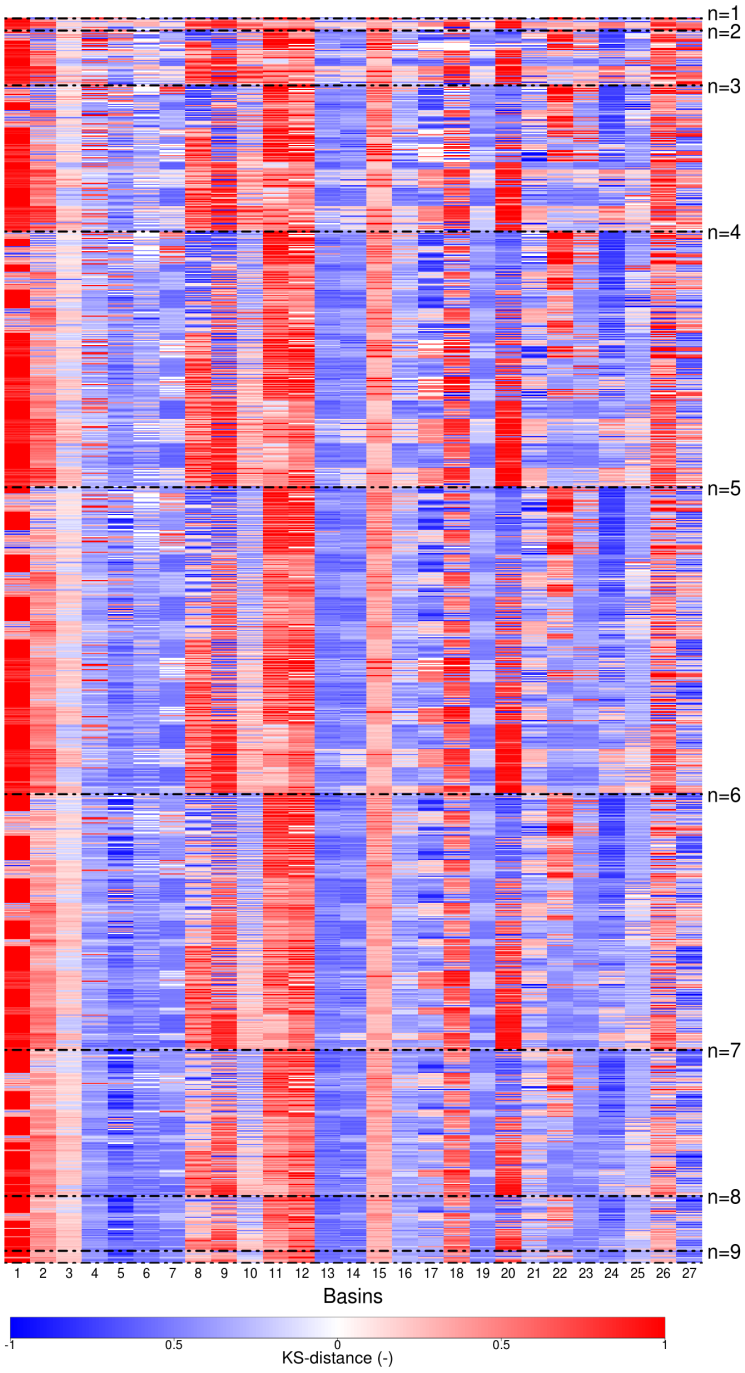

Figure 4.5: Distance between the calibrated and constrained distribution of the Euclidian distance between Nash-Sutcliff of the flows and log of the flows (i.e. Kolmogorov-Smirnoff distance), for each possible combination from n=1 to n=9 included products, for the FLEX model. Blue values indicate a constrained distribution with higher performances than calibrated, red indicates a calibrated distributions with higher performances.

formances obtained by calibration to streamflow (section 4.2.5). When the models are calibrated on streamflow data, they all provide similar model performances, but different patterns emerge for the different catchments and models when constrained with the different data sources as shown in for selected catchments and models in Figure 4.4. See the Supplement Section S2 of (Nijzink et al., 2017) for all model performances in the constrained and calibrated situations. In general, the model performances increase with an increasing number of data sources used for constraining, as expected, but in some occasions no feasible solutions remain.

Some of applications of the FLEX model constrained exclusively by remote sensing data exhibit a similar range of model performances as when calibrated on streamflow. In several cases an equivalent level of model performance was achieved after adding four data sources in the parameter selection procedure, as can be observed from Figure 4.5, where the maximum distance between empirical cumulative distribution curves of the calibration and constrained models (the Kolmogorov-Smirnoff distance) is displayed for all combinations of products and catchments. Furthermore, in Figure 4.4, zigzag patterns of improvements and deteriorations can be noticed after constraining with a larger set of products for some catchments. This is an indication that families of combinations that either include or exclude certain (combinations of) products, can lead to major improvements or strong decreases in performances. It can be clearly observed that the distances vary and are grouped (Figure 4.5), pointing at specific combinations of products that constrain in such a way that the new performance distribution comes close to the calibration distribution, or even improves. For FLEX, only in 4 of the 27 catchments (catchments 1, 11, 12, 18) the constraints lead to performance distributions that are substantially lower than the calibrated results, whilst the other catchments always approach the calibration results more closely, or even show higher performances (catchments 5, 13, 14, 24). Inspection of the combinations reveals that GRACE data are an important contributor to model improvements particularly for a large number of catchments, except for the Gadera (catchment 1), but especially for catchments 7, 14, 17, 19 and 27. Similarly, for the Treene, Modau and Wuemme (catchments 8, 20 and 24) the AMSR-E product is the common factor in the more successful combinations. The Treene catchment, and to a lesser extent also the Wuemme, are peaty lowland catchments, with very moist soils and shallow groundwater tables, which matches well with the information derived from AMSR-E. This seems, however, in contrast to statements from the AMSR-E developers that pixels with a large proportion of open water introduce errors (Owe et al., 2008) or other researchers that suggest that peatland creates errors in soil moisture products (Bartalis et al., 2007). On the other hand, Owe et al. (2008) mention steep mountainous areas as source for error, which these catchments are certainly not. The snow products are included in the more successful combinations for Vils, Grossarler and Grosse Mühl (catchments 4, 5 and 6), which are also the more snow dominated catchments. The evaporation products of MOD16 and LSA-SAF do not show a clear pattern when included or excluded in the parameter selection procedure, pointing at a relatively minor role in determining the performances with regard to streamflow, reflecting results of Oudin et al. (2004). In addition, the number of parameters concerning evaporation

is generally lower, also reducing the importance of these products for constraining.

The results of the FLEXtopo model (Supplement of Nijzink et al. (2017) Figures S10-S13) show in general similar patterns as for FLEX. However, it can be noted that more combinations of constraints also lead to no remaining feasible solutions at all. The problem here is mostly linked to the larger number of free calibration parameters (24) and the resulting under-sampled parameter space. In addition, a similar reasoning can be made for HYMOD (Supplement of Nijzink et al. (2017) Figures S14-S17), but here this is merely caused by the relatively wide prior parameter ranges used for HYMOD, leading to only a relatively small number of solutions with high performances which the constraints cannot easily filter out. These ranges were initially set wide on purpose to assess the power of the constraints, but it largely remains a challenge to obtain similar performances as for the model calibrated on streamflow, when set too wide. Thus, similar performances as in traditional calibration are eventually only achieved for the Tanaro, Fyllean and Deveron (catchments 2, 18 and 25). The variability between the different combinations of constraints is also large in the case of HYMOD, pointing to an extremely high added value of a certain (family of) constraints, which are combinations with the Budyko framework.

Similarly, HYPE (Supplement of Nijzink et al. (2017) Figures S18-S21) has some more difficulties in order to obtain similar performances as when calibrated on streamflow. Even though the performances in calibration are reasonable, the absolute number of feasible solutions is relatively low. Therefore, the constraints from the products need considerably more restrictive power to filter all solutions and to converge to the same performance level as the streamflow calibration. This also relates to the relatively large number of free calibration parameters and thus a larger a priori search space. In other words, too many poor solutions are maintained when the model is constrained on the remote sensing data sources.

The TUW model (Supplement of Nijzink et al. (2017) Figures S22-S25) shows also a clear pattern of strong and weak combinations of remote sensing products, but the variability between the (families of) combinations is generally not very high. The relative importance of each parameter, and thereby each data source connected to it, is lower as TUW has a relatively elevated number of parameters (i.e. 15). Thus, leaving a single parameter of all the TUW parameters unconstrained has less consequences compared to constraining a single parameter from the 8 parameters of FLEX. Nevertheless, the same products (GRACE, AMSR-E, ASCAT) strongly improve the parameterizations of TUW, similar to FLEX, also for the same catchments (such as catchments 8, 20 and 24).

4.3.3. Added value of remote sensing data to reproduce streamflow

Figure 4.6 compares the overall added value of each individual remote sensing product to generate meaningful posterior distributions and thus to provide efficient and effective parameter constraints. This was done as described in section 4.6 by assessing the probability of improving the representation of streamflow when including a specific remote sensing data source for constraining a model than when not including it.

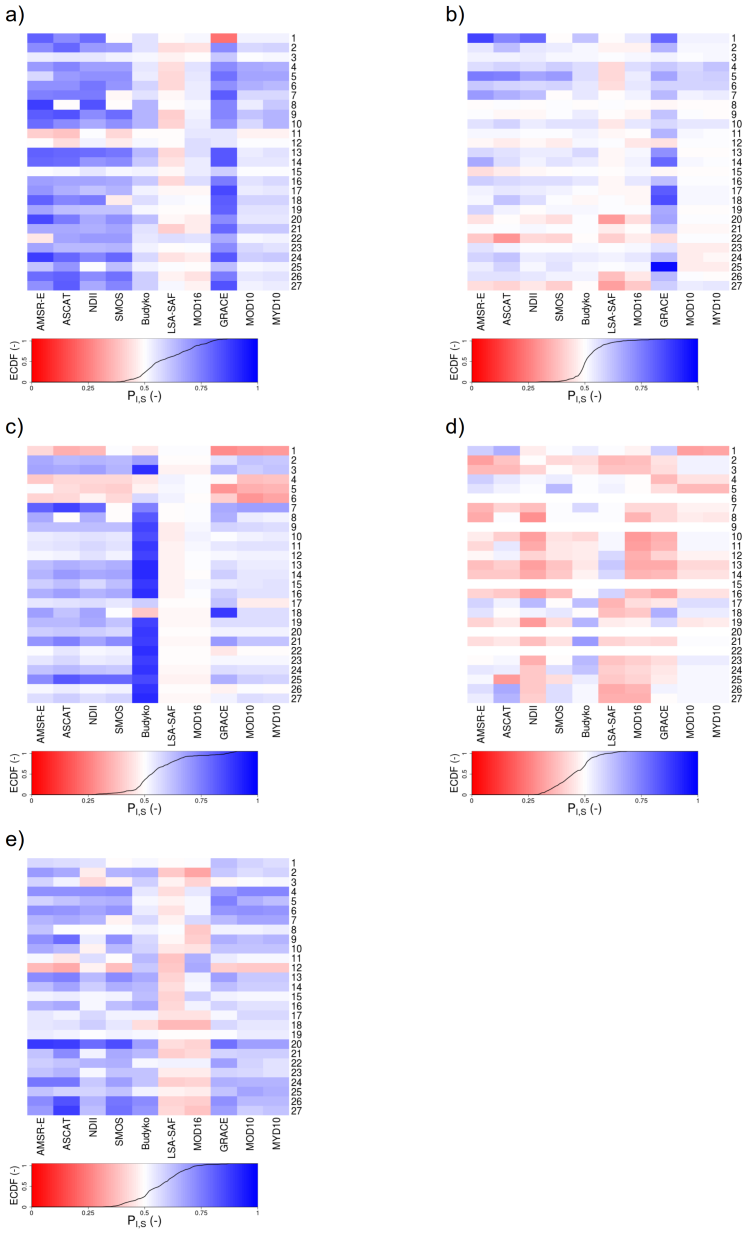

Figure 4.6: Probabilities of improvement for including a specific product in the set of products used for constraining compared to not including this product, with regard to the Euclidian distance between Nash-Sutcliffe of flows and logarithm of the flows, shown for **(a)** FLEX, **(b)** FLEXtopo, **(c)** HYMOD, **(d)** HYPE, **(e)** TUW. High probabilities of improvement plot in increasingly dark shades of blue, while shades of red indicate increasing probabilities of deterioration. The ECDF curves represent the cumulative distribution curves of all the probabilities of improvement for all catchments.

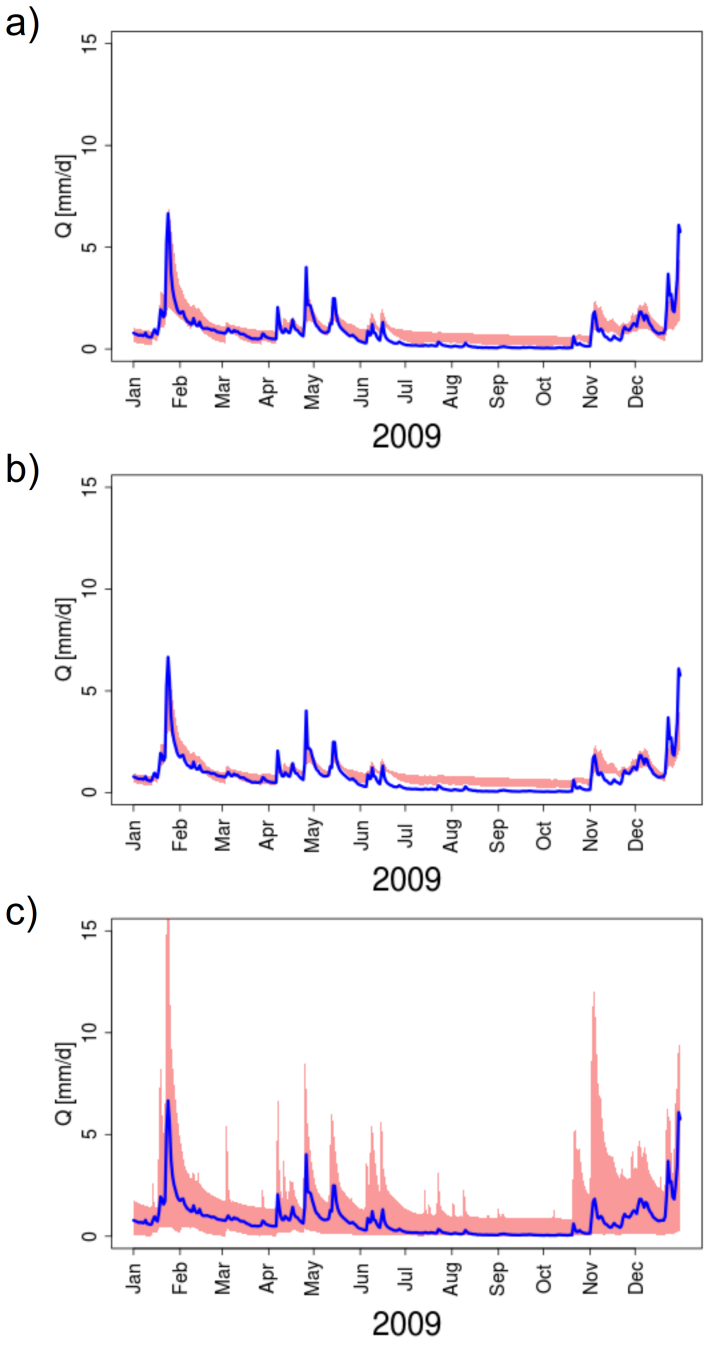

Figure 4.7: Feasible flow ranges for FLEX for a selected time period of catchment 13 obtained with combinations of remote sensing products **(a)** with GRACE and **(b)** without GRACE **(c)** calibration on streamflow. Colored envelopes in a) and b)represent the number of products (n=1 to n=10) used in deriving the posterior parameter distributions and flow ranges, observed discharge is shown in blue and, in **(c)** , calibrated discharge on streamflow in red.

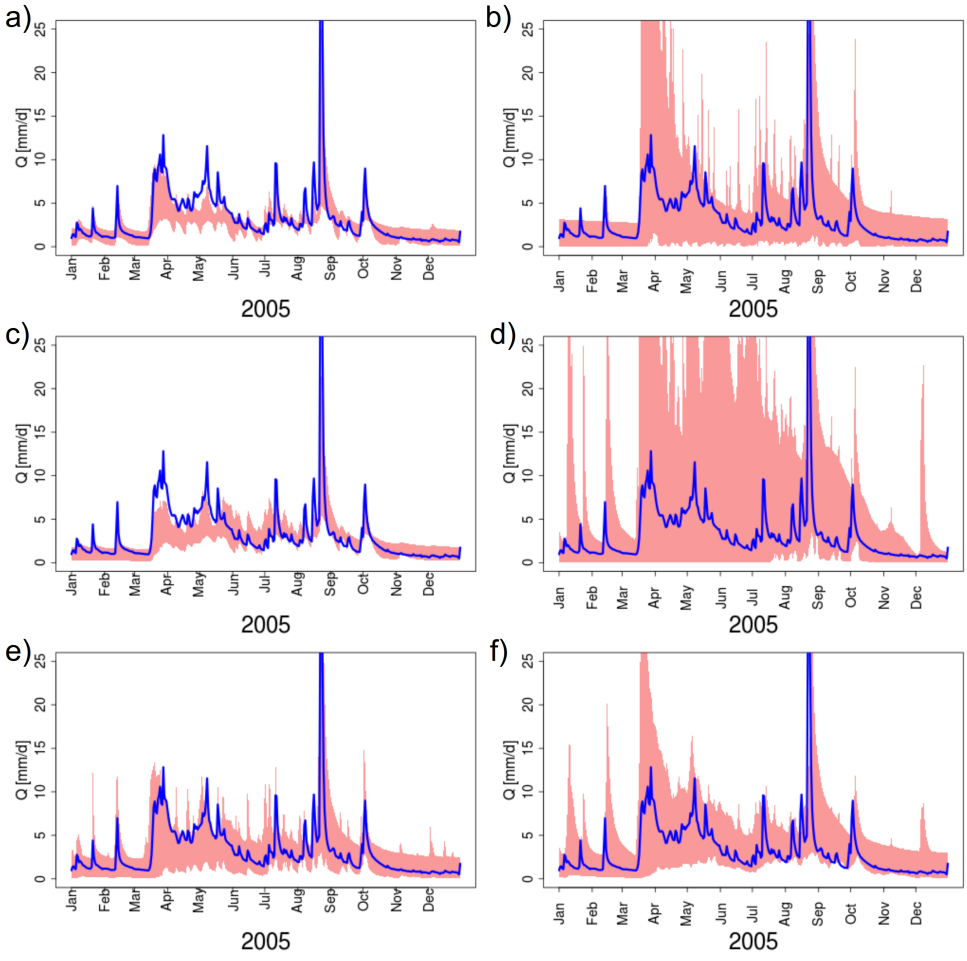

Figure 4.8: Feasible flow ranges for a selected time period of the Vils catchment (catchment 4) obtained with combinations of remote sensing products of a) with MOD10 and FLEX, b) with MOD10 and HYMOD, c) without MOD10 and FLEX and d) without MOD10 and HYMOD, and e) FLEX calibrated on streamflow, f) HYMOD calibrated on streamflow. Observed discharge is shown in blue and the colored envelopes in a), b), c) and d) represent the feasible ranges with 9 products used in deriving the posterior parameter distributions, and, in e) and f), calibrated discharge on streamflow.

For example, in Figure 4.6a it can be noted that high probabilities of improvement are obtained for most data sources for FLEX. Only a few cases exist where one of the data sources provides a low probability of improvement (<0.5), which suggests that constraining on any single remote sensing data source has already considerable constraining power in a wide range of cases. This is for example true for GRACE, in particular when applied with FLEX, where high probabilities of improvement can be observed (see also section 4.3.2). This is similarly illustrated for the hydrograph of the Glane (catchment 13) in Figure 4.7, where especially in combinations with nine products GRACE has the potential to move the lower bound of the uncertainty interval of the modelled hydrograph towards the observations, in particular for the low flows. Further, it can be noted that during the summer period the final set of constraints provides a much narrower uncertainty bound than the calibrated results (Figure 4.7c). Several previous studies similarly suggested that GRACE has a high potential to improve hydrological simulations (Mulder et al., 2015; Rakovec et al., 2016), but this was thought to be true mostly for larger catchments than those under consideration here. In particular the Broye and Dalsaelve (catchments 7, 17) show a high probability of improvement when constraining FLEX with GRACE. For the Broye this is likely to be linked to the adjacent Lake Geneva, which may influence the groundwater tables in the surrounding catchments, leading to similar water storage anomalies of all catchments within the GRACE cells. The larger catchments, such as the Leyre (10; $1587km^2$) or Hunte (23; $1409km^2$) still show high probabilities of improvement, which is related to a signal of water storage anomalies much closer to the GRACE signal. Yet, the probabilities of improvement are high for most other catchments as well, which also includes catchments with areas of less than $100km^2$ (e.g. catchment 20). In addition to GRACE, the soil moisture products of AMSR-E and ASCAT show the highest probabilities of improvement, whereas SMOS has a slightly lower added value. This is in agreement with findings of Wanders et al. (2014), who also applied AMSR-E, ASCAT and SMOS in a hydrological model and found that soil moisture improved the strongest for AMSR-E and ASCAT. Nevertheless, they also found that AMSR-E and ASCAT work best in areas with a pronounced relief, whereas the highest probabilities of improvement are, in our study, obtained for both catchments with low and high elevation differences. The relatively white colors for Budyko, MOD16 and LSA-SAF in Figure 4.6, corresponding to probabilities of around 0.5, indicate that these data sources do not add a lot of constraining power, but also do not have adverse effects when included. These data sources are, apparently, not very important with respect to the streamflow objectives considered here. However, in warmer and more arid climates outside Europe, these products may have significantly more value. The snow products show, as expected, high probabilities of improvement for Vils, Grossarler and Grosse Muehl (catchments 4-6), which are more snow-dominated catchments.

FLEXtopo (Figure 4.6b) shows a similar pattern as for FLEX, but it can also be observed that LSA-SAF evaporation has in several cases a negative influence on the results. This is in line with Figure S6 in the Supplement of Nijzink et al. (2017), where it can be noted that FLEXtopo has difficulties to achieve high correlations between the LSA-SAF evaporation and modeled evaporation, therefore also leading

to rather poor posterior parameter ranges. The low correlations are mainly caused in the landscape class of plateau, as used by the model. Especially this landscape class is often employed for agriculturally used land, which is also the case for the Little Ouse and Stour (catchments 26 and 27, with respectively 64.4% and 86.6% agriculture of the plateau areas), which show very low probabilities of improvement for LSA-SAF evaporation. Either the model or the product has difficulties with reproducing the transpiration signal from the crops, which can even be altered by management decisions to irrigate and/or harvest. Besides this, it was concluded by Hu et al. (2015) that LSA-SAF evaporation has difficulties in arid regions, and the catchments of Little Ouse and Stour have an aridity of 1.05 and 1.03 respectively, which are also the highest aridity values in this analysis (even though located in the UK).

Unlike the results for FLEX and FLEXtopo, the Budyko framework has a big influence on the results for HYMOD (Figure 4.6c). Here, high probabilities of improvement show the importance of the Budyko framework for HYMOD, whereas FLEX and FLEXtopo (and also HYPE and TUW) show an almost white column for Budyko. This indicates that the model has difficulties in reproducing the long-term flux partitioning into streamflow and evaporative fluxes, which can also be seen from Figure S6 in the Supplement of Nijzink et al. (2017). Nevertheless, the Budyko framework was only connected to two parameters (*Sm*, *Beta*) of HYMOD, and therefore, leaving these parameters unconstrained leads to many poor solutions. Based on these results, it can be argued that these parameters must be constrained in all cases. However, the Budyko framework helped here, similar as in the studies of Li et al. (2014) and Gentine et al. (2012), to identify feasible sets of parameters. Another clear distinction with the other models can be found in the probabilities of improvement for the two snow products (MOD10 and MYD10). For all other models at least moderate probabilities are observed, in particular for catchments 4-7, but for HYMOD the probabilities of improvement remain very low. This can also be noted from the hydrographs in Figure 4.8, where for FLEX some snow-peaks are improved when constraining the snow parameters, for HYMOD the small snow peaks in January and February disappear and the large snow peak starting in March remains too high.

Similar to FLEX and FLEXtopo, HYPE (Figure 4.6d) benefits from including ASCAT or AMSR-E, even though the absolute level of improvements remain quite low. NDII shows even lower probabilities for all catchments. It is also interesting to note that the snow products have a similar low probability of improvement for the Vils and Grossarler catchment (catchments 4 and 5) as for HYMOD, whereas FLEX, FLEXtopo and TUW have high probabilities of improvement for applying the snow products in these catchments. These two catchments have the highest number of possible snow days (29.8% and 28.7%), and one would expect high probabilities of improvement for the snow products here for all models, also based on previous work (Parajka and Blöschl, 2008). Figure S6 in the Supplement of Nijzink et al. (2017) also shows that there are no distinct differences between the modeled and observed snow signals between the models and, thus, the low probabilities of improvement, with regard to streamflow, very likely point towards other model

structural deficiencies in HYMOD and HYPE. In other words, the snowmelt may still be better represented when the snow parameters are constrained with the snow products, but how snowmelt water is then routed through the rest of the system may not be adequately represented.

For TUW, the observed patterns in Figure 4.6e are again similar to the patterns for FLEX and FLEXtopo, but with less distinct differences between the different data sources. ASCAT also shows high probabilities of improvement for most catchments, and AMSR-E to a lesser degree. The evaporation products of LSA-SAF and MOD16 have, more than for FLEX or FLEXtopo, rather low probabilities of improvement for the full range of catchments. At the same time, Budyko helps a lot, pointing also here at difficulties of the model to reproduce long-term behavior.

4.3.4. Added value of remote sensing data for hydrological signatures

Figure 4.9 summarizes all empirical cumulative probability distribution curves obtained from the combined probabilities of improvement of all tested hydrological signatures for all combinations of products, relative to the unconstrained models. It can be observed that all models experience, on average, a shift towards higher probabilities of improvement (i.e. to the right – a higher proportion of cases with a probability of improvement > 0.5) when more products are included, also pointing towards improved model internal dynamics.

For FLEX and FLEXtopo (Figures 4.9a and 4.9b) the results suggest that a strong improvement can, on average, be achieved by including more remote sensing products. This can be seen by the large shifts of improvement between the envelopes of one product (gray) and more products (dark red colors). The final set of constraints, with all products included (red in Figure 4.9), is for FLEX close to containing the highest probabilities of improvement (i.e. the curve closest to a probability of improvement of 1 on the x-axis). FLEXtopo also has high probabilities of improvement for the final set of constraints, but this curve (red line) is, also here, not the curve with the highest values. This indicates that at least one of the products is not adding more value, and actually reduces the models ability to reproduce the set of hydrological signatures. Inspection of the individual curves shows that the curve, for nine products included, with the highest probabilities of improvement is the curve without NDII. In Figure 4.6, this can also be observed in some cases, but the negative influence becomes much more apparent when evaluating a set of signatures. The curve with the lowest probabilities of improvement for nine products is however the curve without GRACE, pointing also at the importance of GRACE for reproducing the signatures. The additional gains by including GRACE are also in agreement with findings of Rakovec et al. (2016), who found that evaporation estimates largely improved.

For HYMOD (Figure 4.9c), the solutions obtained from the highest number of remote sensing products used to constrain ($n = 9$) are not the curves with the highest probabilities of improvement. Moreover, the gray envelopes for a low number of products are wide and contain the curves with the highest probabilities of improvement (closest to 1) and lowest (close to 0). The curve most towards high

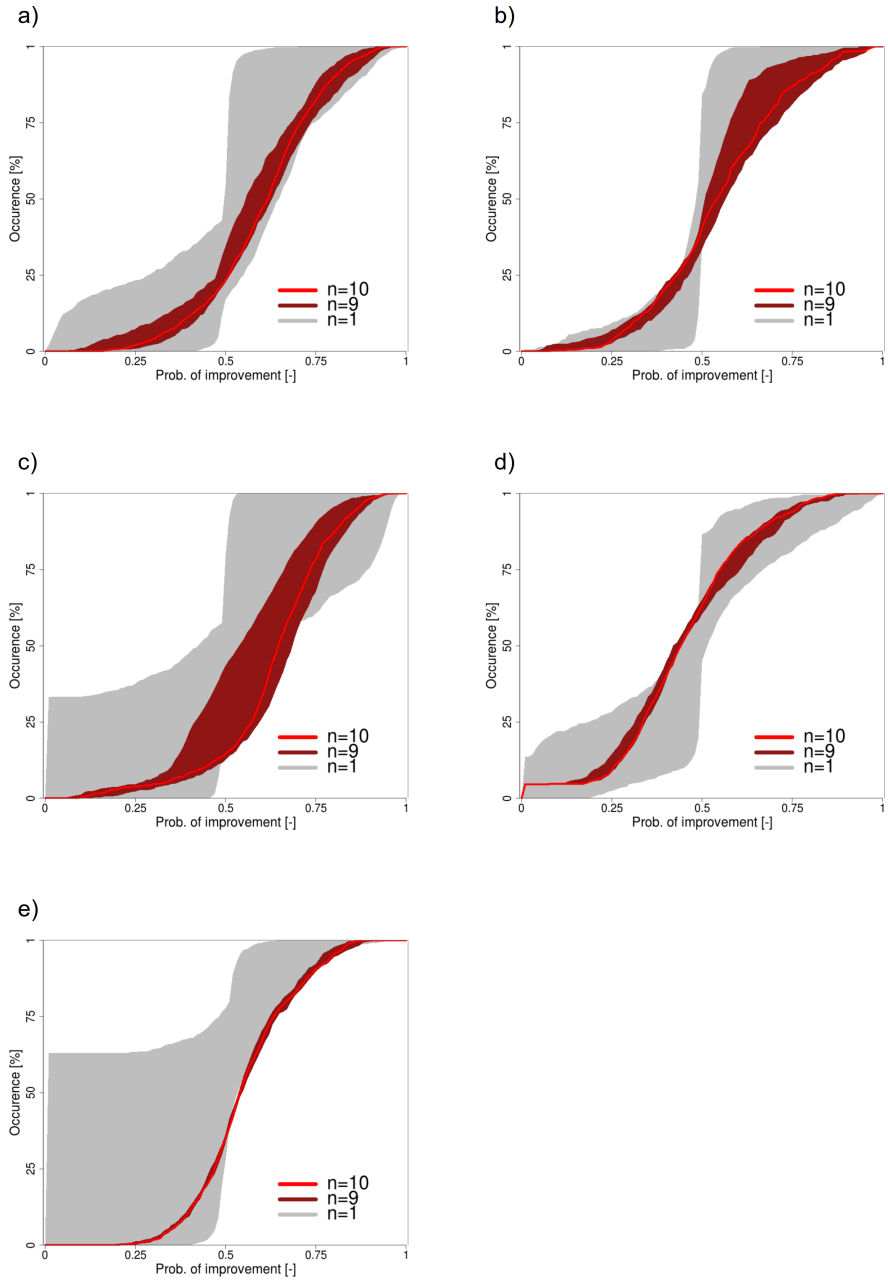

Figure 4.9: Envelopes of all empirical cumulative probability distribution curves for all probabilities of improvement in the representation of catchment signatures (each catchment, each signature) for the different combinations of constraints (i.e. remote sensing products) compared to the unconstrained model: **(a)** FLEX, **(b)** FLEXtopo, **(c)** HYMOD, **(d)** HYPE, and **(e)** TUW. The different colors represent the different number of remote sensing products used to constraining the models.

probabilities when one data source is included (gray) is in this case Budyko, also in agreement with the previous findings for the objective function values. Most of the other curves for the individual products plot around a probability of 0.5, indicating that it does not significantly influence the model results if this data source is included or not. After including Budyko in the combinations, these curves start to shift to the right, eventually leading to curves that end up in the middle of the extremes when combined together in a set of new posterior bounds.

The curves for HYPE (Figure 4.9d) look rather similar to the curves of FLEX and FLEXtopo, but the envelopes of the curves resulting from the use of nine products are much narrower. Thus, the same number of products are much more restrictive for HYPE, leading to reduced uncertainty intervals. Even though the final set of constraints is not the curve with the highest probabilities of improvement, the curves are rather close. Thus, excluding a certain product for the set used for constraining does not make a big difference, pointing at the combined strength of the other remaining nine products.

The very wide envelopes for TUW for the lower probabilities (Figure 4.9e) points at too restrictive constraints from a certain product, which leads to no feasible solutions that remain (around 60% has a probability of 0). When combinations are made, these constraints are relaxed and corrected by including other data sources, leading to a lower start of the curves, thus pointing at more feasible solutions. In the end, the envelopes of the higher number of products are most to the right and quite narrow, indicating that still all data sources helped to improve the representation of catchment signatures. This points at the combined strength of the products, correcting too restrictive constraints and improving the signature representation together.

4.4. Conclusions

27 catchments across Europe were constrained with combinations of nine remotely sensed products and an analytical framework (Budyko). New posterior parameter distributions for five different conceptual hydrological models were derived based on a likelihood weighting procedure, which was specific for each parameter depending on the relevant data sources for that parameter. In this way, all 1023 possible combinations of these 10 data sources could be used to derive new parameter ranges.

It was found that high probabilities of improvement were obtained when combinations included in particular AMSR-E and ASCAT soil moisture data. Surprisingly, considering the relatively small size of the study catchments, also GRACE added considerable value to meaningfully constrain the tested models. In addition, in snow dominated catchments the MODIS snow products were shown to be helpful for some of these models. The evaporation products of LSA-SAF and MOD16 were to a lesser extent important for deriving adequate and meaningful posterior parameter distributions.

A set of 27 hydrological signatures was evaluated for each study catchment and the probability of improvement for reproducing these signatures using only remote sensing data for constraining a model compared to unconstrained models

was analyzed. This showed that all models benefitted from using a combination of remote sensing data for reproducing catchment signatures.

This study illustrates that using combinations of multiple data sources is in most cases valuable to derive reasonably narrow and meaningful posterior parameter distributions. It was shown that the highest gains are, on average, obtained when the soil moisture products AMSR-E and ASCAT as well as the total water storage anomaly of GRACE are included. Using multiple products simultaneously also corrects errors of a single product and including four to five different products is often sufficient to obtain similar performances as in traditional calibration on streamflow. In addition, hydrological signatures were better represented when multiple data sources were used, indicating improved model internal dynamics. In conclusion, adding multiple data sources in parameter selection procedures in an indirect, parameter specific way is a promising way forward in predicting ungauged catchments.

5

Constraining a large-scale landscape-based model

The model search space can substantially be reduced by prior parameter constraints, being also rather similar to model regularization as described before. Nevertheless, qualitative expert-based parameter constraints have the advantage of their easiness to use for different model applications, maintaining also a strong basis in hydrological theory, in contrast for more formal mathematical, and often empirical, model regularization. As especially large-scale models are in need of relatively simple methods to constrain the search space, it is tested whether prior parameter constraints can suffice to obtain improved model formulations.

5.1. Introduction

A vast family of large-scale hydrological models exist as global climate issues are highly prioritized and computational resources increase. Even though Beven (2000) introduced the "uniqueness of place"-concept in hydrology and site-specific and flexible approaches (e.g. de Boer-Euser et al., 2017; Fenicia et al., 2011; Gao et al., 2014a; Uhlenbrook et al., 2004) started to gain attention, many of these models still use a rather simple, single model structure. A large part of the hydrological community still embraces this and supports the idea that the diversity of processes and catchments not means that a single model cannot be applied in all cases (Linsley, 1982). It can also be argued that a strong need for large-scale models with a single model structure will always exist, especially as coupling with climate-models is difficult, but also as time and data are scarce and a generic model structure creates the opportunity to learn from the model and its deficiencies (Le Moine et al., 2007).

Often, these models are developed in order to couple with climate models and model especially the feedback between atmosphere and the land surface, also with regard to the carbon cycle, so-called land surface models (LSM's) such as Tessel (van den Hurk et al., 2000), Orchidee (Krinner et al., 2005), STEAM (Wang-Erlandsson et al., 2014) or LPJ (Sitch et al., 2003). Other large-scale models such as E-HYPE (Lindström et al., 2010), PCRGlob (Sutanudjaja et al., 2014) and mHM (Samaniego et al., 2010b) are more specifically aiming at river discharge, and differ often conceptually from the aforementioned land surface models. Most of these large-scale hydrological models use simple, bucket-like conceptualizations of the land-surface, often similar to the well-known model structure of HBV (Bergström, 1992). In general, these models contain more detailed hydrological processes compared to most LSM's, which mainly model the upper layer of the surface, but these hydrological models can often be considered as (too) simple as well. This has also been recognized by many researchers, as frameworks to evaluate model structures have been proposed (Euser et al., 2013; Gupta et al., 2012; Wagener and Gupta, 2005) and the search for a more unified theory (e.g. Sivapalan, 2006; Troch et al., 2009a) on the catchment scale still continues.

Besides the simplicity, most of the large-scale hydrological models suffer from a lack of calibration due to the long runtimes and the large datasets required. For example, Lopez Lopez et al. (2017) could only calibrate PCRGlob with just 81 models runs, whereas, in comparison, Seibert (1997) calibrated a small scale application of HBV with more than 500000 model runs. Several authors (e.g Gupta and Sorooshian, 1983; Koren et al., 2003; Kuzmin et al., 2008; Renard et al., 2010) argued that better a priori estimates of parameters are needed to avoid these computational issues and have a more aimed optimization approach. In line with this, regionalization techniques were developed that link data to model parameters (e.g. Götzinger and Bárdossy, 2007; Kumar et al., 2013b; Samaniego et al., 2010b). These techniques reduce the need for calibration and limit issues with overparameterization (i.e. equifinality; Beven, 1993), but remain often rather complex, with global parameters that are hard to interpret. Instead, other studies focused more on improved parameter priors by additional data (e.g. Koren et al., 2008; Parajka

and Blöschl, 2008; Winsemius et al., 2008), or by simple prior parameter constraints and process constraints derived from expert-knowledge (Gao et al., 2014a; Gharari et al., 2014; Hrachowitz et al., 2014). Especially by using prior parameter constraints, regionalization remains simple, and the modelers still maintain an understanding of the system. This was only tested for small-scale catchments in Europe, whereas the added value of these approaches mainly concerns cases where calibration options are limited, thus applications in data-scarce regions or large-scale modeling applications.

Large-scale models are not parsimonious in the number of parameters, and yet don't reflect most dominant hydrological processes with high calibration efforts. Therefore, this study focusses on applying a large-scale model with a complex model structure based on the landscape and test whether prior parameter constraints 1) reduce the need for calibration over the full spatial domain and 2) improve process representation.

5.2. Methodology

5.2.1. Model set-up

The large-scale model was based on the principles of FLEXtopo (Savenije, 2010), as also previously applied by (Gao et al., 2014a) and de Boer-Euser et al. (2017). In the original definition of FLEXtopo (Savenije, 2010) three main landscape classes of hillslope, plateau and wetland were defined, which was extended here with classes for urban areas, glaciers and bare rock. Therefore, a global landscape classification was made based on values of the global Height Above Nearest Drainage (HAND; Donchyts et al., 2016) and slopes derived from the SRTM mission (Lehner et al., 2008), but also landuse data from GLOBcover (Arino et al., 2010). The landscape was divided in classes for plateaus (HAND>10m), wetlands (HAND<10m), and north- and south-facing hillslopes (slope>11%) and additionally split in classes with grass, and classes with forest. Three classes were also added for urban areas, glaciers and bare rock, based on the land cover, leading to a total of 11 landscape classes. See also Figure 5.1 for the global landscape classification.

Each landscape class was given a different model formulation, based on a different model structure or model parameterization, following the principles of FLEXtopo (Gao et al., 2014a; Savenije, 2010), see also Figure 5.2. Briefly, each landscape class uses a snow routine, followed by an interception storage. From here, water either evaporates, or, when the interception capacity is exceeded, continues to the soil moisture reservoir. Nevertheless, this only happens if no infiltration excess overland flow occurs, triggered by a rainfall intensity higher than a certain threshold, reflecting the maximum infiltration capacity of the system. After this step, the different landscape classes start to differ more. In the plateau landscapes, infiltration is assumed to be mainly preferential, whereas the fast reservoir generally represents drainage from agricultural land. In the hillslope areas, the infiltration is assumed to be preferential as well, but here a fast reservoir is included that accounts for fast subsurface flow. In addition, a dead storage is added to the fast reservoir, as in karstic areas fissures need to fill first before fast runoff is gener-

ated. In the wetlands, the main difference is introduced by capillary rise, which will occur due to the shallow groundwater tables. In addition to these three main landscape classes, the division in grass and forest was made, as these landcovers introduce strong differences in interception capacity and root-zone storage capacity. The hillslopes were additionally split in south-facing and north-facing hillslopes, due to the different amount of evaporative energy that both categories receive. The snow module of the glacier class differed from the other snow modules as here the snow storage is assumed to be infinite. However, the glacier and bare rock classes used a similar model structure, accounting for mainly fast processes. Here, the unsaturated storage is conceptualized as a threshold process, as in these areas cracks and fissures fill up first, before runoff is generated. See also Figure 5.2 for a schematization of the model structure.

Here, a gridded version of model the model was used, and the landscape classes introduced sub-grid variability for the modelling cells. The model was applied on the European continent from 11^o W until 20.5^o E, and from 35.5^o N until 60^o N, in resolution of 0.25^o following the resolution of the input precipitation data of MSWEP (Beck et al., 2017). Daily potential evaporation was determined using ERA-Interim data (Dee et al., 2011) with the Penman formulation as prescribed by FAO (Allen et al., 1998). The air temperature of the ERA-Interim data was also used for the snow modelling, as well as SRTM elevation data (Lehner et al., 2008), which was used to derive elevation zones of 100m difference with corrected temperature values. The model was run from 1-1-1999 until 31-12-2003, where the first year served as warm-up year and the remaining three years were used for model evaluation.

5.2.2. Model parameter constraints

The model was run with 50 randomly generated parameter sets that fulfilled several semi-quantitative prior parameter constraints, similar to Gharari et al. (2014) and Hrachowitz et al. (2014). These conditions were based on assumed catchment behavior between the different landscape classes, i.e. expert knowledge (Gharari et al., 2014), or remote sensing products such as GRACE (Tapley et al., 2004). The same model was also run 50 times with randomly generated parameters from a uniform prior distribution, which served as a benchmark for assessing the added value of the constraints.

Slow reservoir coefficient

The GRACE satellites provide data about total water storage anomalies, which strongly relate to groundwater. Therefore, it can be argued that this signal can be used in a modified recession analysis. When the signal of GRACE is considered as a proxy for the current total storage in the basin S, the rate of change over time can easily be determined. Only decreasing changes over time were selected, pointing at a period of recession, for the full time series of GRACE. A formal recession analysis uses generally only dry periods, but due to the temporal resolution of GRACE, it was assumed here that precipitation and evaporation are relatively equal over this timestep. Thus, the recessions were only selected based on the require-

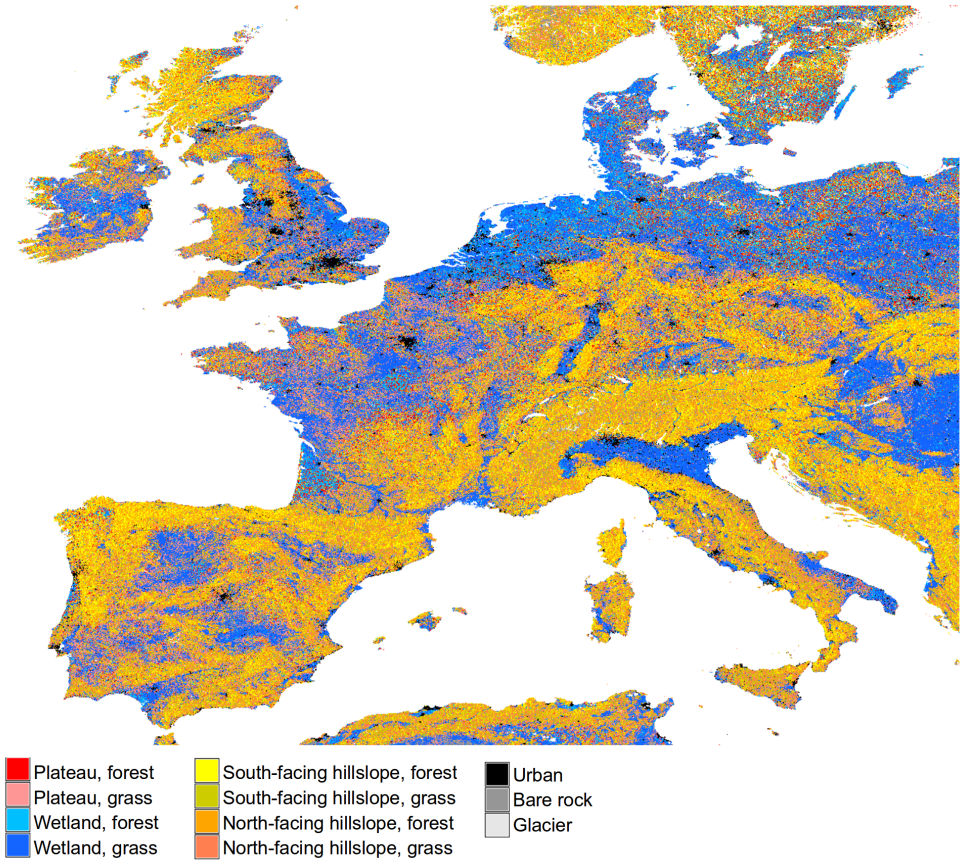

Figure 5.1: Landscape classification of Europe for the different classes of wetland and plateau for grass and forest, hillslope for grass and forest, north- and south-facing, and urban areas.

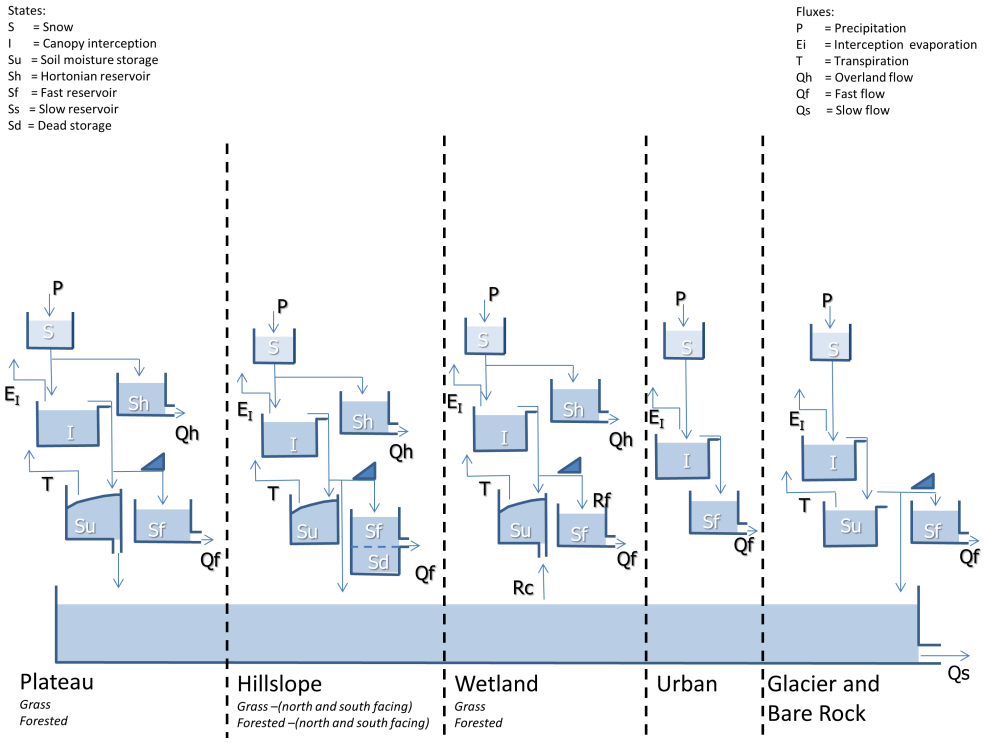

Figure 5.2: Landscape classification of Europe for the different classes of wetland and plateau for grass and forest, hillslope for grass and forest, north- and south-facing, and urban areas.

ment that these had to be decreasing in time, ignoring precipitation and evaporation effects, which led to around 50 – 100 datapoints as input for the recession analysis per GRACE cell. The modified recession analysis can now be carried out as a simple linear regression between the proxy for storage and change in storage (i.e. a linear reservoir):

$$\frac{dS}{dt} = -K_s * S, \tag{5.1}$$

with S the storage, t the time, K_s the reservoir coefficient. In this way, the linear regression between the storage and storage change over time, gives an estimate of K_s. Maximum and minimum values for the prior range were determined based on the standard error from the linear regression. See Figure 5.3 for an example of the procedure and the obtained maximum and minimum values for this parameter.

Root-zone storage capacity

The prior parameter ranges for root-zone storage capacity, $S_{u,max}$, were determined based on the dataset of Wang-Erlandsson et al. (2016). The root-zone storage capacities in this dataset were determined with a similar method as described by Gao et al. (2014a), de Boer-Euser et al. (2016) and Nijzink et al. (2016a), but in a global setting. Briefly, the precipitation data from CRU and CHIRPS were used with the evaporation estimates from SSEBop and MOD16 to determine the deficits between evaporation and precipitation. These deficits, or demands, need to be overcome by the vegetation and vegetation designs it root system following certain return periods of these deficits. Following this, the authors concluded that grassland generally needed root-zone storage capacities with relatively short return periods, whereas forested areas needed longer return periods and thus larger storages. Therefore, the root-zone storage capacity will be conditioned accordingly:

$$S_{R,2yr} < S_{u,max,grass} < S_{R,5yr}, \tag{5.2}$$

$$S_{R,10yr} < S_{u,max,forest} < S_{R,60yr}, \tag{5.3}$$

With $S_{u,max,grass}$ and $S_{u,max,forest}$ the model parameters of maximum root-zone storage capacities for the grassland and forested versions of each landscape class, $S_{R,2yr}$, $S_{R,5yr}$, $S_{R,10yr}$, $S_{R,60yr}$ the root-zone storage capacities with a 2, 5, 10 and 60 year return period respectively. See Figure 5.4 for the maximum and minimum values for grassland and forested areas for this parameter.

Maximum interception capacity

The interception capacity of trees can generally be considered higher than the interception capacity of grassland, due to the increased leaf area. Thus, the following constraint was imposed on the parameter I_{max}:

$$I_{max,grass} < I_{max,forest}, \tag{5.4}$$

With $I_{u,max,grass}$ and $I_{u,max,forest}$ the model parameters of maximum interception capacities for the grassland and forested version of each landscape class.

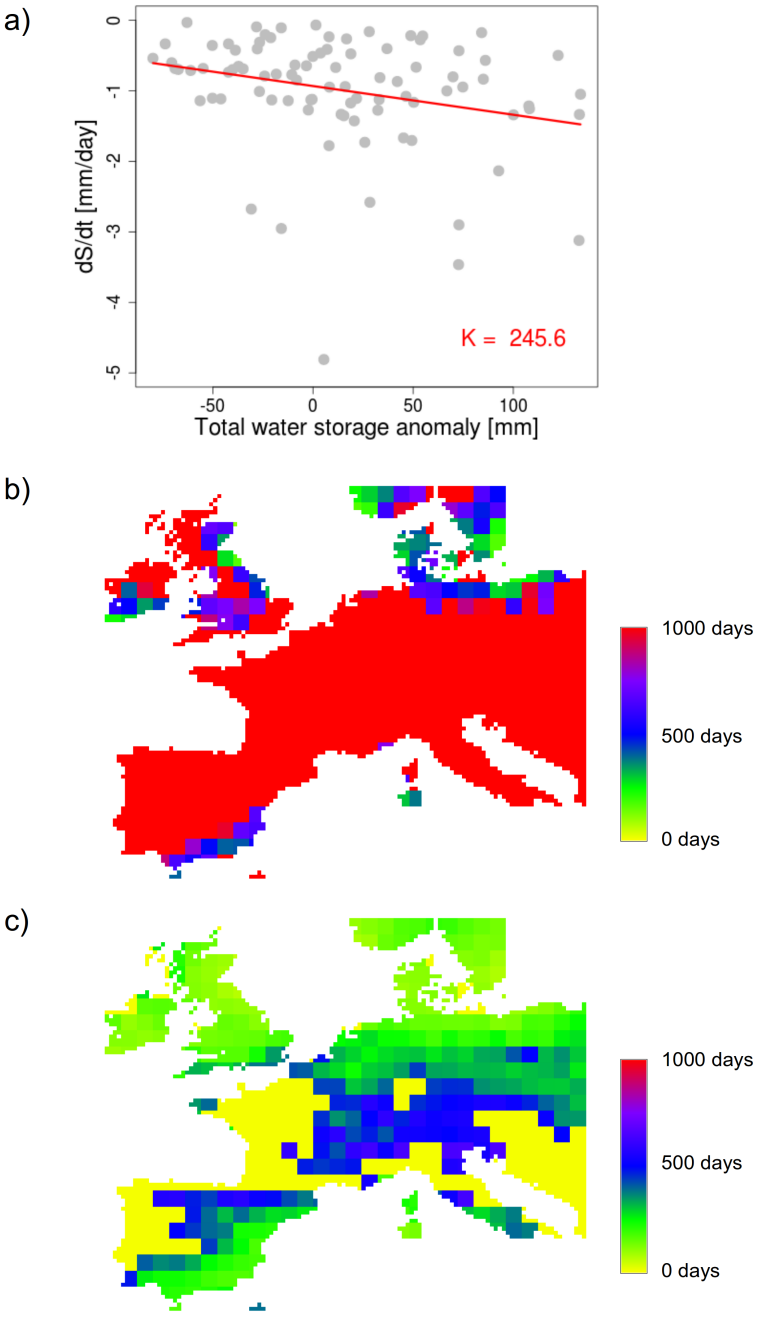

Figure 5.3: Estimates of the slow reservoir coefficient with a) example of regression between GRACE total water storage anomalies and the differences of time of GRACE water storage anomalies, b) maximum values of the slow reservoir coefficient, c) minimum values of the slow reservoir coefficient.

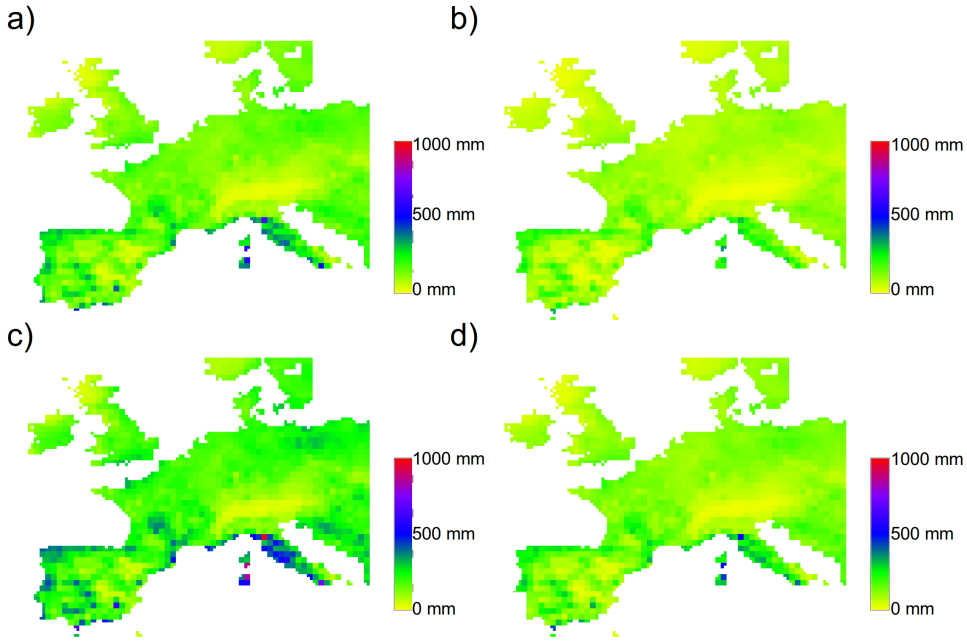

Figure 5.4: Estimation of the root zone storage capacity parameter of $S_{u,max}$ with a) maximum values for grassland, b) minimum values for grassland c) maximum values for forest and d) minimum values for forest.

Hortonian overland flow parameters

Overland flow on a grass area will experience less resistance compared to a forested area. Therefore, the reservoir coefficient for the very fast Hortonian reservoir should follow these concepts as well:

$$K_{H,grass} < K_{H,forest}, \tag{5.5}$$

With $K_{H,grass}$ and $K_{H,forest}$ the reservoir coefficients for the Hortonian reservoir for the grassland and forested version of each landscape class. Similarly, it can also be argued that the threshold intensity before overland flow starts to occur differs for the different land cover types. As trees generate more preferential infiltration, mainly due to their root systems, it can be assumed that the threshold intensity is higher for forested areas:

$$Inf_{max,grass} < Inf_{max,forest}, \tag{5.6}$$

With $Inf_{max,grass}$ and $Inf_{max,forest}$ the rainfall intensity thresholds determining when Hortonian overland starts to occur, for the grassland and forested version of each landscape class.

Lag functions

The model uses several lag functions before the water enters the fast reservoir that represents fast subsurface flow (Figure 5.2). The subsurface flow for forested areas will be relatively fast, where preferential flow paths may exist due to roots of trees. In contrast, the soil layer below grassland may consist of a more homogeneous layer, which introduces more matrix flow with lower velocities. Thus:

$$n_{lag,grass} < n_{lag,forest}, \tag{5.7}$$

5.2.3. Model evaluation

A selection 397 gauging stations from the Global Runoff Data Centre (GRDC) was used for evaluation of the model. This was done in a more traditional way in terms of Nash-Sutcliffe efficiency of the flows and the log of the flows, but also for a larger set of hydrological signatures. As the constrained and unconstrained model both were run 50 times, distributions of model performances exist for both cases. Thus, the probability of improvement can be calculated with respect to the unconstrained model (Nijzink et al., 2016b):

$$\begin{aligned} P_{I,S} &= P\left(S_C > S_U\right) \\ &= \sum_{i=1}^{n} P\left(S_C > S_U \mid S_C = r_i\right) P\left(S_C = r_i\right), \end{aligned} \tag{5.8}$$

where S_C and S_U are the distributions of the signature performance metrics of the constrained model and reference unconstrained model, respectively, for the

set of all 50 solutions, r_i is a single realization from the distribution of S and n is the total number of realizations of the S distribution. For $P_{I,S} > 0.5$ it is then more likely that the model outperforms the reference model with respect to the signature under consideration, and vice versa for $P_{I,S} < 0.5$. A similar approach can be applied for a set of 27 hydrological signatures, see Table 5.1. Also here the probability of improvement can be calculated between the distributions of the constrained and unconstrained models, but then for the objective function values for the hydrological signature under consideration.

5.3. Results and Discussion

5.3.1. Model objective function values

In Figure 5.5, the objective function values are displayed in the constrained and unconstrained situation. Generally, the patterns are in both cases rather variable, with catchments with high performances and low performances. The improvements obtained by constraining are hard to observe, or even absent, and the general pattern for the different catchments remains similar, which is in contrast with several other studies (e.g. Gao et al., 2014a; Gharari et al., 2014; Hrachowitz et al., 2014) that found strong improvements when constraining a model in a similar way. However, deteriorations in performance metrics considering streamflow does not necessarily mean that model internal dynamics are less well represented as well (Pokhrel et al., 2008). Moreover, as the applied parameter constraints force the model towards a more hydrologically consistent behavior, internal model dynamics, or other outgoing fluxes, may still be improved. Following the latter, it can even be argued that the similar model performances in the unconstrained and constrained situation (Figure 5.5) only reflect that values of performance metrics can be deceptive (e.g. de Boer-Euser et al., 2017), as the parameterizations do not necessarily lead to hydrologically consistent model behaviour but a mathematical fit of the data (Kirchner, 2006). In order to obtain a more general conclusion, in Figure 5.6 the model performances for all 397 catchments are merged together for the different objective functions for the constrained and unconstrained situation. Also here, it becomes apparent that there are no major differences between the two approaches. Thus, a more step-wise approach might be needed to achieve a more robust model calibration (Fenicia et al., 2008; Winsemius et al., 2009), which has been carried out for large-scale models as well (e.g. Lopez Lopez et al., 2017; Pechlivanidis and Arheimer, 2015). This can be done in a following more formal calibration step (e.g. Gao et al., 2014a; Kuzmin et al., 2008), but an additional set of constraints which evaluates the model outcomes, i.e. process constraints, may also be considered. Concerning the latter approach, it was shown by Hrachowitz et al. (2014) that process constraints add more information than prior parameter constraints.

In Figure 5.7, the probability of improvement across Europe is shown for the Euclidian distance (Figure 5.7a) between Nash-Sutcliffe of flows (Figure 5.7b) and log of the flows (Figure 5.7c). For all three figures, it can be seen that especially the more urbanized areas have low probabilities of improvement, for example around

Table 5.1: Overview of the hydrological signatures.

Signature	Description	Reference
S_{QMA}	Mean annual runoff	
S_{AC}	One day autocorrelation coefficient	Montanari and Toth (2007)
$S_{\text{AC,summer}}$	One day autocorrelation the summer period	Euser et al. (2013)
$S_{\text{AC,winter}}$	One day autocorrelation the winter period	Euser et al. (2013)
S_{RLD}	Rising limb density	Shamir et al. (2005)
S_{DLD}	Declining limb density	Shamir et al. (2005)
S_{Q_5}	Flow exceeded in 5 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{50}}$	Flow exceeded in 50 % of the time	Jothityangkoon et al. (2001)
$S_{Q_{95}}$	Flow exceeded in 95 % of the time	Jothityangkoon et al. (2001)
$S_{Q_5,\text{summer}}$	Flow exceeded in 5 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{50},\text{summer}}$	Flow exceeded in 50 % of the summer time	Yilmaz et al. (2008)
$S_{Q_{95},\text{summer}}$	Flow exceeded in 95 % of the summer time	Yilmaz et al. (2008)
$S_{Q_5,\text{winter}}$	Flow exceeded in 5 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{50},\text{winter}}$	Flow exceeded in 50 % of the winter time	Yilmaz et al. (2008)
$S_{Q_{95},\text{winter}}$	Flow exceeded in 95 % of the winter time	Yilmaz et al. (2008)
S_{Peaks}	Peak distribution	Euser et al. (2013)
$S_{\text{Peaks,summer}}$	Peak distribution summer period	Euser et al. (2013)
$S_{\text{Peaks,winter}}$	Peak distribution winter period	Euser et al. (2013)
$S_{Q_{\text{peak},10}}$	Flow exceeded in 10 % of the peaks	
$S_{Q_{\text{peak},50}}$	Flow exceeded in 50 % of the peaks	
$S_{Q_{\text{summer,peak},10}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{summer,peak},50}}$	Flow exceeded in 10 % of the summer peaks	
$S_{Q_{\text{winter,peak},10}}$	Flow exceeded in 10 % of the winter peaks	
$S_{Q_{\text{winter,peak},50}}$	Flow exceeded in 50 % of the winter peaks	
S_{SFDC}	Slope flow duration curve	Yadav et al. (2007)
S_{LFR}	Low flow ratio (Q_{90} / Q_{50})	
S_{FDC}	Flow duration curve	Westerberg et al. (2011)

Paris, London and the Ruhr-area. In the center of Europe, around the Alps, several stations with high probabilities of improvement can be observed. In this area the groundwater parameter Ks was constrained relatively strong, based on GRACE (Figure 5.3), and these improvements can probably be contributed to this. The highest probabilities of improvement can however be found in northern Europe, where relatively large parts of the wetland landscape classes are found. Nevertheless, the patterns of high and low probabilities remain rather random, indicating that the applied semi-quantitative prior constraints are not powerful enough in combination with the number of samples (50).

5.3.2. Signature analysis

The probabilities of improvement for a set of 27 hydrological signatures (Table 5.1) show a rather random pattern of improvements and deteriorations as well (Figure 5.8 and 5.7). It can be argued that this points again at an insufficient number of model runs, or constraints which do not contain enough discriminative power. The risk that the parameter search space is not efficiently sampled is also rather big considering the number of parameters (Beven, 2006; Spear et al., 1994), especially as this number is increased due to the additional landscape-classes as sub-grid variability. However, when the catchments are ordered according to size in Figure 5.8, from small catchments on top to large catchments at the bottom, a pattern can be observed. Colors become more distinct blue or red towards larger catchments, which shows that the rather course resolution of $0.25°$ matters. Thus, for the smaller catchments the signatures are represented more equally good or bad, leading to probabilities of around 0.5 (white colors in Figure 5.8), whereas the differences become bigger (i.e. deep blue or deep red) towards larger catchments. Logically, the chosen model resolution seems more applicable for the larger catchments.

In addition to the resolution, in Figure 5.9 the latitude is used to order the catchments. Around higher values of latitudes the most blue areas can be observed, representing high probabilities of improvement. In these areas with a more moderate climate the applied constraints apparently help to improve the signature representation. Similarly, looking more at the individual signatures, it can be noted that especially the winter flows improve ($S_{Q5,winter}$, $S_{Q50,winter}$, $S_{Q95,winter}$ and $S_{Q,winter,peak,10}$, $S_{Q,winter,peak,50}$). In other words, the model shows improvements in a rather wet situation, which is also supported by the observed improvements around latitudes with a more moderate climate. This seems to confirm the findings of (de Boer-Euser et al., 2016), who applied root-zone storage capacities derived in a similar way as the values used here as constraints. Also in this study, relatively wet systems improved by using root-zone storage capacities derived by climate as initially done by (Gao et al., 2014b). Nevertheless, most of the constraints should effect merely fast flows and evaporation (e.g. constraints for $S_{u,max}$, I_{max}) and GRACE may still have an influence on these findings as well. Besides the influence of the constraints, the information content of hydrological signatures is often used to assess the suitability of the model structure (Euser et al., 2013; Hrachowitz et al., 2014). Thus, even though the model structure is complex and covers a large range of processes, it may still be insufficient in a number of cases, and flexible

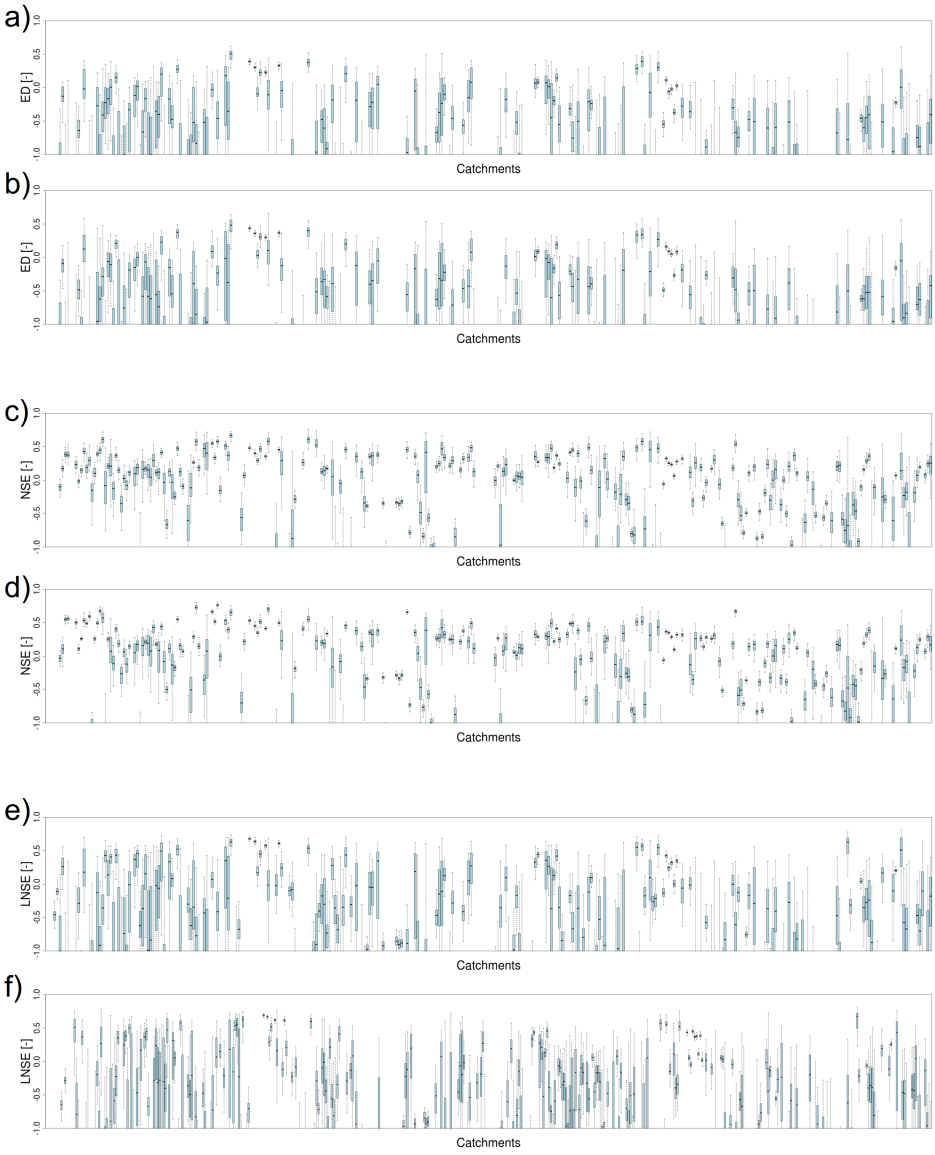

Figure 5.5: Objective function values in terms of Euclidian distance between Nash-Sutcliffe of flows and log of the flows for the a) unconstrained and b) constrained model, Nash-Sutcliffe of flows for the c) unconstrained and d) constrained model, and Nash-Sutcliffe of the log of the flows for the e) unconstrained and f) constrained model.

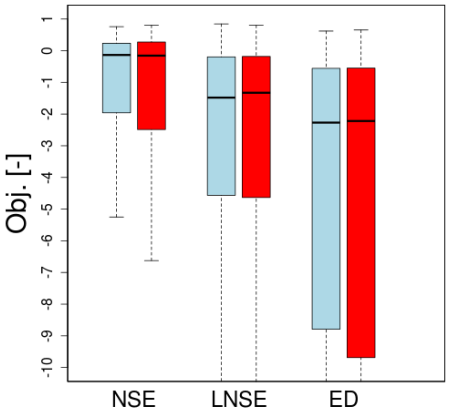

Figure 5.6: Objective function values in terms of the Nash-Sutcliffe of flows and log of the flows and Euclidian distance between both objective function values for the unconstrained (blue) and constrained model (red), for all model realizations.

modelling approaches may be needed (Fenicia et al., 2011). Especially across the European continent, regions may behave hydrologically different, depending also on various explanatory factors (Kuentz et al., 2017). Thus, a South-European wetland versus a wetland in northern Europe may act differently. At the same time, the landscape units applied here were defined with fixed threshold values, whereas these thresholds may differ spatially and temporally, as for example follows from the variable contribution area theory (Dunne et al., 1975).

5.4. Conclusions

A large-scale hydrological model with landscape-derived sub-grid variability was applied for the European continent, where the landscape-based units allowed for semi-quantitative prior parameter constraints for model calibration. The model was run with 50 random parameter sets with and without the prior parameter constraints, which led to a variable pattern in model performances for both approaches. Even though model performances were variable, in terms of Nash-Sutcliffe efficiency of flows and the log of the flows as well as the Euclidian distance between the two metrics, the patterns of high and low performances were consistent over the catchments between the two approaches. Thus, similar performances are obtained without prior parameter constraints, which points at constraints which are too weak. Nevertheless, the ability of hydrologically inconsistent models, i.e. unconstrained models, to obtain similar results, points also at the importance of the constraints to filter out these assumedly incorrect models. In an additional evaluation of 27 hydrological signatures, high probabilities of improvement were found for rather wet, moderate catchments, but also here a rather variable pattern was observed. These findings, with rather random patterns of improvements and deteriorations,

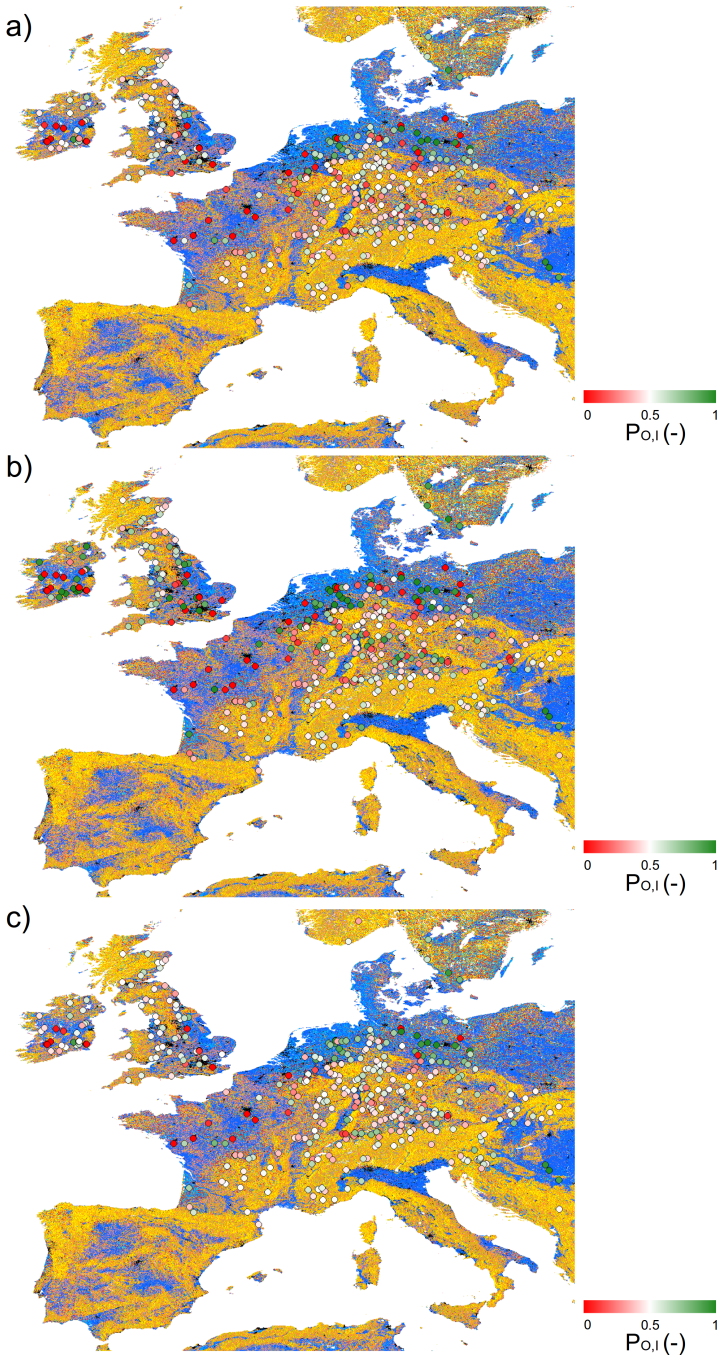

Figure 5.7: Objective function values in terms of the Nash-Sutcliffe of flows and log of the flows and Euclidian distance between both objective function values for the unconstrained (blue) and constrained model (red), for all model realizations.

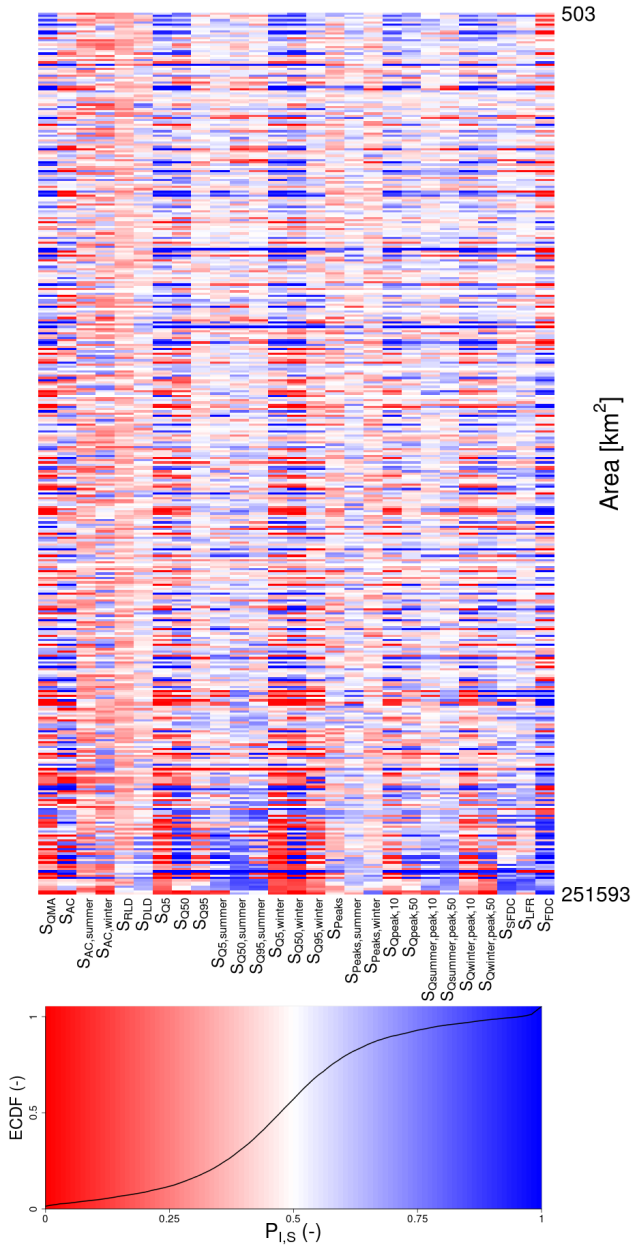

Figure 5.8: Probabilities of improvement for 27 signatures and 397 catchments, ordered from small (top) to large catchments (bottom).

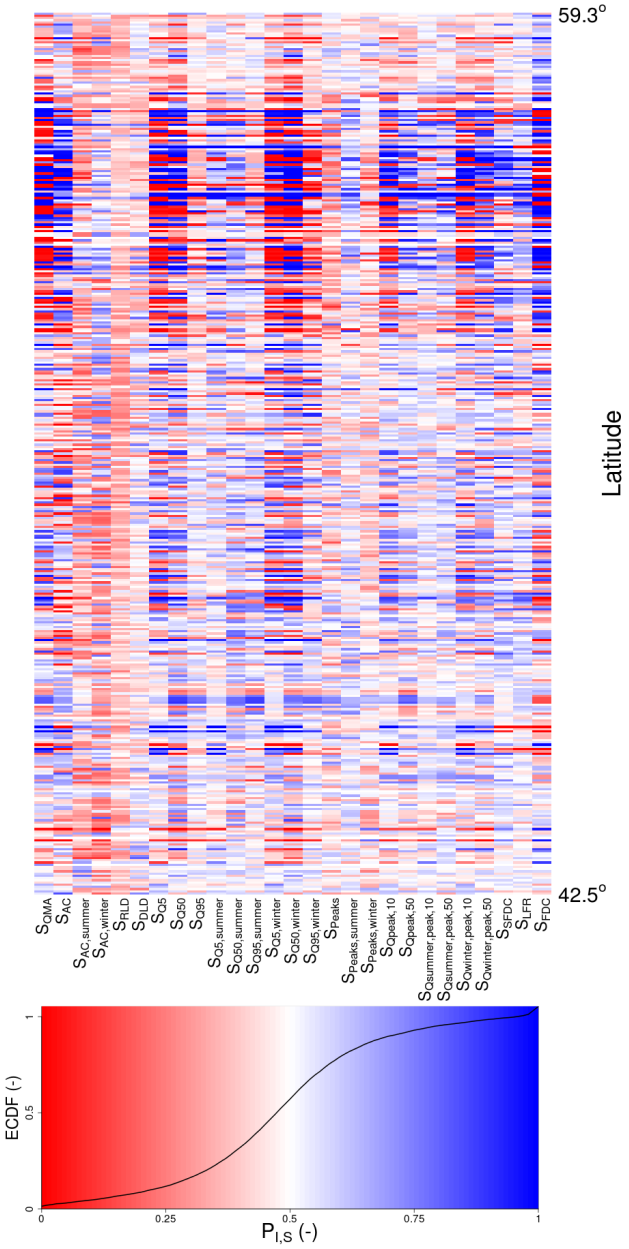

Figure 5.9: Probabilities of improvement for 27 signatures and 397 catchments, ordered from lower latitudes (top) to higher latitudes (bottom).

indicates that the parameter search space was not sampled effectively in combination with the relatively low number of 50 samples, but it also shows that the applied semi-quantitative parameter constraints were not sufficient enough in reducing the size of this search space. Nevertheless, even though no strong improvements were observed, the findings presented here still show that performance measures can be deceptive, as these can also be achieved without conditioning a model with hydrological expert knowledge.

6

Conclusions, implications and outlook

This thesis focused on reducing the need for calibration of different conceptual hydrological models. Model regularization was adjusted for landscape-based modelling, time-dynamic key parameters were formulated and additional data sources were explored, all in order to obtain more robust model conceptualizations and parameterizations.

6.1. Conclusions

In a first step, a regularized hydrological model, mHM, was adjusted in order to account for landscapes as sub-grid variability. The addition of this sub-grid heterogeneity in the existing model was complemented with semi-quantitative constraints, i.e. constraints on (seasonal) runoff coefficients. Seven distinct catchments across Europe were selected as test cases, from which four were employed as 'donor' catchments and the remaining three as 'receiver' catchments. Here, it was shown that adding landscape features as sub-grid heterogeneity moderately improved the existing model mHM. Evaluating a set of hydrological signatures showed that mainly low flow dynamics improved for mHMtopo, caused by the additional interactions between groundwater and soil moisture. The semi-quantitative constraints were more effective and led to a better partitioning between evaporation and runoff, especially in catchments where these are out of phase. The landscape-based sub-grid heterogeneity and semi-quantitative constraints were also shown to be helpful in transferring the models to catchments without any further calibration.

On the more detailed scale of a single parameter, three deforested, experimental catchments (HJ Andrews WS1, Hubbard Brook WS2 and WS5) were used to assess the dynamic character of root-zone storage capacities. Time series of root-zone storage capacities were estimated based on a simple water balance approach and the calibration of four distinct conceptual models for subsequent 2-year windows

after deforestation. The two approaches both showed a strong time-dynamic behavior of root-zone storage capacity after deforestation. At first, a strong decrease was observed, after which a period of gradual recovery started. This recovery took 5-13 years for the catchments under consideration, as found by a trend analysis on the root-zone storage data. Thus, the root-zone parameter of one of the models was adjusted with a time-dynamic formulation. Moderate improvements were obtained in terms of objective function values, but the time-dynamic model showed more improvements in a signature analysis.

In addition to the direct estimation of a single key parameter and the indirect estimation of the full set of parameters by model regularization, 10 different additional information sources were employed to derive model parameter posteriors. The added value of different combinations of products and analytical frameworks was explored for five different conceptual hydrological models for 27 catchments across Europe. The combinations were assessed with regard to objective function values, but also hydrological signatures. In general, most models profited from including more data sources to derive new posterior parameter ranges, but a certain degree of information saturation was obtained after 5-4 additional data sources. The results suggest that including GRACE in the combination of data sources has a strong impact, as high probabilities of improvement were found. The soil moisture products AMSR-E and ASCAT also showed a high added value in the parameter selection procedure. In this way, it was shown that using a combination of data sources simultaneously in an indirect way has a high added value.

In a last step, a large-scale application of a landscape driven model was applied over the European continent. Here, landscape classes on the sub-grid scale allowed for relational conditions between different parameters, i.e. semi-quantitative prior parameter constraints. The model was run with 50 random parameterizations with and without the parameter constraints. A rather variable pattern was observed in performances over a selection of 397 catchments, for both the constrained and unconstrained case. Similarly, a variable pattern in probabilities of improvement was observed for a set of 27 hydrological signatures. The constraining power of the applied constraints in combination with the relatively low number of 50 samples can therefore be considered as not sufficient to reduce the parameter search space. However, the results between the constrained and unconstrained model were similar, pointing also at the deceptive nature of performance metrics, as the unconstrained model still uses hydrologically inconsistent parameter sets.

6.2. Implications and Outlook

Based on the application of a regularized model with additional sub-grid variability, it can be argued that the choice of the model structure remains important. Even though the more mathematical approach of model regularization is extremely powerful in order to apply models in ungauged areas, still the final model performance depends on the included processes for the catchments under consideration. This important role of model structure has been recognized before (e.g. Clark et al., 2016; Euser et al., 2013; Fenicia et al., 2008), but still the use of a single model structure reduces efforts and helps to learn from the known model deficiencies

(Le Moine et al., 2007; Linsley, 1982). The use of regionalization techniques is generally complicated in case of fully flexible model structures, as model structures and parameters may differ per location, but here it is shown that model hypothesis testing with regularized hydrological models is at least very well possible. Still, the value of these regionalization approaches for practical and operational uses is eminent, but more specific catchment understanding can probably only be obtained with more detailed site-specific modeling (i.e. the uniqueness of place; Beven, 2000). On the other hand, links between data, such as land use, soil and vegetation, and model parameters can also be considered as part of the theoretical underpinning of our hydrological models (Clark et al., 2016), and, define the dominant processes, which are now often explicitly included or excluded in site specific modelling. Thus, when all possible combinations of hydrological processes are included in a model, and the right data is used to identify which processes are triggered and which parameterizations are needed for the model, any catchment could be modelled. In other words, a unified catchment theory is found (e.g. Sivapalan, 2006; Troch et al., 2009b), which is currently still rather far away. The results presented here thus reflected one of the many steps towards bridging the gap between site-specific flexible modelling on a small scale and large-scale modeling with a general model structure.

The general applicability of a model structure also strongly depends on the time-dynamic nature of the catchments under consideration. The behavior of model parameters, especially in times of change, is often reflected in the uncertainty of parameter priors (Wagener et al., 2007). Therefore, constraining a key parameter, such as the root-zone storage capacity, effectively and a priori, is highly needed to reduce model uncertainty. The simple water-balance method presented in this thesis and previous studies (de Boer-Euser et al., 2016; Gao et al., 2014b; Wang-Erlandsson et al., 2016) can be extremely valuable for hydrological model applications, as it was shown that effective and even time-dynamic estimates for different models can be achieved in this way. The importance of understanding parameters that may change over time (e.g. Coron et al., 2012; Gharari et al., 2013) may be apparent, especially as an improved understanding of these dynamics may help model formulations. Eventually, instead of just changing the value of a parameter for periods with different circumstances, these systematic changes should directly "emerge" from a suitable parameterization (e.g. Arora, 2002; Westra et al., 2014). The simple formulation of growth in this thesis could be seen as an additional conceptual equation, taking care of (a part of) the vegetation dynamics, but a better understanding of these dynamics are still needed. Several hydrological models exist that incorporate vegetation dynamics, such as TOPOG (Vertessy et al., 1996) or DHSVM (Wigmosta et al., 1994), but these models face the risk of overparameterization (Schulz et al., 2001; Tang and Zhuang, 2008). Simple model formulations that yet reflect the time dynamic character of ecosystems still need to be formulated carefully, and this thesis presents a first simple proof-of-concept of this.

A better understanding of catchment processes can also be achieved by the large number of new datasets, especially from remote sensing, that provide the hydrological sciences with new information sources. Here, the challenge exists to

assess the added value of the new products, and combinations thereof. Single products had been used for hydrological modeling purposes quite extensively (e.g. Brocca et al., 2010; Werth et al., 2009), nowadays the focus starts to shift towards the use of multiple products (e.g. Lopez Lopez et al., 2017; Tian et al., 2017), and in this thesis, nine products and an analytical framework were assessed. New products are however planned and computational resources still increase, hence it can be expected that more approaches will start to appear with even a larger number of products. Following also the reasoning from before, additional data sources in hydrological modelling may help for the underpinning and hydrological theory of the models (Clark et al., 2016; Sivapalan, 2006), but the formulation of functional relations and data may be problematic (Merz and Blöschl, 2004). New innovative approaches are thus needed to efficiently use all the information available, and, at the same time, filter out erroneous information, which in this thesis, as a first suggestion, was both done by the definition of combined objective functions and a likelihood weighting approach. At least, this thesis clearly shows that the use of an ensemble of remote sensing products and/or analytical frameworks provides a way forward for especially predictions in ungauged basins, but also the equifinality issue.

The equifinality issue was also addressed in this thesis with the applied semi-quantitative prior parameter constraints for a large-scale landscape-driven model. However, the prior parameter constraints did not proof effective in combination with a sample size of 50 random parameterizations. It comprised a first attempt in order to apply more complex model structures on a larger scale, that need a strong reduction in the parameter search space. A step-wise calibration might be more effective here (Fenicia et al., 2008; Lopez Lopez et al., 2017) as well as additional process based constraints (Gharari et al., 2014; Hrachowitz et al., 2014) that can be evaluated afterwards. However, the findings here still address the need for hydrological consistency in more complex model settings (e.g. Euser et al., 2013) as similar performances in constrained and unconstrained situations are found.

This thesis showed several options in the process of obtaining more robust model parameterizations, for different conceptual models, spatial scales and changing circumstances, and provides, hopefully, enough new starting points to bridge the gap between site-specific understanding and large-scale modelling. In other words, we learned how to better optimize model formulations without using observed river discharge. Especially in locations with insufficient data (i.e. ungauged catchments), this thesis provides several handles to start from, either with data from other catchments (i.e. model regionalization/regularization), simple (time-dynamic) estimates of parameters, remotely sensed data and hydrological understanding. Even though fully applying models without calibration on river discharge data has not been achieved to a satisfying degree in all cases, this thesis provides new incremental steps to achieve this in the future. Especially as computational resources increase and new remote sensing products become available, the techniques presented in this thesis could be employed on a much larger scale, and, hence, be improved. Especially changes in catchments, either caused by humans or nature, deserve attention, whereas this thesis just provides a first glance of

the time-dynamic character of a specific vegetation parameter. Overall, this thesis shows that model formulations and parameterizations can very well be selected without observed river discharge, and holds a promise for a future with predicting in ungauged catchments.

References

Abdulla, F. and Lettenmaier, D.: 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.

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities, Reviews of Geophysics, 53, 452–480, doi:10.1002/2014RG000456, 2014RG000456, 2015.

Alila, Y., Kuraś, P., Schnorbus, M., and Hudson, R.: Forests and floods: A new paradigm sheds light on age-old controversies, Water Resources Research, 45, W08 416, doi:10.1029/2008WR007207, 2009.

Allen, C., Macalady, A., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D., Hogg, E. H., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S., Semerci, A., and Cobb, N.: A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, Forest Ecology and Management, 259, 660–684, doi:http://dx.doi.org/10.1016/j.foreco.2009.09.001, (Ted), 2010.

Allen, R., Pereira, L., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, D05 109, 1998.

Allison, G. B., Cook, P. G., Barnett, S. R., Walker, G. R., Jolly, I. D., and Hughes, M. W.: Land clearance and river salinisation in the western Murray Basin, Australia, Journal of Hydrology, 119, 1–20, doi:http://dx.doi.org/10.1016/0022-1694(90)90030-2, 1990.

Almeida, S., Le Vine, N., McIntyre, N., Wagener, T., and Buytaert, W.: Accounting for dependencies in regionalized signatures for predictions in ungauged catchments, Hydrology and Earth System Sciences, 20, 887–901, doi:10.5194/hess-20-887-2016, 2016.

Ambroise, B., Beven, K., and Freer, J.: Toward a Generalization of the TOPMODEL Concepts: Topographic Indices of Hydrological Similarity, Water Resources Research, 32, 2135–2145, doi:10.1029/95WR03716, 1996.

Andersson, L. and Arheimer, B.: Consequences of changed wetness on riverine nitrogen – human impact on retention vs. natural climatic variability, Regional Environmental Change, 2, 93–105, doi:10.1007/s101130100024, 2001.

Andréassian, V.: Waters and forests: from historical controversy to scientific de-

bate, Journal of Hydrology, 291, 1–27, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.12.015, 2004.

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.

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.

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

Arora, V.: MODELING VEGETATION AS A DYNAMIC COMPONENT IN SOIL-VEGETATION-ATMOSPHERE TRANSFER SCHEMES AND HYDROLOGICAL MODELS, Reviews of Geophysics, 40, 3–1–3–26, doi:10.1029/2001RG000103, 1006, 2002.

Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrology and Earth System Sciences, 11, 703–710, doi:10.5194/hess-11-703-2007, 2007.

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., and Anderson, C.: Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), Geophysical Research Letters, 34, L20 401, doi:10.1029/2007GL031088, l20401, 2007.

Bathurst, J. C., Ewen, J., Parkin, G., O'Connell, P. E., and Cooper, J. D.: Validation of catchment models for predicting land-use and climate change impacts. 3. Blind validation for internal and outlet responses, Journal of Hydrology, 287, 74–94, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.09.021, 2004.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, doi:10.5194/hess-21-589-2017, 2017.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments. SMHI Reports RHO, Report, SMHI, 1976.

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

Beven, K.: 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.: 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 experimentMOPEXMOPEX workshop, 2006.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation

in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11 – 29, doi:http://dx.doi.org/10.1016/S0022-1694(01)00421-8, 2001.

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 physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43–69, doi:10.1080/02626667909491834, 1979.

Black, P.: WATERSHED FUNCTIONS1, JAWRA Journal of the American Water Resources Association, 33, 1–11, doi:10.1111/j.1752-1688.1997.tb04077.x, 1997.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: Runoff prediction in ungauged basins: synthesis across processes, places and scales, Cambridge University Press, 2013.

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.

Bosch, J. M. and Hewlett, J. D.: A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration, Journal of Hydrology, 55, 3–23, doi:http://dx.doi.org/10.1016/0022-1694(82)90117-2, 1982.

Boyle, D.: Multicriteria calibration of hydrologic models, 2001.

Brath, A., Montanari, A., and Moretti, G.: Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty), Journal of Hydrology, 324, 141–153, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.10.001, 2006.

Bréda, N., Huc, R., Granier, A., and Dreyer, E.: Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences, Ann. For. Sci., 63, 625–644, 2006.

Breuer, L., Eckhardt, K., and Frede, H.-G.: Plant parameter values for models in temperate climates, Ecological Modelling, 169, 237–293, doi:http://dx.doi.org/10.1016/S0304-3800(03)00274-6, 2003.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., and Hasenauer, S.: Improving runoff prediction through the assimilation of the ASCAT soil moisture product, Hydrology and Earth System Sciences, 14, 1881–1893, doi:10.5194/hess-14-1881-2010, 2010.

Brown, A., Zhang, L., McMahon, T., Western, A., and Vertessy, R.: A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, Journal of Hydrology, 310, 28–61, doi:http://dx.doi.org/10.1016/j.jhydrol.2004.12.010, 2005.

Brown, M. E., Escobar, V., Moran, S., Entekhabi, D., O'Neill, P. E., Njoku, E. G., Doorn, B., and Entin, J. K.: NASA's Soil Moisture Active Passive (SMAP) Mission and Opportunities for Applications Users, Bulletin of the American Meteorological

Society, 94, 1125–1128, doi:10.1175/BAMS-D-11-00049.1, 2013.
Brunner, I., Herzog, C., Dawes, M., Arend, M., and Sperisen, C.: How tree roots respond to drought, Frontiers in Plant Science, 6, 547, doi:10.3389/fpls.2015.00547, 2015.
Brunner, P. and Simmons, C.: HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model, Ground Water, 50, 170–176, doi:10.1111/j.1745-6584.2011.00882.x, 2012.
Budyko, M.: Climate and Life, Academic Press, San Diego, California, 1974.
Bulygina, N., McIntyre, N., and Wheater, H.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, Hydrology and Earth System Sciences, 13, 893–904, doi:10.5194/hess-13-893-2009, 2009.
Campbell, J.: Hubbard Brook Experimental Forest (USDA Forest Service): Daily Streamflow by Watershed, 1956 - present, doi:http://www.hubbardbrook.org/data/dataset.php?id=2, 2014a.
Campbell, J.: Hubbard Brook Experimental Forest (US Forest Service): Daily Precipitation Standard Rain Gage Measurements, 1956 - present, doi:http://www.hubbardbrook.org/data/dataset.php?id=13, 2014b.
Campbell, J.: Hubbard Brook Experimental Forest (USDA Forest Service): Daily Maximum and Minimum Temperature Records, 1955 - present, doi:http://www.hubbardbrook.org/data/dataset.php?id=59, 2014c.
Campbell, J.: Hubbard Brook Experimental Forest (USDA Forest Service): Daily Mean Temperature Data, 1955 - present, doi:http://www.hubbardbrook.org/data/dataset.php?id=58, 2014d.
Campbell, J. L., Bailey, A. S., Eagar, C., Green, M. B., and Battles, J. J.: Vegetation treatments and hydrologic responses at the Hubbard Brook Experimental Forest, New Hampshire, vol. Volume 2, pp. Research Paper 013: 1–9, Yale University, Global Institute of Sustainable Forestry, 2013.
Camporese, M., Daly, E., and Paniconi, C.: Catchment-scale Richards equation-based modeling of evapotranspiration via boundary condition switching and root water uptake schemes, Water Resources Research, 51, 5756–5771, doi:10.1002/2015WR017139, 2015.
Cassiani, G., Boaga, J., Vanella, D., Perri, M. T., and Consoli, S.: Monitoring and modelling of soil–plant interactions: the joint use of ERT, sap flow and eddy covariance data to characterize the volume of an orange tree root zone, Hydrol. Earth Syst. Sci., 19, 2213–2225, doi:10.5194/hess-19-2213-2015, hESS, 2015.
Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, Advances in Water Resources, 33, 1235 – 1242, doi:http://dx.doi.org/10.1016/j.advwatres.2010.04.009, special Issue on Novel Insights in Hydrological Modelling, 2010.
Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging, Hydrology and Earth System Sciences,

15, 715–727, doi:10.5194/hess-15-715-2011, 2011.
Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., and Wagener, T.: Virtual laboratories: new opportunities for collaborative water science, Hydrol. Earth Syst. Sci., 19, 2101–2117, doi:10.5194/hess-19-2101-2015, hESS, 2015.
Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, W00B02, doi:10.1029/2007WR006735, w00B02, 2008.
Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resources Research, 52, 2350–2365, doi:10.1002/2015WR017910, 2016.
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, W05 552, doi:10.1029/2011WR011721, 2012.
Crow, W. T. and Ryu, D.: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, Hydrology and Earth System Sciences, 13, 1–16, doi:10.5194/hess-13-1-2009, 2009.
Dahlgren, R. A. and Driscoll, C. T.: The effects of whole-tree clear-cutting on soil processes at the Hubbard Brook Experimental Forest, New Hampshire, USA, Plant and Soil, 158, 239–262, 1994.
Daly, C. and McKee, W.: Meteorological data from benchmark stations at the Andrews Experimental Forest, 1957 to present. Long-Term Ecological Research. Forest Science Data Bank, doi:http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=MS001, 2016.
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.
de Boer-Euser, T., McMillan, H. K., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. G.: Influence of soil and climate on root zone storage capacity, Water Resources Research, p. 2009–2024, doi:10.1002/2015WR018115, 2016.
de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G., and Willems, P.: Looking beyond general metrics for model comparison – lessons from an international model intercomparison study, Hydrology and Earth System Sciences, 21, 423–440, doi:10.5194/hess-21-423-2017, 2017.
Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R.,

Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

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.

Donchyts, G., Winsemius, H., Schellekens, J., Erickson, T and. Gao, H., Savenije, H., and van de Giesen, N.: Global 30m Height Above the Nearest Drainage, in: Geophysical Research Abstracts, vol. 18, EGU General Assembly, 2016.

Donohue, R. J., Roderick, M. L., and Mcvicar, T. R.: On the importance of including vegetation dynamics in Budyko's hydrological model, Hydrology and Earth System Sciences, 11, 983–995, doi:10.5194/hess-11-983-2007, 2007.

Donohue, R. J., Roderick, M. L., and McVicar, T. R.: Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model, Journal of Hydrology, 436–437, 35–50, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.02.033, 2012.

Dooge, J. C. I.: Looking for hydrologic laws, Water Resources Research, 22, 46S–58S, doi:10.1029/WR022i09Sp0046S, URL http://dx.doi.org/10.1029/WR022i09Sp0046S, 1986.

Douglass, J. E.: THE POTENTIAL FOR WATER YIELD AUGMENTATION FROM FOREST MANAGEMENT IN THE EASTERN UNITED STATES1, JAWRA Journal of the American Water Resources Association, 19, 351–358, doi:10.1111/j.1752-1688.1983.tb04592.x, 1983.

Du, E., Link, T. E., Wei, L., and Marshall, J. D.: Evaluating hydrologic effects of spatial and temporal patterns of forest canopy change using numerical modelling, Hydrological Processes, 30, 217–231, doi:10.1002/hyp.10591, 2016.

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.

Dunne, T., Moore, T., and Taylor, C.: Recognition and prediction of runoff-producing zones in humid regions, Hydrological Sciences Bulletin, 20, 305–327, 1975.

Dyrness, C.: Hydrologic properties of soils on three small watersheds in the western Cascades of Oregon, Res. Note PNW-111, US Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station: Portland, OR, 17, 1969.

Eagleson, P.: Climate, soil, and vegetation: 3. A simplified model of soil moisture movement in the liquid phase, Water Resources Research, 14, 722–730, doi:10.1029/WR014i005p00722, 1978.

Eagleson, P.: Ecological optimality in water-limited natural soil-vegetation systems: 1. Theory and hypothesis, Water Resources Research, 18, 325–340, doi:

10.1029/WR018i002p00325, 1982.
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.
Elliott, K. J., Caldwell, P. V., Brantley, S. T., Miniat, C. F., Vose, J. M., and Swank, W. T.: Water yield following forest-grass-forest transitions, Hydrol. Earth Syst. Sci. Discuss., 2016, 1–53, doi:10.5194/hess-2016-548, hESSD, 2016.
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., and Savenije, H. H.: The effect of forcing and landscape distribution on performance and consistency of model structures, Hydrological Processes, 29, 3727–3743, doi:10.1002/hyp.10445, hYP-14-0545.R1, 2015.
FAO/IIASA/ISRIC/ISSCAS/JRC: Harmonized World Soil Database (version 1.2), FAO and IIASA, Rome, Laxenburg, 2012.
Fatichi, S., Pappas, C., and Ivanov, V.: Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale, Wiley Interdisciplinary Reviews: Water, 3, 327–368, doi:10.1002/wat2.1125, 2016.
Feddes, R., Kowalik, P., and Zaradny, H.: Simulation of field water use and crop yield, Centre for Agricultural Publishing and Documentation., 1978.
Federer, A., Flynn, L., Martin, W., Hornbeck, J., and Pierce, R.: Thirty years of hydrometeorologic data at the Hubbard Brook Experiment Forest, New Hampshire, 1990.
Fekete, B. M. and Vörösmarty, C. J.: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements, in: Predictions in Ungauged Basins: PUB kick-off (Proceedings of the PUB kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publication, vol. 349, 2002.
Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Is the groundwater reservoir linear? Learning from data in hydrological modelling, Hydrol. Earth Syst. Sci., 10, 139–150, doi:10.5194/hess-10-139-2006, hESS, 2006.
Fenicia, F., Savenije, H., Matgen, P., and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resources Research, 44, W01 402, doi:10.1029/2006WR005563, 2008.
Fenicia, F., Savenije, H. H. G., and Avdeeva, Y.: Anomaly in the rainfall-runoff behaviour of the Meuse catchment. Climate, land-use, or land-use management?, Hydrol. Earth Syst. Sci., 13, 1727–1737, doi:10.5194/hess-13-1727-2009, hESS, 2009.
Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, W11 510, doi:10.1029/2010WR010174, w11510, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resources Research, 52, 954–989, doi:10.1002/2015WR017398, URL http://dx.doi.org/10.1002/2015WR017398, 2016.

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.

Freer, J., Beven, K., and Ambroise, B.: Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach, Water Resources Research, 32, 2161–2173, doi:10.1029/95WR03723, 1996.

Freer, J., McMillan, H., McDonnell, J., and Beven, K.: Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, Journal of Hydrology, 291, 254 – 277, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.12.037, catchment modelling: Towards an improved representation of the hydrological processes in real-world model applications, 2004.

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., and Blöschl, G.: Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology, Water Resources Research, 48, n/a–n/a, doi:10.1029/2011WR011509, URL http://dx.doi.org/10.1029/2011WR011509, w04511, 2012.

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, 2014GL061668, 2014b.

Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G., and Salvucci, G.: Interdependence of climate, soil, and vegetation as constrained by the Budyko curve, Geophysical Research Letters, 39, L19 404, doi:10.1029/2012GL053492, 2012.

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, Water Resources Research, 45, W04 403, doi:10.1029/2008WR007308, w04403, 2009.

Gerrits, A. M. J., Pfister, L., and Savenije, H. H. G.: Spatial and temporal variability of canopy and forest floor interception in a beech forest, Hydrological

Processes, 24, 3011–3025, doi:10.1002/hyp.7712, 2010.
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., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, Hydrology and Earth System Sciences, 17, 149–161, doi:10.5194/hess-17-149-2013, 2013.
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.
Ghilain, N., Arboleda, A., and Gellens-Meulenberghs, F.: Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data, Hydrology and Earth System Sciences, 15, 771–786, doi:10.5194/hess-15-771-2011, 2011.
Greenstone, R. and King, M.: EOS Science Plan: Executive Summary, Tech. rep., NASA GSFC, Greenbelt, MD, USA, doi:http://www.hubbardbrook.org/data/dataset.php?id=2, 1999.
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.
Gumbel, E. J.: The Return Period of Flood Flows, The Annals of Mathematical Statistics, 12, 163–190, 1941.
Gupta, H., 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., Clark, M., Vrugt, J., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resources Research, 48, W08 301, doi:10.1029/2011WR011044, 2012.
Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, doi:http://dx.doi.org/10.1016/j.jhydrol.2009.08.003, 2009.
Gupta, V. K. and Sorooshian, S.: Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined, Water Resources Research, 19, 269–276, doi:10.1029/WR019i001p00269, URL http://dx.doi.org/10.1029/WR019i001p00269, 1983.
Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, Remote Sensing of Environment, 83, 181 – 194, doi:http://dx.doi.org/10.1016/S0034-4257(02)00095-0, the Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, 2002.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Hydrological Processes, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.

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

Harr, R. D., Harper, W., Krygier, J., and Hsieh, F.: Changes in storm hydrographs after road building and clear-cutting in the Oregon Coast Range, Water Resources Research, 11, 436–444, 1975.

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, D20 119, 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.

Hornbeck, J. W.: Storm flow from hardwood-forested and cleared watersheds in New Hampshire, Water Resources Research, 9, 346–354, doi:10.1029/WR009i002p00346, 1973.

Hornbeck, J. W., Pierce, R. S., and Federer, C. A.: Streamflow Changes after Forest Clearing in New England, Water Resources Research, 6, 1124–1132, doi:10.1029/WR006i004p01124, 1970.

Hornbeck, J. W., Martin, C. W., and Eagar, C.: Summary of water yield experiments at Hubbard Brook Experimental Forest, New Hampshire, Canadian Journal of Forest Research, 27, 2043–2052, doi:10.1139/x97-173, 1997.

Hornbeck, J. W., Eagar, C., Bailey, A., and Campbell, J. L.: Comparisons with results from the Hubbard Brook Experimental Forest in the Northern Appalachians, Long-Term Response of a Forest Watershed Ecosystem: Clearcutting in the Southern Appalachians, p. 213, 2014.

Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, Hydrology and Earth System Sciences, 21, 3953–3973, doi:10.5194/hess-21-3953-2017, 2017.

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

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological Sciences Journal, 58, 1198–

1255, doi:10.1080/02626667.2013.803183, 2013.
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.
Hu, G., Jia, L., and Menenti, M.: Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011, Remote Sensing of Environment, 156, 510 – 526, doi:http://dx.doi.org/10.1016/j.rse.2014.10.017, 2015.
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, J. W. and Fahey, T. J.: Colonization Dynamics of Herbs and Shrubs in a Disturbed Northern Hardwood Forest, Journal of Ecology, 79, 605–616, doi:10.2307/2260656, 1991.
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.
Hundecha, Y., Arheimer, B., Donnelly, C., and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, Journal of Hydrology: Regional Studies, 6, 90 – 111, doi:http://dx.doi.org/10.1016/j.ejrh.2016.04.002, 2016.
Immerzeel, W. and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, Journal of Hydrology, 349, 411 – 424, doi:http://dx.doi.org/10.1016/j.jhydrol.2007.11.017, 2008.
Ivanov, V., Bras, R., and Vivoni, E.: Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks, Water Resources Research, 44, W03429, doi:10.1029/2006WR005588, 2008.
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 plot-scale understanding to the catchment scale, Water Resources Research, 45, W04428, doi:10.1029/2008WR007225, 2009.
Jobbágy, E. G. and Jackson, R. B.: Groundwater use and salinization with grassland afforestation, Global Change Biology, 10, 1299–1312, doi:10.1111/j.1365-2486.2004.00806.x, 2004.
Johnson, C. E., Johnson, A. H., Huntington, T. G., and Siccama, T. G.: Whole-Tree Clear-Cutting Effects on Soil Horizons and Organic-Matter Pools, Soil Science Society of America Journal, 55, 497–502, doi:10.2136/sssaj1991.03615995005500020034x, 1991.
Johnson, S. and Rothacher, J.: Stream discharge in gaged watersheds at the Andrews Experimental Forest, 1949 to present. Long-Term

Ecological Research.Forest Science Data Bank, Forest Science Data Bank, doi:http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=HF004, 2016.

Jones, J. and Post, D.: Seasonal and successional streamflow response to forest cutting and regrowth in the northwest and eastern United States, Water Resources Research, 40, W05 203, doi:10.1029/2003WR002952, 2004.

Jones, J. A. and Grant, G. E.: Peak Flow Responses to Clear-Cutting and Roads in Small and Large Basins, Western Cascades, Oregon, Water Resources Research, 32, 959–974, doi:10.1029/95WR03493, 1996.

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.

Kelleher, C., McGlynn, B., and Wagener, T.: Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding, Hydrology and Earth System Sciences, 21, 3325–3352, doi:10.5194/hess-21-3325-2017, 2017.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Transactions on Geoscience and Remote Sensing, 50, 1384–1403, doi:10.1109/TGRS.2012.2184548, 2012.

Kirchner, J.: 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.

Kleidon, A.: Global Datasets of Rooting Zone Depth Inferred from Inverse Methods, Journal of Climate, 17, 2714–2722, doi:10.1175/1520-0442(2004)017<2714:GDORZD>2.0.CO;2, 2004.

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

Koren, V., Smith, M., and Duan, Q.: Use of a Priori Parameter Estimates in the Derivation of Spatially Consistent Parameter Sets of Rainfall-Runoff Models, Water Sci. Appl.„ 6, 239 – 254, doi:10.1002/9781118665671.ch18, 2003.

Koren, V., Moreda, F., and Smith, M.: Use of soil moisture observations to improve parameter consistency in watershed calibration, Physics and Chemistry of the Earth, Parts A/B/C, 33, 1068 – 1080, doi:https://doi.org/10.1016/j.pce.2008.01.003, from Calibration to Process Understanding in Rainfall-Runoff Modelling, 2008.

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.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, GB1015, doi:10.1029/2003GB002199, gB1015, 2005.

Kuczera, G.: Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest, Journal of Hydrology, 94, 215–236, doi:http://dx.doi.org/10.1016/0022-1694(87)90054-0, 1987.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21, 2863–2879, doi:10.5194/hess-21-2863-2017, 2017.

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.

Kunnath-Poovakka, A., Ryu, D., Renzullo, L., and George, B.: The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction, Journal of Hydrology, 535, 509 – 524, doi:http://dx.doi.org/10.1016/j.jhydrol.2016.02.018, 2016.

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.

Kuzmin, V., Seo, D.-J., and Koren, V.: Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, Journal of Hydrology, 353, 109 – 128, doi:https://doi.org/10.1016/j.jhydrol.2008.02.001, 2008.

Laio, F., Porporato, A., Ridolfi, L., and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: II. Probabilistic soil moisture dynamics, Advances in Water Resources, 24, 707–723, doi:http://dx.doi.org/10.1016/S0309-1708(01)00005-7, 2001.

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, W06 428, doi: 10.1029/2006WR005608, w06428, 2007.

Le Pape, P., Ayrault, S., and Quantin, C.: Trace element behavior 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.

Leavesley, G. and Stannard, L.: Application of remotely sensed data in a distributed-parameter watershed model, in: Proceedings of Workshop on Application of Remote Sensing in Hydrology, edited by Kite G.W., Wankiewicz, A., pp. 47–68, NHRC, Saskatoon, 1990.

Legesse, D., Vallet-Coulomb, C., and Gasse, F.: Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study South Central Ethiopia, Journal of Hydrology, 275, 67–85, doi:http://dx.doi.org/10.1016/S0022-1694(03)00019-2, 2003.

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

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, Water Resources Research, 51, 7309–7342, doi:10.1002/2015WR017616, 2015.

Li, H.-Y., Sivapalan, M., Tian, F., and Harman, C.: Functional approach to exploring climatic and landscape controls of runoff generation: 1. Behavioral constraints on runoff volume, Water Resources Research, 50, 9300–9322, doi: 10.1002/2014WR016307, 2014.

Li, K. Y., Coe, M. T., Ramankutty, N., and Jong, R. D.: Modeling the hydrological impact of land-use change in West Africa, Journal of Hydrology, 337, 258–268, doi:http://dx.doi.org/10.1016/j.jhydrol.2007.01.038, 2007.

Liancourt, P., Sharkhuu, A., Ariuntsetseg, L., Boldgiv, B., Helliker, B., Plante, A., Petraitis, P., and Casper, B.: Temporal and spatial variation in how vegetation alters the soil moisture response to climate manipulation, Plant and Soil, 351, 249–261, doi:10.1007/s11104-011-0956-y, 2012.

Liang, X., Lettenmaier, D., Wood, E., and Burges, S.: 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.

Likens, G.: Biogeochemistry of a forested ecosystem, Springer, New York, 2013.

Likens, G., Bormann, F., Johnson, N., Fisher, D., and Pierce, R.: Effects of Forest Cutting and Herbicide Treatment on Nutrient Budgets in the Hubbard Brook Watershed-Ecosystem, Ecological Monographs, 40, 23–47, doi: 10.2307/1942440, 1970.

Lindström, 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, Hydrology Research, 41, 295–319,

doi:10.2166/nh.2010.007, 2010.
Linsley, R. K.: Rainfall-runoff models: An overview, in: Proceedings of the International Symposium on Rainfall-Runoff Modelling, edited by Singh, V., pp. 3–22, Water Resour. Publ.„ Littleton, Colorado, USA, 1982.
Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43, W07 401, doi: 10.1029/2006WR005756, w07401, 2007.
Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrology and Earth System Sciences, 16, 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.
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.
Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F., and Syed, T. H.: Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, Water Resources Research, 46, W05 517, doi:10.1029/2009WR007855, w05517, 2010.
Lopez Lopez, P., Sutanudjaja, E., Schellekens, J., Sterk, G., and Bierkens, M.: Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products, Hydrology and Earth System Sciences Discussions, 2017, 1–39, doi:10.5194/hess-2017-16, 2017.
Mahe, G., Paturel, J.-E., Servat, E., Conway, D., and Dezetter, A.: The impact of land use change on soil water holding capacity and river flow modelling in the Nakambe River, Burkina-Faso, Journal of Hydrology, 300, 33–43, doi: http://dx.doi.org/10.1016/j.jhydrol.2004.04.028, 2005.
Martin, C.: Soil disturbance by logging in New England–review and management recommendations, Northern Journal of Applied Forestry, 5, 30–34, 1988.
Martin, C. W., Hornbeck, J. W., Likens, G. E., and Buso, D. C.: Impacts of intensive harvesting on hydrology and nutrient dynamics of northern hardwood forests, Canadian Journal of Fisheries and Aquatic Sciences, 57, 19–29, doi: 10.1139/f00-106, 2000.
McDowell, N., Pockman, W., Allen, C., Breshears, D., Cobb, N., Kolb, T., Plaut, J., Sperry, J., West, A., Williams, D., and Yepez, E.: Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought?, New Phytologist, 178, 719–739, doi: 10.1111/j.1469-8137.2008.02436.x, 2008.
McGlynn, B. L. and McDonnell, J. J.: Role of discrete landscape units in controlling catchment dissolved organic carbon dynamics, Water Resources Research, 39, 1090, doi:10.1029/2002WR001525, URL http://dx.doi.org/10.1029/2002WR001525, 1090, 2003.
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, W07 504, doi:10.1029/2003WR002494, 2004.

McMillan, H., Booker, D., and Cattoën, C.: Validation of a national hydrological model, Journal of Hydrology, 541, 800 – 815, doi:http://dx.doi.org/10.1016/j.jhydrol.2016.07.043, 2016.

Mencuccini, M.: The ecological significance of long-distance water transport: short-term regulation, long-term acclimation and the hydraulic costs of stature across plant life forms, Plant, Cell & Environment, 26, 163–182, doi:10.1046/j.1365-3040.2003.00991.x, 2003.

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.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology, 287, 95 – 123, doi:http://dx.doi.org/10.1016/j.jhydrol.2003.09.028, 2004.

Milly, P. C. D.: Climate, soil water storage, and the average annual water balance, Water Resources Research, 30, 2143–2156, doi:10.1029/94WR00586, 1994.

Milly, P. C. D. and Dunne, K. A.: Sensitivity of the Global Water Cycle to the Water-Holding Capacity of Land, Journal of Climate, 7, 506–526, doi:doi:10.1175/1520-0442(1994)007<0506:SOTGWC>2.0.CO;2, 1994.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resources Research, 43, W05 434, doi:10.1029/2006WR005184, 2007.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrological Sciences Journal, 58, 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

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

Mou, P., Fahey, T. J., and Hughes, J. W.: Effects of Soil Disturbance on Vegetation Recovery and Nutrient Accumulation Following Whole-Tree Harvest of a Northern Hardwood Ecosystem, Journal of Applied Ecology, 30, 661–675, doi:10.2307/2404245, 1993.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781 – 1800, doi:http://dx.doi.org/10.1016/j.rse.2011.02.019, 2011.

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.

Mulder, G., Olsthoorn, T. N., Al-Manmi, D. A. M. A., Schrama, E. J. O., and Smidt, E. H.: Identifying water mass depletion in northern Iraq observed by GRACE, Hydrology and Earth System Sciences, 19, 1487–1500, doi: 10.5194/hess-19-1487-2015, 2015.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi: http://dx.doi.org/10.1016/0022-1694(70)90255-6, 1970.

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.

Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, T., McGuire, K., Savenije, H., and Hrachowitz, M.: The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change?, Hydrology and Earth System Sciences, 20, 4775–4799, doi:10.5194/hess-20-4775-2016, 2016a.

Nijzink, R., Almeida, S., Pechlivanidis, I., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Savenije, H., and Hrachowitz, M.: Constraining conceptual hydrological models with multiple information sources, Water Resources Research, p. submitted, 2017.

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, Hydrology and Earth System Sciences, 20, 1151–1176, doi:10.5194/hess-20-1151-2016, 2016b.

Nobel, P. and Cui, M.: Hydraulic Conductances of the Soil, the Root-Soil Air Gap, and the Root: Changes for Desert Succulents in Drying Soil, Journal of Experimental Botany, 43, 319–326, doi:10.1093/jxb/43.3.319, 1992.

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.

North, G. and Nobel, P.: Drought-induced changes in hydraulic conductivity and structure in roots of Ferocactus acanthodes and Opuntia ficus-indica, New Phytologist, 120, 9–19, doi:10.1111/j.1469-8137.1992.tb01053.x, 1992.

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.
Onstad, C. A. and Jamieson, D. G.: Modeling the Effect of Land Use Modifications on Runoff, Water Resources Research, 6, 1287–1295, doi:10.1029/WR006i005p01287, 1970.
Orth, R., Staudinger, M., Seneviratne, S., 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.
Oudin, L., Andréassian, V., Perrin, C., and Anctil, F.: Locating the sources of low-pass behavior within rainfall-runoff models, Water Resources Research, 40, W11 101, doi:10.1029/2004WR003291, 2004.
Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface moisture, Journal of Geophysical Research: Earth Surface, 113, n/a–n/a, doi:10.1029/2007JF000769, f01002, 2008.
Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, Journal of Hydrology, 358, 240 – 258, doi:http://dx.doi.org/10.1016/j.jhydrol.2008.06.006, 2008.
Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrological Processes, 21, 435–446, doi:10.1002/hyp.6253, 2007.
Parajka, J., Naeimi, V., Blöschl, G., and Komma, J.: Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria, Hydrology and Earth System Sciences, 13, 259–271, doi:10.5194/hess-13-259-2009, 2009.
Patric, J. H. and Reinhart, K. G.: Hydrologic Effects of Deforesting Two Mountain Watersheds in West Virginia, Water Resources Research, 7, 1182–1188, doi:10.1029/WR007i005p01182, 1971.
Pechlivanidis, I. G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case, Hydrology and Earth System Sciences, 19, 4559–4579, doi:10.5194/hess-19-4559-2015, 2015.
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, W12 419, doi:10.1029/2007WR006615, 2008.
Porporato, A., Daly, E., and Rodríguez-Iturbe, I.: Soil Water Balance and Ecosystem Response to Climate Change, The American Naturalist, 164, 625–632, doi:10.1086/424970, 2004.
Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.:

A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, Journal of Hydrology, 411, 66–76, doi: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, 2015.

Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resources Research, 52, 7779–7792, doi:10.1002/2016WR019430, 2016.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Dong-Jun Seo, D.-J., and DMIP Participants: 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. and Storm, B.: MIKE SHE, in: Computer models of watershed hydrology, edited by Singh, V. P., vol. 32, pp. 2189–2202, Water Resources Publications, Littleton, doi:10.1029/96WR00896, 1995.

Refsgaard, J. C. and Knudsen, J.: Operational Validation and Intercomparison of Different Types of Hydrological Models, Water Resources Research, 32, 2189–2202, doi:10.1029/96WR00896, 1996.

Reggiani, P. and Rientjes, T. H. M.: Flux parameterization in the representative elementary watershed approach: Application to a natural basin, Water Resources Research, 41, W04 013, doi:10.1029/2004WR003693, 2005.

Reggiani, P., Sivapalan, M., and Hassanizadeh, S.: 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.

Reicher, 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.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Advances in Water Resources, 31, 1411 – 1418, doi:http://dx.doi.org/10.1016/j.advwatres.2008.01.001, hydrologic Remote Sensing, 2008.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, W05 521, doi:10.1029/2009WR008328, w05521, 2010.

Rennó, C., Nobre, A., Cuartas, L., Soares, J., Hodnett, M., Tomasella, J., and Waterloo, M.: 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.

Reynolds, J. F., Kemp, P. R., and Tenhunen, J. D.: Effects of long-term rainfall

variability on evapotranspiration and soil water distribution in the Chihuahuan Desert: A modeling analysis, Plant Ecology, 150, 145–159, doi:10.1023/a:1026530522612, 2000.

Rodríguez-Iturbe, I.: Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies, Water Resources Research, 36, 3–9, doi:10.1029/1999WR900210, 2000.

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.

Rodríguez-Iturbe, I., D'Odorico, P., Laio, F., Ridolfi, L., and Tamea, S.: Challenges in humid land ecohydrology: Interactions of water table and unsaturated zone with climate, soil, and vegetation, Water Resources Research, 43, W09 301, doi:10.1029/2007WR006073, 2007.

Rood, S. B., Braatne, J. H., and Hughes, F. M. R.: Ecophysiology of riparian cottonwoods: stream flow dependency, water relations and restoration, Tree Physiology, 23, 1113–1124, doi:10.1093/treephys/23.16.1113, 2003.

Rothacher, J.: Streamflow from small watersheds on the western slope of the Cascade Range of Oregon, Water Resources Research, 1, 125–134, doi:10.1029/WR001i001p00125, 1965.

Rothacher, J.: Increases in Water Yield Following Clear-Cut Logging in the Pacific Northwest, Water Resources Research, 6, 653–658, doi:10.1029/WR006i002p00653, 1970.

Rothacher, J., Dyrness, C., Fredriksen, R., Forest, P. N., and Station, R. E.: Hydrologic and related characteristics of three small watersheds in the Oregon Cascades, Pacific Northwest Forest and Range Experiment Station, U.S. Dept. of Agriculture, (Portland, Or.), 1967.

Samaniego, L., Bárdossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, Water Resources Research, 46, W02 506, doi:10.1029/2008WR007695, 2010a.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, W05 523, doi:10.1029/2008WR007327, 2010b.

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 "The art of hydrology"*, Hydrology and Earth System Sciences, 13, 157–161, doi:10.5194/hess-13-157-2009, 2009.

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.

Savenije, H. H. G. and Hrachowitz, M.: HESS Opinions "Catchments as meta-organisms – a new blueprint for hydrological modelling", Hydrology and Earth System Sciences, 21, 1107–1116, doi:10.5194/hess-21-1107-2017, 2017.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrology and Earth System Sciences, 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.

Schenk, H. J. and Jackson, R.: Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems, Journal of Ecology, 90, 480–494, doi:10.1046/j.1365-2745.2002.00682.x, 2002.

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.

Schoups, G., Hopmans, J., Young, C., Vrugt, J., and Wallender, W.: Multi-criteria optimization of a regional spatially-distributed subsurface water flow model, Journal of Hydrology, 311, 20–48, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.01.001, 2005.

Schulz, K., Jarvis, A., Beven, K., and Soegaard, H.: The Predictive Uncertainty of Land Surface Fluxes in Response to Increasing Ambient Carbon Dioxide, Journal of Climate, 14, 2551–2562, doi:10.1175/1520-0442(2001)014<2551:TPUOLS>2.0.CO;2, 2001.

Schymanski, S. J., Sivapalan, M., Roderick, M. L., Beringer, J., and Hutley, L. B.: An optimality-based model of the coupled soil moisture and root dynamics, Hydrology and Earth System Sciences, 12, 913–932, doi:10.5194/hess-12-913-2008, 2008.

Seibert, J.: Estimation of Parameter Uncertainty in the HBV Model, Hydrology Research, 28, 247–262, 1997.

Seibert, J. and McDonnell, 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. and McDonnell, J.: Land-cover impacts on streamflow: a change-detection modelling approach that incorporates parameter uncertainty, Hydrological Sciences Journal, 55, 316–332, doi:10.1080/02626661003683264, 2010.

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

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, 1014, 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.
Seibert, J., McDonnell, J., and Woodsmith, R.: Effects of wildfire on catchment runoff response: a modelling approach to detect changes in snow-dominated forested catchments, Hydrology Research, 41, 378–390, 2010.
Seneviratne, S., Corti, T., Davin, E., Hirschi, M., Jaeger, E., Lehner, I., Orlowsky, B., and Teuling, A.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125–161, doi: http://dx.doi.org/10.1016/j.earscirev.2010.02.004, 2010.
Seneviratne, S., Wilhelm, M., Stanelle, T., van den Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M., Meier, A., Brovkin, V., Claussen, M., Ducharne, A., Dufresne, J.-L., Findell, K., Ghattas, J., Lawrence, D., Malyshev, S., Rummukainen, M., and Smith, B.: Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment, Geophysical Research Letters, 40, 5212–5217, doi:10.1002/grl.50956, 2013.
Shafii, M. and Tolson, B. A.: Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, Water Resources Research, 51, 3796–3814, doi:10.1002/2014WR016520, 2015.
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.
Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Change Biology, 9, 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.
Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, Hydrological Processes, 17, 3163–3170, doi:10.1002/hyp.5155, 2003.
Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in: Encyclopedia of Hydrological Sciences, edited by Anderson, M. G., John Wiley & Sons, Ltd, doi:10.1002/0470848944.hsa012, 2006.
Smakhtin, V.: Low flow hydrology: a review, Journal of Hydrology, 240, 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001.
Smith, T., Hayes, K., Marshall, L., McGlynn, B., and Jencso, K.: Diagnostic calibration and cross-catchment transferability of a simple process-consistent hydrologic model, Hydrological Processes, 30, 5027–5038, doi:10.1002/hyp.10955, hYP-15-1062.R2, 2016.

Spaaks, J. H. and Bouten, W.: Resolving structural errors in a spatially distributed hydrologic model using ensemble Kalman filter state updates, Hydrology and Earth System Sciences, 17, 3455–3472, doi:10.5194/hess-17-3455-2013, 2013.

Spear, R. C., Grieb, T. M., and Shang, N.: Parameter uncertainty and interaction in complex environmental models, Water Resources Research, 30, 3159–3169, doi:10.1029/94WR01732, 1994.

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.

Sperry, J. S., Hacke, U. G., Oren, R., and Comstock, J. P.: Water deficits and hydraulic limits to leaf water supply, Plant, Cell & Environment, 25, 251–263, doi:10.1046/j.0016-8025.2001.00799.x, 2002.

Sriwongsitanon, N., Gao, H., Savenije, H. H. G., Maekan, E., Saengsawang, S., and Thianpopirug, S.: Comparing the Normalized Difference Infrared Index (NDII) with root zone storage in a lumped conceptual model, Hydrology and Earth System Sciences, 20, 3361–3377, doi:10.5194/hess-20-3361-2016, 2016.

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.

Sun, W., Ishidaira, H., Bastola, S., and Yu, J.: Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications, Environmental Research, 139, 36 – 45, doi:http://dx.doi.org/10.1016/j.envres.2015.01.002, environmental Research on Hydrology and Water Resources, 2015.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, Water Resources Research, 50, 687–705, doi:10.1002/2013WR013807, 2014.

Swift, L. W. and Swank, W. T.: Long term responses of streamflow following clearcutting and regrowth / Réactions à long terme du débit des cours d'eau après coupe et repeuplement, Hydrological Sciences Bulletin, 26, 245–256, doi:10.1080/02626668109490884, 1981.

Tang, J. and Zhuang, Q.: Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics, Journal of Geophysical Research: Biogeosciences, 113, G04 010, doi:10.1029/2008JG000757, g04010, 2008.

Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., and Weerts, A. H.: Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin, Hydrology and Earth System Sciences, 19, 2079–2100, doi:10.5194/hess-19-2079-2015, 2015.

Tapley, B. D., Bettadpur, S., Watkins, M., and Reigber, C.: The gravity recovery and

climate experiment: Mission overview and early results, Geophysical Research Letters, 31, L09 607, doi:10.1029/2004GL019920, l09607, 2004.

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.

Teuling, A., Van Loon, A., Seneviratne, S., Lehner, I., Aubinet, M., Heinesch, B., Bernhofer, C., Grünwald, T., Prasse, H., and Spank, U.: Evapotranspiration amplifies European summer drought, Geophysical Research Letters, 40, 2071–2075, doi:10.1002/grl.50495, 2013.

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, 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, Hydrology and Earth System Sciences, 10, 619–644, doi:10.5194/hess-10-619-2006, 2006.

Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I. J. M., Walker, J. P., Pauwels, V. R. N., and Allgeyer, S.: Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals, Water Resources Research, 53, 1820–1840, doi: 10.1002/2016WR019641, 2017.

Tolson, B. and Shoemaker, C.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resources Research, 43, W01 413, doi:10.1029/2005WR004723, w01413, 2007.

Tonkin, M. J. and Doherty, J.: A hybrid regularized inversion methodology for highly parameterized environmental models, Water Resources Research, 41, W10 412, doi:10.1029/2005WR003995, 2005.

Tourian, M., Schwatke, C., and Sneeuw, N.: River discharge estimation at daily resolution from satellite altimetry over an entire river basin, Journal of Hydrology, 546, 230 – 247, doi:https://doi.org/10.1016/j.jhydrol.2017.01.009, 2017.

Troch, P., Martinez, G., Pauwels, V., Durcik, M., Sivapalan, M., Harman, C., Brooks, P., Gupta, H., and Huxman, T.: Climate and vegetation water use efficiency at catchment scales, Hydrological Processes, 23, 2409–2414, doi:10.1002/hyp.7358, 2009a.

Troch, P., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J., Roy, T., and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resources Research, 51, 4903–4922, doi:10.1002/2015WR017032, 2015.

Troch, P. A., Carrillo, G. A., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T. H. M., and Yaeger, M.: Dealing with Landscape Heterogeneity in Watershed Hydrology: A Review of Recent Progress toward New Hydrological Theory, Geography Compass, 3, 375–392, doi:10.1111/j.1749-8198.

2008.00186.x, 2009b.
Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resources Research, 42, W02 411, doi:10.1029/2004WR003800, 2006.
Tron, S., Perona, P., Gorla, L., Schwarz, M., Laio, F., and Ridolfi, L.: The signature of randomness in riparian plant root distributions, Geophysical Research Letters, 42, 7098–7106, doi:10.1002/2015GL064857, 2015.
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, catchment modelling: Towards an improved representation of the hydrological processes in real-world model applications, 2004.
van den Hurk, B., Viterbo, P., Beljaars, A., and Betts, A.: Offline validation of the ERA40 surface scheme, Shinfield Park, Reading, 2000.
van Emmerik, T., Mulder, G., Eilander, D., Piet, M., and Savenije, H.: Predicting the ungauged basin: Model validation and realism assessment, Front. Earth Sci, 3, 62, 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., and Bierkens, M.: 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.
Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resources Research, 44, n/a–n/a, doi:10.1029/2008WR006829, w00D06, 2008.
Verstraeten, W. W., Veroustraete, F., and Feyen, J.: Assessment of Evapotranspiration and Soil Moisture Content Across Different Scales of Observation, Sensors, 8, 70–117, doi:10.3390/s8010070, 2008.
Vertessy, R. A., Hatton, T. J., Benyon, R. G., and Dawes, W. R.: Long-term growth and water balance predictions for a mountain ash (Eucalyptus regnans) forest catchment subject to clear-felling and regeneration, Tree Physiology, 16, 221, doi:10.1093/treephys/16.1-2.221, 1996.
Vose, J. M., Miniat, C. F., Luce, C. H., Asbjornsen, H., Caldwell, P. V., Campbell, J. L.,

Grant, G. E., Isaak, D. J., Loheide Ii, S. P., and Sun, G.: Ecohydrological implications of drought for forests in the United States, Forest Ecology and Management, doi:http://dx.doi.org/10.1016/j.foreco.2016.03.025, 2016.

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.

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resources Research, 47, n/a–n/a, doi:10.1029/2010WR009469, w06301, 2011.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, Journal of Hydrology, 320, 132 – 154, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.07.015, the model parameter estimation experiment, 2006.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, Hydrology and Earth System Sciences, 5, 13–26, doi:10.5194/hess-5-13-2001, 2001.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, Geography Compass, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sensing of Environment, 70, 191 – 207, doi:http://dx.doi.org/10.1016/S0034-4257(99)00036-X, 1999.

Waichler, S., Wemple, B., and Wigmosta, M.: Simulation of water balance and forest treatment effects at the H.J. Andrews Experimental Forest, Hydrological Processes, 19, 3177–3199, doi:10.1002/hyp.5841, 2005.

Wanders, N., Bierkens, M. F. P., de Jong, S. M., de Roo, A., and Karssenberg, D.: The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, Water Resources Research, 50, 6874–6891, doi:10.1002/2013WR014639, 2014.

Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J., and Savenije, H. H. G.: Contrasting roles of interception and transpiration in the hydrological cycle – Part 1: Temporal characteristics over land, Earth System Dynamics, 5, 441–469, doi:10.5194/esd-5-441-2014, 2014.

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., J"agermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H. G.: Global root zone storage capacity from satellite-based evaporation, Hydrol. Earth Syst. Sci., 20, 1459–1481, doi:10.5194/hess-20-1459-2016, hESS, 2016.

Weibull, W.: A Statistical Distribution Function of Wide Applicability, Journal of applied mechanics, 18, 293–297, 1951.

Werth, S., Güntner, A., Petrovic, S., and Schmidt, R.: Integration of {GRACE} mass variations into a global hydrological model, Earth and Planetary Science Letters, 277, 166 – 173, doi:https://doi.org/10.1016/j.epsl.2008.10.021, 2009.

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C. Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205–2227, doi: 10.5194/hess-15-2205-2011, hESS, 2011.

Westra, S., Thyer, M., Leonard, M., Kavetski, D., and Lambert, M.: A strategy for diagnosing and interpreting hydrological model nonstationarity, Water Resources Research, 50, 5090–5113, doi:10.1002/2013WR014719, 2014.

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

Wilks, D.: 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.

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resources Research, 45, W12 422, doi:10.1029/2009WR007706, w12422, 2009.

Winter, T.: 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.

Xu, X., Li, J., and Tolson, B. A.: Progress in integrating remote sensing data and hydrologic modeling, Progress in Physical Geography, 38, 464–498, doi:10.1177/0309133314536583, 2014.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756–1774, doi:http://dx.doi.org/10.1016/j.advwatres.2007.01.005, 2007.

Yilmaz, K., Gupta, H., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, W09 417, doi:10.1029/2007WR006716, 2008.

Young, P.: Parallel Processes in Hydrology and Water 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 thermody-

namic 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. 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, Hydrology and Earth System Sciences, 9, 243–261, doi:10.5194/hess-9-243-2005, 2005.

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, Hydrology and Earth System Sciences, 10, 937–955, doi:10.5194/hess-10-937-2006, 2006.

Zhang, S., Yang, H., Yang, D., and Jayawardena, A. W.: Quantifying the effect of vegetation change on the regional water balance within the Budyko framework, Geophysical Research Letters, 43, 1140–1148, doi:10.1002/2015GL066952, 2016.

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

Acknowledgements

This thesis was only possible with the help, advice and support of many people, but, most of all, my promotor and copromotor Huub Savenije and Markus Hrachowitz. They supported me all the time and their office was always open for questions. It's been an interesting and challenging journey, which led to some nice scientific work as well as personal growth.

My colleagues helped a lot in just having a good time at work, preventing me from getting crazy behind my computer. With a good time, I mean the daily coffees with the 'wet-mop-club' under supervision of Wim, but also Petra, Miriam, Martin, Lydia, Maurits. And the apple piece of Betty always helped for some extra energy after lunch. The battles with Boran and Anna in the pipe-fitting-competition (next year we win...) was a great challenge. Fieldwork with Thom, Saket, Bas, Bart, Juan, was also a highlight which I always looked forward to, especially with the great cooks that joined over the years and the terraces without paying with Olivier. My office mates Yingrong and Cesar were always there for a coffee and a chat. Discussions about modelling, hydrology, science and other PhD-stuff during summer schools, conferences, lunches and coffeebreaks were always interesting with my PhD-colleagues Tanja, Hongkai, Lan, Petra, Laurène, but also Yang, Paul, Camille, Zhilin and all others. Playing with Arduinos with Rolf was very nice too, just as the jokes of Erik during lunch. The soccer tournaments with the Dreamteam Water were very challenging, I hope you keep up the increasing 'successes' with this team for the next coming years. The Switch-On 'family' with Lorna, Ilias, Gerben, René, Alberto, Juraj, Thorsten, and of course Berit and many more, made my PhD possible in the first place, and we always had a good time at the general assemblies. Besides all the people mentioned here, all other colleagues, students and friends helped me a lot during these four years.

Additionally, I'd like to thank of course my friends Wesley, Sander, Robin, with whom I made some very nice and very necessary holiday trips, Marc, Lars and Ashwin to just have drinks and watch soccer matches with, and of course many others just to have a good time and get my mind off work. And last, but not least, all my family, my parents and sister.

Curriculum Vitæ

Remko Nijzink

07-06-1988 Born in Leidschendam, The Netherlands.

Education

2001–2007 VWO
Alfrink College, Zoetermeer

2007–2010 BSc Civil Engineering
Delft University of Technology

2010–2013 MSc Civil Engineering
Delft University of Technology

2013 PhD. Doctoral Eduction
Delft University of Technology
Thesis: Hydrological models
Promotor: Prof. dr. ir. H.H.G. Savenije
Copromotor: dr. habil. M. Hrachowitz

List of Publications

First author

3. **Nijzink, R.C.**, Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, T., McGuire, K., Savenije, H., and Hrachowitz, M.: *The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change?*, Hydrol. Earth Syst. Sci., 20, 4775-4799, doi:10.5194/hess-20-4775-2016, 2016.

2. **Nijzink, R.C.**, Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M.: *The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models*, Hydrol. Earth Syst. Sci., 20, 1151-1176, doi:10.5194/hess-20-1151-2016, 2016.

1. **Nijzink, R.C.**, Almeida, S., Pechlivanidis, I., Capell, R., Gustaffsons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Savenije, H., and Hrachowitz, M.: *Constraining conceptual hydrological models with multiple information sources*, Water Resources Research, submitted, 2017.

Co-author

4. Gisen, J. I. A., Savenije, H. H. G., **Nijzink, R.C.**, and Abd. Wahab, A.K.: *Testing a 1-D Analytical Salt Intrusion Model and its Predictive Equations in Malaysian Estuaries*, Hydrological Sciences Journal, 60(1), 156-172, doi:10.1080/02626667.2014.889832, 2015.

3. Gisen, J. I. A., Savenije, H. H. G., and **Nijzink, R. C.**: *Revised predictive equations for salt intrusion modelling in estuaries*, Hydrol. Earth Syst. Sci., 19, 2791-2803, doi:10.5194/hess-19-2791-2015, 2015.

2. Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., **Nijzink, R.**, Parajka, J., Toth, E., Viglione, A., and Wagener, T.: *Virtual laboratories: new opportunities for collaborative water science*, Hydrol. Earth Syst. Sci., 19, 2101-2117, doi:10.5194/hess-19-2101-2015, 2015.

1. 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 Resour. Res., 50, doi:10.1002/2014WR015484, 2014