Parameter’s Controls of Distributed Catchment Models—How Much Information is in Conventional Catchment Descriptors?

Ralf Merz¹, Larisa Tarasova¹, and Stefano Basso¹

¹Department Catchment Hydrology, Helmholtz Centre for Environmental Research, Halle (Saale), Germany

Abstract One major challenge in large scale modeling is the estimation of spatially consistent distributed parameters, that are parameters with a clear functional relationship to climate and landscape characteristics. We present a newly developed Parameter Set Shuffling (PASS) approach, which is able to provide such regionally consistent parameter sets. The PASS method does not require any a priori assumption on the relationship between model parameters and catchment descriptors. Instead, it derives these relationships from observed patterns of calibrated parameters and available catchment descriptors. We tested the PASS approach to derive parameters of a conceptual hydrological model applied to 263 German catchments. The resulting median model efficiencies for training and test catchments are, respectively, 0.74 and 0.72, similar to those obtained by other modeling approaches, which use regional calibration. In this study, a combination of catchment descriptors that clearly controls model parameters is not found. In fact, we show that various regional functional relationships between catchment descriptors and model parameters result in similarly good model performances. Moreover, catchment descriptors used for parameter prediction can be replaced in the parameter prediction, without any decrease in model performance. Our results suggest that by using conventional catchment descriptors based on averages, only the amount of information that is also retained in the existing correlations among climatic and catchment indicators is exploited. Development of a new generation of hydrologically meaningful catchment and climate descriptors is required to further improve our capability of forecasting hydrological dynamics of interest by means of large scale models and regionalization approaches.

1. Introduction

In the recent years, conceptual hydrological models have been increasingly used to provide consistent, long-term predictions of hydrological variables such as streamflow or soil moisture over a large scale (Bierkens, 2015). A major challenge in this field is the choice of a reasonably consistent parameterization (Archfield et al., 2015; Bierkens et al., 2015; Clark et al., 2016; Gupta et al., 2014; Mizukami et al., 2017; Paniconi & Putti, 2015). In fact, applying these models in a distributed mode at a large scale may require the estimation of parameters for tens of thousands of model elements, for gauged as well as for ungauged locations.

In this case traditional model calibration approaches are not feasible and various avenues have been pursued for their parameterization, such as using parameters from a single catchment (Henriksen et al., 2003), regionalizing parameters from observed to ungauged locations (Zhang et al., 2014; Zhu & Lettenmaier, 2007) transferring parameter sets from predefined catchments (Hartmann et al., 2015; Perrin et al., 2008), using uncalibrated models (Holster & Alder, 2016), simultaneously calibrating multiple representative catchments (Fernandez et al., 2000), or combinations thereof. Arheimer et al. (2012), e.g., combined simultaneous calibration in multiple catchments and a step-wise approach applied to groups of parameters regulating specific processes to derive parameters for the HYPE-model at the global scale.

Variations of regional calibration approaches are another alternative. The objective of regional calibration is to find parameter sets for each model element, hydrological response unit or subcatchment, respectively, which have the same functional relationship to catchment descriptors. For the purposes of this paper, we called such parameter sets regionally consistent parameter sets. Hence, regionally consistent parameter sets only vary if the CDs, which are assumed to control the parameter sets, vary whereas they are constant if the respective CDs do not vary.
Several regional calibration approaches have been developed recently. Parajka et al. (2007) used an iterative regional calibration approach to simultaneously calibrate the Hydrologiska Byråns Vattenbalansavdelning (HBV) model parameters of 320 Austrian catchments. The method exploits the spatial correlations of the parameters to constrain their a priori distribution for each catchment. Hundecha and Bardossy (2004) developed a regional calibration approach where, instead of calibrating model parameters directly, a functional dependency of model parameters on catchment descriptors is calibrated for all catchments. This approach, of course, drastically reduced the number of variables to be calibrated. Several studies followed this approach to estimate regionally consistent parameters (see, e.g., Götzinger & Bárdossy, 2007; Pokhrel et al., 2008; Hundecha et al., 2008). Samaniego et al. (2010) introduced more complex relationships between model parameters and catchment descriptors, e.g., based on pedo-transfer functions, to better represent hydrological processes.

In these studies, catchment descriptors that are used to build the regional functional relationship to model parameters are selected a priori, based on some sort of hydrological reasoning. For example, based on a sensitivity analysis of manually calibrated parameters, Hundecha and Bardossy (2004) concluded that snowmelt parameters are linearly dependent on land use, while model parameters controlling soil moisture accounting and evaporation are dependent on soil type and land use. Parameters controlling runoff generation and routing are controlled by soil type, land use, and slope. Samaniego et al. (2010) introduced more complex relationships between model parameters and catchment descriptors, e.g., based on pedo-transfer functions, to better represent hydrological processes.

However, selecting which catchment descriptors should be used to build a regional functional relationship to model parameters is challenging. Ideally, the relationship between catchment descriptors and model parameters should be clearly visible in the data and hydrologically justified. However, clear patterns are often not visible in our data. This becomes evident by looking at the many studies on regionalization of model parameters for ungauged catchments, which are facing the same challenge of finding regional consistent parameter sets. These studies dealing with parameter regionalization often report that unambiguous relationships cannot be found (e.g., Merz & Blöschl, 2004; Parajka et al., 2013; Seibert, 1999; Troch et al., 2003). In fact, often catchment descriptors such as mean annual precipitation and topographical features are highly correlated and, hence, they have therefore equally predictive power for model parameter estimation in ungauged basins (Merz & Blöschl, 2004). Blöschl and Zehe (2005) argued that standard catchment descriptors might not represent the hydrological processes of interest very well. Available catchment descriptors are usually metrics as percentage of soil type, geological unit, land cover, or averaged values of soil properties, such as mean soil depth or averaged bulk density. Clearly, average values or percentage of model element covered by a given unit does not represent the spatial variability in soils, vegetation cover, morphology, or subsurface characteristics, nor do they account for the lateral subgrid connectivity.

The dominant catchment descriptors that control model parameters may also vary with scale. For example, many conceptual hydrological models represent runoff generation as a nonlinear function of soil moisture (e.g., Bergström, 1995), i.e., in a dry state most of the rainfall becomes soil moisture and do not contribute to runoff during the modeling step, while the amount of rainfall contributing to runoff nonlinearly increases with increasing soil moisture. The shape of this function (i.e., the degree of nonlinearity of runoff generation) is a model parameter, either a free model parameter (subject to calibration) or a fixed model parameter one (e.g., Perrin et al., 2003, used a parabola to continuously parameterize the nonlinear dependence of precipitation yield to catchment wetness). Looking at the plot scale, one may argue that this parameter is controlled by soil characteristics, such as soil texture (Samaniego et al., 2010). In larger model elements, such as subcatchments or model grids with a size of several square kilometers, runoff generation is certainly spatially variable, with probably more runoff production in the riparian zones (due to saturation excess) than on the tops of the hills. For example, Freyberg et al. (2014) found that in a small pre-Alpine catchment only a small fraction of the total catchment area (1−26%) generates streamflow during rainfall events. Similarly, McGlynn and McDonnell (2003) showed for a catchment in New Zealand that runoff generated in the riparian zones dominates runoff for smaller events, while hillslope runoff contribution increases for larger events. Hence, one may expect that the distribution of flat (riparian) zones, where saturation excess occurs, controls the nonlinearity parameter. However, the distribution of flat (riparian) zones may be better represented by the topographic wetness index (Sørensen et al., 2006) than by soil characteristics. Looking at the regional
scale, climate may be the dominant control on the variability of runoff generation. For example, the regional patterns of event runoff coefficients derived from observation data in Austria (Merz & Blöschl, 2009) and Germany (Tarasova, Basso, Poncelet, & Merz, 2018; Tarasova, Basso, Zink, & Merz, 2018) can best be explained by the regional patterns of climate characteristics.

Given the findings above, we argue that the traditionally available catchment descriptors for regional modeling studies might not well characterize climate and landscape features which control model parameters. Hence, defining a priori which are the dominant catchment descriptors that control regional patterns of model parameters is challenging. Moreover, we hypothesize that various climate and catchment features can be found, which equally control model parameters and thus result in equally good regional model performances. We also hypothesize that we can exclude various climate and catchment features from the parameter prediction process without considerable deterioration in model performance.

To test our hypothesis, we develop a parameterization strategy which deciphers the regional dependence of model parameters on catchment descriptors from calibrated data. Such a data-driven approach has not been adopted for regional parameterization before. The proposed approach, called Parameter Set Shuffling (PASS) method, is based on a combination of local calibration and regionalization using machine learning tools and does not require the a priori selection of dominant catchment descriptors for each parameter. Using this approach, we generate regionally consistent parameter sets of a conceptual hydrological model for 263 catchments in Germany.

Using this parameterization strategy, we evaluate if different combinations of catchment descriptors can equally well predict model parameters, and finally we analyze if CDs can be replaced by other descriptors in the parameter estimation approach and how this impact model efficiencies.

2. Methodology and Data

2.1. Deriving Regional Parameter Sets Through the Parameter Set Shuffling (PASS) Method

The objective of the parameter estimation approach taken in this study, and called herein PASS method, is to find a set of parameters which guarantees good model performance for runoff simulation in each catchment while having a (yet unknown) consistent relationship with catchment descriptors (CDs) which is valid for all model elements.

The PASS method is based on the assumption that the regional functional relationship between model parameters and CDs used in large scale distributed models can be inferred from the spatial patterns of lumped parameters calibrated for individual catchments.

This assumption is based on the results of Bergström and Graham (1998), who found no scale effect on model parameters of the HBV model in the Baltic Sea region. Similarly, Merz & Blöschl, 2009 showed in a study based on 269 Austrian catchments ranging from 10 to 130,000 km² that most parameters of the HBV model do not change with catchment scale and concluded that regional patterns of climate and landscape characteristics rather than scale are dominant parameter controls. Only some parameters (e.g., those controlling runoff recession) exhibited a trend with catchment area. Since the model used in this study has a structure similar to HBV, we expect the spatial pattern of lumped parameters for smaller catchments (from $10^1$ to $10^3$ km²) to be similar to the spatial pattern of parameters for model elements (which have an area of 64 km² in our study). This assumption is tested within the PASS method by comparing the model efficiency of predicted lumped and distributed model parameters.

The PASS method is inspired by traditional PUB studies (e.g., Parajka et al., 2013), where parameters are first calibrated for single catchments and then the structural dependency of model parameters and CDs is derived. In those studies, only the best calibrated parameter set in each catchment is usually employed for regionalization. However, it is well known that several parameter sets with equally good performance might exist for each catchment due to parameter equifinality (e.g., Beven & Binley, 1992). Using only the best parameter set in each catchment to derive regionalization rules may not be the best choice, because other parameter sets with slightly inferior local model efficiencies may show a better regional relationship to CDs. For this reason, the PASS method considers several parameter sets for each catchment to infer regionalization rules. The combinations of parameter sets are shuffled as long as the regional performance (e.g., median model performance of single catchments) does not further improve.
The PASS method consists of the following repeating steps (see flow chart in Figure 1).

1. Selection of locally calibrated parameter sets to be used for regionalization.

A good (i.e., one out of the best 5%) lumped parameter set derived from local calibration is randomly selected for each catchment and passed to the regionalization routine. Please note that always a complete parameter set is selected for each catchment to account for the internal parameter structure.

What a good parameter set is has been long debated in hydrological literature (e.g., Beven & Freer, 2001; Wagener et al., 2001, Schaefl & Gupta, 2007). In this study we defined all parameter sets with ME > 0.95 * MEmax (where MEmax is the best model efficiency found by local calibration) as good parameter sets. In other words, we assigned the best 5% of calibrated parameters sets in each catchment to a pool of good parameter sets. This is a subjective choice and other criteria could be applied (e.g., all parameter sets with ME higher than a certain threshold). At each iteration step of the PASS method, one parameter set for each catchment is randomly drawn from this pool.

2. Regionalize parameters by means of machine learning techniques.

In the second step the regional functional relationship between CDs and each model parameter is deciphered using data mining tools applied to the selected ensemble of parameter sets.

This study uses the random forest technique. However, any other machine learning method that requires no prior assumptions on the relationship between predictors and response variable can be used. Random forests (Cutler et al., 2007) are a variant of tree-based models (e.g., decision trees, classification trees, regression trees, or bagging decision trees) (Breiman, 2001). Random forest has been successfully applied in different fields of hydrology, e.g., to predict hydrological signatures (Addor et al., 2018; Snelder et al., 2009), to map the flow regime of river networks (Gonzalez-Ferreras & Barquin, 2017), or to model flood damage (Sieg et al., 2017). Singh et al. (2014) used classification and regression trees to relate the similarity in catchments descriptors and stream flow signatures and succeed in transferring hydrological model parameters across the entire United States.
The central idea of tree-based data mining is to recursively split the data space (in our case catchments and CDs) into subspaces according to the behavior of a response variable (i.e., model parameters). Hence, a tree is a set of nodes, where at each node the collections of catchments is divided into two branches according to if a selected CD is smaller or larger than a threshold. The splitting criteria (which CD and which threshold) are selected to minimize the variability of model parameters in each of the two subgroups. Each branch is then iteratively divided into two subbranches. A random forest is simply a collection of decision trees whose results are aggregated into one final result. Random forests apply a bootstrap sampling to compose the training subsamples of the single trees. Only about two thirds of the data sample is used to set up the single trees, while one third of the data is left out. In each bootstrap step data points that are not taken into account for training the classifiers are called Out-of-Bag observations. The Out-of-Bag observations are used internally to calculate quality measures of the resulting model and to estimate the importance of the considered variable. In this study, we used the random forest implementation available in the R statistical software (package “randomForest”). Each forest consists of 200 trees. The number of predictors sampled for splitting at each node is set to the number of available CDs. Each tree is split until a minimum terminal node size of 5.

The data mining approach results in a regional functional relationship between each parameter and CDs. This relationship is then used to estimate regionally consistent parameters, further on called regional parameters.

3. Improve identification of satisfactory regional functional relationships.

Step 3 is optional, but proved able to speed up the PASS method toward identifying satisfactory regional functional relationships.

Some combinations of the local parameter sets randomly selected for each catchment lead to regional functional relationships with low performance, as shown by the explained variance or the mean of squared residuals of predicted and locally calibrated parameters. To improve this performance, the calibrated lumped parameter set which best resembles the one predicted by the functional relationship obtained at the previous regionalization step is selected. Similarity in this study is quantified as the sum of differences between each regionally predicted and locally calibrated parameter. The parameters are normalized with a Min-Max-Normalization using the parameter bounds, which equally weights the contribution of each parameter to the similarity measure. Note that the weights may also be selected according to the sensitivity of each parameter to model performance. However, as step 3 is optional and parameter sensitivity analyses (supporting information Figure S1) show no large difference in the sensitivity of single parameters to model performance, all parameters are equally weighted.

After selecting the most similar parameter set for each catchment, regionalization is repeated. The steps are repeated as long as the variance explained by the decision forest is increasing.

The PASS method tests which combination of good local parameter sets gives the best regional prediction. Hence, the approach strongly depends on the availability for each catchment (i.e., at step 1) of a sufficient number of good lumped parameter sets. To increase their number, each parameter set predicted by the regional functional relationship which also has good local model performance (ME > 0.95 * ME_{max} where ME_{max} is the best model efficiency found by local calibration) is added to the pool of step 1 and further used in the next iterations of the PASS approach.

Please note that if only a small number (i.e., 1 to 3) of different local parameter sets is available at the beginning of the procedure, the approach may tend to identify similar regional functional relationships (and related regional parameter sets) at each iteration step. If these predicted regional parameter sets are retained in the pool of step 1, the approach will likely sample more and more similar local parameter sets. To minimize this effect, only predicted regional parameter sets with a correlation coefficient of $R^2 < 0.95$ to each of the already available local parameter sets are retained. In other words, if two parameter sets are similar we keep only the one with higher model efficiency.

4. Predict regionally consistent parameters for the distributed model.

The regional functional relationship between parameters and CDs obtained at steps 2 and 3 is applied to predict parameters for each element (i.e., an 8 × 8 km grid cell) of a distributed model.
5. Run the distributed model using regional parameter sets.

The distributed model with the predicted regional parameters is used in all catchments. The median of the model efficiencies obtained in each single catchment, termed ME_{reg}, is used as a measure of performance of the predicted regional parameters (i.e., of the predicted functional relationship).

6. Repeating steps 1–5 to improve ME_{reg}

Steps 1 to 5 are repeated until regional mode efficiency ME_{reg} does not further improve. In our study we define regional model efficiency ME_{reg} as the median ME obtained in single catchments using distributed parameters estimated by means of the inferred regional functional relationship. Notice that as a new combination of parameter sets is selected at step 1 for each iteration, always new regionalization rules are derived.

2.2. Hydrological Model

The model used in this study is a distributed conceptual rainfall-runoff model. The model, called SALTO (The SAme Like The Others) model, is a representative of a soil moisture accounting scheme and similar to many other well-known models in hydrology, such as the HBV model (Bergström, 1995) or elements of the SUPERFLEX modeling approach (Fenicia et al., 2011).

A sketch of the model and its equations are shown in Figure 2. The current version has 15 calibration parameters; uses daily time series of precipitation, potential evapotranspiration, and air temperature as inputs; and is applied to each catchment on an 8 × 8 km grid.

Snowfall and snowmelt are calculated by means of a simple degree-day method and a threshold temperature (DeWalle & Rango, 2008). Soil moisture dynamics are represented by one storage unit (Hundecha & Bárdossy, 2004). Rainfall and snowmelt feed the storage, which releases water as actual evaporation, groundwater recharge, and direct or fast runoff. All three pathways of water release are nonlinear functions of the actual soil moisture state.

Direct runoff enters a fast reacting nonlinear reservoir representing fast surface or interflow, which follows a power outflow law (Kavetski & Fenicia, 2011). Groundwater recharge from the soil moisture storage enters a slow reacting nonlinear reservoir (Fenicia et al., 2006). This reservoir (termed groundwater reservoir later on, although it does not represent groundwater storage in a physical sense) can also be fed by groundwater flow from upstream model elements. Water stored in the groundwater reservoir can leave model elements as river runoff in periods when the groundwater reservoir is full (i.e., similarly to what happens when groundwater tables are high). When only little water is stored in the groundwater reservoir (i.e., low groundwater tables), outflow is directly routed into the groundwater reservoir of the next downstream model element and does not contribute to river flow in the considered model element (Le Moine et al., 2007). River water in the model element is therefore composed of runoff from the fast reacting reservoir, groundwater flow and river runoff from the upstream model elements. River runoff is routed by means of a nonlinear reservoir approach.

In the following of the paper a suite of the 15 model parameters is called a parameter set.

2.3. Data and Study Area

The study is based on data from German catchments with areas ranging from 30 to 21,858 km² (Figure 3a). The climate of Germany is temperate. The northern and northwestern parts of the country have a maritime influenced climate, with warm summer and mild winters, while the eastern part is more influenced by continental climate. Precipitation is dominated by westerly circulation patterns and mean annual precipitation ranges from less than 500 to over 2,500 mm/year. Northern Germany is characterized by lowlands while elevation tends to gradually increase toward the pre-alpine/alpine region in the south. About one third of Germany is forested, while another third is agricultural land.

Daily runoff time series in the period from 1979 to 2002 were available for all catchments. Streamflow of many German catchments is dominated by anthropogenic impacts such as reservoirs, dams, or control gates. Based on a regional analysis of the spatiotemporal patterns of runoff event characteristics (Tarasova, Basso, Zink, & Merz, 2018), a survey of metadata on dams and reservoirs (Lehner et al., 2011) and a manual screening of runoff time series we removed all catchments with reported or suspected flow disturbance.
Daily precipitation time series for every catchment have been aggregated from the Regionalisierte Niederschlagshöhen (REGNIE) data set, which is a multiple regression interpolation product based on orography and inverse distance weighing (Rauhe et al., 2013). Daily temperature time series have been produced from station data available from the German Weather Service (DWD) with external drift kriging using elevation as explanatory variable. Potential evapotranspiration has been calculated with the temperature-based method suggested by Oudin et al. (2005).

Various catchment descriptors have been collected or calculated for every catchment. These include descriptors of climate (such as the mean annual precipitation and mean annual air temperature) and indicators of wet and dry spells (e.g., median volume of precipitation events and long-term median duration of dry spells).

Descriptors of catchment and landscape characteristics include information on topography, groundwater recharge, geomorphology, land use, soil texture and soil properties, drainage, rock origin, and geology. A summary of all CDs used in this study is presented in Table S2 in the supporting information. The adopted CDs are all publicly available and regularly used for hydrological studies (e.g., Tarasova, Basso, Zink, & Merz, 2018; Blöschl et al., 2013; Oudin et al., 2010). We assume that a similar composition of climate and catchment indicators is also available for other countries, although country specific variations in types and how indicators are derived are expected. All continuous variables in the set of descriptors have been
aggregated as mean or median for every catchment. Categorical data (e.g., land use data) have been calculated as percent of catchment area covered by specific catchment descriptor. In the remaining of the paper these climate and catchment descriptors are only termed catchment descriptors for the sake of brevity.

2.4. Local Calibration and Selection of Training and Validation Catchments for the PASS Method

As the PASS method is using locally calibrated lumped parameters as starting point for regionalization, each catchment is calibrated individually assuming constant parameters for all model elements within a catchment, and using 10 years of daily runoff observations (from November 1990 to October 1999). The period from January to October 1990 is used as warm-up period. Parameters are optimized using the Self-Organizing Migrating Algorithm (Zelinka, 2004) ("soma" package in R). For a good model fit to high flow as well as to low flow conditions, we used the following objective function of model efficiency (ME):

\[ ME = 0.5 \times KGE_Q + 0.5 \times KGE_{1/Q} \]  

where \( KGE_Q \) is the Kling-Gupta model efficiency (Gupta et al., 2009) calculated on streamflow, while \( KGE_{1/Q} \) is the Kling-Gupta model efficiency computed on inverse streamflow.

MEs for 373 German catchments using locally calibrated lumped parameters are shown in Figure 3. Calibration results in a median ME of 0.81. The 75th and 25th percentiles of ME are 0.74 and 0.85. The best ME achieved is 0.93. These ranges of model efficiency are similar to those obtained by other studies using similar lumped parameter models and large data set. For example, Parajka et al. (2007) found a median Nash-Sutcliffe model efficiency (NSE) of 0.71 using HBV in 320 catchments located in Austria. Arheimer et al. (2012) modeled 318 Swedish catchments with the HYPE model and obtained a median NSE of 0.74. Newman et al. (2015) found that 90% of the catchments had a NSE > 0.55 using the Sacramento Soil Moisture Accounting Model over 671 American catchments. Poncelet et al. (2017) used the GRJ6 model to simulate runoff of over 1,100 French, German, and Austrian catchments, resulting in a mean NSE of 0.81.
To have an independent test to validate the PASS method, we divided our data set into training catchments, used to develop the regionalization rule, and test catchments, which are instead employed to validate it. In particular, we selected all catchments with no major flow disturbance due to reservoirs, lakes, or anthropogenic management (see data section) and with a locally calibrated ME larger than 0.75, to assure that the model is able to reproduce well the hydrological processes of streamflow generation in that catchment. This resulted in a set of 263 catchments. Out of the 263 catchments, we randomly selected as training catchments 100 of them which fulfill the following criteria: (a) at least five independent parameter sets with ME > 0.75 have to be found by local calibration, as the proposed PASS method is based on shuffling parameters sets; (b) catchment area is smaller than 1,000 km², as the PASS method is based on the idea of deriving regionalization rules for distributed models from lumped model parameters calibrated in small catchments (see section 2.1).

The remaining 163 catchments (further on termed test catchments) are used for validation.

2.5. Variable Importance of CDs
Random forests can be used to rank the importance of CDs used for parameter regionalization. During the process of fitting the random forest’s tree to data, the importance of each CD is assessed. Initially, the predictive accuracy of the tree is calculated. Then each predictor is randomly shuffled (permuted) while keeping the other CDs, and the decrease in predictive accuracy is calculated. A strong decrease in predictive accuracy is expected when important CDs are neglected.

3. Results and Discussion
3.1. Model Performance Using Regionally Consistent Parameter Sets
Figure 4a shows the nonexceedance cumulative distribution of MEs of the locally calibrated (light colors) and regionally predicted (dark colors) parameter sets. The “train_cal” (light green) shows MEs obtained by using locally calibrated lumped parameters for the 100 training catchments, whereas “test_cal” (light blue) represents MEs of locally calibrated lumped parameters for the 163 test catchments. Training and test catchments show similar distributions of MEs with medians around 0.86. The “train_reg” (dark green) shows the MEs of regionally predicted distributed parameter sets for the 100 training catchments obtained through the PASS approach. In Figure 4a only the PASS run resulting in the highest median ME for the 100 training catchments is shown. The “test_reg” (dark blue) summarizes MEs of regionally predicted distributed parameter sets for the test catchments. Regional parameters result in a median ME of 0.74 and 0.72, respectively, for the training and test catchments. The 25% quantiles of MEs for the training and test catchment are 0.68 and 0.64, while the 75% quantiles are 0.79 and 0.78.

Figure 4b shows the MEs obtained by running the model with lumped locally calibrated parameters and with regionally predicted distributed parameters for an independent validation period (1980–1990). The median ME using locally calibrated lumped parameters decreases from 0.87 to 0.79 for the training catchments and from 0.86 to 0.75 for the test catchments compared to the calibration period. The decrease in ME from the calibration to the validation period is smaller when regionally predicted distributed parameters are used. ME decreases from 0.74 to 0.69 for the training catchments and from 0.72 to 0.65 for the test catchments. Hence, the regionally consistent distributed parameters seem to be more robust in predicting changed hydro-climatic situations.

Note that only 2,500 combinations of parameter sets for the 100 training catchments have been tested. The overall number of possible combinations is much larger. In fact, after adding potential parameter sets on the run (see section 2.1), on average 55 sets were available for each of the 100 training catchments, which result in 100⁵⁵ possible combinations. However, in our study the model performance only marginally increases after testing about 1,000 combinations. Note that the current PASS method is based on random sampling. Including a more advanced optimization algorithm may result in better model performances. Although only a small portion of all possible combinations has been tested to decipher a regionalization rule, the ranges of model performance obtained from regionally consistent parameter sets are similar to those of other modeling studies of German catchments. For example, Zink et al. (2017) validated their parameterization of the mHM model (4 × 4 km resolution, daily time steps), based on the Multiscale Parameter Regionalization approach (Samaniego et al., 2010), on a partly similar set of German catchments. The model cross-
validation to streamflow observed in 222 catchments resulted in a median ensemble NSE (Nash & Sutcliffe, 1970) of 0.68. Fleischbein et al. (2006) applied the distributed hydrological model J2000 to the German Mulde catchment (including 56 subcatchments) and obtained calibrated NSE of about 0.7 for a 1 year validation period. Huang et al. (2010) applied the semidistributed model SWIM to all major river catchments in Germany. Calibration to five large river catchments provided NSEs between 0.81 and 0.85, while applying the parameterization to 24 validation catchments model efficiencies vary between 0.43 and 0.86, with an average of 0.6. Hundecha and Bárdossy (2004) used a regional functional relationship for modeling 95 Rhine catchments resulting in NSE between 0.81 and 0.93. Götzinger and Bárdossy (2007) used four different parameter regionalization approaches to run a distributed HBV model in 51 subcatchments of the Neckar basin. Their regionalization procedure resulted in median NSE of 0.5–0.53 for the validation period.

As model performances resulting from the PASS method tend to be similar to other regional modeling studies in Germany, the PASS method seems a promising alternative to other regional calibration approaches. However, for a strict comparison of the performance of various parameter regionalization strategies, one would need similar set of catchments and model input applied to the same model. Such a detailed comparison is not the aim of this study.
Similarly to other large scale modeling approaches (e.g., Samaniego et al., 2010; Clark et al., 2016; Mizukami et al., 2017; Arheimer et al., 2012), we use one fixed model structure and data resolution for a wide range of catchment scales. Adapting model structure and data resolution to catchment scale is expected to improve model performance. However, this is beyond the scope of our work. In fact, here we aim to explore the relation of model parameters with climate and catchment characteristics.

In Figures 4a and 4b MEs of only the best run of the PASS method are shown. However, there are many other combinations of parameter sets from which consistent regional parameters with good model performances could be derived. Similar to many other optimization problems in hydrological modeling, there is not only one single optimum. On the contrary, many other regional parameter sets which show similarly good model performance can be found. To get a hint on the spread of different regional parameter sets derived by the PASS method which provide good results, we selected all regional parameter sets for which $\text{ME}_{\text{reg}} > 0.95 * \max (\text{ME}_{\text{reg}})$, where $\text{ME}_{\text{reg}}$ is the median model efficiency for the 100 training catchments and $\max (\text{ME}_{\text{reg}})$ is the maximum $\text{ME}_{\text{reg}}$ of all runs of the PASS method. These constitute the pool of acceptable regional parameters, similar to the definition adopted to define good parameter sets found by local calibration.

In Figure 4c, good locally calibrated and regionally estimated parameter sets for the Alb River at Ettlingen (a tributary of the Rhine River in south-western Germany) are shown as an example. All parameter sets with $\text{ME} > 0.95 * \max (\text{ME}_\text{loc})$ found by local calibration are shown in gray, while regionally predicted parameter sets with $\text{ME}_{\text{reg}} > 0.95 * \max (\text{ME}_{\text{reg}})$ are shown in black. On the y axis, ME (highlighted by a red rectangle) and parameter values (normalized between 0 and 1) are reported. Each parameter sets is connected with a line. In this case study, a dozen of parameter sets with similar model efficiencies ($\text{ME} = 0.85$–$0.89$) have been found by local calibration. Nearly all parameters span the whole possible range of values. Parameters found by regional calibration show a much smaller spread. Decreasing variability of the regional parameter sets has been found in many German catchments, as summarized in Figure 4d. There, box plots of standard deviation of normalized parameters in each of the 100 training catchments are shown. Clearly visible are the much smaller values and spread of the regionally consistent parameters compared to locally calibrated parameters. For the German data set, using regionally consistent parameters considerably reduces the effect of parameter equifinality.

This study deals with finding regionally consistent parameter sets for large scale modeling. However, the challenge of parameterizing large scale models is very similar to estimating model parameters in ungauged basins. While large scale modeling requires the estimation of parameters of gauged and ungauged catchments or model elements, the focus of PUB Studies (see, e.g., Parajka et al., 2013) is on transferring parameters estimated through local calibration to ungauged catchments. Hence, the PASS approach may be advantageous for PUB applications too, as it overcomes the limitation of traditional model parameter regionalization (namely, the use of only the best locally calibrated parameter sets) taking into account also other good parameter sets to set up a functional relationship between model parameters and catchment descriptors. One may even argue from a hydrological perspective that using only the best locally calibrated parameter sets for estimating regional parameters is not a good choice. In fact, calibration of conceptual hydrological models compensates for errors in inputs or streamflow data to a certain extent (Blöschl, 2005). Calibration may also compensate for local hydrological processes or anthropogenic activities not explicitly considered in the model structure (e.g., active drainage systems of agricultural areas would result in simulation of subsurface runoff faster than one would expect given the local topography and soil characteristics). These compensations of data errors and inadequate model structure during local calibration distort the regional relationship derived between model parameters and catchment descriptors.

### 3.2. Controls on the Regional Parameter Distribution

Figure 5 shows spatial patterns of the regionally predicted parameters. Unlike patterns of parameters obtained from calibration to single catchments, spatial patterns in Figure 5 do not display discontinuities at the catchment borders. The spatial patterns are in line with hydro-meteorological and landscape characteristics of Germany. For example, topographic features (Figure 3a) such as the North German Plain, low mountain ranges (e.g., the Black and Thuringian Forests, the Rheinish Massif), or the northern rim of the Alps in the South of the country (or other climate and landscape characteristics that are highly correlated to topography, such as soil types and annual rainfall) are reflected in the spatial patterns of several parameters. This would suggest that topographic and climate features have a strong control on model
parameters, which is in line with the findings of Singh et al. (2014). They used classification and regression trees to relate catchments descriptors and streamflow signatures and found that similarity in elevation and climate is the most important for succeeding in model parameter transfer across the entire United States. However, they also found that controlling catchment characteristics and streamflow indicators vary significantly for different geographic regions.

The variable importance provided as output by the decision forest approach has been used to have a closer look at the control of climate and catchment descriptors on model parameters. In Figure 6 the variable importance of the best regional parameter set obtained from PASS is shown. Red bars display the importance of climate-related indicators, while green, orange, blue, and yellow bars show descriptors, respectively, related to morphology, hydrogeology, soil, and land use. No clear patterns of important and unimportant catchments descriptors have been found for most parameters. This can be explained by the strong correlation among catchment features. If correlated indicators are used to identify regionalization rules, the random forest, which only considers existing relationships in the data, may use one or the other of the correlated indicators.

The analysis of the variable importance suggests that catchment descriptors can be replaced for predicting regionally consistent parameter sets. One would expect that single descriptors can be replaced by others of

Figure 5. Spatial patterns of the regionally predicted model parameters.
the same type (e.g., one soil descriptor by another soil descriptor). However, how will model efficiency change if all descriptors of one type (e.g., all soil-related ones) would not be available? To answer this question, we reran the PASS approach five times, each time leaving out all descriptors of one generic group (i.e., climate, hydrogeology, land use, morphology, and soil). In Figure 7 the nonexceedance cumulative distributions of MEs for both the training (left panel) and test (right panel) catchments are shown. Leaving out a whole generic group of catchment descriptors results in a decrease of median ME of about 0.02 (Figure 7, left panel) for the training catchments. A similar decrease is observed for the test catchments as well (Figure 7, right panel). The little change of model efficiencies highlights that disregarding whole groups of catchment indicators does not significantly deteriorate the overall model efficiency provided by predicted regional parameter set. The information lost by leaving out one generic group of catchment descriptors is replaced by the one contained in highly correlated descriptors of other generic groups.

Climate and catchment descriptors are highly correlated to each other (see, e.g., Tarasova, Basso, Zink, & Merz, 2018, Figure 6) as a result of the coevolution of climate and landscape. Kirkby (2005, p. 58), e.g., notes: “it is clear that the form of the landscape is vitally important in affecting the response of the landscape to a storm event. The form of the landscape, however, is already a product of the hydrology over a long period, determining the soil, vegetation, hillslope forms and channel network morphology.” Given the coevolution
of climate and landscape (Perdigao & Blöschl, 2014) it is clear that, e.g., mean annual rainfall and elevation or morphological indices and soil characteristics show similar features in their regional patterns. So one could expect that some climate and catchment characteristics can be replaced by others in the parameter estimation process to some extent (Merz & Blöschl, 2004). However, our results show that a whole generic group of descriptors (e.g., all soil descriptors) can be replaced by another one, without any considerable decrease in model performance (Figure 7).

Clearly, soil characteristics (or other generic groups) are important controls on the hydrological response of catchments to rainfall events, so why is no decrease of model performance detected when soil characteristics are not available for the parameter estimation process? There may be three reasons for this phenomenon. First, our catchment-scale climate and catchment descriptors are interpolated from point measurements (e.g., climate data from meteorological stations, soil profiles, and geological data from wells). Depending on the density of point information and the available expert knowledge about regional variability, the interpolated patterns contain more or less uncertainty (Beck et al., 2015), which may distort the regional relationship between model parameters and catchment descriptors. Addor et al. (2018) discuss the effect of uncertainty in our catchment descriptors in predicting hydrological signatures across continental United States. They conclude that “It is indeed likely that issues related to data collection (see discussion in Addor et al., 2017) limit the predictive power of soil data. Accounting for the uncertainty of soil characteristics may for instance make the influence of soils on water dynamics clearer and improve the predictions.”

An overly simplified structure of the adopted model might be the second reason of its insensitivity to whole groups of catchment descriptors. However, the model structure used in this study is similar to others used in large scale modeling studies (e.g., Parajka et al., 2007; Arheimer et al., 2012; Newman et al., 2015; Clark et al., 2016; Mizukami et al., 2017). Embracing Popper's falsification approach to science (Popper, 1963), we leave investigations on the effects of model structure to colleagues interested in the topic and instead postulate, as likely reason of our findings, that the problem lies in the use of conventional climate and catchment descriptors.

![Figure 7](image-url)

*Figure 7.* Nonexceedance cumulative distributions of MEs obtained by using regionally predicted model parameters, when climatic (dark yellow), hydrogeological (light yellow), land use (light green), morphological (dark green), and soil (blue) descriptors are disregarded in turn. The MEs of the locally calibrated (red) and regional parameters predicted by using all catchment descriptors (orange) are shown for comparison. Left panel: training catchments; right panel: test catchments.
In fact, our climate and catchment indices are averaged values of these characteristics over subcatchments, hydrological response units, or grid elements, which ignore small scale patterns. Distributed large scale hydrological models are based on grid elements of several square kilometers and use grid-averaged catchments indices for parameter estimation. Smaller scale structures in soil, vegetation, or morphological patterns, which obviously exist, are ignored. One way to resolve this problem is to model with higher spatial resolution. Much higher resolution (at the order of 1 km globally and of 100 m at continental scales) will be soon feasible (Wood et al., 2011). However, finer spatial resolution of model elements comes at the cost of higher data and computational requirements. Also, hydrological patterns are omnipresent at all scales (Bras, 2015) and it is questionable if higher spatial resolution will ever capture the dominant structures.

An alternative way of proceeding is to replace our average-based catchment descriptors with indicators that better represent hydrological functioning along the present structures. An example of such an approach is the use of pedo-transfer functions in relating model parameters to soil characteristics (Mizukami et al., 2017; Samaniego et al., 2010). As typical cell sizes of distributed models are of the order of one to tens of square kilometers and contain many soil types and structures, a scaling procedure helps to represent subgrid variability of soil within cells (Samaniego et al., 2010).

However, the development of indicators that better represent hydrological functioning is still a largely unchartered research field. For example, it is well known in hydrological literature that runoff generation is highly dominated by threshold effects, hydrologic connectivity, and hydrologic hotspots (Jencso & McGlynn, 2011; Lehmann et al., 2007; Rogger et al., 2012; Seyfried et al., 2009; Tromp-van Meerveld & McDonnell, 2006; Zehe & Blöschl, 2004; Zehe & Sivapalan, 2009), but these phenomenon are not accounted for deriving catchment descriptors. In particular, the interplay of hydrologic hotspots and connectivity strongly dominates the catchment and/or model element response to rainfall. Small parts of a catchment (e.g., areas which are prone to saturation) can produce large runoff amounts. If these areas are connected to the river system, large runoff magnitudes should be expected at the catchment or grid cell outlet. However, if no connectivity exists, runoff at the outlet will be significantly smaller. This highlights that not the average connectivity in the model element is critical, but how hydrological hot spots are connected (see, e.g., Merz & Blöschl, 2008). Unfortunately, this kind of subgrid variability is usually not considered in the available catchment descriptors. As also stated by Parajka et al. (2013) in a review of existing methods for predicting runoff hydrographs in ungauged basins, a new search for hydrologically meaningful catchment descriptors that reflect regional climatic patterns and the internal organization of landscapes is paramount to advance toward reliable forecasts of large scale hydrological models.

Although this study is focusing only on model parameter patterns in space, parameters may vary also with climate in time (Merz et al., 2011; Rosero et al., 2010). Hence, functional relationships between model parameters and climate and landscape characteristics should be adapted to a changing climate.

4. Conclusions

The main findings of this paper are as follows:

1. A newly developed PASS approach is an attractive alternative for regional calibration, providing a consistent parameterization of distributed models (i.e., parameters with a functional relationship to climate and landscape characteristics across catchment boundaries are identified). Its application to a data set of river basins in Germany resulted in good runoff simulations for both training and test catchments, with model efficiencies similar to other rainfall-runoff modeling studies which use regionally consistent parameter sets.

2. The PASS method does not require a priori assumptions on the relationships between model parameters and catchment descriptors. It instead derives these relationships from observed patterns of calibrated parameters and catchment descriptors, thus providing flexibility in respect to available catchment descriptors.

3. In this study, a combination of catchment descriptors that clearly control model parameters is not found. In fact, we show that various regional functional relationships between catchment descriptors and model parameters result in similarly good model performances. Whole groups of catchment descriptors can be easily replaced by other ones without any decrease in model performance. We postulate the problem to
be in the use of climate and catchment descriptors that are based on averages rather than on hydrological functioning. As a consequence of it, only the amount of information, which is also retained in the correlation among descriptors, is exploited in parameterization.

Increased effort to develop and test hydrologically meaningful catchment descriptors, that, e.g., account for lateral connectivity and hot spots, is required for both parameterizing large scale distributed models and predicting model parameters in ungauged catchments.

References

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. Water Resources Research, 54, 8792–8812. https://doi.org/10.1002/2018WR022606

Addor, Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences, 21(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., et al. (2015). Accelerating advances in continental domain hydrologic modeling. Water Resources Research, 51, 10,078–10,091. https://doi.org/10.1002/2015WR017498

Arheimer, B., Dahn, J., Donnelly, C., Lindstrom, G., & Stromvist, J. (2012). Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin—Influence of input-data quality and scale. Hydrology Research, 43, 315–329. https://doi.org/10.2166/ nh.2012.010

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2015). Global maps of streamflow characteristics based on observations from several thousand catchments. Journal of Hydro meteorology, 16(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Bergström, S. (1995). The HBV model. In V. Singh (Ed.), Computer Models of Watershed Hydrology (pp. 443–476). Highlands Ranch, Colorado: Water Resour. Publ.

Bergström, S., & Graham, L. P. (1998). On the scale problem in hydrological modelling. Journal of Hydrology, 211(253), 265. https://doi.org/10.1016/S0022-1694(98)00248-0

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249, 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Beven, K. J., & Binley, A. M. (1992). The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6, 279–298.

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., et al. (2015). Hyper-resolution global hydrological modelling: What is next? Hydrological Processes, 29, 310–320. https://doi.org/10.1002/hyp.10391

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. Water Resources Research, 51, 4923–4947. https://doi.org/10.1002/2015WR017173

Blischi, G. (2005). Rainfall-runoff modelling of ungauged catchments. In M. G. Anderson (Ed.), Encyclopedia of Hydrological Sciences (pp. 2061–2080). Chichester, New York: John Wiley.

Blischi, G., Sivapalan, M., Wagenet, T., Viglione, A., & Savenije, H. (Eds.) (2013). Runoff prediction in ungauged basins: Synthesis across processes and scales. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139257661

Blosch, G., & Zehe, E. (2005). On hydrological predictability. Invited commentary. Hydrological Processes, 19(19), 3923–3929.

Bras, R. L. (2015). Complexity and organization in hydrology: A personal view. Water Resources Research, 51, 6532–6548. https://doi.org/10.1002/2015WR016958

Breniman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., et al. (2016). Improving the theoretical underpinnings of process-based hydrologic models. Water Resources Research, 52, 2350–2365. https://doi.org/10.1002/2015WR017910

Cutler, D. R., Edwards, T. C. Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783–2792. https://doi.org/10.1890/07-0539.1

DeWalle, D. R., & Range, A. (2008). Modelling snowmelt runoff. In Principles of Snow Hydrology (pp. 266–305). New York: Cambridge University Press.

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2007a). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. Water Resources Research, 47, W11510.

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling. Hydrology and Earth System Sciences, 10, 139–150. https://doi.org/10.5194/hess-10-139-2006

Fernandez, W., Vogel, R. M., & Sankarasubramanian, A. (2000). Regional calibration of a watershed model. Hydrological Sciences Journal, 45(5), 689–707. https://doi.org/10.1080/02666660000942371

Fleischbein, K., Lindenschmidt, K.-E., & Mertz, B. (2006). Modelling the runoff response in the Mulde catchment (Germany). Advances in Geosciences, 9, 79–84. https://doi.org/10.5194/adgeo-9-79-2006

Freyberg, J., Radny, D., Gall, H. E., & Schirmer, M. (2014). Implications of hydrologic connectivity between hillslopes and riparian zones on streamflow composition. Journal of Contaminant Hydrology, 89, 62–74. https://doi.org/10.1016/j.jconhyd.2014.07.005

Gonzalez-Ferreras, A. M., & Barquin, J. (2017). Mapping the temporary and perennial character of whole river networks. Water Resources Research, 53, 6709–6724. https://doi.org/10.1002/2017WR023090

Götzionger, J., & Börsdorff, A. (2007). Comparison of four regionalisation methods for a distributed hydrological model. Journal of Hydrology, 333, 374–384.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Perrin, C., Bloschli, G., Montanari, A., Kumar, R., Clark, M., & Andrássian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. Hydrology and Earth System Sciences, 18(2), 463–477. https://doi.org/10.5194/hess-18-463-2014

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y., & Wagener, T. (2015). A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. Geoscientific Model Development, 8, 1729–1746. https://doi.org/10.5194/gmd-8-1729-2015
Henzsken, H. J., Trolldborg, L., Nyegaard, P., Sonnenborg, T. O., Refgaard, J. C., & Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. *Journal of Hydrology*, 280, 52–71. https://doi.org/10.1016/S0022-1694(03)00316-0

Holster, S. W., & Alder, J. R. (2016). Implementation and evaluation of a monthly water balance model over the US on an 800 m grid. *Water Resources Research*, 52, 9600–9620. https://doi.org/10.1002/2016WR018665

Huang, S., Krysanova, V., Osterle, H., & Hattermann, F. (2010). Simulation of spatiotemporal dynamics of water fluxes in Germany under climate change. *Hydrological Processes*, 24, 3289–3306. https://doi.org/10.1002/hyp.7553

Hundecha, Y., & Bärdossy, A. (2008). Modeling the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 362, 281–295. https://doi.org/10.1016/j.jhydrol.2004.01.002

Hundecha, Y., Ouarda, T. B. M. J., & Bärdossy, A. (2008). Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space. *Water Resources Research*, 44, W01427. https://doi.org/10.1029/2006WR005439

Jencso, K. G., & McGlynn, B. L. (2011). Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, 47, W11527. https://doi.org/10.1029/2011WR010666

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, 47, W11151. https://doi.org/10.1029/2011WR010748

Kirkby, M. (2005). Organisation and process. Article 4 in: *Encyclopedia of Hydrological Sciences*, M. G. Anderson (Managing Editor) (pp. 41–85). Chichester: J. Wiley & Sons.

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

Lehmann, P., Hinz, C., McGrath, G., Tromp-van Meerveld, H. J., & McDonnell, J. (2007). Rainfall threshold for hillslope runoff: An emergent property of flow pathway connectivity. *Hydrology and Earth System Sciences Discussions*, 11, 1047–1063.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9), 494–502. https://doi.org/10.1890/100125

McGlynn, B. L., & McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, 39(11), 1310. https://doi.org/10.1029/2003WR002901

Merz, R., & Blöschl, G. (2004). Regionalisation of catching model parameter models. *Journal of Hydrology*, 287, 95–123.

Merz, R., & Blöschl, G. (2008). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, 44, W08432. https://doi.org/10.1029/2007WR006744

Merz, R., & Blöschl, G. (2009). A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resources Research*, 45, W01405. https://doi.org/10.1029/2008WR007163

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 47, W02531. https://doi.org/10.1029/2010WR009595

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., et al. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, 53, 8020–8040. https://doi.org/10.1002/2017WR020401

Nash, J., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10, 398–409. https://doi.org/10.1016/0022-1694(70)90255-6

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Ancill, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, 303(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Oudin, L., Kay, A., Andréassian, V., & Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, 46, W11558. https://doi.org/10.1029/2009WR004887

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, 51, 7090–7129. https://doi.org/10.1002/2015WR017780

Parajka, J., Blöschl, G., & Merz, R. (2007). Regional calibration of catchment models: Potential for ungauged catchments. *Water Resources Research*, W06406. https://doi.org/10.1029/2006WR005271

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins: Part 1: Runoff hydrograph studies. *Hydrology and Earth System Sciences*, 17, 1783–1795.

Perdigao, R. A. P., & Blöschl, G. (2014). Spatiotemporal flood sensitivity to annual precipitation: Evidence for landscape-climate coevolution. *Water Resources Research*, 50, 5492–5509. https://doi.org/10.1002/2014WR015365

Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., & Le Moine, N. (2008). Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research*, 44, W08447. https://doi.org/10.1029/2007WR006579

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

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

Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V., et al. (2017). Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. *Water Resources Research*, 53, 7247–7268. https://doi.org/10.1002/2016WR019991

Pepper, K. (1963). *Science as falsification, in Conjectures and Refutations: The growth of scientific knowledge* (pp. 33–39). London: Routledge and Kegan Paul.

Rauhe, M., Steiner, H., Riediger, H., Mazurkiewicz, A., & Gratzi, A. (2013). A Central European precipitation climatology—Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, 22(3), 235–256. https://doi.org/10.1127/0941-2948/2013/0436

Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., et al. (2012). Runoff models and flood frequency statistics for design flood estimation in Austria—Do they tell a consistent story? *Journal of Hydrology*, 456-457, 30–43. https://doi.org/10.1016/j.jhydrol.2012.05.068
Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L. E., Yatheendrudas, S., & Niu, G.-Y. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. *Journal of Geophysical Research*, 115, D03106. https://doi.org/10.1029/2009JD012035

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

Schaeffli, B., & Gupta, H. (2007). Do Nash values have value? *Hydrological Processes*, 21, 2075–2080. https://doi.org/10.1002/hyp.6825

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98, 279–293. https://doi.org/10.1016/S0168-1924(99)00105-7

Seyfried, M. S., Grant, L. E., Marks, D., Winstral, A., & McNamara, J. (2009). Simulated soil water storage effects on streamflow generation in a mountainous snowmelt environment, Idaho, USA. *Hydrological Processes*, 23, 858-873. https://doi.org/10.1002/hyp.7211

Sieg, T., Vogel, K., Merz, B., & Kreibich, H. (2017). Tree-based flood damage modeling of companies: Damage processes and model performance. *Water Resources Research*, 53, 6050–6068. https://doi.org/10.1002/2017WR020784

Singh, R., Archfield, S. A., & Wagener, T. (2014). Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments—A comparative hydrology approach. *Journal of Hydrology*, 57, 985–996. https://doi.org/10.1016/j.jhydrol.2014.06.030

Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Saussart, E., & Shankar, U. (2009). Predictive mapping of the natural flow regimes of France. *Journal of Hydrology*, 373(1–2), 57–67. https://doi.org/10.1016/j.jhydrol.2009.04.011

Sørensen, R., Zinko, U., & Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98, 279–293. https://doi.org/10.1016/S0168-1924(99)00105-7

Troch, P., Paniconi, C., & van Loo, E. (2003). Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hilllopes: 1. Formulation and characteristic response. *Water Resources Research*, 39(11), 1316. https://doi.org/10.1029/2002WR001728

Tromp-van Meerveld, H., & McDonnell, J. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42, W02411. https://doi.org/10.1029/2004WR003890

Wagener, T., Boyle, D. P., Lees, M. J., Wheatier, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, 5, 13–26.

Wood, E. F., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water. *Water Resources Research*, 47, W05301. https://doi.org/10.1029/2010WR010090

Zelinka, I. (2004). SOMA—self-organizing migrating algorithm. In G. C. Onwubolu & B. V. Babu (Eds.), *New optimization techniques in engineering, Volume 141 of “Studies in Fuzziness and Soft Computing*” (pp. 167–217). Berlin, Heidelberg: Springer.

Zhang, X.-J., Tang, Q., Pan, M., & Tang, Y. (2014). A long-term land surface hydrologic fluxes and states dataset for China. *Journal of Hydrometeorology*, 15, 2067–2084. https://doi.org/10.1175/JHM-D-13-0170.1

Zhu, C., & Lettenmaier, D. P. (2007). Long-term climate and derived surface hydrology and energy flux data for Mexico: 1925–2004. *Journal of Climate*, 20, 1936-1946. https://doi.org/10.1175/JCLI4086.1

Zink, M., Rohini, K., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrolgy and Earth System Sciences*. https://doi.org/10.5194/hess-21-1769-2017