# Automatic regionalization of model parameters for hydrological models

Feigl Moritz[1], Thober Stephan[2], Schweppe Robert[3], Herrnegger Mathew[1], Samaniego Luis E.[2], and Schulz Karsten[1]

[1]University of Natural Resources and Life Sciences, Vienna
[2]UFZ-Helmholtz Centre for Environmental Research
[3]Helmholtz Centre for Environmental Research

November 16, 2022

**Abstract**

Parameter estimation is one of the most challenging tasks in large-scale distributed modeling, because of the high dimensionality of the parameter space. Relating model parameters to catchment/landscape characteristics reduces the number of parameters, enhances physical realism, and allows the transfer of hydrological model parameters in time and space. This study presents the first large-scale application of automatic parameter transfer function (TF) estimation for a complex hydrological model. The Function Space Optimization (FSO) method can automatically estimate TF structures and coefficients for distributed models. We apply FSO to the mesoscale Hydrologic Model (mHM, mhm-ufz.org), which is the only available distributed model that includes a priori defined TFs for all its parameters. FSO is used to estimate new TFs for the parameters "saturated hydraulic conductivity" and "field capacity", which both influence a range of hydrological processes. The setup of mHM from a previous study serves as a benchmark. The estimated TFs resulted in predictions in 222 validation basins with a median NSE of 0.68, showing that even with 5 years of calibration data, high performance in ungauged basins can be achieved. The performance is similar to the benchmark results, showing that the automatic TFs can achieve comparable results to TFs that were developed over years using expert knowledge. In summary, the findings present a step towards automatic TF estimation of model parameters for distributed models.

# Automatic regionalization of model parameters for hydrological models

**Moritz Feigl[1], Stephan Thober[2], Robert Schweppe[2], Mathew Herrnegger[1], Luis Samaniego[2]and Karsten Schulz[1]**

[1]University of Natural Resources and Life Sciences, Vienna, Department of Water, Atmosphere and Environment, Institute for Hydrology and Water Management, Muthgasse 18, 1190 Vienna, Austria
[2]Helmholtz Centre for Environmental Research – UFZ, Department Computational Hydrosystems, Permoserstraße 15, 04318, Leipzig, Germany

**Key Points:**

- This study shows the performance of automatic transfer function (TF) estimation with the Function Space Optimization (FSO) method
- We show that FSO is able to estimate parameter regionalization via TFs that perform similar to the mesoscale Hydrologic Model (mHM) TFs in an ungauged setting
- This study represents a step towards automatic TF estimation for physically interpretable model parameters

Corresponding author: Moritz Feigl, `moritz.feigl@boku.ac.at`

**Abstract**

Parameter estimation is one of the most challenging tasks in large-scale distributed modeling, because of the high dimensionality of the parameter space. Relating model parameters to catchment/landscape characteristics reduces the number of parameters, enhances physical realism, and allows the transfer of hydrological model parameters in time and space. This study presents the first large-scale application of automatic parameter transfer function (TF) estimation for a complex hydrological model. The Function Space Optimization (FSO) method can automatically estimate TF structures and coefficients for distributed models. We apply FSO to the mesoscale Hydrologic Model (mHM, `mhm-ufz.org`), which is the only available distributed model that includes a priori defined TFs for all its parameters. FSO is used to estimate new TFs for the parameters "saturated hydraulic conductivity" and "field capacity", which both influence a range of hydrological processes. The setup of mHM from a previous study serves as a benchmark.

The estimated TFs resulted in predictions in 222 validation basins with a median NSE of 0.68, showing that even with 5 years of calibration data, high performance in ungauged basins can be achieved. The performance is similar to the benchmark results, showing that the automatic TFs can achieve comparable results to TFs that were developed over years using expert knowledge. In summary, the findings present a step towards automatic TF estimation of model parameters for distributed models.

## 1 Introduction

Large-domain, spatially contiguous hydrological and land surface models are important tools for managing our water supplies. Hydrological information on the continental or global scale is needed to handle new emerging international and global water management challenges, which include topics like water allocation in international, national, and large river basins, operational flood forecasting services, global water security or the influence of climate extremes on water resources (Archfield et al., 2015). These applications are particularly challenging in areas without hydrologic measurements, which includes a majority of basins worldwide that are effectively ungauged (Hrachowitz et al., 2013). This results in the need for further development in large-domain hydrological modeling to simulate water fluxes and states in both gauged and ungauged basins in different climates in a spatially consistent manner (Rakovec et al., 2019).

In 1982, Jim Dooge stated that "the parameterization of hydrologic processes to the grid-scale of general circulation models is a problem that has not been tackled, let alone solved" (Dooge, 1982) and shortly after that Leavesley et al. (1983) concluded that optimization of distributed parameters of hydrological models is an "ill-posed" problem due to the large number of degrees of freedom. Since then, model parameterization is still one of the major unsolved problems in hydrology (Blöschl et al., 2019). One way to potentially solve this problem is to relate hydrological model parameters/structures to landscape properties (e.g., K. J. Beven & Franks, 1999; K. Beven, 2002; Hundecha & Bárdossy, 2004; Samaniego et al., 2010; Clark et al., 2016). This approach is strongly related to the idea of regionalization - the geographical migration of hydrological model structures (Buytaert & Beven, 2009). This task is however nontrivial and Clark et al. (2017) still described it as one of the unsolved challenges for hydrological model parameter estimation.

One potential solution to this problem are parameter transfer functions (TFs) as mathematical expressions to formulate the relationship between model parameters and physiographic characteristics (e.g. elevation, slope, soil texture, vegetation characteristics, etc.) of the catchment (Hundecha & Bárdossy, 2004; Samaniego et al., 2010; Kumar et al., 2013). By defining TFs for all parameters, we expect to induce three attractive properties into the hydrological model:

1. The model has a significantly lower number of free parameters that is independent of the size of the model domain, facilitating parameter optimization.
2. The model can be transferred across time and space.
3. The model parameters reflect physical properties of the catchment and result in physical meaningful model states.

The first property would lead to a tremendous decrease in effort to set up and run distributed hydrological models. The problem of estimation of distributed parameters - an ill-posed problem because of the large number of model parameters - would potentially be solved. The second property would allow for prediction in ungauged basins and other time periods (Hrachowitz et al., 2013). Finally, the third property would allow the usage of model states and parameter fields to gain further insights into the hydrological properties and status of a catchment, which are generally extremely difficult or impossible to measure over such large areas (e.g. catchment-wide soil moisture, soil properties, evapotranspiration). The first two properties would result from TFs that are only dependent on a few numerical coefficients and predict discharge equally well in calibration and validation basins and time periods. The third property can potentially be achieved by the constrained setting of using TFs for parameters with a clear physical meaning and by relevant physiographic catchment characteristics as inputs.

Generally, the implementation of TFs for the estimation of distributed model parameters can be seen as a promising step towards adequately addressing critical water cycle science questions and global applications of hyperresolution hydrological and land surface models (Wood et al., 2011; K. J. Beven & Cloke, 2012). A corresponding requirement for hyperresolution models was also stated by Bierkens (2015): hydrological models should be able to make predictions "everywhere", but the predictions should be "locally relevant". For these reasons, Bierkens (2015) suggested that the multiscale parameter regionalization technique (MPR), which uses TFs at its input data's native spatial resolution to scale model parameters to the required spatial scale, could be a way forward. Overall, this will be an important step in the direction towards the application and parameterization of global hyperresolution models.

In a previous study, we developed a method to automatically estimate transfer functions from data called Function Space Optimization (FSO) (Feigl et al., 2020), which further developed ideas first proposed by Klotz et al. (2017) and Klotz (2020). FSO is based on a text-generating neural network that is used to transfer the search for a best fitting transfer function in a continuous optimization problem. While other approaches consist of applying or adapting TFs by modifying their parameter (i.e. numeric coefficients) (e.g., Samaniego et al., 2010; Kumar et al., 2013; Imhoff et al., 2020; Pinnington et al., 2021), FSO can additionally change the functional form of the TF. So far, FSO was thoroughly tested on synthetic data by Feigl et al. (2020) and some initial results of a real-world application were presented in Feigl et al. (2021).

The mesoscale Hydrological Model (mHM) (Samaniego et al., 2010; Kumar et al., 2013; Thober et al., 2019) is a distributed hydrological model that was already applied in numerous studies and for a wide range of different tasks, covering different hydroclimatic conditions (e.g., Kumar et al., 2013; Thober et al., 2018; Peichl et al., 2018; Samaniego et al., 2019; Jing et al., 2020; Imhoff et al., 2020; Saha et al., 2021). mHM is unique, as it has already TFs defined for all its parameters, which were chosen by Samaniego et al. (2010); Kumar et al. (2013); Thober et al. (2019) based on pedo-transfer functions from literature, a "step-wise" method (Samaniego & Bárdossy, 2005) and a "trial-and-error" approach. This makes mHM an ideal model for testing FSO because we can compare the automatically estimated TFs with those chosen by expert knowledge and tested rigorously in multiple studies.

Besides the choice of model, choosing a benchmark study that applied mHM over multiple basins and in a prediction in ungauged basins (PUB) setting is important for

objectively assessing the FSO performance. For this purpose, we chose the study by Zink et al. (2017) because it included a state-of-the-art optimization and application of mHM over a large number of basins. Zink et al. (2017) estimated 100 global mHM parameters sets, i.e., the numerical coefficients of all mHM TFs, for 7 large basins located in Germany with 5 years of data, which they then applied on 222 validation basins in Germany with a mean of 42 years years of data.

This study assesses the performance and further develops the FSO approach and thus presents the next step in direction of regularizing the parameter space by using landscape information for distributed hydrological models. Its originality includes (i) further improvements of the FSO method, (ii) a large-scale application of automatic TF estimation using real-world data and benchmark for comparison, and (iii) a detailed description of challenges, potential limitations, and a way forward for automatic TF estimation.

## 2 Function Space Optimization (FSO)

### 2.1 Methodology

FSO is an optimization method for TFs of distributed models introduced by Feigl et al. (2020). It is based on the idea of transferring the search for a mathematical equation into a continuous optimization problem. All steps of the FSO optimization loop are shown in Figure 1. As in any continuous optimization problem, an optimization algorithm (optimizer, Figure 1 top) is used to find the point in a continuous vector space that minimizes or maximizes an objective function. However, the main difference between FSO and continuous optimization is that we are not interested in the numeric values that are optimized. We are only interested in the TFs that can be generated from them with the text generating neural network.

FSO uses the decoder of a variational autoencoder (VAE; Kingma & Welling, 2013) to generate TFs from a numeric vector. The VAE network is trained to encode and decode the information of TF strings and their resulting parameter distribution into a numeric vector. Data for training is generated using a context free grammar (Knuth, 1965), which defines the possible structures, operators, functions, and variables that compose a TF. The variables consist of distributed (e.g. gridded) physiographic properties, that are the basis for parameter estimation. After generating new TFs from Function Space, the hydrological model is applied using the parameter maps generated from these TFs. The prediction of this model can then be used to compute a loss, which is based on a user-defined objective function that is evaluated in each calibration basin, e.g. NSE (Nash & Sutcliffe, 1970), KGE (Gupta et al., 2009). This loss is then used by the optimizer to choose the next point in Function Space for evaluation.

To enable the unbiased estimation of universally applicable TFs, FSO uses two types of scaling. First, all physiographic properties are scaled to [0,1] before being applied in a TF. Second, the resulting TF values are scaled to the parameter bounds. Both scaling operations use min-max scaling, which needs a minimum and maximum value for both the initial value range and the projected value range. The initial value range for the physiographic variables is chosen to be their physical bounds, e.g. sand content [0, 100]. The initial value range for the TF values is chosen using the VAE training data and is the same for all generated TFs. A detailed description of FSO and all its preprocessing steps are given in Feigl et al. (2020).

We further developed the FSO VAE architecture to improve the encoding of long function strings. This should result in a Function Space with a smoother loss surface and thus making it a more adequate search space for optimization. A detailed depiction and description of the new network architecture is shown in Appendix A.

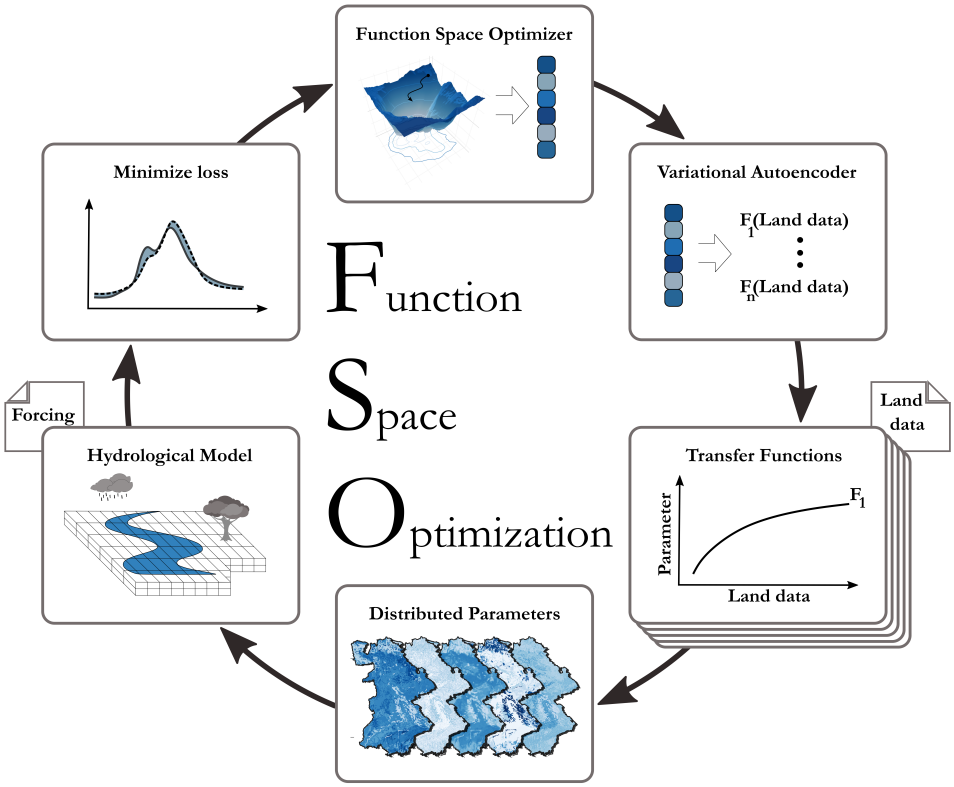


**Figure 1.** A depiction of the Function Space Optimization loop. Starting from the top going clockwise: The optimizer selects the next point in Function Space that should be evaluated. The VAE decoder generates the TFs that are associated with this point in Function Space. The domain of the TFs comprises physiographic properties (e.g. elevation, slope, soil texture, vegetation characteristics) and are used to generate parameter maps. These parameter maps are used as input for the distributed hydrological model to produce predictions (e.g. discharge, evapotranspiration), which results in a loss that represents the aggregated loss of all modeled basins. This loss is then used by the optimizer to decide on the next point in Function Space to be evaluated.

### 2.2 Linking FSO with the mesoscale Hydrologic Model (mHM)

The mesoscale Hydrologic Model (mHM; Samaniego et al., 2010; Kumar et al., 2013; Samaniego et al., 2019; Thober et al., 2019) is a distributed hydrological model that simulates hydrological processes on a multi-layer grid. It uses the multiscale parameter regionalization method (MPR; Samaniego et al., 2010; Schweppe et al., 2021) and thus uses TFs for all its parameters. The applied numerical approximations and conceptualizations are based on the HBV model (Bergström, 1995) and include the processes interception, snow accumulation, snowmelt, infiltration, surface runoff, soil water retention, runoff generation, evaporation, percolation, baseflow, and routing. A detailed description can be found in Samaniego et al. (2010).

In this study we apply FSO to estimate new TFs for the mHM parameters $K_S$ (saturated hydraulic conductivity, [cm/day]) and $FieldCap$ (field capacity, [-]). These parameters affect the storage and conductivity of soil water and have a high sensitivity for streamflow estimation (Cuntz et al., 2015; Höllering et al., 2018). We want to minimize the effect of parameter dependency because this is the first large-scale application of FSO with real-world data. Therefore, we focus only on the estimation of these two TFs, which

allows for a more in-depth analysis of the results. The current version of mHM estimates $K_S$ using a TF that was developed by Cosby et al. (1984):

$$KS - mHM = \gamma_1 \exp(\gamma_2 + \gamma_3 \nu_1 - \gamma_4 \nu_2) \log(10), \tag{1}$$

which is a function of the sand $\nu_1$ and the clay $\nu_2$ content of the soil and numerical coefficients $\gamma_1$to$\gamma_4$. The current mHM parameter $FieldCap$ is estimated using a TF by Twarakavi et al. (2009):

$$FieldCap_{mHM} = ThetaS \exp(\gamma_5(\gamma_6 + log10(K_S)) \log(vGenu_n), \tag{2}$$

where $ThetaS$ is the saturated soil water content, $\gamma_5$, $\gamma_6$ are numerical coefficients and $vGenu_n$ refers to the van Genuchten $n$ model parameter (van Genuchten, 1980). While mHM $K_S$ is only a function of observed physiographic properties, mHM $FieldCap$ is also dependent on the other mHM parameters $ThetaS$ and $vGenu_n$.

# 3 Experimental design

## 3.1 Benchmark and study data

The results of this study are compared to the results of Zink et al. (2017). In their study, they calibrated global mHM parameters, i.e., the numerical coefficients of all mHM TFs, using 5 years of data for 7 large basins in Germany. The calibration was conducted using the DDS optimization algorithm (Tolson & Shoemaker, 2007) with 2000 iterations and was performed 100 times in each of the 7 basins. From these, they selected 100 parameter sets that had a Nash-Sutcliffe model efficiency (Nash & Sutcliffe, 1970) exceeding 0.65 in all 7 basins. Finally, they validated these 100 parameter sets using 42 years of data of 222 smaller basins across Germany. This provides an estimate of the mHM performance in an ungauged setting. We choose the same data and experimental setup as Zink et al. (2017) to make results comparable.

All meteorological forcings, physiographic properties, and discharge observations are taken from Zink et al. (2017), which also includes a detailed description of the preceding data preparation. The study basins consist of 7 large basins used for calibration and 222 smaller validation basins. The calibration basins have a size range of 6 200 km$^2$ to 47 500 km$^2$, while the validation basins have a size range of 100 km$^2$ to 8 500 km$^2$. The area of all calibration basins is shown in Figure 2a and the outlets of all validation basins are shown in Figure 2b. Of these 222 validation basins, 80 are located outside and 142 are located inside the calibration basins area.

The available variables for TF estimation consist of six physiographic properties on a 100 m×100 m grid: sand content in percent ($\nu_1$), clay content in percent ($\nu_2$), mineral bulk density in g/cm$^3$ ($\nu_3$), aspect in degree ($\nu_4$), slope in degree ($\nu_5$) and elevation in m ($\nu_6$). The variables $\nu_4$, $\nu_5$, and $\nu_6$ were derived from a 50 m digital elevation model (DEM) acquired from the German Federal Agency for Cartography. The soil properties are based on a digitalized soil map of the German Federal Institute for Geosciences and Natural Resources (BGR) and contain information for different soil horizons.

The meteorological forcings that are used for running mHM consist of daily fields of precipitation and maximum, minimum, and average temperature, which were derived from local observations from the German Weather Service (Deutscher Wetterdienst, DWD). Daily streamflow data was provided by the European Water Archive (EWA) and the Global Runoff Data Center (GRDC). Land cover information was taken from the CORINE land cover scenes of the years 1990, 2000, and 2006.

## 3.2 mHM setup and objective function

Following Zink et al. (2017), the resolution of the mHM model is 4 km×4 km and each simulation is conducted with a 5-year spin-up period. Calibration data consists of

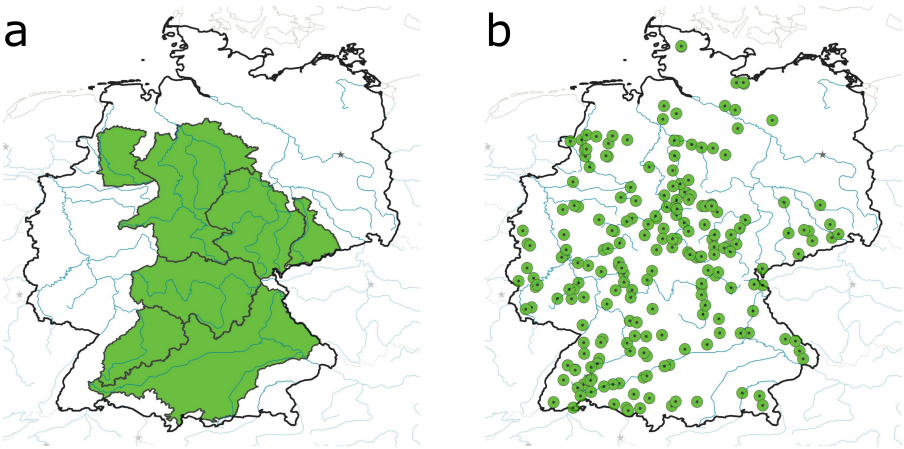


**Figure 2.** Overview of the study basins in Germany. **a** The seven large basins used for calibration. **b** The 222 validation basins shown with the position of their gauges.

the years 2000-2004 of the 7 large basins. The performance of these 7 large basins is validated in the time period 1965-1999. Transfer in space and time is tested by using data between 1955 and 2009 of 222 validation basins, resulting in a mean simulation time period of 42 years and a minimum simulation period of 10 years. The objective function $\Phi$ for evaluating the mHM performance is chosen to be a combination of the NSE and log NSE criteria in form of a power mean:

$$\Phi = (\frac{1}{2}\sum_{i=1}^{2}\phi_i^p)^{\frac{1}{p}}, \tag{3}$$

with $\phi_1 = NSE(Q_{obs}, Q_{sim})$, $\phi_2 = \log NSE(Q_{obs}, Q_{sim})$, $Q_{obs}$ and $Q_{sim}$ the observed and simulated discharge and $p = 6$. This objective function was chosen by Zink et al. (2017) as it ensures equal improvement of both criteria during a multi-objective optimization.

### 3.3 FSO setup

FSO is applied to identify the two TFs and their parameters to regionalize $K_S$ and $FieldCap$ and simultaneously to optimize the numeric coefficients of all other mHM TFs. The resulting numeric vector, which represents all optimizable values, consists of 12 dimensions for the two TFs (two 6-dimensional Function Spaces) and 59 dimensions for all other parameters.

Before applying FSO, it is necessary to train the FSO VAE using a set of function strings and their resulting parameter distribution. This training data was generated using a context free grammar (CFG) that included the physiographic variables, numeric values, the operators +, -, *, /, and a range of mathematical functions: the exponential function, the logarithm function with base $e$ or 10, trigonometric functions (sin, cos, tan), and their arcus and hyperbolic versions, the square root and power functions. The numeric values were chosen to be in the range [-3, 3], discretized with a 0.05 step size. This interval was chosen to allow for a wide range of parameter values to be generated while keeping the search space size manageable. Since we use physiographic variables that are scaled to the range [0, 1], the numeric values thus can be up to three times larger. The CFG is shown in Appendix B, which includes all its building blocks and structural components. This CFG is used to generate 45 million unique TFs that are used to train the FSO VAE. To make our results comparable to Zink et al. (2017), we trained two differ-

ent FSO VAEs, one for each regionalized parameter. The difference between those two are the variables that are used to generate functions: while $K_S$ only uses observed physiographic properties, $FieldCap$ can also use the mHM parameters $K_S$, ThetaS ($\nu_7$) and $vGenu_n$ ($\nu_8$) as potential inputs. Further technical details of the FSO VAE, including all bounds used for scaling, are given in Appendix B.

FSO aims to find the set of TFs that results in the best performance in all available basins. Therefore, Feigl et al. (2020) defined the FSO loss function $f_{loss}$ for $I$ basins to be a weighted mean with more weight on basins with bad performance:

$$f_{loss} = -\frac{\sum_{i=1}^{I} w_i \Phi_i}{\sum_{i=1}^{I} w_i} + \lambda(TFs), \text{ with each } w_i = \sup \Phi - \Phi_i \tag{4}$$

Here, $\Phi_i$ is the value of the objective function for basin $i$, $w_i$ is the corresponding weight and $\lambda(TFs)$ the penalty for TF lengths. The weights are computed by using the supremum of $\Phi$, which is 1 for the above described objective function that is based on NSE and log NSE. $\lambda$ is a function of the number of variables, numerics, operators, parenthesis and functions used in the TFs (the length) and can be computed with $\lambda(TFs) = \frac{1}{M}\sum_{m=1}^{M} \alpha \text{ length}(TF_m)$, where $\alpha$ was chosen to be 0.001.

In our preliminary test runs, we used the dynamically dimensioned search (DDS) algorithm (Tolson & Shoemaker, 2007) for optimizing in Function Space. However, we noticed that it converged extremely fast ($< 500$ iterations) potentially as a result of the algorithm's abidance in local minima. Thus we decided on mainly using the shuffled complex evolution (SCE) algorithm (Duan et al., 1992) for this study. To examine if there is a difference between those two, we include one optimization run with the DDS algorithm. For both optimizers, the optimization is applied for a minimum of 2000 iterations and a maximum of 5000 iterations. It will be stopped if after the minimum required iterations there is no further improvement for 1000 iterations. After one of the stopping criteria was reached, all numeric coefficients that are present in the two TFs are further optimized with 100 iterations of the Genetic Algorithm (Holland, 1975). This additional optimization allows only a $\pm 5\%$ change of the numeric coefficients and represents an adjustment that is not bound by the discretization of the numeric coefficients in the FSO VAE.

### 3.4 Optimization budget and evaluation criteria

The FSO method is applied 5 times to the 7 calibration basins, resulting in 5 independent optimization runs. Four optimization runs will use the SCE algorithm (run 1-4), while one will use the DDS algorithm (run 5). Their performance will be evaluated using the $f_{loss}$, NSE, log NSE, and KGE values.

The validation results of Zink et al. (2017) will be compared to the best performing FSO optimization run. The decision of the best performing optimization run will be based on the NSE, log NSE, KGE and percentage bias (PBIAS; Sorooshian et al., 1993) values in 20 randomly sampled validation basins (approximately 10% of the validation basins). The definition of all performance metrics is given in Appendix C. The random validation sample, which is used to define the best FSO run, will be drawn from a stratified Budkyo curve (Budyko, 1974) to adequately represent the range of different climates in the study area. Since the results of Zink et al. (2017) are the NSE distributions for all validation basins resulting from the 100 parameter sets, the minimum, maximum, 5% quantile, 95% quantile, and median NSE of each basin will be used for comparison.

Different TFs can potentially result in similar parameter values, thus we will also compare the FSO parameter fields to the mHM default parameter fields of the two pa-

rameters $K_S$ and $FieldCap$. For this purpose, the 7 calibration basins will be used, as they present a contiguous field covering a large part of Germany.

# 4 Results

## 4.1 Comparison of optimization runs

This section presents the results of the final parameter sets of each FSO optimization run. The progression during optimization of all 5 FSO runs is shown in Appendix D. Figure 3 shows boxplots with the performance of the final parameter sets of all FSO runs for the calibration time period (2000-2004) and validation time period (1965-1999) in the 7 calibration basins. This also includes the KGE values, which were not part of the loss function and thus had no influence on the optimization. Run 2-4 show very similar calibration performance, which differs from run 1 and 5 performances. The median calibration NSE values of run 1 with 0.76 and run 5 with 0.79 are similar to the run 2-4 medians with a mean of 0.76, but have a much larger variance ($\sigma_{run1\&5} = 0.1$, $\sigma_{run2-4} = 0.03$). This is also the case for run 5 calibration log NSE values, whereas run 1 calibration log NSE values also have a lower median compared to all other runs. Log NSE values are especially similar in runs 2-4 with median values ranging between 0.79 and 0.81. The KGE values show slightly different behavior. Run 2 KGE values for the calibration period are lower with a median of 0.68 compared to run 3 and 4 a mean median KGE of 0.78.

The difference between calibration and validation time period of KGE values is negligible with a median difference of -0.01. The differences of NSE values with a median of -0.03 and the differences of log NSE values with a median of -0.04 are more pronounced.

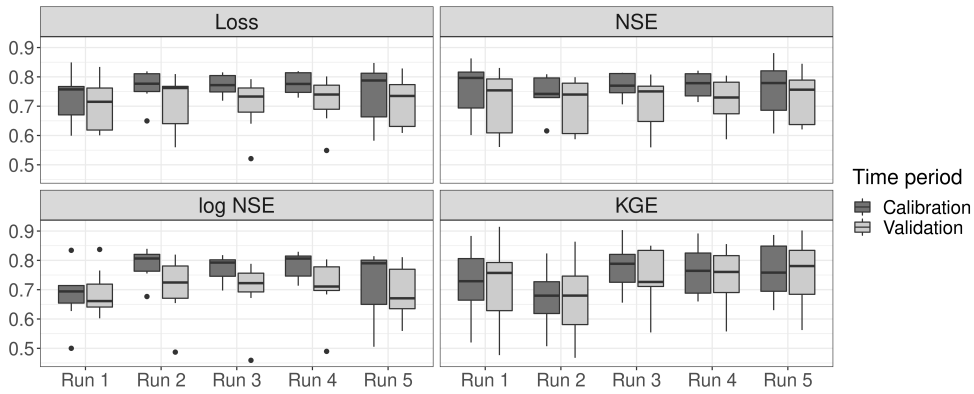


**Figure 3.** Boxplots of different performance metrics of the 5 FSO runs in the 7 training basins for the calibration (2000-2004) and validation (1965-1999) time periods.

Table 1 shows the FSO estimated TFs for all runs. To ease readability, they do not include the scaling factors for variables and TFs which are applied to compute the parameter values. Hence, it is difficult to estimate the value range that results from these TFs solely from the given function, but it shows the complexity, non-linearity, and used physiographic properties.

The estimated TFs show different lengths and levels of complexity. There is no specific physiographic characteristic, which is part of all TFs for either of these two parameters. Even in the TFs of run 2-4, which are very similar in performance, no variable was used in every run. The FSO VAE is constrained to not produce constant functions. In-

terestingly, this was bypassed by generating a function that includes $\nu_4 - \nu_4$, which results in a constant value for the run 4 $FieldCap$ TF.

**Table 1.** FSO estimated TFs for the mHM parameter $K_S$ and $FieldCap$ for the optimization runs 1-5 and their corresponding value ranges. Some functions were simplified to ease readability and thus do not necessarily reflect the direct VAE output. All physiographic properties ($\nu_{1-6}$) and mHM parameters ($K_S$, $\nu_7$) are scaled to [0,1] before being used in these TFs. The values resulting from these TFs are scaled to the parameter value range to generate the final parameter sets for the model. The physiographic properties are sand content ($\nu_1$), clay content ($\nu_2$), bulk density ($\nu_3$), aspect ($\nu_4$), slope ($\nu_5$), elevation ($\nu_6$), saturated hydraulic conductivity ($K_S$) and ThetaS ($\nu_7$).

| Run | $K_S$ (cm/day) | value range | $FieldCap$ (-) | value range |
|---|---|---|---|---|
| 1 | $\nu_6 - 3.009(\nu_3 + \cosh(\nu_5))$ | [1.1, 243.7] | $\frac{\cos(\nu_2)}{\nu_3} + \tanh(\nu_5)$ | [0.250, 0.299] |
| 2 | $\nu_4 - 3.223\nu_3 - 2.72$ | [1.1, 67.6] | $\cos(\cosh(\nu_6)^2)^{2.816}$ | [0.222, 0.228] |
| 3 | $-3.182 - \cosh(\nu_3)$ | [105.4, 117.1] | $K_S - (2.682 + \nu_7 \cosh(\nu_3))$ | [0.100, 0.139] |
| 4 | $log10(\nu_4) - 3.167$ | [1.1, 177] | $\nu_4 - \nu_4 - 3.286$ | [0.100, 0.100] |
| 5 | $\frac{0.101}{(\nu_2 \arcsin(\nu_1) - 1.0))} - (\nu_2 + cos(\nu_6))$ | [16.2, 102.9] | $\nu_5$ | [0.222, 0.311] |

### 4.2 Sampled validation

The run used for the benchmarking test with Zink et al. (2017) is based on validation performance in 20 sampled basins shown in Figure 4. The previous results already showed that there is a distinct difference in performance in run 1 and 5 compared to run 2-4. This is again visible in the results of the sampled validation, especially in the PBIAS values. Run 1 and 5 have large positive median PBIAS values of 22.3% and 27.1%, indicating model overestimation bias. On the other hand, run 2-4 simulations have low median PBIAS values of 5.6%, 4.0%, and 5.7%, respectively.

Run 2 simulations result in the highest NSE values with a median of 0.63. Median log NSE values of runs 2-4 are again very similar with values of 0.55, 0.60, and 0.57, with the main difference being that run 2 is the only run that does not include outliers. Median KGE values of runs 2-4 are also very similar with values of 0.61, 0.60, and 0.62. With the highest NSE, no outliers in the log NSE, and equally high KGE as the other runs, run 2 was chosen as the best model run that will be compared to the benchmark in all 222 validation basins.

### 4.3 Validation and benchmark evaluation

To assess the performance of FSO, the FSO run 2 NSE results are compared to the minimum, maximum, 5% quantile, 95% quantile, and median NSE of the 100 parameter sets applied by Zink et al. (2017) for the 222 validation Basins. Figure 5 shows the resulting violin plots and boxplots of NSE values. The NSE medians are almost equal for both experiments (run 2 = 0.67, Zink et al. (2017) = 0.68). The main difference can be seen in the NSE variance which is lower in the FSO run 2 results ($\sigma_{run2} = 0.008$, $\sigma_{Zinketal.(2017)} = 0.015$) and the numbers of outliers of the Zink et al. (2017) simulations. Testing the differences between these two NSE value distributions using a Kruskal-Wallis test (Kruskal & Wallis, 1952) did not show a significant difference ($p{-}value = 0.457$). While the run 2 results are the NSE values of one specific parameter field, the Zink et al. (2017) NSE values are the results of 100 different parameter sets. Thereby,

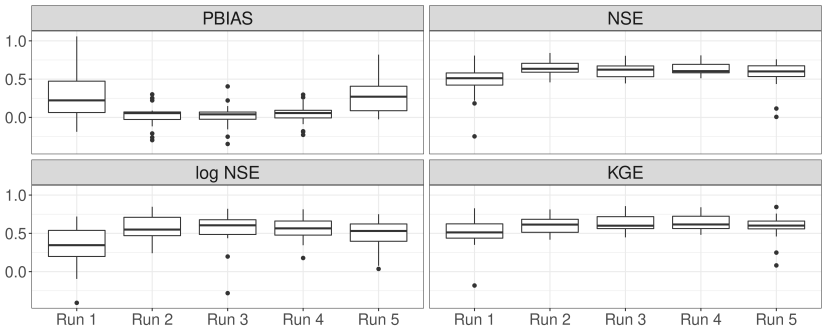


**Figure 4.** Boxplots of different performance metrics of the 5 FSO runs in 20 randomly sampled validation basins. The sampled validation basins were randomly drawn from a stratified Budyko curve and thus represent the full range of climates in the study region.

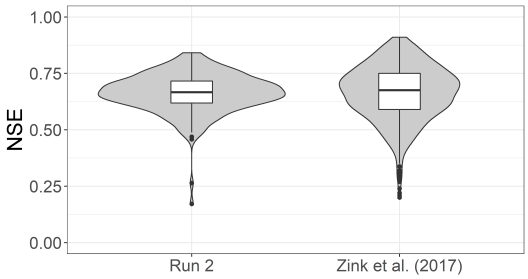


**Figure 5.** Violinplots and boxplots of NSE values of the 222 validation basins for FSO run 2 and the Zink et al. (2017) values. The Zink et al. (2017) values consist of the resulting minimum, maximum, 5% quantile, 95% quantile and median NSE of their applied 100 parameter sets.

they represent the overall behavior of the original mHM TFs, but individual points can not directly be compared to the FSO results.

Figure 6 shows the spatial NSE patterns of FSO run 2 and median Zink et al. (2017) results. The benchmark median NSE seems to be slightly higher in the northernmost basins, while FSO run 2 shows higher NSE values in western and central Germany. Other than that, no distinct spatial pattern can be observed. The median NSE of basins that lie inside the calibration basins is 0.68 and nearly equal to 0.66 for basins outside. The same is true for Zink et al. (2017) results with a median NSE inside the calibration basins of 0.68 and outside of 0.67.

To further assess the FSO performance, the relationship between resulting NSE values and the validation basins climate, basin area and mean altitude is analyzed. For this task, the climate is represented by the basins' Aridity index (PET/P). Testing for dependency using a linear regression shows a significant positive association of NSE with the basin area (coefficient $= 2.155 \times 10^{-5}$, p-value $< 10^{-7}$) and a negative association of NSE with the mean altitude (coefficient $= -1.2 \times 10^{-4}$, p-value $= 0.006$), but no existing association of NSE with aridity (p-value $= 0.113$).

### 4.4 Comparison of parameter fields

Finally, the FSO generated parameter fields are examined and compared to the default mHM parameter fields. In addition to the best performing run (run 2), we will ex-

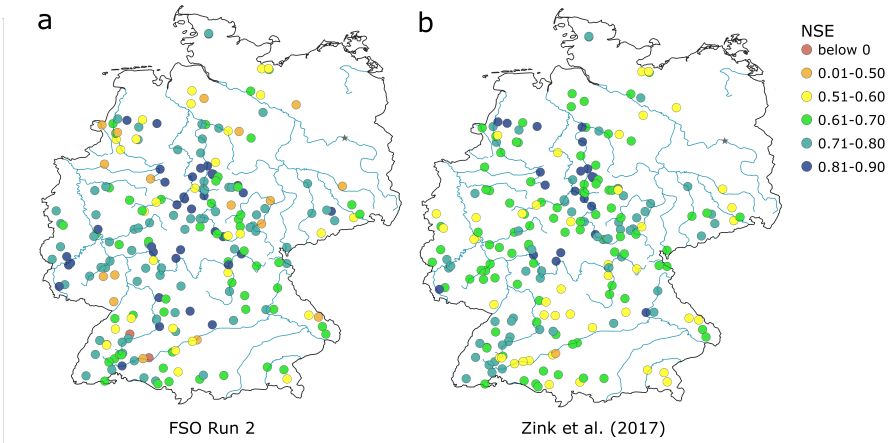


**Figure 6.** Maps of NSE for the 222 validation basins in Germany. **a** FSO run 2 results. **b** Zink et al. (2017) median results.

amine the resulting parameter fields of run 3 and 4, which produced nearly equally good simulations. Figure 7a shows the $K_S$ [cm/day] parameter distributions and Figure 7b the corresponding parameter fields. These parameter fields have very different characteristics. Only the spatial patterns of run 2 and the default mHM parameter is somewhat similar, however, run 2 does not have areas with high $K_S$ values ($> 150 cm/day$). The TF of run 3 produces nearly constant $K_S$ values of around 110 cm/day, while the run 4 TF results in values mostly between 150-200 cm/day with isolated lower outliers distributed over the whole area.

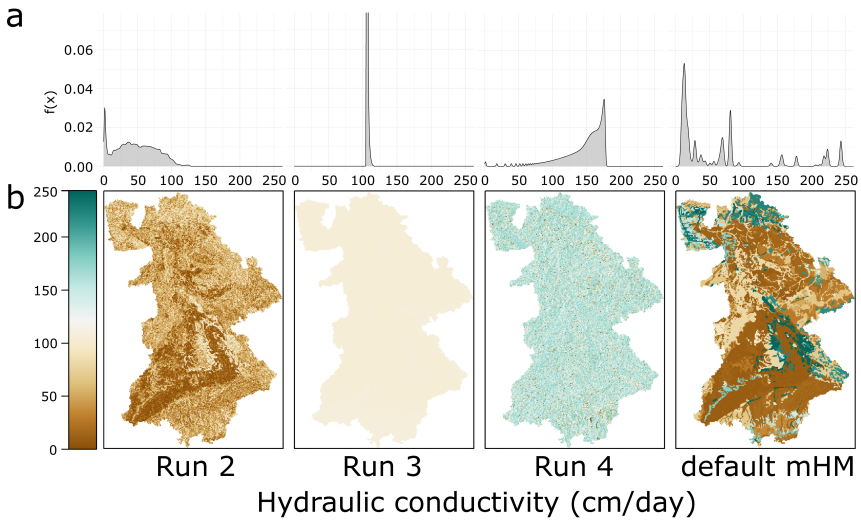


**Figure 7.** Resulting parameter values for the mHM parameter $K_S$ (saturated hydraulic conductivity, cm/day) for three FSO runs and the default mHM parameter set for the 7 calibration basins. **a** Parameter distributions of FSO run 2-4 and mHM default parameter and **b** parameter fields of the 7 calibration basins.

Figure 8a shows the $FieldCap$ [-] parameter distributions and Figure 8b the corresponding parameter fields. The TFs of two of the FSO runs, run 2 and run 4, predict a constant value: $FieldCap_{run2} = 0.222$, $FieldCap_{run4} = 0.100$. Run 3 values show more variability with values in the range of [0.10, 0.14]. Default mHM values have a median of 0.20 are thus generally higher than run 3 and run 4 FSO values, but lower than run 2 values. The default mHM TF for $FieldCap$ results in values with a much higher variance compared to all FSO estimated $FieldCap$ TFs.

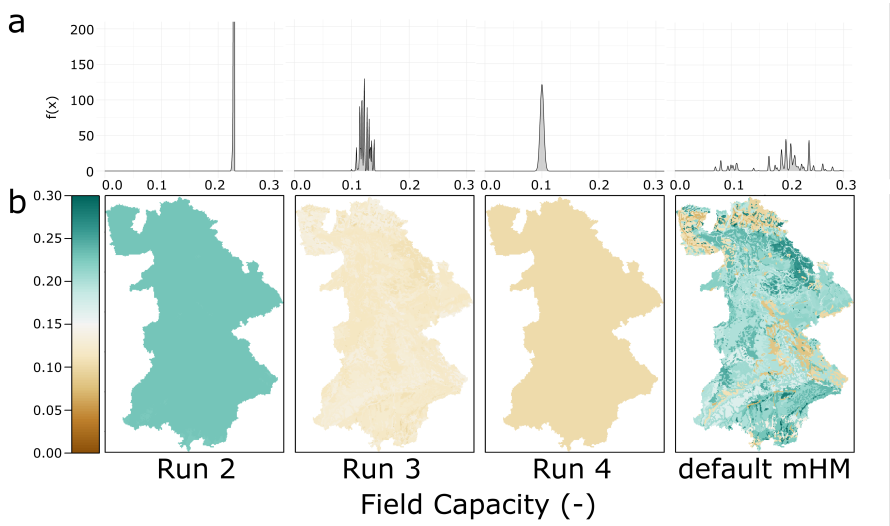


**Figure 8.** Resulting parameter values for the mHM parameter $FieldCap$ (field capacity, -) for three FSO runs and the default mHM parameter set for the 7 calibration basins. **a** Parameter distributions of FSO run 2-4 and mHM default and **b** parameter fields of the 7 calibration basins.

## 5 Discussion

Estimating distributed parameter fields for hydrological models with TFs is potentially able to significantly reduce the number of free model parameters, make the model transferable across time and space and produce physically meaningful parameters and model states. The presented results show that with the current FSO approach we are able to induce the first two properties in the model. Using the FSO estimated TFs greatly reduces the complexity of optimization compared to a model whose parameters are estimated for each pixel or each spatial modeling unit. Furthermore, the model had only a slight reduction in performance when applied in validation basins. This is interesting and notable as it represents a prediction in ungauged basins (PUB) problem that is inherently difficult for hydrological models to solve. Both these goals are reached by using only 5 years of data for training. However, these years include a year with high-impact flood events in central Europe (2002) and a year with a significant drought event (2003). The findings regarding the third property will be discussed in a following paragraph.

Of the 5 FSO runs, 3 performed nearly equally well, while 2 runs had lower performance and already showed issues with convergence during optimization. One of these 5 runs used the DDS, which we already expected to have convergence issues. This shows that optimization in Function Space is difficult and could potentially still be improved by a VAE that is able to produce a smoother Function Space for optimization. The current version leads to similar performance results in three out of four times when using the SCE for optimization. Reducing the stopping criteria to 500 iterations without improvement would certainly result in discarding a run with convergence issues in an early

stage of the optimization. Comparison of the convergence behavior of FSO to other studies on large-scale parameter estimation is not possible, since they are neither reported for mHM (e.g., Rakovec et al., 2016; Dembélé et al., 2020) nor for other models (e.g., Mizukami et al., 2017; López López et al., 2017).

The FSO estimated TFs perform as well as the benchmark predictions by Zink et al. (2017) which is an astonishing result considering its implications. The TFs used by Zink et al. (2017) were developed over a span of 10 years using a large amount of literature and expert knowledge and were developed in some of the German basins that are part of this study (Samaniego et al., 2010; Kumar et al., 2013; Thober et al., 2019; Samaniego et al., 2019). Furthermore, the benchmark represents the results of the best 100 parameter sets after an overall of 1.4 million optimization iterations. Hence, this benchmark is more like an upper bound that we aim to reach with an automatic setting, rather than a threshold we certainly have to cross. The current state of FSO is able to estimate TFs that reach this upper bound and would allow for estimating TFs for other models without the work that was necessary for developing the mHM TFs. This will drastically reduce the time to implement TFs for a new model that is coupled with the FSO and MPR method.

Hydrological model performances are strongly dependent on the basins and the available data quality. Therefore, it is important to use a benchmark to adequately interpret the results of a study. For this reason, we chose the study of Zink et al. (2017), as it had the necessary scope and used the same model. The Zink et al. (2017) benchmark setup was the best possible available option but increased the difficulty of the task for two reasons. One reason was the fact that only quantiles of performance values were available from Zink et al. (2017). We do not know whether run 2 actually has a higher performance than each individual parameter set of Zink et al. (2017). The second reason can be found in the selection of calibration basins. Zink et al. (2017) used 7 very large basins for calibration because they only had to optimize a small set of numerical parameters of existing TFs. This is a much more constrained optimization problem compared to the estimation of structure and numerical coefficients of TFs. As each basin results in one loss value, the most pronounced feedback for optimization would be reached by using a set of basins that have a high variance of physiographic properties between them, but a low variance inside each basin. Hence, only a few extremely large basins with a high internal variance of physiographic properties are not the best starting point. Interestingly, FSO-mHM still performs as well as the median of 100 parameter sets with this non-ideal selection of calibration basins, pointing to the fact that there is still potential for improvement from a different calibration setup.

From the results, it is evident that physical interpretation of the parameter fields is still difficult at this point. This is indicated by the three FSO runs (run 2-4) having a very similar discharge simulation performance, while having very different parameter fields. Looking at different parameter values estimated by the FSO TFs, it is interesting that all of them produce constant values or nearly constant values for the field capacity parameter. Looking into the model states, we could observe that soil water content was usually above field capacity in the FSO models. This is equivalent to a model simplification, showing that the study region, which does not include arid basins, can be represented without using the parameter $FieldCap$. These parameter sets will most likely perform poorly in arid regions. Similarly, the predicted saturated hydraulic conductivity values were diverse, with only FSO run 2 showing similar patterns compared to the default mHM parameters. $K_S$, unlike $FieldCap$, is not directly used by mHM but is used to compute the model parameter $Kperco$ - the K factor for percolation that controls the amount of water that flows between soil layers. This conceptualization of the parameter is on one hand influencing its physical interpretation, which potentially differs from the definition of the saturated hydraulic conductivity, but on the other hand, potentially makes it easier to find a suitable relationship to the physiographic proper-

ties of a catchment. This is especially important as saturated hydraulic conductivities at larger scales are difficult to measure or estimate with Pedotransfer functions (Zhang & Schaap, 2019).

While we could have just chosen run 2 parameter fields for comparison, we included the slightly lower performing runs as well to show that parameter equifinality still exists when applying FSO. This may be primarily a result of using only large basins for calibration, which only give coarse feedback during optimization. In our opinion, this equifinality will most likely be strongly reduced if an appropriate set of calibration basins are selected. This was not possible in this study because we wanted to have a comparable benchmark. As mentioned above - ideally, these calibration basins should consist of a larger number of basins with a high variance of physiographic properties between them, but a low variance inside each basin. Still, further constraints using additional boundary conditions, e.g. soil moisture or ET-fluxes, are certainly helpful for predicting physically sound parameter values. Additionally, a wider range of physiographic properties used in FSO TFs would produce a larger search space and potentially better performing TFs. This could include inputs derived from existing ones, e.g. the Topmodel index $ln(\frac{a}{slope}$ (K. J. Beven & Kirkby, 1979), inputs that have shown relevance for other hydrological or soil science prediction tasks, e.g. the relevant inputs used in the SoilGrids regression trees (Poggio et al., 2021), or other gridded inputs that represent vegetation, soil and climatic properties of the catchments. Furthermore, for future studies, the number of TFs estimated with FSO should be increased because we could show that optimizing two TFs is feasible.

Recently, there have been two other studies that derived distributed model parameters from physiographic attributes which both applied the Variable Infiltration Capacity model (VIC, Liang et al., 1994) over the contiguous United States (COThetaS): Mizukami et al. (2017) and Tsai et al. (2021). Mizukami et al. (2017) used an MPR based approach (MPR-flex) which uses TFs chosen from literature. They concluded that TFs with global parameters lead to improve spatial fields, that there is still a large gap in performance between a global parameter set and individual basin calibration for the chosen TFs, and "though not trivial" different forms of TFs should be evaluated. Overall, Mizukami et al. (2017) shows the advantage of the MPR approach, while being limited by TFs from literature that were not developed for large-scale modeling systems. Especially since VIC uses multiple conceptual parameters, which limits the use of literature-based TFs, this study would have benefited from the FSO approach. Tsai et al. (2021) developed a deep learning approach for parameter estimation called differentiable parameter learning (dPL). dPL estimates parameters by optimizing a neural network that generates model parameters. To optimize this neural network, it is necessary to either have a fully differentiable hydrological model or to use a surrogate neural network instead of the hydrological model. They show that the dPL approach improves the discharge prediction results of Mizukami et al. (2017) from a median NSE of 0.32 to 0.44. Tsai et al. (2021) argue that dPL produces better generalizability and physical coherence of the derived parameters based on the fact that dPL uses dynamic and static catchment attributes as inputs and improves soil moisture simulation compared to a model optimization using calibration with the SCE-UA algorithm. dPL is dependent on a well-performing surrogate model, or on rewriting the existing model to make it differentiable, which is certainly not feasible for most hydrological models. By comparison, we demonstrate that FSO can identify TFs that can be applied in an ungauged setting without requiring a differentiable model. While MPR-flex and dPL were only tested with the VIC model in a gauged setting, they show the current state of approaches for deriving model parameters based on physiographic properties of catchment and highlight the complexity of this task.

This study has some limitations. First, while Zink et al. (2017) only used data from the calibration basins to derive the parameter sets, we also used 20 sampled validation basins to find the best FSO run. Since we applied FSO multiple times and were also in-

terested in the variance of FSO performance, comparing them with data outside the calibration basins was necessary. Nevertheless, these sampled validation results showed comparable performance to the calibration and consequently did not strongly influence our choice of the best run. Another limitation is the fact that we only use 5 FSO runs, which is due to very practical reasons: the resources of the cluster that we used for computation are limited. However, we do believe that the 5 runs provide a useful estimate of the TF variability. Another limitation is the fact that we compare the performance of one FSO derived parameter set with the median performance of 100 parameter sets derived by Zink et al. (2017). This certainly shows the general performance of the original mHM TFs, but will also lead to a reduction in variance in the performance of Zink et al. (2017) due to aggregation. Therefore, from the presented results we cannot conclude that there is one specific Zink et al. (2017) parameter set that performs equal, better, or worse than the FSO parameter set, but shows that their performance is comparable.

## 6 Summary and Conclusions

In this study, we presented the first large-scale application of FSO for automatic transfer function (TF) estimation of a complex distributed hydrological model. We assessed the performance variability of the FSO method by applying it 5 times, which resulted in 3 nearly equally well-performing sets of TFs and two with a slightly lower performance. The final selected TFs resulted in predictions in 222 validation basins with a median NSE of 0.68. The performance was equal to the median performance of 100 predictions of the benchmark study of Zink et al. (2017).

Overall, this study is a proof-of-concept where we showed that FSO is able to produce state-of-the-art results when applied to a complex distributed model, but more work is needed to derive physically meaningful parameter fields. We see some important aspects that have the potential to greatly improve TF estimation. First, this includes a careful selection of calibration basins, ideally with a wide range of physiographic characteristics but a low internal variance of these characteristics. Second, it is important to include further constraints during optimization in form of additional boundary conditions, e.g., simultaneously optimizing discharge and evapotranspiration, to further constrain the optimization and allow for physically sound parameter fields. Finally, an extension of available physiographic properties available for FSO will potentially allow finding a better representation for a larger number of model parameters.

The multiscale parameter regionalization technique (MPR), which uses TFs, was described as a promising way forward for global hydrological and Land Surface models. With FSO we now have a method that can automatically estimate these TFs for any model, which will make it possible to apply global hyperresolution models "everywhere" in the future.

## Appendix A FSO VAE architecture

The new encoder network of the VAE consists of a word embedding layer (Mikolov et al., 2013), bidirectional long short-term memory (LSTM) layers (Hochreiter & Schmidhuber, 1997), highway layers (Srivastava et al., 2015) using feedforward neural networks (FNN) (White & Rosenblatt, 1963) and the SELU (scaled exponential unit) activation function (Klambauer et al., 2017). The new VAE decoder consists of Temporal Convolutional Network (TCN) layers (Bai et al., 2018) and a FNN using a softmax activation function. The function space is chosen to be 6-dimensional and normal distributed. A detailed depiction of the VAE architecture is shown in the A1.

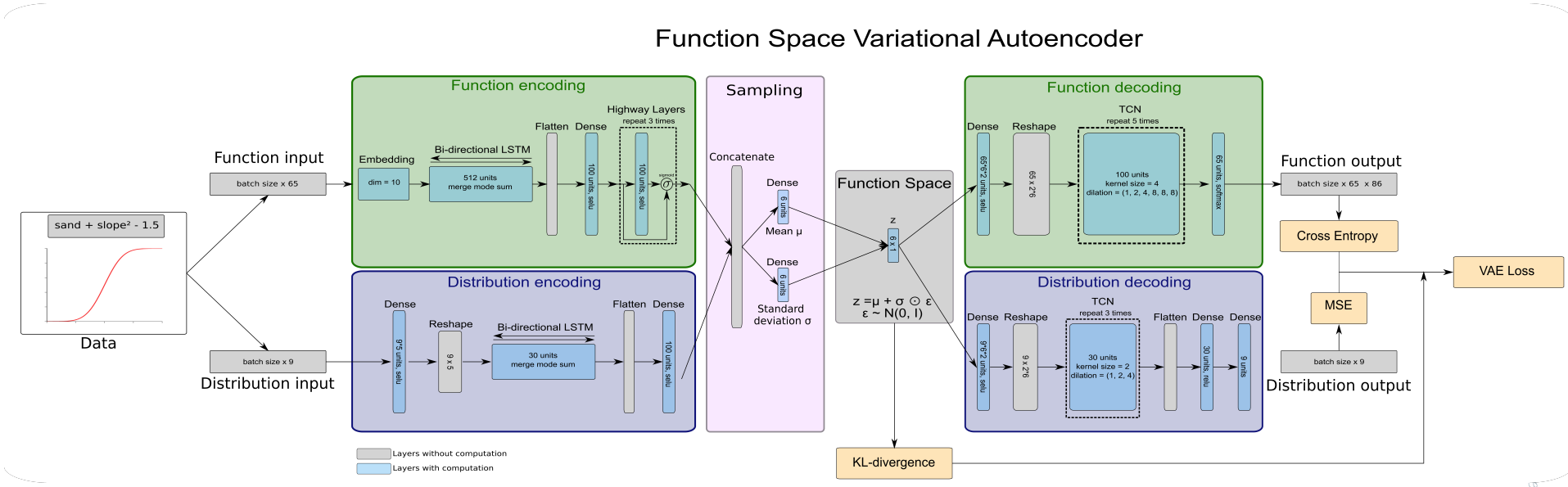


**Figure A1.** Function Space variational autoencoder architecture. Starting from the left: the data is separated into the function strings (Function input) and function quantiles (Distribution inputs). Each input has its own encoder network (Function encoding, Distribution encoding). The encoded ionputs are concatenated and used as basis for the mean and standard deviation sampling in the Sampling layer. These are then used to generate the normal distributed Function Space. The Function Space (6 dim vector) is used by the Function decoding and Distribution decoding networks to generate the Function and Distribution outputs, which are reconstructions of the inputs. The loss is computed with the sum of KL-divergence of the Function Space to a standard normal distribution, the cross-entropy of function string reconstruction and the MSE of function quantile reconstruction.

# Appendix B FSO setup details

## B1 Context free grammar (CFG) used to generate training data:

$$
\begin{aligned}
\textbf{eq} &= \textbf{eq op eq} \mid \textbf{eq op numeric} \mid \textbf{eq op var} \mid \textbf{eq op (eq)} \mid \textbf{var} \mid \\
&\quad \textbf{f(var)} \mid \textbf{f(eq)} \mid \textbf{(eq)}\hat{}\textbf{(pm numeric)} \mid \textbf{numeric} \\
\textbf{op} &= + \mid - \mid * \mid / \\
\textbf{pm} &= + \mid - \\
\textbf{f} &= exp \mid log10 \mid log \mid sin \mid cos \mid tan \mid asin \mid acos \mid atan \mid tanh \mid \\
&\quad cosh \mid sinh \mid sqrt \mid abs \\
\textbf{var} &= bd \mid sand \mid clay \mid slope \mid aspect \mid dem \mid ThetaS \mid K_S \mid vGenu_n \\
\textbf{numeric} &= 0.05 \mid 0.1 \mid ... \mid 2.95 \mid 3.0
\end{aligned}
$$

## B2 Scaling bounds:

The parameter bounds for the two estimated parameter were chosen to be $K_S = [1.1, 1000]$ and $FieldCap = [0.01, 0.55]$. The scaling bounds for the physiographic catchment properties is shown in table B1.

**Table B1.** Scaling bounds for scaling physiographic catchment properties and mHM parameters to [0,1].

| Physiographic property | Bounds |
|---|---|
| *slope* | [0, 90] |
| *aspect* | [0, 360] |
| *bd* | [0, 2.3] |
| *sand* | [0, 100] |
| *clay* | [0, 100] |
| *dem* | [0, 4000] |
| $vGenu_n^*$ | [1, 2] |
| $ThetaS^*$ | [0.24, 0.51] |

*Estimated from default mHM parameter values.

The bounds for scaling TF outputs to the parameter range are automatically estimated using the FSO VAE training data. The lower bound is the *median* 10% *quantile*+ 3×*mad*(10% *quantile*) and the upper bound is the *median* 90% *quantile*+3×*mad*(90% *quantile*), where *mad* denotes the median absolute deviation. Resulting in the ranges $K_S = [-5.606, 8.481]$ and $FieldCap = [-5.498, 8.50]$.

## B3 Hyperparameter of optimization algorithms:

The used DDS and SCE algorithms are our own R implementation and the GA algorithm was taken from the *GA* R package (Scrucca, 2013).

- DDS: $r = 0.2$, $iterations = 5000$
- SCE: $ncomplex = 5$, $pointscomplex = 50$, $pointssimplex = 10$, $cce.iter = 5$, $elitism = 1$, $initsample = "latin"$, $iterations = 5000$
- GA: $pop\_size = 10$, $iterations = 100$

## Appendix C Performance metrics overview

$$NSE(Q_{obs}, Q_{sim}) = 1 - \frac{\sum_t Q^t_{sim} - Q^t_{obs}}{\sum_t Q^t_{obs} - \bar{Q}_{obs}} \tag{C1}$$

$$logNSE(Q_{obs}, Q_{sim}) = 1 - \frac{\sum_t \log Q^t_{sim} - \log Q^t_{obs}}{\sum_t \log Q^t_{obs} - \log \bar{Q}_{obs}} \tag{C2}$$

$$\begin{aligned} KGE(Q_{obs}, Q_{sim}) &= 1 - ED \\ ED &= \sqrt{((r-1))^2 + ((\alpha-1))^2 + ((\beta-1))^2} \\ r &= \text{Pearson correlation coefficient} \\ \alpha &= \sigma_{sim}/\sigma_{obs} \\ \beta &= \mu_{sim}/\mu_{obs} \end{aligned} \tag{C3}$$

$$PBIAS(Q_{obs}, Q_{sim}) = 100 \times (\sum_t Q^t_{sim} - Q^t_{obs}) / \sum_t Q^t_{obs} \tag{C4}$$

## Appendix D Optimization progress

Figure D1 shows the progression during optimization of all 5 FSO runs. The DDS optimizer (run 5) converges rapidly to a comparatively low performance, similar to what we have already observed in our preliminary tests. The SCE optimizer (run 1-4) seems to have a much more continuous improvement, except for run 1, which shows a similar behaviour to the DDS run. These optimization performances suggest that discarding run 1 and run 5 would be a reasonable choice.

**Acknowledgments**

This work was funded by the Austrian Science Fund FWF, project P 31213. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).

The study data that was used in this study was previously made freely accessible by Zink et al. (2017) under Creative Commons license at `www.ufz.de/index.php?en=41160`. The mHM is available in Samaniego et al. (2019). The scripts used to produce all results of this study are available under Feigl et al. (2022), which also use functions from Feigl (2022).

## References

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., ... Over, T. (2015). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, *51*(12), 10078–10091. doi: 10.1002/2015WR017498

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv*. Retrieved from `http://arxiv.org/abs/1803.01271`

Bergström, S. (1995). The HBV model. *Computer models of watershed hydrology.*, 443–476.

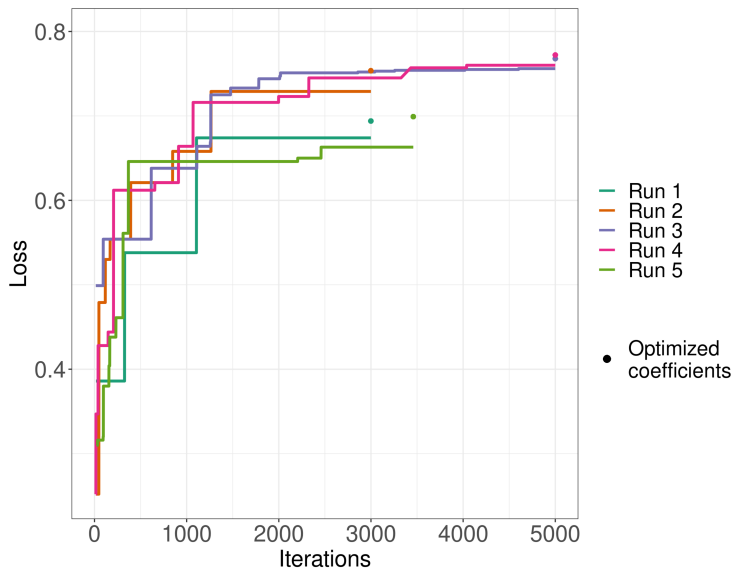


**Figure D1.** Overview of loss per iteration of all FSO optimization runs during calibration. Only values above 0 are shown to increase readability. The optimization was applied for a minimum of 2000 iterations and for a maximum of 5000 iterations. It is stopped if after the minimum required iterations there is no further improvement for 1000 iterations. After reaching one of these stopping criteria, only the numeric coefficients of the TFs were optimized for 100 iterations, leading to additional improvement at the end of each run. This additional improvement is depicted as a point at the end of each run.

Beven, K. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes*, *16*(2), 189–206. doi: 10.1002/HYP.343

Beven, K. J., & Cloke, H. L. (2012). Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al. *Water Resources Research*, *48*(1). doi: 10.1029/2011WR010982

Beven, K. J., & Franks, S. W. (1999). Functional similarity in landscape scale SVAT modelling. *Hydrology and Earth System Sciences*, *3*(1), 85–93. doi: 10.5194/hess-3-85-1999

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. doi: 10.1080/02626667909491834

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, *51*(7), 4923–4947. doi: 10.1002/2015wr017173

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., ... Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. doi: 10.1080/02626667.2019.1620507

Budyko, M. I. (1974). *Climate and life*. Academic Press. Retrieved from `https://agris.fao.org/agris-search/search.do?recordID=US201300514816`

Buytaert, W., & Beven, K. (2009). Regionalization as a learning process. *Water Resources Research*, *45*(11), 11419. doi: 10.1029/2008WR007359

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., ... Peters-Lidard, C. D. (2017). The evolution of process-based

hydrologic models: historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. doi: 10.5194/hess-21-3427-2017

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., ... Ceola, S. (2016, mar). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, *52*(3), 2350–2365. doi: 10.1002/2015WR017910@10.1002/(ISSN)1944-9208.COMHES1

Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resources Research*, *20*(6), 682–690. doi: 10.1029/WR020i006p00682

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., ... Samaniego, L. (2015, aug). Computationally inexpensive identification of noninformative model parameters by sequential screening. *Water Resources Research*, *51*(8), 6417–6441. Retrieved from `http://doi.wiley.com/10.1002/2015WR016907` doi: 10.1002/2015WR016907

Dembélé, M., Hrachowitz, M., Savenije, H. H., Mariéthoz, G., & Schaefli, B. (2020, jan). Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. *Water Resources Research*, *56*(1), e2019WR026085. doi: 10.1029/2019WR026085

Dooge, J. C. I. (1982). Parameterization of hydrologic processes. *Land surface processes in atmospheric general circulation models*, 243–288.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. doi: 10.1029/91WR02985

Feigl, M. (2022). *MoritzFeigl/FSO: v0.1.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.5833794` doi: 10.5281/zenodo.5833794

Feigl, M., Herrnegger, M., Klotz, D., & Schulz, K. (2020). Function Space Optimization: A Symbolic Regression Method for Estimating Parameter Transfer Functions for Hydrological Models. *Water Resources Research*, *56*(10). doi: 10.1029/2020WR027385

Feigl, M., Herrnegger, M., Schweppe, R., Thober, S., Klotz, D., Samaniego, L., & Schulz, K. (2021). Regionalisierung hydrologischer Modelle mit Function Space Optimization. *Österreichische Wasser- und Abfallwirtschaft 2021 73:7*, *73*(7), 281–294. doi: 10.1007/S00506-021-00766-0

Feigl, M., Schweppe, R., & Thober, S. (2022). *MoritzFeigl/FSO_mHM: Submission.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.5833785` doi: 10.5281/zenodo.5833785

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*(1-2), 80–91. doi: 10.1016/j.jhydrol.2009.08.003

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Holland, J. (1975). Adaptation in natural and artificial systems. an introductory analysis with applications to biology, control and artificial intelligence. *Ann Arbor: University of Michigan Press, 1975*.

Höllering, S., Wienhöfer, J., Ihringer, J., Samaniego, L., & Zehe, E. (2018). Regional analysis of parameter sensitivity for simulation of streamflow and hydrological fingerprints. *Hydrol. Earth Syst. Sci*, *22*, 203–220. doi: 10.5194/hess-22-203-2018

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., ... Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, *58*(6), 1198–1255. doi: 10.1080/02626667.2013.803183

Hundecha, Y., & Bárdossy, A. (2004). 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*(1-4), 281–295. doi: 10.1016/J.JHYDROL.2004.01.002

Imhoff, R. O., van Verseveld, W. J., van Osnabrugge, B., & Weerts, A. H. (2020). Scaling Point-Scale (Pedo)transfer Functions to Seamless Large-Domain Parameter Estimates for High-Resolution Distributed Hydrologic Modeling: An Example for the Rhine River. *Water Resources Research*, *56*(4), e2019WR026807. doi: 10.1029/2019WR026807

Jing, M., Kumar, R., Heße, F., Thober, S., Rakovec, O., Samaniego, L., & Attinger, S. (2020). Assessing the response of groundwater quantity and travel time distribution to 1.5, 2, and 3thinsp;°C global warming in a mesoscale central German basin. *Hydrology and Earth System Sciences*, *24*(3), 1511–1526. doi: 10.5194/HESS-24-1511-2020

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. Retrieved from `http://arxiv.org/abs/1312.6114`

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-Normalizing Neural Networks. Retrieved from `http://arxiv.org/abs/1706.02515`

Klotz, D. (2020). *Systematic estimation of transferfunctions for the parameterization of spatially distributed rainfall-runoff models.* Retrieved from `https://permalink.obvsg.at/AC16121579`

Klotz, D., Herrnegger, M., & Schulz, K. (2017). Symbolic Regression for the Estimation of Transfer Functions of Hydrological Models. *Water Resources Research*, *53*(11), 9402–9423. doi: 10.1002/2017WR021253

Knuth, D. E. (1965). On the translation of languages from left to right. *Information and Control*, *8*(6), 607–639. doi: 10.1016/S0019-9958(65)90426-2

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. doi: 10.1080/01621459.1952.10483441

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*(1), 360–379. doi: 10.1029/2012WR012195

Leavesley, G. H., Lichty, R. W., Troutman, B. M., & Saindon, L. G. (1983). Precipitation-runoff modeling system: User's manual. *Water-resources investigations report*, *83*(4238), 207.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994, jul). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428. doi: 10.1029/94JD00483

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrol. Earth Syst. Sci*, *21*, 3125–3144. doi: 10.5194/hess-21-3125-2017

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Retrieved from `http://arxiv.org/abs/1301.3781`

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., ... Samaniego, L. (2017, sep). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040. doi: 10.1002/2017WR020401

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. doi: 10.1016/0022-1694(70)90255-6

Peichl, M., Thober, S., Meyer, V., & Samaniego, L. (2018). The effect of soil moisture anomalies on maize yield in Germany. *Natural Hazards and Earth System*

*Sciences*, *18*(3), 889–906. doi: 10.5194/NHESS-18-889-2018

Pinnington, E., Amezcua, J., Cooper, E., Dadson, S., Ellis, R., Peng, J., ... Quaife, T. (2021). Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. *Hydrology and Earth System Sciences*, *25*(3), 1617–1641. doi: 10.5194/HESS-25-1617-2021

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B., Kempen, B., Ribeiro, E., & Rossiter, D. (2021, jun). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. doi: 10.5194/SOIL-7-217-2021

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*(10), 7779–7792. doi: 10.1002/2016WR019430

Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., ... Samaniego, L. (2019). Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13991–14007. doi: 10.1029/2019JD030767

Saha, T. R., Shrestha, P. K., Rakovec, O., Thober, S., & Samaniego, L. (2021). A drought monitoring tool for South Asia. *Environmental Research Letters*, *16*(5), 054014. doi: 10.1088/1748-9326/ABF525

Samaniego, L., & Bárdossy, A. (2005). Robust parametric models of runoff characteristics at the mesoscale. *Journal of Hydrology*, *303*(1-4), 136–151. doi: 10.1016/j.jhydrol.2004.08.022

Samaniego, L., Kaluza, M., Kumar, R., Rakovec, O., Schüler, L., Schweppe, R., ... Attinger, S. (2019). mesoscale Hydrologic Model. doi: 10.5281/ZENODO.3239055

Samaniego, L., Kumar, R., & Attinger, S. (2010, may). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5). Retrieved from `http://doi.wiley.com/10.1029/2008WR007327` doi: 10.1029/2008WR007327

Schweppe, R., Thober, S., Kelbling, M., Kumar, R., Attinger, S., & Samaniego, L. (2021). MPR 1.0: A stand-alone Multiscale Parameter Regionalization Tool for Improved Parameter Estimation of Land Surface Models. *Geoscientific Model Development Discussions*, *2021*, 1–40. Retrieved from `https://gmd.copernicus.org/preprints/gmd-2021-103/` doi: 10.5194/gmd-2021-103

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, *53*(4), 1–37. doi: 10.18637/JSS.V053.I04

Sorooshian, S., Duan, Q., & Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, *29*(4), 1185–1194. doi: 10.1029/92WR02617

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway Networks. Retrieved from `https://arxiv.org/abs/1505.00387v2`

Thober, S., Cuntz, M., Kelbling, M., Kumar, R., Mai, J., & Samaniego, L. (2019). The multiscale routing model mRM v1.0: Simple river routing at resolutions from 1 to 50 km. *Geoscientific Model Development*, *12*(6), 2501–2521. doi: 10.5194/gmd-12-2501-2019

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., ... Zink, M. (2018). Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming. *Environmental Research Letters*, *13*(1), 014003. doi: 10.1088/1748-9326/AA9E35

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1). doi: 10.1029/2005WR004723

Tsai, W. P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., ... Shen, C. (2021, oct). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications 2021 12:1*, *12*(1), 1–13. doi: 10.1038/s41467-021-26107-z

Twarakavi, N. K. C., Sakai, M., & Šimůnek, J. (2009). An objective analysis of the dynamic nature of field capacity. *Water Resources Research*, *45*(10). doi: 10.1029/2009WR007944

van Genuchten, M. T. (1980). A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, *44*(5), 892–898. doi: 10.2136/sssaj1980.03615995004400050002x

White, B. W., & Rosenblatt, F. (1963). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Vol. 76) (No. 4). doi: 10.2307/1419730

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., ... Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, *47*(5). doi: 10.1029/2010WR010090

Zhang, Y., & Schaap, M. G. (2019, aug). Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. *Journal of Hydrology*, *575*, 1011–1030. doi: 10.1016/J.JHYDROL.2019.05.058

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, *21*(3), 1769–1790. doi: 10.5194/hess-21-1769-2017