AUTOMATIC ESTIMATION OF PARAMETER TRANSFER FUNCTIONS FOR DISTRIBUTED HYDROLOGICAL MODELS

A CASE STUDY WITH THE mHM MODEL

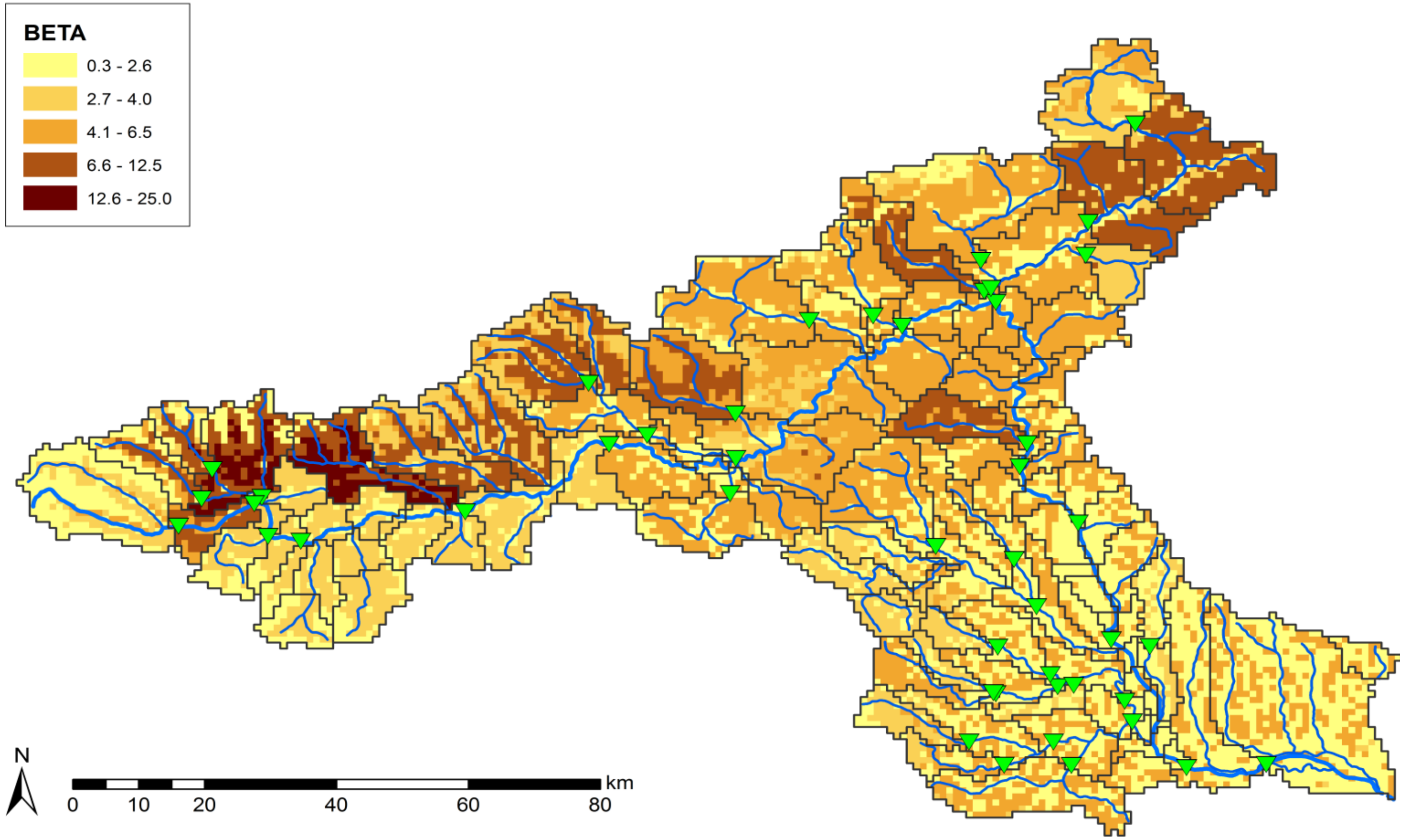
M. Feigl[1], S. Thober[2], R. Schweppe[2], M. Herrnegger[1], L. Samaniego[2] and K. Schulz[1]

[1]Institute for Hydrology and Water Management (HyWa), University of Natural Resources and Life Sciences, Vienna, Austria

[2]Department of Computational Hydrosystems, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany

Parameter estimation for distributed models

In a typical water resource managment problem, we are faced with:



Catchment size ≈ 10 000 km²
Model resolution ≈ 1 km²
Model parameters ≈ 10–50

Parameters to optimize ≈ 100 000 – 500 000

→ equifinality
→ “ill-posed problem”
→ reduced physical meaning of parameters

Parameter estimation for distributed models

In summary:

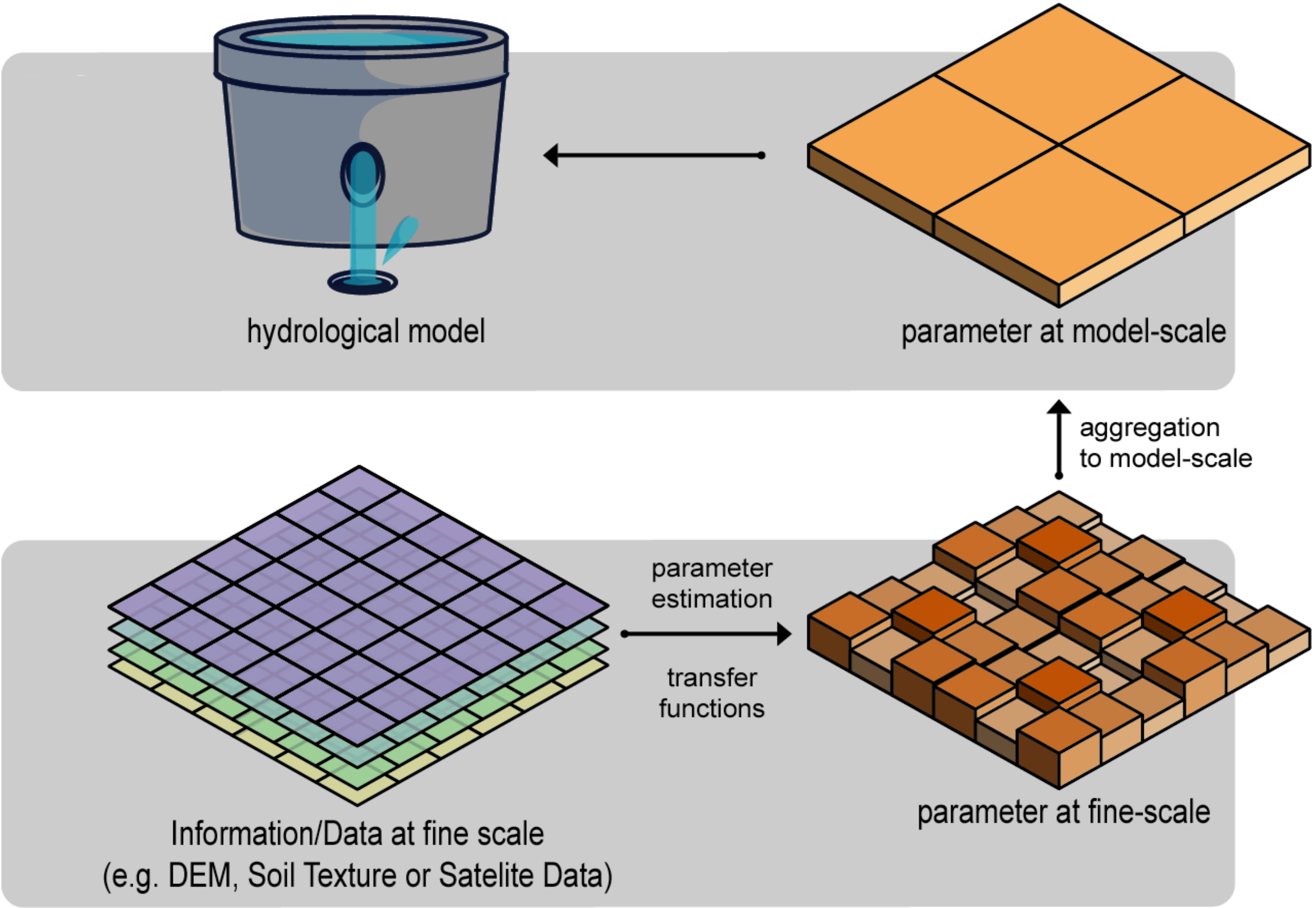
- Number of parameters to optimize for distributed models is very large:
  - → Difficult to optimize
  - → Time-consuming & computational exhaustive
- Optimized parameters are not transferable to other locations
- Often results in patch-work like parameter fields
- Reduced physical meaning of parameters due to parameter equifinality

Possible solution:

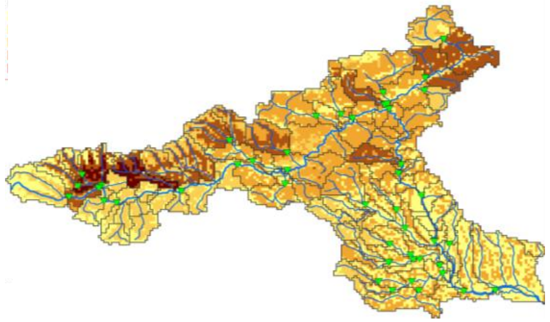
Define model parameters with relationship to catchment characteristics, e.g. soil, vegetation, and geological properties.

→ Multiscale Parameter Regionalization (MPR)

Multiscale Parameter Regionalization (MPR)



=

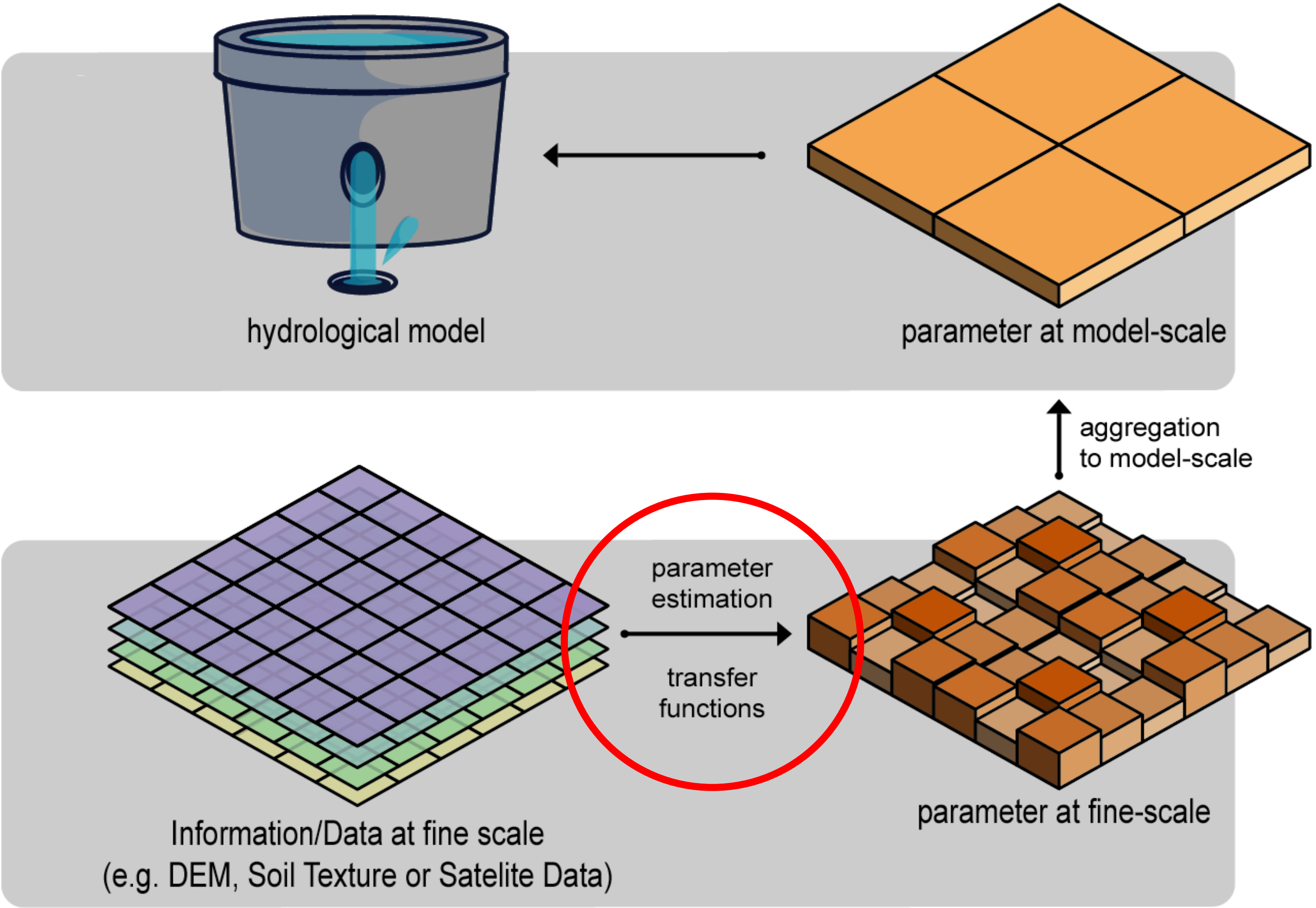


MPR utilizes observable basin physical characteristics (e.g. soil, vegetation, and geological properties) to infer the spatial variability of the required parameters.

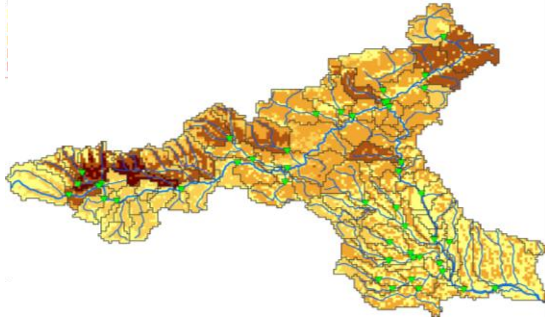
A regionalization method developed by [Samaniego et al. (2010)]

→ available as [stand-alone software package]

Multiscale Parameter Regionalization (MPR)



=



→ Main difficulty for application: Estimation of the structure of transfer functions

A regionalization method developed by [Samaniego et al. (2010)]

→ available as [stand-alone software package]

Automatic estimation of transfer functions

Klotz et al. (2017) were the first to investigate a symbolic regression approach to automatically estimate transfer functions. By using a simple model and synthetic data, Klotz et al. (2017) showed that it is possible to automatically estimate transfer functions from stream data in a virtual setting.

The term symbolic regression refers to methods that search the space of mathematical expressions while minimizing some error metrics.

Recently, we introduced a new approach for the automatic estimation of parameter transfer function, named Function Space Optimization (FSO, Feigl et al., 2020).

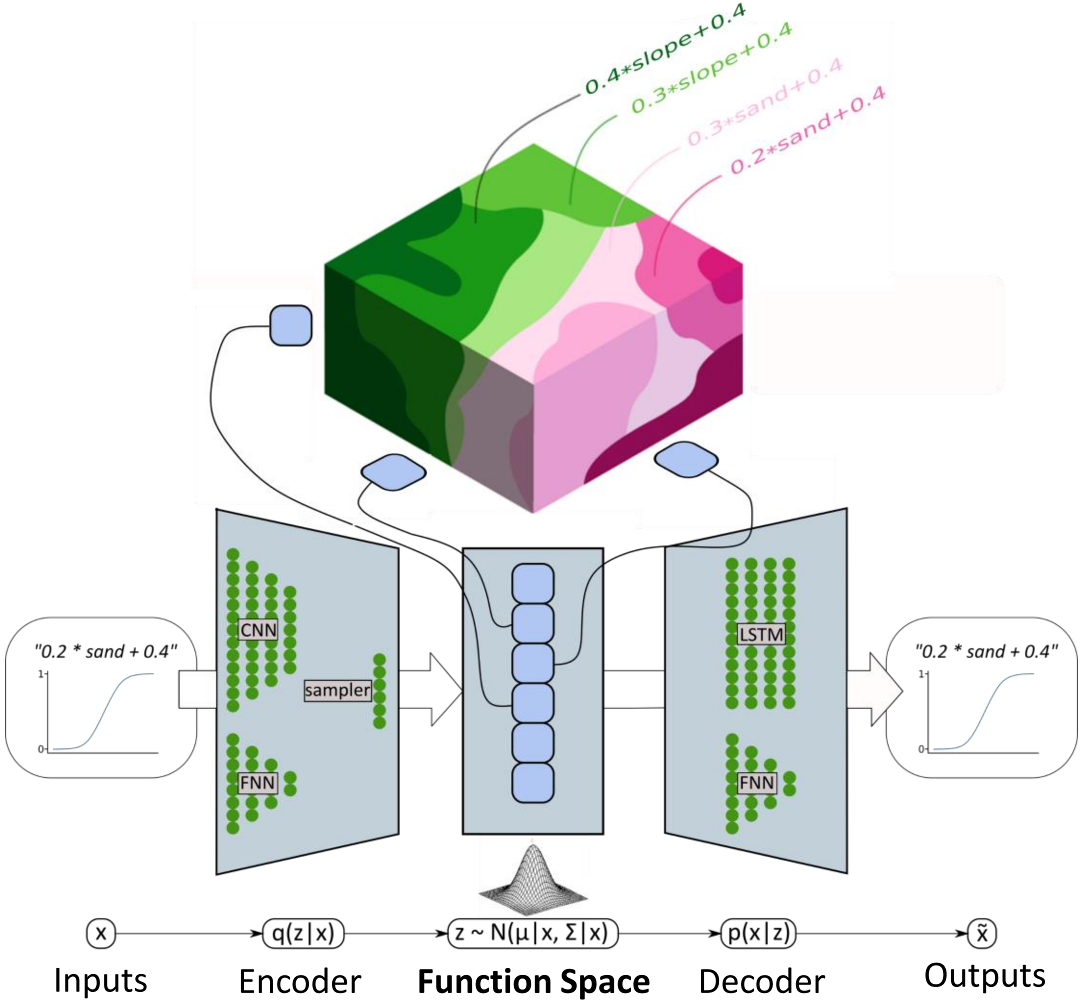
Function Space Optimization (FSO)

- FSO is a symbolic regression optimization method for estimating parameter transfer functions for distributed hydrologic models.

- FSO uses a variational autoencoder (VAE) to transfer the search of a best fitting transfer function into a continuous numerical vector space (Function Space).

- So far, FSO has only been applied in a synthetic setting which showed its potential ability to solve the problem of transfer function estimation.

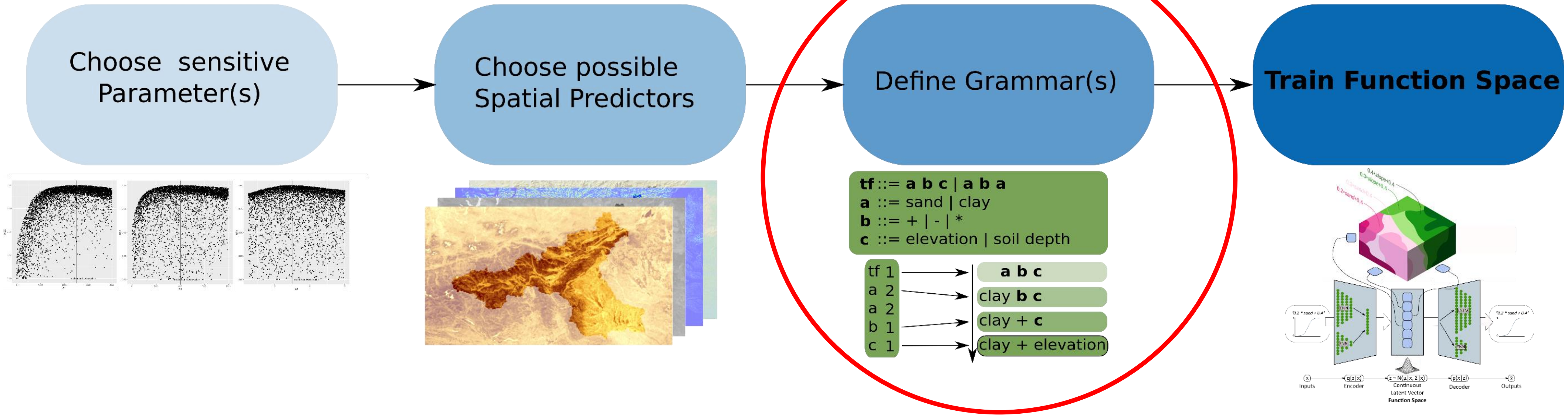
- For methodological details refer to Feigl et al. (2020)



FSO variational autoencoder

FSO workflow

1. Assembly Phase



FSO generates possible functions from a Context free grammar (CFG). A CFG defines rules, variables, functions and operations for creating transfer functions. The operators and functions used in this study are:

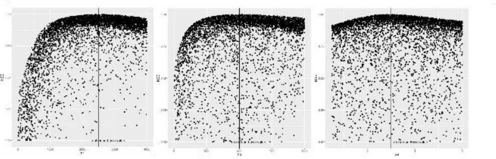
Operators: +, - , *, /, ^

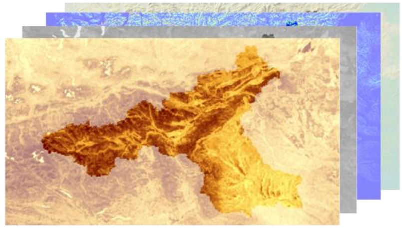
Functions: exp, log10, log, sin, cos, tan, abs, acos, asin, atan, cosh, sinh, sqrt

FSO workflow

1. Assembly Phase

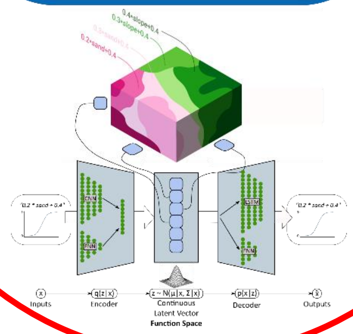
Choose sensitive Parameter(s) → Choose possible Spatial Predictors → Define Grammar(s) → **Train Function Space**





tf ::= **a b c** | **a b a**
a ::= sand | clay
b ::= + | - | *
c ::= elevation | soil depth

tf	1	**a b c**
a	2	clay **b c**
a	2	clay + **c**
b	1	clay + elevation
c	1	



The assembly phase results in a trained variational autoencoder that can generate transfer functions from a continuous vector space, called Function Space.

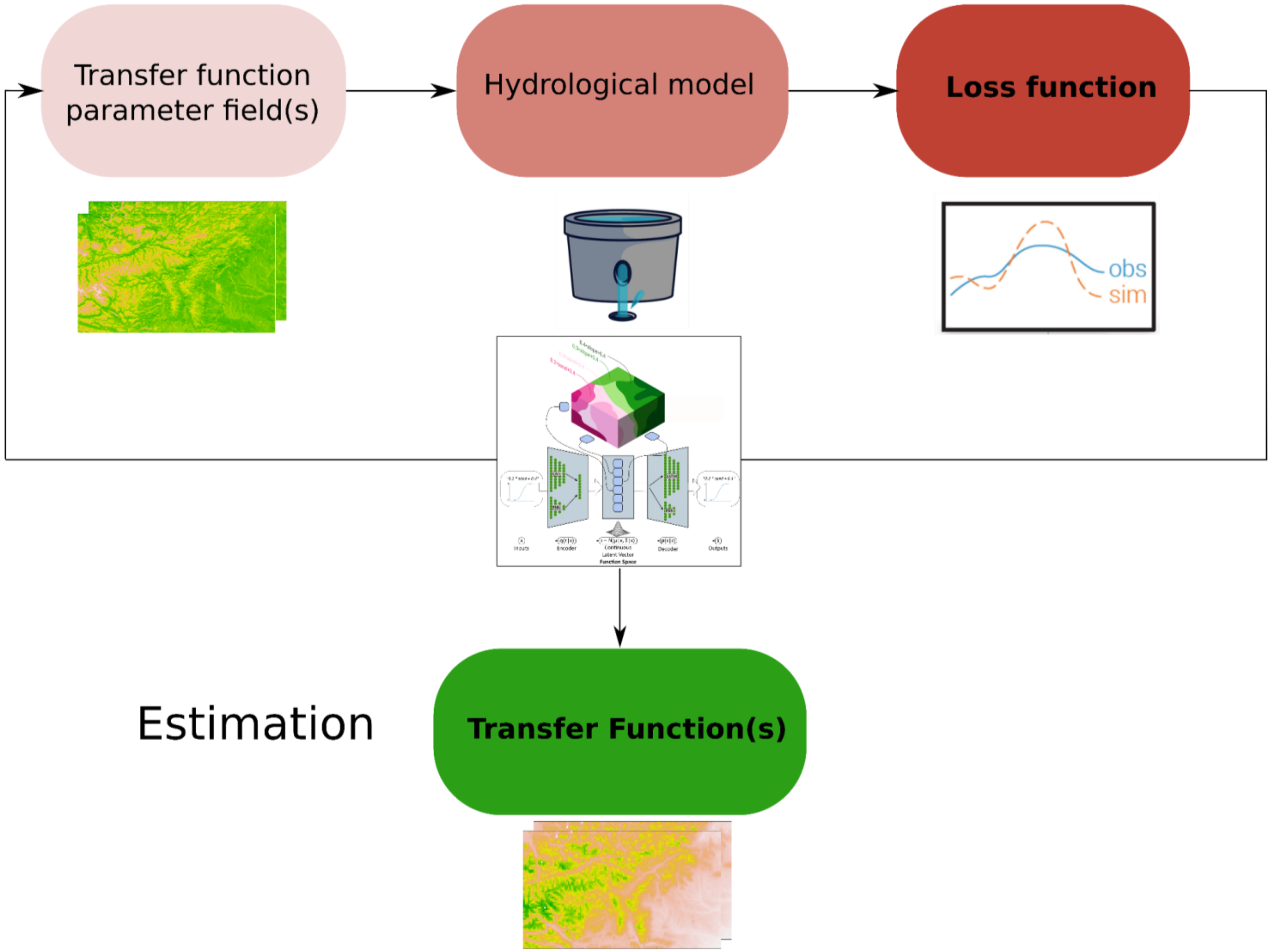
FSO workflow

- The DDS algorithm (Tolson & Shoemaker, 2007) for continuous optimization is used to search through Function Space.

- For each iteration a new transfer function is generated from the FSO autoencoder and used in the hydrologic model.

- The optimization loop is stopped after convergence or after reaching the maximum number of iterations.

2. Optimization Phase



FSO parameter scaling

Scaling

- Scaling of the geo-physical catchment properties is necessary to make them comparable and to not induce a bias resulting from different scales.
- In FSO, all geo-physical properties are scaled to the interval [0, 1] by their physical bounds, e.g. sand content from the range [0, 100] to [0, 1]. These scaled variables are denoted as e.g. $sand_{01}$.
- The values of the FSO transfer functions are then scaled to the parameter bounds before applying them in the model:

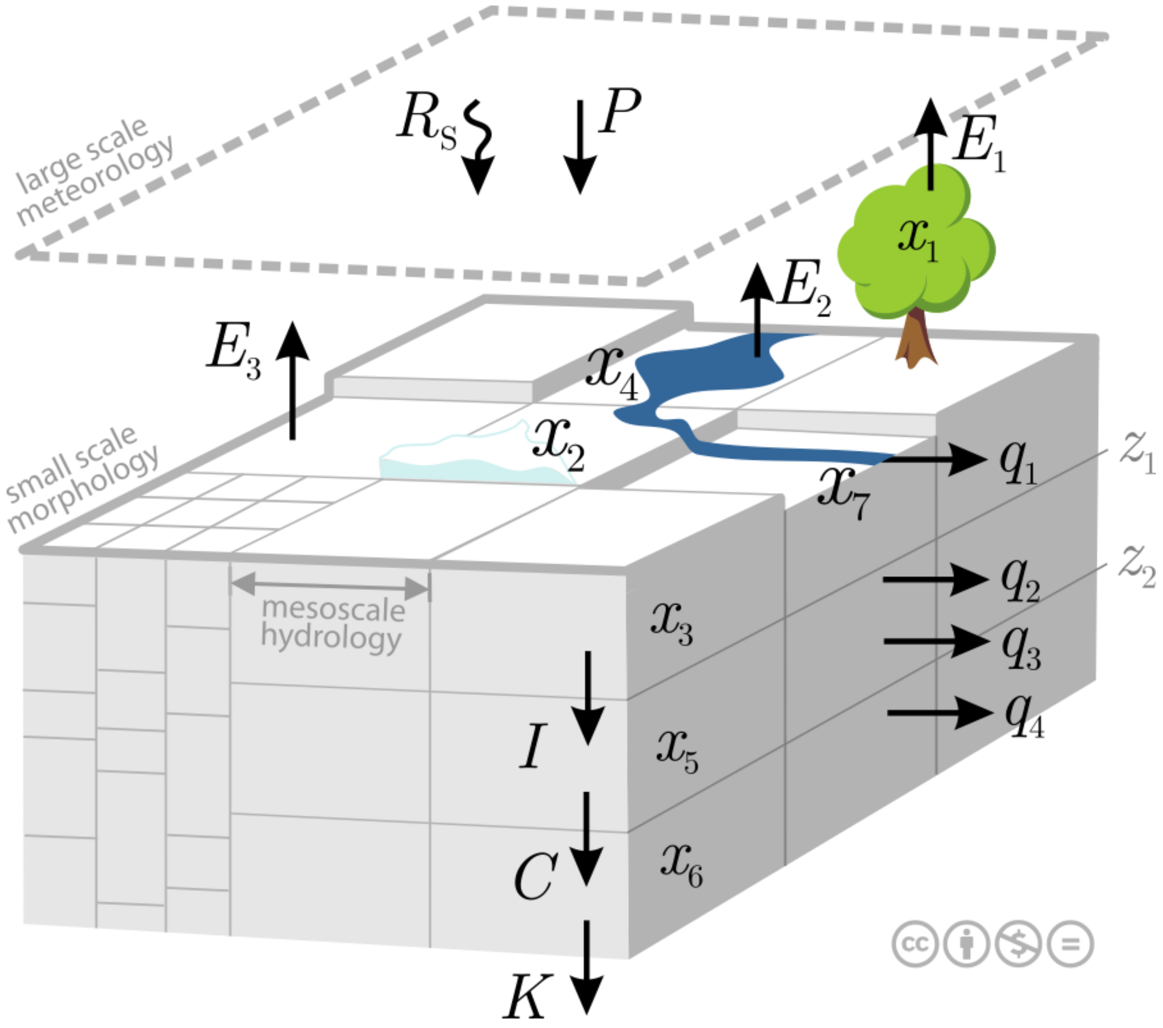
$$x_{[a,b]} = a + \frac{(x - \min(x))(b - a)}{\max(x) - \min(x)}, \; with\ parameter\ bounds\ a, b$$

FSO-mesoscale Hydrologic Model (mHM) case study

Aim:
To assess FSO‘s ability to estimate transfer functions in a real world setting using a model that already defines all its parameters by using transfer functions.

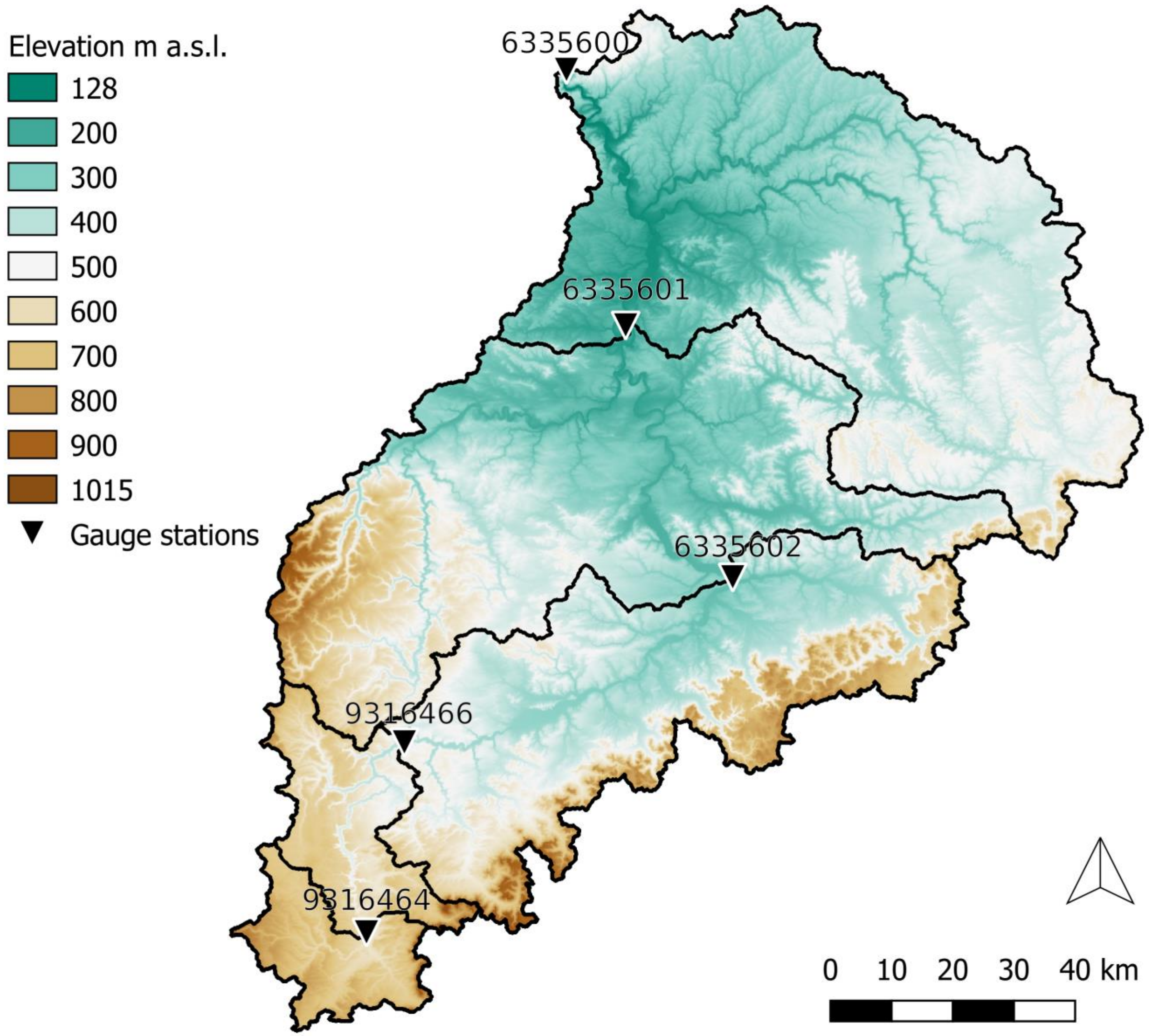
Setup:
- Apply mHM on 5 basins in the Neckar catchment
- Catchment data with 100 m and parameter fields with 4 km resolution
- 10 years each for training & testing and 1800 days spin-up
- Optimize 2 transfer functions: saturated hydraulic conductivity & Field capacity
- Compare FSO results to mHM parameter optimization results
- Standard values are used for all other model parameters

The mesoscale Hydrological Model (mHM)



- Developed by Samaniego et al. (2010)
- Spatially explicit distributed model
- Uses grid cells as primary units
- **Defines parameter fields with the Multiscale Parameter Regionalzation method (MPR)**
- The applied version of MPR allows to flexibly set transfer functions
- The model setup used in this study is equivalent to the one used in Zink et al. (2017)

Case study – Neckar catchment



Resolution:
- Spatial predictors: 100 x 100 m
- Model grid: 4 x 4 km

Basin data:
- 5 basins
- Training: 10 yrs data
- Testing: 10 yrs data
- Spin-up: 1800 days

Spatial predictors:
- Mean sand percentage
- Mean clay percentage
- Mineral bulk density
- Aspect
- Terrain slope

Please refer to Zink et al. (2017) for the data sources.

Case study - loss function

For this case study a weighted mean of all basin KGEs (Gupta et al., 2009) was chosen as objective function for the FSO optimization:

$$mean\ weighted\ KGE = \frac{\sum_{i=1}^{m} w_i KGE(Q_{o,i}, Q_{p,i})}{\sum_{i=1}^{m} w_i}$$

with the weights $w_i = 1 - KGE(Q_{o,i}, Q_{p,i})$ for all $i \in \{1, \ldots, m\}$. $Q_{o,i}$ and $Q_{p,i}$ are the observed and predicted time series of discharge for basin i, respectively.

The resulting applied loss function consists of the mean weighted KGE and a penalty for function length to avoid overfitting with complex functions:

$$loss = mean\ weighted\ KGE - function\ length * 0.001$$

This loss function aims to produce equally good results in all basins and therefore complies with the goal of finding a global transfer function.

Case study – transfer functions

The original mHM transfer functions of the parameters are:

Saturated hydraulic conductivity (cm/d):

$$KSat = \gamma_1 * exp(\gamma_2 + \gamma_3 * sand - \gamma_4 * clay) * \log(10)$$

– Only dependent on geo-physical catchment properties.
– 4 optimizable parameters: $\gamma_1, \gamma_2, \gamma_3, \gamma_4$

Field Capacity (-):

$$FieldCap = ThetaS * \exp(\gamma_5 * \left(\gamma_6 + log10(KSat)\right) * \log(vGenu\_n)$$

– 2 optimizable parameters: γ_5, γ_6
– Dependent on 3 other mHM parameters: ThetaS, Ksat, v_Genu_n

→ FSO also uses these 3 parameters as possible input variables for Field Capacity

Preliminary results – overview

mHM (MPR)

Time period	mean KGE	Basin 6335600	Basin 6335601	Basin 6335602	Basin 9316464	Basin 9316466
Training	0.702	0.469	0.777	**0.878**	**0.748**	0.656
Testing	0.704	0.512	0.826	**0.840**	**0.664**	0.650

FSO-mHM (MPR)

Training	**0.742**	**0.643**	**0.863**	0.839	0.674	**0.692**
Testing	**0.734**	**0.671**	**0.877**	0.760	0.656	**0.708**

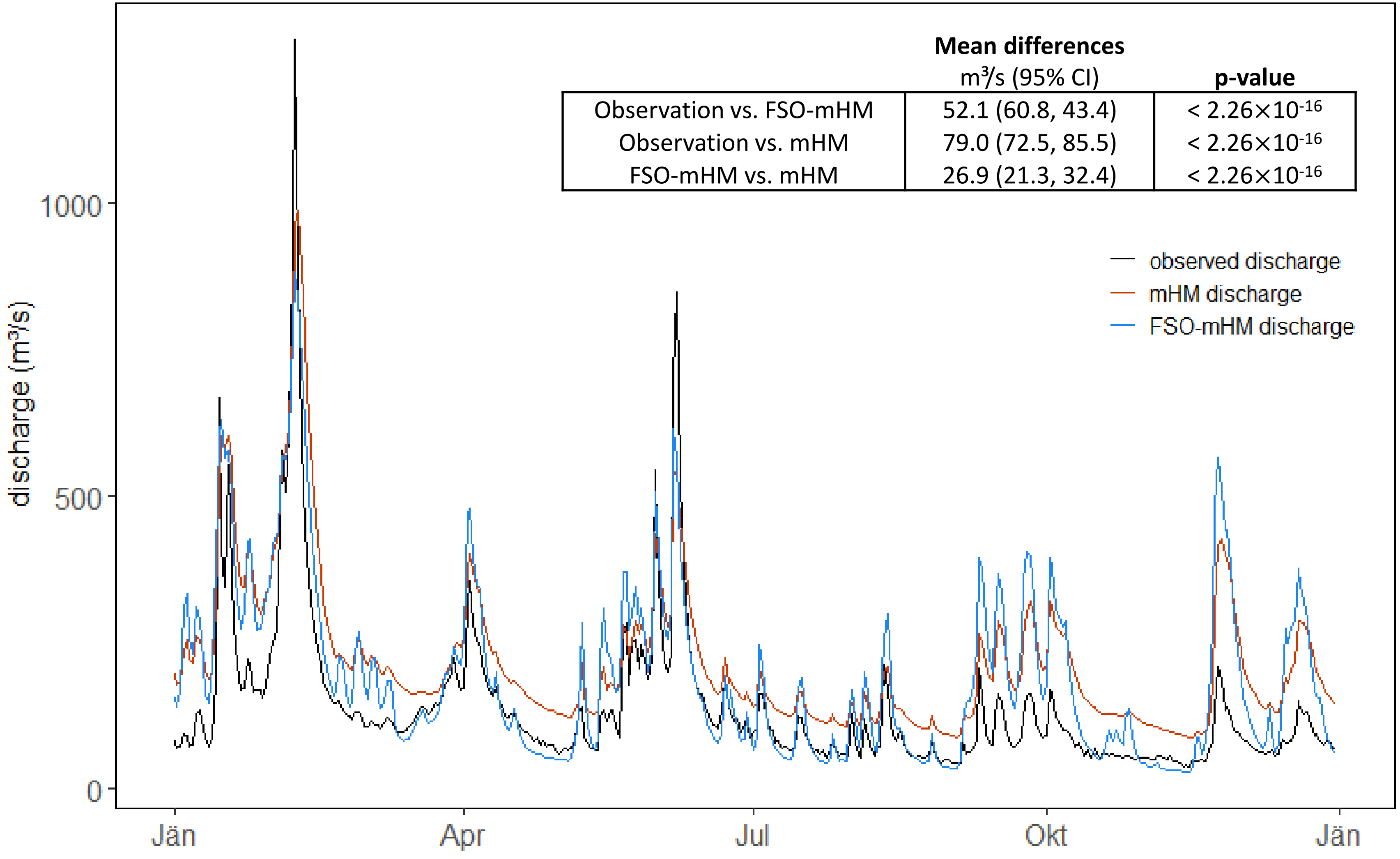
Mean testing differences

Difference in m³/s (95% CI)	31.8 (30.1, 33.7)	14.5 (13.5, 15.5)	8.8 (8.3, 9.2)	0.7 (0.5, 0.8)	6.0 (5.7, 6.3)
p-value	$< 2.26 \times 10^{-16}$	$< 2.26 \times 10^{-16}$	$< 2.26 \times 10^{-16}$	$< 2.26 \times 10^{-16}$	$< 2.26 \times 10^{-16}$

All mHM results (without FSO) are generated by optimizing the parameters for KSat and FieldCap and using the standard values for all other parameters. FSO-mHM transfer functions increased the performance of three basins, while two were slightly decreased. Performance in training and testing time period is similar for all basins, except basin 633560 which showed a decrease of about 0.08 KGE in the testing time period. The FSO-mHM predicted time series in the testing time period are significantly different from the mHM time series.

Preliminary results – discharge time series

	Mean differences m³/s (95% CI)	p-value
Observation vs. FSO-mHM	52.1 (60.8, 43.4)	$< 2.26\times10^{-16}$
Observation vs. mHM	79.0 (72.5, 85.5)	$< 2.26\times10^{-16}$
FSO-mHM vs. mHM	26.9 (21.3, 32.4)	$< 2.26\times10^{-16}$



Observed and simulated discharges for a year of the testing period (1984) for basin 6335600

The resulting FSO parameter fields improved the performance especially in the largest basin (6335600). Standard values for all parameters combined with FSO estimated Ksat and Field capacity parameters resulted in reasonable discharges.

Preliminary results – transfer functions

Saturated hydraulic conductivity (cm/d):

mHM (MPR) : $KSat = \gamma_1 * exp(\gamma_2 + \gamma_3 * sand - \gamma_4 * clay) * \log(10)$

FSO-mHM (MPR) : $KSat = \frac{clay_{01} * aspect_{01}}{log10(bulk\ density_{01})} + \frac{\mathrm{atan}(bulk\ density_{01})}{\sqrt{clay_{01}}}$

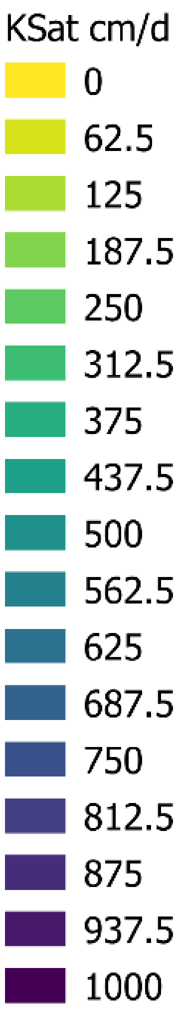
Field Capacity (-):

mHM (MPR) : $FieldCap = ThetaS * \exp(\gamma_5 * (\gamma_6 + log10(KSat)) * \log(vGenu\_n)$

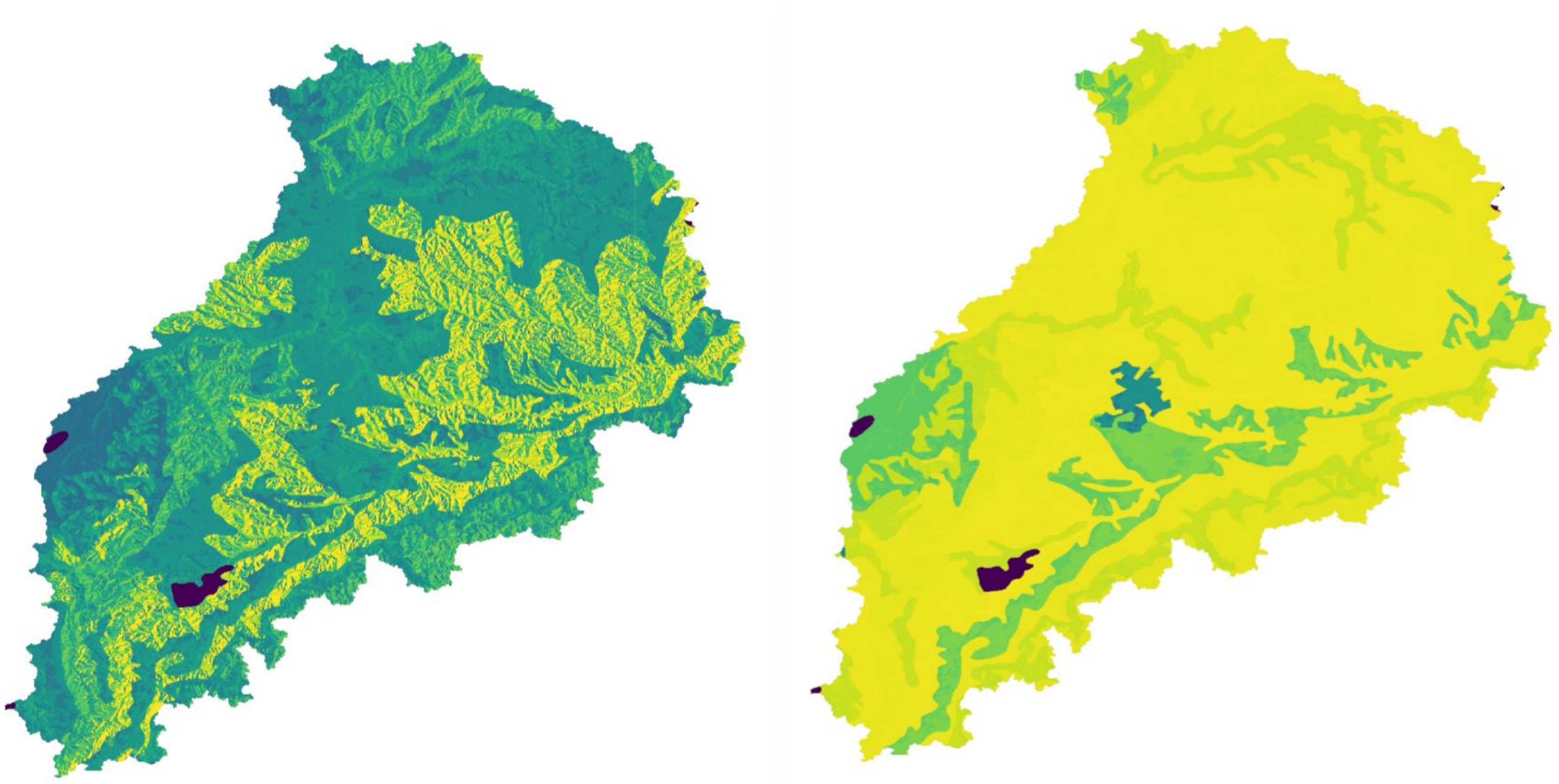
FSO-mHM (MPR) : $FieldCap = \frac{1.71 * slope_{01}^{0.93}}{clay_{01}} - bulk\ density_{01}$

The FSO estimated KSat function uses additional inputs compared to the mHM function. The FSO estimated Field capacity is dependent on soil properties but does not use any model parameters.

Preliminary results – saturated hydraulic conductivity



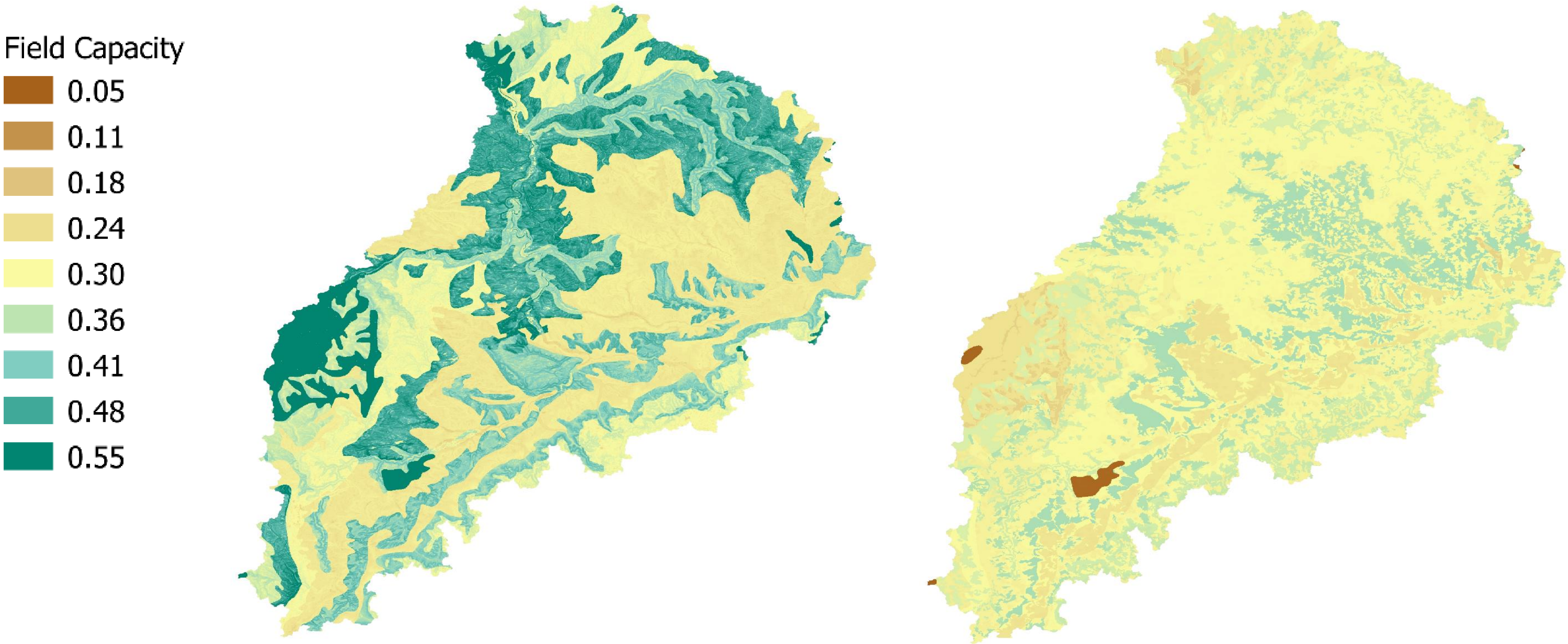
FSO-mHM (MPR) | mHM (MPR)



Resulting KSat parameter fields on the 100 x 100 m grid for the top layer of the model (tillage layer, first 20 cm)

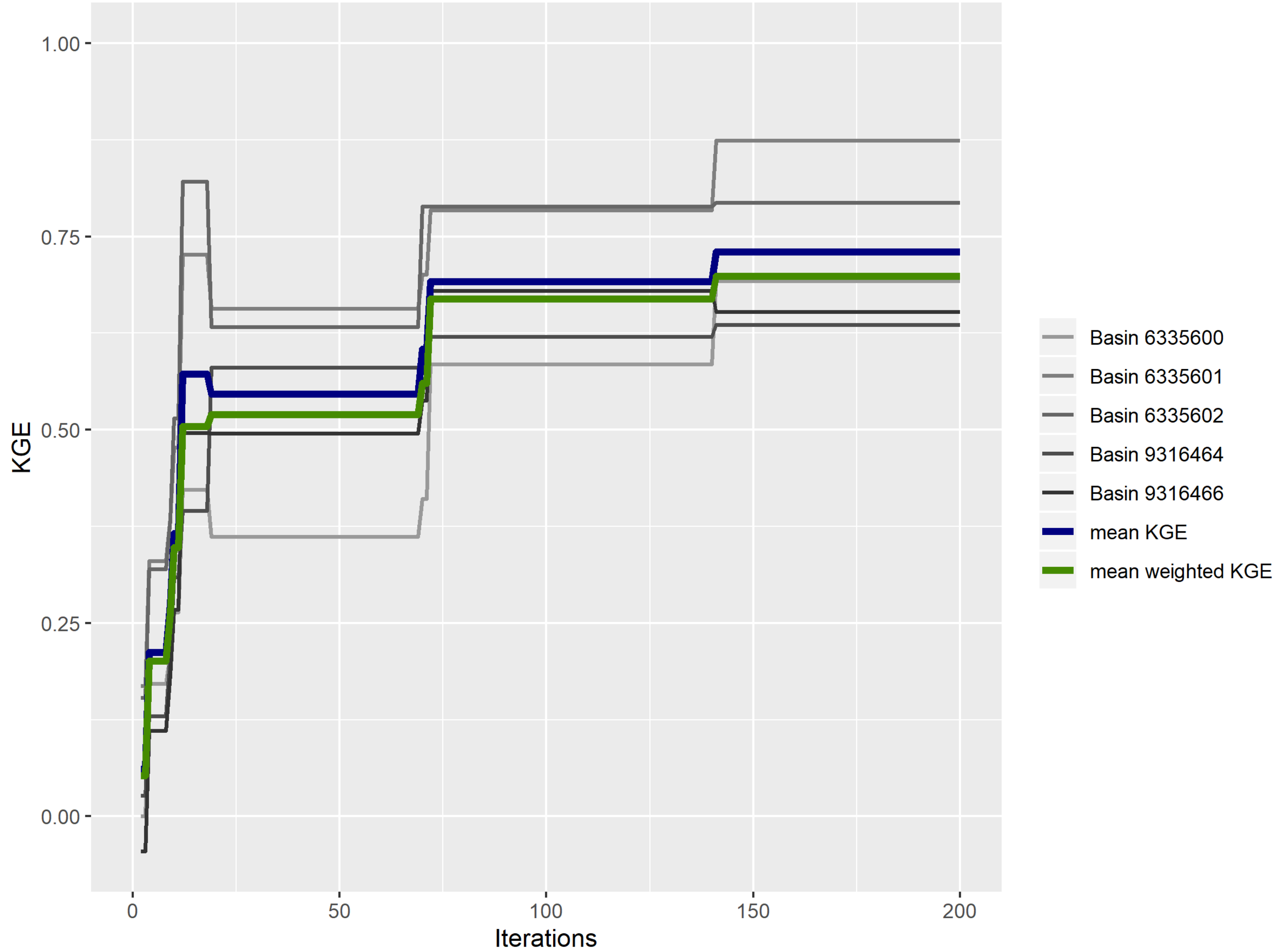
Preliminary results – Field Capacity

FSO-mHM (MPR) mHM (MPR)



Resulting FieldCap parameter fields on the 100 x 100 m grid for the top layer of the model (tillage layer, first 20 cm)

Preliminary results – Optimization



The FSO optimization run, showing the developement of KGE values with numbers of iterations.

Discussion

- The presented preliminary result are based on less than 300 iterations of the FSO optimization, which limits their interpreation.
- The FSO parameter fields show a higher variation of values on the small scale information.
- The Field capacity values estimated by FSO show a negative correlation with values calculated by standard mHM functions.
- FSO already showed improvement in the model fit after 200 iterations. However, the estimation procedure is based on only 5 basins. Also, the transferability of the optimized functions to other basins was not tested.
- This study shows promising first results for automatically estimating transfer functions with real world data.

UFZ HELMHOLTZ Centre for Environmental Research HyWa

References

Feigl, M., Herrnegger, M., Klotz, D., & Schulz, K. (2020). Function Space Optimization: A symbolic regression method for estimating parameter transfer functions for hydrological models. https://doi.org/10.1002/essoar.10502385.1

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. https://doi.org/10.1016/j.jhydrol.2009.08.003

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. https://doi.org/10.1002/2017WR021253

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

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

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017).). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrol. Earth Syst. Sci*, *21*, 1769-1790. https://doi.org/10.5194/hess-21-1769-2017