



An integrated multi-fingerprint sensitivity-nested approach for regional model parameter estimation and catchment similarity assessment

Simon Höllering¹, Jürgen Ihringer¹, Luis Samaniego², and Erwin Zehe¹

¹Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

²Helmholtz Centre of Environmental Research, Leipzig, Germany

Correspondence to: Simon Höllering (simon.hoellering@kit.edu)

Abstract. The present study provides a novel approach to the challenge of identifying behavioural parameters in the context of parameter sensitivity and related hydrologic similarity classification. A methodical framework is presented wherein global sensitivity analysis of a spatially distributed conceptual hydrologic model within 14 different mesoscale headwater catchments is combined with a parameter estimation scheme based upon both classification by (1) physiographic and climate and (2) related dynamic response characteristics represented by hydrologic signatures (fingerprints) creating an interface between hydrologic variables of observed and simulated origin. Changing ranks in (3) partial parameter sensitivities within the catchments indicate that hydrologic dynamics might be governed by different hydrologic processes. Model simulated and the respective observed response fingerprints are found to cluster within typical sample regions. These findings show a general model adequacy to represent mesoscale streamflow response processes that relate temporally dominant parameters and allow a reasonable constraint on the parameter space. The sensitivity-nested approach may be useful to calibrate hydrologic models sequentially on streamflow sections as well as on constraining (observable) single or combined hydrologic fingerprints and also to transfer results to similar sites, ungauged or anthropogenically altered.

1 Introduction

1.1 Defining hydrologic similarity

Assessment of catchment similarity is a cardinal challenge in hydrologic research. Although it might have been motivated by similarity theory in hydraulics and fluid mechanics (Monin and Obukhov, 1954), according to Dyck and Peschke (1995) it is extraordinarily more complex than its hydraulic archetype. Hydrologic similarity comprises physiographic, meteorologic, ecologic and geologic factors, notwithstanding human influences. As stressed by Oudin et al. (2010) one may approach the similarity problem from different angles. The classical regionalisation approach aims to represent similar catchments by similar behavioural parameter sets within a hydrologic model (Hirsch, 1979; Castellarin et al., 2001; Göttinger and Bárdossy, 2007; Patil and Stieglitz, 2012). The proposed measures for catchment similarity in related studies range from spatial proximity (Patil and Stieglitz, 2012; Vandewiele and Elias, 1995), which is often a good predictor for a similar climate setting, to similarity



of various physiographic catchment properties (Parajka et al., 2005) or combinations thereof (Merz and Blöschl, 2005; Oudin et al., 2008). Physiographic catchment similarity is, except with respect to geomorphic descriptors, often characterised by categorical data on soil, landuse and geology. These data sources are frequently translated into metric catchment descriptors by means of their areal share (e.g. Ali et al. (2012); Merz and Blöschl (2003); Hundecha and Bárdossy (2004); He et al. (2011)) and related to behavioural model parameters, for instance by means of variance reduction (He et al., 2011) or multiple regressions (Merz, 2006), which can be further constrained by a Lipschitz condition (Hundecha and Bárdossy, 2004; Götzinger and Bárdossy, 2007). Establishing a link between observable physiographic catchment properties and parameters of conceptual hydrologic models (CHMs) requires an extended calibration-oriented context. Correlation analysis, including model parameters and their estimation or even transfer to ungauged sites, is generally complicated by uncertainty inherent in parameters of CHMs (Bárdossy, 2006; Fenicia, 2008) and equifinality in acceptable model results obtained by different parameter sets (Beven and Freer, 2001; Beven, 2006). For this reason, Bárdossy (2006) proposed to use vectors containing pre-selected parameter sets for a reasonable model calibration in ungauged catchments.

However, parameter regionalisation feasibly represents "similar" catchments by similar parameters in conceptual models, this method of defining similarity/dissimilarity is restricted to those differences that can still be represented by the chosen model structure. This might imply that two catchments are treated as similar in a semi-lumped model structure, but will appear as dissimilar in a fully distributed model structure.

Alternatively, Wagener et al. (2007) proposed to distinguish similarity with respect to climate and physiographic catchment characteristics from dynamic response measures characterising streamflow release from or water storage within catchments. The ultimate vision is to relate similarity of independent properties, sorted in a multidimensional space, and dependent functional catchment descriptors in regionally specific classification schemes.

1.2 Classifying similar catchments

A generally accepted classification system in hydrology is defined as a *crucial step towards hydrologic synthesis* (Ali et al., 2012), a *fundamental step towards improved catchment hydrology science* (He et al., 2011) or as an *important organising principle, complementing the concept of the hydrologic cycle with the principle of mass conservation* (McDonnell and Woods, 2004) or even with energetic optimality (Zehe et al., 2013; Kleidon et al., 2013). The feasibility and usefulness of any classification scheme depends crucially on a) the identification of similarity measures/fingerprints of catchment behaviour and b) a successful explanation of differences in hydrologic functioning by differences in the catchment climate and physiographic setting. In this context McDonnell and Woods (2004) proposed dimensionless numbers, dominant hydrologic processes or functional traits as assortative elements.

A well known and successful classification system is the Hydrology of Soil types (HOST) in the United Kingdom (Boorman et al., 1995). It builds on the insight that the physical properties of soils and particularly soil structure, crucially influence



catchment hydrologic functioning in this region. Another regionally specific scheme, is the River Environment Classification (REC) in New Zealand (Snelder et al., 2004), which is based on the insight that catchment climate, topography, geology and land cover are decisive for riverine ecology. Winter (2001) proposed that Fundamental Hydrologic Landscape Units (FHLU) can be defined and distinguished based on their land surface form, hydraulic properties, geologic characteristics and climatic setting. In accordance with this Zehe et al. (2014) proposed a hierarchy of functional units that may guide experimental characterisation and development of appropriate process models for lower mesoscale catchments.

Alternatively, there have been several recent attempts to classify catchment dynamics often based on streamflow response characteristics. Sawicz et al. (2011) used six hydrologic response characteristics derived from precipitation, temperature and streamflow data to classify 280 catchments in the eastern USA by means of a Bayesian clustering algorithm. In line with this, Carrillo et al. (2011) relates the climate setting and physiographic characteristics of several US catchments to their hydrologic behaviour by means of dimensionless numbers derived from process-based signatures. Merz and Blöschl (2003) proposed a catchment-based framework, mainly relying on the classification of climate and meteorologic controls to classify the flood generation for the whole of Austria. Gharari et al. (2011) distinguished the three hydrologic response units - wetland, hillslope and plateau - with distinct runoff generation process (saturation excess overland flow, storage excess sub-surface flow, deep percolation) as building blocks for hydrologic modelling. To this end they adopted a landscape classification based on the Height Above Nearest Drainage proposed by Rennó et al. (2008). Di Prinzio et al. (2011) employed Self Organising Maps (SOM) in combination with a Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) for classifying 300 Italian catchments, using indices of streamflow regime and geomorphoclimatic characteristics. In another study, also based on SOM, Ley et al. (2011) performed a cluster analysis to identify, investigate and delineate hydrologically similar catchments considering their response behaviour in terms of more than 8200 event runoff coefficients (ERCs) and flow duration curves of 53 gauged catchments in Germany.

1.3 The inconsistency of functional and physiographic similarity - a new paradox?

Notwithstanding, the appeal and success of classifying catchments into "functionally similar" groups, it is not straight forward to explain these similarities based on catchment characteristics. A comparison of structural and functional similarity indices of 36 Scottish catchments by means of the affinity propagation clustering algorithm by Ali et al. (2012) revealed, for instance, a rather inconsistent picture between flow-derived similarity and physiographic similarity indicators. This is in line with the results of several other studies on similarity metrics which e.g. concluded that the overlap of the two groups with 60 % (Oudin et al., 2010) and 67 % respectively (Ley et al., 2011) is relatively small but can be raised e.g. by an analysis with Self-Organising Maps (SOM).

Parts of this inconsistency may be explained by the inherent equifinality and non-uniqueness of most of our governing equations (Beven, 1990; Zehe et al., 2014). Particularly, runoff as a mass flux leaving a catchment control volume is a non-unique



product of a drying potential gradient and the control volume conductance/resistance. Systems that distinctly differ with respect to the topographic controls on the driving gradients and the pedo-/geologic controls on control volume conductance may nevertheless produce runoff in a similar fashion (Wienhöfer and Zehe, 2014; Binley and Beven, 2003). Topographic and pedologic controls on runoff generation must thus be interpreted as a group, to judge how they jointly control runoff behaviour.

5 The necessary metric data sets on topography and soil water characteristics, are unfortunately not at hand. In line with this (Ley et al., 2011) attributes the inconsistency of functional and physiographic catchment similarity to the equifinality of runoff processes where a similar hydrologic response behaviour can be triggered by special (unique) physiographic catchment properties. Furthermore Oudin et al. (2010) invoked unique hydrologic settings of catchments and an absence of descriptors that are capable of linking the physiographic and functional catchment similarity as reasons.

10

1.4 Functional fingerprints for constraining hydrologic models

Fingerprints of catchment functioning may, furthermore, serve as additional measures to test and improve the adequacy of hydrologic models (Oudin et al., 2010), which might ultimately imply a better representation of catchment hydrologic functioning in hydrologic models. Castiglioni et al. (2010) proposed a calibration methodology for conceptual models which should generally be both as parsimonious as possible as well as adaptable to the degree of available data in the catchment. For their spectral, multiple objective calibration they suggested the use of catchment specific signatures derived as statistical river flow indices such as the standard deviation and the autocorrelation of the river flow data in order to result in an uncertainty envelope hydrograph formed by the multiple behavioural parameter sets. Yadav et al. (2007) and several other authors (Shamir et al., 2005a, b; Zhang et al., 2008) used dynamic response characteristics of watersheds to relate them to physically measurable catchment characteristics in order to derive regression relationships to raise the information considered in the parameter transfer and finally to reasonably constrain the parameter space. Vrugt and Sadegh (2013) also preferred to use indices based on hydrologic signatures to evaluate hydrologic models instead of the commonly applied standard efficiency measures based on residuals.

20

25 2 Objective and driving research questions

In line with the recommendation of Vrugt and Sadegh (2013) this study explores the consistency between catchment structural and functional similarity indices and particularly the value of the latter for constraining distributed hydrologic models. Specifically, we address three main research questions to contribute to clarification within the interrelated fields of hydrologic similarity, catchment classification and objective catchment functional parameter estimation:

30

- Can we constrain the feasible parameter space of a distributed hydrologic model in a consistent manner using fingerprints characterising streamflow generation during high, average and low flow conditions?



- May we employ regional differences in "dependent" parameter sensitivity of a distributed model to learn about regional differences in catchment functioning?
- Can we explain differences in catchment behaviour and parameter sensitivity based on typically available physiographic and meteorologic catchment descriptors?

5 We employed the mesoscale Hydrologic Model (mHM) (Samaniego et al., 2010b) as state-of-the-art distributed hydrologic model and the Fourier Amplitude Sensivity Test (FAST) (Cukier et al., 1973; Reusser, 2010) to explore temporal dependent parameter sensitivity of simulated streamflow within different subcatchments of the Ruhr basin in Germany. In particular, we employed the FAST method to explore parameter sensitivity of the selected fingerprints, within the range of their independent variable.

10 3 Concept and methodological approach

3.1 General considerations

Our approach rests on three pillars, 1) indices characterising the landscape properties, 2) streamflow generation and 3) dependent partial parameter sensitivities of mHM of simulated streamflow and of the related fingerprints. We explicitly focus on the headwaters of the strongly regulated Ruhr basin, to analyse, so far as possible, unaffected hydrologic regimes, thereby
15 conserving the ability to extend and test the approach to more strongly altered downstream locations. Our analysis, which as implemented in the Ruhr catchment using mHM, maybe generalised to any model or catchments using the following steps:

- Combining the application of a hydrologic model as learning tool and forward operator with global sensitivity analysis;
- Selections and comparison of physiographic and dynamic response fingerprints within the target catchment along with exploration of possible linkages;
- 20 – Comparison of simulated and observed dynamic fingerprints as respective constraints;
- Check and assess consistency of parameter sets to reproduce several fingerprints. Evaluate the most consistent ones against observed streamflow time series;
- Evaluation of dependent parameter sensitivities with respect to streamflow and the selected fingerprints;

In the following we give further details of the model mHM, the FAST method and our selected similarity indices.

25 3.2 The mesoscale Hydrologic Model (mHM)

The mesoscale Hydrologic Model, conceptualised on the basis of grid cells applied to a wide range of mesoscale river catchments ($10^1 - 10^4 km^2$) (Cuntz et al., 2015; Samaniego et al., 2010a, b, 2011; Kumar et al., 2010) accounts for diverse processes of the hydrologic cycle: Canopy interception, evapotranspiration, snow, soil moisture dynamics, infiltration and overland flow,



interflow, subsurface storage and groundwater recharge, baseflow, discharge attenuation as well as flood routing (Samaniego et al., 2010b; Kumar et al., 2010). Three model levels of gridded information are implemented in mHM (for more details, see Kumar et al. (2010)): morphology, hydrology, meteorology, with $l_0 \ll l_1 \leq l_2$ denoting the relative sizes of the grid cells at the respective data level. In this study both l_1 and l_2 were set to a spatial resolution of 1000 m, whereas for level-0 with the morphological catchment characteristics, a finer resolution of $l_0 = 200$ m was selected as an adequate spatial discretisation at this hydrologic scale. The mHM parameterisation is substantially based upon a simultaneous regionalisation technique (see also Hundecha and Bárdossy (2004); Götzinger and Bárdossy (2007)) called multiscale parameter regionalisation (MPR) (Samaniego et al., 2010b) to account for the physiographic sub-grid and hydrologic process variability. Level-1 hydrologic process parameters are derived from (pedo-)transfer functions with coefficients (in the following also referred to as (global) mHM parameters) related to physiographic characteristics at level-0. Hence, mHM is calibrated indirectly, altering the level-0 parameters instead of hydrologic level-1 parameters. This procedure not only reduces the problem of overparameterisation (52 global parameters) or the dependence on specific hydrologic scales (Beven, 2001) but also diminishes the amount of time to be spent on grid-wise calibration.

3.3 Dependent parameter sensitivity

3.3.1 The Fourier Amplitude Sensitivity Test (FAST)

The Fourier Amplitude Sensitivity Test (FAST) introduced and tested by Cukier et al. (1973, 1975); Schaibly and Shuler (1973) is a partial variance-based method to determine global sensitivities of parameter changes on the outcome of monotonic and non-monotonic numeric models (Frey and Patil, 2002; Song et al., 2013) with a limited number of evaluation runs. For a detailed explanation of the FAST method the reader might refer to Reusser et al. (2011).

20

The method was computationally enhanced to extended FAST (eFAST) by Saltelli et al. (1999) merging the advantages of FAST (high efficiency) and Sobol's technique (high order interaction terms and total sensitivity effects) for global sensitivity analysis (Sobol, 1993, 2001). FAST was originally applied to study parametric model sensitivities of chemical reaction systems. The method has since been used in several fields such as atmospheric (Rodríguez-Camino and Avissar, 1998), ecologic forestry (Song et al., 2013), hydrogeologic (Fontaine et al., 1992) or geologic nuclear waste disposal modelling (Lu and Mohanty, 2001). More recently, a few studies also include the application of (e)FAST to hydrologic modelling (Guse et al., 2014; Reusser et al., 2011; Reusser and Zehe, 2011; Zajac, 2010). Sanadhya et al. (2013) employed eFAST in combination with the Soil Water Assessment Tool (SWAT) (Arnold et al., 1998) to analyse critical interactions of snow-related parameters in mountainous watersheds.

30



3.3.2 FAST implementation for mHM in the study area

In this study, we implemented the FAST method (Reusser, 2014) to initially conduct a sensitivity analysis on mHM with a focus on six global mHM parameters (see Table 1) explaining a substantial part of the model output variance (see also Cuntz et al. (2015)) in the upper Lenne basin at gauge Bamenohl (BAM). The parameters were pre-selected by local sensitivity analysis in terms of different indices for streamflow dynamics and water balance. Subsequently, a pre-run of FAST including 14 parameters previously identified as most sensitive was performed at BAM. For this study FAST was applied again on a subset of six mHM parameters showing highest sensitivity coefficients in the pre-run.

Table 1 presents both these six mHM parameters and their value ranges determining the number of FAST variations (i.e. mHM model runs). The parameter intervals were basically selected from the mHM literature (Samaniego et al., 2014) and then slightly extended based on results from numeric experiments within local sensitivity analysis. In our case, the FAST method requires altogether 91 model runs (variations) to assign sensitivity coefficients to each of the six mHM parameters (see Fig. 1). Parameter-related sensitivity values have been evaluated for each of the 14 headwater subcatchments (see section 5.4).

All other mHM parameters were fixed on pre-calibrated reference values found via global automatic optimisation using the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker, 2007) at gauges BAM and Wenholthausen (WEN). DDS calibration results coming from two different precipitation data sets: 12 precipitation stations with Thiessen polygons (Thiessen, 1911) and HYRAS (Rauthe et al., 2013), for the six study-relevant mHM parameters are also depicted in Fig. 1. Covering the whole of Germany and several adjacent regions the HYRAS data set was originally derived from 6200 rain gauges for the period from 1951 to 2006. With 1 km² grid size a comparably highly resolved daily precipitation data set was available for pre-calibration and FAST-mHM simulations.

Apart from global sensitivity analysis, FAST can be used to span an ensemble spectrum of streamflow simulations. In this study FAST variations provide an ensemble of 91 different streamflow time series for further analysis of model based hydrodynamic response and subsequent parameter identification (see section 3.4.2 and section 5.4).

Based on the ensemble of 91 streamflow simulations we derived ensembles of 91 realisations for dynamic fingerprints characterising different streamflow attributes, to assess their dependent parameter sensitivity as elaborated in the next section.

3.4 Fingerprints for regional parameter estimation

3.4.1 Relevant physiographic characteristics

On the basis of the available physiographic and meteorologic data a set of 12 indices characterising similarity of the climate and the physiographic setting (see Table 2 and Fig. 2) was selected to



1. capture characteristics that control relevant hydrologic functions of a mesoscale catchment as defined e.g. by Black (1997) such as partition, storage and release of water,
2. relate physiographic and climatic characteristics to sensitivity-confined hydrodynamic response fingerprints,
3. keep the option open for additional research to perform cluster analysis for catchment similarity classification and,
- 5 4. to further investigate and avoid the hydrologic similarity paradox described in section 1.3.

Each of these fingerprints were assigned to one of five classes proposed by Yadav et al. (2007) representing the physiographic functions in the panoply of hydrologic processes. The choice of physiographic fingerprints originates from correlation analysis of catchment descriptors within each category (Yadav et al., 2007), multivariate statistical analysis techniques (Di Prinzio et al., 2011), from practical experience with slope-averaging measures in regionalisation models (Plate et al., 1988) and from comparison of methods for baseflow separation (Duband et al., 1993; Eckhardt, 2008, 2012). For subsequent research on the topic of classification using non-redundant hydrologic similarity indices, an alternative composition of physical fingerprints might certainly be conceivable. Ley et al. (2011) for example based a classification study on a set of six physical catchment characteristics including a fingerprint characterising zones of significant groundwater recharge. Likewise, additional similarity indices coping with anthropogenically altered hydrologic settings might be considered in the context of future socio-hydrologic studies.

15

3.4.2 Selected fingerprints for streamflow generation

Six hydrologic fingerprints reflecting the integral hydrodynamic response characteristics of the mesoscale headwaters were derived from streamflow and precipitation data (see Table 2).

One dynamic fingerprint was chosen for this study, each characterising streamflow generation during high (High Flow Discharge), average (Runoff Ratio) and low (Baseflow Index) flow conditions, the variability of streamflow (Coefficient of Variation), the frequency of flow events (High Pulse Count) as well as the rate streamflow changes in a catchment (Slope of Flow Duration Curve). This subset of hydrologic fingerprints is regarded as a reasonable prelude open for later expansion to further response signatures. The fingerprints were selected and classified in accordance with Olden and Poff (2003) and Yadav et al. (2007) who took the study of redundancies via linear and Spearman rank correlation (Yadav et al., 2007) and multivariate PCA (Olden and Poff, 2003) of indices as a basis for their choice of fingerprints.

25

The response indices were simultaneously determined from the observed and from the simulated ensemble of 91 streamflow time series originating from FAST-mHM parameter variations (see Fig. 2 and section 3.3.2). In addition to the physiographic and climatic characteristics of headwaters, the dynamic response fingerprints were used as a means to constrain the FAST parameter space keeping the observed quantities as a benchmark in the first instance. This provides an insight into the model performance based on FAST ensembles, i.e. multiple mHM parameter sets.

30



Whether weak or strong, constraints have to be thoughtfully considered, since too strong constraints can result in an overconditioning of the parameter space, leading to a situation with non-resolvable constraints (Zhang et al., 2008) which can highly limit the predictive power of the model causing a complete rejection of all parameter sets. If confining limits are chosen that are too weak the identification of behavioural model set-ups will be complicated and equifinality of retained models (Beven, 2006) might be introduced. If the performance of the streamflow ensemble simulation is rated as very poor, i.e. the distance to the observation is large, a re-run of FAST might be helpful in the search for behavioural parameter sets. The closest simulation results in terms of the underlying observed response fingerprints could be taken to determine the parameter range needed for revised sensitivity analysis.

10 A special focus was set on the observed flow duration curves (FDCs) of the 14 catchments serving as constraints for the parameter space of the ensemble simulation (see Fig. 3).

Slope values between 33 % and 66 % flow exceedence were derived as fingerprints that represent the rate streamflow changes for intermediate flow events. The consideration of FDCs as are commonly employed to indicate and classify catchment response functioning (Yilmaz et al., 2008) enables analysis of observed and simulated streamflow data - free of autocorrelation and time dependency e.g. in contrast to classical hydrograph inspection. A pronounced sensitivity in terms of the vertical redistribution of soil moisture was further recognised by Yilmaz et al. (2008). Sequential calibration can be performed on curve sections of special importance in terms of catchment response for flow segments of high, medium and low magnitude. Westerberg et al. (2011) showed the advantages to use FDCs as an objective function for model calibration with behavioural parameter sets setting up an uncertainty framework based on a limit of acceptability GLUE approach (see also Blazkova and Beven 2009)) for FDC calibration on selected evaluation points (EPs) of the FDCs. Singh et al. (2014) point out that depending on physiographic similarity classification of 83 gauged US catchments with rather few anthropogenic influences on the hydrologic regime, the slope of the FDCs in addition with a few other streamflow response characteristics (e.g. baseflow index and runoff ratio) can constitute a useful constraint for parameter transfer to ungauged sites. Yu and Yang (2000) employed a methodology whereby synthetic flow duration curves of ungauged Taiwanese sites were derived in order to use them as an objective function preconditioning the parameter space of the applied HBV model.

4 Study area

4.1 The Ruhr catchment - An anthropogenically influenced system

The Ruhr, with a catchment area of 4.485 km², originates from a spring about 670 m a.s.l. on the northern slope of the Ruhrkopf in the Hochsauerland (about 842 m a.s.l.) and joins the Rhine at Duisburg-Ruhrort (about 20 m a.s.l.) after 219 km. The landscape characteristics of the Ruhr catchment range from densely wooded and scarcely populated lower mountain ranges in the Sauerland to widely impervious urban areas in the river valleys and in the western part close to the mouth. The areas can be attributed to the geology and geography of the Rhenish Slate Mountains to the east of the Rhine (Brudy-Zippelius, 2003). The



average discharge at its mouth joining the Rhine is about $80.5 \text{ m}^3 \text{ s}^{-1}$ (Bode et al., 2003).

With a total of 8 dams and 5 reservoirs the Ruhr with its tributaries forms a very complex hydrologic system. The total stored water surface area of just under 35 km^2 equates to about 480 million m^3 as the volume of all the water retained behind the damming structures (Ruhrverband, 2011). An intensive use of water resources (e.g. reservoirs, barrages, sewage plants, withdrawals, inlets, etc.) supplies almost 5 million people with drinking and processing water along the Ruhr and within its catchment area and because of this intensive use, a large amount of water is lost completely (310 million m^3) (Bode et al., 2003). Annual evaporation of about 35 million m^3 can be estimated for the known inundated areas by using a simple empirical method e.g. by Haude (1955). The large anthropogenic influence on the hydrologic cycle, particularly water export, low water elevation and evaporation from the reservoirs, causes significant discrepancies and an overestimation of the water balance when applying conceptual hydrologic models to river reaches downstream of the reservoirs e.g. at gauge Villigst (see Fig. 4). Employing model efficiency measures without taking account of anthropogenic changes in the water balance due to damming and water management (Morgenschweis et al., 2003) is highly inappropriate for estimating the quality of a model in the Ruhr catchment. Comparisons with the clearly affected observed time series generally rule out high model efficiencies as the model reproduces the unaffected hydrologic system without the use of water resources.

4.2 Ruhr headwater catchments

The present study concentrates on 14 headwater subcatchments of the river Ruhr and upper reaches of its tributaries (e.g. Bigge, Lenne and Möhne) where the assumption of unaffected hydrologic regimes is justified to a much higher extent (Fig. 4). The pre-selection of headwaters is an essential factor in avoiding additional uncertainty in model parameterisation (Liu and Gupta, 2007; Fenicia et al., 2008).

The headwaters are situated in the eastern rural part of the Ruhr basin with higher altitudes, covering an area of 1989.2 km^2 in total, where individual catchment sizes range from 28.7 km^2 at gauge Amecke (AME) to 453.1 km^2 at gauge Bamenohl (BAM). Average catchment slopes vary between 10.8 % (Rüblinghausen) and 26.1 % (Kickenbach). The dominant form of land cover is forest (39.7 % - 87.3 %) followed by pasture (0.8 % - 47.5 %), cropland (7.6 % - 43.9 %) plus a few predominantly dispersed settlements (0.0 % - 13.2 %) with a subordinate role with regard to the integral hydrologic response characteristics (see Fig. 2a). The climatic conditions are humid-warm-temperate (Göppert et al., 1998) with warm summers and moderate winters, with annual mean temperature values between $8.45 \text{ }^\circ\text{C}$ and $5.45 \text{ }^\circ\text{C}$ from lower to higher altitudes in the eastern part of the headwater catchments. Annual precipitation ranges from 959 mm in the northern to 1473 mm in the south-western catchments (Ruhrverband, 2013).



5 Results

5.1 Constraints on parameter space

As described above in section 3.4.2 the aggregated dynamic response of headwater catchments has been derived in the form of six streamflow indices considered as regionally typical hydrologic fingerprints. Fingerprints were evaluated for observed and simulated time series. The former act as constraining quantities for the latter where an ensemble of model runs was made available from 91 different parameter combinations spanning the FAST-mHM parameter space. As well as box plotted catchment physiographic and climate characteristics, Fig. 2b shows the results in an overview for all six dynamic fingerprints considered in the form of the simulated ensemble and of observed values at the 14 gauges. Figure 2 also includes the baseflow index (BFI) as an intermediate hydrologic catchment signature between a physiographic descriptor and a dynamic response characteristic. For the next step, the observed and simulated values of two response fingerprints were plotted as scatter points against each other. Figure 5a-d shows the values of observations and ensemble simulations for four pairs of dynamic fingerprints at seven selected gauges. The observed quantities are plotted as circles where the centre represents the actual 2-d combined fingerprint value and the radius serves as a qualitative constraint on the ensemble of Fast-mHM simulations. For each radius a scaling factor is introduced which can be regarded as a measure of distance orientated towards the overall data spread of the 14 derived values of one fingerprint. The factor is determined as a multiple n of the average of the two standard deviations σ of the pairs of fingerprints (FP) considered in Fig. 5a-d.

Simulated value pairs which plot inside the circles can be rated as behavioural, assuming the hydrologic response similarity of the model and the hydrologic system in terms of the two combined response indices (Blöschl et al., 2013). Thus, the underlying parameter set can be shortlisted potentially together with other behavioural parameter combinations. Likewise, Euclidean or Chebychev distance values between the 2-d combined observed and each of the simulated plots might serve as feasible measures of functional performance to evaluate FAST-mHM ensembles (see section 5.3).

Consider the case where the coefficient of variation (CV) as a hydrologic fingerprint of longterm streamflow variability is combined with the slope of the respective flow duration curve (SLFDC) indicating the rate of change in streamflow of average magnitudes (33 % - 66 % exceedance) (see Fig. 5c). The first point to recognise is that for each catchment the simulated FAST-mHM ensemble clusters in an arrow-like sample space showing relatively small scatter along the two branches. The head is positioned at the right-hand end of the horizontal axis with larger values of SLFDC of about 2.1 % vertically central at a CV -value of 1.4 connecting the two branches of the arrow. The lower branch where the simulation produced positively correlated scatter between both fingerprints is regarded as the behavioural subset of the model runs. The observed CV values cluster in a comparably dense range of 1.1 to 1.3 in the right half of the lower branch of the arrow head implying similar hydrologic functioning for this response signature of streamflow variability.



Most of the circles contain at least a couple of corresponding simulated data points. However, in the case of Fig. 5c (CV against SLFDC) the number of behavioural model runs varies among the seven catchments, which portends that the constraining effect by a uniform distance measure (circle radius) diversifies catchment specifically. Gauge Meschede (MES) exhibits only one model run plotting at the low central edge of the circle (medium brown colour), still underestimating streamflow variability of the relatively large catchment with the highest elevation range, a steep catchment slope and a lower weighted slope along the main flow path, all at the same time. Fast reacting interflow components capture a more substantial part of the hydrograph which can be verified by the baseflow index based on the observed data that shows the lowest value ($BFI=0.3$) among all 14 catchments (see Fig. 2a). In other more frequent cases such as gauge Kickenbach (KIC) (grey) several parameter sets produce results close to the observed reference values (circle centre). Similar results were obtained from the combination of High Pulse Count (HPC) with SLFDC plotted in Fig. 5b. Circles can clearly constrain the bulk of model simulations with varying strength.

The combined scatter of CV against HPC shows a kind of expected image of correlated fingerprints (Fig. 5d). Streamflow variability and the yearly frequency of flow events above three times the average streamflow represented by CV and HPC exhibit a strong positive correlation with $\rho = 0.84$ for the observed fingerprints at the 14 headwater gauges. The ensemble of model runs even indicates a higher strength of linear relationship of the two fingerprint variables ($\rho = 0.96$). This clear linear correlation can be rated as a sign of redundancy of two fingerprints describing similar response characteristics. Obviously, the seven circles in Fig. 5d act as constraints for the scatter of simulations wherein all observations are centrally positioned, indicating intermediate streamflow dynamics.

In exactly the same manner dynamic response fingerprints were combined with the intermediate fingerprint BFI (introduced in Table 2 and Fig. 2) both for observed values and for the model simulated ensemble output. The pairwise analysis consistently including BFI shows somewhat similar findings as illustrated and presented here. Likewise, for example the contribution of SLFDC leads to an arrow-like simulated sample space and combinations with CV exhibit a linear correlation to a certain degree that here is distinctly stronger for the simulated ensemble ($\rho = 0.91$) than for the observed data (0.52). The coupling of BFI with HPC, different from the results in Fig. 5b, though less marked, also shows this correlation ($\rho = 0.87$; 0.62). Similar site specifically varying constraining effects on the ensemble by the observed combined fingerprints were found in this analysis based on BFI.

5.2 Linkages of physiographic characteristics and hydrodynamic response

Different from those itemised in section 5.1, fingerprints that characterise the hydrologic response of a catchment were only plotted on the ordinate, whereas both physiographic and climate catchment characteristics were plotted on the abscissa. In Fig. 6a-d SLFDC representing the rate of change in average streamflow was combined with four physiographic attributes for forest land cover (FOR), average catchment slope (SLOPE), longest drainage path (LDP) and the flow accumulation (FACC)



describing land use, topography and landform, respectively.

For most of the gauges (11) more than half of the ensemble of 91 simulations shows an underestimation of the observed SLFDC values which vary only slightly from 1.3 % at Meschede (MES) to 1.9 % upstream the Möhne reservoir in Neuhaus (MOE). Nevertheless, differences in SLFDC, though weakly pronounced, indicate a diverse composition of hydrologically relevant structural elements. Trying to explain in this context certain salience at selected headwaters, the four highest SLFDC values were found at MOE, HER, KIC and BAM (Fig. 6). These are the gauges exhibiting both highest percentage of forest land cover (FOR) (66.8 % - 87.3 %) (Fig. 6a) and, with the exception of MOE, steepest average catchment slopes (21.1 % - 26.1 %) (Fig. 6b). Other dynamic response fingerprints such as HPC and HFD (see Fig. 2b) plot at values close to the overall average of all headwaters which portray jointly a moderate hydrologic response for these four catchments in terms of high flow conditions.

5.3 Consistency and performance of behavioural parameter sets

In this subsection behavioural FAST-mHM parameter sets (see also section 5.1) were analysed and evaluated against observed streamflow time series. The evaluation of behavioural parameter sets was done with respect to their consistency in reproducing one fingerprint of hydrodynamic response, and combinations of two, both at different locations with slightly different physiographic settings and for other signatures, which reflect further major response characteristics of the mesoscale headwaters.

In Fig. 7a the Nash-Sutcliffe model efficiency (NSE) criterion (Nash and Sutcliffe, 1970) is plotted against the relative error between simulation and observation. In this case, evaluation is shown for the best performing parameter sets, in terms of each dynamic fingerprint at seven selected headwater gauges. Different colours were chosen to mark the headwaters, whereas the distinguishable form types of the markers indicate individual fingerprints. Each marker is labelled by a number signifying the parameter set in the ensemble found to perform most behaviourally in the reproduction of the fingerprint value. Additionally, Fig. 7b illustrates a subset of Fig. 7a condensed to the two fingerprints Runoff Ratio (RR) and the slope of the flow duration curve (SLFDC) at two gauges Amecke (AME) and WEN, extended by the results for the whole of the simulated ensemble of FAST-mHM parameter sets (model runs).

It can be frequently observed, that single dynamic response fingerprints are most behaviourally reproduced by fingerprint specific parameter sets. Furthermore, in this case, the fingerprint specific parameter sets seem to be unique for distinct headwater gauges, fairly closely located in some cases. Figure 7b also shows that for the simulation of RR the parameter sets with smallest relative errors plot close to the optimal solution within the ensemble concerning NSE, while both best SLFDC results show negative NSE values and are way out, especially at the gauge AME.

Contrarily to the findings for one single fingerprint, the reproduction of two combined dynamic fingerprints (section 5.1) can more frequently result in a consistency of most behavioural parameter sets plotting most closely to the observed pair of



fingerprints. The scatter plots in Fig. 7c and Fig. 7d evaluate the simulated ensemble for combined fingerprints (see also Fig. 5, section 5.1). NSE values are plotted against the Euclidean distances (D_E) between four pairs of observed and corresponding simulated fingerprints. Parameter sets with smallest D_E , in the following are shortlisted as the most behavioural for the reproduction of combined fingerprints, and are labelled and markers are bold-framed. Figure 7c illustrates a fairly consistent picture
5 whereby parameter set 78 appears to be the best fit to the combination of SLFDC and HPC at both gauges and at least for WEN in terms of the pair HPC-CV. Figure 7d only partly conforms to this image of consistency, revealing solely set 78 as the most behavioural for RR combined with SLFDC at gauge AME.

Up to this point the evaluation of model performance via FAST-mHM ensembles with observed values indicates that most
10 behavioural parameter sets, which best reproduce single or combined dynamic fingerprints, might not necessarily lead to best results among all ensemble members in terms of the NSE performance measure. This assumption is supported by Fig. 7e-f, where the overall five most behavioural parameter sets for combined fingerprints (referred to the seven shown headwaters), the Spearman's rank correlation coefficient ρ_s (e) and the water balance percentage error (WBE) (f) are scatter plotted against NSE within the ensemble.

15

Sets identified as behavioural combinations of six mHM parameters were further analysed in terms of their distances among themselves in the 6-d vector space. Predominantly, we used the Chebychev distance to detect the placement of a specific set within the entire FAST-mHM ensemble. Chebychev distances, along with the corresponding Euclidean values in reference to behavioural set 78 for combined dynamic fingerprints, were incorporated for all parameter sets. Generally, Chebychev
20 distances (also Euclidean) of five best parameter sets (of combined fingerprints) vary relatively with the chosen reference parameter set. Though placed very closely to 78, parameter set 64 does not necessarily show the most similar performance values (see Fig. 7e-f). This finding might be partly triggered by the nature of the maximum norm underlying the Chebychev distance metric emphasising the largest differences in higher dimensioned parameter values without necessarily having the largest influence (sensitivity) on runoff generation. Generally, Chebychev distances vary site-specifically as well as among the
25 different fingerprints. In our case, distances are generally found to be smaller between the parameter sets which best reproduce combined fingerprints (18.7) than between site and fingerprint specific sets that reproduce a single fingerprint value most adequately (23.5 (WEN), 25.8 (AME)).

Since simulations based on behavioural parameter sets frequently showed rather poor results, if performance measures like
30 NSE and ρ_s were considered (Fig. 7), visual inspection of time series was conducted. This allows better to judge why these sets do not perform well during simulation, whether e.g. peaks are too small or timing errors possibly affect the results. As illustrated in Fig. 8a and 8b streamflow hydrographs that originate from the most behavioural parameter sets in terms of the reproduction of two combined dynamic response fingerprints (SLFDC-HPC, HPC-CV, RR-SLFDC and SLFDC-CV, see also Fig. 5 and Fig. 7) were further analysed and compared to the hydrographs derived from observational data as well as to the



simulation revealing highest Nash-Sutcliffe model efficiency values within the FAST-mHM ensemble.

Both examples in Fig. 8 show that two hydrographs plot at the lower, under- (SLFDC-CV) and upper, overestimating (SLFDC-HPC) end of the ensemble spectrum during average and low flow periods. Streamflow spectrums are simulated with high dynamics (SLFDC-CV) or high damping (SLFDC-HPC), leading to an over- (SLFDC-CV) or underestimation (SLFDC-HPC) of observed high flow conditions. The combination of RR-SLFDC exhibits intermediate (Fig. 8a) or similar highly damped (Fig. 8b) streamflow characteristics, whereas parameter sets reproducing HPC-CV most adequately consistently result in highly damped hydrographs with low dynamics at both gauges (Fig. 8a and Fig. 8b). Surprisingly the damping sets 78 and 85 produce contrary results in terms of the WBE at the two gauges (Fig. 7f). The model runs based on the two parameter sets 78 and 85 show water balance percentage errors of very low 0.95 % and 3.49 % at WEN and high 48.1 % and 52.9 % at AME.

Parameter set 60 in the ensemble of 91 FAST-mHM sets frequently produces highest NSE model efficiency values and reaches 0.75 at WEN and 0.64 at AME, respectively. Generally, NSE values do not exceed the 0.77 mark at all headwater gauges, constituting an acceptable adjustment, if kept in mind that the six considered parameters were not calibrated but varied according to FAST sensitivity analysis. Nevertheless, certain inadequacies can be observed, especially high flow peaks are commonly not properly matched, low flow recessions are underestimated or model timing errors can be detected (Fig. 8a and Fig. 8b), preventing higher efficiencies, likewise for the results based on behavioural parameter sets.

5.4 Parameter identification via sensitivities

5.4.1 Sensitivity reference

The application of parameter sensitivities as a means to identify and focus more specifically on the most influential subset of the parameter space constitutes the third column of the approach. Global sensitivity analysis with FAST provides sensitivity index values (see Table 1 for sensitivity ranges) for the six analysed mHM parameters and 14 considered headwaters. The sum over all six sensitivity values per timestep reaches a maximum value of 0.43 at HER and can, in principal, maximally equal 1.0 for the entire parameter space (Cukier et al., 1975; Reusser et al., 2011), i.e. for the whole of 52 global mHM parameters in this case.

For every timestep the observed streamflow value can be related to sensitivity values assigned to each model parameter. Time dependent graphs are depicted in Fig. 9a and Fig. 9b showing sensitivity dynamics for the six mHM parameters at the gauges WEN and HUE in the hydrologic year 2009/10. WEN and HUE were selected as representatives of two major groups showing essentially distinct sensitivity properties of parameters. From a preliminary result of visual inspection and in accordance with the findings in section 5.3, the model performance is obviously higher for the larger WEN catchment and in an acceptable range, in contrast to the poor ensemble simulation results at HUE shown in Fig. 9c. For that reason and for reasons of legibility,



some of the plots presented in this section relating streamflow to sensitivity indices were depicted for both gauges in the case of observed streamflow but for the ensemble simulations only for gauge WEN.

Sensitivity hydrographs plotted as solid lines for WEN in Fig. 9a-b indicate some kind of reverse patterns between some of the six parameters, while AspectcorrPET, representative of others, and correcting potential evapotranspiration data aspect specifically, shows an opposite sensitivity behaviour to RechargeCoeff, determining the percolation rate to the groundwater reservoir. The former one allows the inclusion of the exposition of slopes, controlling insulation, into evapotranspiration estimations. It essentially influences the water balance - and thus streamflow behaviour -, whereby, as expected, highest sensitivity values are reached in summer periods. Conversely, RechargeCoeff exhibits its largest influence on streamflow generation during winter, with negligible evapotranspiration rates and substantial amounts of groundwater recharge in this region.

In the next instance, at each modelled timestep, FAST parameter sensitivity indices were ranked in descending order to obtain parameters with the highest influence on the generation of streamflow for every simulated day. The most sensitive parameter at each timestep was then related to observed streamflow hydrographs plotted in Fig. 10 together with the FAST-mHM ensemble. The hydrograph for WEN is mainly dominated by the colours of the two opposing parameters AspectcorrPET (blue) and RechargeCoeff (orange). Indeed, shortly recessed by soil moisture parameters ThetaSDb (light green), OMcontforest (green) or Ksconst (light blue), both in the summer period of 2010 and on the limbs of flow recessions, AspectcorrPET, controlling evapotranspiration, has the highest influence on simulated streamflow generation. During the winter season RechargeCoeff reveals the longest periods of highest sensitivity values where percolation and groundwater recharge dominates.

At streamflow gauge HUE essentially distinct sensitivity properties of parameters can be observed. Hydrograph sequences of the highest parameter sensitivities are only attained by Ksconst, AspectcorrPET and especially ThetaSconst, a global soil moisture pedotransfer-parameter used to estimate water content at soil saturation in order to calculate the soil water content via the equation proposed by van Genuchten (1980) and discussed by Zacharias and Wessolek (2007), seem to strongly influence streamflow dynamics at this site, interacting inversely with Ksconst (Fig. 9 and Fig. 10).

Parameter sensitivities were also related to FDCs to sequentially evaluate portions of the streamflow spectrum and to identify process-relevant parameters at different flow magnitudes. Four sequences with different intervals of streamflow exceedance for the observed (WEN and HUE) and simulated ensemble (WEN) of FDCs are pictured in Fig. 11a-d. Observed FDCs, provided with parameter information of highest sensitivity at every point of flow, give an insight into differences of dependent parameter sensitivity for the whole spectrum of streamflow, revealing regional differences in both model and catchment functioning at distinct headwaters.

For WEN, RechargeCoeff clearly dominates the first five percent of streamflow exceedance with highest flow values (Fig. 11a), whereas the evapotranspiration parameter AspectcorrPET is most influential on runoff generation in the lowest 5 percent of



flow (Fig. 11d). In the large mid section of the FDC for WEN (Fig. 11b and Fig. 11c) a rather heterogeneous composition and permanent change in the highest parameter sensitivity index values becomes obvious. In the headwater catchment of HUE the soil moisture parameter ThetaSconst predominates in high and low flow sequences as well as partly so in the mid sections, mostly alternating with Ksconst, another soil moisture regression parameter, to estimate the saturated vertical hydraulic conductivity (Cosby et al., 1984) within field capacity calculations.

Although some of the parameters named above frequently show the highest values of partial parameter sensitivity, the influence on model-generated streamflow is certainly a complex interplay of different sensitive parameters, which act as a team and conceptually reflect the diverse composition of physiographic and climatic factors affecting the natural hydrologic functioning of catchments at various scales. Nevertheless, the reference to parameters with the highest partial sensitivity provides a feasible means to set the focus more precisely on calibration parameters having essential influences on streamflow generation.

5.4.2 Sensitivity duration

Besides the reference of highest parameter sensitivity values to streamflow hydrographs and FDCs we also employed Sensitivity Duration Curves (SDCs) to be able to further specify the parameter identification process. Analogous to FDCs, a SDC is a cumulative frequency curve that each applies to one parameter, to one (type of) catchment and to the period in which sensitivity analysis is performed. SDCs were developed by arranging the assigned daily FAST sensitivity index values according to their magnitude in ascending order and by plotting as a line against the percentage of time during which the sensitivity equalled or exceeded the specified values. The implementation of SDCs helps both to identify varying functions of parameters more easily and to better concentrate on parameters with higher sensitivity duration, exhibiting more substantial influences on simulation results.

SDCs are plotted in Fig. 12a-d for four of six analysed mHM parameters at each of the 14 headwater gauges covering the simulation period from 2003 to 2011. OMcontforest, which describes the portion of organic matter (OM) in soil to account for effects of organic matter on soil water retention characteristics in forested areas (Zacharias and Wessolek, 2007), shows a similar sensitivity behaviour in all 14 headwaters (Fig. 12a). The highest sensitivities are reached during intermediate streamflow conditions at rising and falling limbs (see also Fig. 9 and Fig. 10). A maximum sensitivity value of 0.176 is reached in the headwater catchment of MOE with the highest percentage of forest cover (87.3 %) (Fig. 2a), which indicates the direct connection of the parameter to forest as a significant land cover type in this region. Nevertheless, large periods exist when streamflow generation is not crucially influenced by this predictor, in our case with soils showing low OM: Zacharias and Wessolek (2007) concluded likewise, even proposing to renounce OM as an input variable, marginally affecting simulation output.

AspectcorrPET (Fig. 12b) shows similar progressions of sensitivity values at all gauges within the lower 50 %. Conversely, from this point on and for the upper half of the curves, a spread in SDCs can be recognised. For some of the headwaters the



slopes of the SDC decrease and the highest sensitivities do not exceed 0.08, whereas the rest of the SDCs keep slopes constant beyond the split point and reach, with a maximum of 0.132 for WEN, generally higher magnitudes of sensitivity index values. Especially in the south-western headwaters (e.g. BOE or HUE) with lower relief energy, the aspect as a physiographic predictor seems to exhibit smaller influences on evapotranspiration and water balance than in catchments showing a bold relief, with 5 valleys more deeply incised where the factor of exposure dominates evapotranspiration processes.

This split picture is also confirmed by Fig. 12c and Fig. 12d where SDCs are plotted for ThetaSconst and RechargeCoeff. For the south-western catchments, including the largest headwater BAM where a more intermediate sensitivity behaviour shows up, ThetaSconst (Fig. 12c) has an increased influence in large parts of the spectrum, while RechargeCoeff (Fig. 12d) 10 seems to be steadily more sensitive in the eastern and northern headwaters. For the reproduction of observed streamflow, soil water retention characteristics along with runoff generation from the unsaturated zone, seem to play a major role in the former rather than in the latter headwaters where the percolation into the deeper saturated zone and subsequent baseflow prevail in the interplay of hydrologic processes.

15 6 Discussion and conclusions

The main objective of this study was to approach three embedded research questions, whether (1) fingerprints of streamflow response characteristics can act as constraints on the parameter space of a distributed hydrologic model, (2) regional differences in parameter sensitivity can serve as an indicator to learn about spatial differences in catchment functioning and (3) differences in hydrologic catchment behaviour and parameter sensitivities can be applied as explanatory variables for physiographic or 20 climate attributes.

(1) The approach based on six streamflow fingerprints, representing the observed catchment response functioning of 14 Ruhr headwaters, allowed us to effectively constrain the parameter space of a distributed hydrologic model to the most behavioural subset in terms of the specific reproduction of these indices. Observed response fingerprints were found to cluster within 25 typical catchment sample regions. The respective ensemble of simulated counterparts, initially model-specifically grouped, was successfully constrained to the observed sample regions of the hydrologic system. We further found most of the behavioural parameter sets within the constrained subsets to frequently change along with each fingerprint and location. The singularity of parameter sets for fingerprints or location is partly attenuated if two fingerprints are considered as constraint which we defined as a measure of the minimal distance in the two fingerprint sample space. Streamflow hydrographs based both on the combined 30 and fingerprint specific behavioural parameter sets often show particular dynamics with varying or even opposed results if standard efficiency criteria are considered.

These differences might be partially explained by the fact that, contrary to the three chosen measures, dynamic response fingerprints aim at absolute performance values of a specific hydrologic functioning, less normalised than e.g. the NSE, with



higher explanatory value, objectively interpretable, which was also noted e.g. by Schaeffli and Gupta (2007). Notwithstanding, results in model performance (NSE, ρ_s and WBE) of parameter sets, outlined in Fig. 7, can be (highly) improved or further corrupted by randomly selecting parameter sets from the group of the five most behavioural ones both within the same catchment (e.g. WEN) or by transferring the selected set to another site (AME).

5

(2) Region specific differences were not only detected among the dynamic response fingerprints but also for dependent parameter sensitivity by employing FAST on the mesoscale Hydrologic Model. Headwaters can be grouped by sensitivity properties of parameters into two major categories, that also exhibit high and low (five southwestern headwaters) model performances in our example. Parameter effects (sensitivities) on the hydrologic response behaviour can be detected far easier by the option of a reference for streamflow hydrographs or flow duration curves by highest sensitivities. The composition of parameters identified as the most influential for streamflow sequences at different points in time or at variable magnitudes (free of autocorrelation and time dependency) was greater in number and variety in headwaters where the distributed model was found to perform reasonably well in terms of streamflow dynamics and response fingerprints. Sensitivity dynamics were also significantly higher in the eastern headwaters. Sensitivity duration curves allowed us to distinguish differences in parameter sensitivity exceedance probability progress between the major groups. SDCs provided a convenient means for studying varieties of sensitivity characteristics of parameters at sites often revealing crucial hydrologic differences.

These local differences in sensitivity dynamics, variety and duration of parameters might be partly triggered by model independent uncertainties in the spatial distribution of meteorologic forcing data (Song et al., 2013), but can also be interpreted as hint of different hydrologic behaviour caused by factors such as physiographic settings. The RechargeCoeff as a conceptual proxy of catchment structures controlling percolation and groundwater recharge either temporarily exhibits high influences on streamflow generation or solely plays a tangential role which in the southwestern headwaters is compensated for instead by proxies for soil moisture dynamics (Ksconst) or evapotranspirative processes (AspectcorrPET) (Fig. 9a-b and 10). Soil moisture controlled conditions can be associated with the generation of fast reacting or slightly deferred components of streamflow which dominate during high flow periods. Response fingerprints describing the natural characteristics of highflow events (HPC and HFD) and the variability of streamflow (CV) showed highest values in the five southwestern catchments (Fig. 2b). The ranking of highest parameter sensitivities seems to generally capture this special local functioning of flashy response while physiographic characteristics (Fig. 2a) do not intrinsically differ or cluster apart from the eastern group of headwaters.

(3) A selection comprising the most integral catchment attributes or reliable derivatives (Fig. 2a) was set up within catchment functional categories (Table 2) serving as possible structural and climatic similarity metrics for later cluster analysis. Physical descriptors were primarily employed to explore basic linkages to potentially corresponding dynamic response fingerprints connecting observable with predictable hydrologic response variables.



Specifically we found that four highest SLFDC values were found in headwaters (MOE, HER, KIC and BAM), characterised by both highest percentage of forest land cover and, with the exception of MOE, steepest average catchment slopes. Other dynamic response fingerprints plot at values close to the overall average of all headwaters which pictures a jointly moderate hydrologic response character among these four catchments in terms of high flow response functioning. In a qualitative sense and in this specific case, the vast forest coverage (damping) and marked relief intensity via higher slope averages (stimulating) in the catchment seem to play a certain role for the rate of change in average streamflow events whereas catchment form descriptors such as the longest drainage path (LDP) (Fig. 6c) or the topographic descriptor flow accumulation (FACC) (Fig. 6d) do not seem to have further explanatory value.

Since the nature of hydrologic response is generally composed of and generated by the complex interplay of multiple catchment specific factors (see section 1.3), the explanation of a categoric dynamic fingerprint value at a specific site by one physiographic or climate attribute is crucially hampered. Site and fingerprints specific most behavioural parameter sets further underline the heterogeneity in the interplay of catchment conditions determining hydrologic processes and consequently complicate a reliable model-based mapping of streamflow and its derivatives. As also discussed in section 1.3, conclusions about hydrologic similarities cannot be easily drawn if the process behaviour represented by dynamic response fingerprints does not conform to causative physiographic catchment descriptors impeding model parameter estimation, not to mention parameter transfer based on feasible regionalisation representable by the given model structure (section 1.3). However, global sensitivity analysis spanning an ensemble of parameter combinations that reproduces streamflow of headwaters in many different ways can give further support in the explanation of similar or distinct response functioning.

This paper has implemented some promising steps toward an analytic framework to (1) investigate hydrologically relevant structural and functional attributes in terms of consistency and feasibility in classifying similar catchments, (2) assess the value of functional constraints for the parameter spaces of distributed hydrologic models, (3) diagnose and explain regional and temporal differences of model dynamics and performance based on global sensitivity analysis. The concept and systematic approach thus presented, along with our findings in this study, motivate starting points for additional research and some methodological revision.

(1) A lack of diversity in terms of some of the absolute values of physiographic fingerprints (e.g. wetness index) both unveils similar structural conditions or forcings of proximal locations that helps to sort out redundant catchment characteristics but also limits the scope within this study area for catchment classification purposes. Contrarily, other physiographic indices such as the forest cover or the average catchment slope were better able to capture important regional characteristics. Scale dependent limitations have to be kept in mind to avoid a levelling out of the explanatory value for a physiographic index (Blöschl and Sivapalan, 1995). Cluster analysis, including mesoscale headwaters with different physiographic and climatic settings, certainly further enables classification by structural similarities of watersheds, which is an imminent development. Some further fingerprints should replace those found to be redundant in order to capture all hydrologically behavioural components relevant



within the functions of watersheds (Black, 1997) such as the recession behaviour of streamflow or by indices that evaluate timing errors of simulated streamflow (see e.g. Yilmaz et al. (2008)). This step might further clarify potential linkings of functional and structural characteristics and could implicate an additional gain in model performance with regard to hydrograph dynamics and will add explanatory value for sequential calibration on different streamflow magnitudes.

5

(2) The consideration of combined fingerprint calibration constraints led to a higher degree of consistency in parameter sets which can serve as a feasible trade-off to map physiographic and functional heterogeneity by the given model structural concept. Behavioural parameter vectors can then be classified by their effect on streamflow dynamics within specific physiographic settings with reference to partial sensitivities in the set.

10

(3) In addition, more parameters maybe analysed within FAST to raise the sum of sensitivity index values per timestep significantly, explaining an even larger majority of variances in streamflow. This step involves the ensemble spectrum to be widened, potentially leading to more behavioural solutions involving the risk for equifinality. Global sensitivity analysis can be directly applied to catchment class-specific dynamic fingerprints, in order to establish feasible relationships of individual

15 (vectors of) parameters and functional attributes.

van Griensven et al. (2006) similarly remarked that local differences indicate that results of global sensitivity analysis for one (group of) catchments cannot be applied directly to other sometimes even closely located places but maybe used as a reasonable estimate within the same catchment category of parameter sensitivity. Therefore, classification by sensitivity duration and

20 dynamics are further conceivable and required as subsequent steps. As correctly recommended by (Song et al., 2013) for a complete analysis of parameter effects on model output, influences due to meteorologic forcings might also be assessed in order to employ the multi-fingerprint approach to ungauged sites or those under substantial anthropogenic influences.

Acknowledgements. We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.



References

- Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., and Capell, R.: A comparison of similarity indices for catchment classification using a cross-regional dataset, *Advances in Water Resources*, 40, 11–22, doi:10.1016/j.advwatres.2012.01.008, 2012.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment Part I: Model development, *Journal Of The American Water Resources Association*, 34, 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.
- 5 Bárdossy, a.: Calibration of hydrological model parameters for ungauged catchments, *Hydrology and Earth System Sciences Discussions*, 3, 1105–1124, doi:10.5194/hessd-3-1105-2006, 2006.
- Beven, K.: Changing ideas in hydrology — The case of physically based models — Comment, *Journal of Hydrology*, 120, 405–407, doi:10.1016/0022-1694(90)90161-P, <http://linkinghub.elsevier.com/retrieve/pii/002216949090161P>, 1990.
- 10 Beven, K.: How far can we go in distributed hydrological modelling?, *Hydrology and Earth System Sciences*, 5, 1–12, doi:10.5194/hess-5-1-2001, <http://www.hydrol-earth-syst-sci.net/5/1/2001/hess-5-1-2001.html>, 2001.
- Beven, K.: A manifesto for the equifinality thesis, *Journal of Hydrology*, 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249, 11–29, doi:10.1016/S0022-1694(01)00421-8, 2001.
- 15 Binley, A. and Beven, K.: Vadose zone flow model uncertainty as conditioned on geophysical data, *Ground Water*, 41, 119–127, 2003.
- Black, P.: Watershed functions', *Journal Of The American Water Resources Association*, 33, 1997.
- Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resources Research*, 45, 1–12, doi:10.1029/2007WR006726, 2009.
- Blöschl, G. and Sivapalan, M.: Scale Issues in Hydrological Modeling, *Hydrological Processes*, 9, 251–290, doi:0885-6087/95/03025 1-40, 20 1995.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. H. G.: Runoff prediction in ungauged basins, 2013.
- Bode, H., Evers, P., and Albrecht, D. R.: Integrated water resources management in the Ruhr River Basin, Germany, *Water Science and Technology*, 47, 81–86, 2003.
- Boorman, D., Hollis, J., and Lilly, A.: Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom., *Tech. Rep. 126*, Institute of Hydrology, <http://nora.nerc.ac.uk/7369/>, 1995.
- Brudy-Zippelius, T.: Wassermengenbewirtschaftung im Einzugsgebiet der Ruhr : Simulation und Echtzeitbetrieb, Ph.D. thesis, Universität Karlsruhe (TH), 2003.
- Carrillo, G., Troch, P. a., Sivapalan, M., Wagener, T., Harman, C., and Sawicz, K.: Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrology and Earth System Sciences*, 15, 3411–3430, doi:10.5194/hess-15-3411-2011, 2011.
- 30 Castellarin, a., Burn, D., and Brath, a.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *Journal of Hydrology*, 241, 270–285, doi:10.1016/S0022-1694(00)00383-8, <http://linkinghub.elsevier.com/retrieve/pii/S0022169400003838>, 2001.
- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Advances in Water Resources*, 33, 1235–1242, doi:10.1016/j.advwatres.2010.04.009, <http://dx.doi.org/10.1016/j.advwatres.2010.04.009>, 2010.
- 35 Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, *Water Resources Research*, 20, 682–690, 1984.



- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, a. G., and Schaibly, J. H.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, *Journal of Chemical Physics*, 59, 3873, doi:10.1063/1.1680571, <http://link.aip.org/link/?JCP/59/3873/1>{&}Agg=doi, 1973.
- Cukier, R. I., Schaibly, J. H., and Shuler, K. E.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations Theory, *Journal of Chemical Physics*, 63, 1140, doi:10.1063/1.431440, <http://link.aip.org/link/?JCP/59/3873/1>{&}Agg=doi, 1975.
- Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S., and Samaniego, L.: Computationally inexpensive identification of noninformative model parameters by sequential screening, *Water Resources Research*, 51, 6417–6441, doi:10.1002/2015WR016907, 2015.
- 10 Di Prinzio, M., Castellarin, a., and Toth, E.: Data-driven catchment classification: Application to the pub problem, *Hydrology and Earth System Sciences*, 15, 1921–1935, doi:10.5194/hess-15-1921-2011, 2011.
- Duband, D., Oblé, C., and Rodriguez, J.: Unit hydrograph revisited: an alternate iterative approach to UH and effective precipitation identification, *Journal of Hydrology*, 150, 115–149, doi:10.1016/0022-1694(93)90158-6, 1993.
- Dyck, S. and Peschke, G., eds.: *Grundlagen der Hydrologie*, Verlag für Bauwesen, Berlin, 3 edn., 1995.
- 15 Eckhardt, K.: A comparison of baseflow indices, which were calculated with seven different baseflow separation methods, *Journal of Hydrology*, 352, 168–173, doi:10.1016/j.jhydrol.2008.01.005, 2008.
- Eckhardt, K.: Technical note: Analytical sensitivity analysis of a two parameter recursive digital baseflow separation filter, *Hydrology and Earth System Sciences*, 16, 451–455, doi:10.5194/hess-16-451-2012, 2012.
- Fenicia, F.: Understanding catchment behavior through stepwise model concept improvement, Ph.D. thesis, Delft University of Technology (TU Delft), doi:10.1029/2006WR005563, 2008.
- Fenicia, F., Savenije, H. H. G., and Winsemius, H. C.: Moving from model calibration towards process understanding, *Physics and Chemistry of the Earth*, 33, 1057–1060, doi:10.1016/j.pce.2008.06.008, 2008.
- Fontaine, D. D., Havens, P. L., Blau, G. E., and Tillotson, M.: Groundwater Risk Modeling for Pesticides¹, *Weed Technology*, 6, 716–724, 1992.
- 25 Frey, H. C. and Patil, S. R.: Identification and review of sensitivity analysis methods, *Risk Analysis*, 22, 553–578, doi:10.1111/0272-4332.00039, 2002.
- Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: Investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, *Hydrology and Earth System Sciences*, 15, 3275–3291, doi:10.5194/hess-15-3275-2011, 2011.
- 30 Göppert, H., Ihringer, J., Plate, E. J., and Morgenschweis, G.: Flood forecast model for improved reservoir management in the Lenne River catchment, Germany, *Hydrological Sciences Journal*, 43, 215–242, doi:10.1080/02626669809492119, 1998.
- Göttinger, J. and Bárdossy, a.: Comparison of four regionalisation methods for a distributed hydrological model, *Journal of Hydrology*, 333, 374–384, doi:10.1016/j.jhydrol.2006.09.008, 2007.
- Guse, B., Reusser, D. E., and Fohrer, N.: How to improve the representation of hydrological processes in SWAT for a lowland catchment - temporal analysis of parameter sensitivity and model performance, *Hydrological Processes*, 28, 2651–2670, doi:10.1002/hyp.9777, 2014.
- 35 Haude, W.: Zur Bestimmung der Verdunstung auf möglichst einfache Weise, *Mitt. Dtsch. Wetterd.*, 2, 35 818, 1955.
- He, Y., Bárdossy, a., and Zehe, E.: A catchment classification scheme using local variance reduction method, *Journal of Hydrology*, 411, 140–154, doi:10.1016/j.jhydrol.2011.09.042, 2011.



- Hirsch, R. M.: An evaluation of some record reconstruction techniques, *Water Resources Research*, 15, 1781, doi:10.1029/WR015i006p01781, 1979.
- Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *Journal of Hydrology*, 292, 281–295, doi:10.1016/j.jhydrol.2004.01.002, 2004.
- 5 Kleidon, A., Zehe, E., Ehret, U., and Scherer, U.: Thermodynamics, maximum power, and the dynamics of preferential river flow structures at the continental scale, *Hydrology and Earth System Sciences*, 17, 225–251, doi:10.5194/hess-17-225-2013, <http://www.hydrol-earth-syst-sci.net/17/225/2013/>, 2013.
- Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, *Journal of Hydrology*, 392, 54–69, doi:10.1016/j.jhydrol.2010.07.047, <http://dx.doi.org/10.1016/j.jhydrol.2010.07.047>, 2010.
- 10 Ley, R., Casper, M. C., Hellebrand, H., and Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM), *Hydrology and Earth System Sciences*, 15, 2947–2962, doi:10.5194/hess-15-2947-2011, 2011.
- Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resources Research*, 43, 1–18, doi:10.1029/2006WR005756, 2007.
- 15 Lu, Y. C. and Mohanty, S.: Sensitivity Analysis of a complex, proposed geological waste disposal system using the Fourier Amplitude Sensitivity Test Method, *Reliability Engineering and System Safety*, 72, 275–291, 2001.
- McDonnell, J. J. and Woods, R.: On the need for catchment classification, *Journal of Hydrology*, 299, 2–3, doi:10.1016/j.jhydrol.2004.09.003, 2004.
- Merz, R.: Methoden der hydrologischen Regionalisierung, *Wiener Mitteilungen*, 197, 109–130, 2006.
- 20 Merz, R. and Blöschl, G.: A process typology of regional floods, *Water Resources Research*, 39, 1–20, doi:10.1029/2002WR001952, 2003.
- Merz, R. and Blöschl, G.: Flood frequency regionalisation - Spatial proximity vs. catchment attributes, *Journal of Hydrology*, 302, 283–306, doi:10.1016/j.jhydrol.2004.07.018, 2005.
- Monin, A. and Obukhov, A.: Basic laws of turbulent mixing in the surface layer of the atmosphere, *Contrib. Geophys. Inst. Acad. Sci. USSR*, 24, 163–187, 1954.
- 25 Morgenschweis, G., Brudy-Zippellus, T., and Ihringer, J.: Operational water quality management in a river basin, *Water Science and Technology*, 48, 111–118, 2003.
- Nash, E. and Sutcliffe, V.: River flow forecasting through conceptual models part I- A discussion of principles, *Journal of Hydrology*, 10, 282–290, 1970.
- Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, *River Research and Applications*, 19, 101–121, doi:10.1002/rra.700, 2003.
- 30 Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resources Research*, 44, 1–15, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resources Research*, 46, 1–15, doi:10.1029/2009WR008887, 2010.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrology and Earth System Sciences Discussions*, 2, 509–542, doi:10.5194/hessd-2-509-2005, 2005.



- Patil, S. and Stieglitz, M.: Controls on hydrologic similarity: Role of nearby gauged catchments for prediction at an ungauged catchment, *Hydrology and Earth System Sciences*, 16, 551–562, doi:10.5194/hess-16-551-2012, 2012.
- Plate, E. J., Ihringer, J., and Lutz, W.: Operational models for flood calculations, *Journal of Hydrology*, 100, 489–506, 1988.
- Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology - Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), *Meteorologische Zeitschrift*, 22, 235–256, doi:10.1127/0941-2948/2013/0436, 2013.
- Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, *Remote Sensing of Environment*, 112, 3469–3481, doi:10.1016/j.rse.2008.03.018, 2008.
- 10 Reusser, D.: Implementation of the Fourier Amplitude Sensitivity Test (FAST), 2014.
- Reusser, D. E.: Combining smart model diagnostics and effective data collection for snow catchments, 2010.
- Reusser, D. E. and Zehe, E.: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resources Research*, 47, n/a–n/a, doi:10.1029/2010WR009946, <http://doi.wiley.com/10.1029/2010WR009946>, 2011.
- Reusser, D. E., Buytaert, W., and Zehe, E.: Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test, *Water Resources Research*, 47, doi:10.1029/2010WR009947, 2011.
- 15 Rodríguez-Camino, E. and Avissar, R.: Comparison of three land-surface schemes with the Fourier amplitude sensitivity test (FAST), *Tellus, Series A: Dynamic Meteorology and Oceanography*, 50, 313–332, doi:10.3402/tellusa.v50i3.14529, 1998.
- Ruhrverband: Ruhrwassermenge 2011, Tech. rep., Ruhrverband, Essen, 2011.
- Ruhrverband: Ruhrwassermenge 2013, Tech. rep., Ruhrverband, Essen, 2013.
- 20 Saltelli, a., Tarantola, S., and Chan, K.: A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, 41, 39–56, doi:10.1080/00401706.1999.10485594, <http://amstat.tandfonline.com/doi/full/10.1080/00401706.1999.10485594>, 1999.
- Samaniego, L., Bárdossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, *Water Resources Research*, 46, doi:10.1029/2008WR007695, 2010a.
- 25 Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, 46, 1–25, doi:10.1029/2008WR007327, 2010b.
- Samaniego, L., Kumar, R., and Jackisch, C.: Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data, *Hydrology Research*, 42, 338–355, doi:10.2166/nh.2011.156, <http://www.iwaponline.com/nh/042/nh0420338.htm>, 2011.
- 30 Samaniego, L., Cuntz, M., Craven, J., Dalmaso, G., Kumar, R., Mai, J., Musuza, J., Prykhodko, V., Schneider, C., Spieler, D., Thober, S., and Zink, M.: multiscale Hydrologic Model mHM -Documentation for version 5.1., Tech. rep., Helmholtz Centre for Environmental Research - UFZ, 2014.
- Sanadhya, P., Gironás, J., and Arabi, M.: Global sensitivity analysis of hydrologic processes in major snow-dominated mountainous river basins in Colorado, *Hydrological Processes*, 28, 3404–3418, doi:10.1002/hyp, <http://jamsb.austms.org.au/courses/CSC2408/semester3/resources/ldp/abs-guide.pdf>, 2013.
- 35 Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. a., and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrology and Earth System Sciences*, 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.



- Schaefli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrological Processes*, 21, 2075–2080, doi:10.1002/hyp.6825, <http://jamsb.austms.org.au/courses/CSC2408/semester3/resources/ldp/abs-guide.pdf>, 2007.
- Schaibly, J. H. and Shuler, K. E.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications, *Journal of Chemical Physics*, 59, 3879, doi:10.1063/1.1680572, <http://link.aip.org/link/?JCP/59/3873/1> {&}Agg=doi, 1973.
- 5 Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, *Water Resources Research*, 41, 1–16, doi:10.1029/2004WR003409, 2005a.
- Shamir, E., Imam, B., Morin, E., Gupta, H. V., and Sorooshian, S.: The role of hydrograph indices in parameter estimation of rainfall-runoff models, *Hydrological Processes*, 19, 2187–2207, doi:10.1002/hyp.5676, 2005b.
- Singh, R., Archfield, S. a., and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged
10 catchments - A comparative hydrology approach, *Journal of Hydrology*, 517, 985–996, doi:10.1016/j.jhydrol.2014.06.030, <http://dx.doi.org/10.1016/j.jhydrol.2014.06.030>, 2014.
- Snelder, T. H., Cattaneo, F., Suren, A. M., and Biggs, B. J. F.: Is the River Environment Classification an improved landscape-scale classification of rivers?, *Journal of the North American Benthological Society*, 23, 580–598, doi:10.1899/0887-3593(2004)023<0580:ITRECA>2.0.CO;2, 2004.
- 15 Sobol, I. M.: Sensitivity Estimates for Nonlinear Mathematical Models, *Mathematical Modeling and Computation*, 1, 407–414, 1993.
- Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, 55, 271–280, doi:10.1016/S0378-4754(00)00270-6, 2001.
- Song, X., Bryan, B. a., Almeida, A. C., Paul, K. I., Zhao, G., and Ren, Y.: Time-dependent sensitivity of a process-based ecological model, *Ecological Modelling*, 265, 114–123, doi:10.1016/j.ecolmodel.2013.06.013, 2013.
- 20 Thiessen, A. H.: Precipitation Averages for Large Areas, *Monthly Weather Review*, 39, 1082–1089, doi:10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2, 1911.
- Tolson, B. a. and Shoemaker, C. a.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, 43, 1–16, doi:10.1029/2005WR004723, 2007.
- van Genuchten, M. T.: A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1,
25 doi:10.2136/sssaj1980.03615995004400050002x, 1980.
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of multi-variable catchment models, *Journal of Hydrology*, 324, 10–23, doi:10.1016/j.jhydrol.2005.09.008, 2006.
- Vandewiele, G. and Elias, A.: Monthly water balance of ungauged catchments obtained by geographical regionalization, *Journal of Hydrology*, 170, 277–291, doi:10.1016/0022-1694(95)02681-E, 1995.
- 30 Vrugt, J. a. and Sadegh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resources Research*, 49, 4335–4345, doi:10.1002/wrcr.20354, 2013.
- Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1, 1–31, doi:10.1111/j.1749-8198.2007.00039.x, 2007.
- Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C. Y.: Calibration of hydrological
35 models using flow-duration curves, *Hydrology and Earth System Sciences*, 15, 2205–2227, doi:10.5194/hess-15-2205-2011, 2011.
- Wienhöfer, J. and Zehe, E.: Predicting subsurface stormflow response of a forested hillslope - the role of connected flow paths, *Hydrology and Earth System Sciences*, 18, 121–138, doi:10.5194/hess-18-121-2014, <http://www.hydrol-earth-syst-sci.net/18/121/2014/>, 2014.



- Winter, T. C.: The concept of hydrologic landscapes', *Journal Of The American Water Resources Association*, 37, 335–349, doi:10.1111/j.1752-1688.2001.tb00973.x, <http://doi.wiley.com/10.1111/j.1752-1688.2001.tb00973.x>, 2001.
- Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Advances in Water Resources*, 30, 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007.
- 5 Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resources Research*, 44, 1–18, doi:10.1029/2007WR006716, 2008.
- Yu, P. S. and Yang, T. C.: Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites, *Hydrological Processes*, 14, 117–133, doi:10.1002/(SICI)1099-1085(200001)14:1<117::AID-HYP914>3.0.CO;2-Q, 2000.
- Zacharias, S. and Wessolek, G.: Excluding Organic Matter Content from Pedotransfer Predictors of Soil Water Retention, *Soil Science Society of America Journal*, 71, 43, doi:10.2136/sssaj2006.0098, 2007.
- 10 Zajac, Z. B.: Global sensitivity and uncertainty analysis of spatially distributed watershed models, Ph.D. thesis, University of Florida, 2010.
- Zehe, E., Ehret, U., Blume, T., Kleidon, A., Scherer, U., and Westhoff, M.: A thermodynamic approach to link self-organization, preferential flow and rainfall–runoff behaviour, *Hydrology and Earth System Sciences*, 17, 4297–4322, doi:10.5194/hess-17-4297-2013, <http://www.hydrol-earth-syst-sci.net/17/4297/2013/>, 2013.
- 15 Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, *Hydrology and Earth System Sciences*, 18, 4635–4655, doi:doi:10.5194/hess-18-4635-2014, <http://www.hydrol-earth-syst-sci-discuss.net/11/3249/2014/>, 2014.
- 20 Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resources Research*, 44, 1–13, doi:10.1029/2008WR006833, 2008.

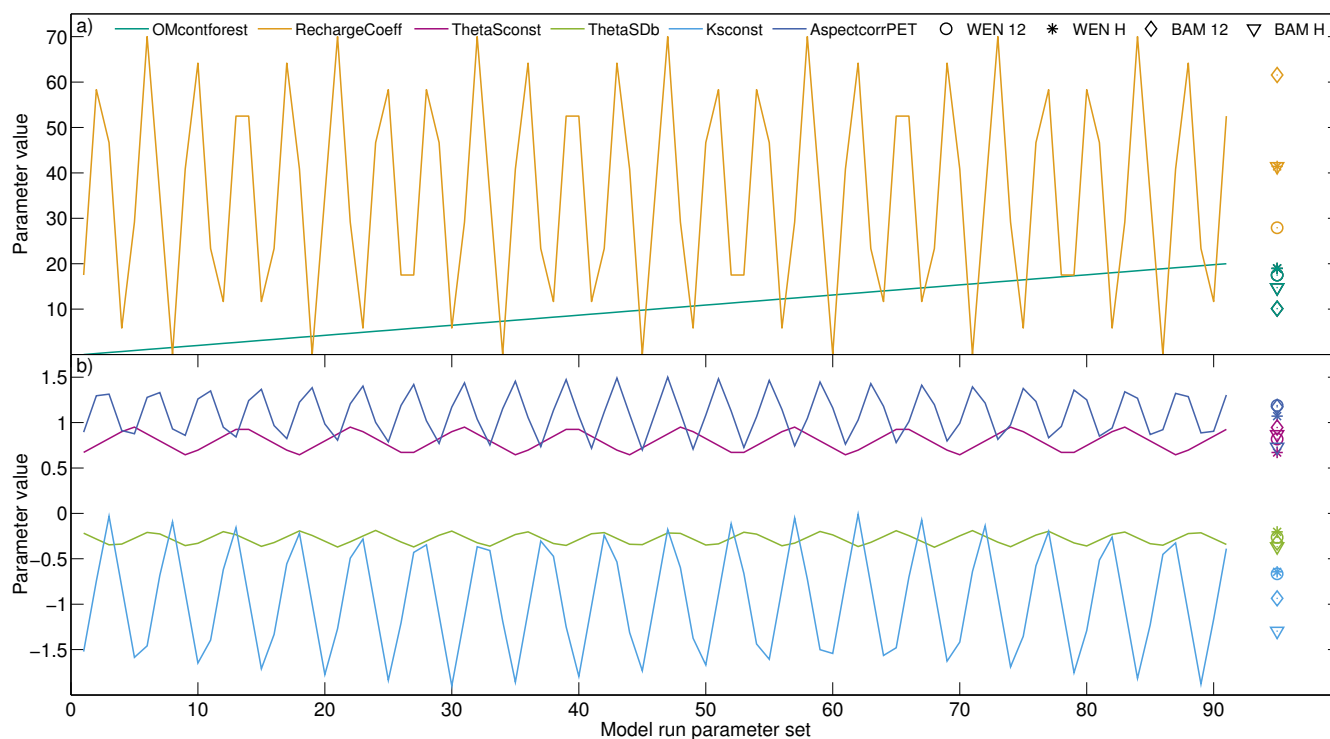


Figure 1. 91 FAST variations of the six selected global mHM parameters (see Table 1) plotted as connected curves and mHM parameter values obtained by calibration using the dynamically dimensioned search (DDS) algorithm at gauges Bamenohl (BAM) (\diamond , ∇) and Wenholthausen (WEN) (\circ , $*$) (see Fig. 4 for gauge locations). DDS calibration was based on two different precipitation inputs: 12 DWD stations with Thiessen polygons (12) and HYRAS gridded precipitation data set (H).

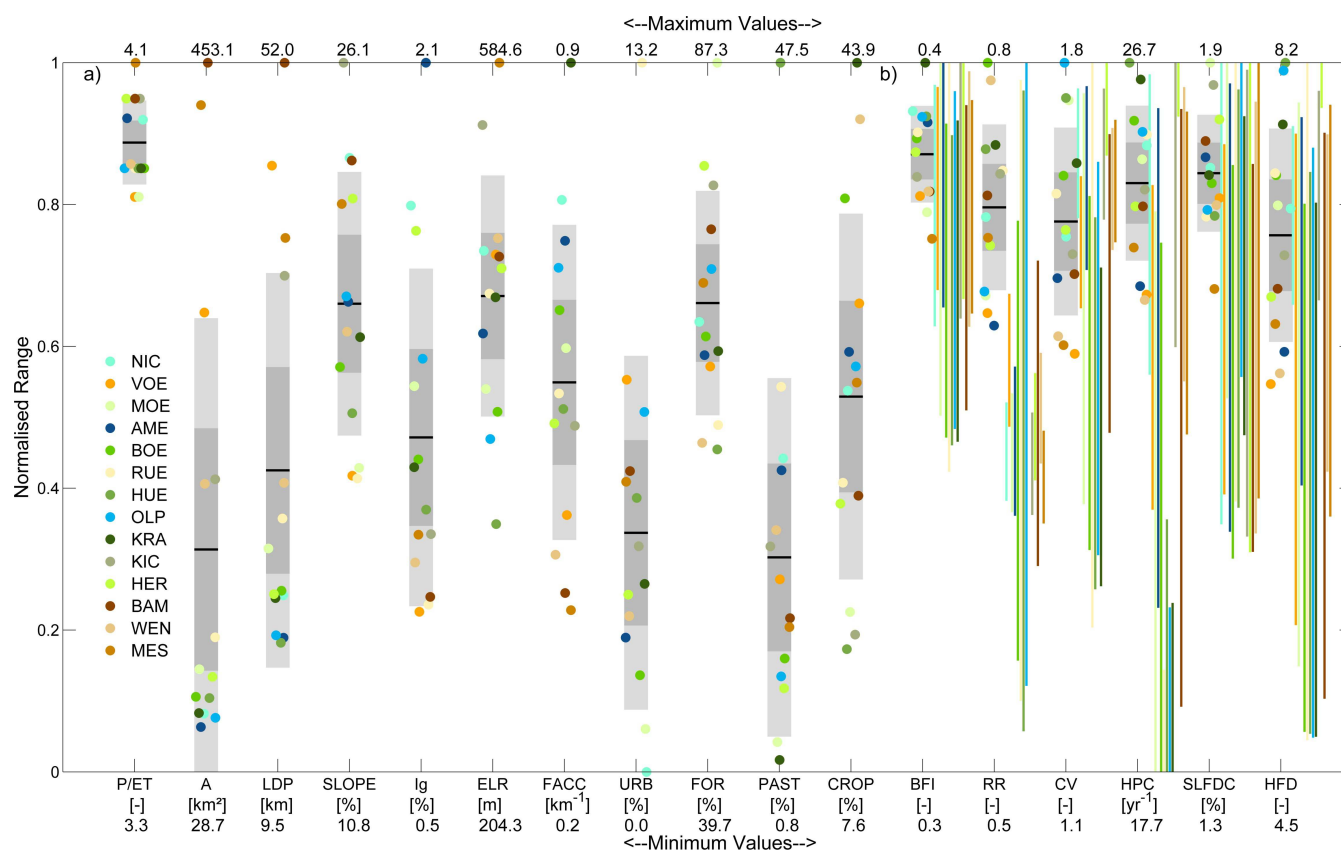


Figure 2. Plots of derived hydrologic fingerprints of physiographic and climate characteristics (a) & dynamic response (b) for 14 headwaters (colour coded). Observed fingerprints as shown as dots juxtaposed with corresponding simulated FAST-mHM ensembles depicted as vertical lines in the case of dynamic response fingerprints (b). Observed values superimpose boxes showing appertaining averages (black line), 95 % confidence intervals (grey) & standard deviations σ (light grey). Numbers on the lower and upper axis denote minimum and maximum values of fingerprints among the 14 watersheds. In the case of dynamic response fingerprints the extreme values are denoted for indices derived from observed streamflow data (b). Plots are normalised by the respective maximal values among the 14 headwaters.

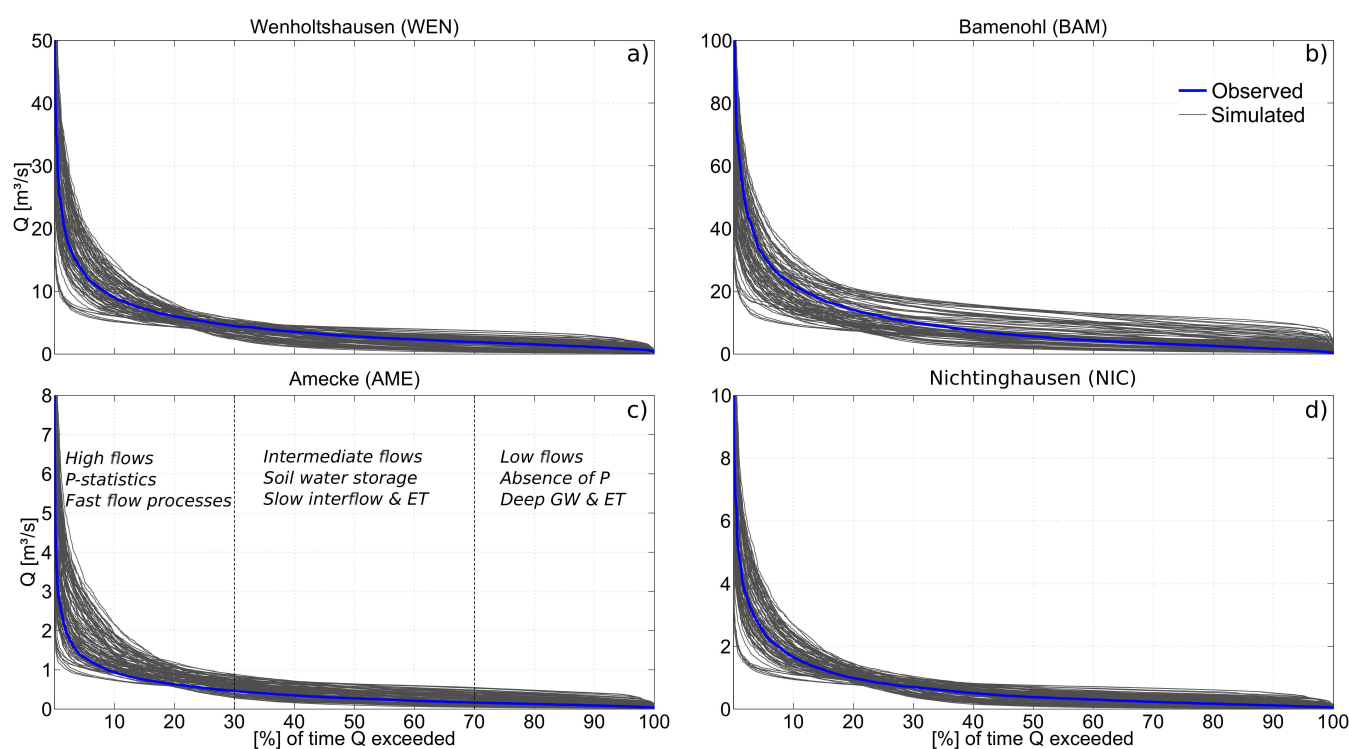


Figure 3. Ensemble of FAST-mHM simulated & observed FDCs of 4 selected Ruhr headwater catchments WEN (a), BAM (b), AME (c) and NIC (d) (daily flow, 2002-2011). Process controls in qualitative sequences of the FDCs proposed by Yilmaz et al. (2008) (c).

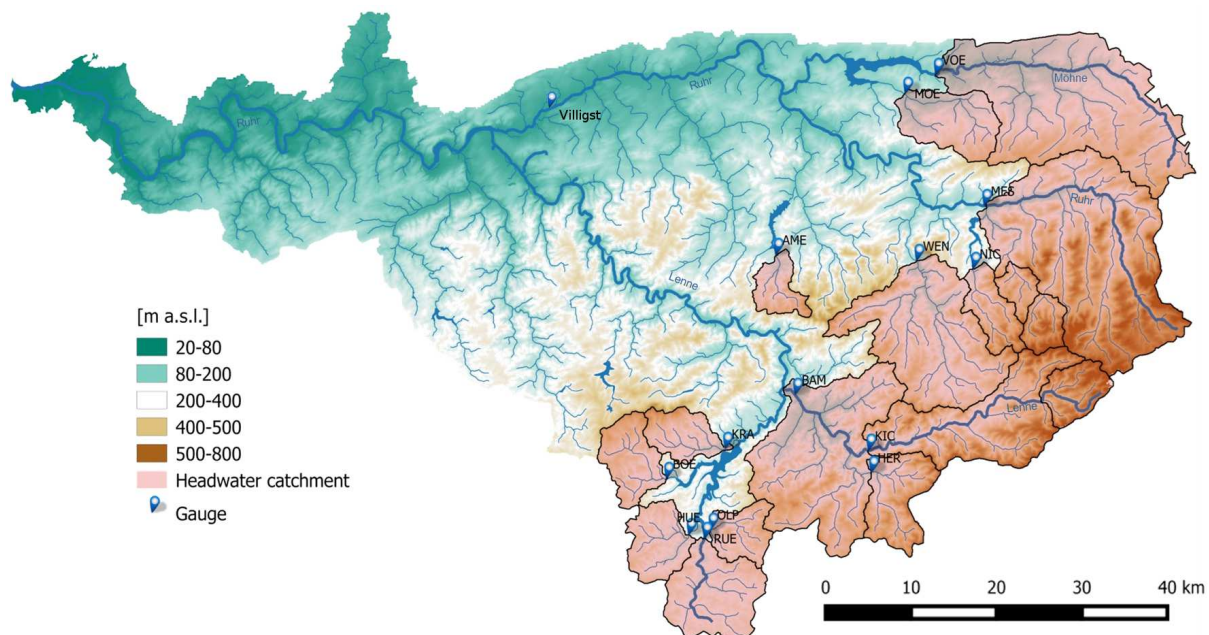


Figure 4. The Ruhr catchment with altitudinal zones, river network and 14 gauged headwater catchments and downstream gauge Villigst.

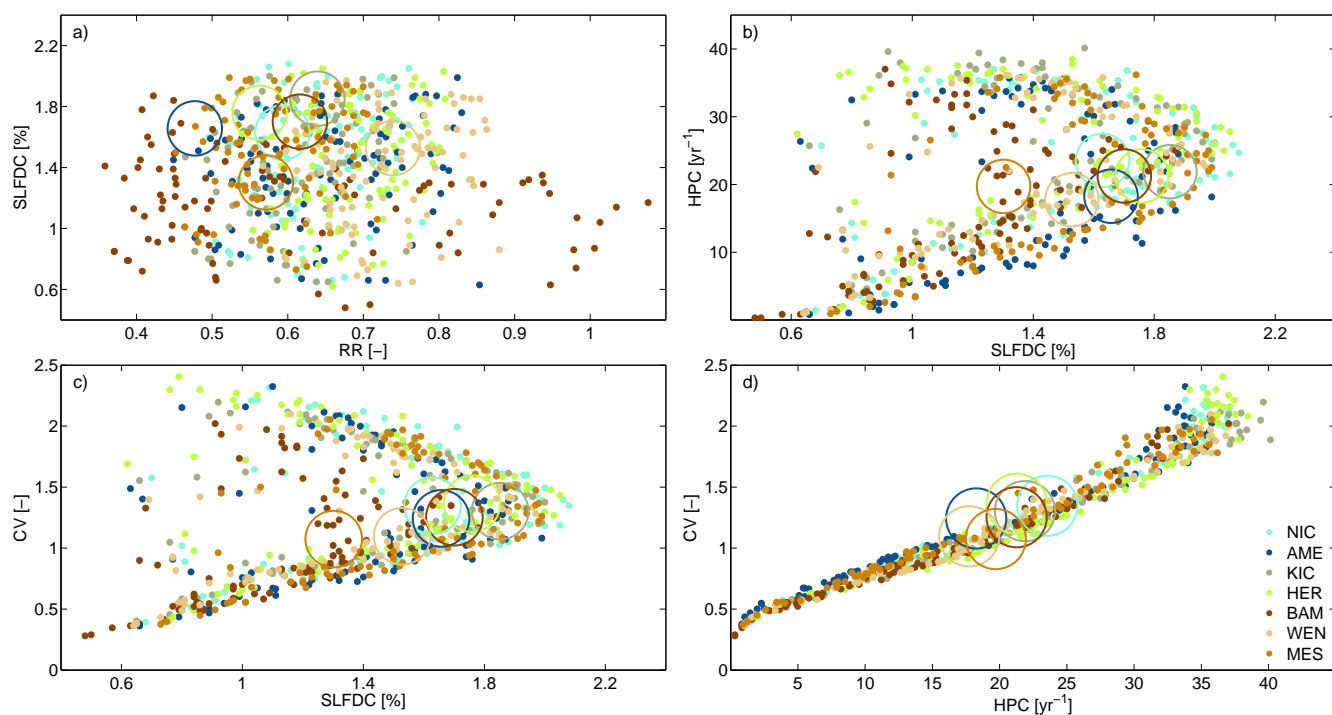


Figure 5. (a-d) Observed and scatterplots of the ensemble of FAST-mHM simulated dynamic response fingerprints. The centers of large circles depict observed values of combined response fingerprints. Circles' radii were scaled by the average values of the respective mean σ of both fingerprints considered in each subplot qualitatively illustrating a possible constraint for the FAST-mHM parameter space.

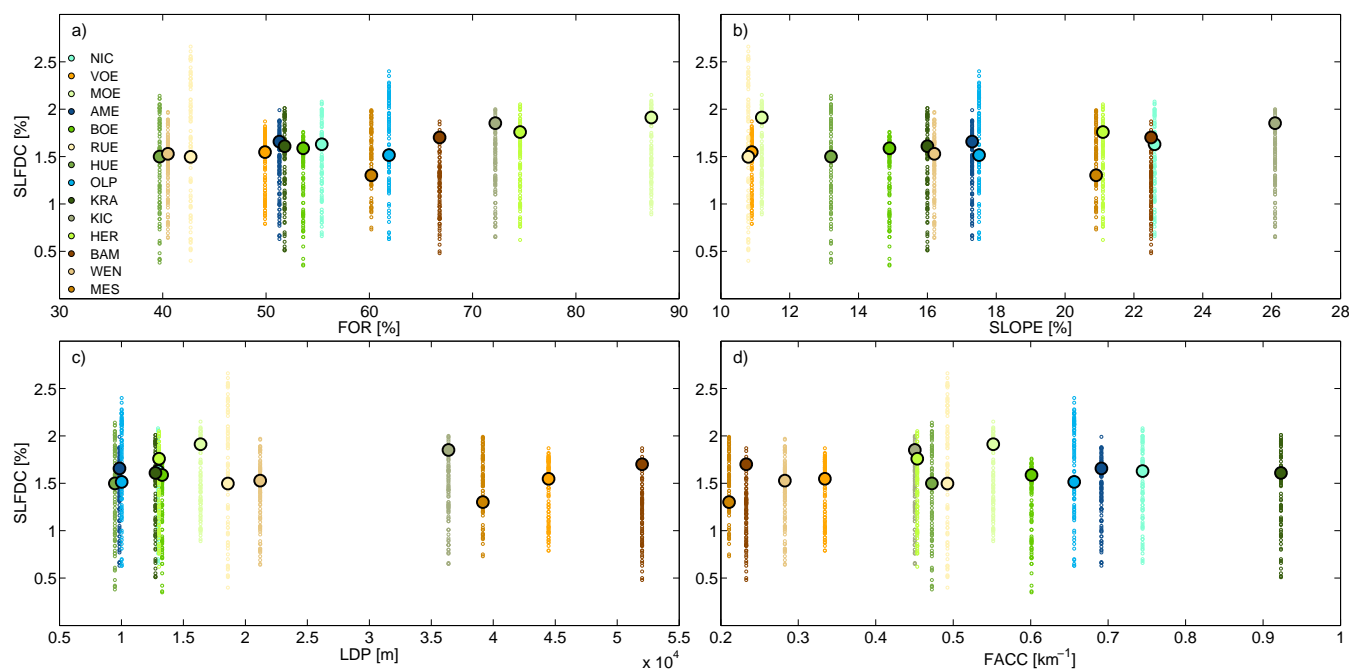


Figure 6. Scatterplots of observed (filled large circles) and ensemble of the simulated dynamic response fingerprint SLFDC against the physiographic fingerprints land cover type forest (FOR) (a), average catchment slope (SLOPE) (b), Longest Drainage Path (LDP) (c) & Flow Accumulation (FACC) (d) at 14 headwaters of the Ruhr basin. Circles of observed quantities indicate the sample space to be considered as constraint for the FAST-mHM parameter space.

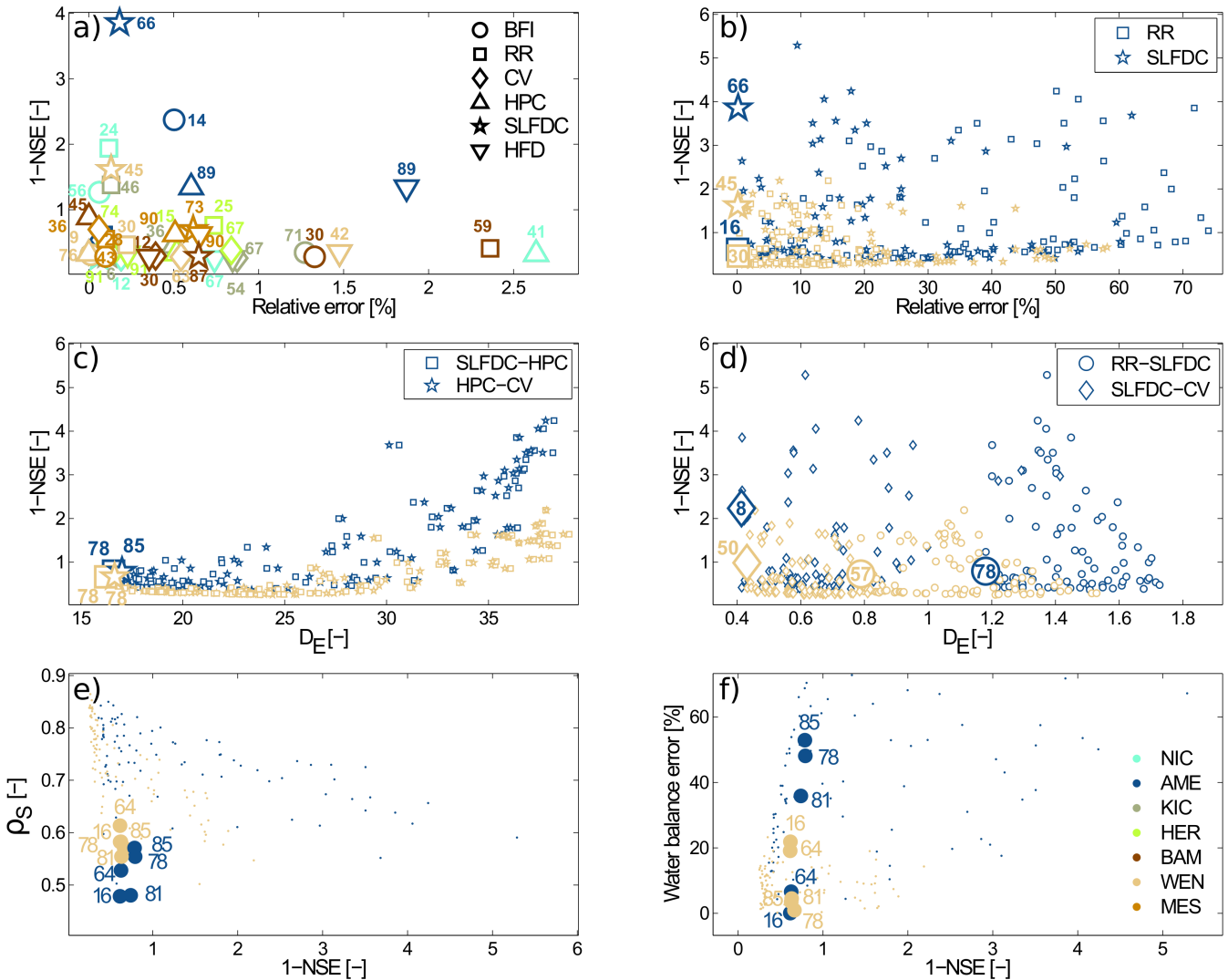


Figure 7. Performance of most behavioural parameter sets for each dynamic response fingerprint at seven headwater gauges (a) and for Runoff Ratio (RR) and Slope of Flow Duration Curve (SLFDC) with corresponding ensembles of FAST-mHM simulated fingerprints at gauges AME and WEN (b). NSE against Euclidean distance (D_E) between four observed and ensemble simulated pairs of dynamic fingerprints (b-c). Spearman's rank correlation (e) and water balance error (f) against NSE for ensemble simulations with a focus on the overall 5 most behavioural parameter sets of 4 analysed pairs of fingerprints. Most behavioural parameter sets are plotted with their model run number and as large markers (a-f).

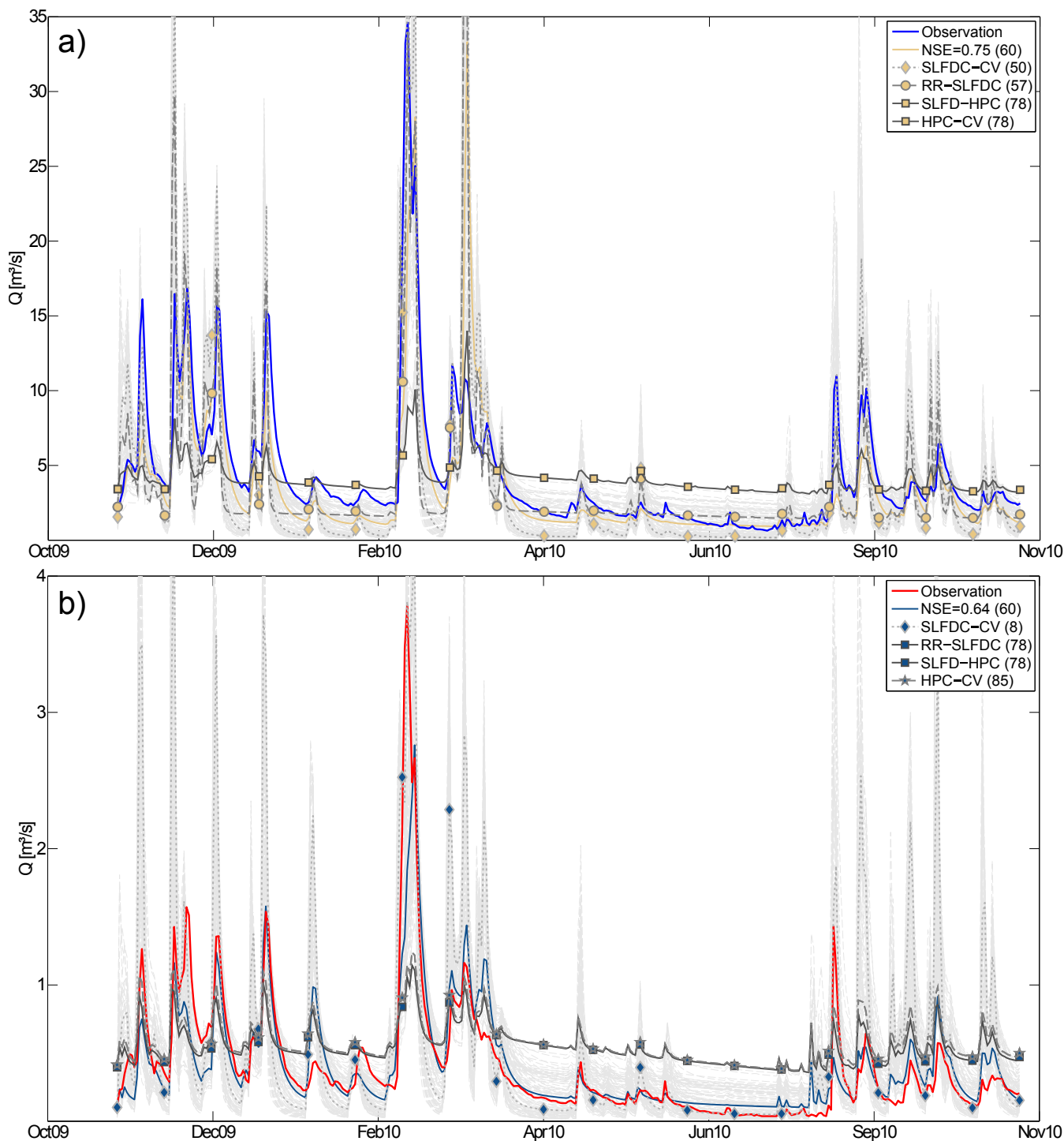


Figure 8. Comparison of streamflow hydrographs corresponding to different derivation: Most behavioural FAST-mHM parameter sets of 4 combinations of dynamic fingerprints, FAST-mHM parameter set with the highest Nash-Sutcliffe model efficiency (NSE) and observation at gauges WEN (a) and AME (b).

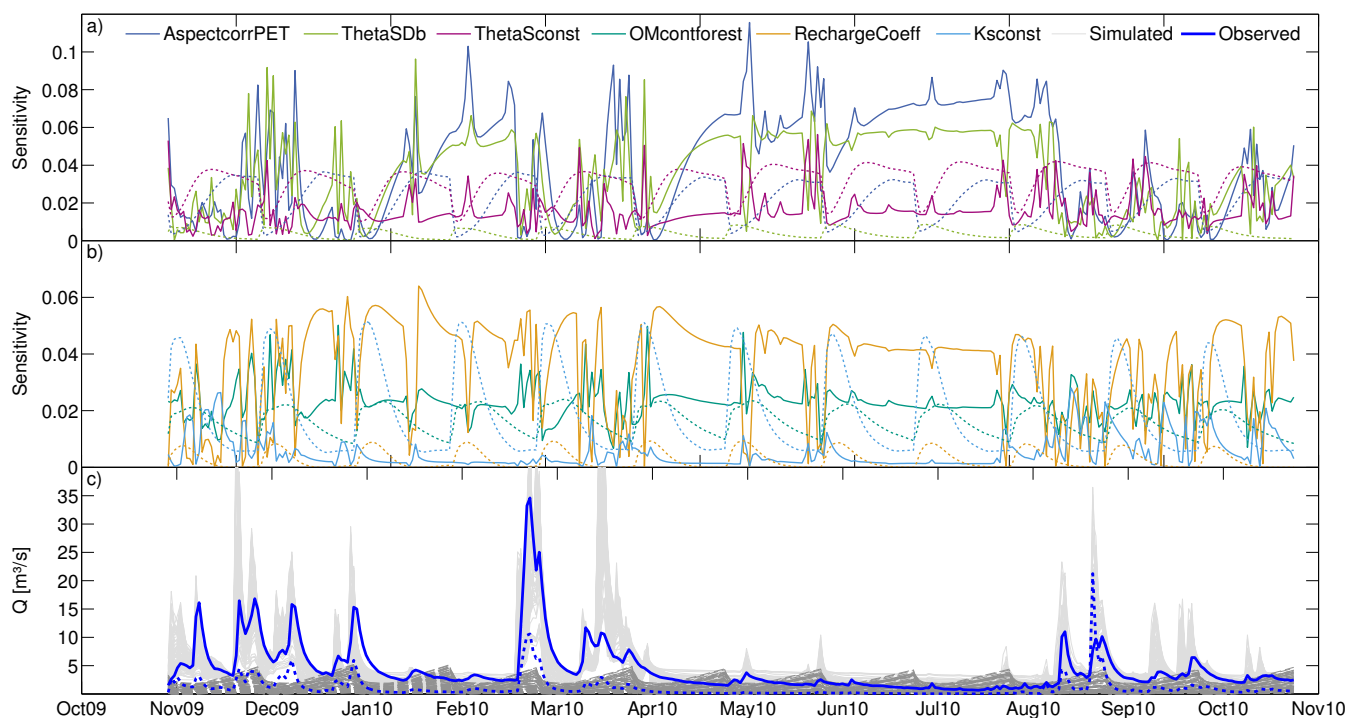


Figure 9. Time-dependent FAST sensitivities of six global mHM parameters for gauges WEN (solid lines) and HUE (dashed) (a and b). Observed and FAST-mHM simulated streamflow ensembles at gauges WEN (solid blue & light grey) and HUE (dashed blue & grey) (c).

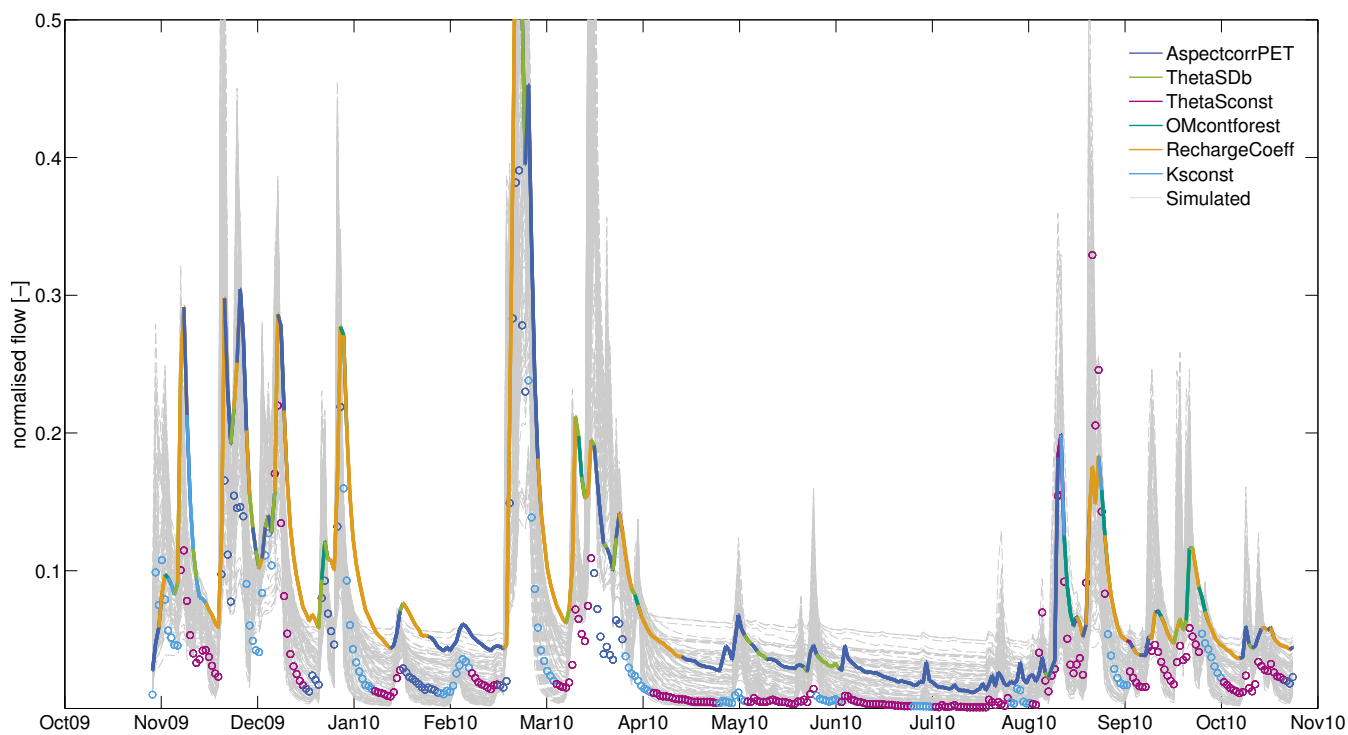


Figure 10. Highest parameter sensitivity related observed and corresponding ensemble of FAST-mHM simulated streamflow hydrographs at gauge Wenholthausen (WEN) (solid line) and highest parameter sensitivity related observed hydrograph at gauge Hüppcherhammer (HUE) (circles).

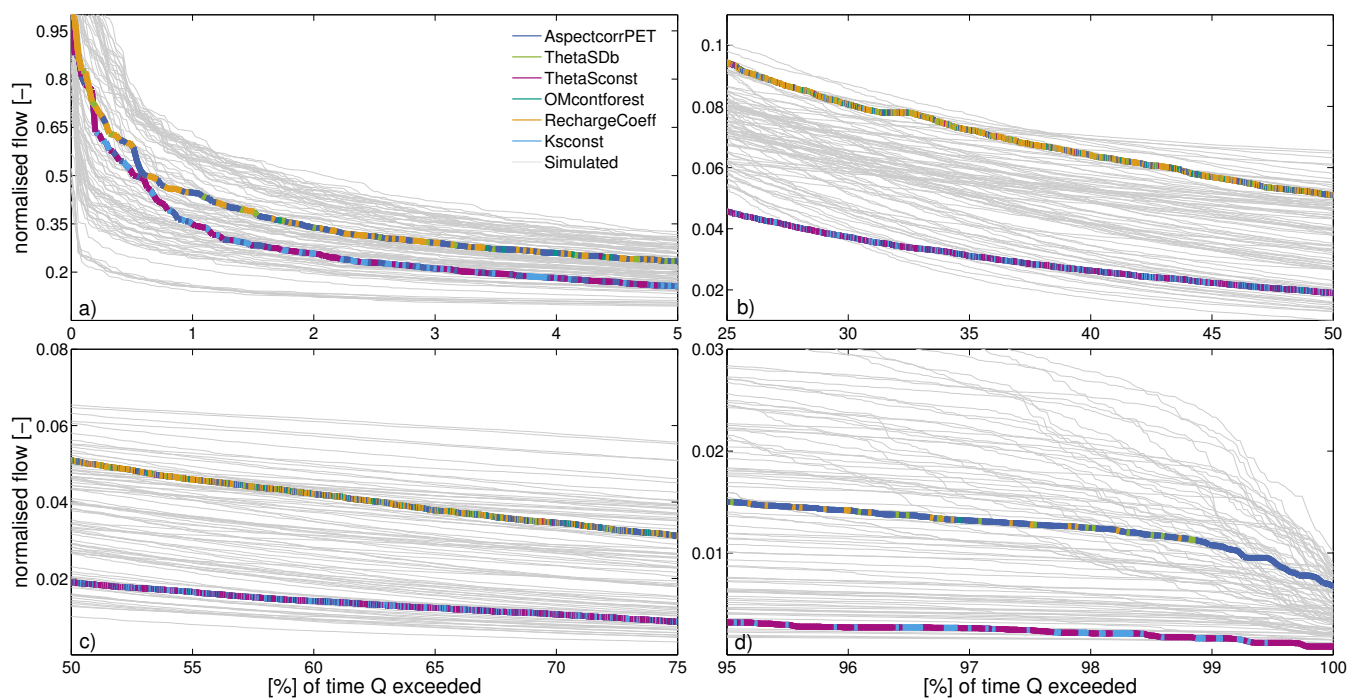


Figure 11. (a-d) 4 sequences of streamflow exceedance of highest parameter sensitivity related observed and corresponding ensemble of FAST-mHM simulated FDCs at gauge Wenholthausen (WEN) (upper FDC) and observed FDC at gauge Huppcherhammer HUE (lower FDCs).

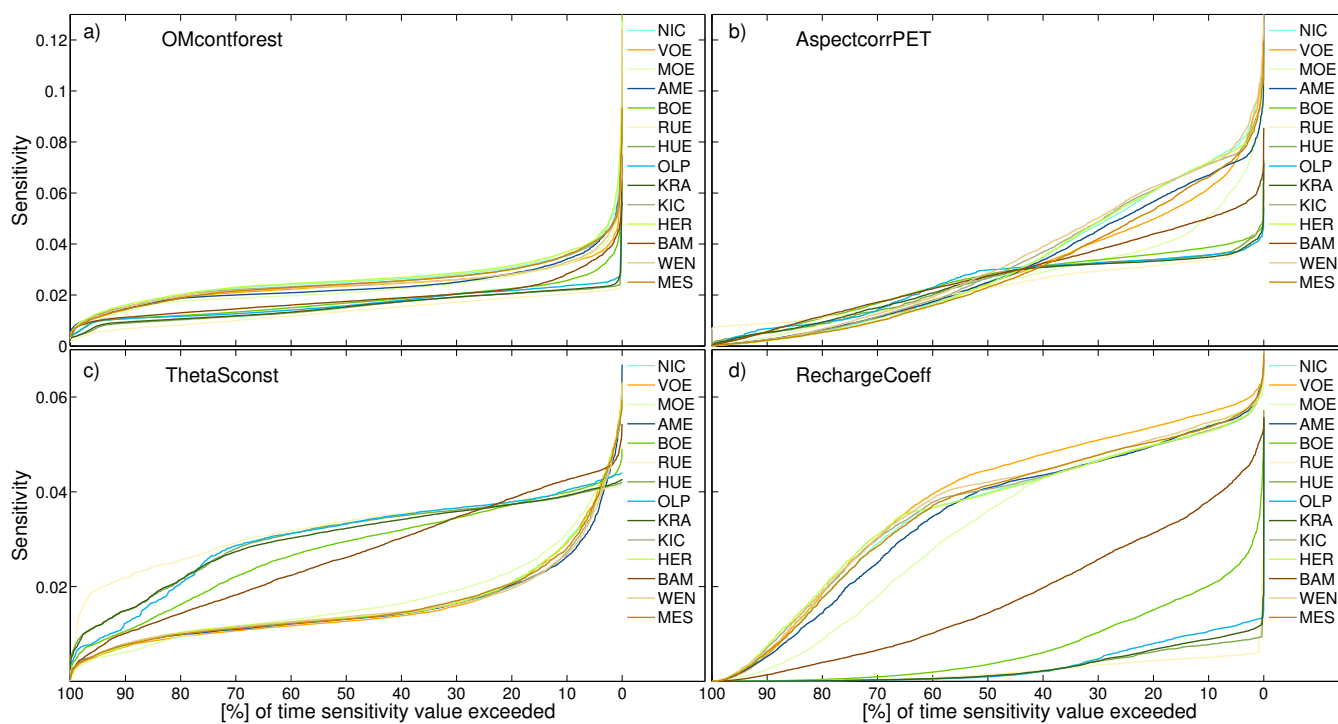


Figure 12. Sensitivity duration curves (SDCs) for 4 mHM parameters OMcontforest (a), AspectcorrPET (b), ThetaSconst (c) and Recharge-Coeff (d) in 14 Ruhr headwater catchments (2003-2011).



Table 1. Six global mHM parameters, function, value and FAST sensitivity ranges (2002-2011) for 14 Ruhr headwater catchments.

Parameter	Process	Description	Value Range [-]	Sensitivity Range [-]
AspectcorrPET	meteo correction	Account for aspect [$^{\circ}$] dependent correction of PET [$mm d^{-1}$] [$m^3 m^{-3}$]	0.70 - 1.50	0 - 0.132
ThetaSDB	soil moisture	Estimation of water content of saturated soil (Bulk density) [$g cm^{-3}$]	-0.37 - -0.19	0 - 0.259
ThetaSconst	soil moisture	Estimation of water content at saturation of soil (Constant part)	0.65 - 0.95	0 - 0.067
OMcontforest	soil moisture	Portion of organic matter [%] in soil to account for effects of organic matter	0 - 20	0.002 - 0.176
Ksconst	soil moisture	Estimation of saturated vertical hydraulic conductivity [$cm d^{-1}$]	-1.9 - 0.0	0 - 0.074
RechargeCoeff	percolation	Determination of percolation [$mm d^{-1}$] coefficient [d^{-1}]	0 - 70	0 - 0.069



Table 2. Physiographic, climatic and dynamic response fingerprints to classify headwater catchments and constrain parameter space of mHM.

Fingerprint Category	Functional Class	Hydologic Fingerprint
Physiography and climate	Climate	Wetness Index (P/ET)
	Landform	Area (A)
	Topography	Longest Drainage Path (LDP) (Weighted) Slope (SLOPE, I_g) Elevation Range
	Land cover	Flow Accumulation (FACC) Forest (FOR), Urban (URB), Pasture (PAST), Cropland (CROP)
	Soil	Baseflow Index (BFI)
Dynamic response	Average streamflow	Runoff Ratio (RR) Total Q / Total P
	Streamflow variability	Coefficient of Variation (CV) σ/μ
	Frequency of flow events	High Pulse Count (HPC) Number of timesteps $Q > 3*Q(\text{mean})$ / years
	Rate of change in streamflow	Slope of Flow Duration Curve (SLFDC) Between 33 % & 66 % exceedance
	High flow events	High Flow Discharge (HFD) $Q(5\text{th percentile}) / Q(\text{median})$