# Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case

**I.G. Pechlivanidis and B. Arheimer**

Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
*Correspondence to:* I.G. Pechlivanidis (ilias.pechlivanidis@smhi.se)

**ABSTRACT**

The Prediction in Ungauged Basins (PUB) scientific initiative (2003-2012 by IAHS) put considerable effort into improving the reliability of hydrological models to predict flow response in ungauged rivers. PUB's collective experience advanced hydrologic science and defined guidelines to make predictions in catchments without observed runoff data. At present, there is a raised interest in applying catchment models for large domains and large data samples in a multi-basin manner, to explore emerging spatial patterns or learn from comparative hydrology. However, such modelling involves additional sources of uncertainties caused by the inconsistency between input datasets, i.e. particularly regional and global databases. This may lead to inaccurate model parameterisation and erroneous process understanding. In order to bridge the gap between the best practices for flow predictions in single catchments and multi-basins at the large scale, we present a further developed and slightly modified version of the recommended best practices for PUB by Takeuchi et al. (2013). By using examples from a recent HYPE hydrological model set-up across 6 000 subbasins for the Indian subcontinent, named India-HYPE v1.0, we explore the PUB recommendations, indicate challenges and recommend ways to overcome them. We describe the work process related to: (a) errors and inconsistencies in global databases, unknown human impacts, poor data quality; (b) robust approaches to identify model parameters using a stepwise calibration approach, remote sensing data, expert knowledge and catchment similarities; and (c) evaluation based on flow signatures and performance metrics, using both multiple criteria and multiple variables, and independent gauges for "blind tests". The results show that despite the strong physiographical gradient over the subcontinent, a single model can describe the spatial variability in dominant hydrological processes at the catchment scale. In addition, spatial model deficiencies are used to identify potential improvements of the model concept. Eventually, through simultaneous calibration using numerous gauges, the median Kling-Gupta Efficiency for river flow increased from 0.14 to 0.64. We finally demonstrate the potential of multi-basin modelling for comparative hydrology using PUB, by grouping the 6 000 subbasins based on similarities in flow signatures to gain insights in spatial patterns of flow generating processes at the large scale.

**Keywords**
Multi-basin modelling, large-scale hydrology, PUB, HYPE, model set-up, parameter constraints, flow signatures, spatial patterns, India

## 1. INTRODUCTION

Numerical hydrological models have been used world-wide for operational needs and scientific research since the early 1970s (e.g. Hrachowitz et al., 2013; Pechlivanidis et al., 2011; Refsgaard et al., 2010; Singh, 1995). The Prediction in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences (IAHS) was launched in 2003 to improve the reliability of models at ungauged regions, overcome the fragmentation in catchment hydrology, and advance the collective understanding (Sivapalan et al., 2003). PUB highlighted the need to: 1) characterise the data and model information content, 2) examine the extent to which a model can be reconciled with observations, and 3) point towards model structural improvements (Gupta et al., 2008). In this regard, several approaches (e.g. multi-objectives, signature measures, information-based metrics, sub-period evaluation) have been applied to reveal significant information about the hydrological systems and indicate perceived model structural errors (Hrachowitz et al., 2013). The use of parameter constraints has also been a significant advancement since such an approach can increase model consistency and reliability (Bulygina et al., 2009; Hrachowitz et al., 2014). Constraints are generated by independent information via either additional data, i.e. remote sensing, tracers, quality, multiple-variables, etc. (Arheimer et al., 2011; Finger et al., 2011; McDonnell et al., 2010; McMillan et al., 2012; Samaniego et al., 2011) and/or expert knowledge (Bulygina et al., 2012; Fenicia et al., 2008; Gao et al., 2014).

It is apparent that the PUB community made significant progress towards these scientific objectives; however the investigations were normally conducted at only one or a limited number of catchments (Hrachowitz et al., 2013). Such an approach is indeed focused on detailed process investigation but is limited when it comes to generalisation of the underlying hydrological hypotheses; to advance science in hydrology, much can be gained by comparative hydrology to search for robustness in hypothesis (Blöschl et al., 2013; Falkenmark and Chapman, 1989). The need for a large sample of process understanding and model evaluation has also been highlighted in the new 2013-2022 IAHS scientific initiative named “Panta Rhei – Everything Flows” (Montanari et al., 2013).

Multi-basin modelling at the large scale complement the “deep” knowledge from single catchment modelling, whilst the current release of open and global datasets has given new opportunities for catchment hydrologists to contribute (Andreassian et al., 2006; Arheimer and Brandt, 1998; Gupta et al., 2014; Johnston and Smakhtin, 2014). Hydrological modelling at the large scale has the potential to encompass many river basins, cross regional and international boundaries and represent a number of different physiographic and climatic zones (Alcamo et al., 2003; Raje et al., 2013; Widén-Nilsson et al., 2007). Application of multi-basin modelling at the large scale can be used to predict the hydrological response at interior ungauged basins (Arheimer and Lindström, 2013; Donnelly et al., 2015; Samaniego et al., 2011; Strömqvist et al., 2012). The use of large sample of gauges can also allow exploration of emerging patterns (e.g. climate change impacts) and facilitate comparative hydrology allowing to test hypothesis for many catchments with a wide range of environmental conditions (Blöschl et al., 2013; Donnelly et al., 2015; Falkenmark and Chapman, 1989).

Modelling at the large scale, however, includes additional model uncertainties. Physical properties (e.g. topography, vegetation and soil type) in large systems generally show higher spatial variability and thus larger heterogeneity in system behaviour (Coron et al., 2012; Sawicz et al., 2011), which in turn affects model parameters (Kumar et al., 2013). In addition, large river basins are often strongly influenced by human activities, such as irrigation, hydropower production, and groundwater use, for which information is rarely available at high resolution in global databases. This introduces additional uncertainty regarding process understanding and description at the large scale. Moreover, the topographic and forcing data of global datasets (i.e. water divides, weather and climatic data) are more likely to be inconsistent, erroneous, and/or only available at a coarse resolution (Donnelly et al., 2012; Kauffeldt et al., 2013).

Applying catchment models at the continental scale in a multi-basin manner is a way to introduce catchment modelling approaches to the existing global hydrological models, i.e. land-surface schemes and global water-allocation concepts. In this paper, we pose the following scientific questions: 1) to what extent are the PUB recommendations for catchment scale also relevant for hydrological modelling at the large scale? and 2) how have the scientific advancements during the PUB decade improved the potential for process-based hydrological modelling at the large scale? To address these questions, we: (a) identify specific challenges at the large scale (uncertain/erroneous basin delineation and routing, errors in global datasets, human impact (i.e. reservoir/dams)) and exemplify on how to overcome them, (b) further develop and modify the PUB best practices to be applicable at the large scale, (c) illustrate the improvement on parameter identification by using remote sensing data and expert knowledge, (d) cluster catchments based on physiographic similarity and their hydrological functioning, (e) ensure model reliability using flow signatures and temporal variability of multiple modelled variables, (f) detect links between model performance and physiographical characteristics to understand model inadequacies along the gradient, and finally (g) discuss how process understanding can benefit from multi-basin modelling and what hydrological insights can be gained by analysing spatial patterns from large-scale predictions in ungauged basins. We use examples from the recent HYPE model set-up of the Indian subcontinent, which experiences unique and strong hydro-climatic and physiographic characteristics and poses extraordinary scientific challenges to understand, quantify and predict hydrological responses.

## 2. BEST PRACTICES FOR PUB WHEN MODELLING MULTI-BASINS AT THE LARGE SCALE

Takeuchi et al. (2013) recommend a six step procedure for predicting runoff at locations where no observed runoff data are available (Fig. 1A). This best practice recommendation is intended for single catchments, and requires modification when applied to multi-basins at the large scale (Fig. 1B). Big datasets are subject to uncertainty and identification of errors is usually time consuming. Analysis of each dataset or catchment may be impractical and risk focusing on details instead of the most crucial overall hydrological functioning across the model domain. We therefore recommend starting with a top-down approach, in which the model is setup directly before proceeding with the PUB recommendations (circle of steps in Fig. 1). The hydrological model needs to include the description of most water fluxes, storages and anthropogenic influences that can be relevant and satisfy the modelling objectives. In addition, we recommend using a model that is familiar to the modeller and open for changes, allowing coherent hydrological interpretations and code adjustments to cope with the region's spatial heterogeneity and hydrological features. Setting-up the model system includes to: (i) acquire readily available datasets that cover the entire geographical domain or merge datasets to get a full coverage; (ii) define calculation points and river network, by taking into account the location of gauges, major landscape features, user requests, catchment borders and routing; (iii) make a first set of model input-data files and make the first model run for the model domain with a multi-basin resolution. The analysis of results from the first model run will indicate major obstacles, such as systematic errors in input data or model structural limitations. Moreover, by having the technical system in place immediately facilitates an incremental and agile approach to model set-up, with direct feedback on model performance at many gauges. We then recommend starting to improve the performance according to the six steps of best practices for predictions in ungauged basins, using a bottom-up approach to refine input data, model structure and parameter values.

### 2.1. **Read the landscape**: *"Go out to your catchment, look around…!"* (cit: page 385 in Blöschl et al. (2013))

It is practically impossible to visit all the basins in a large-scale domain, so instead we recommend: (i) navigate on hard-copies, digitised maps and webpages (e.g. Google Earth) to check landscape characteristics; (ii) review the literature for dominant processes and well-known features or hydrological challenges in the region; (iii) proceed with quality checks and cross-validations towards other data sources (i.e. sources that contain limited in space but local information); (iv) validate the basin delineation and routing using archived metadata from other available datasets; (v) check quality of observed discharge data to assure coherence of time-series; and finally, (vi) check the spatiotemporal information of meteorological datasets after transformation from the grid to the subbasin scale. It is important to get an understanding of the entire domain and ensure that the datasets correspond to this understanding, and hence tackling systematic errors in the data.

### 2.2. Runoff signatures and processes: *"Analyse all runoff signatures in nearby catchments to get an understanding…!"* (cit: page 385 in Blöschl et al. (2013))

Detailed inspection of flow signatures for each gauging station from large datasets (often in the range of thousand stations, see http://hypeweb.smhi.se/) is best done by using clustering techniques to discover spatial similarities (Sawicz et al., 2011). It is then important to use many flow signatures for each site to fully capture the characteristics of the hydrographs. We also recommend searching for statistical relationships between the observed flow signatures and basin characteristics (both physiography and human alteration) across the model domain. This will increase our understanding of dominant processes and fitness of model structure (Donnelly et al., 2015).

### 2.3. Process similarity and grouping: *"…find similar gauged catchments to assist in predicting runoff in the ungauged basin!"* (cit: page 385 in Blöschl et al. (2013))

In most process-based models, the modeller has some freedom to define the characteristics of the smallest calculation units, which are normally linked to physiography to account for spatial distribution of for instance soil properties or land use. When producing these calculation units both technical (e.g. computational efficiency) and conceptual (e.g. restrictions with the number of classes) concerns must be taken into account. However, lakes, wetlands, glacier, and urban areas should be respected since even small proportions can significantly alter the flow regime. When calculation units are defined, we recommend clustering the basins/gauges with similar upstream characteristics and/or system behaviour to isolate key processes for regionalisation of parameter values during calibration. We finally suggest checking the spatial distribution by plotting the catchment characteristics of subbasins on maps and compare to other or original data sources.

***2.1-3. Quality checks:*** This is an additional step in the procedure accounting for repetition of step 1-3 in an iterative way to ensure quality in the required input data and files of the model prior to parameter tuning (Fig. 1); it is easy to fail and introduce errors when handling large datasets by automatic scripts (generalisation of scripts is not always straightforward and some manual adjustment is usually required) and/or human error (particularly when many modellers collaborate), which can lead to erroneous assumptions on hydrological processes during calibration. We recommend to analyse flow time-series as follows: (i) compare modelled to observed time-series and signatures; (ii) check water-volume errors and their distribution in space; (iii) inspect the spatial distribution of model dynamics to correct spatial patterns from systematic errors; and (iv) search for errors in the model set-up (routing, meteorological input etc.).

### 2.4. Model - Right for the right reasons: *"Build… model for the signature of interest… regionalise the parameters from similar catchments…more information than the hydrograph…!"* (cit: page 385 in Blöschl et al. (2013))

In here, it is crucial that the model structure represents the modeller's perception of how the hydrological system is organized and how the various processes are interconnected. For the model set-up to be "right for the right reasons", we recommend to: (i) constrain relevant parameters to alternative data than just time-series of river discharge (e.g. snowmelt parameters to snow depths, evapotranspiration parameters to data from flux towers and satellites) or select a subset of gauges representing different flow generating processes; (ii) apply expert knowledge when analysing internal variables to ensure that the model structure reflects the understanding of flow paths and their interconnections; (iii) change the model algorithms or structure if tuning of parameters is not enough to reflect the perception of the hydrological system; (iv) include specific rating curves of lakes and reservoirs wherever available, and tune parameters for irrigation and dam regulation to fit the flow dynamics at downstream gauges; and (v) assimilate observed data if possible, e.g. snow, upstream discharge, and regulation rules in reservoirs.

### 2.5. Hydrological interpretation: *"Interpret the parameters… and justify their values against what was learnt during field trips and other data…!"* (cit: page 385 in Blöschl et al. (2013))

Although, hydrological interpretation has been present in every step of the model set-up procedure described here, this step includes the overall synthesis and analysis of results both at the large scale and for single catchments in the multi-basin approach. For *spatial interpretation*, we recommend plotting maps with multi-basin outputs for several variables, performance criteria and signatures across the model domain. This allows checking model's coherency at various landscape features, e.g. spatial patterns of vegetation, geology, climate, population density, and human alterations. The objective is to understand the drivers that influence flow, find rational reasons behind the hydrological heterogeneity, and identify knowledge gaps or model limitations. For *temporal interpretation*, we recommend plotting time-series for some basins in each group of similar landscape units and catchment response. This is to make sure that the model reflects our perception and assists to better understand the dominant drivers of the flow generation processes and water dynamics in the region.

### 2.6. Uncertainty – local and regional: *"… by combining error propagation methods, regional cross-validation and hydrological interpretation…!"* (cit: page 385 in Blöschl et al. (2013))

Multi-basin models are more computationally demanding than single basin models and it is therefore not always feasible to explicitly address all uncertainties from all sources. To explore the model performance in ungauged basins, we recommend dividing the set of gauging stations into those used in calibration and independent "blind-tests", respectively. Cross-validation, e.g. using the jackknife procedure (Good, 2006), is practically difficult in process-based modelling of multi-basins. To examine uncertainties we recommend to: (i) use several performance (diagnostic) criteria and many flow signatures; (ii) relate the spatial distribution of model performance to physiographical variables; and (iii) check model performance for independent gauging sites and new datasets.

The major spatiotemporal deviations found between modelled and observed data should be the focus for the next round in the circle of steps for better predictions. We recommend reading the landscape and searching for local knowledge again to elaborate new hypotheses of hydrological functioning and data sources. We also recommend documenting and version-managing each model set-up before looping into step 1, to ensure knowledge accumulation for a broader audience and to make the set-up process transparent. This sets a baseline for the next round of improvements.

## 3. DATA AND METHODS

### 3.1. Study area and data description

India is considered the seventh largest country by area and the second-most populous country with over 1.2 billion people. The country covers an area of about 3.3 million km$^2$ and some of its river basins cover several countries in the area (i.e. China, Nepal, Pakistan, and Bangladesh; see Fig. 2). The spatiotemporal variation in climate is perhaps greater than any other area of similar size in the world. The climate is generally strongly influenced by the Himalayas and the Thar Desert in the northwest, both of which contribute to drive the summer and winter monsoons (Attri and Tyagi, 2010). Four seasons can be distinguished: winter (January-February), pre-monsoon (March-May), monsoon (June-September), and post-monsoon (October-December). The temperature varies between seasons ranging from mean temperatures of about 10 °C in winter to about 32 °C in pre-monsoon season. In terms of spatial variability, the rainfall pattern roughly reflects the different climate regimes of the country, which vary from humid in the northeast (rainfall occurs about 180 days/year), to arid in Rajasthan (20 days/year). Accordingly, river flow show large spatial and seasonal variability across the sub-continent (Fig. 2b), e.g. the Ganga River has an intra-annual amplitude in monthly river discharge of 50 000 m$^3$/s.

For the hydrological model set-up, we use global datasets to extract the input data (see Table 1). APHRODITE (Yatagai et al., 2009, 2012) and AphroTEMP (Yasutomi et al., 2011) are the only long-term continental-scale datasets that contain a dense network of daily data (in here, only daily precipitation and mean temperature are required) for Asia including the Himalayas. Data of land use and soil type were aggregated into fewer classes than in the original databases. Discharge data are available from the Global Runoff Data Centre (GRDC) at 42 sites limited to monthly values in the period 1971-1979. More discharge data are held in the Indian government agencies but are not open to the public. Consequently, in this application, flow information (Table 2) is available only for a small fraction of the subcontinent, which makes the region a great example for PUB. Monthly potential evapotranspiration (pot. E) data were obtained for the period 2000-2008 from the Moderate Resolution Imaging Spectroradiometer (MODIS) global dataset (Mu et al., 2007, 2011). The dataset covers the domain in a spatial resolution of 1 km and is derived based on the Penman-Monteith (Penman, 1948) approach.

Water divides and catchment characteristics were appointed for each subbasin by using the World Hydrological model Input Set-up Tool (WHIST; http://hype.sourceforge.net/WHIST/). This is a spatial information tool from SMHI to transform data and create input files for hydrological models, from different types of databases. From the information of topographic databases, for example, WHIST can delineate the subbasins and the linking (routing) between them. This is also the tool for allocating information of soil, vegetation, surface water, regulation and irrigation to each calculation unit. For the Indian subcontinent, we chose to work with some 6 000 points for calculations of runoff in the river network (i.e. 6 000 subbasins).

### 3.2. A multi-basin hydrological model for large-scale applications - the HYPE model

The Hydrological Predictions for the Environment (HYPE) model is a dynamic rainfall-runoff model, which describes the hydrological processes at the catchment scale (Lindström et al., 2010). The model represents processes for snow/ice, evapotranspiration, soil moisture and flow paths, groundwater fluctuations, aquifers, human alterations (reservoirs, regulation, irrigation, abstractions), and routing through rivers and lakes. The HYPE source code is continuously developed and released in new versions for open access at http://hype.sourceforge.net/, where also model descriptions, manuals and file descriptions can be downloaded.

HYPE is most often run at a daily time-step and simulates the water flow paths in soil for Hydrological Response Units (HRU), which are defined by gridded soil and land-use classes and can be divided in up to three layers with a fluctuating groundwater table. The HRUs are further aggregated into subbasins based on topography. Elevation is also used to get temperature variations within a subbasin to influence the snow melt and storage as well as evapotranspiration. Glaciers have a variable surface and volume, while lakes are defined as classes with specified areas and variable volume. Lakes receive runoff from the local catchment and, if located in the subbasin outlet, also the river flow from upstream subbasins. On glaciers and lakes, precipitation falls directly on the surfaces and water evaporates at the potential rate. Each lake has a defined depth below an outflow threshold. The outflow from lakes is determined by a general rating curve unless a specific one is given or if the lake is regulated. Regulated lakes and man-made reservoirs are treated equally but a simple regulation rule can be used, in which the outflow is constant or follows a seasonal function (as it is often the case with hydropower) for water levels above the threshold. A rating curve for the spillways can be used when the reservoir is full.

Irrigation is simulated based on crop water demands calculated either with the FAO-56 crop coefficient method (Allen et al., 1998) or relative to a reference flooding level for submerged crops (e.g. rice). The demands are withdrawn from rivers, lakes, reservoirs, and/or groundwater within and/or external to the subbasin where the demands originated and are constrained by the water available at these sources. After subtraction of conveyance losses, the withdrawn water is applied as additional infiltration to the irrigated soils. The agriculture and irrigation datasets (see Table 1) are used to define irrigated area, crop types, growing seasons, crop coefficients, irrigation methods and efficiencies, and irrigation sources. The irrigation parameters regulating water demand and abstraction are usually manually calibrated using discharge stations in irrigation-dominated areas.

River discharge is routed between the subbasins along the river network and may also pass subbasins, flow laterally in the soil between subbasins or interact with a deeper groundwater aquifer in the model. For the study in this paper, the HYPE model version 4.5.0 was set up for the entire Indian subcontinent (4.9 million $km^2$) with a resolution of 6 010 subbasins, i.e. on average 810 $km^2$, and is referred to as India-HYPE version 1.0.

### 3.3. Model calibration and regionalisation

The calibration objective was to derive a reliable model of adequately representing the temporal dynamics of flow (high flows, timing, variability and volume) across the Indian river systems. With such a model set-up, we can identify spatial patterns of hydrologic similarity across the subcontinent, and also analyse impacts of environmental change on water resources. The HYPE model has many rate coefficients, constants and parameters, which in theory could be adjusted, but in practice some 20 are tuned during calibration. Many of the parameters are linked to physiographic characteristics in the landscape, such as soil type and depths (soil dependent parameters) or vegetation (land use dependent parameters), while others are assumed to be general to the entire domain (general parameters) or specific to a defined region or river (regional parameters). Parameters for each HRU are calibrated for representative gauged basins and then transferred to similar HRUs, which are gridded with higher resolution than the subbasins across the whole domain to account for spatial variability in soil and land use. Using the distributed HRU approach in the multi-basin concept is thus one part of the regionalisation method for parameter values. Some other parameters, however, are either estimated from literature values and from previous modelling experiences (a priori values) or identified in the (automatic or manual) calibration procedure. Slightly different methods for regionalisation of parameter values have been used when setting up the different HYPE model applications, depending on access to gauging stations, additional data sources and expert knowledge. The following procedure was used for India-HYPE v.1.0:

*Stepwise, iterative calibration of parameter groups*

To tackle, to a certain extent, the equifinality problem in this processed-based model, the parameters (general, soil and land use dependent, specific or regional) are calibrated in a progressive way, i.e. stepwise calibration (Arheimer and Lindström, 2013) using different subsets of the gauging station in each step. In this way, errors induced by inappropriate parameter values in some model processes are not compensated for by introducing errors in other parts of the model. Hence, groups of parameters responsible for certain flow paths or processes (e.g. soil water holding capacity) are calibrated first and then kept constant when the second group of parameters (e.g. river routing) is calibrated. However, stepping downstream along the model code includes some reconsideration about chosen parameter values in an iterative procedure. For each step and group of parameters, a subset of representative gauging stations is used in simultaneous calibration, which means that no gauging station is calibrated individually. This is to get parameters that are robust also for ungauged basins. Model performance in specific sites is thus traded against average performance across the full model domain or regions.

For the Indian subcontinent, the following groups of HYPE parameters were calibrated stepwise: (i) general parameters (e.g. precipitation and temperature correction factors with elevation etc.), which significantly affect the water balance in the system, snow pack and distribution, and regional discharge; (ii) Soil and land use dependent parameters (e.g. field capacity, rate of potential evapotranspiration etc.), which can influence the dynamics of the flow signal, groundwater levels and transit-time, (iii) Regional parameters, which are applied as multipliers to some of the general-soil-land use parameters and may be seen as downscaling parameters as they compensate for the scaling effects and/or other types of uncertainty. The multipliers are either specific for a region or a river-basin.

*Expert knowledge for parameter constraints*

During this progressive stepwise calibration approach, constraints based on expert knowledge and basin similarity are introduced. As an example, we apply a constraint imposed on the *mactrsm* soil dependent parameter (*mactrsm* is the threshold soil water for macropore flow and surface runoff). In the first run, during the calibration procedure the parameter is allowed to vary freely within the parameter range and all distributions for the soil types are acceptable (unconstrained sets). We then apply expert knowledge on the parameter distribution and agree that a model will only be retained as feasible if it can satisfy the constraint:

$$mactrsm_{Coarse} > mactrsm_{Medium} > mactrsm_{Fine}$$

The *mactrsm* values for the remaining two soil types in the India-HYPE model domain, i.e. organic and shallow, are expected to be close to the corresponding values for the coarse soil; although the value for shallow soil is constrained to be less than *mactrsm* for organic soils.

*Spatial clustering based on catchment similarities*

We assume hydrologic similarity across the region on the basis of similarity in physiographic characteristics. We applied a k-means clustering approach within the 17-dimensional space, consisting of: 5 soil types, 7 land use types, mean annual precipitation, mean temperature, mean slope, mean elevation, and basin area. This separated the subbasins into homogeneous classes. A silhouette analysis was used to overcome the subjectivity on the determination of the number of clusters. The catchment similarity approach significantly reduces the number of parameters, while it allows regionalisation of parameters, which are assumed to be robust enough also for ungauged basins.

*Spatiotemporal calibration and evaluation*

India-HYPE was calibrated and evaluated in a multi-basin approach by considering the median performance in all selected stations. 30 stations were selected for model calibration and 12 "blind" stations for spatial validation. The years 1969-1970 are used as a model warm-up period, the next 5 years for model calibration (1971-1975) and the final 4 years for temporal performance evaluation (1976-1979).

The Differential Evolution Markov Chain (DE-MC; Ter Braak, 2006) optimisation algorithm is used to explore the feasible parameter space and to investigate parameter sensitivity. DE-MC was applied at each step of the iterative calibration procedure (to optimise the general, soil and land use dependent, and regional parameters) with 200 generations of 100 parallel chains each being explored respectively. The Kling-Gupta Efficiency, KGE (Gupta et al., 2009), was used to define the performance of the model towards the observed discharge. KGE allows a multi-objective perspective by focusing to separately minimise the correlation (timing) error, variability error, and bias (volume) error. We also investigated the relative influence of timing, variability and volume error on the KGE value. To do this, we transformed the three components to result into a consistent range of possible values (the metrics are named as *cc*, *alpha* and *beta* corresponding to timing, variability and volume errors respectively; see Appendix A).

### 3.4. Evaluation beyond standard performance metrics

*Evaluation based on flow signatures*

The model was further evaluated on its ability to capture spatial and temporal variability in discharge by comparing modelled flow signatures and monthly simulations with observed data. Here, three flow signatures are calculated for each gauging station to illustrate different aspects of the flow variability and the hydrograph characteristics (Appendix A): the mean annual specific runoff (*Qm*, mm $yr^{-1}$), the normalised high flow statistic (*q05*, -) and the slope of the flow duration curve (*mFDC*, -).

*Multi-variable evaluation*

To judge model credibility, other observed variables than river discharge are used, for instance from satellite products. For India-HYPE, these included evaluations against estimated snow areal extent and snow water equivalent from the GlobSnow system and potential evapotranspiration (pot. E) from the MODIS system. The assumption is that MODIS pot. E can be used as reference to calibrate the HYPE parameters that control pot. E; this refers only to the *cevp* land-use dependent parameter, which is a coefficient of potential evapotranspiration (mm/d °C) (Lindström et al., 2010). The *cevp* parameter was optimised for each land use type so that HYPE modelled annual pot. E matches the MODIS annual pot. E at the entire model domain. A Monte Carlo uniform random search was used to explore the feasible *cevp* parameter space (constant for each land use type; 0.15-0.30) and to investigate parameter identifiability and interdependence (10 000 samples). The Root Mean Square Error (RMSE) and Absolute Bias (Bias) were used as objective functions in this analysis; 0 values indicate a perfect model with no errors for both criteria. Note that the analysis was conducted in the 2000-2008 period during which MODIS data were available. We therefore assume that the *cevp* parameter is static in time and representative also for the 1971-1979 period.

*Linking performance to physiographical characteristics*

To better understand the model performance and identify potential for model improvements, we apply classification and regression trees (CART; Breiman et al., 1984). CART is a recursive-partitioning algorithm that classifies the space defined by the input variables (i.e. physiographic-climatic characteristics) based on the output variable (i.e. KGE model performance). The tree consists of a series of nodes, where each node is a logical expression based on a similarity metric in the input space (physiographic-climatic characteristics). In this case, we divided the KGE performance into three groups – bad (KGE < 0.4), medium (0.4 < KGE < 0.7), and good (KGE > 0.7), which were termed C0, C1 and C2 respectively. A terminal leaf exists at the end of each branch of the tree, where the probability of belonging to any of the three output groups can be inspected. Here we summarised the physiographic-climatic characteristics of the basin into 5 soil types (coarse, medium, fine, organic and shallow), 7 land use types (crops, forest, open land with vegetation, urban, bare/desert, glacier, water), mean annual precipitation and mean temperature.

### 3.5. Catchment functioning across gradients

We finally explored the spatial runoff patterns across the entire subcontinent by analysing the flow characteristics in all modelled 6 000 catchments. In here, we used the modelled discharge and calculated 12 flow signatures for each subbasin (see Appendix A): Mean annual specific discharge (mm $yr^{-1}$); Range of Pardé coefficient (-); Slope of FDC (-); Normalised low flow (-); Normalised high flow (-); Coefficient of variation (-); Flashiness defined as 1-autocorrelation (-); Normalised peak distribution (-); Rising limb density (-); Declining limb density (-); Long term mean discharge ($m^3/s$); Normalised relatively low flow (-). We then applied a k-means clustering approach within the 12-dimensional space (consisting of the 12 calculated flow signatures) to categorise the subbasins based on their combined similarity in flow signatures. Through the mapping of the spatial pattern we gained insight in similarities of catchment functioning and could identify the dominant flow generating processes for specific regions. To further highlight the hydrological insights gained during model identification, we conducted the clustering analysis on two different steps of the model calibration and explored the sensitivity of calibration on the spatial patterns of flow signatures.

## 4. RESULTS AND DISCUSSION

The very first model set-up to establish a technical model infrastructure of the Indian subcontinent showed very poor model performance, with an average and median KGE for all stations of -0.02 and 0.0 respectively. This was expected and the baseline for improvements following the six steps of the modified PUB best practices.

### 4.1. Read the landscape

Background knowledge was firstly acquired via visual and/or numerical analysis of available maps that describe the spatial patterns of land use, soil and climate, and study of the scientific literature on regional hydrological investigations, which enabled identification of dominant physical processes and flow paths. Such soft information was useful for turning on/off processes and selecting relevant algorithms, i.e. management, snow melting. Communication with local scientists (i.e. governmental hydrological institutes), managers (i.e. regional water authorities) and end-users (i.e. agricultural sector) enabled knowledge exchange and justified the model approach. Three extensive field trips provided important soft information about system behaviour in the semi-arid northwest and humid subtropical northeast parts of the country (i.e. identification of sources to irrigate water for agricultural needs and estimation of water losses due to faults in the irrigation systems).

Analysis of the topographic data was of major importance since they affected the subbasin delineation and routing. Although Hydrosheds are based on high-resolution elevation layers, which are hydrologically conditioned and corrected, there are still many errors. Merging Hydrosheds with GRDC (hence forcing the delineation at subbasins where GRDC stations are available) involved some mismatches in terms of the size of upstream areas between the subbasin delineations and the GRDC metadata. As an example, the location of the Dundeli station in the Kali Nadi river basin (asterisk 1 in Fig. 2) was adjusted to match the underlying topography and drainage accumulation data based on published and computed upstream areas respectively (see Fig. 3a). The consequent change in the routing resulted in a considerable improvement in the model performance (KGE improved from -0.51 to 0.30; see Fig. 3b). Many similar corrections had to me made.

To make corrections also for ungauged basins and major rivers, the delineated basins were additionally evaluated using a shapefile of basin areas reported by Gosain et al. (2011). Some minor corrections had to be done in the routing to achieve similarly delineated basins, particularly in the northwest region, where mean elevation at the subbasin scale does not show much variability.

### 4.2. Runoff signatures and processes

As recommended, several flow signatures were extracted for the gauging stations across India to be compared to physiographical patterns. Flow signatures were also used for model evaluation to find potential for improvements. The analysis was done at different stages in the model set-up, and finally, there was a relatively good agreement of the observed and modelled flow signatures (Fig. 4). In general, poor agreement was found in mountains and in semi-arid regions, which are characterised by local, convective rainfall events during the monsoon season. No clear pattern is found between signature agreement and basin scale for calibrated river gauges.

We also explored how flow signatures can be affected by human impacts by analysing modelled responses considering and omitting the human influence. Fig. 5 highlights the significant effect reservoirs have to dampen hydrographs and control discharge variability; hence various flow signatures. The model can fairly well represent the reservoir routing and KGE improved from 0.37 to 0.48 after introducing a regulation scheme. The model improved on capturing the seasonality of regulation; however at this modelling state it was not able to represent the monthly peaks. Note that model results are subject to the general rating curve generalised to all reservoirs; there were no downstream data available to calibrate the parameters specifically for a given reservoir/dam.

### 4.3. Process similarity and grouping

After having identified relevant HRUs, reclassified them into suitable calculation units and inserted major features as lakes and dams, we identified basin similarities to drive the identification of the model's regional parameters. The cluster analysis was applied to all 6 010 subbasins of the domain within the 17-dimensional space (see section 3.3). We identified 13 different classes of varying size (Fig. 6) out of 42 values, which is the number of gauged river-basins in the domain, yet with relatively high class strength (i.e. the variability of characteristics within each cluster is relatively low). It is important to note that the physiographic (soil and land use) characteristics had more influence on the clustering as opposed to the climatic properties; the clustering was repeated without climatic information but the spatial pattern of the clusters remained. In the last stage of the stepwise calibration procedure, the regional model parameters were estimated for each cluster region. When using the clustering for regional calibration (Section 5.4), however, it could not significantly improve the overall model performance but nevertheless, the model consistency at all stations was improved. Overall, we found a high potential of catchment similarity concepts to drive parameter identification in the ungauged basins.

### 4.1-3. Quality checks

Steps 1-3 of our best practices were performed in an iterative procedure including checking against independent data sources that resulted in reconsiderations of assumptions and corrections of input data. For instance, the proportion of each land use type driven by GLC2000 was calculated and compared to soft information from official governmental reports. According to GLC2000 11% of the country is forest, which contradicts the estimated 22% based on reports from the Ministry of Water Resources (India-WRIS, 2012, River Basin Atlas of India, RRSC-West, NRSC, ISRO, Jodpur, India). To address this, forest information from the Global Irrigated Area Mapping (GIAM; Thenkabail et al., 2009) was merged with GLC2000. Although the proportion of forest areas was corrected, this merging consequently changed the proportion of open land with vegetation and crops from 14 and 68% to 12 and 59% respectively.

In addition, several modelled and observed flow signatures were compared repetitively at every stage of model refinement. We found it valuable to adjust as much as possible before starting to work on parameter values and model algorithms. For instance, the analysis of flow time series and signatures during the first model runs showed consistent underestimation of runoff in the Himalayan-fed basins. A comparison of the mean annual precipitation between Aphrodite and national precipitation gridded data provided by the Indian Meteorological Department, showed an underestimation of the Aphrodite precipitation in the mountainous regions; the Aphrodite precipitation network is sparse over Himalaya (Yatagai et al., 2012). To overcome this underestimation, a correction factor was applied to precipitation (in HYPE, this was a multiplier of 4% per 100 m) at regions with elevation greater than 400 m. Allowing such modification in the data, we expected that calibration of model parameters could further compensate precipitation uncertainty.

### 4.4. Model – Right for the right reasons

When setting up India-HYPE we considered realism in the process calculations by using parameter constraints. We did not have to adjust the model structure and we did not assimilate data or rating curves as we did not have access to such observations.

*Additional data sources*

The calibration of pot. E model routine against the MODIS pot. E data resulted in a well identified coefficient of potential evapotranspiration (*cevp*) values for most land use types. Analysis of the

Monte Carlo results presents an initial screening of parameter sensitivities (Fig. 7). Results show that the different objective functions extract different information from the pot. E spatial pattern. As expected, *cevp* values for crops, forest and open land with vegetation types are the most sensitive to both objective functions, since these land use types dominate the region (60, 23 and 11% of India respectively) and hence significantly affect pot. E. Overall India-HYPE was lower in pot. E at the arid regions and over the Himalayas (on average by 15%), whereas it was higher in pot. E along the western and eastern coast lines (on average by 12%). Although the two estimates do not fully match, the use of additional information to constrain parameters (hence constraining the model's results for specific processes) is promising. However, the uncertainty of MODIS results was not examined and more data sources should be included.

*Expert knowledge*

Expert knowledge was applied to filter out unrealistic relationships of the *mactrsm* parameter for different soil types (see section 3.3). Both the constrained and unconstrained models resulted in a comparable calibration performance; median KGE was 0.48 and 0.49 for the constrained and unconstrained models respectively. The optimum set for the unconstrained model gave an unrealistic distribution of the parameter values for the coarse and medium soil types (Fig. 8). However, the optimum values are within the parameter range defined in the constrained calibration approach. The slight increase is due to the free calibration parameters whose values and/or distributions are allowed to compensate for errors/uncertainties at other processes. In such cases it is important to select the model which performs well and respects the theoretical understanding of the system. This illustrates the value of the recommendations to constrain parameters based on expert knowledge – the right model for the right reason.

*Stepwise calibration procedure*

The predictability of the model with prior parameter values was very poor (Fig. 9), highlighting the limitations when parameters are regionalised from a donor system of strongly different hydro-climatic characteristics (e.g. Sweden). A significant improvement in the performance is achieved in both calibration and evaluation period after the calibration of the general parameters due to a better representation of the water volume in the rivers (*beta* in KGE improved from 0.51 to 0.78). Calibration of the soil and land use parameters further improved the performance; however KGE was slightly decreased at the poorly performed basins of the previous calibration step. Using the clusters based on catchment similarities for regional calibration did not significantly improve the overall model performance, however, the model consistency at all stations was improved in both calibration and evaluation periods.

### 4.5. Hydrological interpretation

The temporal interpretation was done by analysing interacting dynamics of internal model variables, i.e. precipitation (P, mm), snow depth (SD, mm), temperature (T, °C), evapotranspiration (E, mm), soil moisture deficit (SMDF, mm), and discharge (Q, $m^3/s$). These are checked visually in a set of validation basins, to avoid unrealistic model behaviour due to parameter setting. Results from this point onwards correspond to the calibrated India-HYPE model (after step 3 in Fig. 9). Results in the Chenab River at the Akhnoor station (branch river of the Indus system; asterisk 3 in Fig. 2) show that the snow melt characterises the monthly hydrograph (Fig. 10). Snow accumulation/melting processes occur at the headwaters of the basin which experience T below 0 °C during the winter and pre-monsoon period and above 0 °C during the rest of the months ("Up" black-dashed T series in Fig. 10). P also varies in space while it exhibits strong seasonal variability according to the location ("Up" black-line and "Down" blue envelope in the P series). Spatiotemporal analysis of P allows a better understanding of the snow depth temporal distribution; in the model, snow depth increases when precipitation occurs and temperature is below 0 °C. Given the model's evapotranspiration module, potential E varies depending on mean temperature. However the distribution of actual E is dependent on the water availability in the soil, which further justifies the strong (negative) correlation between actual E and SMDF.

For spatial interpretation of flow predictions, we investigated potential relationships between model performance and physiographic-climatic characteristics; hence identify the controls of poor model performance. Fig. 11 shows the classification tree obtained when relating the KGE performance with physical and climatic characteristics across the domain. Results show that the dominant variables resulting in poor/good model performance are soil (medium and shallow) and climate (mean precipitation and temperature). Despite the relatively small sample is this analysis, results are insightful and show that poor performance (KGE<0.4) is generally achieved at basins with shallow soil type greater than 13%. The probability of obtaining poor performance is also highest for basins with medium soil type greater than 34% and precipitation less than 1038 mm. Consequently, emphasis should be given to parameters for medium and shallow soils in a future effort to improve the model performance.

### 4.6. Uncertainty – local and regional

The India-HYPE model was calibrated and validated in space and time and the overall model performance (at the end of the stepwise approach) in terms of KGE (Gupta et al., 2009) and its decomposed terms is presented in Table 3. India-HYPE achieved an acceptable performance and is therefore considered adequate to describe the dominant hydrological processes in the subcontinent. However, the performance decreased (from KGE=0.64 to KGE=0.44) when the model is evaluated for gauges, which are independent both in space and time. This shows that the model still needs improvements to be equally reliable for predictions in ungauged basins at independent time-periods. The decomposed KGE terms show that the model during the validation period and for the validation stations cannot fully capture the variability of the observed data (described by the *alpha* term). *alpha* decreases during the validation period at the validation stations from 0.78 to 0.58 which consequently affects the KGE values. However other flow characteristics, i.e. timing and volume, are well represented also during the validation period.

To search for major uncertainties and potential for improvements, we finally analyse the model performance in both the calibration and validation stations across the domain. The ability of the model to reproduce the monthly variability of discharge varies regionally as shown by the KGE (Fig. 12). Performance is generally poor in the mountainous and semi-arid regions (western and eastern Himalayas and northwest India respectively). The Indian river-basins are also regulated limiting the model's predictive power; regulation strategies are irregular and difficult to reproduce. The KGE's decomposed terms (*cc*, *alpha* and *beta*) can reveal the causes for the model errors. For example, the poor performance at the Indus river system (north India) is due to the poor representation of the observed variability of discharge, which is probably related to parameterisation in the model's snow accumulation/melting component. In addition, mass volume error seems to be the main cause of poor KGE performance in the south-western rivers. This seems to be due to the under-estimation of precipitation and/or over-estimation of actual evapotranspiration; comparison of APHRODITE data against precipitation data from the Indian Meteorological Department showed underestimation of precipitation in this region. Conclusions are similar for the stations used in calibration and validation analysis; hence justify the model's spatial consistency in the region.

### 4.7. Spatial flow pattern across the subcontinent and dominant processes

Although the India-HYPE model has limitations, we identified potential for further improvements during the set-up procedure. The present version already demonstrated the usefulness of multi-basin modelling for comparative hydrology and how to gain insights in spatial patterns of flow generating processes at the large scale. The final clustering analysis of the 12 flow signatures from India-HYPE version 1 resulted in six different classes of varying size (Fig. 13) with different distribution in signatures (Fig. 14). Similarity in catchment behaviour for each class was interpreted and dominant flow generating processes could be distinguished as follows:

Catchments in *cluster 3* are located in the Himalayan region and in the western Indian coast (Western Ghats) and are characterised by high ranges of annual specific runoff (Qm) due to high precipitation occurring in these regions, and variable flow regime (high mFDC). Variability is dependent on snow/ice processes which are important in controlling the flow regime, at least in the Himalayan region (c.f. annual cycle in the Indus River in Fig. 2). Flow is also characterised by high rising and declining limb densities (RLD and DLD). The climate in catchments of *cluster 3* is humid subtropical and tropical with high evapotranspiration. Catchments in the north-western part of India (*cluster 4*; arid regions including the Thar Desert) are characterised by high intra-annual variability (DPar) and low values of flow (q95). Ephemeral rivers exist in this region due to high evaporation rate (e.g. Luni river), and generate runoff mainly during the monsoon period. The high variability in the flow regime is also shown by the high values of CV, Flash and RLD signatures. Similar flow characteristics are observed for the catchments located in the semi-arid regions (*cluster 1*), yet not at the same range of signature values as for *cluster 4*. The catchments in *cluster 1* are also fast responsive and their flow shows strong dynamics, in terms of rising (RLD) and declining limb densities (DLD). Catchments in *cluster 2* are located in the tropical climate and their runoff response is mainly driven by rainfall. Although these catchments receive less precipitation compared to other regions, their normalised high flow statistic (q05) is the highest of any cluster group. Moreover, catchments in *cluster 5* are located at the downstream areas of the Indus River distinguished for their high values of low flows. Finally, catchments in *cluster 6* are characterised for their high mean annual discharge values and are located at the downstream areas of the large river systems (Indus, Ganga and Brahmaputra). Note also that only few catchments belong to these cluster groups; 112 and 57 catchments in *cluster 5* and *6* respectively.

Repeating the clustering analysis at two different steps of the calibration procedure can assess changes in the understanding of hydrological response in the region. Fig. 13 shows that parameterisation can affect the spatial pattern of clusters in terms of catchment functioning. In particular, clusters after calibration (Regional step) seem to have a consistent spatial structure; this also justifies the validity of parameter regionalisation approaches based a spatial proximity between catchments. Results from clustering based on physiography show spatial consistency in the arid region (Thar Desert) and the western coast (Western Ghats) respectively. This affected identification of the regional parameters (multipliers of precipitation and evapotranspiration) applied at the subbasin scale, which consequently led to a more consistent spatial structure in the mapping (c.f. Fig. 13a and 13b). Finally, calibration of the soil and land use parameters led to a better representation of snow processes and hence affected the flow signatures in the Himalayan region (*cluster 3*).

### 4.8. Performance in India-HYPE v1.0 and future model refinements

Many other catchment-scale and multi-basin hydrological models have been applied in (parts of) the Indian subcontinent. However, it is generally common that only results from success stories are reported which limits the potential for comparative analyses and hence improving process understanding. Here, we presented results from all 42 Indian GRDC stations including both failure and success. We closed the adjustments of the first model version and documented the India-HYPE version 1.0 providing also guidelines on how to start working on the next version, looping back to step 1 again. Overall, India-HYPE performed well for most river systems with the performance being comparable to other studies, in which a model was applied at the large scale. Application of the VIC hydrological model resulted in a similar performance for the large systems of Ganges, Krishna and Narmada (Raje et al., 2013) with the Nash-Sutcliffe Efficiency, NSE (Nash and Sutcliffe, 1970) varying between 0.44 and 0.94 (at the same stations India-HYPE achieved NSE between 0.45 and 0.94). In contrast to previous studies, our contribution lies in the fact that anthropogenic influences (i.e. reservoirs and irrigation) are simulated, as those have been shown to be very important controlling the amplitude, phase and shape of the hydrograph. Other models, i.e. SWAT, have also been applied in India to assess the impacts of climate change; however the parameters have been estimated empirically from the literature, whilst the performance was not reported (Gosain et al., 2006, 2011).

Catchment-scale hydrological models from India have generally been achieving high performance (Arora, 2010; Patil et al., 2008), mainly due to the local gauged data used; usually the data are governmental and confidential with high spatiotemporal resolution and less uncertainty/error. In addition, model parameters in single catchments are normally transferred along a smoother hydro-climatic gradient and are calibrated for individual gauging stations. Nevertheless, catchment-scale studies set a benchmark of performance and provide deeper knowledge of process description which further leads to refinements in multi-basin modelling. Of particular interest are the investigations about the western Himalayas, in which India-HYPE performed poorly. Studies by Singh and Bengtsson (2004), Singh and Jain (2003) and Singh et al. (2006) highlight the importance of accumulation/melting processes in the snow-/glacier-fed parts of the region accounting for 17% each to total discharge; however for other regions of the Indus system higher contributions from snow and ice are reported (Immerzeel et al., 2009). The poor model performance in terms of *alpha* (variability) and *beta* (volume) highlights the need to refine the current snow/glacier algorithms, and/or improving the parameters by using this soft information in model evaluation. Similar model needs can be concluded when assessing the India-HYPE performances at the Ganges and Brahmaputra basins based on previous literature (Arora, 2010; Nepal et al., 2014). Finally results for the arid northwest and mountainous regions highlighted the need to refine the pot. E algorithm. Most regional hydrological studies considered relationships including extraterrestrial radiation and relative humidity, i.e. Hargreaves-Samani or Penman-Monteith, which are expected to improve the magnitude and variability of evapotranspiration losses (Samaniego et al., 2011). Therefore the pot. E model component will be further investigated and refined in the next version of India-HYPE.

## 5. CONCLUSIONS

When investigating the modified recommendations for predictions in ungauged basins across the Indian subcontinent, we found that:

- Each step in the best practice procedure was relevant and we could find methods that also work at the large scale using the knowledge derived for catchments during the PUB decade. We argue to adapt an incremental and agile approach to model set-up, which requires frequent testing to get feedback on introduced changes. The large-scale modelling is more prone to technical problems and data inconsistencies that become apparent when running the model and therefore it should be done early in the model set-up process.
- Multi-basin modelling of ungauged rivers at the large scale reveals insight in spatial patterns and dominating flow processes. Indian catchments can be categorised into 6 clusters based on their flow similarity. River flow varies spatially in terms of flow means, variability, extremes and seasonality. Catchments in the Himalayan region and the Western Ghats seem to respond similarly and are characterised by high mean annual specific runoff values and variable flow regime. Response of the catchments in the tropical zone is characterised by high peaks, while catchments in the dry regions show very strong flow variability and respond quickly to rainfall.
- Overall the model showed high potential to represent the hydrological response across the region despite the strong hydro-climatic gradient. However, the India-HYPE v.1.0 still needs to be improved to be equally reliable for predictions in ungauged basins as for gauged rivers. The model set-up procedure according to the PUB recommendations brought insights on where the single model structure did not perform well. Based on this, future model improvements will mainly focus on the western Himalayas and arid regions by refining the hypothesis of snow/glacier processes and the evapotranspiration algorithm.

**ACKNOWLEDGEMENTS**

We are very grateful for the funding of this research by the Swedish International Development Cooperation Agency (Sida) through the India-HYPE project (AKT-2012-022) and the Swedish Research Council (VR) through the WaterRain-Him project (348-2014-33). The investigation was performed at the SMHI Hydrological Research unit, where much work is done jointly. We would especially like to acknowledge contributions from David Gustafsson, Göran Lindström, Jafet Andersson, Kean Foster, Kristina Isberg, and Jörgen Rosberg for assistance with background material for this study. The authors would finally like to express their sincere gratitude to two anonymous reviewers for their constructive comments. Their detailed suggestions have resulted in an improved manuscript. The HYPE model code is open source and can be retrieved with manuals at

http://hype.sourceforge.net/. Time-series and maps from the India-HYPE model (including climate change impact studies) are available for inspection at http://hypeweb.smhi.se. The work contributes to the decadal research initiative “Panta Rhei” by the International Association of Hydrological Sciences (IAHS) under Target 2 “Estimation and Prediction” and its two working groups on Large Samples and Multiple ungauged basins, respectively.

## APPENDIX A: DEFINITION OF PEFORMANCE METRICS AND FLOW SIGNATURES

The Kling-Gupta Efficiency (KGE) is defined as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$$

where $r$ is the linear cross-correlation coefficient between observed and modelled records, $\alpha$ is a measure of variability in the data values (equal to the standard deviation of modelled over the standard deviation of observed), and $\beta$ is equal to the mean of modelled over the mean of observed. For a perfect model with no data errors, the value of KGE is 1; hence $r$, $\alpha$ and $\beta$ are also 1. In addition, we transform the three KGE components to results into a consistent range of possible values. Consequently we consider:

$$cc = 1 - \sqrt{(r-1)^2}$$

$$alpha = 1 - \sqrt{(\alpha-1)^2}$$

$$beta = 1 - \sqrt{(\beta-1)^2}$$

where the range of values for each term varies between -∞ and 1 with 1 being the optimum.

In this paper we quantify the signatures by single values. Given the time series of observed (or modelled) specific daily runoff $Q_d(t)$ (mm $d^{-1}$), the calculated signatures are given in Table A1.

Table A1. Flow signatures used for model evaluation and catchment functioning.

| Signature | Abbreviation | Reference |
|---|---|---|
| Mean annual specific runoff | Qm | (Viglione et al., 2013) |
| Normalised high flow | q05 | (Viglione et al., 2013) |
| Normalised low flow | q95 | (Viglione et al., 2013) |
| Normalised relatively low flow | q70 | (Viglione et al., 2013) |
| Slope of flow duration curve | mFDC | (Viglione et al., 2013) |
| Range of Pardé coefficient | DPar | (Viglione et al., 2013) |
| Coefficient of variation | CV | (Donnelly et al., 2015) |
| Flashiness | Flash | (Donnelly et al., 2015) |

| | | |
|---|---|---|
| Normalised peak distribution | PD | (Euser et al., 2013) |
| Rising limb density | RLD | (Euser et al., 2013) |
| Declining limb density | DLD | (Euser et al., 2013) |
| Long term mean discharge | Qdm | (Donnelly et al., 2015) |

**REFERENCES**


Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T. and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrol. Sci. J., 48(3), 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration, Guidelines for computing crop water requirements, in FAO Irrigation and drainage paper 56, Rome., 1998.

Andreassian, V., Hall, A., Chahinian, N. and Schaake, J.: Large Sample Basin Experiment for Hydrological Model Parameterization: Results of the Model Parameter Experiment - MOPEX, IAHS Publication 307, Wallingford., 2006.

Arheimer, B. and Brandt, M.: Modelling nitrogen transport and retention in the catchments of southern Sweden, Ambio, 27(6), 471–480, 1998.

Arheimer, B., Dahné, J., Lindström, G., Marklund, L. and Strömqvist, J.: Multi-variable evaluation of an integrated model system covering Sweden (S-HYPE), IAHS Publ., 345, 145–150, 2011.

Arheimer, B. and Lindström, G.: Implementing the EU Water Framework Directive in Sweden, in Runoff Predictions in Ungauged Basins - Synthesis across processes, places and scales, edited by G. Blöschl, M. Sivapalan, T. Wagener, and A. Viglione, pp. 353–359, Cambridge University Press, Cambridge, UK., 2013.

Arora, M.: Estimation of melt contribution to total streamflow in river Bhagirathi and river DhauliGanga at Loharinag Pala and Tapovan Vishnugad project sites, J. Water Resour. Prot., 02(07), 636–643, doi:10.4236/jwarp.2010.27073, 2010.

Attri, S. D. and Tyagi, A.: Climate profile of India, in Government of India Ministry of Earth Sciences, p. 129, New Delhi., 2010.

Bartholomé, E., Belward, A. S., Achard, F., Bartalev, S., Carmona Moreno, C., Eva, H., Fritz, S., Grégoire, J.-M., Mayaux, P. and Stibig, H.-J.: GLC 2000 Global Land Cover mapping for the year 2000, European Commission, DG Joint Research Centre, EUR 20524 EN, Ispra., 2002.

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

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J.: Classification and Regression Trees, CRC Press, Wadsworth, Belmont, CA., 1984.

Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G. and Wheater, H.: Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis, Water Resour. Res., 48(6), W06519, doi:10.1029/2011WR011207, 2012.

Bulygina, N., McIntyre, N. and Wheater, H. S.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, Hydrol. Earth Syst. Sci., 13, 893–904, 2009.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments, Water Resour. Res., 48, W05552, doi:10.1029/2011WR011721, 2012.

Donnelly, C., Andersson, J. C. M. and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, Hydrol. Sci. J., doi:10.1080/02626667.2015.1027710, 2015.

Donnelly, C., Rosberg, J. and Isberg, K.: A validation of river routing networks for catchment modelling from small to large scales, Hydrol. Res., 44(5), 917–925, doi:10.2166/nh.2012.341, 2013.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17(5), 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.

Falkenmark, M. and Chapman, T.: Comparative hydrology: An ecological approach to land and water resources, UNESCO, Paris, France., 1989.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44(1), 1–13, doi:10.1029/2006WR005563, 2008.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S. and Burlando, P.: The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, Water Resour. Res., 47, W07519, doi:10.1029/2010WR009824, 2011.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S. and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18(5), 1895–1915, doi:10.5194/hess-18-1895-2014, 2014.

Good, P. I.: Resampling methods: A practical guide to data analysis, 3rd ed., Birkhäuser, Boston., 2006.

Gosain, A., Rao, S. and Arora, A.: Climate change impact assessment of water resources of India, Curr. Sci., 101(3), 356–371, 2011.

Gosain, A., Rao, S. and Basuray, D.: Climate change impact assessment on hydrology of Indian river basins, Curr. Sci., 90(3), 346–353, 2006.

Gupta, H. V, Wagener, T. and Liu, Y.: Reconciling theory with observations : elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802– 3813, doi:10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18(2), 463–477, doi:10.5194/hess-18-463-2014, 2014.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G. and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resour. Res., 50, 7445–7469, doi:10.1002/2014WR015484, 2014.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Immerzeel, W. W., Droogers, P., de Jong, S. M. and Bierkens, M. F. P.: Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing, Remote Sens. Environ., 113(1), 40–49, doi:10.1016/j.rse.2008.08.010, 2009.

Johnston, R. and Smakhtin, V.: Hydrological modeling of large river basins: how much is enough?, Water Resour. Manag., 28(10), 2695–2730, doi:10.1007/s11269-014-0637-8, 2014.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y. and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol. Earth Syst. Sci., 17(7), 2845–2857, doi:10.5194/hess-17-2845-2013, 2013.

Kumar, R., Livneh, B. and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resour. Res., 49(9), 5700–5714, doi:10.1002/wrcr.20431, 2013.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, J. Hydrol., 296(1-4), 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N. and Wisser, D.: Global Reservoir and Dam (GRanD) database - Technical documentation version 1.1, Bonn., 2011.

Lehner, B., Verdin, K. and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, Trans. AGU, 89(10), 93–94, doi:10.1029/2008EO100001, 2008.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41(3-4), 295–319, doi:10.2166/nh.2010.007, 2010.

McDonnell, J. J., McGuire, K., Aggarwal, P., Beven, K. J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., Wörman, A. and Wrede, S.: How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis, Hydrol. Process., 24(12), 1745–1754, doi:10.1002/hyp.7796, 2010.

McMillan, H., Tetzlaff, D., Clark, M. and Soulsby, C.: Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach, Water Resour. Res., 48(5), W05501, doi:10.1029/2011WR011688, 2012.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z. and Belyaev, V.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrol. Sci. J., 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

Mu, Q., Heinsch, F. A., Zhao, M. and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, doi:10.1016/j.rse.2006.07.007, 2007.

Mu, Q., Zhao, M. and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115(8), 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Nachtergaele, F., van Velthuizen, H., Verelst, L. and Wiberg, D.: Harmonized world soil database version 1.2, FAO, Rome and IIASA, Laxenburg, Austria., 2012.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, J. Hydrol., 10, 282–290, 1970.

Nepal, S., Krause, P., Flügel, W.-A., Fink, M. and Fischer, C.: Understanding the hydrological system dynamics of a glaciated alpine catchment in the Himalayan region using the J2000 hydrological model, Hydrol. Process., 28(3), 1329–1344, doi:10.1002/hyp.9627, 2014.

Patil, J. P., Sarangi, A., Singh, A. K. and Ahmad, T.: Evaluation of modified CN methods for watershed runoff estimation using a GIS-based interface, Biosyst. Eng., 100(1), 137–146, doi:10.1016/j.biosystemseng.2008.02.001, 2008.

Pechlivanidis, I. G., Jackson, B., McIntyre, N. and Wheater, H. S.: Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, Glob. NEST J., 13(3), 193–214, 2011.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proc. R. Soc. London Ser. A Math. Phys. Sci., 193, 120–145, 1948.

Portmann, F. T., Siebert, S. and Döll, P.: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochem. Cycles, 24(1), GB1011, doi:10.1029/2008GB003435, 2010.

Raje, D., Priya, P. and Krishnan, R.: Macroscale hydrological modelling approach for study of large scale hydrologic impacts under climate change in Indian river basins, Hydrol. Process., 28(4), 1874–1889, doi:10.1002/hyp.9731, 2013.

Refsgaard, J. C., Storm, B. and Clausen, T.: Système Hydrologique Europeén (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modelling, Hydrol. Res., 41(5), 355, doi:10.2166/nh.2010.009, 2010.

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, Hydrol. Res., 42(5), 338–355, doi:10.2166/nh.2011.156, 2011.

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, Hydrol. Earth Syst. Sci., 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P. and Portmann, F. T.: Groundwater use for irrigation - A global inventory, Hydrol. Earth Syst. Sci., 14(10), 1863–1880, doi:10.5194/hess-14-1863-2010, 2010.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K. and Feick, S.: Development and validation of the global map of irrigation areas, Hydrol. Earth Syst. Sci., 9, 535–547, doi:10.5194/hessd-2-1299-2005, 2005.

Singh, P., Arora, M. and Goel, N. K.: Effect of climate change on runoff of a glacierized Himalayan basin, Hydrol. Process., 20(9), 1979–1992, doi:10.1002/hyp.5991, 2006.

Singh, P. and Bengtsson, L.: Hydrological sensitivity of a large Himalayan basin to climate change, Hydrol. Process., 18(13), 2363–2385, doi:10.1002/hyp.1468, 2004.

Singh, P. and Jain, S. K.: Modelling of streamflow and its components for a large Himalayan basin with predominant snowmelt yields, Hydrol. Sci. J., 48(2), 257–276, doi:10.1623/hysj.48.2.257.44693, 2003.

Singh, V.: Computer models of watershed hydrology, Water Resources Publications, Littleton, Colorado., 1995.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrol. Sci. J., 48(6), 857–880, 2003.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. and Lindström, G.: Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale, Hydrol. Sci. J., 57(2), 229–247, doi:10.1080/02626667.2011.637497, 2012.

Takeuchi, K., Blöschl, G., Savenije, H. H. G., Schaake, J., Sivapalan, M., Viglione, A., Wagener, T. and Young, G.: Recommendations, in Runoff Predictions in Ungauged Basins - Synthesis across processes, places and scales, edited by G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. H. G. Savenije, pp. 384–387, Cambridge University Press, Cambridge, UK., 2013.

Ter Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces, Stat. Comput., 16(3), 239–249, doi:10.1007/s11222-006-8769-1, 2006.

Thenkabail, P., Biradar, C., Noojipady, P., Dheeravath, V., Li, Y. J., Velpuri, M., Reddy, G. P. O., Cai, X. L., Gumma, M., Turral, H., Vithanage, J., Schull, M. and Dutta, R.: A Global Irrigated Area Map (GIAM) Using Remote Sensing at the End of the Last Millennium, Int. J. Remote Sens., 30(14), 3679–3733, 2009.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M. and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, Hydrol. Earth Syst. Sci., 17(6), 2263–2279, doi:10.5194/hess-17-2263-2013, 2013.

Widén-Nilsson, E., Halldin, S. and Xu, C.: Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, J. Hydrol., 340(1–2), 105–118, doi:10.1016/j.jhydrol.2007.04.002, 2007.

Yasutomi, N., Hamada, A. and Yatagai, A.: Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation, Glob. Environ. Res., 3, 165–172, 2011.

Yatagai, A., Arakawa, O. and Kamiguchi, K.: A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Sola, 5, 137–140, doi:10.2151/sola.2009–035, 2009.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. and Kitoh, A.: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bull. Am. Meteorol. Soc., 93(9), 1401–1415, doi:10.1175/BAMS-D-11-00122.1, 2012.

Table 1. Data sources and characteristics of the India-HYPE v.1.0 model set-up.

| *Characteristic/Data type* | *Info/Name* | *Provider* |
|---|---|---|
| Total area ($km^2$) | 4.9 million | - |
| Number of subbasins | 6 010 (mean size 810 $km^2$) | - |
| Topography (routing and delineation) | Hydrosheds (15 arcsec) | Lehner et al. (2008) |
| Soil characteristics | Harmonised World Soil Database (HWSD) | Nachtergaele et al. (2012) |
| Land use characteristics | Global Land Cover 2000 (GLC2000) | Bartholomé et al. (2002) |
| Reservoir and dam | Global Reservoir and Dam database (GRanD) | Bernhard et al. (2011) |
| Lake and wetland | Global Lake and Wetland Database (GLWD) | Lehner and Döll (2004) |
| Agriculture | MIRCA2000 | Portmann et al. (2010) |
| Irrigation | Global Map of Irrigation Areas (GMIA) | Siebert et al. (2005)<br>Siebert et al. (2010) |
| Discharge | Global Runoff Data Centre (GRDC; 42 stations) | http://www.bafg.de/GRDC |
| Precipitation | APHRODITE (0.25° × 0.25°) | Yatagai et al. (2012) |
| Temperature | AphroTEMP (0.5° × 0.5°) | Yasutomi et al. (2011) |
| Potential evapotransp. | MODIS pot. E (1 km) | Mu et al. (2011) |

Table 2. Statistics for the 42 gauging stations of river discharge used in the model evaluation.

| | *Percentiles* | | | | | |
|---|---|---|---|---|---|---|
| | 5% | 25% | Median | 75% | 95% | Mean |
| Basin surface ($km^2$) | 2 062 | 12 691 | 32 770 | 68 522 | 294 524 | 75 493 |
| Mean annual runoff (*Qm*, mm) | 40 | 168 | 377 | 648 | 2 090 | 582 |
| *Inter-annual variability of runoff (%) | 20 | 28 | 40 | 61 | 102 | 48 |

**Values of inter-annual variability correspond to coefficients of variation calculated on 9 year periods*

Table 3. Median model performance for calibration and validation stations and periods.

| Space | Time | KGE | *cc* (timing) | *alpha* (variability) | *beta* (volume) |
|---|---|---|---|---|---|
| Cal. (30 stations) | Cal. (1971-1975) | 0.64 | 0.93 | 0.78 | 0.75 |
| | Val. (1976-1979) | 0.62 | 0.92 | 0.81 | 0.80 |
| Val. (12 stations) | Cal. (1971-1975) | 0.64 | 0.91 | 0.78 | 0.79 |
| | Val. (1976-1979) | 0.44 | 0.84 | 0.58 | 0.75 |

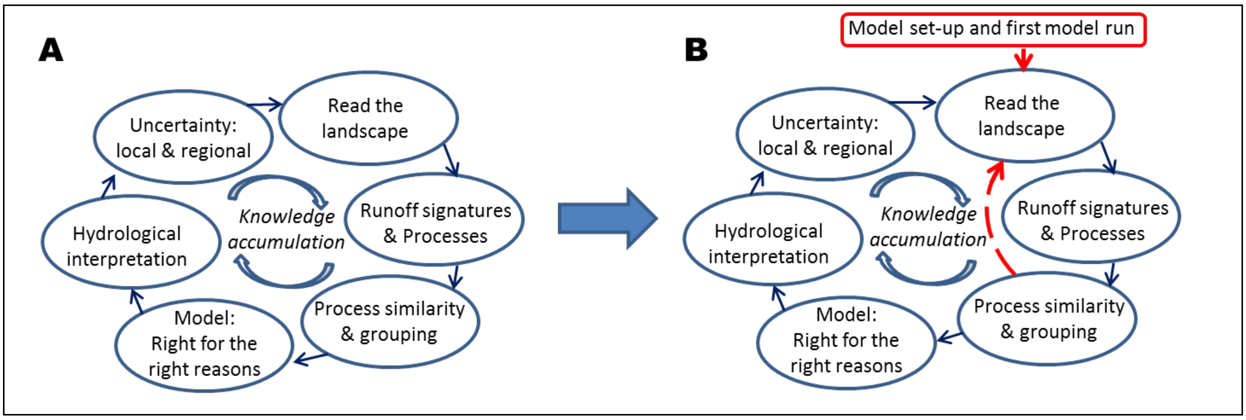


Figure 1. Best practices for predictions in ungauged basins: A) according to Fig. 13.1 by Takeuchi et al. (2013) in Blöschl et al. (2013), and B) modified version for multi-basin applications at the large scale.

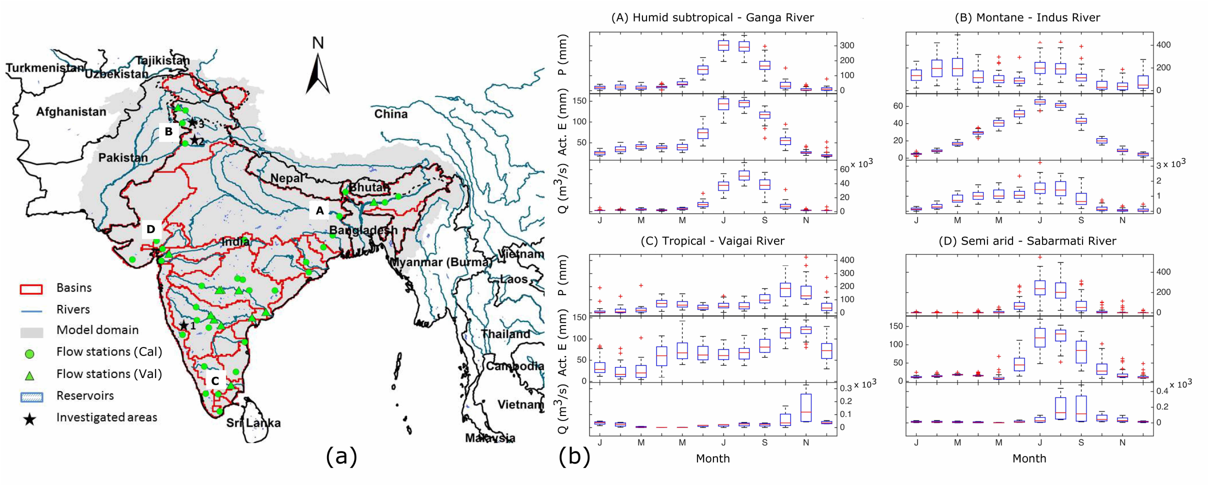


Figure 2. (a) Map of the Indian subcontinent (model domain). Results will be shown from investigation areas with a star in the order of their numbering. (b) Annual cycles (1971-1979) at four river systems (A-D) of various climate (P – observed precipitation, Act. E – modelled actual evapotranspiration, Q – observed discharge).

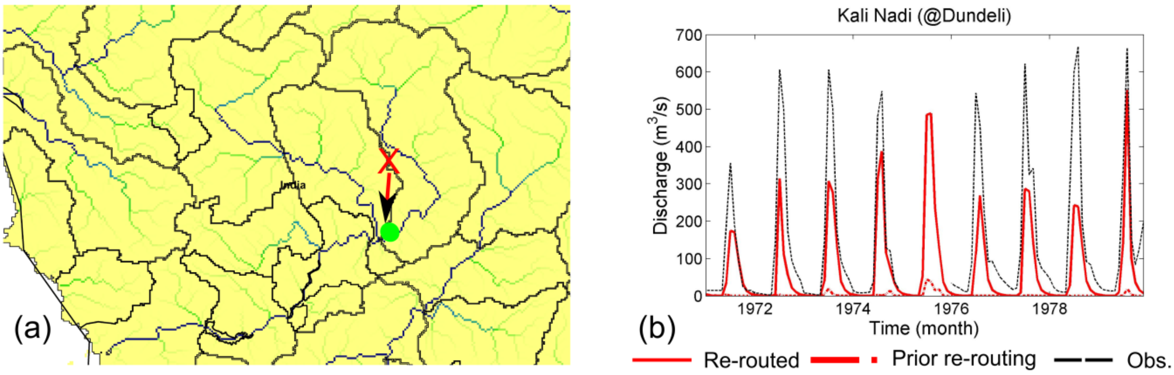


Figure 3. Example of the impact of basin delineation and routing on model behaviour: (a) correction in the location (red x and green circle is prior and after the correction respectively) of the Dundeli discharge station (Kali Nadi river basin), and (b) the corresponding modelled discharge before and after the correction. In (a) the subbasins and flow accumulation are also depicted.

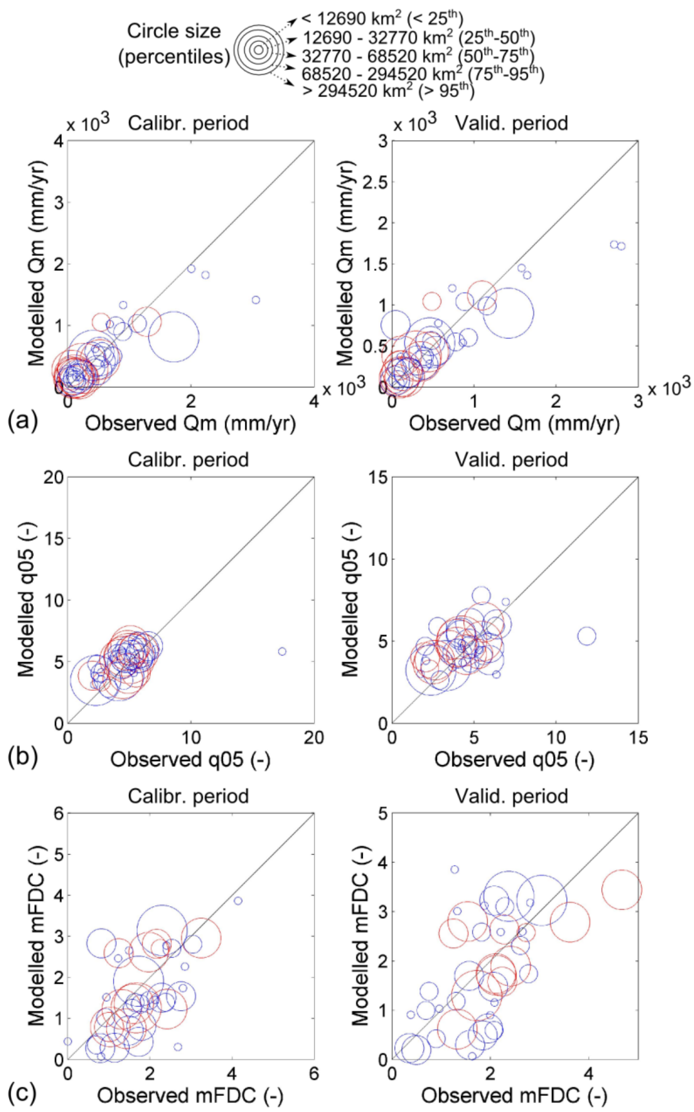


Figure 4. Signature analysis in the spatiotemporal model evaluation: (a) the mean annual specific runoff, (b) the normalised high flow statistic, and (c) the slope of the flow duration curve. Blue and red circles are used for the calibration and evaluation stations respectively.

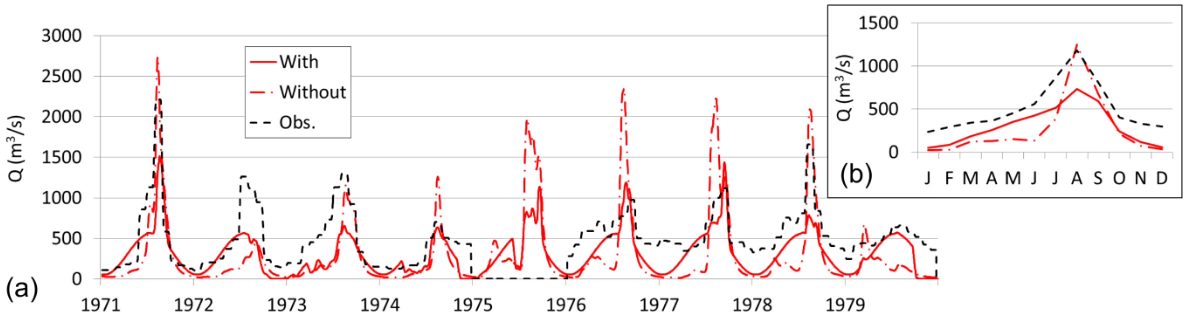


Figure 5. Impact of model parameterisation of reservoir regulation on discharge for (a) monthly streamflow, and (b) annual hydrograph, showing naturalised (without) and regulated (with) conditions at the basin outlet (located at asterisk 2 in Fig. 2).

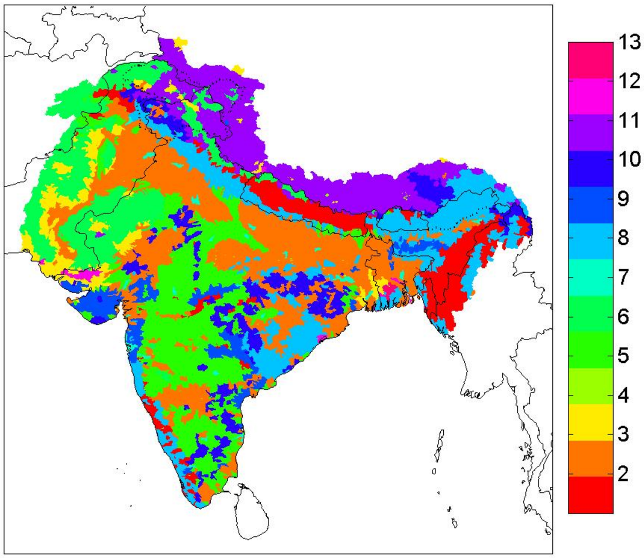


Figure 6. Subbasin clusters using a k-means clustering approach based on physiographical characteristics.

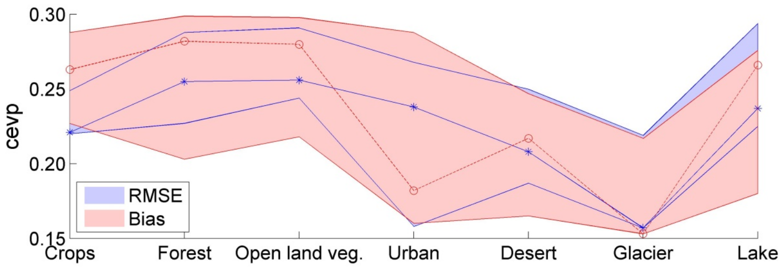


Figure 7. Coefficient of potential evapotranspiration (*cevp*) parameter as identified (the range is derived from the 100 parameter sets that perform best, and the optimum set) for different objective functions (RMSE and Bias) and land use type. Lines with markers present the optimum parameter values for different objective functions.

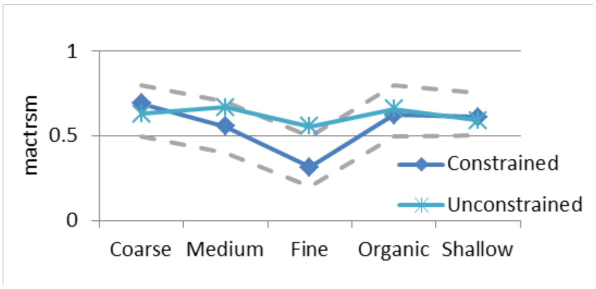


Figure 8. Constraints (grey dashed lines) and optimum (solid lines) values of the *mactrsm* soil dependent model parameter based on process understanding.

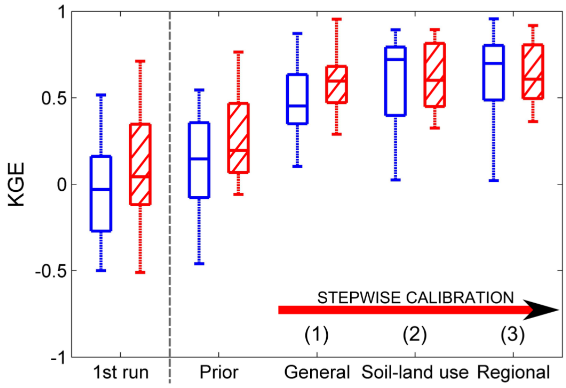


Figure 9. Improvements in model performance (average KGE for 30 stations) during the stepwise calibration approach (steps 1-3 correspond to general, soil-land use, and regional calibration as described in section 3.3). "1$^{st}$ run" corresponds to model performance of the very first model set-up to establish a technical model infrastructure. "Prior" corresponds to model performance before parameter calibration and after overcoming routing errors. The evaluation is conducted at the calibration (blue) and the validation (red shaded) period.

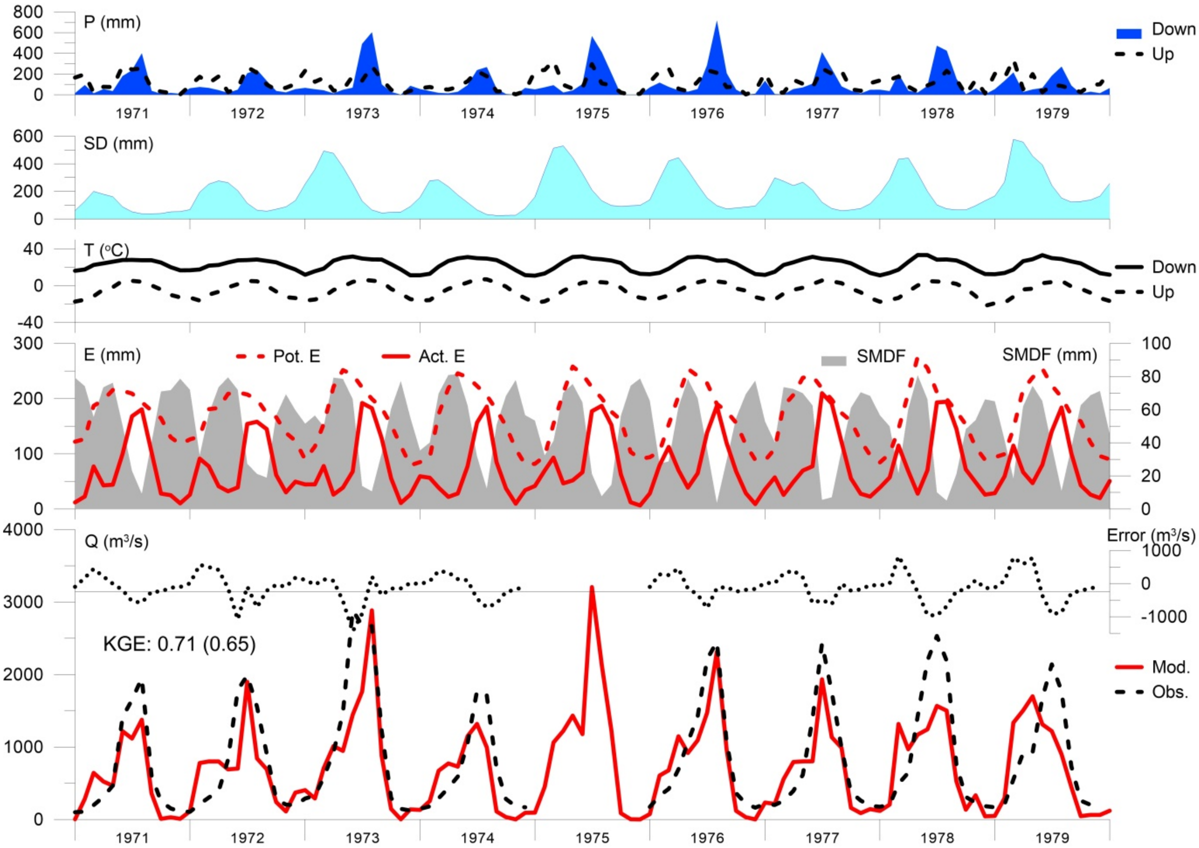


Figure 10. Analysis of model variables at the Akhnoor station (Chenab River; asterisk 3 in Fig. 2): P, SD, T, E, SMDF and Q. E corresponds to potential (Pot.) and actual (Act.) evapotranspiration, and Q corresponds to modelled (Mod.) and observed (Obs.) discharge). Note that P and T series are plotted at the outlet of the basin (Down) and the most upstream subbasin (Up).

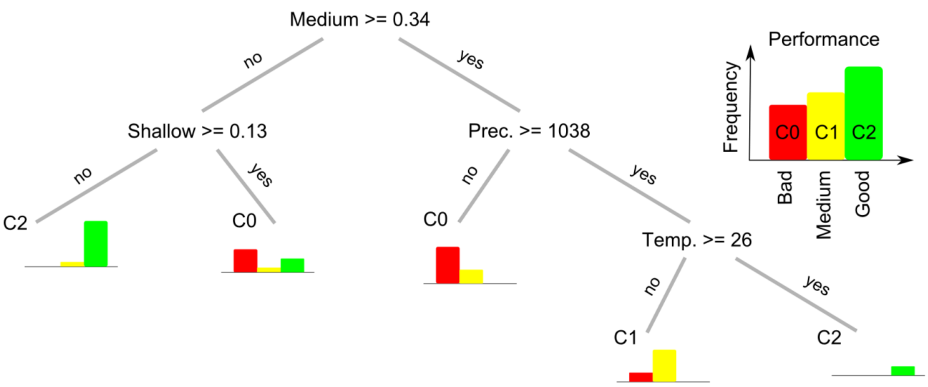


Figure 11. Classification trees relating regions of different KGE performance with physical and climatic characteristics. The bars represent the probability of a performance resulting in any of the three performance classes (C0, C1 or C2).

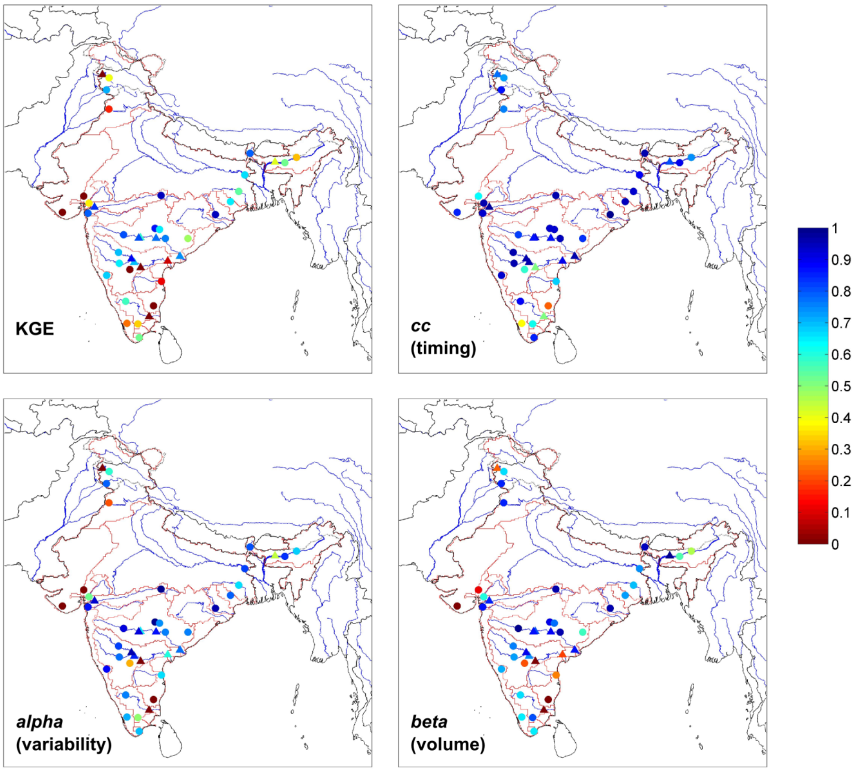


Figure 12. Spatial variability of KGE (and its decomposed terms) model performance for the calibration (circle) and validation (triangle) stations.

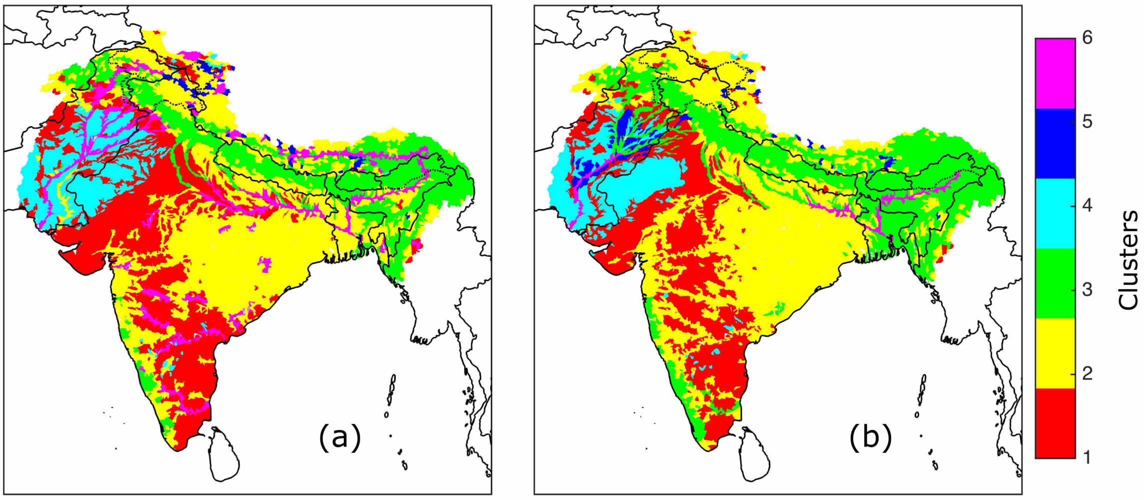


Figure 13. Subbasin clusters based on flow signatures at different stages of the model set-up: (a) Prior, and (b) Regional.

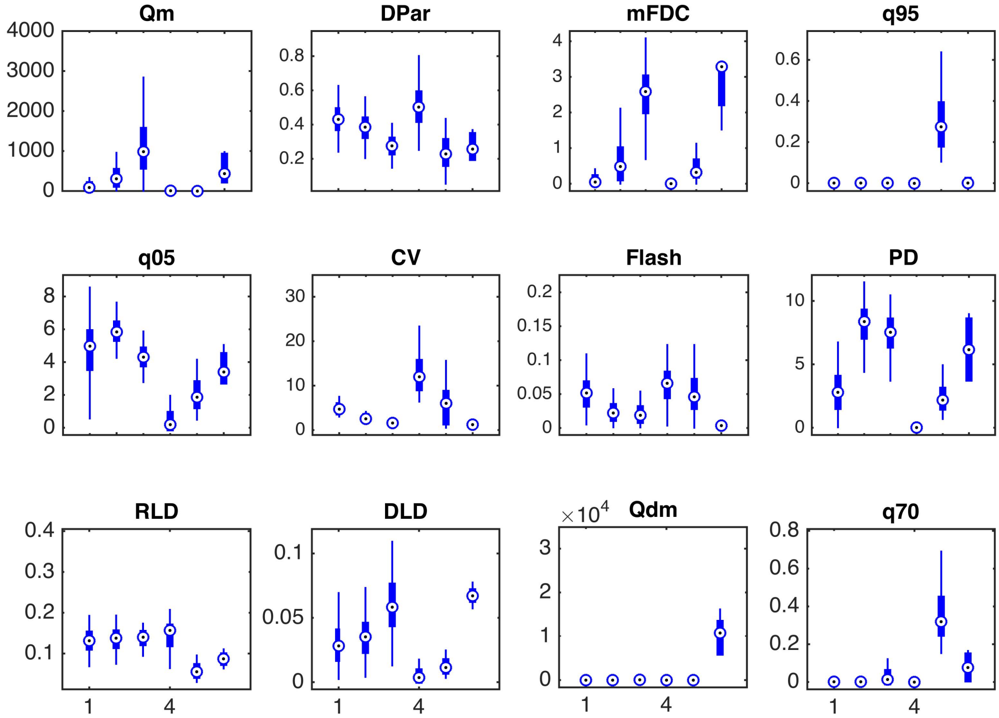


Figure 14. Distribution of signature values for each cluster (at Regional step). The flow signatures are described in Appendix A.