Hydrol. Earth Syst. Sci. Discuss., 12, 2885–2944, 2015
www.hydrol-earth-syst-sci-discuss.net/12/2885/2015/
doi:10.5194/hessd-12-2885-2015

Hydrology and Earth System Sciences
Discussions
Open Access

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

Received: 13 February 2015 – Accepted: 26 February 2015 – Published: 10 March 2015

Correspondence to: I. G. Pechlivanidis (ilias.pechlivanidis@smhi.se)

Published by Copernicus Publications on behalf of the European Geosciences Union.

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. However, such modelling involves several sources of uncertainties, which may be caused by the imperfectness of input data, i.e. particularly regional and global databases. This may lead to inaccurate model parameterisation and incomplete process understanding. In order to bridge the gap between the best practices for single catchments and large-scale hydrology, 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 on the Indian subcontinent, named India-HYPE v1.0, we explore the recommendations, indicate challenges and recommend quality checks to avoid erroneous assumptions. We identify the obstacles, ways to overcome them and describe the work process related to: (a) errors and inconsistencies in global databases, unknown human impacts, poor data quality, (b) robust approaches to identify 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 hydro-climatic gradient over the subcontinent, a single model can adequately describe the spatial variability in dominant hydrological processes at the catchment scale. Eventually, during calibration of India-HYPE, the median Kling–Gupta Efficiency for river flow increased from 0.14 to 0.64. To sum up, we demonstrate that by using the further developed PUB recommendations in processed-based large-scale models, the predictions can be consistent in both space and time for multiple basins. We describe and argue for the suggested work process

HESSD

12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

when approaching the large scale with multi-basins and big datasets. Some useful methods are presented and examples of results are given.

1 Introduction

Numerical hydrological models have been used world-wide for operational needs and scientific research since the early 1970s. In an effort to improve the reliability when modelling catchments without observed runoff data, the Prediction in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences (IAHS) was launched in 2003. In general, PUB aimed towards overcoming the fragmentation in catchment hydrology and advancing the collective understanding (Sivapalan et al., 2003). PUB highlighted the need to move beyond a model calibration philosophy towards a diagnostic evaluation approach that aims to: (i) characterise the information contained in the data and in the model, (ii) examine the extent to which a model can be reconciled with observations, and (iii) point towards the aspects of the model that need improvement (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 increasing the information content via either additional data, i.e. remote sensing, tracers, quality etc. (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 (Falkenmark and Chapman, 1989; Blöschl et al., 2013). 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). This initiative has given new momentum on putting science into practice by linking change in hydrology and society.

Multi-basin modelling can complement the "deep" knowledge from catchment-based modelling when applied to a large geographical domain covering a large sample of observations (Andreassian et al., 2006; Gupta et al., 2014; Johnston and Smakhtin, 2014). However, the majority of basins world-wide are effectively ungauged, as are also, in relative terms, the subbasins (defined here as prediction points in the model set-up) in a high resolution multi-basin model at the large scale. 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). Traditionally, the resolution was poor in such models or meteorological land-surface schemes were normally applied, but the current release of open and global datasets gives new opportunities for catchment modelling. 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). In addition, it can facilitate comparative hydrology, and explore hydrological hypothesis in catchments with a wide range of environmental conditions by analysing dominant processes at each river system (Donnelly et al., 2015; Blöschl et al., 2013; Falkenmark and Chapman, 1989).

Modelling at the large scale, however, includes additional model uncertainties. Physical properties (e.g. vegetation and soil type) in large systems generally exhibit 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

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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. Finally, 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., 2013; Kauffeldt et al., 2013).

To sum up, multi-basin modelling at the large scale gives new opportunities but also challenges for hydrological research. In this paper, we present a set of examples on how the scientific advancements during the PUB decade have improved the potential for process-based hydrological modelling at the large scale. Our objective is to test the recommendations presented by Takeuchi et al. (2013) and identify specific challenges for multi-basin modelling at the large-scale when using process-based models. Furthermore, we exemplify on how to overcome these problems by using various tools and methods. So far, we have experience in setting up large-scale multi-basin models for Sweden, Europe, the Arctic basin, La Plata and Niger River basins, the Middle East and North Africa (MENA) region and the subcontinents of India (see http://hypeweb.smhi.se). One major issue we want to stress is the importance of frequent quality checks in large-scale modelling, as there are more unknowns at the larger scale and potential disinformation in the global datasets. The scientific questions in this paper are:

1. To what extent are the PUB recommendations for catchment scale also relevant for multi-basin modelling at the large scale?

2. Which obstacles can be identified when using the PUB recommendations at the large scale?

3. How can a processed-based modeller overcome these obstacles by using complementing data and evaluation methods?

HESSD
12, 2885–2944, 2015
The India-HYPE case
I. G. Pechlivanidis and B. Arheimer

To test the PUB recommendations we first had to further develop and slightly modify the best practices to be applicable at the large scale. For obstacles at the large scale, we particularly address potential failures in capturing runoff response due to uncertain/erroneous basin delineation and routing, errors in global datasets and human impact (i.e. reservoir/dams). We also illustrate the potential improvement on parameter identification by using remote sensing data and expert knowledge. We further show how regions can be grouped based on physiographic-climatic similarity, and how flow signatures and temporal variability of other modelled variables, apart from discharge, can be used to ensure "right for the right reasons" in data sparse regions. In addition, we investigate potential links between model performance and physiographic-climatic characteristics to understand model inadequacies along a physiographic-climatic gradient. 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

In the PUB book by Blöschl et al. (2013), 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 must be slightly modified and/or complemented when applied to multi-basins and using process-based models at the large scale (Fig. 1b and text below). According to the recommendations, the modeller should ideally go through all the six steps quickly and start over again in a repetitive way (Fig. 1a). In this way an understanding is developed incrementally and knowledge is accumulated as the structure, input data and parameter values are gradually improved and/or revised in the work process. Findings in all steps should be communicated in such a way that they contribute to the global and national body of knowledge in hydrology, especially process knowledge (Takeuchi

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

et al., 2013). In this section, we present our best-practice recommendations for large-scale applications of process-based models. It is based on our interpretation of the best practices and previous experience from PUB in multi-basin applications (e.g. Andersson et al., 2014; Arheimer et al., 2012; Donnelly et al., 2015; Strömqvist et al., 2012).

The first deviation from the original best practices for single catchments is that for large scale predictions with multi-basin resolution, we recommend to set-up the model directly before going into the circle of steps (Fig. 1). When doing this, it is necessary to first decide upon a model, calculation units and sites for runoff predictions. As we focus on process-based modelling, we recommend not picking just any model from the shelf, but choosing a model that includes the description of most water fluxes, storages and anthropogenic influences that can be relevant for the large geographical domain, satisfies the modelling objectives, is familiar to the modeller and can be easily set up and run for a large domain. It is also wise to have access and competence to adjust the model code as the process descriptions might need adjustments to cope with the spatial heterogeneity and all hydrological features a region contains. Then we suggest to: (i) download readily available datasets needed that covers the entire geographical domain or merge datasets to get a full coverage, (ii) define basin outlets and boundaries, e.g. using Hydrosheds (Lehner et al., 2008), taking into account coordinates of gauges, major landscape features or challenges and user requests, (iii) make a first set of model input data and make the first model run for the full domain with multi-basin resolution. This approach of setting up the full system at once facilitates findings of major obstacles as systematic errors in input data or model limitations. By getting the technical system in place immediately, it will also start delivering results at once. This facilitates an incremental and agile approach to model set-up, which helps in focusing on the most critical processes or data sources. When the model runs as a first version for the full domain, we recommend starting to improve the predictions according to the following six steps.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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 the full variety of basins in a large-scale model domain, so instead we recommend using the following methods to understand the hydrology and its dominating processes: (i) navigate on hard-copies, digitised maps and webpages (e.g. Google Earth) to check landscape characteristics, (ii) check the literature and websites for dominant processes and well-known features or hydrological challenges in the region. Thereafter, it is important to read the downloaded datasets used in the model, to ensure that they correspond to our understanding of the landscape as errors may appear when handling large datasets. Therefore: (iii) make quality checks and cross-validation towards other data sources (i.e. containing limited in space but local information), (iv) check catchment boundaries so they fit with points of interests for predictions, (v) check consistency between calculated upstream areas at selected gauges against metadata and move or remove stations with, e.g. > 10 % difference (see Donnelly et al., 2012), (vi) check quality of runoff data to assure coherence of time-series, (vii) check the meteorological datasets and the transfer from grid to subbasins; apply corrections for altitude if necessary, and finally, (viii) check the resulting long-term mean values of meteorological variables from the hydrological model towards maps from other data sources to evaluate the accuracy of the spatial pattern.

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)

Analysing large samples of time-series from numerous gauging stations is difficult as inspection of all runoff signatures in detail for each gauging station can be very time-consuming if they are in their thousands. Instead, we recommend analysing the spatial

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

variation of mean values in various flow signatures. First, we suggest to correlate observed flow signatures to up-stream catchment characteristics across the geographical domain to check if there are significant relationships, which can support our understanding of the hydrology and justify a specific model concept (Donnelly et al., 2015). It is important to explore all dominant landscape features in this process and also recognise human impact, as this is normally reflected in the observed time-series of runoff at the large scale. Second, we recommend checking the spatial correlation between averages in observed and modelled flow signatures for each gauging site across the model domain. At this step, it is also important to decide which stations to use for calibration and validation, respectively. Cross-validation, e.g. using the jackknife procedure (Good, 2005), is practically not feasible in process-based modelling of multi-basins as it would be too time-consuming. Instead we recommend hiding a part of the gauging stations to only be used for final validation in "blind tests" so that these stations are independent from any calibration or model tuning.

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 calculation units, such as Hydrological Response Units (HRUs) and other types of physiography, which are used as input data to the model. When producing these calculation units for large domains, we need to be restrictive with the number of classes and we normally remove small calculation units to speed up the model run times; both technical and conceptual concerns must be taken into account. We recommend to group landscape units and catchments as we: (i) combine dominant soil and land cover classes for the full domain and distribute their proportion into subbasins, (ii) remove classes with less than 5 % in each catchment (except for lakes, wetlands, glacier, and urban areas) and redistribute this area to the remaining classes, (iii) sepa-

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

rate lake/reservoir area into internal and main-stream outlet lakes, (iv) check that the model classes are still relevant for the purpose of modelling and external needs from end-users. When this is done, we recommend to, (v) cluster the gauges with similar up-stream characteristics (e.g. Donnelly et al., 2015) and/or system behaviour (this study) to isolate key processes for regionalisation of parameter values in the calibration procedure. This can be done by using different cluster analysis methods, and after clustering, we suggest carefully checking the spatial patterns by plotting the catchment categories on maps and compare to other data sources.

Quality checks

This is an additional step in the procedure when setting up a model for a large region in a multi-basin manner (Fig. 1). When using large datasets it is important to repeat step 1–3 in an iterative way to ensure quality in the required input data and files of the model; it is easy to fail and introduce errors when handling big 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). Moreover, the data quality should be relevant for the catchment resolution and the potential bias from transformation of data into hydrological units must be corrected. We recommend to do this by analysing runoff time-series from all stations available in the model domain, as follows: (i) run the model and compare simulated to all observed time-series, (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 hydrological network, locations and area of lakes/reservoirs, precipitation patterns, etc. It is important to remove as many errors as possible in the input data before starting to tune parameters, otherwise the calibration may lead to erroneous assumptions on hydrological processes to compensate for input data errors.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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)

When the technical model system is in place and input data seem to be relevant, the modeller can start tuning the parameters, so that the model structure represent the modeller's perception of how the hydrological system is organized and how the various processes are interconnected. In this step the modeller's competence and experience with the chosen model becomes important, as well as other data sources than traditional river discharge at gauges. For a complex process-based model it can be worthwhile to use a stepwise approach to separate out processes and choose a subset of gauges representing specific processes, so that for instance recession parameters are not hidden by lake parameters during calibration. For the model set-up to be right for the right reason we recommend to: (i) use a model we understand and can change, (ii) constrain relevant parameters to alternative data than just time-series of river discharge (e.g. snowmelt parameters to snow depths, evapotranspiration parameters to flux tower and satellite data) or select a subset of gauges representing different flow generating processes so that internal modelled variables/fluxes also represent our understanding of the hydrological system, (iii) apply a calibration procedure constrained to expert knowledge ensuring that the model structure reflects our understanding of flow paths and their interconnections, (iv) improve model structure by adding or changing the model algorithms if tuning of parameters is not enough to reflect the perception of the hydrological system, (v) include specific rating curves of lakes and reservoirs wherever available, tune routines for irrigation and dam regulation to fit with dynamics in downstream gauges, and (vi) assimilate observed data if possible, e.g. snow, upstream discharge, or regulation rules in reservoirs. The large scale system should be coherent with local processes in the multi-basin approach; we therefore recommend

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

combining bottom-up and top-down analysis to improve process understanding and make the model more reliable.

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 realistic results both at the large scale and for single catchments in the multi-basin approach. The model results should be analysed and understood in relation to catchment characteristics, human alterations and forcing data. *For spatial interpretation*, we recommend plotting maps with multi-basin outputs for several variables and signatures across the model domain, to check their coherence with various landscape features, such as spatial patterns of vegetation, geology, climate, precipitation, population density, and human alterations. This is to make sure that the model reflects our understanding. In addition, performance criteria for flow predictions should be plotted for each gauging site. This will help the modeller to understand the drivers and find logical reasons behind the hydrological heterogeneity, but also to identify knowledge gaps or model limitations. *For temporal interpretation*, we recommend to plot time-series for some basins in each group of similar landscape units and catchments, to check how several model variables interact to get an understanding of the process interactions in the model. 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)

HESSD
12, 2885–2944, 2015

Multi-basin models covering a large area with high spatial resolution are normally more computationally demanding than single basin models. It is therefore not always feasible to explicitly address all uncertainties from all sources or perform proper cross-validation. Andersson et al. (2014) showed that uncertainties in the input data and/or routing are often more important than parameter uncertainty when modelling at the large scale. Nevertheless, parameter uncertainties may be examined for specific process descriptions (see Sect. 2.4) and for various flow signatures (Donnelly et al., 2015). To examine uncertainties we recommend to: (i) use several performance criteria, (ii) check their spatial distribution across the model domain, and (iii) evaluate model performance for independent gauging sites and new datasets. The major deviations found between modelled and observed data in time and space should be the focus for the next round in the circle of steps for better predictions. It is then important to start reading the landscape and literature again and find new hypotheses of hydrological functioning and data sources to improve the model performance. We recommend to document and version-manage each model set-up before going into step 1 again, to make sure that knowledge is perceived and to make the set-up process more transparent. It is important to get a new baseline for the next round of improvements.

3 Data and methods

The further developed and modified recommendations (Sect. 2), which are based on the best practice recommendation for PUB (Takeuchi et al., 2013), were tested for predictions of ungauged basins across the Indian subcontinent. A process-based model was set up according to the six steps above for runoff predictions in some 6000 subbasins, where gauged time-series were only available at some 40 sites. Most catchments can thus be considered as ungauged in this work. Examples were extracted from each of the six recommended steps, to illustrate how we applied the recommended best practices and how they affected the quality of the predictions. The

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

geographical domain and methods used for modelling, regionalisation and evaluation during the exercise are described more in detail below.

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 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). 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 yr^{-1}), to arid in Rajasthan (20 days yr^{-1}). Moreover, India is characterised by strong temperature variations in different seasons ranging from mean temperatures of about 10 °C in winter to about 32 °C in pre-monsoon season.

The monsoon season is very important for water resources (and in addition to others, power generation, agriculture, economics and ecosystems) since 75 % of the annual rainfall (around 877 out of 1182 mm) is received in this period (Mall et al., 2006). In particular, India's mean monthly rainfall during July (286.5 mm) is highest and contributes about 24.2 % of annual rainfall.

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 for Asia including the Himalayas. 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 agen-

HESSD
12, 2885–2944, 2015

cies but are not released to the public domain; this generally sets a constraining factor for a model set-up. Consequently, in this application, flow information is available only for a small fraction of the subcontinent, which makes the region a great example for the application of PUB's best practices. Statistics of the basin areas and runoff for the entire set of gauged stations are presented in Table 2. Monthly potential evapotranspiration (PET) 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 6000 points for calculations of runoff in the river network (i.e. 6000 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, semi-distributed and process-based rainfall–runoff model, which describes the hydrological processes at the catchment scale (Lindström et al., 2010). The model represents processes for snow/ice accumulation and melting, evapotranspiration, soil moisture, discharge generation, groundwater recharge, aquifer discharge, some human alterations, 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.

HESSD
12, 2885–2944, 2015

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 ET conditions. 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. Precipitation falls directly on glacier and lake 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. 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 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. The demands are constrained by the water availability at these sources. After subtraction of conveyance losses, the withdrawn water is applied as additional infiltration to the irrigated soils from which the demands originated. 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 (see Table 1).

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

3.3 Model calibration and regionalisation

In total, 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 HRUs, which are linked to soil type and depths (soil dependent parameters) and vegetation (land use dependent parameters), while others are assumed to be general to the entire domain (general parameters) or specific to a defined region (regional parameters). Parameters are calibrated for representative gauged basins for each HRU and then transferred to similar HRUs across the whole domain. Using the 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 available. The following procedure was used for India-HYPE v.1.0.

3.3.1 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). 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, after which another group of parameters (e.g. river routing) is calibrated. As the model concept follows the flow paths, the headwaters are calibrated first, then streams, lakes, rivers and finally the overall outlet to the sea is checked. In the step-wise procedure, each step downstream in the model code in-

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

cludes some reconsideration about chosen parameter values in an iterative procedure. The lake/reservoir and irrigation parameters may be difficult to identify, particularly at ungauged areas, and usually require manual calibration driven by known regulation routines or observed time series at the outlet of the lakes/reservoirs, and soft information about water extraction at the regional scale.

For the Indian subcontinent, the following groups of HYPE parameters were calibrated step-wise: (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.

3.3.2 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:

$$\mathrm{mactrsm}_{\mathrm{Coarse}} > \mathrm{mactrsm}_{\mathrm{Medium}} > \mathrm{mactrsm}_{\mathrm{Fine}}$$

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

HESSD
12, 2885–2944, 2015

coarse soil; although the value for shallow soil is constrained to be less than mactrsm for organic soils.

3.3.3 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.

3.3.4 Spatiotemporal calibration approach

The calibration objective in the large-scale applications is not to provide an optimal model for a specific catchment, but rather to identify a robust model that performs well for multiple basins. In relation to the PUB concept, we assume that identified parameters and their regionalisation to ungauged regions are acceptable if the model performs adequately in the gauged basins of the domain. This assumption has been tested in other regions with similar performance also for independent gauges, thus representing ungauged conditions (Arheimer and Lindström, 2013). 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 stations for evaluation. 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 independent 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

HESSD
12, 2885–2944, 2015

sensitivity. DE-MC was applied at each step of the iterative calibration procedure 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:

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

where r is the linear cross-correlation coefficient between observed and modelled records, α is a measure of variability in the data values (equal to the SD of modelled over the SD of observed), and β 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, α and β are also 1. KGE allows a multi-objective perspective if the three components are treated separately, by focusing on the correlation (timing) error, variability error, and bias (volume) error as separate aspects to be minimised. We further investigate the relative influence of timing, variability and volume error on the KGE value; hence have them as diagnostic metrics. Therefore, we firstly transform the three KGE components to results into a consistent range of possible values. Consequently we consider:

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

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

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

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

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

3.4 Evaluation beyond standard performance metrics

3.4.1 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 (Q_m, mm yr^{-1}), the normalised high flow statistic (q05, –) and the slope of the flow duration curve (mFDC, –).

3.4.2 Multi-variable evaluation

To judge model credibility, other observed variables than river discharge are used. For India-HYPE, these included evaluations against "observed" snow areal extent and snow water equivalent from the GlobSnow system and potential evapotranspiration (PET) from the MODIS system. The assumption is that MODIS PET can be used as reference to calibrate the HYPE parameters that control PET; this refers only to the cevp land-use dependent parameter, which is a coefficient of potential evapotranspiration (mm d^{-1} °C^{-1}) (Lindström et al., 2010). To test this assumption and avoid introducing misinformation in our analysis, the MODIS PET estimates were integrated into long-term values, i.e. annual, and over the subbasin resolution. Consequently, the objective is to optimise the cevp parameter for each land use type so that HYPE modelled annual PET matches the MODIS annual PET 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) are used as objective functions in this analysis; 0 values indicate a perfect model with no errors for both criteria. Note that the analysis is conducted in the 2000–

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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.

3.4.3 Linking performance to physiographic-climatic characteristics

To better understand the controls of a poor (or good) model performance and point towards aspects of the model that need improvement, 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). CART also provides information on the probabilities of different output groups at each leaf node. In this case, we divide the KGE performance into three groups – bad (KGE < 0.4), medium (0.4 < KGE < 0.7), and good (KGE > 0.7), which are 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.

4 Results

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; see boxplots of "1st run" in Fig. 9. This was expected and clearly indicated major focus areas for improvements, such as adjustment in routing, regulation for reservoirs, precipitation correction based on eleva-

HESSD
12, 2885–2944, 2015

tion, and parameter tuning. This was valuable information for further model refinement, which was done according to the interpreted and further developed "best practices for predictions in ungauged basins" (Sect. 2). Examples of results and major reflections on the procedure and methods used are given below for each of the steps.

4.1 Read the landscape

Background knowledge was firstly acquired via analysis of available maps (i.e. Google Earth, digitised and hard copies from credible websites and governmental reports) that describe the spatial patterns of land use, soil and climate. Depending on the form of available data, the analysis was visual (i.e. identifying the similarity in spatial patterns) and/or numerical (i.e. calculating metrics of agreement between the datasets). Moreover, study of the scientific literature on regional hydrological investigations enabled identification of dominant physical processes and flowpaths. Such soft information was useful for turning on/off processes and selecting algorithms in the HYPE code, 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, whereas 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).

In addition, analysis of the topographic data has been very important since they affect the subbasin delineation and routing. Although Hydrosheds are based on elevation layers, which are hydrologically conditioned and corrected, their spatial resolution can affect basin discretisation. Here, merging Hydrosheds with GRDC (hence forcing WHIST to generate subbasins where GRDC stations are available) involved some mismatches in size of upstream areas between the subbasin delineations and the metadata of catchments for some stations. 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

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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). Hence, the recommendation to carefully check the subbasin delineation is of major importance for multi-basin model performance.

In addition, the delineated basins generated by WHIST were evaluated using a shapefile of basin areas reported by Gosain et al. (2011); these are presented in Fig. 2 (in red). 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. 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. 5) in long-term averages of discharge (Q_m) and high flow (q05). 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. However, the agreement seems to be generally reduced in the validation period when the values of the two observed signatures increase, indicating that the model has lower skill to capture extremes. The slope of the flow duration curve (mFDC) was more difficult to represent, especially for "smaller" basins of in the model ($< 32\,770\,\mathrm{km}^2$), which may also reflect the difficulties to capture the extremes; this is normally related to size of river basin as influence from many processes evens out in a large catchment while smaller catchments show a more peaky behaviour. It is finally important to note that these results are consistent for gauging stations used both in calibration and independent validation (blue and red circles in

Fig. 4), which highlights the potential of the regionalisation methods used in this multi-basin modelling across India.

We explored how flow signatures can be significantly affected by human impacts by analysing modelled responses considering and omitting the human influence. Figure 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. Results presented are after the calibration of the general parameters (step 1 in Fig. 9), however further improvements were achieved by the end of the calibration procedure. To sum up, it was very useful to analyse flow signatures in both space and time for understanding processes and model skills across the domain.

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 aim to identify basin similarities to drive the identification of the model's regional parameters. The cluster analysis was applied to all 6010 subbasins of the domain within the 17 dimensional space (see Sect. 3.3). The analysis 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 and the spatial pattern of the clusters was rather similar to Fig. 6. Furthermore, the pattern of clusters follows well the spatial pattern of soil types highlighting the latter's importance in hydrological clustering

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

analysis. 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 (Sect. 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.

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

HESSD
12, 2885–2944, 2015

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 based on: (i) additional data sources for parameter identification, using MODIS PET data, (ii) expert knowledge on parameter inter-relations; and (iii) a stepwise calibration procedure. 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.

4.4.1 Additional data sources

The calibration of PET model routine against the MODIS PET 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. The range of the top 100 cevp values for each land use type and objective function is presented in Fig. 7. Results show that cevp sensitivity to different land use types depends on the objective function (RMSE and Bias); different objective functions extract different information from the PET 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 PET. The other types cover a very small proportion of the entire area and therefore do not significantly contribute to the annual PET. cevp values for these types were slightly modified at a later stage (introducing expert knowledge on expected cevp values based on previous model set-ups). Overall India-HYPE underestimated PET at the arid regions and over the Himalayas (on average by 15 %), whereas the model overestimated PET along the western and

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

eastern coast lines (on average by 12 %). This highlights limitations of the simple PET algorithm used in this model set-up. Although the model is not fully capable of matching MODIS PET over the entire domain, the use of additional information to constrain parameters (hence constraining the model's results for specific processes) is promising. To draw more robust conclusions on the identification of the appropriate cevp parameter value, we investigate further the parameter inter-dependence between land use types (not shown). Trends in the HYPE model – MODIS PET agreement could be identified for some of the land use types. A weak but statistically significant relationship is observed between crops and forest for both objective functions; the correlation coefficient is −0.34 and −0.87 for RMSE and Bias respectively. No other relationship is identified, probably due to the less important contribution of the other land use types in the region's PET.

4.4.2 Expert knowledge

Expert knowledge was applied to filter out unrealistic relationships of the mactrsm parameter for different soil types (see Sect. 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.

HESSD
12, 2885–2944, 2015

4.4.3 Stepwise calibration procedure

The stepwise calibration follows the way water is moving through the landscape, using a selection of representative gauged basins across the whole geographical domain for groups of model parameters simultaneously. For India-HYPE, three calibration steps were considered (see Sect. 3.3) while the prior used HYPE parameter values for Sweden. Figure 9 shows the relative importance of each parameter group on model's predictive power and consistency. The predictability of the model with prior parameter values is very poor 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, $m^3 s^{-1}$). These are checked visually in a test bench of 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 dur-

HESSD
12, 2885–2944, 2015

ing 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. Figure 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 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 (KGE equal to 0.44) when the model is evaluated for gauges, which are independent both in space

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

and time. This shows that the model still needs improvements to be equally reliable for predictions in ungauged basins. The diagnostic 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 (Gupta et al., 2009). 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. Based on this analysis, we closed the adjustments of the first model version and documented the India-HYPE version 1.0. From the uncertainty analysis we got some guidelines on how to start working on the next version in step 1, namely to investigate alternative evapotranspiration algorithms (e.g. Jensen–Haise, Priestly–Taylor), and refine the snow accumulation/melting component. Results from the model and its application in climate change impact studies can be inspected and downloaded from http://hypeweb.smhi.se.

5 Discussion

5.1 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 hydrology and hence improving process understanding. Here, we presented results from all 42 Indian GRDC stations including both failure and success stories. 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 flow. 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 in India have generally been achieving high performance (Arora, 2010; Patil et al., 2008), mainly due to the local forcing data used; usually the data are governmental and confidential with high spatiotemporal resolution and less uncertainty/error. In addition, model parameters are required to be transferable along a smooth hydro-climatic gradient in contrast to the usually strong gradient

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

of large systems. 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 PET 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 PET model component will be further investigated and refined in the next version of India-HYPE.

5.2 Pros and cons in the methods used

The set of tools and numerical/graphical methods were selected based on our experience on multi-basin large-scale modelling and large data analyses. Firstly, the WHIST tool was shown to be very useful for the transformation of large datasets into model input files and also useful for the delineation and linking of subbasins. However, as in every tool, the various functions are based on numerical algorithms which under applications with low resolution data, could result in artefacts, i.e. incorrect linking between subbasins in very flat regions. Consequently quality control is still recommended. With the generated input files, we run the model to test potential failures in the mod-

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

elling chain and eventually set the baseline performance prior to further refinements in model parameters and structure. We found this to be a good way to familiarise ourselves with the modelled responses across the region and correct obvious technical errors. Analysis of flow signatures allowed a direct evaluation of long-term and seasonal patterns to judge the model's predictability across different environmental conditions as well as spatial and temporal scales. Here, we only presented three flow signatures; however others are additionally recommended for comparative hydrology (see Viglione et al., 2013). We also found that the k-means clustering provided important information for the regionalisation of model parameters (here the regional parameters). However, the potential of the catchment similarity concept should be further investigated on the regionalisation of more model parameters, i.e. general parameters. It would also be interesting to assess the sensitivity of clustering to catchment characteristics by introducing indices that describe variability in the dependent variables (i.e. relief ratio, precipitation seasonality index); some preliminary analysis has already been conducted as explained in Sect. 4.3.

In our approach, the calibration is focused on process understanding and aimed to ensure "right for the right reasons" model results. The stepwise semi-automatic calibration, expert knowledge and/or information from multiple variables, other than river discharge, were useful to improve model consistency within limited ranges of uncertainty. The use of remote sensing data to identify the parameters should be explored further particularly in data sparse regions. Conventional modelling approaches driven by ground-based observation could be complemented (without obviating the need for ground observations) by allowing assimilation of satellite data which capture spatial variability better than ground observations. We believe that model evaluation using internal model variables, in combination with river discharge, ensures model realism. We also found the analysis of time series from different flowpaths and classification trees to be useful diagnostic tools, which can point towards falsification of models. However, here the number of stations was not statistically significant to fully explore the potential of classification trees analysis. Finally, we highlighted the potential of the KGE metric by

HESSD
12, 2885–2944, 2015

decomposing its terms and identifying the flow characteristics that result in a good/bad KGE values; note that a similar analyses could be conducted with other diagnostic metrics. In multi-basin modelling we still miss a statistical metric that can consider both the spatial and temporal performance for a given model domain. Currently, we use mean, median and percentiles of e.g. NSE and KGE (e.g. Arheimer et al., 2012; Strömqvist et al., 2012) but it would be useful to have a metric where both dimensions can be compiled in one value. This will ease the comparison of one large-scale model set-up using many gauges with another, or allow an easy follow-up of the progress in model performance as the model is refined and improved in the six step approach (Fig. 1) for a given model domain.

5.3 PUB recommendations: catchment-based modelling vs. large-scale modelling

The presented examples support the overall relevance of PUB's best practices for multi-basin modelling also at the large scale. We think that the modification we suggested speeds up the development of robust model set-ups and also represents human impacts and water resources management. The quick development of a first model set-up allows generation of preliminary results, probably of an inadequate model performance, which was found to be very useful to detect errors in input data. However, uncertainties still remained and post-processing analysis pointed both towards input data errors and/or model limitations (e.g. flow signatures and CART). Although, we also highlighted the importance of quality checks, the use and pre-/post-processing analysis of large datasets (either raw or repurposed) is not always straightforward with their quality also affected by various technical and/or numerical obstacles. Finally, we stress the need to represent human impacts, i.e. irrigation, lake rating curve, regulation of reservoirs, and ensure realism on multi-basin models. However, available information on human influence is generally limited whilst model structures are often insufficient to reproduce all aspects of regulated flow response (Nazemi and Wheater, 2015a). Nevertheless, opportunities to improve the representation of water resources management in multi-

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD
12, 2885–2944, 2015
The India-HYPE case
I. G. Pechlivanidis and B. Arheimer

basin modelling at the large scale are discussed in Nazemi and Wheater (2014b) and will be tested in the next version of India-HYPE.

6 Conclusions

In this study we show that advances in PUB (Blöschl et al., 2013) are relevant also for multi-basin modelling at the large scale. Thus, the best practices for predictions in ungauged basins (Takeuchi et al., 2013) should be followed independently of the scale. However, we suggest a slightly modified interpretation of each step and stress the need to set up the model system before starting to work on analysing the landscape and/or evaluating the input data. We argue that many technical problems and data inconsistencies become apparent when running the model and therefore it should be done early in the model set-up process. In general, input data may be more erroneous at the large scale and more efforts on quality checks are therefore needed using global datasets. We also stress the need to include human alterations, which are crucial at the large scale.

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

- Each step in the 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. Some useful methods were for instance: the stepwise semi-automatic calibration approach, diagnostic metrics to improve the model's predictability and consistency, evaluation methods based on many data sources, combination of traditional performance metrics with analysis of flow signatures, multiple variables and the CART.

- Parameter constraints based on either expert knowledge or remote sensing data (MODIS PET) did not improve the skill to reproduce the overall system response; however they set a system understanding-based strategy of filtering out unsuit-

HESSD
12, 2885–2944, 2015
The India-HYPE case
I. G. Pechlivanidis and B. Arheimer

able models and hence ensuring the right model for the right reasons. Identification of regions with similar physiographic-climatic characteristics allowed calibration of the regional model parameters and improvement of the model's spatial consistency.

- 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. 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.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Appendix A: Definition of flow signatures

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 three signatures are calculated:

Q_m: the arithmetic mean annual specific runoff (mm yr^{-1}):

$$Q_m = 365 \cdot \overline{Q_d} = \frac{365}{T} \sum_{t=1}^{T} Q_d(t)$$

where $\overline{Q_d}$ is the mean daily specific runoff (mm d^{-1}) and T (days) is the record length (corresponding to 9 years in this study).

q05: the normalized high flow statistic (–)

$$\text{q05} = \frac{Q_{5\,\%}}{\overline{Q_d}}$$

where $Q_{5\,\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 5 % of the time.

mFDC: the slope of the flow duration curve (–)

$$\text{mFDC} = 100 \cdot \frac{Q_{30\,\%} - Q_{70\,\%}}{40 \cdot \overline{Q_d}}$$

where $Q_{30\,\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 30 % of the time, $Q_{70\,\%}$ 70 % of the time. mFDC is a measure of slope of the central part of the flow duration curve and indicates the percentage of increase of runoff, with respect to the annual mean, for 1 % decrease of exceedance probability (Viglione et al., 2013).

HESSD
12, 2885–2944, 2015
The India-HYPE case
I. G. Pechlivanidis and B. Arheimer

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 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 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.

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, Hydrolog. Sci. J., 48, 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, FAO Irrigation and drainage paper 56, Rome, Italy, 1998.

Andersson, J. C. M., Pechlivanidis, I. G., Gustafsson, D., Donnelly, C., and Arheimer, B.: Key factors for improving large-scale hydrological model performance, Eur. Water, in review, 2014.

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 Proceedings and Reports No. 307, Wallingford, UK, 2006.

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: Blöschl, G., Sivapalan, M., Wagener, T., and Viglione, A., Cambridge University Press, Cambridge, UK, 353–359, 2013.

HESSD
12, 2885–2944, 2015

Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., and Strömqvist, J.: Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea Basin – influence of input-data quality and scale, Hydrol. Res., 43, 315–329, doi:10.2166/nh.2012.010, 2012.
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., 2, 636–643, doi:10.4236/jwarp.2010.27073, 2010.
Attri, S. D. and Tyagi, A.: Climate Profile of India, Gov. India Minist. Earth Sci., 129, New Delhi, India, 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, Italy, 2002.
Bernhard, L., Liermann, C., 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, Germany, 2011.
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, USA, 1984.
Bulygina, N., McIntyre, N., and Wheater, H.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, Hydrol. Earth Syst. Sci., 13, 893–904, doi:10.5194/hess-13-893-2009, 2009.
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, W06519, doi:10.1029/2011WR011207, 2012.
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.

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

Donnelly, C., Rosberg, J., and Isberg, K.: A validation of river routing networks for catchment modelling from small to large scales, Hydrol. Res., 44, 917–925, doi:10.2166/nh.2012.341, 2013.
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, Hydrolog. Sci. J., accepted, 2015.
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, 1–13, 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, 1895–1915, doi:10.5194/hess-18-1895-2014, 2014.
Good, P. I.: Resampling Methods: a Practical Guide to Data Analysis, 3rd edn., Birkhäuser, Boston, 2005.
Gosain, A., Rao, S., and Basuray, D.: Climate change impact assessment on hydrology of Indian river basins, Curr. Sci. India, 90, 346–353, available at: http://www.iisc.ernet.in/currsci/feb102006/346.pdf (last access: 28 January 2014), 2006.
Gosain, A., Rao, S., and Arora, A.: Climate change impact assessment of water resources of India, Curr. Sci. India, 101, 356–371, 2011.
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, 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, 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, 463–477, doi:10.5194/hess-18-463-2014, 2014.

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

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, Hydrolog. Sci. J., 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

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.

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, 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, 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, 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, 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, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos T. Am. Geophys. Un., 89, 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, 295–319, doi:10.2166/nh.2010.007, 2010.

Mall, R. K., Singh, R., Gupta, A., Srinivasan, G., and Rathore, L. S.: Impact of climate change on Indian agriculture: a review, Climatic Change, 78, 445–478, doi:10.1007/s10584-005-9042-x, 2006.

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, 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, 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, Hydrolog. Sci. J., 58, 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, 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.

Nazemi, A. and Wheater, H. S.: On inclusion of water resource management in Earth system models – Part 1: Problem definition and representation of water demand, Hydrol. Earth Syst. Sci., 19, 33–61, doi:10.5194/hess-19-33-2015, 2015a.

Nazemi, A. and Wheater, H. S.: On inclusion of water resource management in Earth system models – Part 2: Representation of water supply and allocation and opportunities for improved modeling, Hydrol. Earth Syst. Sci., 19, 63–90, doi:10.5194/hess-19-63-2015, 2015b.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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, 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, 137–146, doi:10.1016/j.biosystemseng.2008.02.001, 2008.

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

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, 1874–1889, doi:10.1002/hyp.9731, 2013.

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, 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., 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/hess-9-535-2005, 2005.

Singh, P. and Bengtsson, L.: Hydrological sensitivity of a large Himalayan basin to climate change, Hydrol. Process., 18, 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, Hydrolog. Sci. J., 48, 257–276, doi:10.1623/hysj.48.2.257.44693, 2003.

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

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, Hydrolog. Sci. J., 48, 857–880, 2003.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

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, Hydrolog. Sci. J., 57, 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: Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. H. G., Cambridge University Press, Cambridge, UK, 384–387, 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, 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, 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, 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, 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, B. Am. Meteorol. Soc., 93, 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	6010 (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)
Irrigation	Global Map of Irrigation Areas (GMIA)	Siebert et al. (2005)
Discharge	Global Runoff Data Centre (GRDC; 42 stations)	http://www.bafg.de/GRDC
Precipitation	APHRODITE ($0.25^{\circ} \times 0.25^{\circ}$)	Yatagai et al. (2012)
Temperature	AphroTEMP ($0.5^{\circ} \times 0.5^{\circ}$)	Yasutomi et al. (2011)
Potential evapotransp.	MODIS PET (1 km)	Mu et al. (2011)

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

Table 2. Statistics over the entire basin set.

	Percentiles					
	5 %	25 %	Median	75 %	95 %	Mean
Basin surface (km^2)	2062	12 691	32 770	68 522	294 524	75 493
Mean annual runoff (Q_m, 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.

HESSD
12, 2885–2944, 2015

Table 3. Median model performance for calibration and evaluation 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

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

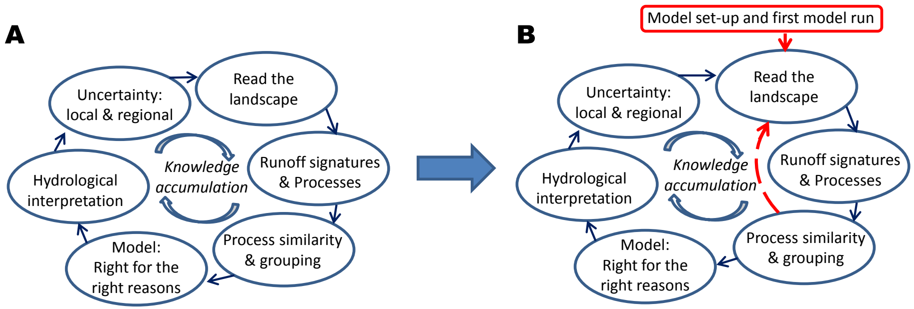

Figure 1. Best practices for predictions in ungauged basins: **(a)** according to Fig. 13.1 by Takeuchi et al. (2013) in Blöschl et al. (2013), and **(b)** modified version for multi-basin applications at the large scale.

HESSD
12, 2885–2944, 2015

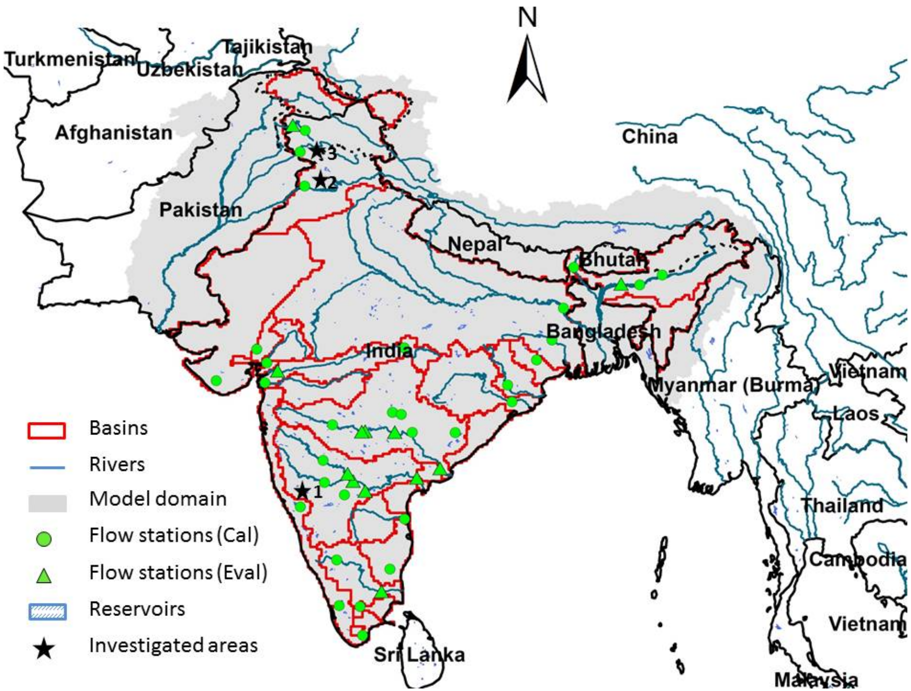

Figure 2. Map of the Indian subcontinent (model domain). Specific investigations are conducted at river-basins with a star in the order of their numbering.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and
B. Arheimer

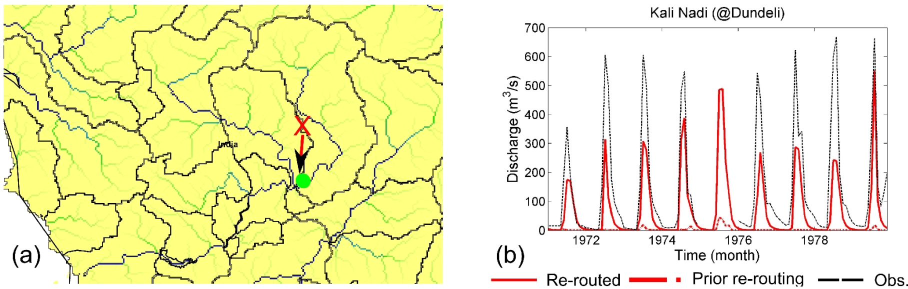

Figure 3. Example of the impact of catchment delineation and routing on model behaviour: **(a)** correction in the location (red "x" and green circle is prior and after the correction respectively) of the Dundeli discharge station (Kali Nadi River Basin), and **(b)** the corresponding modelled discharge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

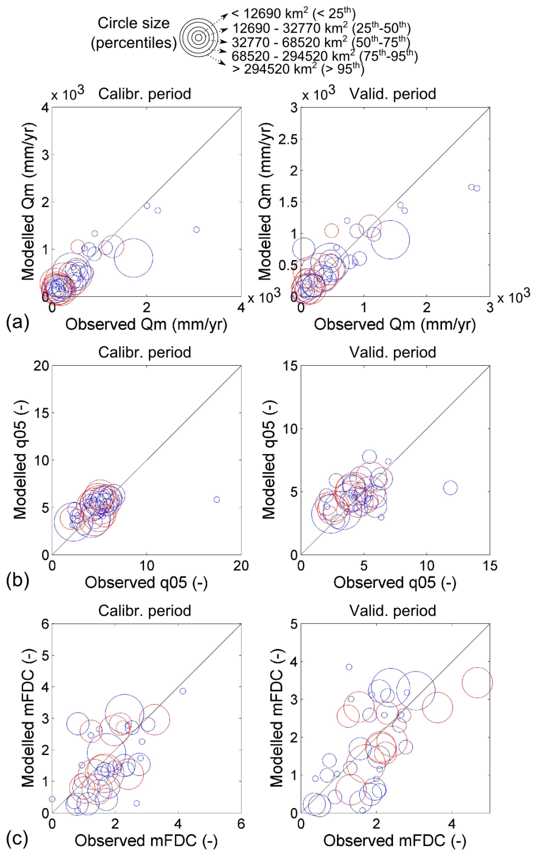

Figure 4. Signature analysis in the spatiotemporal model evaluation: **(a)** the mean annual specific runoff, **(b)** the normalised high flow statistic, and **(c)** the slope of the flow duration curve. Blue and red circles are used for the calibration and evaluation stations respectively.

HESSD

12, 2885–2944, 2015

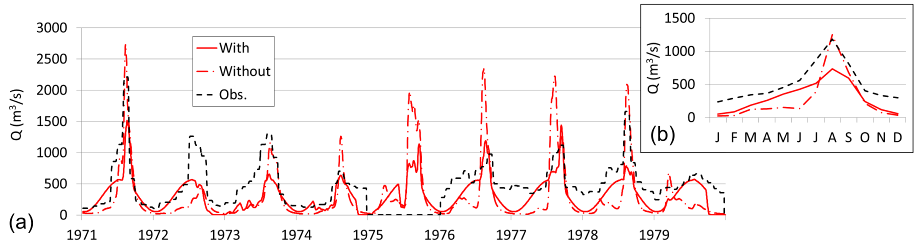

Figure 5. Impact of model parameterisation of reservoir regulation on discharge for **(a)** monthly streamflow, and **(b)** annual hydrograph, showing naturalised (without) and regulated (with) conditions at the basin outlet (located at asterisk 2 in Fig. 2).

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

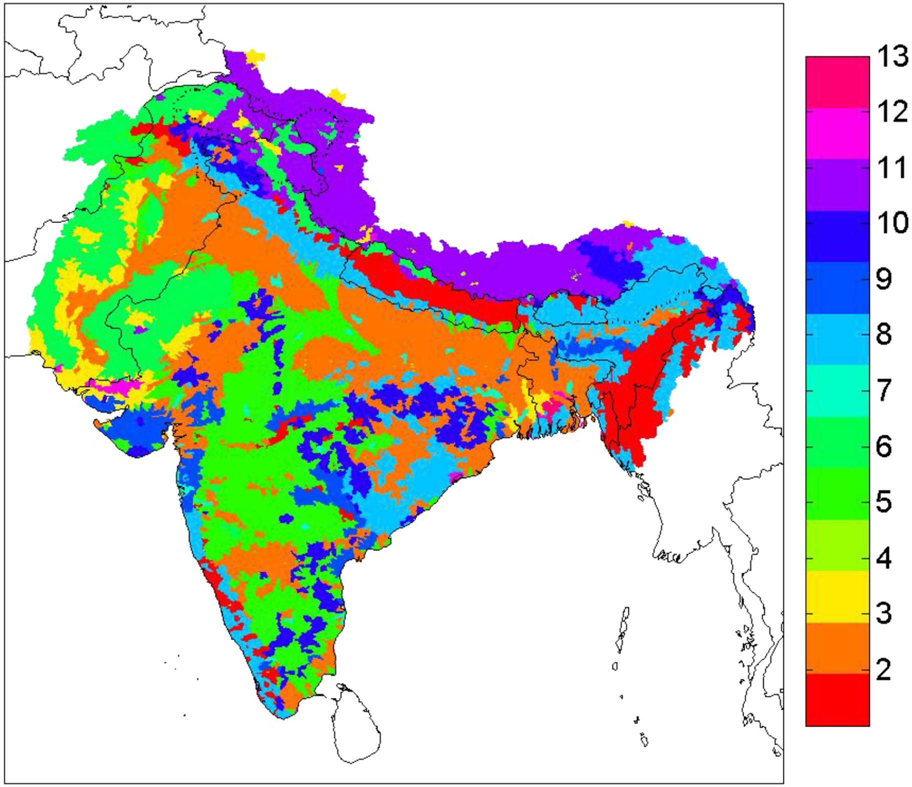

Figure 6. Subbasin clusters using a *k*-means clustering approach based on physiographic-climatic characteristics.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and
B. Arheimer

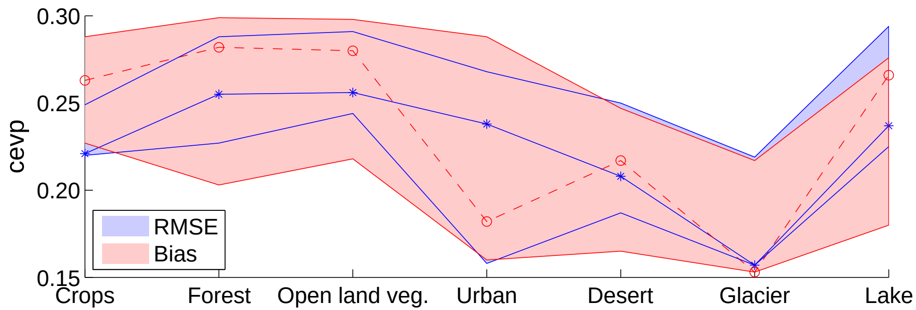

Figure 7. Coefficient of potential evapotranspiration (cevp) parameter as identified (behavioural range and optimum set) for different objective functions (RMSE and Bias) and land use type. Lines with markers present the optimum parameter values for different objective functions.

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and B. Arheimer

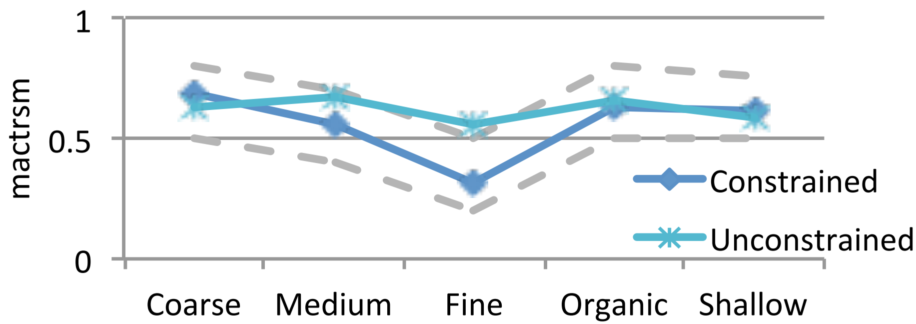

Figure 8. Constraints (grey dashed lines) and optimum (solid lines) values of the mactrsm soil dependent model parameter based on process understanding.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

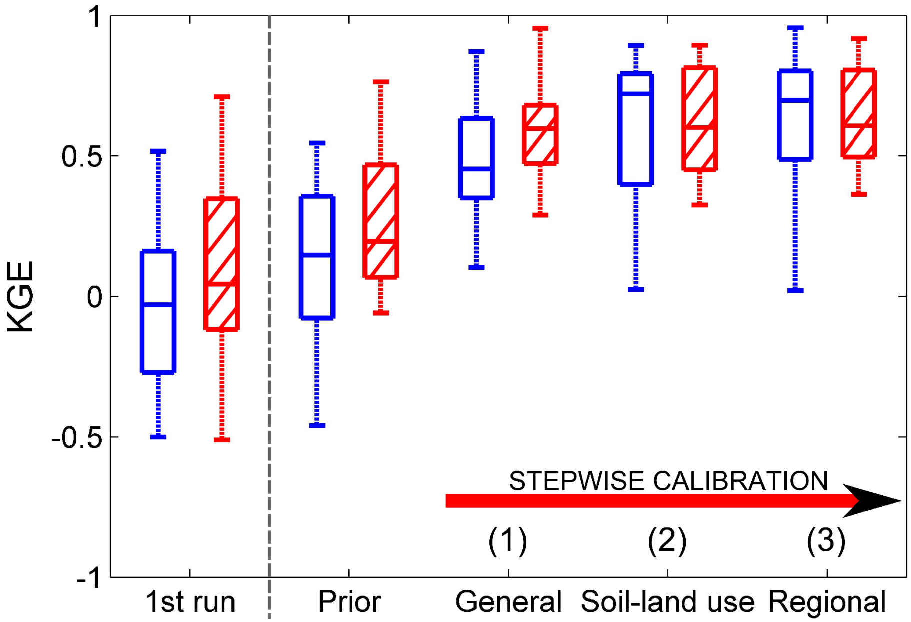

Figure 9. Improvements in KGE model performance during the stepwise calibration approach (steps 1–3 correspond to general, soil-land use, and regional calibration as described in Sect. 3.3). "1st run" corresponds to model performance of the very first model set-up to establish a technical model infrastructure. "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.

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and B. Arheimer

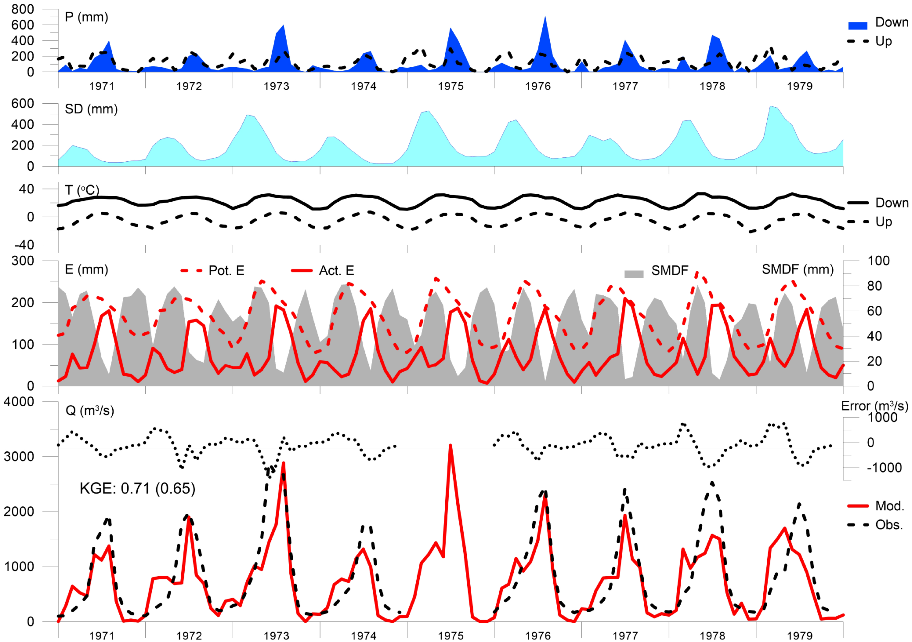

Figure 10. Analysis of model variables: P, SD, T, E, SMDF and Q. E corresponds to potential (Pot.) and actual (Act.) evapotranspiration, and Q corresponds to modelled (Mod.) and observed (Obs.) discharge. Note that P and T series are plotted at the outlet of the subbasin (Down) and the most upstream subbasin (Up).

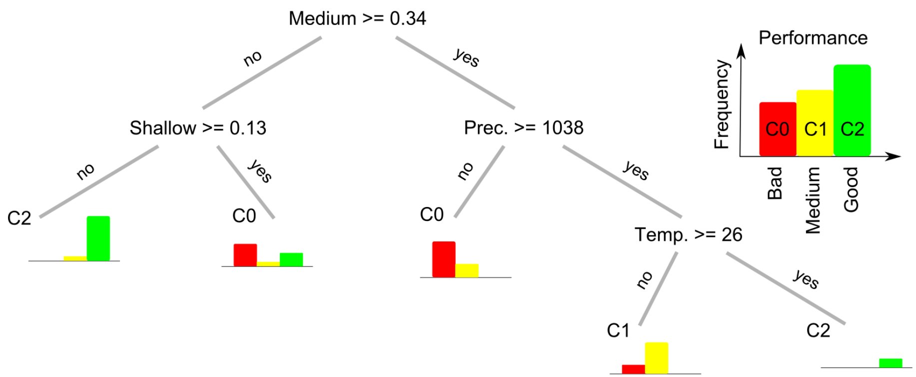

Figure 11. Classification trees relating regions of different KGE performance with physical and climatic characteristics. The bars represent the probability of a performance resulting in any of the three performance classes (C0, C1 or C2).

HESSD
12, 2885–2944, 2015

I. G. Pechlivanidis and
B. Arheimer

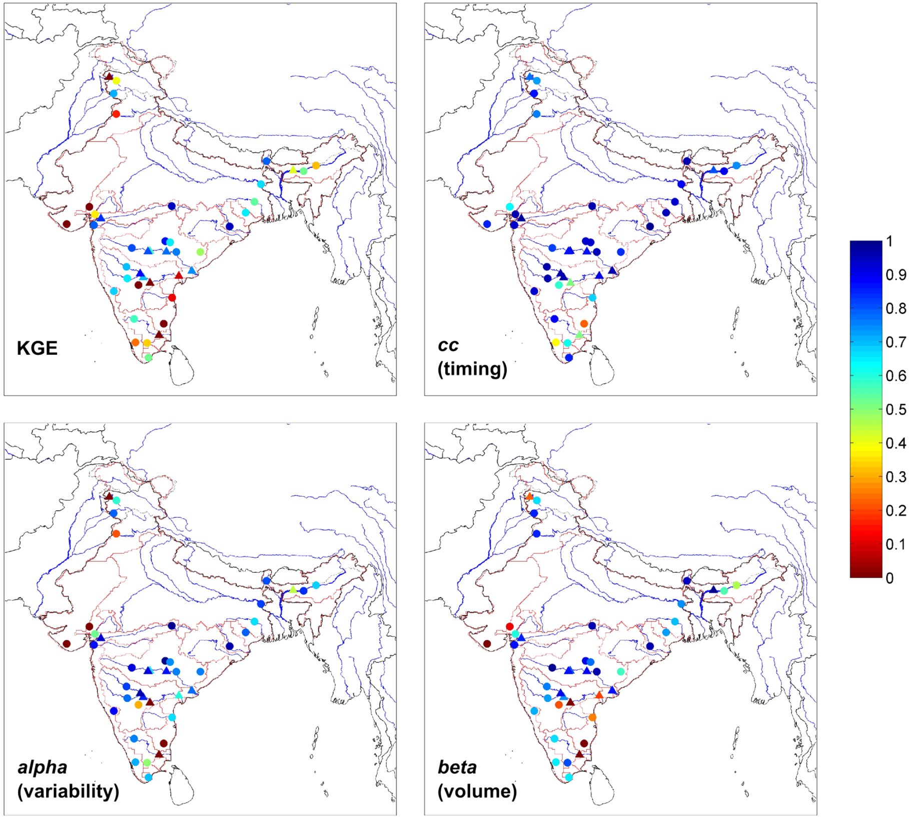

Figure 12. Spatial variability of KGE (and its decomposed terms) model performance for the calibration (circle) and evaluation (triangle) stations.

HESSD
12, 2885–2944, 2015

The India-HYPE case

I. G. Pechlivanidis and
B. Arheimer