hess-2015-53 Author's Response to Referees

"Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case" by I.G. Pechlivanidis and B. Arheimer

Dear Dr. Ross Woods (Editor of the HESS journal),

We would first like to thank you for the attention you paid to our paper. In addition, we would like to gratefully acknowledge the two referees for their constructive comments. We are happy to submit a substantially improved manuscript! In the following, we present their comments in italics with our responses distinguished by red colour.

Table of Contents

Response to Referees
Referee #1

This paper represents the first formal application of the general recommendations from the final chapter of the PUB Book (Takeuchi et al., 2013; Bloeschl et al., 2013), and that too so soon after the book was published and even more importantly in the Indian subcontinent. More than anything it demonstrates that these recommendations and the PUB Book itself are indeed practically relevant, and the effort that went into them has been worthwhile.

In order to accommodate application to over 6000 catchments around India, they made one modification or compromise: they chose to apply one common model to all the catchments. From then on, they claim that the implementation of the recommendations helped to improve NSE (model performance) improved from 0.14 to 0.64.

This is impressive. Those are the positives, and I do want the paper to be eventually published in HESS. Having said this, I must also point to areas of potential improvement.

We would like to thank for these very positive judgments and we are happy to know that the first reviewer recommend eventual publication in HESS.

Highlighting Hydrological Insights Gained

First of all, in spite of the value and relevance of the Takeuchi et al. recommendations, the presentation, especially in relation to the implementation of these recommendations came across

as rather pedantic (or abstract). I would have gained more from the paper if these had been illustrated by substantive illustrations that have provided hydrologic insight.

Indian hydrology must exhibit enormous heterogeneity, and hence process complexity: dominant processes must vary substantially across the country, and when one applies the same model everywhere, model performance must vary also tremendously. The authors missed an opportunity to highlight many of these issues through their modeling (PUB) experience. Absence of this gives me an empty feeling, of not learning much hydrological from this extensive application of the model across the whole subcontinent. I have to believe that when one improves performance from 0.16 to 0.64, one must have learned a lot from that experience.

We realize now that the Section 2 "Modified PUB recommendations" came out far too detailed in the first version of the manuscript. We have therefore significantly reduced this section to make it less pedantic and more relevant. In the section 4 "Results", we show illustrations for each step in the "Modified PUB recommendations" (section 2), to highlight model improvements, dominant processes and spatial variability of results. However, we realize that also this section came out too long and detailed in the first version of the manuscript – it was very difficult for the reader to "see the forest for the trees". We have therefore shortened it considerably to more clearly show the highlights of hydrologic insights from the model set-up procedure.

Large Sample Application

The authors claimed that they modified the PUB recommendations to tackle simultaneous prediction in over 6000 catchments. However, their statement that they applied one common model in all catchments, I missed anything else they did to deal with multiple basins.

When you apply the same model to many catchments, it provides an opportunity to compare the functioning of many catchments across gradients of climate and land- scape properties, to pool catchments together in terms of similar behavior, make connections between catchment classes and regional climatic and landscape patterns etc. One also could have discovered that model is deficient in certain respects in different parts of the country, which calls for improvements in model conceptualization (not just model parameterization). Bringing these out would have been also interesting and insightful.

We fully agree that the first manuscript was mostly focusing on methods and modelling procedures not on overall insights in Indian hydrology. To also include this aspect we have now introduced two new sections with additional model results; the Methods used to extract the new results are described in the section 3.5 "Catchment functioning across gradients" and the new results in the section 4.7 "Spatial flow pattern across the subcontinent and dominant processes". We hope this illustrates the usefulness of multi-basin large-scale modeling and show the knowledge we gained in Indian hydrology from modeling 6000 catchments across the entire subcontinent. To show this, we included two new figures, which illustrate: 1) a map of runoff characteristics across India and a sensitivity analysis of how we improve our understanding of this pattern during the model set-up, using PUB methods (Fig. 13), and 2) the flow characteristics in the hydrological regions across the subcontinent, expressed as a number of flow signatures (Fig. 14).

To make the figures, we have extracted daily discharge series from every single subbasin and calculated 12 flow signatures for every single subbasin (the flow signatures are presented in Appendix A). We then applied a k-means clustering approach. The methodology is presented in section 3.5.

Recommendations

Overall, while I am really impressed with the effort that went into the paper, much more could have been brought out. The paper is already long enough and I can imagine the authors might resist adding more material of the sort I have called for. However, if the authors want the readers to take interest in their paper, they must make at least some effort to present some hydrological details and insights along the lines of my remarks. I will leave it to the discretion of the authors and the editor as to how much of this needs to/can be done in this paper.

To sum up, we have significantly shortened the text (especially section 2 and 4) to better emphasise the major highlights. The Discussion section is now merged with the Results section, while sections 5.2 and 5.3 from the first version of the manuscript are now completely removed to avoid repetition. Moreover, we have inserted new findings on similarities in runoff patterns, mostly from ungauged basins, across the entire sub-continent and an analysis of dominant drivers. We would like to thank Referee #1 for suggesting these improvements to our manuscript. We think this ended up in a much better paper, which hopefully should be interesting to a much broader audience.

Referee #2

1 General comments

This manuscript reports on the process of implementing a hydrological model for the Indian subcontinent and reflects upon this process in the context of the recent recommendations of PUB on hydrological modelling. It is well written and well presented.

But to be honest, I find this paper of a lightness that verges on the unbearable. It surely touches upon some of the fundamental issues of modelling in a complex and challenging area such as India. But it feels strongly as if the authors really wanted to convey their entire thought process in the smallest detail, and to interweave that with each relevant concept of current hydrological thinking that springs to mind. This makes the manuscript rather tedious and long, especially because it regularly gets trapped in generalities and even clichés. In the specific comments below, I have tried to highlight some sections that I think need particular attention but it is in no means exhaustive.

We realise that the original manuscript was too long and detailed, which exhausted the reader. We have therefore significantly condensed the paper and removed details in order to focus more on the overall message and highlight findings.

Our novelty is that we apply the catchment modeling approach to the continental scale, which has been requested in numerous opinion papers lately. We believe that such model setups at the large scale is currently lacking in the scientific literature. With our effort, we want to encourage the catchment modeling community to reach out for the larger scale, where land-surface schemes and global water-allocation models operate. We think that the catchment modelers have a lot to contribute with in this context, especially when it comes to uncertainty analysis and quality controls. This has now been emphasized in the Introduction section.

In addition, I understand that authors use the model implementation as an illustration to make their point, rather than as a scientific experiment that generates insights in the local hydrology (or the model behaviour). Nevertheless, in my opinion it is not a very powerful illustration, mainly because of:

- *- the lack of purpose of the model. Without a particular purpose, it is very hard to make a point about the quality of a model implementation. It is a bit of a cliché in itself, but Box's*

aphorism that all models are wrong but some may be useful, is relevant here: there is a convincing case here that the model is wrong, but because of lack of purpose the argument that the model is adequate is far less convincing. A more concrete hypothesis to test, or use case, would help a lot.

We fully agree with this statement and therefore we have now included the use case of mapping spatial patterns of runoff characteristics across the Indian subcontinent and analyzing dominant flow generating processes in various regions. We think this is a good example to illustrate the usefulness of a large-scale model with relatively high resolution using the multi-basin approach of mostly ungauged basins. This is now clarified in the Introduction chapter and we have inserted new sections in the Methods and Results chapters, respectively, and 2 more figures in the Results and a paragraph in the Conclusions. We think this part of the manuscript is essential and will make the paper interesting for a much broader audience.

- *- The discussion on uncertainties and data errors, while exhaustive, is almost entire qualitative. With the very many methods for uncertainty analysis available, a more comprehensive uncertainty analysis would sure make a strong point.*

This is generally a good point and uncertainties should certainty be considering in modeling studies, but we disagree that our approach is "almost entirely qualitative". On the contrary:
- Figure 4 shows quantitatively uncertainty for various signatures during calibration and validation periods and "blind" catchments that were not used in calibration.
- Figure 5 shows quantitatively the uncertainties due to reservoir regulation.
- Figure 7 shows quantitatively the behavioural range (defined from the 100 best parameter sets for each objective function) from a Monte Carlo approach exemplified for one parameter.
- Figure 9 shows quantitatively the progress in model performance and consistency, and corresponding decrease in errors during the stepwise calibration.
- Figure 12 shows quantitatively the spatial model performance using different metrics across the subcontinent.

A comprehensive overall uncertainty analysis (of all the sources of uncertainty present at such a large scale) requires numerous model realisations, which are practically very difficult (but we guess not impossible) for a computationally heavy model. To implement such a methodology at 6000 subbasins in a distributed model is really not straightforward and would end-up in a new study. For this manuscript, there are no intentions to proceed in such experiment, as this is not the objective of this study.

- *- Lastly, the monsoon climate characteristics of the study region do not help. With such an extreme seasonality, and only monthly data available, the modelling challenge essentially boils down to a yearly water balance prediction, which is almost entirely dominated by uncertainties in the precipitation data and the evapotranspiration parameterisation. Under such conditions it is hard to do proper model diagnostics.*

To illustrate that the observed monthly data of discharge do capture seasonal variability of the systems (and not only the yearly water balance) we have now extended Fig. 2 to also include intra-annual cycles of precipitation, evapotranspiration and discharge at 4 river basins with significantly different flow regimes. We hope this makes it clearer to the reader. Moreover, our results shows that for instance soil characteristics and human alterations also had large impact on model results, not only precipitation and evaporation.

Based on this, I strongly recommend the authors to rethink the message they want to convey (and its novelty) and how they'd convey it.

To sum up, we have significantly shortened the text and re-written the text (especially section 2 and 4) to clarify our message and better emphasise the major highlights. The Discussion section is now merged with the Results section, while sections 5.2 and 5.3 from the first version of the manuscript are now completely removed to avoid repetition. Moreover, we have inserted new findings on similarities in runoff patterns, mostly from ungauged basins, across the entire sub-continent and an analysis of dominant drivers. We would like to thank Referee #2 for suggesting these improvements to our manuscript. We think this ended up in a much better paper, which hopefully should be interesting to a much broader audience.

2 Specific comments

2886/23: adequate is about as vague as it gets. Of course the model performance is discussed in more detail further on in the manuscript, but I think it would still be very useful to give some form of purpose and fitness for purpose evaluation to the model (see general comments).

We agree that the word "adequate" is not appropriate given the lack of stating the modeling purpose. The sentence is now modified as "*... The results show that despite the strong hydro-climatic gradient over the subcontinent, a single model can describe the spatial variability in dominant hydrological processes at the catchment scale. In addition, spatial model deficiencies are used to identify potential improvements of the model concept...*"

2886/27: consistent: that is of course nice, but again very vague. One the one hand, of course many model structures are applied over a large variety of hydrologies (global models being the extreme case) so it is not really that unusual to expect models to be consistent over the variety of hydrological processes considered in this paper. On the other extreme, it would be easy to rebuke the claim of consistency, given that there are so many known and unknown processes such as water abstraction, irrigation, etc. that are not or only very limitedly represented by the model. So again, I feel that I am missing the point. Did you expect the model not to be consistent? This goes very much against hypothesis driven research setup.

The sentence is removed to avoid confusion. However, we would like to note that human impacts are modelled to some extent (see model description section).

2887/20: "increasing the information content": of what? Of your model? Not really. Perhaps the best formulation is "Constraints are generated by independent information..."

The sentence is modified accordingly.

2888/6: "putting science into practice": I find it hard to find any evidence of this in this manuscript.

This sentence is no longer in the new version.

2889/1: "Influenced by human activities": This is the point, isn't it? But the presented study does not do this. I think it is one of the many missed opportunities.

We agree that investigating human impacts can be very promising (see the opinion papers by Nazemi and Wheater, (2015a;2015b)), however we are limited by a single paper which aims to exemplify the usefulness of PUB recommendations in multi-basin modeling at the large scale.

2889/11: "test the recommendations": can you really test recommendations? I can't help but being a bit sceptical.

The paragraph is reformatted and the sentence is removed to avoid confusion.

2889/18: "frequent quality checks": arguably with all models. The case for large-scale models is not made very convincingly here.

We have further elaborated throughout the manuscript on why this is even more important at the large scale as mismatches of datasets and errors are even more difficult to find and trace using big data.

2890/3-5: again, these are hardly unique to the large-scale. Of course data collection will be harder, but could even turn the argument around and claim that large-scale modelling is easier compared to a data-scarce small catchment, because of the availability of global datasets and the fact that many small scale errors are smoothed out.

This paper focuses on the large-scale and the problems involved in multi-basin modelling. Of course, all problems are not unique but we have tried to emphasize the new problems arising at this scale. The new manuscript is re-written to highlight this more.

2891/9-14: apart from being quite a cliché, I would disagree with "familiar to the modeller" and "can be easily set up and run"

The last part is removed as it might not always be true, but we would really like to emphasize the role of the modeler and using a model to which you are already familiar with, as this significantly helps for the understanding of the results.

2892/4-20: again I would challenge the authors to go beyond the commonplace. Moreover, further on the manuscript the authors mention that they did 3 extensive field visits to the basin and go on to state that they have been very useful for the modelling exercise but. But again there is no discussion as to how it has been useful beyond some general statements on source of irrigation water! I think that such a discussion could be highly informative, but it would have to be much more in-depth. Here is your chance to show really the value of (expensive and time consuming!) field visits.

Unfortunately, the field visits did not cover experiments and/or data acquisition. The usefulness of the field trips was more for soft information and understanding (i.e. to translate the databases to landscapes) - it is indeed challenging to find out the real value. We would like to start collecting such examples and save this to another paper dedicated entirely to this topic. This is an interesting idea for a study!

2892/24-26: This is a rather platitude statement, isn't it? Not only don't you have thousands of stations in your case study, in the (luxury) case where such information is available, it should not be too hard to filter unrealistic values plot spatial and temporal patterns etc. After all, from a statistical perspective it would not be such an unusual large dataset!

We agree that in this investigation we did not have many stations and hence an inspection of time-series was not time-consuming. However, Section 2 provides recommendation for general applications on multi-basin modeling and is not focused in India (apart from the Introduction, reference to the Indian application starts in section 3). For instance, analyses that are recommended in section 2.2 are conducted in the HYPE model setup at the pan-European scale, in which 2600 discharge stations were available. The burden here is not the computation effort but rather the time-consumption. To avoid confusion we re-wrote this sentence as "*Detailed inspection of flow signatures for each gauging station from large datasets (often in the range of thousand stations, see http://hypeweb.smhi.se/) is best done by…*" and we have inserted a

6

reference to http//hypeweb.smhi.se/ where model domains with more gauging stations can be found.

2893/11: "practically not feasible": A rather strong statement in this era of cloud computing and big data.

The problem is that most catchment models are not parallelized and it is not always straight forward to do this for a complex processed-based model with river routing. We are working on it but the run times are still rather long for multi-basin modeling using HYPE.

To avoid confusion we now state "*practically difficult…*". Note that the sentence is moved in section 2.6.

2894/10 - 26: I would think that these recommendations belong to basic quality control practice, and surely are not specific or particularly important to large scale modelling.

As mentioned earlier, we agree that quality checks are also conducted in catchment scale modeling. However, the necessity, the required time, and the approaches (given that we deal with large datasets) differ between catchment scale and multi-basin modeling. In this section, we aimed to highlight the need for quality checks (note that although been obvious to some, this step was missing from the PUB best practices) and list methods that in our experience have been proven useful for quality checks.

2895/13: "lake parameters": this comes a bit out of the blue. Why lakes? Surely they influence the hydrological response, as do probably a plethora of other processes (e.g., water intakes, swamps, unusual groundwater systems, …). Again it is a pretty generic (though not always the most useful) strategy in analysing large datasets to start with a subset…

This sentence is removed. Section 2.4 is re-written to avoid repetition and generalizations.

2896/9-10: Again a rather unbearable generalization. Who could argue against this at any scale?

To avoid generalizations, this sentence is removed.

2896/section 2.6: despite featuring here rather prominently, the absence of a rigorous uncertainty analysis (apart from a lonely mention of the "behavioural range" of parameters in the caption of figure 7) is in my opinion one of the major shortcomings of the paper.

Please see our response to the general comment about lack of comprehensive uncertainty analysis.

2908/23: "of in" -> in

Corrected accordingly.

2939: "behavioural": this is the only mention of behavioural in the entire manuscript. How where they obtained?

To avoid confusion, this is now changed into "*… the range is derived from the 100 parameter sets that perform best…*"

2942: "outlet of the subbasin" -> outlet of the basin? If not, what subbasin?

This is now corrected as "outlet of the basin"

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

I.G. Pechlivanidis and B. Arheimer

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

ABSTRACT

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

1. INTRODUCTION

Numerical hydrological models have been used world-wide for operational needs and scientific research since the early 1970s (e.g. Hrachowitz et al., 2013; Pechlivanidis et al., 2011; Refsgaard et al., 2010; Singh, 1995). 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 independent information via either additional data, i.e. remote sensing, tracers, quality, multiple-variables, etc. (Arheimer et al., 2011; Finger et al., 2011; McDonnell et al., 2010; McMillan et al., 2012; Samaniego et al., 2011) and/or expert knowledge (Bulygina et al., 2012; Fenicia et al., 2008; Gao et al., 2014).

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

2

113   Multi-basin modelling complement the "deep" knowledge from single catchment modelling when

114   applied to a large geographical domain covering a large sample of observations (Andreassian et al.,

115   2006; Arheimer and Brandt, 1998; Gupta et al., 2014; Johnston and Smakhtin, 2014). However, the

116   majority of basins world-wide are effectively ungauged, as are also the subbasins (defined here as

117   prediction points in the model set-up) in a high resolution multi-basin model at the large scale.

118   Hydrological modelling at the large scale has the potential to encompass many river basins, cross

119   regional and international boundaries and represent a number of different physiographic and climatic

120   zones (Alcamo et al., 2003; Raje et al., 2013; Widén-Nilsson et al., 2007). Traditionally, the

121   performance and the spatiotemporal resolution in such models was poor, but the current release of

122   open and global datasets has given new opportunities for catchment hydrologists to contribute.

123   Application of multi-basin modelling at the large scale can be used to predict the hydrological

124   response at interior ungauged basins (Arheimer and Lindström, 2013; Donnelly et al., 2015;

125   Samaniego et al., 2011; Strömqvist et al., 2012). The use of large sample of gauges can also facilitate

126   comparative hydrology allowing to test hypothesis for many catchments with a wide range of

127   environmental conditions (Blöschl et al., 2013; Donnelly et al., 2015; Falkenmark and Chapman,

128   1989). In addition, the multi-basin approach can be used to map spatial variability and explore

129   emerging patterns of for instance climate change (see http://hypeweb.smhi.se/).

130

131   Modelling at the large scale, however, includes additional model uncertainties. Physical properties

132   (e.g. topography, vegetation and soil type) in large systems generally show higher spatial variability

133   and thus larger heterogeneity in system behaviour (Coron et al., 2012; Sawicz et al., 2011), which in

134   turn affects model parameters (Kumar et al., 2013). In addition, large river basins are often strongly

135   influenced by human activities, such as irrigation, hydropower production, and groundwater use, for

136   which information is rarely available at high resolution in global databases. This introduces additional

137   uncertainty regarding process understanding and description at the large scale. Moreover, the

138   topographic and forcing data of global datasets (i.e. water divides, weather and climatic data) are more

139   likely to be inconsistent, erroneous, and/or only available at a coarse resolution (Donnelly et al., 2012;

140   Kauffeldt et al., 2013).

141

142   Applying catchment models at the continental scale in a multi-basin manner is a way to introduce

143   catchment modelling approaches to the existing global hydrological models, i.e. land-surface schemes

144   and global water-allocation concepts. In this paper, we therefore present a set of examples on how the

145   scientific advancements during the PUB decade have improved the potential for process-based

146   hydrological modelling at the large scale. We identify specific challenges at the large scale and

147   exemplify on how to overcome them. In here, we further develop and slightly modify the PUB best

148   practices to be applicable at the large scale. We use examples from the recent HYPE model set-up of

149   the Indian subcontinent, which experiences unique and strong hydro-climatic and physiographic

3

characters and poses extraordinary scientific challenges to understand, quantify and predict hydrological responses. We particularly address 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 improvement on parameter identification by using remote sensing data and expert knowledge. We further show how regions can be grouped based on physiographic 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 this data sparse region. In addition, we investigate potential links between model performance and physiographical characteristics to understand model inadequacies along the gradient. Finally we cluster the catchments based on their hydrological functioning and discuss how process understanding can benefit from multi-basin modelling and what hydrological insight can be gained by analysing spatial patterns from large-scale predictions in ungauged basins.

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

Takeuchi et al. (2013) recommend a six step procedure for predicting runoff at locations where no observed runoff data are available (Figure 1A). This best practice recommendation is intended for single catchments, and requires modification when applied to multi-basins at the large scale (Figure 1B). In this section, we present our best-practice recommendations for large-scale applications of process-based models. They are based on our interpretation of the best practices and previous experience from PUB in multi-basin applications (e.g. Andersson et al., 2015; Arheimer et al., 2012; Donnelly et al., 2015; Strömqvist et al., 2012), which are visualised at http://hypeweb.smhi.se/.
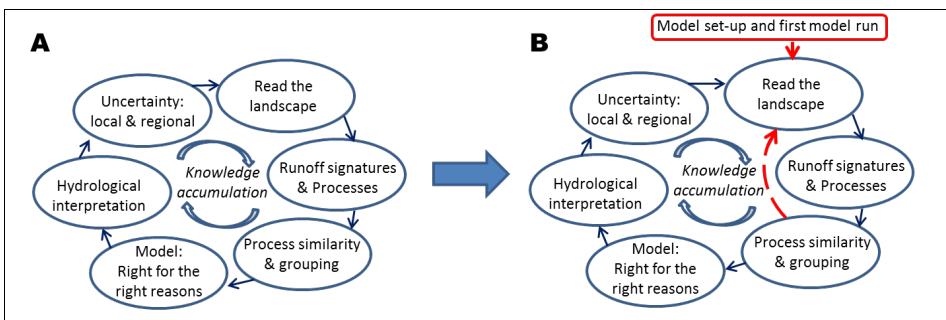
Figure 1. Best practices for predictions in ungauged basins: A) according to Fig. 13.1 by Takeuchi et al. (2013) in Blöschl et al. (2013), and B) modified version for multi-basin applications at the large scale.

4

264

265 Many sources of uncertainties/errors appear when handling big datasets and may be time consuming

266 to be discovered. Analysis of each dataset or catchment may be impractical and risk focusing on

267 details instead of the most crucial overall hydrological functioning across the model domain. We

268 therefore recommend starting with a top-down approach, in which the model is setup directly before

269 proceeding with the PUB recommendations (circle of steps in Fig. 1). The hydrological model needs

270 to include the description of most water fluxes, storages and anthropogenic influences that can be

271 relevant and satisfy the modelling objectives. In addition, we recommend to use a model that is

272 familiar to the modeller and open for changes, to allow coherent hydrological interpretations and code

273 adjustments to cope with the region's spatial heterogeneity and hydrological features. Setting-up the

274 model system includes to: (i) acquire readily available datasets that cover the entire geographical

275 domain or merge datasets to get a full coverage; (ii) define calculation points and river network, by

276 taking into account the location of gauges, major landscape features, user requests, catchment borders

277 and routing; (iii) make a first set of model input data files and make the first model run for the model

278 domain with a multi-basin resolution. The analysis of preliminary results from setting up the full

279 system at once will indicate major obstacles, such as systematic errors in input data or model

280 structural limitations. Moreover, by having the technical system in place immediately facilitates an

281 incremental and agile approach to model set-up, with direct feed-back on model performance at many

282 gauges. Once the model runs for the full domain, we recommend starting to improve the performance

283 according to the six steps of best practices for predictions in ungauged basins, using a bottom-up

284 approach to refine input data, model structure and parameter values.

285

286 **2.1. Read the landscape**: *"Go out to your catchment, look around…!"* (cit: page 385 in Blöschl et

287 al. (2013))

288 It is practically impossible to visit the full variety of basins in a large-scale model domain, so instead

289 we recommend: (i) navigate on hard-copies, digitised maps and webpages (e.g. Google Earth) to

290 check landscape characteristics; (ii) review the literature for dominant processes and well-known

291 features or hydrological challenges in the region; (iii) proceed with quality checks and cross-

292 validations towards other data sources (i.e. sources that contain limited in space but local

293 information); (iv) validate the basin delineation and routing using archived metadata from other

294 available datasets; (v) check quality of observed discharge data to assure coherence of time-series; and

295 finally, (vi) check the spatiotemporal information of meteorological datasets after transformation from

296 the grid to the subbasin scale. It is important to get an understanding of the full domain but also to

297 ensure that the datasets correspond to this understanding, as errors often appear when handling and

298 interpreting large datasets.

299

300 **2.2. Runoff signatures and processes:** *"Analyse all runoff signatures in nearby catchments to get*

5

386   *an understanding…!"* (cit: page 385 in Blöschl et al. (2013))

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

394

395   **2.3. Process similarity and grouping:** *"…find similar gauged catchments to assist in predicting*
396   *runoff in the ungauged basin!"* (cit: page 385 in Blöschl et al. (2013))

397   In most process-based models, the modeller has some freedom to define the characteristics of the
398   smallest calculation units, which is normally linked to physiography, to account for spatial distribution
399   of for instance soil properties or land use. When producing these calculation units for large domains,
400   we need to be restrictive with the number of classes and we normally redistribute small calculation
401   units to speed up the model run times; both technical and conceptual concerns must be taken into
402   account. However, lakes, wetlands, glacier, and urban areas should remain as even small proportions
403   can significantly alter the flow regime. When calculation units are defined, we recommend clustering
404   the basins/gauges with similar upstream characteristics and/or system behaviour to isolate key
405   processes for regionalisation of parameter values during calibration. We finally suggest checking the
406   spatial distribution by plotting the catchment characteristics of subbasins on maps and compare to
407   other or original data sources.

408

409   *2.1-3. Quality checks:* This is an additional step in the procedure accounting for repetition of step 1-3
410   in an iterative way to ensure quality in the required input data and files of the model (Figure 1); it is
411   easy to fail and introduce errors when handling large datasets by automatic scripts (generalisation of
412   scripts is not always straightforward and some manual adjustment is usually required) and/or human
413   error (particularly when many modellers collaborate). It is important to remove as many errors as
414   possible in the input data before starting to tune parameters; otherwise the calibration may lead to
415   erroneous assumptions on hydrological processes to compensate for input data errors. We recommend
416   to analyse flow time-series as follows: (i) compare modelled to observed time-series and signatures;
417   (ii) check water-volume errors and their distribution in space; (iii) inspect the spatial distribution of
418   model dynamics to correct spatial patterns from systematic errors; and (iv) search for errors in the
419   model set-up (routing, meteorological input etc.).

420

421   **2.4. Model - Right for the right reasons:** *"Build… model for the signature of interest…*
422   *regionalise the parameters from similar catchments…more information than the*

6

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

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

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

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

Multi-basin models are more computationally demanding than single basin models and it is therefore not always feasible to explicitly address all uncertainties from all sources. To explore the model performance in ungauged basins, we recommend dividing the set of gauging stations into those used in calibration and validation, respectively. Cross-validation, e.g. using the jackknife procedure (Good, 2005), is practically difficult in process-based modelling of multi-basins. Instead we recommend
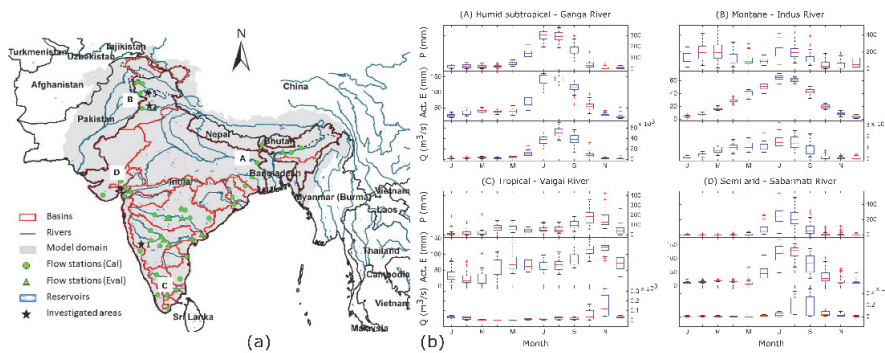
7

using a subset of the validation gauges for "blind tests", to be independent from any calibration or model tuning. To examine uncertainties we recommend to: (i) use several performance (diagnostic) criteria, and many flow signatures; (ii) relate the spatial distribution of model performance to physiographical variables; and (iii) check model performance for independent gauging sites and new datasets.

The major 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 search for local knowledge again to elaborate new hypotheses of hydrological functioning and data sources. We recommend to document and version-manage each model set-up before looping into step 1, to ensure knowledge accumulation for a broader audience and to make the set-up process transparent. This sets a baseline for the next round of improvements.

3. DATA AND METHODS

3.1. Study area and data description

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

8

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

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 agencies but are not open to the public. Consequently, in this application, flow information (Table 2) is available only for a small fraction of the subcontinent, which makes the region a great example for PUB. Monthly potential evapotranspiration (pot. E) data were obtained for the period 2000-2008 from the Moderate Resolution Imaging Spectroradiometer (MODIS) global dataset (Mu et al., 2007, 2011). The dataset covers the domain in a spatial resolution of 1 km and is derived based on the Penman-Monteith (Penman, 1948) approach.

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

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

Characteristic/Data type	Info/Name	Provider
Total area (km^2)	4.9 million	-
Number of subbasins	6 010 (mean size 810 km^2)	-
Topography (routing and delineation)	Hydrosheds (15 arcsec)	Lehner et al. (2008)
Soil characteristics	Harmonised World Soil Database (HWSD)	Nachtergaele et al. (2012)
Land use characteristics	Global Land Cover 2000	Bartholomé et al. (2002)

	(GLC2000)	
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° × 0.25°)	Yatagai et al. (2012)
Temperature	AphroTEMP (0.5° × 0.5°)	Yasutomi et al. (2011)
Potential evapotransp.	MODIS pot. E (1 km)	Mu et al. (2011)

749

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

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

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

752

753 ### 3.2. A multi-basin hydrological model for large-scale applications - the HYPE model

754 The Hydrological Predictions for the Environment (HYPE) model is a dynamic rainfall-runoff model,

755 which describes the hydrological processes at the catchment scale (Lindström et al., 2010). The model

756 represents processes for snow/ice, evapotranspiration, soil moisture and flow paths, groundwater

757 fluctuations, aquifers, human alterations (reservoirs, regulation, irrigation, abstractions), and routing

758 through rivers and lakes. The HYPE source code is continuously developed and released in new

759 versions for open access at http://hype.sourceforge.net/, where also model descriptions, manuals and

760 file descriptions can be downloaded.

761

762 HYPE is most often run at a daily time-step and simulates the water flow paths in soil for

763 Hydrological Response Units (HRU), which are defined by gridded soil and land-use classes and can

764 be divided in up to three layers with a fluctuating groundwater table. The HRUs are further

765 aggregated into subbasins based on topography. Elevation is also used to get temperature variations

766 within a subbasin to influence the snow melt and storage as well as evapotranspiration. Glaciers have

767 a variable surface and volume, while lakes are defined as classes with specified areas and variable

10

volume. Lakes receive runoff from the local catchment and, if located in the subbasin outlet, also the river flow from upstream subbasins. On glaciers and lakes, precipitation falls directly on the surfaces and water evaporates at the potential rate. Each lake has a defined depth below an outflow threshold. The outflow from lakes is determined by a general rating curve unless a specific one is given or if the lake is regulated. 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 (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. After subtraction of conveyance losses, the withdrawn water is applied as additional infiltration to the irrigated soils. River discharge is routed between the subbasins along the river network and may also pass subbasins, flow laterally in the soil between subbasins or interact with a deeper groundwater aquifer in the model. For the study in this paper, the HYPE model version 4.5.0 was set up for the entire Indian subcontinent (4.9 million km^2) with a resolution of 6 010 subbasins, i.e. on average 810 km^2, and is referred to as India-HYPE version 1.0.

3.3. Model calibration and regionalisation

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

11

Stepwise, iterative calibration of parameter groups

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

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

Expert knowledge for parameter constraints

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

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

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: .

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: , after which another

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: 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

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: in

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: The lake/reservoir

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: irrigation

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: may be difficult to identify, particularly at ungauged areas, and usually require manual

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: driven by known regulation routines

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: observed time series at the outlet of the lakes/reservoirs, and soft information about water extraction at the regional scale

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: step-wise

12

882 The *mactrsm* values for the remaining two soil types in the India-HYPE model domain, i.e. organic
883 and shallow, are expected to be close to the corresponding values for the coarse soil; although the
884 value for shallow soil is constrained to be less than *mactrsm* for organic soils.
885

886 *Spatial clustering based on catchment similarities*
887 We assume hydrologic similarity across the region on the basis of similarity in physiographic
888 characteristics. We applied a k-means clustering approach within the 17-dimensional space, consisting
889 of: 5 soil types, 7 land use types, mean annual precipitation, mean temperature, mean slope, mean
890 elevation, and basin area. This separated the subbasins into homogeneous classes. A silhouette
891 analysis was used to overcome the subjectivity on the determination of the number of clusters. The
892 catchment similarity approach significantly reduces the number of parameters, while it allows
893 regionalisation of parameters, which are assumed to be robust enough also for ungauged basins.
894

895 *Spatiotemporal calibration and evaluation*
896 India-HYPE was calibrated and evaluated in a multi-basin approach by considering the median
897 performance in all selected stations. 30 stations were selected for model calibration and 12 "blind"
898 stations for spatial validation. The years 1969-1970 are used as a model warm-up period, the next 5
899 years for model calibration (1971-1975) and the final 4 years for temporal performance evaluation
900 (1976-1979).
901

902 The Differential Evolution Markov Chain (DE-MC; Ter Braak, 2006) optimisation algorithm is used
903 to explore the feasible parameter space and to investigate parameter sensitivity. DE-MC was applied
904 at each step of the iterative calibration procedure with 200 generations of 100 parallel chains each
905 being explored respectively. The Kling-Gupta Efficiency, KGE (Gupta et al., 2009), was used to
906 define the performance of the model towards the observed discharge. KGE allows a multi-objective
907 perspective by focusing to separately minimise the correlation (timing) error, variability error, and
908 bias (volume) error. We also investigated the relative influence of timing, variability and volume error
909 on the KGE value. To do this, we transformed the three components to result into a consistent range
910 of possible values (the metrics are named as *cc*, *alpha* and *beta* corresponding to timing, variability
911 and volume errors respectively; see Appendix A).
912

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

913
914

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

13

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

3.4. Evaluation beyond standard performance metrics

Evaluation based on flow signatures

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

Multi-variable evaluation

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

Linking performance to physiographical characteristics

To better understand the model performance and identify potential for model improvements, we apply classification and regression trees (CART; Breiman et al., 1984). CART is a recursive-partitioning algorithm that classifies the space defined by the input variables (i.e. physiographic-climatic characteristics) based on the output variable (i.e. KGE model performance). The tree consists of a series of nodes, where each node is a logical expression based on a similarity metric in the input space (physiographic-climatic characteristics). In this case, we divided the KGE performance into three groups – bad (KGE < 0.4), medium (0.4 < KGE < 0.7), and good (KGE > 0.7), which were termed C0, C1 and C2 respectively. A terminal leaf exists at the end of each branch of the tree, where the

14

probability of belonging to any of the three output groups can be inspected. Here we summarised the physiographic-climatic characteristics of the basin into 5 soil types (coarse, medium, fine, organic and shallow), 7 land use types (crops, forest, open land with vegetation, urban, bare/desert, glacier, water), mean annual precipitation and mean temperature.

3.5. Catchment functioning across gradients

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

4. RESULTS AND DISCUSSION

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

4.1. Read the landscape

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

1077 needs and estimation of water losses due to faults in the irrigation systems).

1079 Analysis of the topographic data was of major importance since they affected the subbasin delineation
1080 and routing. Although Hydrosheds are based on high-resolution elevation layers, which are
1081 hydrologically conditioned and corrected, there are still many errors. Merging Hydrosheds with
1082 GRDC (hence forcing the delineation at subbasins where GRDC stations are available) involved some
1083 mismatches in terms of the size of upstream areas between the subbasin delineations and the GRDC
1084 metadata. As an example, the location of the Dundeli station in the Kali Nadi river basin (asterisk 1 in
1085 Figure 2) was adjusted to match the underlying topography and drainage accumulation data based on
1086 published and computed upstream areas respectively (see Figure 3a). The consequent change in the
1087 routing resulted in a considerable improvement in the model performance (KGE improved from -0.51
1088 to 0.30; see Figure 3b). Many similar corrections had to me made.
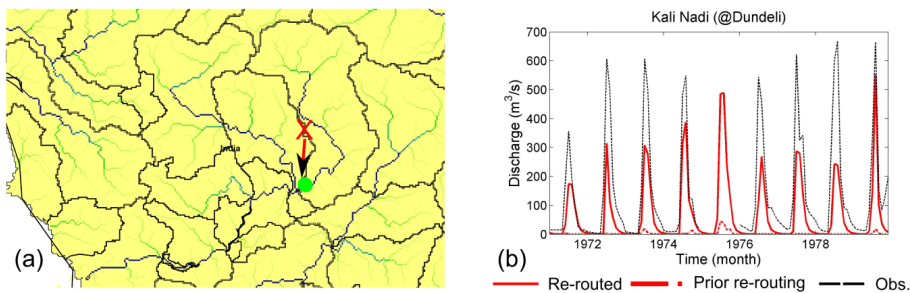
Figure 3. Example of the impact of basin delineation and routing on model behaviour: (a) correction
in the location (red x and green circle is prior and after the correction respectively) of the Dundeli
discharge station (Kali Nadi river basin), and (b) the corresponding modelled discharge before and
after the correction. In (a) the subbasins and flow accumulation are also depicted.

1097 To make corrections also for ungauged basins and major rivers, the delineated basins were
1098 additionally evaluated using a shapefile of basin areas reported by Gosain et al. (2011). Some minor
1099 corrections had to be done in the routing to achieve similarly delineated basins, particularly in the
1100 northwest region, where mean elevation at the subbasin scale does not show much variability.

4.2. Runoff signatures and processes

1103 As recommended, several flow signatures were extracted for the gauging stations across India to be
1104 compared to physiographical patterns. Flow signatures were also used for model evaluation to find
1105 potential for improvements. The analysis was done at different stages in the model set-up, and finally,
1106 there was a relatively good agreement of the observed and modelled flow signatures (Fig. 4). In

16

Margin comments:

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: In addition, analysis

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: has been very important

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: affect

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: their spatial resolution can affect basin discretisation. Here, merging

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: WHIST to generate

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: of catchments for some stations

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: Fig. 2

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: Fig. 3

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: Fig. 3

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: b). Hence, the recommendation to carefully check the subbasin delineation is of major importance for multi-basin model performance

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: catchment

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: In addition

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: generated by WHIST

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: ; these are presented in Fig. 2 (in red).

Ilias Pechlivanidis 7/27/2015 17:41
**Formatted**

Ilias Pechlivanidis 7/27/2015 17:41
Formatted: Heading 2, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.27 cm

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: 5) in long-term averages of discharge (Qm) and high flow ($q05$).

1126   general, poor agreement was found in mountains and in semi-arid regions, which are characterised by

1127   local, convective rainfall events during the monsoon season. No clear pattern is found between

1128   signature agreement and basin scale for calibrated river gauges.

1129

1130
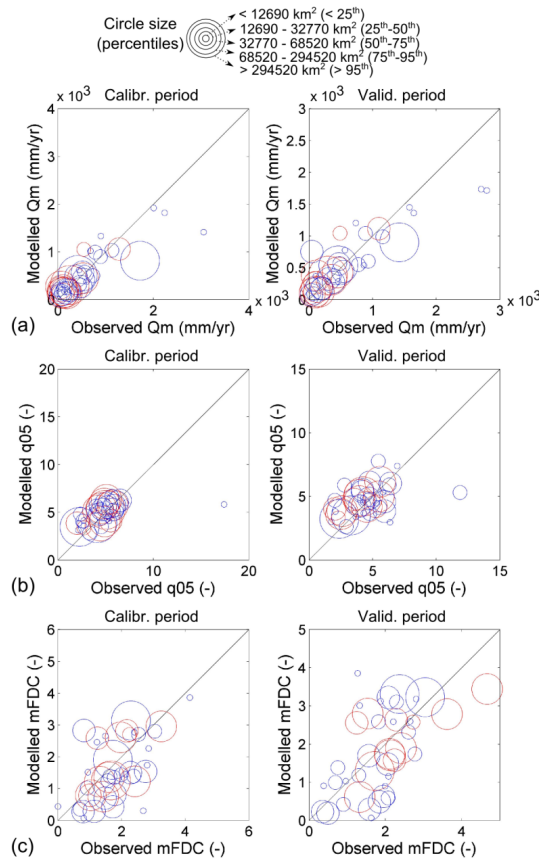
1131

1132   Figure 4. Signature analysis in the spatiotemporal model evaluation: (a) the mean annual specific

1133   runoff, (b) the normalised high flow statistic, and (c) the slope of the flow duration curve. Blue and

1134   red circles are used for the calibration and evaluation stations respectively.

1135

1136   We also explored how flow signatures can be affected by human impacts by analysing modelled

1137   responses considering and omitting the human influence. Figure 5 highlights the significant effect

1138   reservoirs have to dampen hydrographs and control discharge variability; hence various flow

1139   signatures. The model can fairly well represent the reservoir routing and KGE improved from 0.37 to

1140   0.48 after introducing a regulation scheme. The model improved on capturing the seasonality of

1160 regulation; however at this modelling state it was not able to represent the monthly peaks. Note that
1161 model results are subject to the general rating curve generalised to all reservoirs; there were no
1162 downstream data available to calibrate the parameters specifically for a given reservoir/dam.
1163
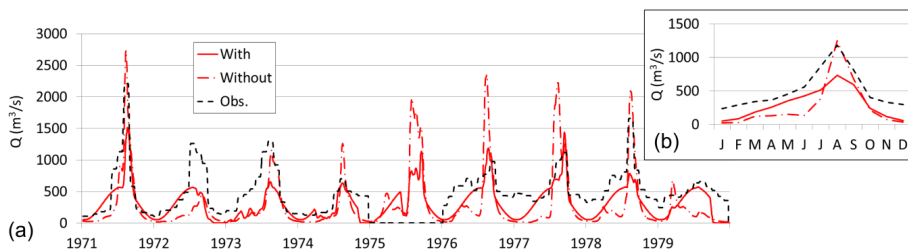
1164
1165 Figure 5. Impact of model parameterisation of reservoir regulation on discharge for (a) monthly
1166 streamflow, and (b) annual hydrograph, showing naturalised (without) and regulated (with) conditions
1167 at the basin outlet (located at asterisk 2 in Fig. 2).
1168

4.3. Process similarity and grouping

1170 After having identified relevant HRUs, reclassified them into suitable calculation units and inserted
1171 major features as lakes and dams, we identified basin similarities to drive the identification of the
1172 model's regional parameters. The cluster analysis was applied to all 6 010 subbasins of the domain
1173 within the 17-dimensional space (see section 3.3). We identified 13 different classes of varying size
1174 (Figure 6) out of 42 values, which is the number of gauged river-basins in the domain, yet with
1175 relatively high class strength (i.e. the variability of characteristics within each cluster is relatively
1176 low). It is important to note that the physiographic (soil and land use) characteristics had more
1177 influence on the clustering as opposed to the climatic properties; the clustering was repeated without
1178 climatic information but the spatial pattern of the clusters remained. In the last stage of the stepwise
1179 calibration procedure, the regional model parameters were estimated for each cluster region. When
1180 using the clustering for regional calibration (Section 5.4), however, it could not significantly improve
1181 the overall model performance but nevertheless, the model consistency at all stations was improved.
1182 Overall, we found a high potential of catchment similarity concepts to drive parameter identification
1183 in the ungauged basins.
1184

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

4.1-3. Quality checks

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

In addition, several modelled and observed flow signatures were compared repetitively at every stage of model refinement. We found it valuable to adjust as much as possible before starting to work on parameter values and model algorithms. For instance, the analysis of flow time series and signatures during the first model runs showed consistent underestimation of runoff in the Himalayan-fed basins. A comparison of the mean annual precipitation between Aphrodite and national precipitation gridded data provided by the Indian Meteorological Department, showed an underestimation of the Aphrodite precipitation in the mountainous regions; the Aphrodite precipitation network is sparse over Himalaya (Yatagai et al., 2012). To overcome this underestimation, a correction factor was applied to

19

precipitation (in HYPE, this was a multiplier of 4% per 100 m) at regions with elevation greater than 400 m. Allowing such modification in the data, we expected that calibration of model parameters could further compensate precipitation uncertainty.

4.4. Model – Right for the right reasons

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

Additional data sources

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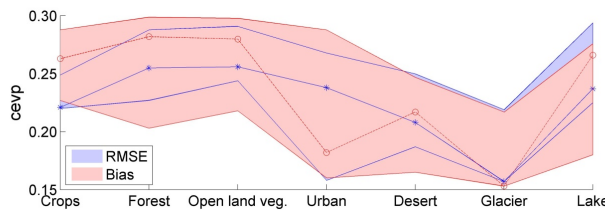

Figure 7. Coefficient of potential evapotranspiration (*cevp*) parameter as identified (the range is derived from the 100 parameter sets that perform best, and the optimum set) for different objective functions (RMSE and Bias) and land use type. Lines with markers present the optimum parameter values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

Expert knowledge

20

Expert knowledge was applied to filter out unrealistic relationships of the *mactrsm* parameter for different soil types (see section 3.3). Both the constrained and unconstrained models resulted in a comparable calibration performance; median KGE was 0.48 and 0.49 for the constrained and unconstrained models respectively. The optimum set for the unconstrained model gave an unrealistic distribution of the parameter values for the coarse and medium soil types (Figure 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.
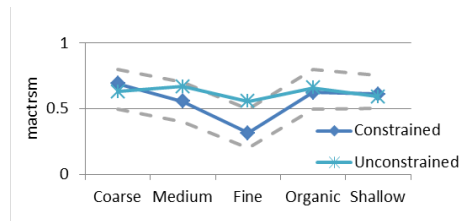
Figure 8. Constraints (grey dashed lines) and optimum (solid lines) values of the *mactrsm* soil
dependent model parameter based on process understanding.

Stepwise calibration procedure

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

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: 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 section 3.3) while the prior used HYPE parameter values for Sweden. Fig. 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
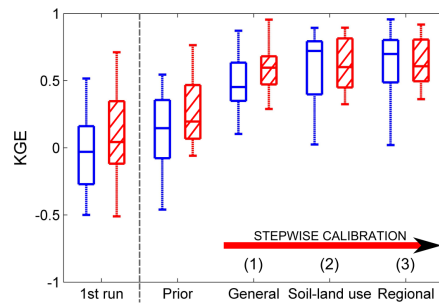
Figure 9. Improvements in model performance (average KGE for 30 stations) during the stepwise calibration approach (steps 1-3 correspond to general, soil-land use, and regional calibration as described in section 3.3). "1st run" corresponds to model performance of the very first model set-up to establish 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.

4.5. Hydrological interpretation

The temporal interpretation was done by analysing interacting dynamics of internal model variables, i.e. precipitation (P, mm), snow depth (SD, mm), temperature (T, °C), evapotranspiration (E, mm), soil moisture deficit (SMDF, mm), and discharge (Q, m^3/s). These are checked visually in a set of validation basins, to avoid unrealistic model behaviour due to parameter setting. Results from this point onwards correspond to the calibrated India-HYPE model (after step 3 in Figure 9). Results in the Chenab River at the Akhnoor station (branch river of the Indus system; asterisk 3 in Figure 2) show that the snow melt characterises the monthly hydrograph (Figure 10). Snow accumulation/melting processes occur at the headwaters of the basin which experience T below 0 °C during the winter and pre-monsoon period and above 0 °C during the rest of the months ("Up" black-dashed T series in Figure 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.

22

1379
1380
1381　　Figure 10. Analysis of model variables: P, SD, T, E, SMDF and Q. E corresponds to potential (Pot.)
1382　　　　and actual (Act.) evapotranspiration, and Q corresponds to modelled (Mod.) and observed (Obs.)
1383　　　　　discharge). Note that P and T series are plotted at the outlet of the basin (Down) and the most
1384　　　　　　　　　　　upstream subbasin (Up).
1385

1386　For spatial interpretation of flow predictions, we investigated potential relationships between model
1387　performance and physiographic-climatic characteristics; hence identify the controls of poor model
1388　performance. Figure 11 shows the classification tree obtained when relating the KGE performance
1389　with physical and climatic characteristics across the domain. Results show that the dominant variables
1390　resulting in poor/good model performance are soil (medium and shallow) and climate (mean
1391　precipitation and temperature). Despite the relatively small sample is this analysis, results are
1392　insightful and show that poor performance (KGE<0.4) is generally achieved at basins with shallow
1393　soil type greater than 13%. The probability of obtaining poor performance is also highest for basins
1394　with medium soil type greater than 34% and precipitation less than 1038 mm. Consequently, emphasis
1395　should be given to parameters for medium and shallow soils in a future effort to improve the model
1396　performance.
1397

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: subbasin

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: Fig. 11

23
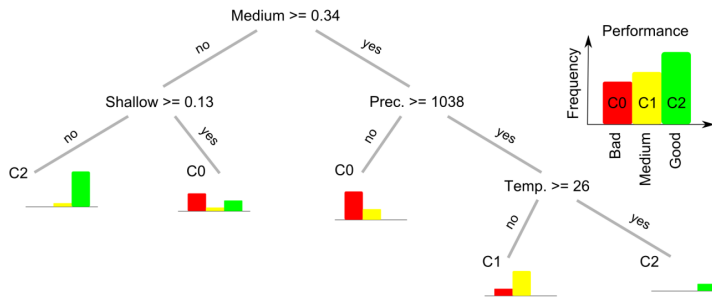
Figure 11. Classification trees relating regions of different KGE performance with physical and climatic characteristics. The bars represent the probability of a performance resulting in any of the three performance classes (C0, C1 or C2).

4.6. Uncertainty – local and regional

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

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

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

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

24

to reproduce the monthly variability of discharge varies regionally as shown by the KGE (Figure 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.
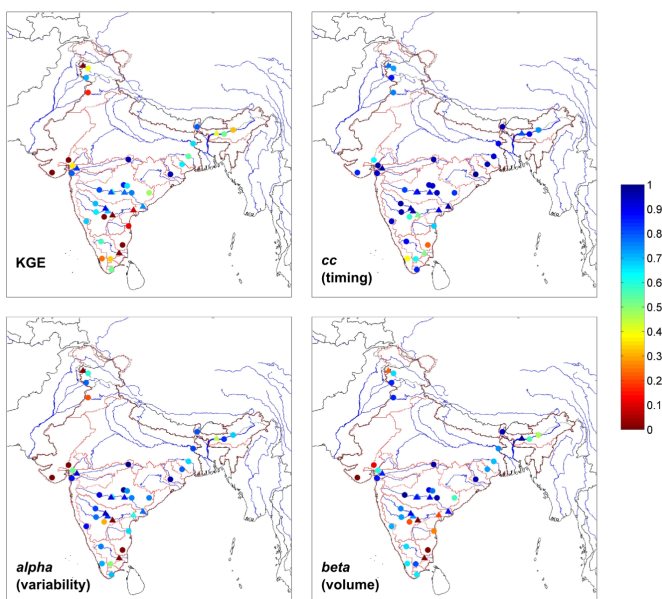
Figure 12. Spatial variability of KGE (and its decomposed terms) model performance for the calibration (circle) and validation (triangle) stations.

4.7. Spatial flow pattern across the subcontinent and dominant processes

Although the India-HYPE model has limitations, we identified potential for further improvements during the set-up procedure. The present version already demonstrated the usefulness of multi-basin modelling for comparative hydrology and how to gain insights in spatial patterns of flow generating

25

processes at the large scale. The final clustering analysis of the 12 flow signatures from India-HYPE version 1 resulted in six different classes of varying size (Fig. 13) with different distribution in signatures (Fig. 14). Similarity in catchment behaviour for each class was interpreted and dominant flow generating processes could be distinguished as follows:

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

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

26
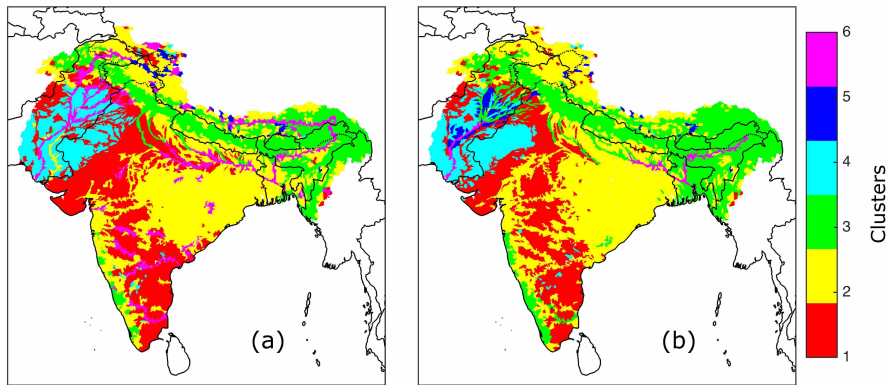
Figure 13. Subbasin clusters based on flow signatures at different stages of the model set-up: (a) Prior, and (b) Regional.
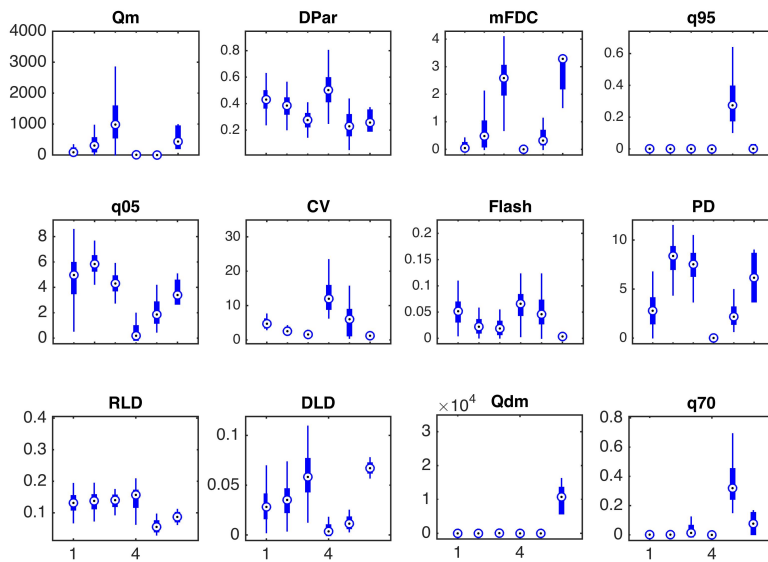
Figure 14. Distribution of signature values for each cluster (at Regional step). The flow signatures are described in Appendix A.

27

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

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

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

28

studies considered relationships including extraterrestrial radiation and relative humidity, i.e. Hargreaves-Samani or Penman-Monteith, which are expected to improve the magnitude and variability of evapotranspiration losses (Samaniego et al., 2011). Therefore the pot. E model component will be further investigated and refined in the next version of India-HYPE.

5. CONCLUSIONS

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

- Each step in the best practice procedure was relevant and we could find methods that also work at the large scale using the knowledge derived for catchments during the PUB decade. We argue to adapt an incremental and agile approach to model set-up, which requires frequent testing to get feedback on introduced changes. The large-scale modelling is more prone to technical problems and data inconsistencies that become apparent when running the model and therefore it should be done early in the model set-up process.

- Multi-basin modelling of ungauged rivers at the large scale reveals insight in spatial patterns and dominating flow processes. Indian catchments can be categorised into 6 clusters based on their flow similarity. River flow varies spatially in terms of flow means, variability, extremes and seasonality. Catchments in the Himalayan region and the Western Ghats seem to respond similarly and are characterised by high mean annual specific runoff values and variable flow regime. Response of the catchments in the tropical zone is characterised by high peaks, while catchments in the dry regions show very strong flow variability and respond quickly to rainfall.

- Overall the model showed high potential to represent the hydrological response across the region despite the strong hydro-climatic gradient. However, the India-HYPE v.1.0 still needs to be improved to be equally reliable for predictions in ungauged basins as for gauged rivers. The model set-up procedure according to the PUB recommendations brought insights on where the single model structure did not perform well. Based on this, future model improvements will mainly focus on the western Himalayas and arid regions by refining the hypothesis of snow/glacier processes and the evapotranspiration algorithm.

29

performed at the SMHI Hydrological Research unit, where much work is done jointly. We would especially like to acknowledge contributions from David Gustafsson, Göran Lindström, Jafet Andersson, Kean Foster, Kristina Isberg, and Jörgen Rosberg for assistance with background material for this study. The authors would finally like to express their sincere gratitude to two anonymous reviewers for their constructive comments. Their detailed suggestions have resulted in an improved manuscript. The HYPE model code is open source and can be retrieved with manuals at http://hype.sourceforge.net/. Time-series and maps from the India-HYPE model (including climate change impact studies) are available for inspection at http://hypeweb.smhi.se. The work contributes to the decadal research initiative "Panta Rhei" by the International Association of Hydrological Sciences (IAHS) under Target 2 "Estimation and Prediction" and its two working groups on Large Samples and Multiple ungauged basins, respectively.

APPENDIX A: DEFINITION OF PEFORMANCE METRICS AND FLOW SIGNATURES

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

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

where r is the linear cross-correlation coefficient between observed and modelled records, α is a measure of variability in the data values (equal to the standard deviation of modelled over the standard deviation 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. In addition, we transform the three KGE components to results into a consistent range of possible values. Consequently we consider:

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

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

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

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

Signature	Abbreviation	Reference
Mean annual specific runoff	Qm	(Viglione et al., 2013)
Normalised high flow	q05	(Viglione et al., 2013)

30

Normalised low flow	q95	(Viglione et al., 2013)
Normalised relatively low flow	q70	(Viglione et al., 2013)
Slope of flow duration curve	mFDC	(Viglione et al., 2013)
Range of Pardé coefficient	DPar	(Viglione et al., 2013)
Coefficient of variation	CV	(Donnelly et al., 2015)
Flashiness	Flash	(Donnelly et al., 2015)
Normalised peak distribution	PD	(Euser et al., 2013)
Rising limb density	RLD	(Euser et al., 2013)
Declining limb density	DLD	(Euser et al., 2013)
Long term mean discharge	Qdm	(Donnelly et al., 2015)

REFERENCES

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

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

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, 2015.

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

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

Arheimer, B., Dahné, J., 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(4), 315–329, doi:10.2166/nh.2012.010, 2012.

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

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

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

31

1701   Attri, S. D. and Tyagi, A.: Climate profile of India, in Government of India Ministry of Earth
1702   Sciences, p. 129, New Delhi., 2010.

1703   Bartholomé, E., Belward, A. S., Achard, F., Bartalev, S., Carmona Moreno, C., Eva, H., Fritz, S.,
1704   Grégoire, J.-M., Mayaux, P. and Stibig, H.-J.: GLC 2000 Global Land Cover mapping for the year
1705   2000, European Commission, DG Joint Research Centre, EUR 20524 EN, Ispra., 2002.

1706   Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H.: Runoff prediction in
1707   ungauged basins. Synthesis across processes, places and scales, Cambridge University Press,
1708   Cambridge, UK., 2013.

1709   Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J.: Classification and Regression Trees,
1710   CRC Press, Wadsworth, Belmont, CA., 1984.

1711   Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G. and Wheater, H.: Integrating different types of
1712   information into hydrological model parameter estimation: Application to ungauged catchments and
1713   land use scenario analysis, Water Resour. Res., 48(6), W06519, doi:10.1029/2011WR011207, 2012.

1714   Bulygina, N., McIntyre, N. and Wheater, H. S.: Conditioning rainfall-runoff model parameters for
1715   ungauged catchments and land management impacts analysis, Hydrol. Earth Syst. Sci., 13, 893–904,
1716   2009.

1717   Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash
1718   testing hydrological models in contrasted climate conditions: an experiment on 216 Australian
1719   catchments, Water Resour. Res., 48, W05552, doi:10.1029/2011WR011721, 2012.

1720   Donnelly, C., Andersson, J. C. M. and Arheimer, B.: Using flow signatures and catchment similarities
1721   to evaluate the E-HYPE multi-basin model across Europe, Hydrol. Sci. J.,
1722   doi:10.1080/02626667.2015.1027710, 2015.

1723   Donnelly, C., Rosberg, J. and Isberg, K.: A validation of river routing networks for catchment
1724   modelling from small to large scales, Hydrology Research, special issue, Large-Scale Hydrology.
1725   doi:10.2166/nh.2012.341, 2012.

1726   Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H. H. G.: A
1727   framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst.
1728   Sci., 17(5), 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.

1729   Falkenmark, M. and Chapman, T.: Comparative hydrology: An ecological approach to land and water
1730   resources, UNESCO, Paris, France., 1989.

1731   Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through
1732   stepwise model concept improvement, Water Resour. Res., 44(1), 1–13, doi:10.1029/2006WR005563,
1733   2008.

1734   Finger, D., Pellicciotti, F., Konz, M., Rimkus, S. and Burlando, P.: The value of glacier mass balance,
1735   satellite snow cover images, and hourly discharge for improving the performance of a physically
1736   based distributed hydrological model, Water Resour. Res., 47, W07519, doi:10.1029/2010WR009824,
1737   2011.

1738   Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S. and Savenije, H. H. G.: Testing the realism of a
1739   topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol.
1740   Earth Syst. Sci., 18(5), 1895–1915, doi:10.5194/hess-18-1895-2014, 2014.

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: , Hydrol. Res., 44(5), 917–925,

Ilias Pechlivanidis 7/27/2015 17:41
Deleted: , 2013

1743   Good, P. I.: Resampling methods: A practical guide to data analysis, 3rd editio., Birkhäuser, Boston.,
1744   2005.

1745   Gosain, A., Rao, S. and Arora, A.: Climate change impact assessment of water resources of India,
1746   Curr. Sci., 101(3), 356–371, 2011.

1747   Gosain, A., Rao, S. and Basuray, D.: Climate change impact assessment on hydrology of Indian river
1748   basins, Curr. Sci., 90(3), 346–353 [online] Available from:
1749   http://www.iisc.ernet.in/currsci/feb102006/346.pdf (Accessed 28 January 2014), 2006.

1750   Gupta, H. V, Wagener, T. and Liu, Y.: Reconciling theory with observations : elements of a
1751   diagnostic approach to model evaluation, Hydrol. Process., 22, 3802– 3813, doi:10.1002/hyp.6989,
1752   2008.

1753   Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error
1754   and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1-
1755   2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

1756   Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.:
1757   Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18(2), 463–
1758   477, doi:10.5194/hess-18-463-2014, 2014.

1759   Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G.
1760   and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert
1761   knowledge, and process complexity, Water Resour. Res., 50, 7445–7469,
1762   doi:10.1002/2014WR015484, 2014.

1763   Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W.,
1764   Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V.,
1765   Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S.,
1766   Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. and Cudennec, C.: A decade of Predictions in
1767   Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255,
1768   doi:10.1080/02626667.2013.803183, 2013.

1769   Immerzeel, W. W., Droogers, P., de Jong, S. M. and Bierkens, M. F. P.: Large-scale monitoring of
1770   snow cover and runoff simulation in Himalayan river basins using remote sensing, Remote Sens.
1771   Environ., 113(1), 40–49, doi:10.1016/j.rse.2008.08.010, 2009.

1772   Johnston, R. and Smakhtin, V.: Hydrological modeling of large river basins: how much is enough?,
1773   Water Resour. Manag., 28(10), 2695–2730, doi:10.1007/s11269-014-0637-8, 2014.

1774   Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y. and Westerberg, I. K.: Disinformative data in large-
1775   scale hydrological modelling, Hydrol. Earth Syst. Sci., 17(7), 2845–2857, doi:10.5194/hess-17-2845-
1776   2013, 2013.

1777   Kumar, R., Livneh, B. and Samaniego, L.: Toward computationally efficient large-scale hydrologic
1778   predictions with a multiscale regionalization scheme, Water Resour. Res., 49(9), 5700–5714,
1779   doi:10.1002/wrcr.20431, 2013.

1780   Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and
1781   wetlands, J. Hydrol., 296(1-4), 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

1782 Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan,
1783 M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N. and Wisser, D.:
1784 Global Reservoir and Dam (GRanD) database - Technical documentation version 1.1, Bonn., 2011.

1785 Lehner, B., Verdin, K. and Jarvis, A.: New global hydrography derived from spaceborne elevation
1786 data, Eos, Trans. AGU, 89(10), 93–94, doi:10.1029/2008EO100001, 2008.

1787 Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B.: Development and testing of the
1788 HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales,
1789 Hydrol. Res., 41(3-4), 295–319, doi:10.2166/nh.2010.007, 2010.

1790 Mall, R. K., Singh, R., Gupta, A., Srinivasan, G. and Rathore, L. S.: Impact of Climate Change on
1791 Indian Agriculture: A Review, Clim. Change, 78(2-4), 445–478, doi:10.1007/s10584-005-9042-x,
1792 2006.

1793 McDonnell, J. J., McGuire, K., Aggarwal, P., Beven, K. J., Biondi, D., Destouni, G., Dunn, S., James,
1794 A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe,
1795 A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P.,
1796 Weiler, M., Western, A., Wörman, A. and Wrede, S.: How old is streamwater? Open questions in
1797 catchment transit time conceptualization, modelling and analysis, Hydrol. Process., 24(12), 1745–
1798 1754, doi:10.1002/hyp.7796, 2010.

1799 McMillan, H., Tetzlaff, D., Clark, M. and Soulsby, C.: Do time-variable tracers aid the evaluation of
1800 hydrological model structure? A multimodel approach, Water Resour. Res., 48(5), W05501,
1801 doi:10.1029/2011WR011688, 2012.

1802 Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis,
1803 D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey,
1804 M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P.,
1805 Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M.,
1806 Viglione, A., McMillan, H., Characklis, G., Pang, Z. and Belyaev, V.: "Panta Rhei—Everything
1807 Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrol. Sci. J.,
1808 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

1809 Mu, Q., Heinsch, F. A., Zhao, M. and Running, S. W.: Development of a global evapotranspiration
1810 algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536,
1811 doi:10.1016/j.rse.2006.07.007, 2007.

1812 Mu, Q., Zhao, M. and Running, S. W.: Improvements to a MODIS global terrestrial
1813 evapotranspiration algorithm, Remote Sens. Environ., 115(8), 1781–1800,
1814 doi:10.1016/j.rse.2011.02.019, 2011.

1815 Nachtergaele, F., van Velthuizen, H., Verelst, L. and Wiberg, D.: Harmonized world soil database
1816 version 1.2, FAO, Rome and IIASA, Laxenburg, Austria., 2012.

1817 Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, J. Hydrol., 10,
1818 282–290, 1970.

1819 Nazemi, A. and Wheater, H. S.: On inclusion of water resource management in Earth System models
1820 – Part 1: Problem definition and representation of water demand, Hydrol. Earth Syst. Sci., 19, 33–61,
1821 doi:10.5194/hess-19-33-2015, 2015a.

1822 Nazemi, A. and Wheater, H. S.: On inclusion of water resource management in Earth System models
1823 – Part 2: Representation of water supply and allocation and opportunities for improved modeling,
1824 Hydrol. Earth Syst. Sci., 19, 63–90, doi:10.5194/hess-19-63-2015, 2015b.

1825 Nepal, S., Krause, P., Flügel, W.-A., Fink, M. and Fischer, C.: Understanding the hydrological system
1826 dynamics of a glaciated alpine catchment in the Himalayan region using the J2000 hydrological
1827 model, Hydrol. Process., 28(3), 1329–1344, doi:10.1002/hyp.9627, 2014.

1828 Patil, J. P., Sarangi, A., Singh, A. K. and Ahmad, T.: Evaluation of modified CN methods for
1829 watershed runoff estimation using a GIS-based interface, Biosyst. Eng., 100(1), 137–146,
1830 doi:10.1016/j.biosystemseng.2008.02.001, 2008.

1831 Pechlivanidis, I. G., Jackson, B., McIntyre, N. and Wheater, H. S.: Catchment scale hydrological
1832 modelling: A review of model types, calibration approaches and uncertainty analysis methods in the
1833 context of recent developments in technology and applications, Glob. NEST J., 13(3), 193–214, 2011.

1834 Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proc. R. Soc. London Ser. A
1835 Math. Phys. Sci., 193, 120–145, 1948.

1836 Raje, D., Priya, P. and Krishnan, R.: Macroscale hydrological modelling approach for study of large
1837 scale hydrologic impacts under climate change in Indian river basins, Hydrol. Process., 28(4), 1874–
1838 1889, doi:10.1002/hyp.9731, 2013.

1839 Refsgaard, J.C., Storm, B. and Clausen, T.: Système Hydrologique Europeén (SHE): review and
1840 perspectives after 30 years development in distributed physically-based hydrological modelling,
1841 Hydrology Research, 41(5), 355-377. 2010.

1842 Samaniego, L., Kumar, R. and Jackisch, C.: Predictions in a data-sparse region using a regionalized
1843 grid-based hydrologic model driven by remotely sensed data, Hydrol. Res., 42(5), 338–355,
1844 doi:10.2166/nh.2011.156, 2011.

1845 Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A. and Carrillo, G.: Catchment classification:
1846 empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol.
1847 Earth Syst. Sci., 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.

1848 Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K. and Feick, S.: Development and
1849 validation of the global map of irrigation areas, Hydrol. Earth Syst. Sci., 9, 535–547,
1850 doi:10.5194/hessd-2-1299-2005, 2005.

1851 Singh, V. (Ed.): Computer Models of Watershed Hydrology. Water Resources Publications, Littleton,
1852 Colorado, 1995.

1853 Singh, P., Arora, M. and Goel, N. K.: Effect of climate change on runoff of a glacierized Himalayan
1854 basin, Hydrol. Process., 20(9), 1979–1992, doi:10.1002/hyp.5991, 2006.

1855 Singh, P. and Bengtsson, L.: Hydrological sensitivity of a large Himalayan basin to climate change,
1856 Hydrol. Process., 18(13), 2363–2385, doi:10.1002/hyp.1468, 2004.

1857 Singh, P. and Jain, S. K.: Modelling of streamflow and its components for a large Himalayan basin
1858 with predominant snowmelt yields, Hydrol. Sci. J., 48(2), 257–276, doi:10.1623/hysj.48.2.257.44693,
1859 2003.

35

1860   Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X.,
1861   McDonnell, J. J., Mendiondo, E. M., O'Connell, P., Oki, T., Pomeroy, J. W., Schertzer, D.,
1862   Uhlenbrook, S. and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012:
1863   Shaping an exciting future for the hydrological sciences, Hydrol. Sci. J., 48(6), 857–880, 2003.

1864   Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. and Lindström, G.: Water and nutrient
1865   predictions in ungauged basins: set-up and evaluation of a model at the national scale, Hydrol. Sci. J.,
1866   57(2), 229–247, doi:10.1080/02626667.2011.637497, 2012.

1867   Takeuchi, K., Blöschl, G., Savenije, H. H. G., Schaake, J., Sivapalan, M., Viglione, A., Wagener, T.
1868   and Young, G.: Recommendations, in Runoff Predictions in Ungauged Basins - Synthesis across
1869   processes, places and scales, edited by G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. H.
1870   G. Savenije, pp. 384–387, Cambridge University Press, Cambridge, UK., 2013.

1871   Ter_Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm Differential
1872   Evolution: easy Bayesian computing for real parameter spaces, Stat. Comput., 16(3), 239–249,
1873   doi:10.1007/s11222-006-8769-1, 2006.

1874   Thenkabail, P., Biradar, C., Noojipady, P., Dheeravath, V., Li, Y. J., Velpuri, M., Reddy, G. P. O.,
1875   Cai, X. L., Gumma, M., Turral, H., Vithanage, J., Schull, M. and Dutta, R.: A Global Irrigated Area
1876   Map (GIAM) Using Remote Sensing at the End of the Last Millennium, Int. J. Remote Sens., 30(14),
1877   3679–3733, 2009.

1878   Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M. and Blöschl, G.:
1879   Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria,
1880   Hydrol. Earth Syst. Sci., 17(6), 2263–2279, doi:10.5194/hess-17-2263-2013, 2013.

1881   Widén-Nilsson, E., Halldin, S. and Xu, C.: Global water-balance modelling with WASMOD-M:
1882   Parameter estimation and regionalisation, J. Hydrol., 340(1–2), 105–118,
1883   doi:10.1016/j.jhydrol.2007.04.002, 2007.

1884   Yasutomi, N., Hamada, A. and Yatagai, A.: Development of a long-term daily gridded temperature
1885   dataset and its application to rain/snow discrimination of daily precipitation, Glob. Environ. Res., 3,
1886   165–172, 2011.

1887   Yatagai, A., Arakawa, O. and Kamiguchi, K.: A 44-year daily gridded precipitation dataset for Asia
1888   based on a dense network of rain gauges, Sola, 5, 137–140, doi:10.2151/sola.2009–035, 2009.

1889   Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. and Kitoh, A.: APHRODITE:
1890   Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain
1891   gauges, Bull. Am. Meteorol. Soc., 93(9), 1401–1415, doi:10.1175/BAMS-D-11-00122.1, 2012.

1892

36

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:

> To what extent are the PUB recommendations for catchment scale also relevant for multi-basin modelling at the large scale?
>
> Which obstacles can be identified when using the PUB recommendations at the large scale?
>
> How can a processed-based modeller overcome these obstacles by using complementing data and evaluation methods?

To test the PUB recommendations we first had to

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

from the shelf, but choosing a model that includes

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

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

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

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

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

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

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

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

Page 5: [2] Deleted   **Ilias Pechlivanidis**   **7/27/15 17:41**

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

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

using the following methods to understand the hydrology and its dominating processes

| Page 5: [3] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

using the following methods to understand the hydrology and its dominating processes

| Page 5: [3] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

| Page 5: [3] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

| Page 6: [4] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

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 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., accepted). 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.

| Page 6: [5] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

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.

| Page 6: [6] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

such as Hydrological Response Units (HRUs) and other types of

| Page 6: [7] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

, which are used as input data to the model

| Page 6: [8] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

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

) and redistribute this area to the remaining classes, (iii) separate 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.

procedure. This can be done by using different cluster analysis methods, and after clustering, we

a large region in a multi-basin manner (Fig. 1). When using large datasets it is important to repeat

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.

represent

represent

represent

represent

represent

represent

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [13] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

represent

| Page 7: [14] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

, to check their coherence with

| Page 7: [14] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

, to check their coherence with

| Page 7: [14] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

, to check their coherence with

| Page 7: [15] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

covering a large area with high spatial resolution are normally

| Page 7: [15] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

covering a large area with high spatial resolution are normally

| Page 7: [15] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

covering a large area with high spatial resolution are normally

| Page 7: [16] Deleted | Ilias Pechlivanidis | 7/27/15 17:41 |

. Andersson et al. (2014) showed that uncertainties in the input data and/or routing are often more important than parameter uncertainty when

. Andersson et al. (2014) showed that uncertainties in the input data and/or routing are often more important than parameter uncertainty when

The further developed and modified recommendations (section 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 6 000 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 geographical domain and methods used for modelling, regionalisation and evaluation during the exercise are described more in detail below.

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.
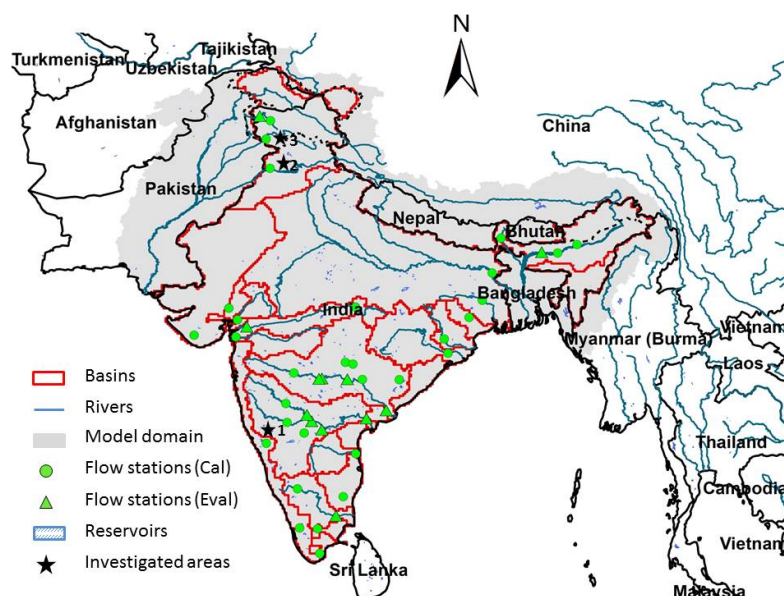
To draw more robust conclusions on the identification of the appropriate *cevp* parameter value, we investigate further the parameter inter-dependence between land use types (not shown). Trends in the HYPE model – MODIS PET agreement could be identified for some of the land use types. A weak but statistically 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.

Page 25: [20] Deleted	Ilias Pechlivanidis	7/27/15 17:41

DISCUSSION

1.1.

Page 29: [21] Deleted	Ilias Pechlivanidis	7/27/15 17:41

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 modelling 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 section 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 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.

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

Page 29: [22] Deleted	Ilias Pechlivanidis	7/27/15 17:41

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.

Page 30: [23] Deleted	Ilias Pechlivanidis	7/27/15 17:41

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

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

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 (-)

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

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 (-)

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

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).