

## ***Interactive comment on “Calibration approaches for distributed hydrologic models using high performance computing: implication for streamflow projections under climate change” by S. Wi et al.***

**S. Wi et al.**

sungwookwi@gmail.com

Received and published: 6 January 2015

We would like to thank the editor, Dr. Efrat Morin, and the two referees, Dr. Andreas Efstratiadis and an anonymous reviewer, for their constructive suggestions on our manuscript. The comments greatly improved our manuscript. Please find detailed answers to all the comments of the reviewers. We also enclose a word file in which all the changes are highlighted to be easily tracked.

Response to Dr. Efstratiadis

C5911

### General comments:

The paper is well-structured, well-written and easy to follow. I am very happy with the experience gained from this exhaustive modelling experiment, which reveals the superiority of pooled calibration (i.e. estimation of model parameters on the basis of flow data at multiple sites across the basin) over the stepwise strategy, and also reveals the advantages of semi-distributed over (non-parsimonious) fully-distributed schematizations, by means of improved predictive capacity and reduced parameter uncertainty. These outcomes are in agreement with the “holistic” approach proposed by Nalbantis et al. (2011) and other researchers of the same philosophy, which recognize that: (a) model complexity should be as high as allowed by the available information, and (b) all available information – even a single measurement – is valuable and should be accounted for in calibration. Unlikely, this is not the dominant philosophy among modellers, thus I believe that this paper will be a significant contribution to both hydrological science and practice. By reading this very good paper, I detected some issues to be clarified or further discussed, thus my recommendation is for a minor revision. In the following list, please find my specific comments as well as some technical corrections, to be addressed in your revision.

We are very pleased that you enjoyed reading our manuscript and found it useful. We appreciate your encouraging and constructive comments. Below, please find our responses to all your specific comments and technical corrections.

### Specific comments:

1. p. 10275, lines 15-17: “Importantly, distributed hydrologic models can evaluate hydrological response at interior ungaged sites, a benefit not afforded by conceptual, lumped models.” Please, remove “conceptual”, which refers to the modelling approach behind the formulation of the governing equations and not the spatial discretization of the model domain. Apart from lumped models, semi-distributed schemes are also by definition conceptual. Quoting Beven (1989), even a fully-distributed physically-based

C5912

model can be regarded as conceptual, at the grid scale.

We removed “conceptual” in that sentence.

2. p. 10275, line 27 to p. 10276, line 2: “Parameters can be discretized across the watershed in several ways: uniquely for each grid cell (fully distributed), based on hydrologic response units (semi-distributed), or in the simplest case, a single parameter set for all model grid cells (lumped).” In hydrologic models, hydrologic response units (HRUs) are mainly used for distributed and less often for semi-distributed schemes (e.g. Efstratiadis et al., 2008). The concept of HRUs was introduced by Flugel (1995) to characterize homogeneous areas with similar geomorphologic and hydrodynamic properties. The one-to-one correspondence of HRUs and sub-basins could be considered a specific case, which is however not consistent with the rationale of HRUs, as far as sub-basins have arbitrary boundaries that do not necessarily ensure homogenous characteristics.

The sentence has been rewritten with new references as follows. “Parameters can be discretized across the watershed in several ways (Flugel, 1995; Efstratiadis et al., 2008; Khakbaz, et al., 2012): uniquely for each grid cell or hydrologic response unit (fully distributed), based on sub-basins whose boundaries do not necessarily ensure homogenous characteristics (semi-distributed), or in the simplest case, a single parameter set for all model grid cells (lumped).”

3. p. 10276, lines 26-30: “Many studies have reported that distributed models calibrated at the basin outlet are less accurate at interior locations (Anderson et al., 2001; Cao et al., 2006; Wang et al., 2012), but the extent of the error and uncertainty is unknown due to the computational expense needed to explore this issue.” To my opinion (and my experience), the accuracy of predictions of runoff at interior points mainly depends on the local characteristic of the basin. In the case of strongly heterogeneous basins, it is far from reasonable to make estimations based on the lumped information obtained at the basin outlet. On the other hand, if the key properties of the basin that

C5913

influence runoff generation (e.g., permeability, vegetation, slope) do not vary significantly, such estimations could be quite reliable. However, the latter is not the rule.

We agree with this point. The sentence has been rewritten as follows. “In the case of significant spatial variability in the basin properties that influence runoff generation (e.g., permeability, vegetation, slope, etc.), accurate runoff predictions are unlikely at interior locations based only on the lumped information obtained at the basin outlet (Anderson et al., 2001; Cao, et al., 2006; Breuer et al., 2009; Lerat et al., 2012; Smith et al., 2012; Wang, et al., 2012). The extent of this error and uncertainty is not well understood for heterogeneous basins due to the computational expense required to explore this issue.”

4. p. 10277, lines 1-2: “. . . for an alternative climate, which is required in climate change impact studies”. My impression is that climate change studies over broader areas refer to systematic deviations from the average climatic conditions, and not to “alternative climates”.

Instead of using “alternative climate”, we used “possible future climate conditions”. Also, we changed the same term in the later part of the manuscript.

5. p. 10277, lines 22-24: “Water resources from the basin are shared by Afghanistan and Pakistan and serve as a water supply source for more than 20 million people.” How significant are water abstractions in this basin? Are they accounted for in the modelling scheme? Are there any important regulations that modify the flow regime across the basin?

We completely agree with reviewer’s concern about human interfere. The Kabul River has the largest flow of all of Afghanistan’s rivers, but it can irrigate only a limited area because there is little land suitable for agriculture in the Afghan part of the basin (Ahmad and Wasiq, 2004) – for the most part, the river flows through mountainous or rocky areas. According to World Bank, (2010), about 2,927 km<sup>2</sup> (4.3% of the total basin area) is agricultural land and the average annual flow of the Kabul River is ap-

C5914

proximately 24,000 million cubic meters (MCM). Irrigation is a large water demand since the annual water demand estimate for the agricultural use is about 2,000 MCM, or about 8.3% of the total annual flow. In our hydrologic modelling process, the water consumed by irrigated croplands is implicitly accounted for by the evapotranspiration module. We note that the degree of irrigation impact during the time frame used for calibration (1960-1981) is likely much smaller than the current level.

The Naglu dam, which is located in the western part the Kabul River basin (upstream of the Daronta streamflow gage), forms the largest and most important storage among dams in the basin (World Bank, 2010). The live storage of the Naglu dam is 379 MCM. We expect that using monthly data for calibration somewhat reduces the bias from human interference, particularly the daily operations of Naglu dam. Nevertheless, the calibration results for the gage below this dam (Daronta), and to a lesser extent the basin outlet (Dakah), should be approached with caution. Given that a majority of the gages examined in this study are on an underdeveloped branch of the Kabul River, issues of human interference on calibration are somewhat mitigated. We also note that the poor performance at Daronta is likely due in part to the impacts of water abstraction and the operation of Naglu dam.

This information has been provided accordingly in the text. "Similar to most other hydrological models (Efstratidis et al., 2008), HYMOD\_DS is not designed to model water abstractions for agricultural lands and dam operations within the basin. According to World Bank (2010), water demand for agricultural use is about 2,000 MCM (million cubic meters), or about 8.3% of the total annual flow. The Naglu dam (Figure 1) upstream of the Daronta streamflow gage forms the largest and most important reservoir in the basin, with an active storage of 379 MCM. In our hydrologic modelling process, the water consumed by irrigated croplands is implicitly accounted for by the evapotranspiration module. We note that the degree of irrigation impact during the time frame used for calibration (1960-1981) is likely much smaller than the current level. We also expect that using monthly data for calibration somewhat reduces the bias from human

C5915

interference, particularly the daily operations of Naglu dam. Nevertheless, the calibration results for the gage below this dam (Daronta), and to a lesser extent the basin outlet (Dakah), should be approached with caution. Given that a majority of the gages examined in this study are on an underdeveloped branch of the Kabul River, issues of human interference on calibration are somewhat mitigated."

6. Section 2 (Study area): Here you should add information about the flow stations and the available data, and also provide synoptic statistical information about the hydrological characteristics of the basin, e.g. mean annual flow at the seven stations of interest, mean precipitation over the sub-basins, etc. (you can add these data to Table 1). It would also be useful to refer to the physiographic properties of the basin and the dominant runoff mechanisms, which are essential to interpret the model results and plausibility of the optimized parameter values. It is also essential to explain to which extend is this basin heterogeneous, thus justifying the implementation of each parameterization approach and better explain the model results.

Table 1 has been updated with basin climate information (mean annual precipitation, mean temperature, and flow) and geographic properties (drainage area, glacier area, mean elevation). This additional information is also discussed in the section describing the study area.

"The streamflow regime of the Kabul River can be classified as glacial with maximum streamflow in June or July and minimum streamflow during the winter season. Approximately 70% of annual precipitation (475mm) falls during the winter season (November to April). While the dominant source of streamflow in winter is baseflow and winter rainfall, glaciers and snow cover are the most important long-term forms of water storage and, hence, the main source of runoff during the ablation period for the basin (Shakir et al., 2010). In total 2.9% (1954km<sup>2</sup>) of the basin is glacierized based on the Randolph Glacier Inventory version 3.2 (Pfeffer, et al., 2014). The melt water from glaciers and snow produce the majority (75%) of the total streamflow (Hewitt, et al., 1989). Table 1 provides the climates and geophysical properties of each sub-watershed delineated by

C5916

the stations located inside the Kabul Basin (Figure 1). Two different climate patterns are distinguishable across the sub-basins. The sub-basins on the Kunar River tributary (Kama, Asmar, Chitral, Gawardesh, and Chaghasarai) receive moderate annual precipitation and are highly affected by snow and glacier covers. All of these sub-basins have high ratios of mean annual flow to mean annual precipitation, with the ratios for the Kama, Asmar, Chitral, and Chaghasarai sub-basins larger than 1. Conversely, the Daronta sub-basin contains only minimal glacial cover, and is relatively dry. Daronta is also much less productive, with annual streamflow far below the other sub-basins with an average of only 165 mm/year.”

7. p. 10279, lines 4-6: “However, in this particular study daily hydrologic model simulations can only be compared against available monthly streamflow records”. It is not clear whether monthly streamflows are averaged values of daily (or hourly) observations or instantaneous values, gather e.g. from direct flow measurements. Such clarification is very important.

Unfortunately, the only observations that are available for public use are monthly. There is a report (Olson and Williams-Sether, 2010) clarifying that each monthly streamflow is the mean of the daily values for the month, and monthly values are calculated from daily values for all complete months of record. However, the daily values are not made available because there are political issues surrounding the trans-boundary use of the river’s waters and potential projects planned on the river.

We have added the following details in the manuscript to clarify the immediate question regarding the data. “Streamflow data were not collected in Afghanistan after September 1980 until recently because streamgaging was discontinued soon after the Soviet invasion of Afghanistan in 1979 (Olson and Williams-Sether, 2010). Though measurements were taken at a daily time step, data are only made available for public use at monthly aggregated levels, calculated using the mean of the daily values.”

8. p. 10279, lines 18-20: “No matter the parameterization scheme, the model structure

C5917

follows the climate input grids, i.e. the hydrological water cycle within each grid cell is modelled separately.” In the revised paper, I suggest also employing the simplest of model configurations, assuming a lumped structure for both model inputs and parameters (i.e. using the averaged precipitation over the basin). This classical lumped approach considering 15 (or less) parameters would provide, in theory, the optimal results at the basin outlet with minimal computational burden, to be considered as “baseline scenario”.

We understand the reviewer’s suggestion, and initially considered this ourselves. However, we wanted the comparisons in this paper to isolate the effects of calibration uncertainty rather than address the structural uncertainties surrounding the model grid distribution (or lack thereof). Also, since a large focus of our study is on ungaged, interior point streamflow estimation, a lumped model structure would not really be appropriate (unless there was some scaling from flow estimates at the outlet of the basin to the interior points). With that said, we do agree with the reviewer that this issue should at least be addressed. Therefore, we now include a preliminary test of the basin outlet model from the lumped HYMOD without the gridded structure (13 parameters and basin-averaged climate inputs). The 2 parameters associated with the river routing models are dropped due to its lumped structure. We have added a description regarding a preliminary performance comparison between this model and its analogue with a gridded structure. Since the distributed model outperformed, we used this as a justification to set our “baseline” model as having a distributed structure. We decided that a figure is necessary for this additional part and a new figure has been provided as another supplementary material (Figure S3). We summarize these details in a new paragraph in the manuscript.

New paragraph: “We note that a lumped model structure (i.e., no gridded or sub-unit structure) has often been considered as a baseline model formulation in the assessment of distributed modelling frameworks (e.g., see Simith et al., 2013). However, the focus of our study is on ungaged interior site streamflow estimation, making this

C5918

formation somewhat inappropriate. Further, preliminary tests comparing streamflow simulations at the basin outlet (Dakah) between a gridded and basin-averaged structure, both with a lumped parameter formation, support the use of the distributed grid structure (Figure S3)"

New figure:

Figure S3. (a) Basin outlet (Dakah) simulations of HYMOD and MYMOD\_DS (with the lumped parameterization) from 50 trials of calibration. The Box plots provide the performance evaluation on 50 simulations of both models for both calibration and validation periods. (b) Performances of the models at the interior points of the watershed are assessed.

9. p. 10279, lines 21-24: "The parameter complexity will vary depending on the calibration experiment being conducted, but for each experiment regardless of the parameterization, the optimization is implemented 50 times using the GA algorithm to explore parameter uncertainty." Parameter uncertainty is a combined effect of multiple causes, one of which is inefficient calibrations (i.e. calibrations trapped to local optima). Even the use of robust and sophisticated evolutionary algorithms cannot remedy this problem, especially when a large number of parameters are considered. However, there are also other sources of parameter uncertainty, associated with data errors, unknown boundary conditions, etc. In this context, I propose avoiding the general term "parameter uncertainty" and focus to "calibration uncertainty", which is very well represented in your work, by implementing 50 independent runs for each optimization problem.

Throughout the manuscript, we replaced the term "parameter uncertainty" with "calibration uncertainty".

10. Section 3.1 (Multisite calibration): There are some important issues that are mentioned in next parts of the document, yet they should be also highlighted in this section. In order to better follow the modelling experiment, is essential to explain the sequence of sub-catchments, which strongly affects the outcomes of stepwise calibration (thus I

C5919

propose moving Fig. S1 from the supplement to the main text). Another missing issue is the lack of overlapping data periods among most of stations, which is a bad coincidence, since this weakens the multisite calibration approach: in fact, you do not have simultaneous information on the basin responses, which would allow account for the heterogeneity of the associated hydrological processes.

First, we have changed the paper structure to address some of these issues. Now, the Section "Data and Models" is placed ahead of Section "Methods" so that the reader sees some pertinent information regarding the basin before being introduced to the details of the modeling experiments. In the methods section, we moved Fig. S1 (now Figure 5) to the main text in section 4.1 (multisite calibration) to make sure the reader understands the sequence of sub-basins in the stepwise calibration. Also, at the end of this section, we now include a brief discussion of the second point made by the reviewer:

"It is important to note that the evaluation of these multisite calibration strategies is somewhat weakened because of the lack of overlapping data periods among most of the stations (Figure 2). This drawback prevents the calibration methods from accounting for simultaneous information from different tributaries, which, if available, would better enable the calibration methods to account for heterogeneity of hydrological processes across the sub-basins."

11. p. 10281, lines 23-27: ".. the lumped version of the HYMOD\_DS contains a single, 15-member parameter set applied to all model grid cells. The semi-distributed conceptualization of HYMOD\_DS contains a single parameter set for each sub-basin, totaling 75 parameters. In the distributed parameterization ... the number of parameters requiring calibration reaches 2400." Here it is worthy reminding that for the transformation of rainfall to hydrograph at the basin outlet, only 5 to 6 parameters can be identified on the basis of a single observation set (cf. Wagener et al., 2001). Under this premise, the number of parameters for the lumped scheme is realistic, taking into account that snow, glacier and flow routing processes are also modelled. For the semi-distributed

C5920

approach, the number of parameters remains realistic, since external information is increased by accounting for interior flow data in calibrations. However, the distributed approach, with 2400 parameters to be optimized, is far from acceptable, and any attempt to interpret the outcomes of calibration is unreasonable.

Thank you for pointing out this issue and the useful references. We expanded our discussion section with this issue.

"It is worth noting that for the transformation of rainfall to runoff, up to five or six parameters can be identified on the basis of a single hydrograph (Wagner et al., 2001). Under this premise, the number of the HYMOD\_DS parameters being calibrated in the semi-distributed approach remains realistic, but the fully distributed parameterization scheme likely causes poor identifiability of the parameters. Thus, pursuing a parsimonious configuration (e.g. optimization for a small portion of the parameters) with an effort to increase the amount of information (e.g. multivariable/multisite) is critical in the calibration of watershed system models (Gupta et al., 1998; Efstratiadis et al., 2008)."

12. p. 10284, lines 11-12: "Monthly streamflow observations for seven locations in the Kabul River basin (Fig. 1) were gathered between calendar years 1961–1980". The same equation with comment 6: why monthly flow data and how are these data extracted?

Please refer to the answer for the comment 7.

13. p. 10285, line 14-15: "The overall model structure of the HYMOD\_DS and its 15 parameters are described in Fig. 4 and Table 2 respectively." The feasible ranges that are employed for the model parameters are extremely large thus resulting to huge parameter uncertainty (at least, *a priori* uncertainty). For instance, the maximum soil moisture capacity ranges from 5 to 1500 mm. I would expect that an experienced hydrologist would propose much more narrow bounds, taking into account the physical interpretation of those parameters and the local characteristics of the specific study area. I strongly believe that a hydrological model is not a mathematical game, and

C5921

calibration is not a black-box exercise. In contrast, model parameters should always have some correspondence to the physical properties of the basin, which is yet not reflected in this work. In addition, a substantial reduction of feasible ranges would be beneficial for the calibration effort, which is tremendous (1000 parallel processors running for 7 days!).

Our main focus is to explore a variety of calibration strategies which becomes a computationally exhaustive task but can be implemented with the aid of parallel computing power. We noticed that there might be an advantage of having wide feasible parameter ranges; we can expect to avoid *a priori* errors that could be caused by inappropriately narrowing down the ranges. We decided to embrace the computational cost owing to the wide parameter ranges and then try to solve this issue with the high computation power available from the MGHPC.

Nonetheless, this is a very good point which is worth a further discussion. "We also note the important role of experienced hydrologists in designing a parsimonious hydrologic calibration (e.g. Boyle et al., 2000). In this study, the feasible ranges of the HYMOD\_DS parameters were kept wide (as is often done in automatic hydrologic calibrations) without consideration of the physical properties of the basin; the judgment of local hydrologic experts could help reduce the feasible ranges used during the calibration and thus contribute to a reduction of calibration uncertainty."

14. p. 10286, line 15: The Hamon method for PET estimations is not widely known. Please, provide one or two sentences with a very synoptic description of this method (rationale, input data). Is this method suitable for the climatic regime of the study area?

We provided more information on the Hamon method with an additional equation. Please refer to the following for the changes made in the text: "The potential evapotranspiration (PET) is derived based on the Hamon method (Hamon, 1961), in which daily PET in mm is computed as a function of daily mean temperature and hours of daylight:

C5922

where,  $L_d$  is the daylight hours per day,  $T$  is the daily mean air temperature ( $^{\circ}\text{C}$ ), and  $\text{Coeff}$  is a bias correction factor. The hours of daylight is calculated as a function of latitude and day of year based on the daylight length estimation model (CBM model) suggested by Forsythe et al. (1995)." Is this method suitable for the climatic regime of the study area? As explained, the Hamon is a temperature based method. Despite its simplicity relative to more input-detailed models, some studies identified the model as a method that produce satisfactory estimates of PET. Here's some examples. Vorosmarty et al. (1998) compared to 11 different PET models for a wide range of climatic conditions across the conterminous US and found that the Hamon model is comparable to more input-detailed models, such as the Shuttleworth-Wallance. In a study of 5 PET models for use with global water balance models Federer et al. (1996) found that estimates of PET from the Hamon model agreed with estimates from other models across a wide range of climates. From a comparison of six PET models, Lu et al. (2005) recommended the Hamon method for regional applications in the southeastern US based on the criteria of availability of input data and correlations with actual ET values.

15. p. 10290, lines 16-17: "High accuracy holds even under the Lump\_Outlet, which is somewhat surprising given the spatial heterogeneity of the basin." I do not agree that this is a surprising conclusion. The lumped configuration of HYMOD\_DS has 15 parameters, which are far from sufficient to represent hydrographs of any complexity.

We understand the point here. We have changed the wording accordingly to now read: "High accuracy holds even under the Lump\_Outlet, despite the spatial heterogeneity of the basin."

16. p. 10290, lines 25-27: ". . .the HYMOD\_DS significantly overestimated streamflow at Daronta and underestimated flow at three sites in the eastern part of the basin" This is a strong evidence of the heterogeneity of the basin. Please, provide some information on the properties of the basin (e.g. geology) that would justify these differences.

C5923

We have updated Table 1 with the information to support the heterogeneity of the basin and also include new information on the basin heterogeneity in the updated Figure 1. Please refer to the answer for the comment 6.

17. p. 10292, lines 7-9: "On the other hand, temperature clearly shows an upward trend for both radiative forcing scenarios. The average changes in annual temperature are  $+2.2^{\circ}\text{C}$  and  $+2.8^{\circ}\text{C}$  for RCP4.5 and RCP8.5, respectively". Which are the impacts of such difference in PET estimations?

Changes in temperature are important in the PET estimation. The Hamon PET calculation is a function of temperature and daylight length. Since the daylight hours is a time-invariant variable, temperature changes will be the only factor affecting PET changes under the warming conditions. For an example, we took the grid cell covering Kabul city to calculate Hamon PET values under historical condition,  $+2.2^{\circ}\text{C}$ , and  $+2.8^{\circ}\text{C}$ . The value of calibrated  $\text{Coeff}$  (bias correction parameter) is 1.007 for this grid and the result is shown in the figure attached below. The average annual PET calculations are 804mm, 916mm, and 940mm under historical condition,  $+2.2^{\circ}\text{C}$ , and  $+2.8^{\circ}\text{C}$ , respectively. The percent changes of PET relative to the historical PET in the warming conditions of  $+2.2^{\circ}\text{C}$  and  $+2.8^{\circ}\text{C}$  are approximately +14% and +18%, respectively.

We have included a brief addition to the line in question for clarification: "On the other hand, temperature clearly shows an upward trend for both radiative forcing scenarios. The average changes in annual temperature are  $+2.2^{\circ}\text{C}$  and  $+2.8^{\circ}\text{C}$  for RCP4.5 and RCP8.5, which, using the Hamon method, correspond to an increase in annual PET by approximately 100mm and 150mm, respectively."

18. p. 10292, lines 17-19: "For the historical time period, all calibration schemes match the observed climatology at Dakah well, but monthly streamflow is underestimated in most of months at Kama and Asmar under the basin outlet calibrations". If I understood well, you used as meteorological inputs the average projections of the 36 climate models during the period of observations. In that case, it is not clear whether

C5924

the underestimation of monthly flows is due to inappropriate representation of past precipitation and temperature data by climate models or due to inappropriate calibrations at the specific flow stations. For this reason, it is essential providing results on model bias (apart from NSE and KGE).

The 50 runs for the historical period have nothing to do with the climate model outputs. The observed climate is the only input that is used to derive the monthly streamflow estimates for the historical period with the 50 calibrated parameter sets. On the other hand, for the future period, 36 runs are related to a single parameter set because of 36 different GCM climate inputs. For each of the 50 parameter sets, we average out the uncertainty from the 36 future climate time series. In this way, the uncertainty ranges shown in Figure 10 both are composed of 50 different values, as described in the text “The whisker bars indicate the range across the 50 calibration trials; for the future scenarios, the whisker bars are derived by averaging over the 36 different climate projections for each of the 50 trials.”

We have rewritten the part for a better clarification on this as follows. “Figure 10 shows the monthly streamflow estimates for the historical period with the whisker bars indicating the uncertainty range across the 50 calibration trials. The monthly streamflow predictions are also provided for the 2050s under the RCP 4.5 and 8.5 scenarios. For the future scenarios, the whisker bars are derived by averaging over the 36 different climate projections for each of the 50 trials.”

19. p. 10293, lines 26-27: “Another clear point is that the uncertainty resulting from different climate change scenarios substantially outweighs that from parameter uncertainty.” This is of course a very important conclusion, and would deserve further discussion about the misuse of such scenarios as “deterministic” projections.

We discussed more about it in the section Discussion and Conclusion. “We evaluated the separate and joint influence of uncertainties in parameter estimation and future climate on projections of seasonal streamflow and 100-year daily flood across calibration

C5925

schemes and found that the uncertainty resulting from variations in projected climate between the CMIP5 GCMs substantially outweighs the calibration uncertainty. These results agree with other studies showing the dominance of GCM uncertainty in future hydrologic projections (Chen et al., 2011; Exbrayat et al., 2014). While the GCM-based simulations still have widespread use in assessing the impacts of climate change on water resources availability, the bounds of uncertainty resulting from an ensemble of GCMs cannot be well-defined because of the low credibility with which GCMs are able to produce timeseries of future climate (Koutsoyiannis et al., 2008). This issue hinders a straightforward appraisal of future water availability under climate change and has motivated other efforts; e.g. performance-based selection of GCMs (Perez et al., 2014).”

20. p. 10294, lines 10-14: “While no observed data is available against which to compare the results, an inter-model comparison is useful to distinguish the differences between the parameterization schemes.” Since observed flood data are missing, these comparisons are little safe. You may use them in the context of a theoretical calibration exercise, but definitely not for decision-making purposes.

Yes, we agree. We changed the sentence as follows. “Although the inter-model comparison is intended to be a useful addition that provides a distinction between the parameterization schemes in the pooled calibration approach, results from this analysis should be viewed in the context of a theoretical calibration exercise, not for decision-making purposes, because no observed daily streamflow is available against which to compare the estimated 100-year daily flood events.”

Technical corrections:

1. p. 10278, lines 23, 24: Please, change to read “Sutcliffe”.

Done.

2. p. 10292, line 18: Term “observed climatology” is unclear. Climatology is defined

C5926

as "the study of climate", while climate is defined as "as weather conditions averaged over a period of time" (<http://en.wikipedia.org/wiki/Climatology>).

We changed it to "average monthly streamflow estimates"

3. p. 10292, line 21: Similarly, term "historical streamflow climatology" is not valid. I suppose that you refer to average monthly flow data?

We changes it to "historical average monthly flow estimates" Throughout the manuscript we tried to correct the parts where the term "climatology" is used.

4. p. 10304, Table 1: Please, use common symbols for dates, e.g. YYYY/M or M/YYYY (not YYYY.M).

Now it is in "YYYY/M"

5. p. 10316, Fig. 10: The coefficient of variation of which quantity is represented in the graphs? (similar for Fig. 12).

For Fig. 10, it is for "Coefficient of variation of average season flow predictions" For Fig. 12, it is for "Coefficient of variation of 100-year flood estimations" We changed the y-axis label to reflect these clarifications in Fig. 10 and 12. Also, the captions for those figures are changed for more clear description of the figures.

Thank you.

#### References

While we were revising our manuscript, references listed below were added accordingly in the text.

Ahmad, M., and Wasiq, M.: Water resources development in Northern Afghanistan and its implications for Amu Darya Basin, The World Bank, Washington, D.C., 2004.

Boscarello, L., Ravazzani, G., and Mancini, M.: Catchment multisite discharge measurements for hydrological model calibration, *Procedia Environmental Sciences*, 19,

C5927

158-167, 2013.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resources research*, 36(12), 3663-3674, 2000.

Breuer, L., Huisman J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H. G., Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M., and Viney, N. R.: Assessing the impact of land use change on hydrology by ensemble modeling (LUChEM). I: Model intercomparison with current land use, *Advances in Water Resources*, 32, 129-146, 2009.

Brown, C., Ghile, Y., Laverty, M., and Li, K.: Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector, *Water Resources Research*, 48, W09537, 2012.

Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, *Water Resources Research*, 47, W12509, 2011.

DAWN: Pakistan, Afghanistan mull over power project on Kunar River, available at: <http://www.dawn.com/news/1038435>, last access: 2 January 2015, 2013.

Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., and Koutsoyiannis, D.: HYDROGEIOS: a semi-distributed GIS-based hydrological model for modified river basins, *Hydrol. Earth Syst. Sci.*, 12, 989-1006, doi:10.5194/hess-12-989-2008, 2008.

Exbrayat, J. F., Buytaert, W., Timbe, E., Windhorst, D., and Breuer, L.: Addressing sources of uncertainty in runoff projections for a data scarce catchment in the Ecuadorian Andes, *Climatic Change*, 125, 221-235, 2014.

FAO: Global Land Cover Share Database version 1.0, available at: <http://www.fao.org/geonetwork>, last access: 2 January 2015, 2013.

C5928

Federer C. A., Vorosmarty C., Fekete B.: Intercomparison of methods for calculating potential evaporation in regional and global water balance models. *Water Resour Res* 32:2315–2321, 1996.

Flugel, W. A.: Delineating Hydrological Response Units (HRU's) by GIS analysis for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Brol, Germany, *Hydrol. Processes*, 9, 423-436, 1995.

Forsythe, W. C., Rykiel Jr., E. J., Stahl, R. S., Wu, H., Schoolfield, R. M.: A model comparison for daylength as a function of latitude and day of year, *Ecological Modelling*, 80, 87-95, 1995.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resources Research*, 34, 751-763, 1998.

Immerzeel, W. W., Petersen, L., Ragettli, S., and Pellicciotti, F.: The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas, *Water Resour. Res.*, 50, 2212–2226, 2014.

IUCN: Towards Kabul Water Treaty: Managing Shared Water Resources – Policy Issues and Options, IUCN Pakistan, Karachi, 11 pp, 2010.

Koutsoyiannis, D., Efstratiadis, A., Mamassis, N., and Christofides, A.: On the credibility of climate predictions, *Hydrological Sciences Journal*, 53(4), 671-684, 2008.

Lerat, J., Andreassian V., Perrin, C., Vaze, J., Perraud J. M., Ribstein, P., and Loumagne C.: Do internal flow measurements improve the calibration of rainfall-runoff models?, *WATER RESOUR RES*, 48, W02511, 2012.

Lu J, Sun G, McNulty S. G., Amataya D. M.: A comparison of six potential evapotranspiration methods for regional use in the southeastern United States. *J Am Water Resour Assoc* 3:621–633, 2005.

C5929

Olson, S. A., and Williams-Sether, T.: Streamflow characteristics at streamgages in Northern Afghanistan and selected locations, U. S. Geological Survey, Reston, Virginia, 2010.

Perez, J., Menendez, M., Mendez, F. J., and Losada, I. J.: Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region, *Climate Dynamics*, 43, 2663-2680, 2014.

Shakir, A. S., Rehman, H., and Ehsan, S.: Climate change impact on river flows in Chitral watershed, *Pakistan Journal of Engineering and Applied Sciences*, 7, 12-23, 2010.

Steinschneider, S., Wi, S., and Brown, C.: The integrated effects of climate and hydrologic uncertainty on future flood risk assessments, *Hydrological Processes*, DOI: 10.1002/hyp.10409, 2014.

USDA-NRCS: Global Soil Regions Map and Global Soil Suborder Map Data from US Department of Agriculture, Natural Resource Conservation Service, 2007.

Vorosmarty C. J., Federer C. A., Schloss A. L.: Potential evaporation functions compared on US watersheds: possible implications for global-scale water balance and terrestrial ecosystem modeling. *J Hydrol* 207:147–169, 1998.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, *Hydrology and Earth System Sciences*, 5(1), 13-26, 2001.

World Bank: Afghanistan – Scoping strategic options for development of the Kabul River Basin: a multisectoral decision support system approach, World Bank, Washington, D. C., 2010.

Please also note the supplement to this comment:

<http://www.hydrol-earth-syst-sci-discuss.net/11/C5911/2015/hessd-11-C5911-2015>

C5930

supplement.zip

---

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, 11, 10273, 2014.

C5931