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

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 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 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; Simith 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^2 (4.3% of the total basin area) is agricultural land and the average annual flow of the Kabul River is approximately 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 (Efstratisdis 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 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^2) 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 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 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 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:

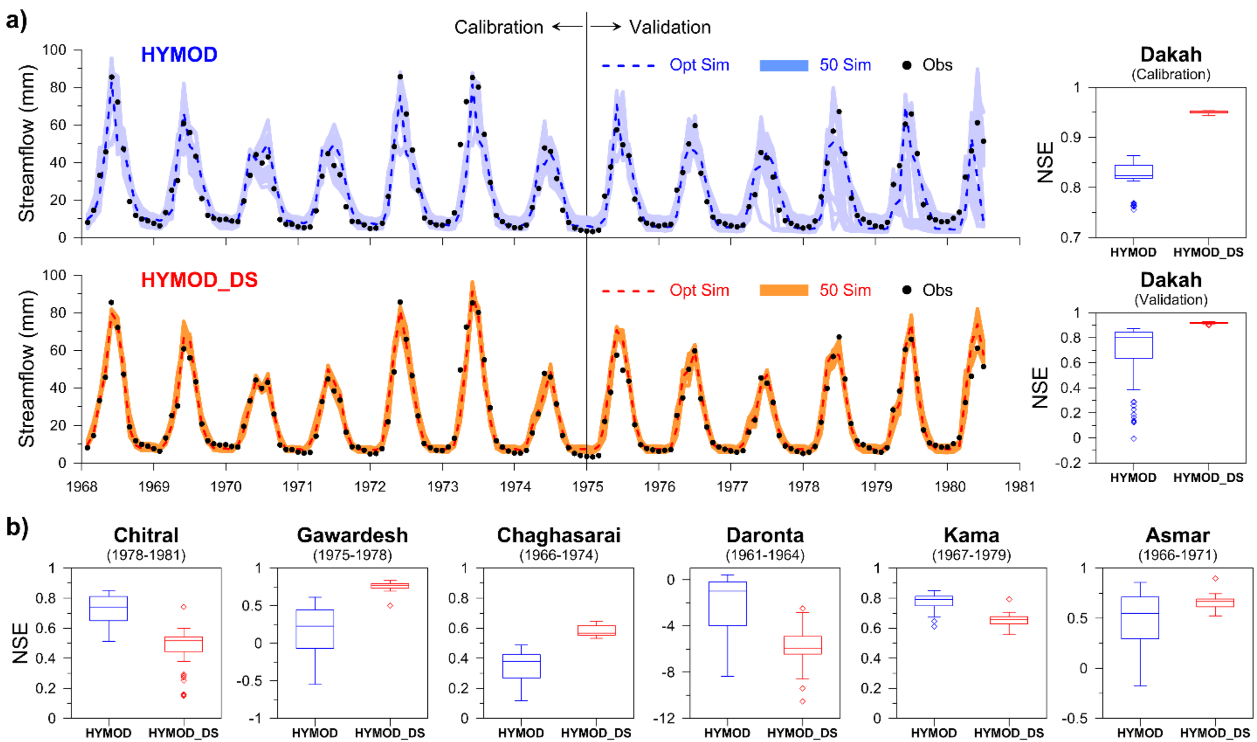


Figure S3. (a) Basin outlet (Dakah) simulations of HYMOD and MYMOD_DS (with the lumped parameterization) from 50 trials of calibration. The Box plots provide the performance evaluation on 50 simulations of both models for both calibration and validation periods. (b) Performances of the models at the 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 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 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 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 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 MGHPCC.

    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:

$$\mathrm{PET} = Coeff \cdot 29.8 \cdot L_{\mathrm{d}} \cdot \frac{0.611 \times \exp\left(17.27 \cdot T/(T+273.3)\right)}{T+273.3}$$

where, L_d is the daylight hours per day, T is the daily mean air temperature (^{o}C), and 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.”

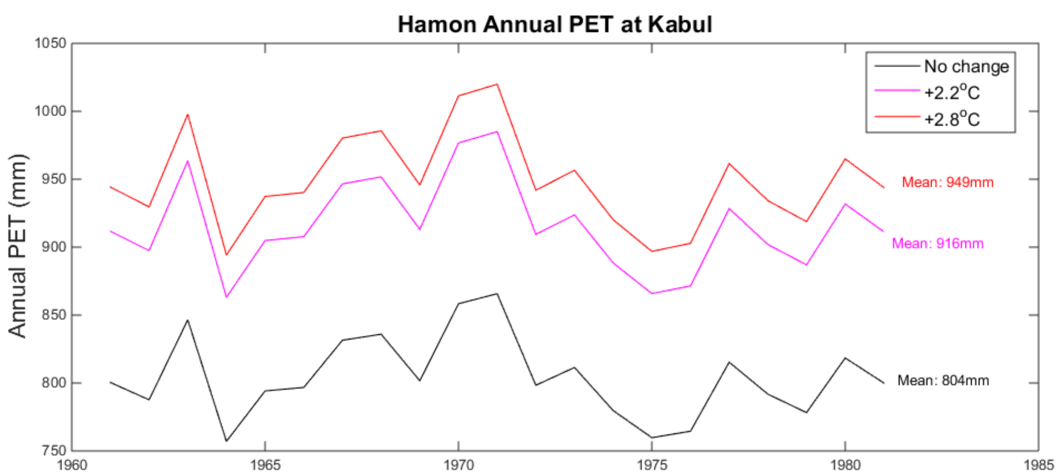
16. p. 10290, lines 25-27: “. . .the HYMOD_DS significantly overestimated streamflow at Daronta and underestimated flow at three sites in the eastern part of the basin” This is a strong evidence of the heterogeneity of the basin. Please, provide some information on the properties of the basin (e.g. geology)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

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 +2.2oC and +2.8oC 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^{o}C, and +2.8^{o}C. The value of calibrated 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^{o}C, and +2.8^{o}C, respectively. The percent changes of PET relative to the historical PET in the warming conditions of +2.2^{o}C and +2.8^{o}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^{o}C and +2.8^{o}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 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 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 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.

Response to Anonymous Referee #2

General comments:

- I see one major limitation of the paper that leads me to ask for at least minor, if not major revisions: there is not much of a scientific discussion. The authors discuss their results most of all “with themselves” by comparing the various results they obtained. The discussion is short of any discussion with findings by other authors (e.g. on P10294 L3 the authors cite other work for the first time in the results and discussion section. This is on the last page of an eight pages long results and discussion section). There is plenty of published work about the effect of parameterization and their spatial variation, lumped vs distributed calibration approaches, performances of models in simulating interior gauges not considered in calibration, see for example results of the DMIP and LUCHEM projects, amongst others. Additionally, climate change effects on discharge in Central Asian catchments has been in the focus of many, many studies – how do these related to the results obtained here?

  Thank you for pointing out this. We also realized that there were not much discussion in the section “Results and Discussion”. To try to follow the reviewer’s suggestion, we expanded our discussion. First, we decided to focus on our results in the result section and change the paper’s structure accordingly. Now we combined the discussion section with the conclusion part. Also, we expanded our discussion by introducing additional references in relevance to our work as suggested by the reviewer. Please find the revisions made in the section “Discussion and Conclusion” and detailed answers to all the specific comments in the following.

Specific comments:

- Title: High performance computing is mentioned in the title, but hardly presented in the method section, and not at all in the discussion. HPC in this paper is used as a technique to be able to run a large number of models, but it is not in the center of research as indicated by the title. I suggest to change the title.

  We understand your concern. We have changed the title to highlight our focus on a poorly gaged basin (which we feel is the more important emphasis of this work anyway). However, we do feel that the use of high performance computing is an important component of this work, so we tried to emphasize the necessity of exploiting parallel computing power to implement this kind of study in the abstract:

  “To address the research questions, high performance computing is utilized to manage the computational burden that results from high-dimensional optimization problems.”

- P10276 L26 There are a number of papers which looked at model performance when excluding/including interior gauging stations during model calibration and validation; see e.g. the DMIP projects (Reed et al., 2004; Smith et al., 2012), the LUCHEM project (Breuer et al., 2009) or work by others (Andersen et al., 2001; Lerat et al., 2012).

  Thank you. We have added the recommended references.

- P10277 L1 You might want to have a closer look to a recent paper by Exbrayat et al. (2014) who investigated the contribution of uncertain model structures versus the impact of uncertain climate change projection to the global predictive model uncertainty. Even though not directly comparable to what the authors show here, it is worth considering and can be used in the discussion, which is lacking other researchers work (see general comment).

  Thank you for suggesting this useful reference. We expanded our discussion with the suggested reference.
  "These results agree with other studies showing the dominance of GCM uncertainty in future hydrologic projections (Chen et al., 2011; Exbrayat et al., 2014). …
  In addition to the uncertainties surrounding model parameters and future climate explored in this study, there is also significant uncertainty in streamflow projections stemming from structural differences between applied hydrologic models, which can be especially pertinent where robust calibration is hampered by the scarcity of data (Exbrayat et al., 2014). Further, the residual error variance of hydrologic model simulations would increase the effects of hydrologic model uncertainty as compared to that of the climate projections (Steinschneider et al., 2014). These issues need to be addressed in future work for exploring a comprehensive uncertainty assessment of climate change risk for poorly monitored hydrologic systems."

- P10277 L18 I do not agree that HPC is so new in hydrological modeling. I rather think that many researcher use HPC without highlighting it. Also in the work presented here, HPC is a tool that is used, but not a method that is further developed or presented in detail.

  We understand and have removed the language suggesting HPC is new in hydrological modeling. While we still feel that the use of HPC is uncommon and adds new possibilities for research questions, we agree that we are using HPC as a tool – it is not the focus of our study.

- P10278 L3 Is the annual precipitation 475 mm or are the 475 mm the 70% of total precipitation? Overall, the study area description is very short. Some more information about topography, soils/geology, flow characteristics, specific discharges from the subcatchments, and land use/management would be helpful to better understand some of the results.

  We dropped the number in the text to avoid any confusion caused by that. The number was meant to be for annul precipitation and is now provided in the updated Table 1.
  Figure 1 has been updated with more information (topography, soil types, and vegetation cover). We expanded the study area description accordingly.

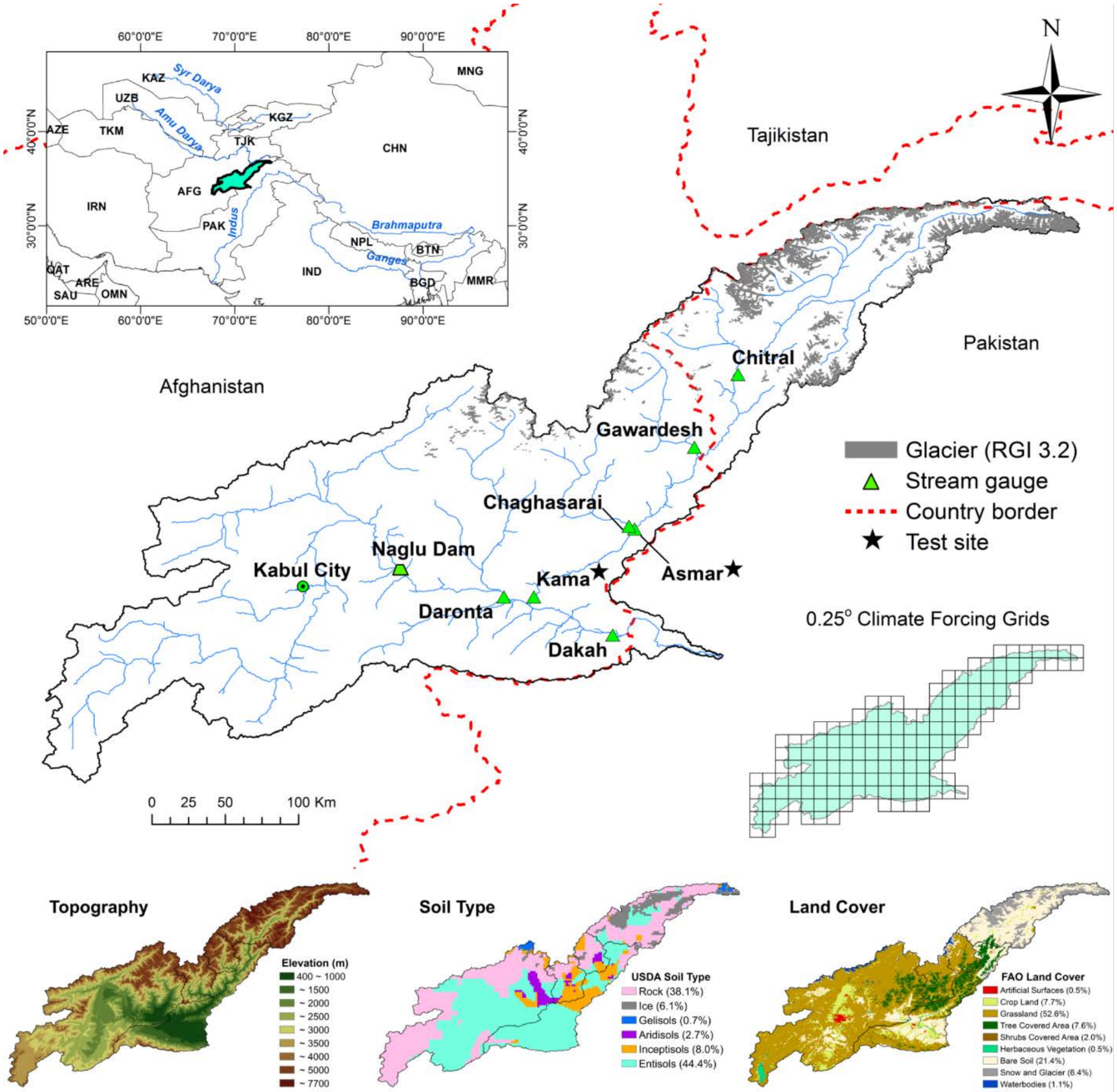


Figure 1. Kabul River Basin.

How about irrigation? Is it an important land management and if so, how did you deal with water abstraction. Looking at the often poor model performance in the western part of your catchment around Kabul I assume that missing information on water abstraction substantially influences your model performance.

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^2 (4.3% of the total basin area) is agricultural land and the average annual flow of the Kabul River is approximately 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 (Efstratisdis 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 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."

- P10278 L21 Should it not be "a genetic algorithm" as there are many kinds of genetic algorithms available for model calibration? Or you should state "the genetic algorithm introduced by Wang et al. 1991".

We made this clearer as suggested.

- P10279 L5 I wonder how these monthly streamflow values were calculated if not from daily measurements. If there are only monthly data available, I also wonder if the NSE is the best choice for goodness of fit criteria. Nevertheless, I like the argumentation given for choosing NSE but suggest to also mentioning here the use of KGE as another goodness of fit criterion for model evaluation (so far, KGE is introduced in chapter 5 in the discussion and not in the methods section).

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.

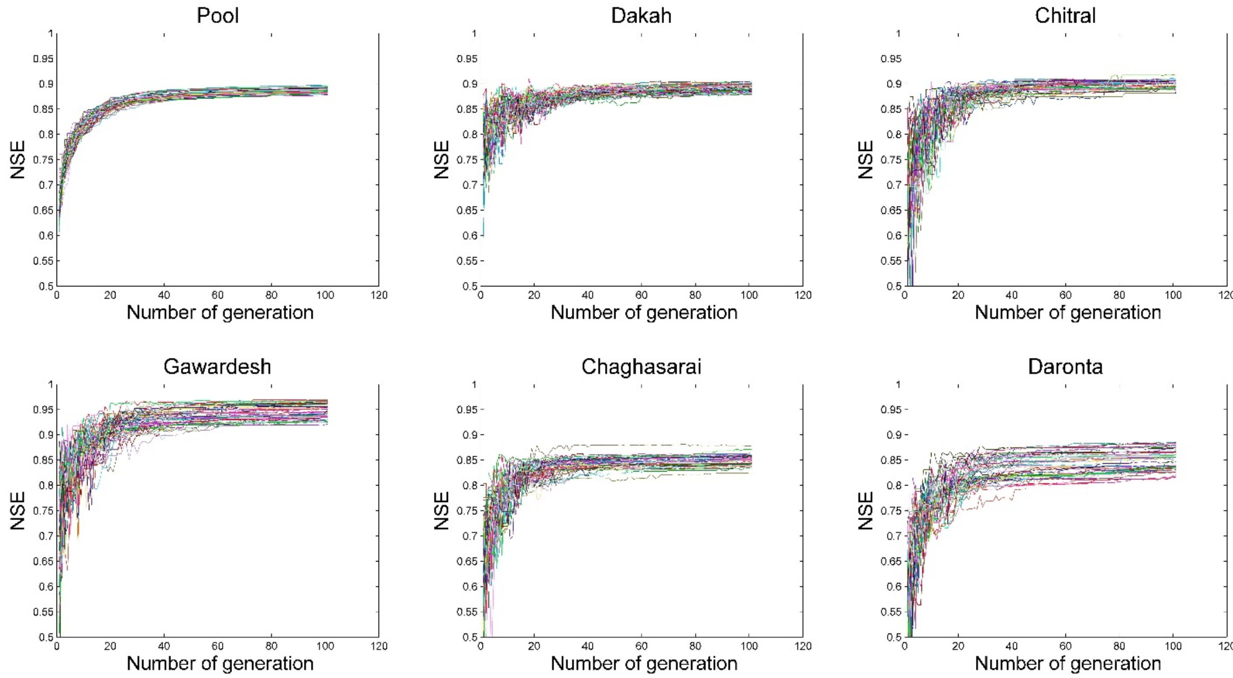
We have added the following details in the manuscript to clarify the immediate question regarding the data:
"Streamflow data were not collected in Afghanistan after September 1980 until recently because stream gaging was discontinued soon after the Soviet invasion of Afghanistan in 1979 (Olson and Williams-Sether, 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."

We acknowledged the limitation of the use of NSE for a model evaluation metric by writing this:
"However, in this particular study daily hydrologic model simulations can only be compared against available monthly streamflow records, reducing the number of viable objectives against which to calibration. That is, statistics representing peak flows, extreme low flows, and other daily flow regime characteristics often used in multi-objective optimization approaches are unavailable. We believe that the use of a monthly NSE value as a single objective, while coarse, does not inhibit our ability to provide insight into the research questions posed."

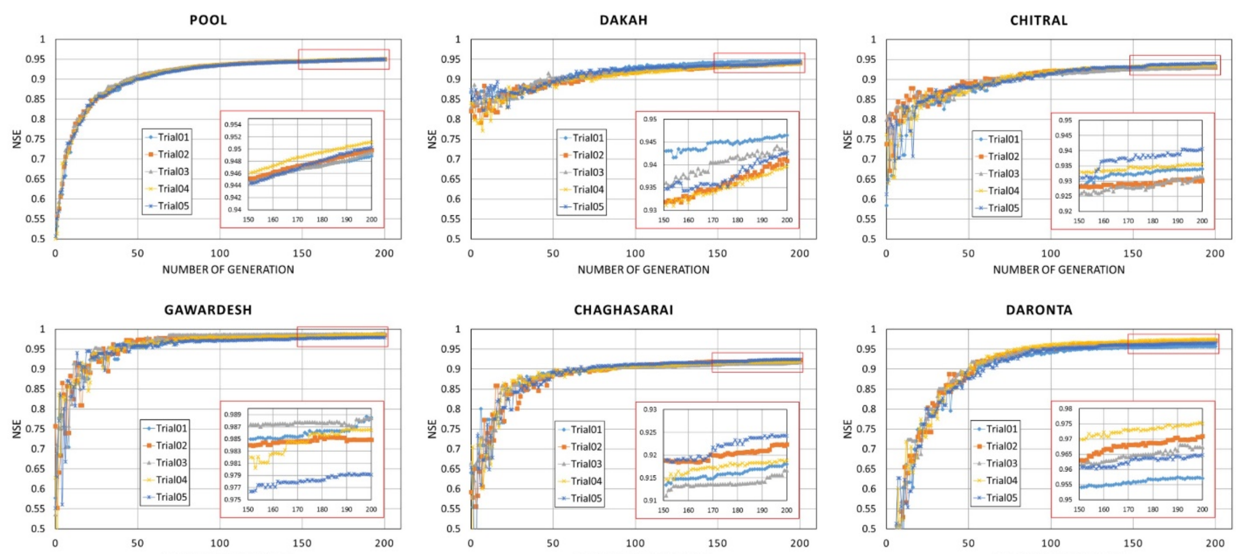
Also, we now introduce the KGE earlier in the Methods section to make clear that we are considering more than just the NSE for model diagnostics.

- P10282 L3 Are the numbers correct? The page before you present 15, 75 and 2400 parameter values being searched for in the various spatial set ups. Should it then not be 15x100 and 75x100? And why is 2400 multiplied by 200 and not by 100 as the others? Even though you state in the next sentence that the population/generation sizes were supported by convergence tests, the generation of numbers given here remains unclear.

We set up different numbers of population and generation in the GA algorithm according to the complexity of parameterization scheme. For instance, for the lumped parameterization, the number of parameter to be optimized is 15 and we considered 150 parameter sets. Those 150 parameter sets evolve through 100 generations, and the result of our convergence test showed a convergence while going through 100 generations. For the distributed parameterization scheme, there are more number of parameters to be calibrated. We considered 2400 parameter sets to calibrated 2400 parameters. Although it can be argued that having 2400 parameter sets to optimize 2400 parameters is not enough, we confirmed from the convergence test that this calibration setup shows a convergence behavior with 200 generations. Below, we enclosed the convergence test results.



GA convergence for the semi-distributed parameterization scheme with 750 parameter sets (population) and 100 iteration (generation)



GA convergence for the distributed parameterization scheme with 2400 parameter sets (population) and 200 iteration (generation)

- P10283 L11 step-wise (not step-wide)

Done.

- P10284 L12 the period “1960-1981” better covers all available discharge measurements given in Table 1.

  Yes, you are right. We changed it.

- P10294 L6 is shown in: : : (not was shown)

  We corrected it.

- Section 6 Conclusion P10295 L8 until P10296 L16 This is an extended summary of the results presented rather than a conclusion of the work. I think more effort should be put into real conclusions – what do we learn from the study, what are suggestions for future research, are results transferable to other regions or modelling approaches?

  As suggested, we tried to focus on the points that should be addressed in the conclusion part.

- Sections 5.2 and 5.3 The model performances for the upper subcatchments Kama and Asmar are generally very good. This is the same for Dakha (Figs 6 and 7). Glaciers have the largest extend in these subcatchments and I assume that they therefore contribute large volumes of water to total discharge at Dakah. Further, I assume that western catchments contribute only minor to total discharge as rainfall input is comparatively low (information on specific discharges for the various subcatchments would be helpful for a quick comparison). As you optimize your model using NSE, with NSE putting emphasis in matching peak flows, it does not come as a surprise to obtain good results for Dakah as long as subcatchments Kama and Asmar are calibrated sufficiently well.

  We updated Table 1 with more contents including the information on specific discharges for the sub-watersheds.
  In our study, we always treated Kama and Asmar as ungagged sub-watersheds, which means that we never tried to calibrate those two sites. All the available data at those sites were used for the validation purpose only. Dakah (the basin outlet) is the one against which the model calibrated. One of the main ideas we try to show in Sections 5.2 and 5.3 is that the calibration based on only the basin outlet does not provide a good performance at Kama and Asmar, while the pooled calibration does.

- Furthermore, the model performance of the ungauged sites Kama and Asmar are often very similar. Looking at the choice of stations that you treated ungauged and the general location of available gauging stations, I wonder why you have selected the Kama and Asmar, which belong to the same eastern area of the catchment. Why have you not selected the one in the west as a second interior test station (i.e. Daronta), or at least two subcatchments which are not draining into each other (e.g. Chaghasari and Asmar) and therefore being more independent than Kama and Asmar.

  The Government of Afghanistan with the support of the international donors (e.g. The World Bank) has developed comprehensive plans for the development of new hydro-power projects, irrigation schemes and rehabilitation of old schemes on various rivers including the Kabul River (IUCN, 2010). Recently, Afghanistan and Pakistan reached an agreement in working on a 1,500MW hydropower project on Kunar River as part of the joint management of common rivers between the two countries (DAWN, 2013). For this study, Kama and Asmar were chosen and treated as ungagged sites in the processes of multisite calibrations because they align with the potential dam project.

This information has been provided accordingly in the text.
"The Government of Afghanistan has developed comprehensive plans for new hydropower projects on the Kabul River owing to its advantageous topography for the development of water storage and hydropower (IUCN, 2010), and recently reached an agreement with the Pakistan government to work on a 1,500MW hydropower project on the Kunar River (one of major tributary in the Kabul River basin) as part of the joint management of common rivers between the two countries (DAWN, 2013). …
Kama and Asmar stations are treated as ungaged sites because they align with the potential dam project on the Kunar River tributary."

- Section 5.4 Do you assume constant glacier volume to be discharging or are glaciers prone to glacier melt, resulting in smaller volume and spatial extend in the future and during your climate change simulation period. What are the expectations in glacier extend for the end of your simulation period in your catchment? Are calibrated model parameters still valid under these new boundary conditions? I expect not, as glacier melt is an important process, described by various parameters (Table 2) and needs rigorous calibration.

The hydrologic model (HYMOD_DS) used in this study does account for the changes in volume but has no ability to trace explicitly the spatial extend of glaciers. At the beginning of the simulation we were informed by the glacier volume (the amount water stored in the glaciers) which is provided by RGI3.2 and the area-volume relationship. A simple and possible way to trace the glacier extend from this study is to back-calculate the area with volume remaining at the end of simulation using the area-volume relationship. The model parameters related to the temperature-index glacier model stay the same once those are calibrated. Therefore, water from glacier melt with respect to a temperature above the threshold temperature will be same as long as glacier keep existing. We agree that it is hard to expect the calibrated parameters to be valid under new glacier conditions.
For our 20-year historical model simulation, we checked that the glacier volume decreases due to the ablation of glaciers larger than accumulation in the sub-watersheds that produce annual total flow larger than annual total precipitation as shown in the new Table 1. We argue that the high ratio of streamflow to precipitation is unrealistic and might be caused by error in precipitation data used in this study since precipitation measurement in high mountain areas is highly uncertain (Immerzeel et al., 2014). What we checked for the 20-year historical simulation and 30-year future simulation is that glaciers still stored enough water at the end of the simulations.
In our discussion for future work, we note the necessity of exploiting remote sensing and satellite products with which the evaluation of distributed hydrologic models with respect to model internal processes (e.g. snow, evapotranspiration, and glacier) becomes possible.

- S2 Please describe the meaning of abbreviations in the legend or figure caption

  We put the description in the figure caption.

- S8 Is this a simulation of the 100 yr flood event, at least this is what I understand from the text (P10294 L6 and following).

  We assumed that the reviewer meant Figure S6, not S8.
  No, this figure is showing the variability of optimum parameters derived from 50 trials of semi-distributed and distributed pooled calibrations. Here, we tried to explore the variability of 100-year flood estimates using 50 calibrated parameter sets for each calibration approach. Specifically, every time when the model was run with an optimum parameter set, we estimated the 100-year flood using the Log-Pearson III distribution for three locations (the basin outlet and 2 ungagged sites). With 50 100-year flood estimates for each calibration approach, we then examined the influence of the parameter variability on the flood estimates by comparing the flood estimates resulting from two calibration approaches.

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, 158-167, 2013.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the stregths 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 Koutsoviannis, 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.

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 Jounal, 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.

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.

Calibration approaches for distributed hydrologic models ~~using high performance computing~~ in poorly gaged basins: Implication for streamflow projections under climate change

S. Wi[1], Y. C. E. Yang[1], S. Steinschneider[1], A. Khalil[2] and C. M. Brown[1]

[1] Department of Civil and Environmental Engineering, University of Massachusetts Amherst, USA

[2] The World Bank, USA

Correspondence to: S. Wi (email: sungwookwi@gmail.com)

Submitted to Hydrology and Earth System Sciences

January 6, 2015

Abstract

This study ~~utilizes high performance computing to~~ tests the performance and uncertainty of calibration strategies for a spatially distributed hydrologic model in order to improve model simulation accuracy and understand prediction uncertainty at interior ungaged sites of a sparsely-gaged watershed. The study is conducted using a distributed version of the HYMOD hydrologic model (HYMOD_DS) applied to the Kabul River basin. Several calibration experiments are conducted to understand the benefits and costs associated with different calibration choices, including 1) whether multisite gaged data should be used simultaneously or in a step-wise manner during model fitting, 2) the effects of increasing parameter complexity, and 3) the potential to estimate interior watershed flows using only gaged data at the basin outlet. The implications of the different calibration strategies are considered in the context of hydrologic projections under climate change. To address the research questions, high performance computing is utilized to manage the computational burden that results from high-dimensional optimization problems. Several interesting results emerge from the study. The simultaneous use of multisite data is shown to improve the calibration over a step-wise approach, and both multisite approaches far exceed a calibration based on only the basin outlet. The basin outlet calibration can lead to projections of mid-21st century streamflow that deviate substantially from projections under multisite calibration strategies, supporting the use of caution when using distributed models in data-scarce regions for climate change impact assessments. Surprisingly, increased parameter complexity does not substantially increase the uncertainty in streamflow projections, even though parameter equifinality does emerge. The results suggest that increased (excessive) parameter complexity does not always lead to increased predictive uncertainty if structural uncertainties are present. The

largest uncertainty in future streamflow results from variations in projected climate between climate models, which substantially outweighs the calibration uncertainty.

1. Introduction

In an effort to advance hydrologic modelling and forecasting capabilities, the development and implementation of physically-based, spatially distributed hydrologic models has proliferated in the hydrologic literature, supported by readily available geographic information system (GIS) data and rapidly increasing computational power. Distributed hydrologic models can account for spatially variable physiographic properties and meteorological forcing (Beven, 2012), improving simulations compared to conceptual, lumped models for basins where spatial rainfall variability effects are significant (Ajami, et al., 2004; Koren, et al., 2004; Reed, et al., 2004; Khakbaz, et al., 2012; Smith, et al., 2012) and for nested basins (Bandaragoda, et al., 2004; Brath, et al., 2004; Koren, et al., 2004; Safari, et al., 2012; Smith, et al., 2012). The benefits of distributed modeling have been recognized by the U. S. National Oceanic and Atmospheric Administration's National Weather Service (NOAA/NWS) and demonstrated in the Distributed Model Intercomparison Project (DMIP) (Reed, et al., 2004; Smith, et al., 2004; Smith, et al., 2012; Smith, et al., 2013). Importantly, distributed hydrologic models can evaluate hydrological response at interior ungaged sites, a benefit not afforded by ~~conceptual,~~ lumped models. The use of distributed hydrologic modelling for interior point streamflow estimation is particularly relevant for poorly gaged river basins in developing countries, where reliable predictions at interior sites are often required to inform water infrastructure investments. As international development agencies begin to integrate climate change considerations into their decision-making processes (e.g., Yu et al., 2013), these investments need to be robust under both current climate conditions and ~~alternative~~ possible future climate regimes.

Despite their roots in physical realism, distributed hydrologic models can suffer from substantial uncertainty. A major source of uncertainty originates from the proper identification of parameter values that vary across the watershed, especially when observed streamflow data is only available at one or a few points (Exbrayat et al., 2014). 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 ~~hydrologic response units~~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). With limited data, the parameter identification problem, particularly for the fully distributed case, can be impractical or infeasible (Beven, 2001). The parameterization challenge has spurred substantial advances in understanding appropriate calibration techniques for distributed hydrologic models. Many studies have attempted to reduce the dimensionality of the calibration problem to alleviate the issue of equifinality (Beven & Freer, 2001), which is the phenomenon whereby multiple parameter sets produce indistinguishable model performance. This work has found favorable results when the parametric complexity of the distributed model is aligned with the data available for calibration (Leavesley, et al., 2003; Ajami, et al., 2004; Eckhardt, et al., 2005; Frances, et al., 2007; Zhu & Lettenmaier, 2007; Cole & Moore, 2008; Pokhrel & Gupta, 2010; Khakbaz, et al., 2012). There has also been extensive research exploring the use of multiple objectives and different operational procedures to understand parameter estimation tradeoffs and identifiability for distributed model calibration, with great success (Madsen, 2003; Efstratiadis & Koutsoyiannis, 2010; Li, et al., 2010; Kumar, et al., 2013).

Despite these advances, important questions still persist. It still remains difficult to compare the uncertainty that emerges from different operational calibration procedures for multisite applications (i.e. whether gages in series should be used sequentially or simultaneously for calibration) and under different levels of parametric complexity. Due to the computational burden required to calibrate distributed models, this uncertainty is problematic to explore. Further, in poorly gaged basins, it is challenging to quantify the lost accuracy and increased uncertainty for interior flow estimation when a distributed model is calibrated only at an outlet gage (which is often all that is available in developing country river basins). 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 ~~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; Breuer et al., 2009; Lerat et al., 2012; Simith et al., 2012; Wang, et al., 2012). ~~but t~~The extent of ~~the~~ this error and uncertainty is not well understood for heterogeneous basins ~~unknown~~ due to the computational expense required~~needed~~ to explore this issue. Finally, rarely have the implications of these calibration issues been explicitly examined for ~~an alternative climate~~ possible future climate conditions, which is required in climate change impact studies. This question has been explored for lumped, conceptual models (Wilby, 2005; Steinschneider, et al., 2012), but has been difficult to evaluate for computationally expensive distributed models.

This study addresses the above research challenges by focusing on the following four questions: 1) How does calibration procedure for using multisite data effect the accuracy and uncertainty of distributed models used for streamflow predictions at ungaged sites, 2) what effects do increased parameter complexity have on distributed model calibration and prediction, 3) how much degradation in model accuracy and uncertainty can be expected for interior flow estimation based on a calibration procedure using only the basin outlet, and 4) how do different calibration formulations for a distributed model alter projections of streamflow at ungaged sites under climate change conditions? These questions are considered in an application of a distributed version of the daily HYMOD hydrologic model to the Kabul River basin in Afghanistan and Pakistan. To address these research questions, high performance computing is utilized to manage the computational burden that often hinders such explorations~~, a relatively recent technique employed in hydrological modeling research~~ (Laloy & Vrugt, 2012; Zhang, et al., 2013).

2. Study area

The Kabul River basin (67,370km^2) is a plateau surrounded by mountains located in the eastern central part of Afghanistan (Figure 1). It is the most important river basin of Afghanistan, containing 35 percent of the country's population. While it encompasses just 12 percent of the area of Afghanistan, the basin's average annual streamflow (about 24 billion cubic meters) is about 26 percent of the country's total streamflow volume (World Bank, 2010).

Water resources from the basin are shared by Afghanistan and Pakistan and serve as a water supply source for more than 20 million people. The shared use of transboundary water between these two countries is central in establishing regional water resources development for this area (Ahmad, 2010). It is crucial to develop tools that can support engineering plans for existing and potential water infrastructure to take full advantage of the water resources in the basin. The Government of Afghanistan has developed comprehensive plans for new hydropower projects on the Kabul River owing to its advantageous topography for the development of water storage and hydropower (IUCN, 2010), and recently reached an agreement with the Pakistan government to work on a 1,500MW hydropower project on the Kunar River (one of major tributary in the Kabul River basin) as part of the joint management of common rivers between the two countries (DAWN, 2013).

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~~ 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 ~~5.7~~2.9% (~~3813km^2~~1954km^2) 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 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.

Issues of shared water resources between Afghanistan and Pakistan in the Kabul River basin are becoming complex due to the impacts of climatic variability and change (IUCN, 2010). ~~In recent years, most of the world's mountain glaciers have shown negative mass balance and rapid decrease in glacier area and volume (Dyurgerov & Meier, 2005), while in the Himalayan region trends depend on location (Bolch et al., 2012).~~ The vulnerability of glacial streamflow regimes to changes in temperature and precipitation (Stahl, et al., 2008; Immerzeel, et al., 2012; Radic et al.,

2014) highlights the need to assess the impact of climate change on ~~water resources~~ future water availability in this area ~~(Immerzeel, et al., 2010; Immerzeel, et al., 2013; Molg, et al., 2014; Radic, et al., 2014)~~.

3. Data and Models

3.1. Data

Gridded daily precipitation and temperature products with a spatial resolution of 0.25^{o} were gathered between calendar years 1961-2007 from the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation *(*APHRODITE) dataset (Yatagai, et al., 2012). There has been some concern regarding underestimation of precipitation in APHRODITE for some regions of Asia (Palazzi, et al., 2013); our preliminarily data analysis (intercomparison of precipitation products between 5 different databases) confirmed this for the Kabul River basin (shown in Figure ~~S2~~S1). Thus, the APHRODITE precipitation was bias-corrected by the precipitation product from the University of Delaware global terrestrial precipitation (UD) dataset (Legates & Willmott, 1990). Daily series of bias-corrected APHRODITE precipitation were coupled with APHRODITE temperature for 160 0.25^{o} grid cells to produce a climate forcing dataset for the distributed domain of the Kabul River basin model.

This study used the set of global climate change simulations from the World Climate Research Programme's Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model ensemble (Talyor, et al., 2012). Monthly climate outputs of GCMs were downscaled to a daily temporal resolution and 0.25^{o} spatial resolution based on the bias-correction spatial disaggregation (BCSD) statistical downscaling method introduced by Wood et al. (2004).

Monthly streamflow observations for seven locations in the Kabul River basin (Figure 1) were gathered between calendar years ~~1961~~1960-~~1980~~ 1981 from two data sources: the Global Runoff Data Centre (GRDC) database and the United States Geological Survey (USGS) database (Table 1). 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. The available monthly ~~-~~streamflow observations at each station were used for calibrating and validating the distributed hydrologic model (Figure ~~3~~2). Kama and Asmar stations are treated as ungaged sites because they align with the potential dam project on the Kunar River tributary. ~~and~~ The two gage stations are left out of the processes of multisite calibrations in order to evaluate the model's ability to predict streamflow at interior ungaged sites. Furthermore, half of the record at the Dakah station, located at the basin outlet, is also used for validation purposes.

The Randolph Glacier Inventory version 3.2 (RGI 3.2) dataset (Pfeffer, et al., 2014) was used to extract glacial coverage in the Kabul River basin, which totaled 5.7% of the basin area (Figure ~~S3~~S2). In the hydrological modeling process, the model needs to be informed by reliable estimates on volume of water retained in glaciers, especially for future simulations under warming conditions. We followed the method proposed in Grinsted (2013), which uses multivariate scaling relationships to estimate glacier and ice cap volume based on elevation range and area. Specifically, the scaling law including area and elevation range factors was applied to estimate glacier/ice cap volume when the glacier depth exceeded 10m. Otherwise, glacier/ice cap volume was estimated with the area-volume scaling law. The elevation range spanned by each individual glacier is estimated using the global digital elevation model (DEM) from the shuttle radar topography mission (SRTMv4) in 250m resolution (Jarvis, et al., 2008). Density of ice (0.9167 g/cm^3) is applied to calculate glacier/ice cap volume in meters of water equivalent.

The database for land covers and soil types of the Kabul River basin (Figure 1) are provided by the Food and Agriculture Organization of the United Nations (Latham, 2014) and United States Department of Agriculture-Natural Resources Conservation Service Soils (USDA-NRCS, 2005), respectively.

3.2. Distributed Hydrologic Model (HYMOD_DS)

In this study the lumped conceptual hydrological model HYMOD (Boyle, 2001) is coupled with a river routing model to be suitable for modelling a distributed watershed system. We name it HYMOD_DS denoting the distributed version of HYMOD. Snow and glacier modules have been introduced to enhance the modelling process for glacier and snow covered areas within the Kabul River basin. The HYMOD_DS is composed of hydrological process modules that represent soil moisture accounting, evapotranspiration, snow processes, glacier processes and flow routing. The model operates on a daily time step and requires daily precipitation and mean temperature as input variables. The overall model structure of the HYMOD_DS and its 15 parameters are described in Figure ~~4~~ 3 and Table 2 respectively. Further details are provided below.

The HYMOD conceptual watershed model has been extensively used in studies on streamflow forecasting and model calibration (Wagener, et al., 2004; Vrugt, et al., 2008; Kollat, et al., 2012; Gharari, et al., 2013; Remesan, et al., 2013). The HYMOD is a soil moisture accounting model based on the probability-distributed storage capacity concept proposed by Moore (1985). This conceptualization represents a cumulative distribution of varying storage capacities (C) with the following function:

$$F(C) = 1 - (1 - \frac{C}{C_{\max}})^{B} \quad 0 \le C \le C_{\max} \tag{1}$$

where the exponent B is a parameter controlling the degree of spatial variability of storage capacity over the basin and $C_{\max}$ is the maximum storage capacity. The model assumes that all storages within the basin are filled up to the same critical level ($C^{*}(t)$), unless this amount exceeds the storage capacity of that particular location. With this assumption, the total water storage $S(t)$ contained in the basin corresponds to

$$S(t) = \frac{C_{\max}}{B+1} \cdot \left(1 - \left(1 - \frac{C^{*}(t)}{C_{\max}}\right)^{B+1}\right) \tag{2}$$

Consequently, two parameters are introduced for the runoff generation process with two components:

$$Runoff_1 = \begin{cases} P(t) + C^{*}(t-1) - C_{\max} & \text{if } P(t) + C^{*}(t-1) \ge C_{\max} \\ 0 & \text{if } P(t) + C^{*}(t-1) < C_{\max} \end{cases} \tag{3}$$

$$Runoff_2 = \begin{cases} \left(P(t) - Runoff_1\right) - \left(S(t) - S(t-1)\right) & \text{if } P(t) - Runoff_1 \ge S(t) - S(t-1) \\ 0 & \text{if } P(t) - Runoff_1 < S(t) - S(t-1) \end{cases} \tag{4}$$

where $P(t)$ is precipitation, $Runoff_1$ is surface runoff, and $Runoff_2$ is subsurface runoff. A parameter (α) is introduced to represent how much of the subsurface runoff is routed over the fast (Q_{fast}) and slow (Q_{slow}) pathway:

$$Q_{\text{fast}} = Runoff_1 + \alpha \cdot Runoff_2 \tag{5}$$

$$Q_{\text{slow}} = (1-\alpha) \cdot Runoff_2 \tag{6}$$

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:

$$\text{PET} = Coeff \cdot 29.8 \cdot L_{\text{d}} \cdot \frac{0.611 \times \exp(17.27 \cdot T/(T+273.3))}{T+273.3} \tag{7}$$

where, L_d is the daylight hours per day, T is the daily mean air temperature (°C), and 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).

The HYMOD_DS includes snow and glacier modules with separate runoff processes, i.e., the runoff from the glacierized area is calculated separately and added to runoff generated from the soil moisture accounting module coupled with the snow module. The implicit assumption here is that there is no interchange of water between soil layers and glacial area and runoff from glacial areas is regarded as surface flow. The runoff from each area is weighted by its area fraction within the basin to obtain total runoff.

The time rate of change in snow and glacier volume governed by ice accumulation and ablation (melting and sublimation) is expressed by the Degree Day Factor (DDF) mass balance model (Moore, 1993; Stahl, et al., 2008). The dominant phase of precipitation (snow vs. rain) is determined by a temperature threshold (T_{th}). The snow melt M_{s} and glacier melt M_{g} is calculated as:

$$M_{\mathrm{s}} = DDF_{\mathrm{s}} \times \left(T - T_{\mathrm{s}}\right) \qquad (\cancel{7}\underline{8})$$

$$M_{\mathrm{g}} = DDF_{\mathrm{g}} \times \left(T - T_{\mathrm{g}}\right) \qquad (\cancel{8}\underline{9})$$

with DDF_{s} (T_{s}) and DDF_{g} (T_{g}) applied separately for snow and glacier modules, respectively. To account for the higher melting rate of glacier than snow owing to the low albedo (Konz & Seibert, 2010; Kinouchi, et al., 2013), we introduced a parameter $r > 1$ to constrain DDF_{g} to be larger than DDF_{s} (i.e. $DDF_{\mathrm{g}} = r \times DDF_{\mathrm{s}}$). For the rain that falls on the glacierized area, the glacier parameter K_{g} determines the portion of rain becoming surface runoff as a multiplier for the rainfall. The remaining rainfall is assumed to be accumulated to the glacier store.

The within-grid routing process for direct runoff is represented by an instantaneous unit hydrograph (IUH) (Nash, 1957), in which a catchment is depicted as a series of N reservoirs each having a linear relationship between storage and outflow with the storage coefficient of K_{q}. Mathematically, the IUH is expressed by a gamma probability distribution:

$$u(t) = \frac{K_{\mathrm{q}}}{\Gamma(N)} \left(K_{\mathrm{q}} t\right)^{N-1} \exp\left(-K_{\mathrm{q}} t\right) \qquad (\cancel{9}\underline{10})$$

where, Γ is the gamma function. The within-grid groundwater routing process is simplified as a lumped linear reservoir with the storage recession coefficient of K_{s}.

The transport of water in the channel system is described using the diffusive wave approximation of the Saint-Venant equation (Lohmann, et al., 1998):

$$\frac{\partial Q}{\partial t} + C \frac{\partial Q}{\partial x} - D \frac{\partial^2 Q}{\partial x^2} = 0 \qquad (\cancel{10}\underline{11})$$

where C and D are parameters denoting wave velocity (*Velo*) and diffusivity (*Diff*) respectively.

Similar to most other hydrological models (Efstratisdis 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 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.

4. Methods

The purpose of this study is to explore the implications of different calibration strategies and choices for a computationally expensive distributed hydrologic model. A variety of calibration experiments are conducted, with the results from preceding experiments informing choices made for subsequent ones. All calibration approaches are tested in terms of their ability to predict flows at interior site gages that were left out of the calibration process. In all cases, the Genetic Algorithm (GA) introduced by Wang (1991) is used as an optimization method for model parameter calibration ~~(Wang, 1991; Zhang, et al., 2008; Kollat, et al., 2012)~~, and the objective function is based simply on the Nash Sutcliffe efficiency (NSE) (Nash & Sutcliff, 1970), which is by far the most utilized performance metric in hydrological model applications (Biondi et al., 2012). A multisite average of the NSE is used when evaluating performance across multiple sites. We fully recognize that the use of one objective, such as the NSE, is inferior compared to multi-objective approaches that can identify Pareto optimal solutions that provide good model performance across different components of the flow regime (Madsen, 2003; Efstratiadis & Koutsoyiannis, 2010; Li, et al., 2010; Kumar, et al., 2013). However, in this particular study daily hydrologic model simulations can only be compared against available monthly streamflow records, reducing the number of viable objectives against which to calibration. That is, statistics representing peak flows, extreme low flows, and other daily flow regime characteristics often used in multi-objective optimization approaches are unavailable. We believe that the use of a monthly NSE value as a single objective, while coarse, does not inhibit our ability to provide insight into the research questions posed. In addition to the NSE, the Kling-Gupta efficiency (KGE) (Gupta et al., 2009) is adopted as an alternative model performance metric, which equally weights model mean bias, variance bias, and correlation with observations.

In this study, three levels of parameter complexity are considered: lumped, semi-distributed, and fully distributed formulations (Figure ~~2~~4). The different levels of parameter complexity are defined according to the spatial distribution of unique hydrologic model parameters. In the lumped formulation a single parameter set is applied to the entire basin. In the semi-distributed formulation, a unique parameter set is assigned to each sub-basin, defined based on the location of available streamflow gaging sites. The fully distributed parameter structure follows the spatial discretization of climate input grids, allowing a unique parameter set for each grid cell. No matter the parameterization scheme, the model structure follows the climate input grids, i.e. the hydrological water cycle within each grid cell is modelled separately. 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 Smith et al., 2013). However, the focus of our study is on ungaged interior site streamflow estimation, making this 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 formulation, support the use of the distributed grid structure (Figure S3).

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~~ calibration uncertainty. The considerably high computational cost required to perform a large number of calibrations is managed using the parallel computing power provided by the Massachusetts Green High Performance Computing Center (MGHPCC), from which several thousands of processors are available.

In the first modeling experiment, we explore two calibration strategies for using multisite streamflow data, a stepwise and pooled approach. In the stepwise calibration, parameters are calibrated for upstream gaged sub-catchments and subsequently fixed during calibration of downstream points, while for the pooled approach, parameters are calibrated for multiple sub-catchments simultaneously. Both approaches are assessed for the semi-distributed formulation. The better of the two methods is identified for use in the second experiment, where the effects of increased parameter complexity are tested in terms of streamflow prediction accuracy and uncertainty. In the third experiment, we consider the situation where there is only ~~gaged location~~data at the basin outlet for calibration. Here, the model is calibrated against the outlet gage under all levels of parameter complexity and is compared against the best combination of calibration strategy (step-wise or pooled) and parameter complexity (lumped, semi-distributed, or fully distributed) identified in the previous experiments. Finally, a subset of the calibration approaches deemed worthy of further investigation are compared in terms of their projections of future streamflow under climate change to highlight how model calibration differences can alter the results of a climate change assessment for water resources applications. These experiments are described in further detail below.

4.1. Multisite Calibration: Stepwise and Pooled Approaches

In the first experiment, the semi-distributed parameterization concept is compared under alternative multisite calibration strategies, the stepwise and pooled calibration approaches. To conduct the stepwise calibration, a nested class of sub-basins is defined corresponding to multiple gaging stations. In the first step of the stepwise calibration, the optimization process is carried out with nested sub-basins at the lowest level (i.e., the most upstream sites). Once parameters of nested sub-basins are determined, the parameters are fixed, and the calibration procedure proceeds with nested basins at upper levels until parameters for the entire basin are determined. In this particular application to the Kabul River basin, 5 gaged sub-basins were selected and the stepwise calibration procedure for those sub-basins followed this direction: Chitral → Gawardesh → Chaghasarai → Daronta → Dakah (Figure ~~S1~~5). The stepwise calibration approach involves a number of GA implementations corresponding to the number of gaging sites. The GA optimization was carried out a total of 250 times in this application, with 50 optimization runs containing GA implementations for 5 sub-basin regions.

The pooled calibration strategy involves calibrating all parameters of the model domain simultaneously against multiple streamflow gages within the watershed. This approach aims at looking for suitable parameters that are able to produce satisfactory model results at all gaging stations in a single implementation of GA optimization. That is, the GA searches the entire parameter space at once to maximize the average NSE across all sites. This operational feature reduces the processing time spent on the GA implementation compared to the stepwise calibration strategy. To identify the better of the two multisite calibration approaches, the comparison focused on their ability to predict streamflow and calibration uncertainties at two interior site gages (Kama and Asmar) that were assumed to be ungaged (Figure ~~S1~~5), as well as for validation data at the basin outlet.

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.

4.2. Increased Parameter Complexity

In the second experiment, the better of the two approaches (step-wise or pooled) identified in the first experiment is further tested with respect to the three different levels of parameter complexity. In addition to the semi-distributed parameter formulation considered in the first experiment, lumped and fully-distributed parameter formulations are calibrated for the selected approach to investigate the gain or loss arising from different levels of parameter complexity. Since the hydrologic model HYMOD employed in this study involves 15 parameters, 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 increases dramatically. With 160 0.25^{o} grid cells, the number of parameters requiring calibration reaches 2,400. As the number of parameters increase across the parameterization schemes, calibration becomes increasingly computationally expensive. The number of model runs used in the GA optimization algorithm for the lumped, semi-distributed, and distributed parameterization schemes are 15,000 (150 populations × 100 generations), 75,000 (750 × 100), and 480,000 (2400 × 200), respectively. These population/generation sizes were supported using convergence tests for each calibration. Again, 50 separate GA optimizations were used to explore calibration uncertainties for each parameterization scheme. To give a sense of the computational burden of this experiment, we note that 50 trials of the HYMOD_DS calibration under the distributed conceptualization required 1,000 processors over 7 days on the MGHPCC system.

4.3. Basin Outlet Calibration

The third experiment considers the situation where there is only gaged data at the basin outlet (Dakah) for calibration, a common situation when calibrating hydrologic models in data-scarce river basins. Here, we evaluate the potential of the basin outlet calibration to estimate interior watershed flows in terms of both accuracy and precision at all gaging stations. All levels of parameter complexity are considered for this calibration. The main purpose of this experiment is to compare the veracity of a distributed hydrologic model calibrated only using basin outlet data with results from multisite calibrations to better understand the degradation in model performance under data scarcity. Other than the use of an NSE objective only at the basin outlet, all other GA settings for each level of parameter complexity are same as the settings used in the second experiment.

4.4. Climate Change Projections of Streamflow

The fourth experiment investigates how the choice of calibration approach can alter the projections of future streamflow under climate change. To explore this question, streamflow simulations for the 2050s, defined as the 30-year period spanning from 2036 to 2065, are carried out using climate projections from the ~~World Climate Research Programme's Coupled Model Intercomparison Project Phase 5 (~~CMIP5~~)~~ (Talyor, et al., 2012). A total of 36 different climate models run under two future conditions of radiative forcing (RCP 4.5 and 8.5) are used. Streamflow projections are developed for the basin outlet (Dakah) and two interior gages left out of the calibration (Kama and Asmar). By using 36 different General Circulation Models (GCMs) and 50 optimization trials for each calibration scheme, this analysis compares the uncertainty in future streamflow projections originating from uncertainty in different hydrologic model parameterization schemes and under alternative future climates.

Streamflow projections are considered under all three parameterization schemes (lumped, semi-distributed, and fully distributed) for both the basin outlet model and the best multi-site calibration approach (step-~~wide~~ wise or pooled). Multiple streamflow characteristics are evaluated, including monthly streamflow~~ climatology~~, wet (April-September) and dry (October-March) season flows, and daily peak flow response. The differences and uncertainty in these metrics across calibration approaches will highlight the importance of calibration strategy for evaluating future water availability and flood risk.

5. Results ~~and Discussion~~

For the remaining part of the paper, we introduce the following shorthand: Lump, Semi, and Dist indicate the lumped, semi-distributed, and fully distributed parameterization schemes, and Outlet, Stepwise, and Pooled correspond to basin outlet, stepwise, and pooled calibrations. The comparison between different calibration strategies is based on the model performance evaluated with the NSE, as well as an alternative metric, the ~~Kling-Gupta efficiency (~~KGE~~)~~. ~~(Gupta et al., 2009), which equally weights model mean bias, variance bias, and correlation with observations.~~

5.1. Pooled Calibration vs. Stepwise Calibration

This section reports the results from the first experiment comparing the stepwise and pooled calibration approaches for the semi-distributed model parameterization. Figure ~~5~~ 6 shows the comparison between the Semi-Stepwise and Semi-Pooled with boxplots representing the 50 trials of calibration. Under the stepwise calibration the results for 4 sub-basins (Chitral, Gawardesh, Chaghasarai, and Daronta) are optimal because there is no interaction between those sub-basins. However, the calibrated parameter sets of each sub-basin act as constraints in the last step of the Semi-Stepwise resulting in the degradation of model skill at the basin outlet (Dakah) and two left-out gages (Asmar and Kama). This becomes apparent when comparing the Semi-Stepwise to the Semi-Pooled results. The model skill under the Semi-Pooled is similar to that from the Semi-Stepwise with respect to the 4 upstream sub-basins, but it outperforms at the verification gages. This is particularly true for the Asmar gage, which exhibits a downward bias and substantial variability in performance under the Semi-Stepwise. The Semi-Pooled results suggest that small sacrifices of model performance at certain sites can improve and stabilize basin-wide performance. Expected values of KGE from 50 calibrations are also provided (values in parenthesis in the bottom of Figure ~~5~~6) and this performance metric also leads to the same conclusion. Therefore, the Semi-Pooled was selected as the better multisite calibration strategy and is considered for further analyses in the following sections.

5.2. Pooled Calibration with Alternative Parameterizations

Here we examine results for the three levels of parameter complexity applied to the pooled calibration approach. Figure ~~6~~ 7 shows the comparison of the pooled calibrations. Unsurprisingly, streamflow predictions from the Lump-Pooled have the lowest accuracy and largest uncertainty at the calibration sites, particularly for the Chaghasarai and Daronta sites. This demonstrates the well-known difficulty in representing flow characteristics of a spatially variable system with a homogenous parameter set (Beven, 2012). The pooled calibration substantially improves with increasing parameter complexity at the calibration sites. Both the Semi-Pooled and Dist-Pooled produce NSE values above 0.8 for all calibration sites, with the Dist-Pooled showing somewhat higher performance, undoubtedly from its greater freedom to over-fit to the calibration data. However, the advantage of the Dist-Pooled with respect to streamflow predictions at validation sites becomes less clear. Only the Dist-Pooled at Kama shows marginally better predictions, while the results are ambiguous at Dakah and Asmar. Overall, this likely suggests that the fully distributed conceptualization leads to over-fitting of the model as compared to the Semi-Dist conceptualization. We reached the same conclusion when examining the KGE values, which rise with greater parameter complexity at calibration sites but no longer follow this pattern strictly at validation sites.

Interestingly, the Lump-Pooled performs well at the verification sites despite its poor performance at calibration sites. The Lump-Pooled does not show significant degradation in skill at Kama compared to the more complex parameterizations, and the flow prediction at Asmar actually exhibits the best performance of all three model variants. A partial reason for this unexpected result arises from different overlapping periods in the calibration and validation data (see Figure ~~3~~2). The periods used for the calibration for Chitral (1978-1981) and Gawardesh (1975-1978) have no overlapping periods with the one for Asmar (1966-1971), which encompasses those two sub-basins. Instead, the validation at Asmar is mostly affected by the calibration to Dakah because of the overlapping 4 years (1968-1971) between those two sites. This explains the reason why the Lump-Pooled shows high skill at Asmar despite the low skill at its sub-basins. However, the low model skill at Chaghasarai from the Lump-Pooled propagates to the validation result at Kama, as these two sites have a relatively long overlapping period (8 years from 1967-1974).

5.3. Limitations of the Basin Outlet Calibration

In the third experiment the HYMODS_DS was calibrated only to data at the basin outlet under all levels of parameter complexity, and streamflow records for all 6 sub-basins, as well as flows at Dakah not used during calibration, are used for model validation. First, we consider the flows at Dakah. During the calibration period, all three parameterization schemes produce very accurate streamflow predictions with NSE (KGE) values above 0.95 (0.96) (Figure ~~7~~8). High accuracy holds even under the Lump_Outlet, ~~which is somewhat surprising given~~despite the spatial heterogeneity of the basin. While NSE and KGE values at Dakah rise marginally with greater parameter complexity during calibration, this no longer holds during the validation period, suggesting no benefit with an increase in parameter complexity.

The validation results for the 6 sub-basins demonstrate the danger in relying on outlet data alone when calibrating a distributed model for flow prediction at interior points. Streamflow predictions at interior sites exhibit low accuracy and high uncertainty, with the worst performance at the Daronta site (all NSEs and KGEs are negative). We note that the poor performance at Daronta is likely due in part to the impacts of water abstraction and the operation of Naglu dam. Further examination (Figure S4) showed that the HYMOD_DS significantly overestimated streamflow at Daronta and underestimated flow at three sites in the eastern part of the basin (Chitral, Gawardesh, and Chaghasarai). Model performance at Kama and Asmar is somewhat better than the other validation sites, although improvements are not the same across all parameterizations. The Lump-Outlet predictions at these sites still have low average accuracy (average NSE < 0.7 and average KGE < 0.6), while the Semi-Outlet exhibits large uncertainty in performance across the 50 optimization trials. Surprisingly, the over-parameterized Dist-Outlet shows promising results with high expected accuracy at Kama and Asmar (mean NSE (KGE) of 0.84 (0.71) and 0.90 (0.88), respectively) and comparable performance at many of the other sites. One exception is Gawardesh, where the Lump-Outlet outperforms the other model variants, although the reason for this is not immediately clear. Overall, the results indicate that any calibration based on basin outlet data should be used with substantial caution when predicting flows at interior basin sites.

After reviewing all of the calibration experiments, it becomes clear that the Semi-Pooled and Dist-Pooled calibrations provide more robust performance compared to the basin outlet calibrations due to their improved representation of internal hydrologic processes across the basin.

To further compare these calibration strategies against one another, we evaluate the variability in optimal parameters resulting from the 50 trials of the GA algorithm. Figure ~~8~~ 9 shows the coefficient of variation (CV) of Cmax (a parameter for the soil moisture account module) over the basin from all combinations of calibration approaches (the outlet and pooled) and 3 parameterization schemes. A clear pattern of increasing variability (higher uncertainty in Cmax) emerges as parameter complexity increases for both the outlet and pooled calibration strategies. That is, the semi- and fully-distributed parameterizations lead to significantly variable parameter sets that produce similar representations of the observed basin response. Figure ~~8~~ 9 also suggests that the equifinality can be alleviated to an extent by pooling data across sites. The pooled calibration approaches consistently show lower variability in Cmax compared to the outlet calibration at the same level of parameter complexity. These results are relatively consistent across the remaining 14 HYMOD_DS parameters. The implications of parameter stability on streamflow projections under climate change is addressed in the next section.

5.4. Climate Change Projections of Streamflow with Uncertainty

Here we explore how projections of future water availability and flood risk under climate change are influenced by the choice of calibration approach. For the Kabul River basin, the CMIP5 GCM projections of monthly total precipitation and mean temperature are shown in Figure S5. According to the CMIP5 ensemble, precipitation projections show no clear trend; the average precipitation change in monthly total precipitation fluctuates between -10mm and 10mm. On the other hand, temperature clearly shows an upward trend for both radiative forcing scenarios. The average changes in annual temperature are +2.2^{o}C and +2.8^{o}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.

We first examine average monthly streamflow estimates ~~climatology~~ across four calibration strategies: the Semi-Pooled and Dist-Pooled (most promising calibration strategies), as well as the Lump-Outlet (as a baseline) and Dist-Outlet (the best outlet calibration strategy). Figure ~~9~~ 10 shows the monthly streamflow ~~predictions~~ estimates for the historical period with ~~and the 2050s under the RCP 4.5 and 8.5 scenarios. The~~ the whisker bars indicating~~e~~ 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. For the historical time period, all calibration schemes match the observed monthly streamflow ~~climatology~~ at Dakah well, but monthly streamflow is underestimated in most of months at Kama and Asmar under the basin outlet calibrations, particularly by the Lump-Outlet. The historical monthly streamflow estimates ~~streamflow climatology~~ from the outlet calibration strategies also tends to be highly uncertain for the months of June, July, August, and September, especially compared to the SemiPool and DistPool.

Under future climate projections for the 2050s, the four calibration strategies show similar changes in ~~climatology~~ monthly streamflow at Dakah, but the magnitudes of change are somewhat different. All calibration strategies suggest reduction in streamflow for June, July, and August under both RCP4.5 and RCP8.5 scenarios. Also, the peak monthly flow, which occurred in June or July in the historical period, is shifted to May at Dakah. However, the Lump-Outlet predicts less reduction of flow in June and July and a greater reduction in August and September as compared to the other three calibrations. Considering that all calibration schemes had similar levels of good performance at this site for both calibration and validation periods, it is notable that they project future streamflow ~~climatology~~ somewhat differently.

Future monthly streamflow predictions ~~climatology~~ at Kama and Asmar vary widely between the four calibration schemes, mostly an artifact of their historic differences (Figure ~~9~~10). Streamflow projections under the outlet calibration strategies tend to show large uncertainties at these two sites, particularly the Lump-Outlet calibration. For three months, July through September, the outlet calibration and pooled calibration strategies provide substantially different insights about future water availability at Kama and Asmar. The outlet calibrations suggest less water with large uncertainties for those months as compared to the pooled calibrations. At Kama, the pooled calibrations suggest significant changes in the pattern of peak monthly flow timing under both RCP scenarios; instead of having a clear peak in July, streamflow from May to August show similar amounts of water.

To further understand the sources of uncertainty in future water availability, we evaluate the separate and joint influence of uncertainties in parameter estimation and future climate on seasonal streamflow projections across all calibration schemes. Figure ~~10~~ 11 represents the uncertainty of wet and dry seasonal streamflow at Dakah from three sources: 1) ~~parameter~~ calibration uncertainty across the 50 trials, with future climate uncertainty averaged out for each trial, 2) future climate uncertainty across the 36 projections, with ~~parameter~~ calibration uncertainty averaged out across the 50 trials, and 3) the combined uncertainty across all 1800 (50×36) simulations. The results suggest somewhat surprisingly that uncertainty reduction can be expected as parameter complexity increases, and less surprisingly, by applying pooled calibration approaches. Another clear point is that the uncertainty resulting from different climate change scenarios substantially outweighs that from ~~parameter~~ calibration uncertainty.

Up to this point, there has been little difference between the Semi-Pooled and Dist-Pooled model variants. These two versions were further analyzed with respect to extreme streamflow to see if distinguishing characteristics emerge. It has been demonstrated that clear gains in predicting peak flows from distributed models are noticeable (Reed et al., 2004) and spatial variability in model parameters significantly influence the runoff behavior (Brath and Montanari, 2000; Pokhrel and Gupta, 2011). The spatial variability of optimal parameters derived from the Semi-Pooled and Dist-Pooled ~~was~~ is shown in Figure S6, with larger variability across all parameters for the Dist-Pooled than for the Semi-Pooled. To understand the effects of ~~parameter~~ spatial variability and calibration uncertainty of parameters on extreme event estimation, the ~~100-year flood~~100-year daily flood event was calculated under the Semi-Pooled and Dist-Pooled for each of the 50 historic simulations and 1800 future simulations across both RCP scenarios. 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. ~~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.~~ Projections of the ~~100-year flood~~100-year daily flood, estimated using a Log-Pearson type III distribution fit to annual peaks of 30 years, differ somewhat between the Semi-Pooled and Dist-Pooled (Figure ~~11~~12). At 3 validation sites, extreme floods are consistently larger under the Semi-Pooled than the Dist-Pooled, and the mean difference in the ~~100-year flood~~100-year daily flood estimate between the two calibration approaches grows between the historic runs and the RCP 4.5 and 8.5 scenarios. This suggests that the flood-generation process is fundamentally different between the two parameterizations, with the Semi-Pooled formalization magnifying the effect of climate change on extremes. Furthermore, there is substantially more uncertainty in the ~~100-year flood~~100-year daily flood estimate under the Semi-Pooled. Figure ~~11~~ 12 shows the combined uncertainty across both climate projections and calibrations, but this uncertainty is broken down further in Figure ~~12~~13. Similar to Figure ~~10~~11, 3 sources of uncertainty are evaluated for the ~~100-year flood~~100-year daily flood, including ~~parameter~~ calibration uncertainty alone, climate projection uncertainty alone, and their combined effect. For both the Semi-Pooled and Dist-Pooled, ~~parameter~~ calibration uncertainty has a smaller influence than projection uncertainties, and for all sites, the Dist-Pooled has a smaller uncertainty range than the Semi-Pooled, even for ~~parameter~~ calibration uncertainty alone. This was a truly surprising result, given the parametric freedom in the Dist-Pooled model and the fact that no daily data was ever used in the calibration of either model. It appears that a lack of model parsimony does not necessarily lead to greater uncertainty in model simulations under different climate conditions, somewhat counter to what would be expected of over-fit models. One possible reason for this result would be if increased parametric freedom somehow offset the effects of structural deficiencies in the model. However, further research is needed to investigate this issue.

6. Discussion and Conclusion

In this study we examined a variety of calibration experiments to better understand the benefits and costs associated with different calibration choices for a complex, distributed hydrologic model in a data-scarce region. The goal of these experiments was to provide insight regarding the use of multisite data in calibration, the effects of parameter complexity, and the challenges of using limited data for distributed model calibration, all in the context of projecting future streamflow under climate change.

This study tested two multi-site calibration strategies, the stepwise and pooled approaches, finding that the pooled approach using all data simultaneously provides improved calibration results. This suggests that small sacrifices of model performance at certain sites can improve and stabilize basin-wide performance. The pooled calibration substantially improves with increasing parameter complexity at the calibration sites, but ~~the~~ similar streamflow predictions at the validation sites between the semi-distributed and distributed pooled calibrations were found, suggesting over-fitting of the model from the fully distributed conceptualization. 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). 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.

Calibration only based on data at the basin outlet is all too common in hydrologic model applications and is sometimes considered comparable to multisite calibrations even for predictions at interior gauges (Lerat et al., 2012). In contrast, others have reported improvements in interior flow predictions by using internal flow measurements (Anderson et al., 2001; Wang et al., 2012; Boscarello et al., 2013). This is in agreement with the finding from this study, demonstrating the superiority of the pooled calibration approach to the basin outlet calibration in terms of its ability to represent interior hydrologic response correctly. This study shows the danger in relying on an outlet calibration for interior flow prediction.

~~It is difficult to expect hydrologic models to yield reliable streamflow estimates at interior locations of a watershed when calibration is only based on data at the basin outlet, yet this is all too common in hydrologic model applications. The pooled calibration approach is superior to the basin outlet calibration in terms of its ability to represent interior hydrologic response correctly. This study shows the danger in relying on an outlet calibration for interior flow prediction.~~

It was shown that caution is needed when using an outlet calibration approach for streamflow predictions under future climate conditions. This study showed that the basin outlet calibration can lead to projections of mid-21st century streamflow that deviate substantially from projections under multisite calibration strategies. From the test of implications of the pooled calibration in the context of climate change, it was found that applying the pooled calibration with semi-distributed and distributed parameter formulations showed clear gains in reducing uncertainties in predictions of monthly and seasonal water availability as compared to the basin outlet calibrations. Surprisingly, increased parameter complexity in the calibration strategies ~~does~~ did not increase the uncertainty in streamflow projections, even though parameter equifinality ~~does~~

did emerge. The results suggest that increased (excessive) parameter complexity does not always lead to increased uncertainty if structural uncertainties in the model are present.

The semi-distributed pooled and distributed pooled calibrations are very similar for monthly streamflow projections, yet differ in their projections of extreme flows in part due to their differences in the spatial variability of optimal parameters, with the distributed pooled calibration showing less uncertainty for ~~100 year flood~~100-year daily flood events. We evaluated the separate and joint influence of uncertainties in parameter estimation and future climate on projections of seasonal streamflow and ~~100 year flood~~100-year daily flood across calibration 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).

In addition to the uncertainties surrounding model parameters and future climate explored in this study, there is also significant uncertainty in streamflow projections stemming from structural differences between applied hydrologic models, which can be especially pertinent where robust calibration is hampered by the scarcity of data (Exbrayat et al., 2014). Further, the residual error variance of hydrologic model simulations would increase the effects of hydrologic model uncertainty as compared to that of the climate projections (Steinschneider et al., 2014). These issues need to be addressed in future work for exploring a comprehensive uncertainty assessment of climate change risk for poorly monitored hydrologic systems.

Successful automatic calibration algorithms for hydrologic models are based primarily on global optimization algorithms that are computationally expensive and require a large number of function evaluations (Kuzmin et al., 2008). Although the speed and capacity of computers have increased multi-fold in the past several decades, the time consumed by running hydrological models (especially complex, physically based, distributed hydrological models) is still a concern for hydrology practitioners. A single trial of parameter optimization of HYMOD_DS associated with 100,000 runs can take 28 days on a single processor (Figure S7). Accordingly, ~~T~~the use of high performance computing power was essential in this study to better understand the implications of different calibration choices and their associated uncertainty for streamflow projections. Enhanced data with high spatial and temporal resolution are increasingly available from remote sensing and satellite products. In the future, remote sensing and satellite information can be integrated into calibration approaches to develop more robust estimates of spatially distributed parameter values, ~~for~~ enabling internal consistency of distributed hydrological modeling. Significant progress has been made toward this end (Tang et al., 2009; Khan et al., 2011; Thirel et al., 2013). Future work will consider using ~~advanced computing techniques~~high performance computing power (e.g. Laloy and Vrugt, 2012; Zhang et al., 2013) to understand how such information can enhance the hydrologic simulation at ungaged sites and reduce the ~~parameter~~ calibration uncertainty of distributed hydrologic models in data-scarce regions.

Acknowledgements

The authors are grateful to Dr. Efrat Morin, Dr. Andreas Efstratiadis, and one anonymous reviewer for their constructive suggestions for improving this manuscript.

This research is funded by a World Bank grant: “Hydro-Economic Modeling for Brahmaputra and Kabul River.” The views expressed in this paper are those of the authors and do not necessarily reflect the views of the World Bank.

We acknowledge the use of the supercomputing facilities managed by the Research Computing department at the University of Massachusetts.

References

Ahmad, S.: Towards Kabul Water Treaty: Managing Shared Water Resources - Policy Issues and Options, Karachi, Pakistan, 2010.

Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, J HYDROL, 298, 112-135, 2004.

Anderson, J., Refsgaard, J. C., and Jensen, K. H.: Distributed hydrological modeling of the Senegal river basin - model construction and validation, J HYDROL, 247(3-4), 200-214, 2001.

Bandaragoda, C., Tarboton, D. G., and Woods, R.: Application of TOPNET in the distributed model intercomparison project, J HYDROL, 298, 178-201, 2004.

Beven, J. K.: Rainfall-Runoff Modelling: The Primer, 2nd Edition, Wiley-Blackwell, Chichester, 2012.

Beven, K.: How far can we go in distributed hydrological modelling?, HYDROL EARTH SYST SC, 5(1), 1-12, 2001.

Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J HYDROL, 249, 11-29, 2001.

Boscarello, L., Ravazzani, G., and Mancini, M.: Catchment multisite discharge measurements for hydrological model calibration, Procedia Environmental Sciences, 19, 158-167, 2013.

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., and Montanari, A.: Validation of hydrological models: conceptual basis, methodological approaches and a proposal for a code of practice, Phys. Chem. Earth, 42-44, 70-76, 2012

~~Bolch, T., Kulkarni, A., Kaab, A., Hyggel, C., Paul, F., Cogley, J. G., Frey, H., Kargel, J. S., Fujita, K., Scheel, M., Bajracharya, S., and Stoffel, M.: The State and Fate of Himalayan Glaciers, SCIENCE, 336, 310-314, 2012~~

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

Boyle, D. P.: Multicriteria calibration of hydrologic models, Ph.D. thesis, Department of Hydrology and Water Resources Engineering, The University of Arizona, USA, 2001.

Brath, A., and Montanari, A.: The effects of the spatial variability of soil infiltration capacity in distributed flood modelling, HYDROL PROCESS, 14, 2779-2794, 2000.

Brath, A., Montanari, A., and Toth, E.:. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model, J HYDROL, 291, 232-253, 2004.

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

Cao, W., Bowden, W. B., Davie, T., and Fenemor, A.: Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability, HYDROL PROCESS, 20, 1057-1073, 2006.

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.

Cole, S. J., and Moore, R. J.: Hydrological modelling using raingauge- and radar-based estimators of areal rainfall, J HYDROL, 358(3-4), 159-181, 2008.

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

~~Dyurgerov, M. B., and Meier, M. F.: Glaciers and the changing Earth system: a 2004 snapshot, Boulder: Institute of Arctic and Alpine Research, University of Colorado, Boulder, 2005.~~

Eckhardt, K., Fohrer, N., and Frede, H. G.: Automatic model calibration, HYDROL PROCESS, 19, 651-658, 2005.

Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, HYDROLOG SCI J, 55(1), 58-78, 2010.

Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., and Koutsoviannis, 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.

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.

Frances, F., Velez, J. I., and Velez, J. J.: Split-parameter structure for the automatic calibration of distributed hydrological models, J HYDROL, 332(1-2), 226-240, 2007.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, HYDROL EARTH SYST SC, 17, 149-161, 2013.

Grinsted, A.: An estimate of global glacier volume, The Cryosphere, 7, 141-151, 2013.

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.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J HYDROL, 377, 80-91, 2009.

Hamon, W. R.: Estimating potential evapotranspiration, J HYDR ENG DIV-ASCE, 87, 107-120, 1961.

Hewitt, K., Wake, C. P., Young, G. J., and David, C.: Hydrologicl investigations at Biafo glacier, Karakoram Himalaya, Pakistan: An important source of water for the Indue River, ANN GLACIOL, 13, 103-108, 1989.

~~Immerzeel, W. W., Pellicciotti, F., and Bierkens, M. F. P.: Rising river flows throughout the twenty-first century in two Himalayan glacierized watersheds, NAT GEOSCI, 6, 742-745, 2013.~~

~~Immerzeel, W. W., van Beek, L. P. H., and Bierkens, M. F. P.: Climate Change Will Affect the Asian Water Towers, SCIENCE, 328(5984), 1382-1385, 2010.~~

Immerzeel, W. W., van Beek, ~~J~~L. P. H., Konz, M., Shrestha, A. B., and Bierkens, M. F. P.: Hydrological response to climate change in a glacierized catchment in the Himalayas, CLIMATIC CHANGE, 110, 721-736, 2012.

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

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available at: http://srtm.csi.cgiar.org, last access: 2 January 2015, 2008.

Khakbaz, B., Imam, B., Hsu, K., and Sorooshian, S.: From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models, J HYDROL, 418-419, 61-77, 2012.

Khan, S. I., Yang, H., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policell, F., Habib, S., and Irwin, D.: Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: Implications for hydrologic prediction in ungauged basins, IEEE T GEOSCI REMOTE, 49, 85-95, 2011.

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

Kinouchi, T., Liu, T., Mendoza, J., and Asaoka, Y.: Modeling glacier melt and runoff in a high-altitude headwater catchment in the Cordillera Real, Andes, HYDROL EARTH SYST SC, 10, 13093-13144, 2013.

Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, WATER RESOUR RES, 48(3), W03520, 2012.

Konz, M., and Seibert, J.: On the value of glacier mass balances for hydrological model calibration, J HYDROL, 385(1-4), 238-246, 2010.

Koren, V., Reed, S., Smith, M., Zhang, Z., and Seo, D. J.: Hydrology laboratory research modeling system (HL-RMS) of the US national weather service, J HYDROL, 291, 297-318, 2004.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, WATER RESOUR RES, 49(1), 360-379, 2013.

Kuzmin, V., Seo D., and Koren V.: Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, J HYDROL, 353, 109-128, 2008.

Laloy, E., and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, WATER RESOUR RES, 48(1), W01526, 2012.

Latham, J., Cumani, R., Rosati, I., and Bloise, M.: Global Land Cover SHARE (GLC-SHARE) database Beta-Release Version 1.0, available at: http://www.glcn.org/databases/lc_glcshare_en.jsp, last access: 2 January 2015, 2014.

Leavesley, G. H., Hay, L. E., Viger, R. J., and Markstrom, S. L.: Use of Priori Paramter-Estimation Methods to Constrain Calibration of Distributed-Parameter Models, WATER SCI APPL, 6, 255-266, 2003.

Legates, D. R., and Willmott, C. J.: Mean seasonal and spatial variability in gauge-corrected, global precipitation, INT J CLIMATOL, 10(2), 111-127, 1990.

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.

Li, X., Weller, D. E., and Jordan, T. E.: Watershed model calibration using multi-objective optimization and multi-site averaging, J HYDROL, 380(3-4), 277-288, 2010.

Lohmann, D., Raschke, R., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, HYDROLOG SCI J, 43(1), 131-141, 1998.

Madsen, H.: Parameter estimation in distributed hydrologicl catchment modelling using automatic calibration with multiple objectives, ADV WATER RESOUR, 26(2), 205-216, 2003.

~~Molg, T., Maussion, F., and Scherer, D.: Mid-latitude westerlies as a driver of glacier variability in monsoonal High Asia, NATURE, 4, 68-73, 2014.~~

Moore, R. D.: Application of a conceptual streamflow model in a glacierized drainage basin, J HYDROL, 150(1), 151-168, 1993.

Moore, R. J.: The probability-distribted principle and runoff production at point and basin scales, HYDROLOG SCI J, 30(2), 273-297, 1985.

Nash, J. E.: The form of the instantaneous unit hydrograph, International Association of Science and Hydrology, 3, 114-121, 1957.

Nash, J. E., and Sutcliff, J. V.: River flow forecasting through conceptual models: Part 1. A discussion of priciples, J HYDROL, 10(3), 282-290, 1970.

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

Palazzi, E., von Hardenberg, J., and Provenzale, A.: Precipitation in the Hindu-Kush Karakoram Himalaya: Observations and future scenarios, J GEOPHYS RES, 118(1), 85-100, 2013.

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.

Pfeffer, T. W., Arendt, A. A., Bliss, A., Bolch, T., Cogley J. G., Gardner, A. S., Hagen, J. O., Hock R., Kaser, G., Kienholz, C., Miles E. S., Moholdt, G., Molg, N., Paul, F., Radic, V., Rastner, P., Raup, B. H., Rich, J., Sharp, M. J., and The Randolph Consortium: The Randolph Glacier Inventory, J GLACIOL, 60(221), 537-552, 2014.

Pokhrel, P., and Gupta, H. V.: On the use of spatial regularization strategies to improve calibration of distributed watershed models, WATER RESOUR RES, 46, W01505, 2010.

Pokhrel, P., and Gupta, H. V.: On the ability to infer spatial catchment variability using streamflow hydrographs, WATER RESOUR RES, 47, W08534, 2011.

Radic, V., Bliss, A., Beedlow, A. C., Hock, R., Miles, E., and Cogley, J. G.: Regional and global projections of twenty-first century glacier mass changes in response to climate scenarios from global climate models, CLIM DYNAM, 42(1-2), 37-58, 2014.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. J., and DMIP Participants: Overall distributed model intercomparison project results, J HYDROL, 298, 27-60, 2004.

Remesan, R., Bellerby, T., and Frostick, L.: Hydrological modelling using data from monthly GCMs in a regional catchment, HYDROL PROCESS, 28(8), 3241-3263, 2013.

Safari, A., De Smedt, F., and Moreda, F.: WetSpa model application in the Distributed Model Intercomparison Project (DMIP2), J HYDROL, 418-419, 77-89, 2012.

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.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project - Phase 2: Motivation and design of the Oklahoma experiments, J HYDROL, 418-419, 3-16, 2012.

Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, J HYDROL, 298, 4-26, 2004.

Smith, M., Koren, V., Zhang, Z., Moreda, F., Cui, Z., Cosgrove, B., Mizukami, N., Kitzmiller, D., Ding, F., Reed, S., Anderson, E., Schaake, J., Zhang, Y., Andreassian, V., Perrin, C., Coron, L., Valery, A., Khakbaz, b., Sorooshian, S., Behrangi, A., Imam, B., Hsu, K. L., Todini, E., Coccia, G., Mazzetti, C., Andres, E. O., Frances, F., Orozco, I., Hartman, R., Henkel, a., Fickenscher, P., and Staggs, S.: The distributed model intercomparison project - Phase 2: Experiment design and summary results of the western basin experiments, J HYDROL, 507, 300-329, 2013.

Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., and Cannon, A. J.: Coupled modelling of glacier and streamflow response to future climate scenarios, WATER RESOUR RES, 44(2), W02422, 2008.

Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H.: Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change, WATER RESOUR RES, 48(11), W11525, 2012.

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.

Talyor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, B AM METEOROL SOC, 93, 485-498, 2012.

Tang, Q., Gao, H., Lu, H., and Lettenmaier, D. P.: Remote sensing: hydrology, PROG PHYS GEOG, 33, 490-509, 2009.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter, Remote Sensing, 5, 5825-5850, 2013

USDA-NRCS: FAO-UNESCO Soil Map of the World, available at: http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013, last access: 2 January 2015, 2005.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, STOCH ENV RES RISK A, 23, 1011-1026, 2008.

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.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperical College Press, London, 2004.

Wang, Q. J.: The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models, WATER RESOUR RES, 27(9), 2467-2471, 1991.

Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y., and Yao, A.: Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China, HYDROL EARTH SYST SC, 16, 4621-4632, 2012.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, HYDROL PROCESS, 19(16), 3201-3219, 2005.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs, CLIMATIC CHANGE, 62(1-3), 189-216, 2004.

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.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges, B AM METEOROL SOC, 93(9), 1401-1415, 2012.

Yu, W., Yang, Y. C. E., Savitsky, A., Alford, D., Brown, C., Wescoat, J., Debowicz, D., and Robinson, S.: The Indus Basin of Pakistan: The Impacts of Climate Risks on Water and Agriculture, World Bank, Washington DC, 2013.

Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R. C., Sadeghi, A., Thomson, A. M., Sahajpal, R., Srinivasan, R., and Arnold, J. G.: Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python, ENVIRON MODELL SOFTW, 46, 208-218, 2013.

Zhang, X., Srinivasan, R., and Van Liew, M.: Multi-site calibration of the SWAT model for hydrologic modeling, T ASABE, 51(6), 2039-2049, 2008.

Zhu, C., and Lettenmaier, D. P.: Long-Term Climate and Derived Surface Hydrology and Energy Flux Data for Mexico: 1925-2004, J CLIMATE, 20, 1936-1946, 2007.

Tables

Table 1 Streamflow gaging stations in the Kabul River basin.

~~Name~~	~~River~~	~~Station ID~~	~~Drainage Area (km^2)~~	~~Data Period~~	
				~~Start~~	~~End~~
~~Dakah~~	~~Kabul~~	~~USGS 341400071020000/ GRDC 2240100~~	~~67,370~~	~~1968.2~~	~~1980.7~~
~~Pul-i-Kama~~	~~Kunar~~	~~USGS 342800070330000/ GRDC 2240200~~	~~26,005~~	~~1967.1~~	~~1979.9~~
~~Asmar~~	~~Kunar~~	~~USGS 345300071100000~~	~~19,960~~	~~1960.3~~	~~1971.9~~
~~Chitral~~	~~Kunar~~	~~GRDC 2340200~~	~~11,396~~	~~1978.1~~	~~1981.12~~
~~Chaghasarai~~	~~Pech~~	~~USGS 345400071080000/ GRDC 2240210~~	~~3855~~	~~1960.2~~	~~1979.2~~
~~Gawardesh~~	~~Landaisin~~	~~USGS 352300071320000~~	~~3,130~~	~~1975.5~~	~~1978.6~~
~~Daronta~~	~~Kabul~~	~~USGS 342800070220000/ GRDC 2240101~~	~~34,375~~	~~1959.10~~	~~1964.9~~

~~Dual station ID for stations archived in both USGS and GRDC database~~

Data Source	**Station Name**	**River**	**Data Period**		**Physiographic Property**			**Basin Climate**		
			Start	**End**	**Drainage Area (km^2)**	**Glacier Area (%)**	**Mean Elev (m)**	**Mean Annual Prcp (mm)**	**Mean Annual Mean Temp (°C)**	**Mean Annual Flow (mm)**
USGS/ GRDC	Dakah	Kabul	1968/2	1980/7	67,370	2.9	2,883	418	7.7	282
USGS/ GRDC	Pul-i-Kama	Kunar	1967/1	1979/9	26,005	7.3	3,446	446	5.6	573
USGS	Asmar	Kunar	1960/3	1971/9	19,960	9.4	3,716	483	4.1	651
GRDC	Chitral	Kunar	1978/1	1981/12	11,396	14.4	4,126	518	2.1	698
USGS	Gawardesh	Landaisin	1975/5	1978/6	3,130	2.1	3,707	555	4.5	521
USGS/ GRDC	Chaghasarai	Pech	1960/2	1979/2	3,855	0.4	3,141	482	7.4	535

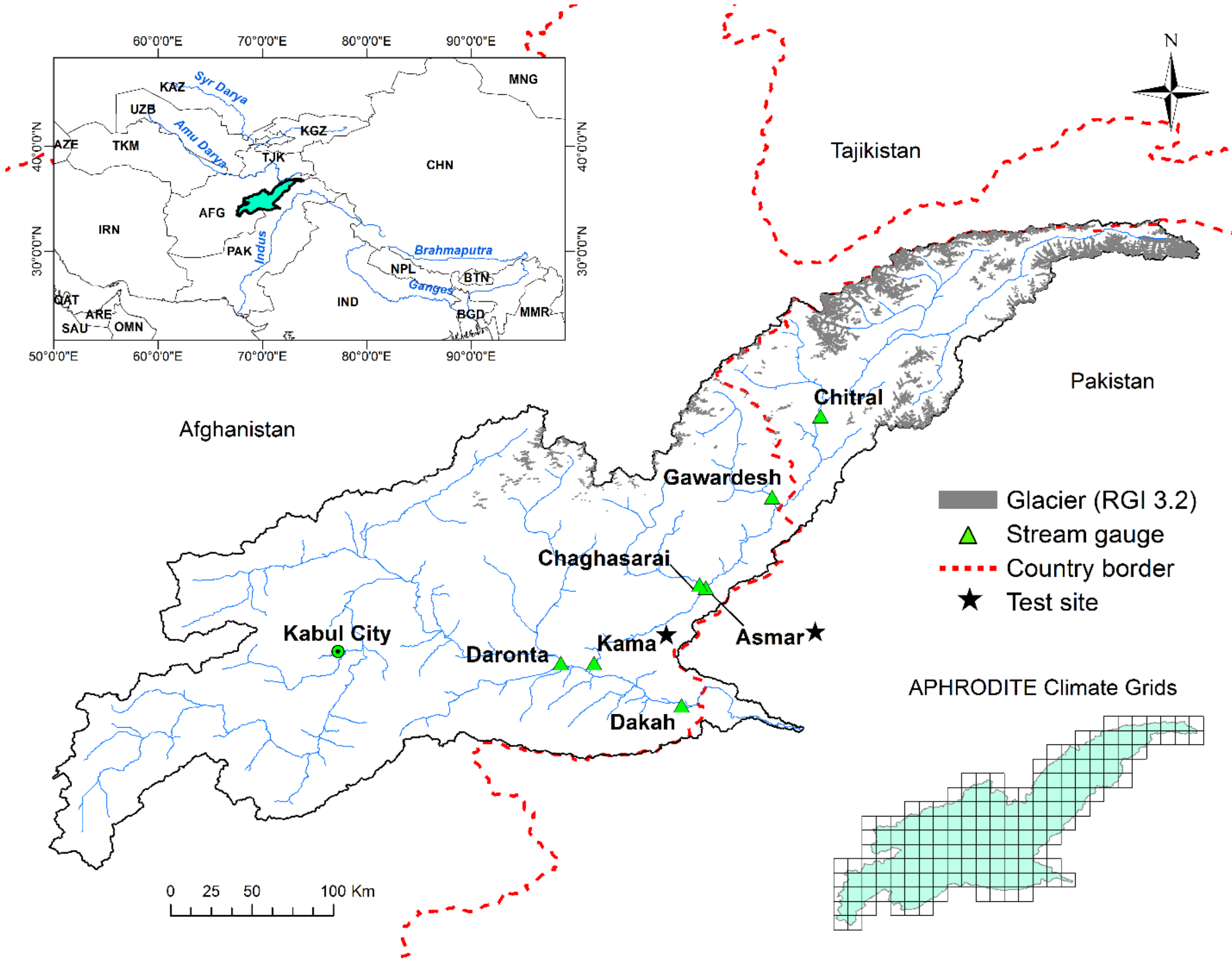
USGS/ GRDC	Daronta	Kabul	1959/10	1964/9	34,375	0.3	2,722	350	8.0	165

988

Table 2 HYMOD_DS parameters.

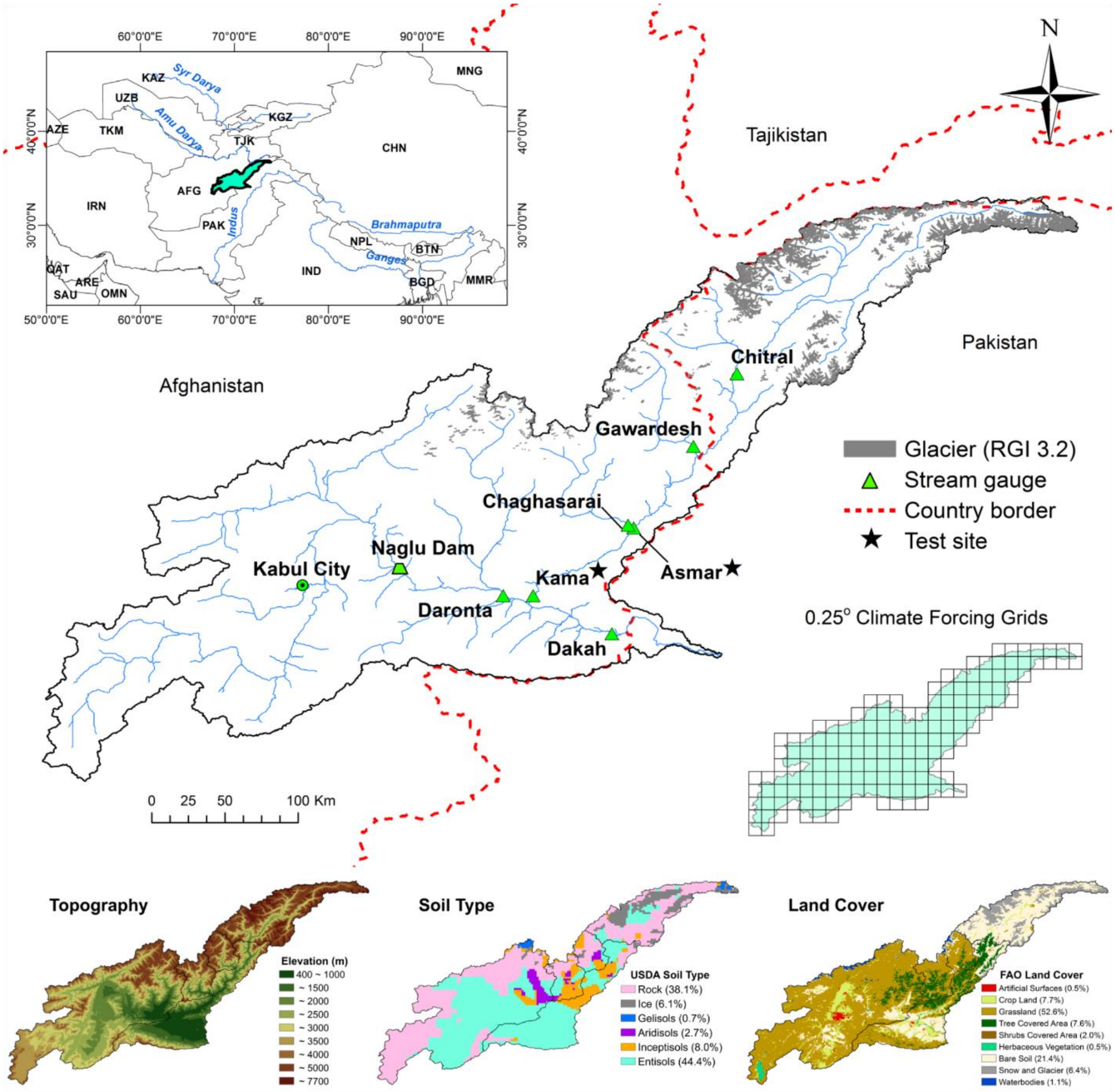
Parameter Name	Description	Feasible Range	
		Lower Bound	**Upper Bound**
Coeff	Hamon potential evapotranspiration coefficient	0.1	2
C_{max}	Maximum soil moisture capacity [mm]	5	1500
B	Shape for the storage capacity distribution function	0.01	1.99
α	Direct runoff and base flow split factor	0.01	0.99
K_s	Release coefficient of groundwater reservoir	0.00005	0.001
DDF_s	Degree day snow melt factor [$mm \cdot {}^{o}C \cdot day^{-1}$]	0.001	10
T_{th}	Snow melt temperature threshold [°C]	0	5
T_s	Snow/rain temperature threshold [°C]	0	5
r	Glacier melt rate factor	1	2
K_g	Glacier storage release coefficient	0.01	0.99
T_g	Glacier melt temperature threshold [°C]	0	5
N	Unit hydrograph shape parameter	1	99
K_q	Unit hydrograph scale parameter	0.01	0.99
Velo	Wave velocity in the channel routing [$m \cdot s^{-1}$]	0.5	5
Diff	Diffusivity in the channel routing [$m^2 \cdot s^{-1}$]	200	4000

993 **Figures**



994

995



996 Figure 1. Kabul River basin.

997

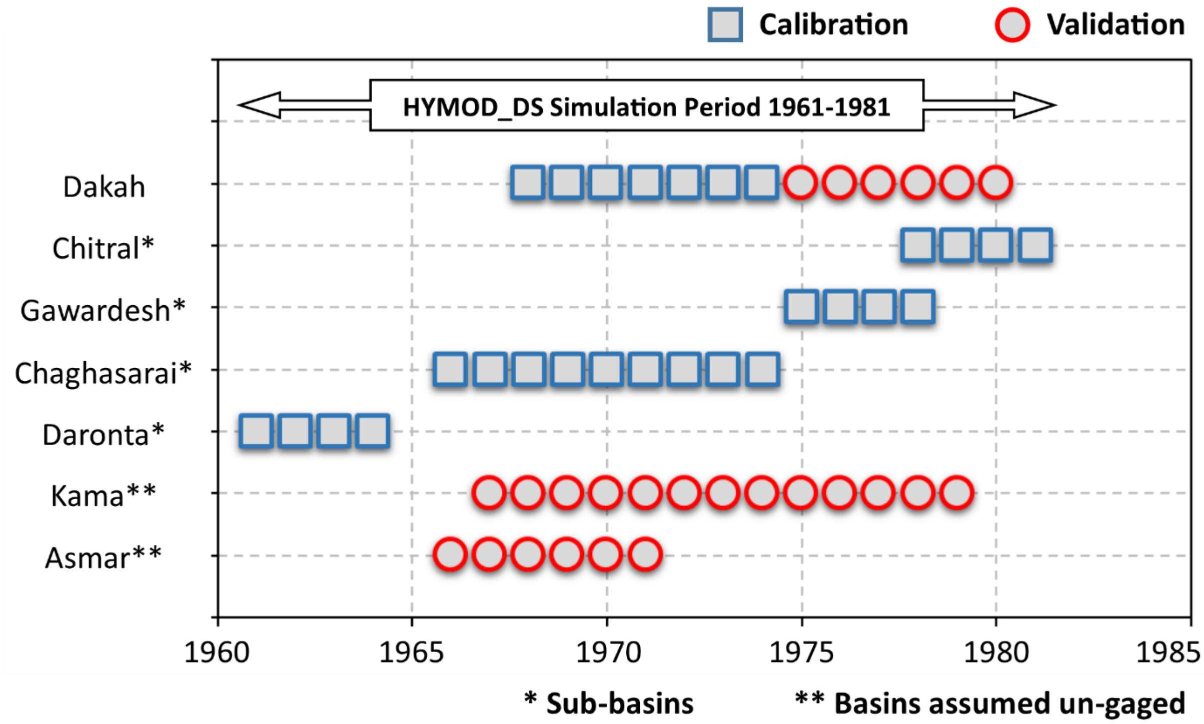


Figure ~~3~~2. Streamflow data usage for the model calibration and validation.

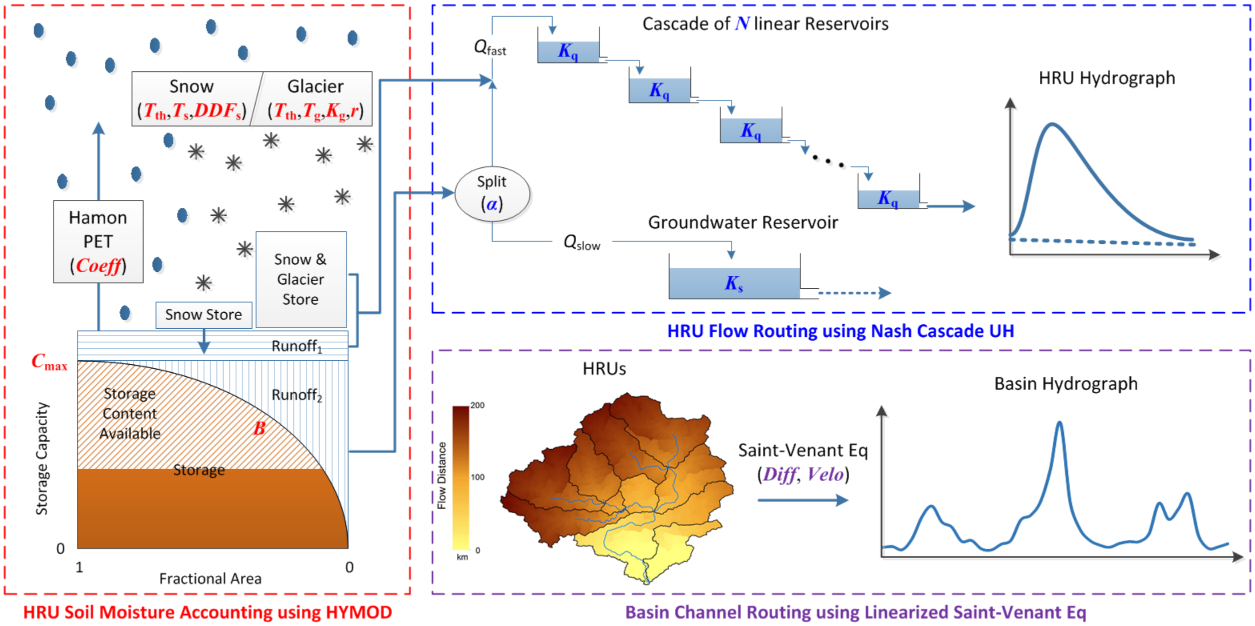


Figure ~~4~~3. Distributed version of HYMOD model (HYMOD_DS).

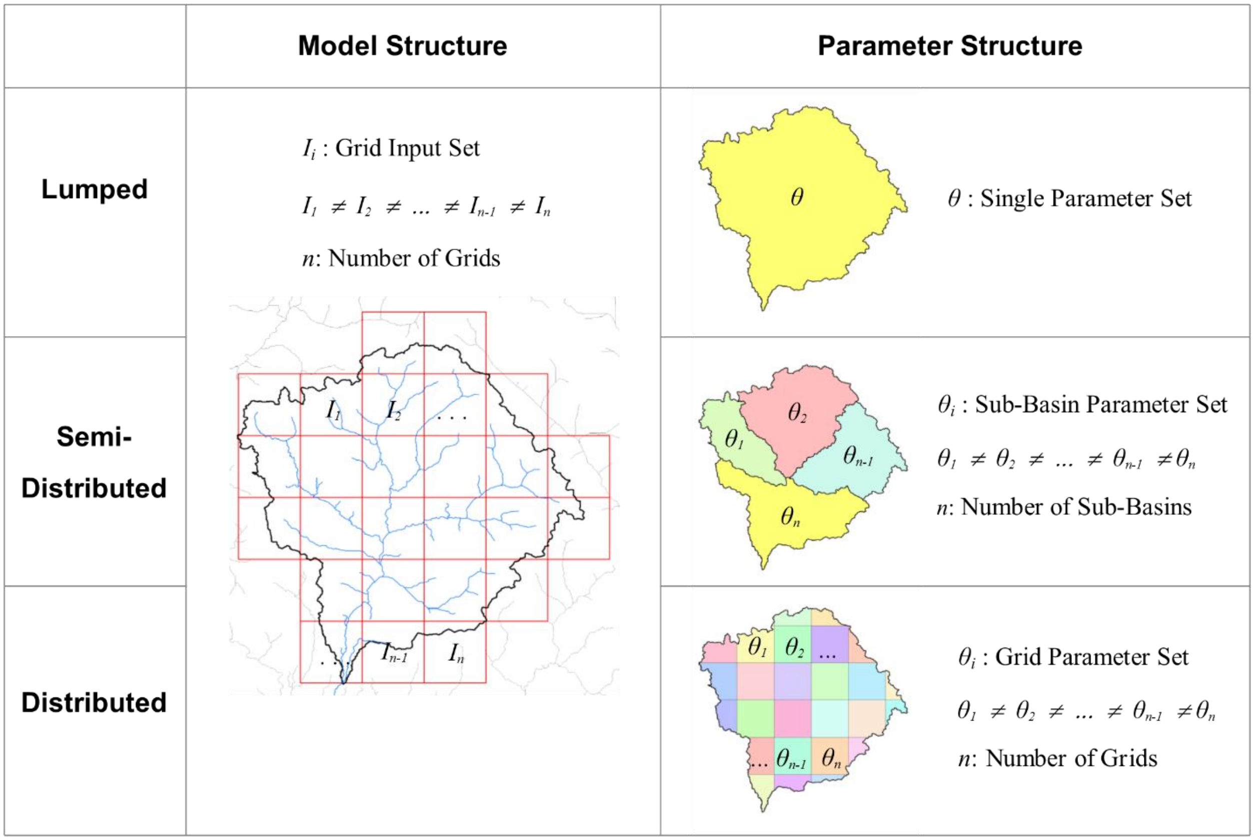


Figure ~~2~~4. Model structure based on climate input grids and three different parameterization concepts.

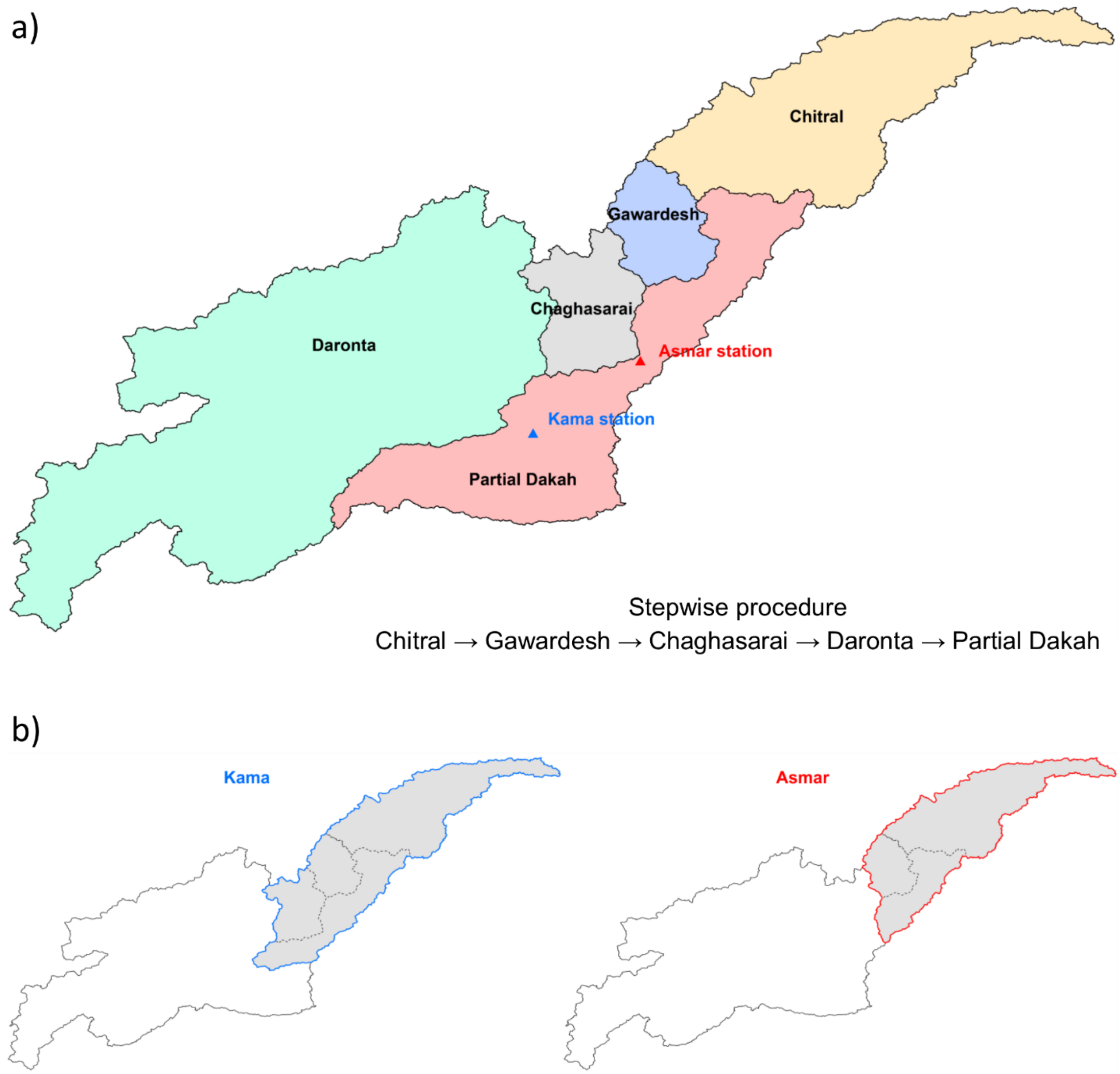


Figure 5. (a) Sub-basins corresponding to five gaging stations are used for the multisite calibrations. (b) Two sub-basins (Kama and Asmar) are assumed to be ungaged and used for evaluating the calibration approaches.

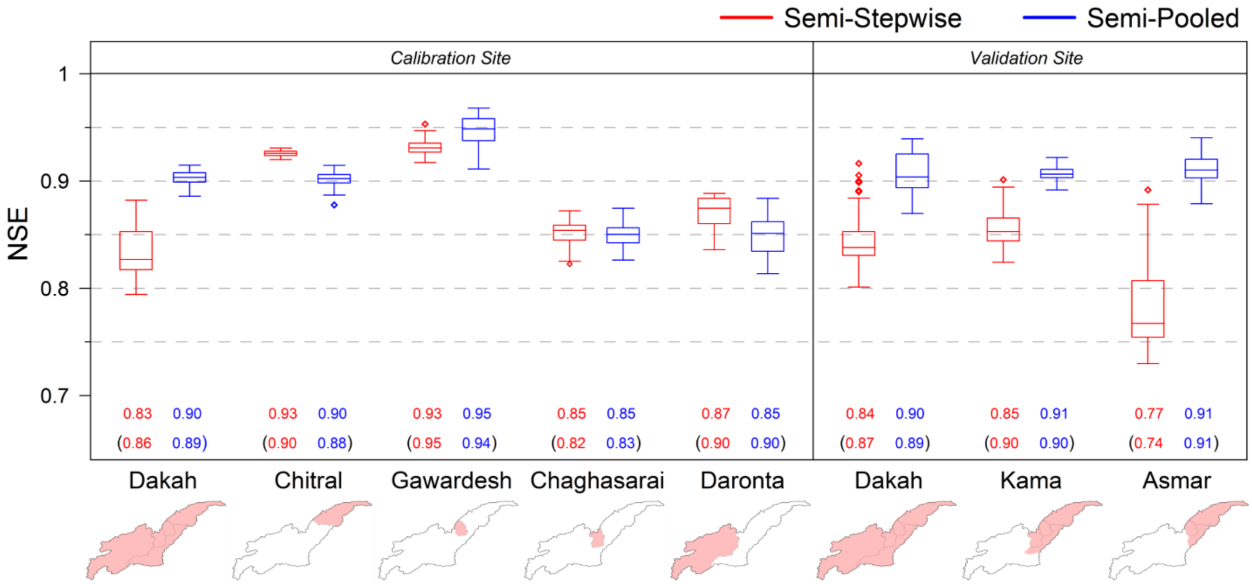


Figure ~~5~~6. Comparison of the stepwise and pooled calibrations under the semi-distributed parameterization. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

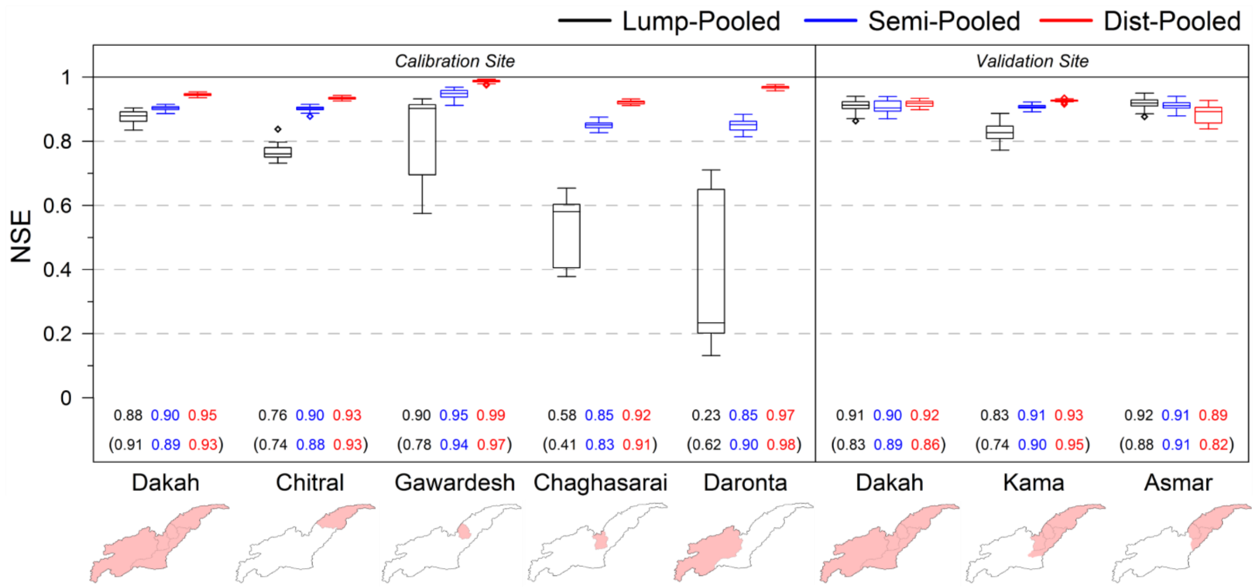


Figure ~~6~~7. Comparison of the pooled calibrations for the 3 parameterizations of lumped, semi-distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

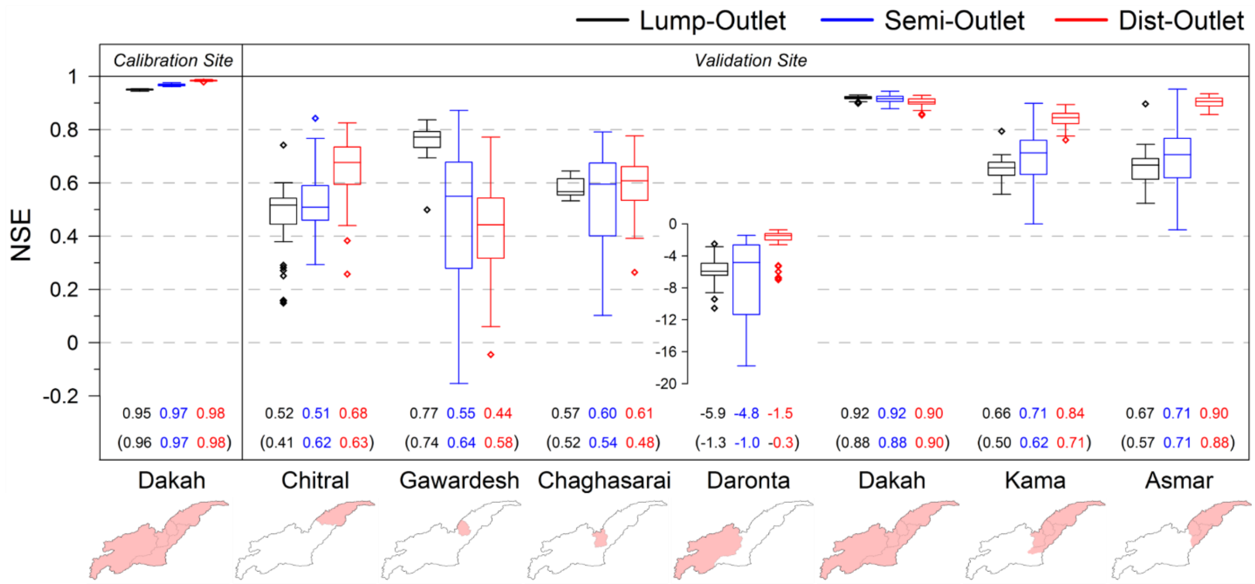


Figure ~~7~~8. Comparison of the basin outlet calibrations for the 3 parameterizations of lumped, semi-distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

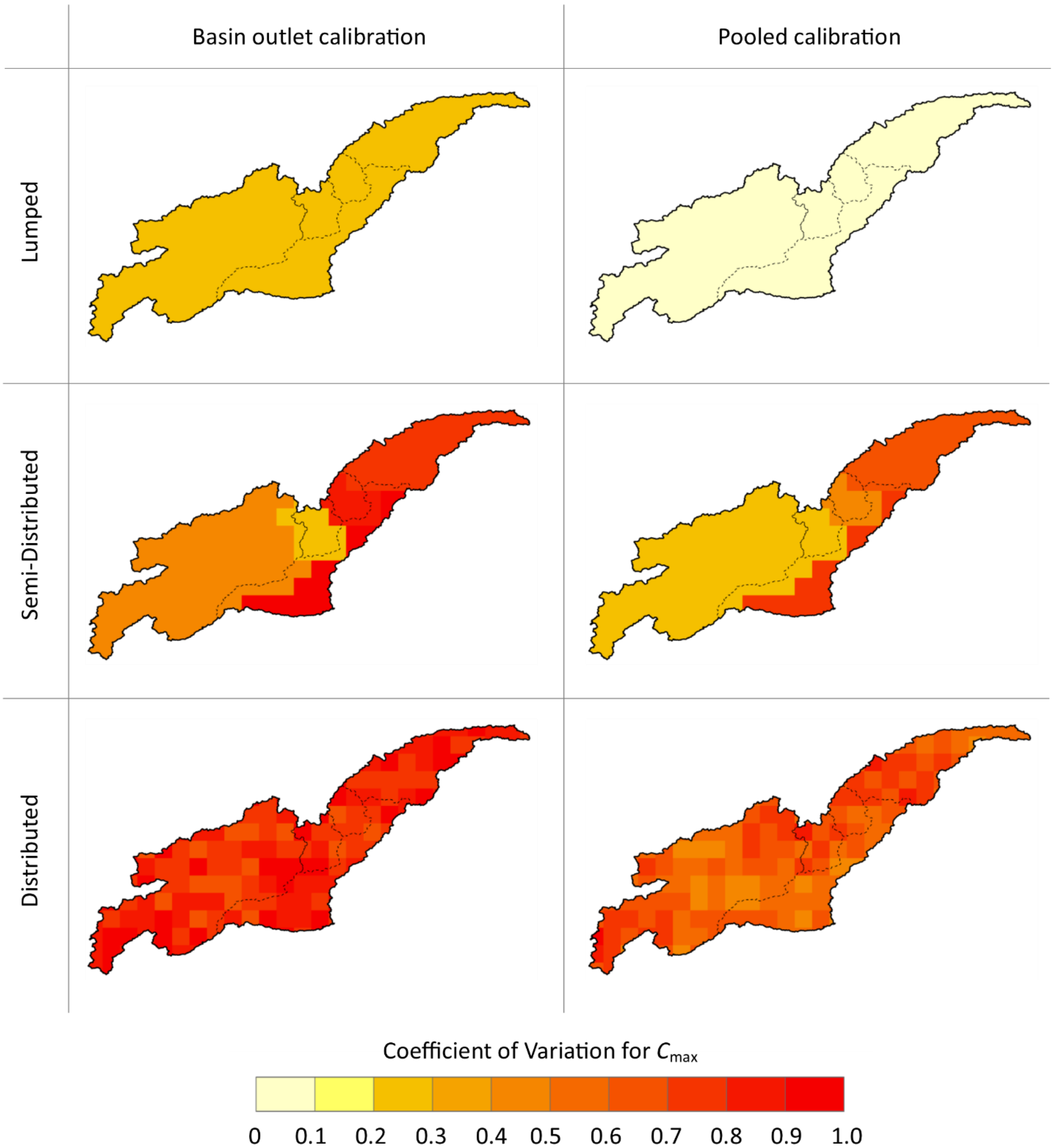


Figure ~~8~~9. Coefficient of variation (CV) of 50 optimal values of C_{max} (parameter for the soil moisture accounting module in the HYMOD_DS) from the basin outlet calibrations (left panel) and the pooled calibrations (right panel).

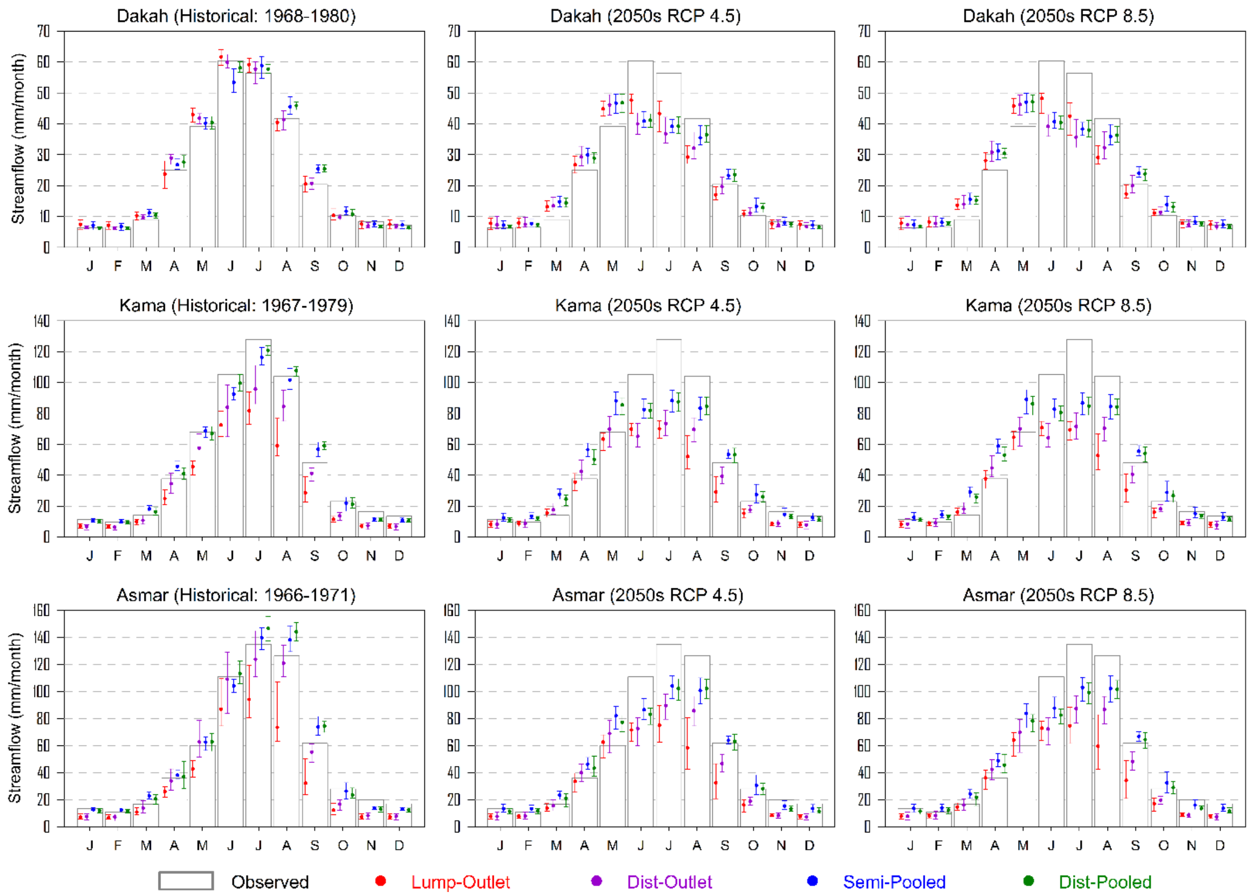
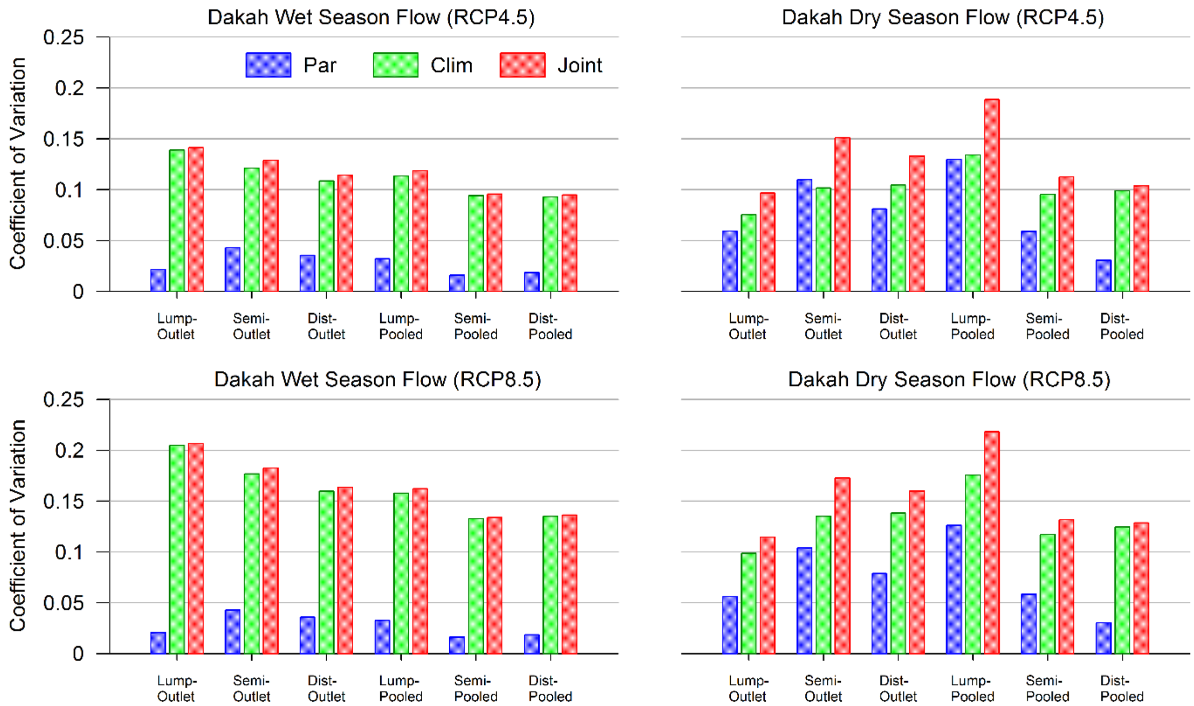


Figure ~~9~~10. Historical and 2050s average monthly streamflow ~~climatology~~ predictions at Dakah, Kama, and Asmar under 4 calibration strategies: Lump-Outlet, Dist-Outlet, Semi-Pooled, and Dist-Pooled. The error bars represent the streamflow ranges resulting from 50 trails of the HYMOD_DS calibration. For each of the 50 trials, the 2050s streamflow predictions are averaged over 36 GCM climate projections.



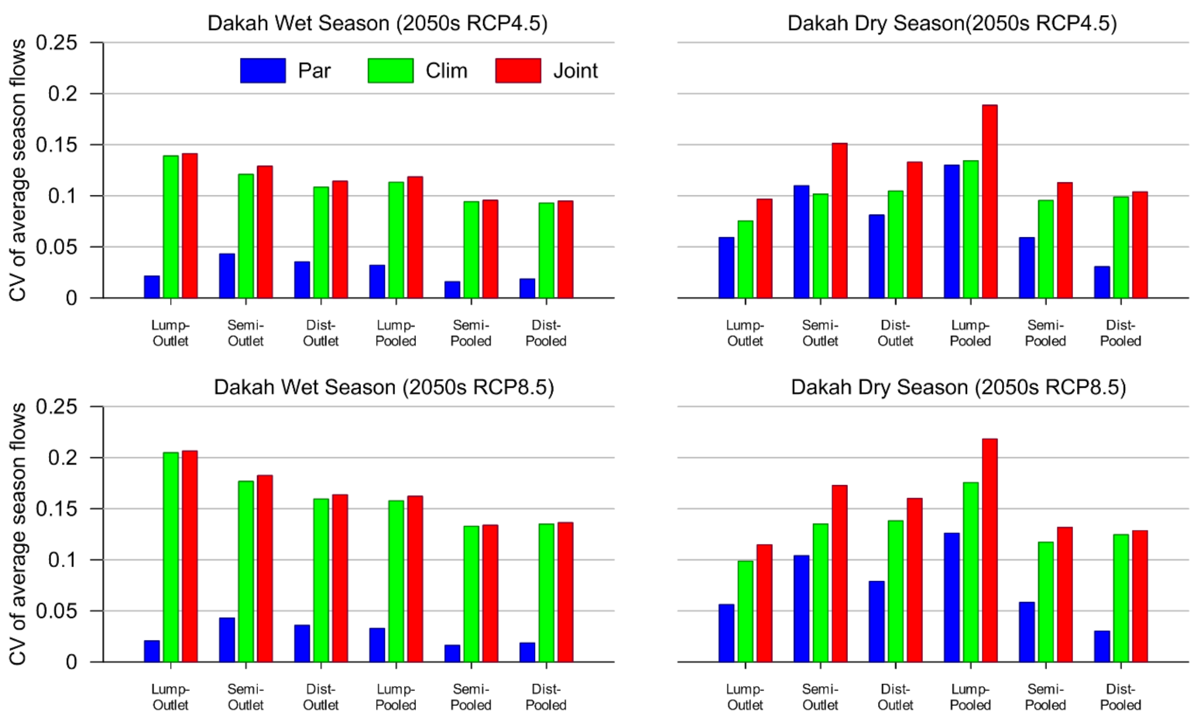


Figure ~~10~~11. Uncertainties in wet and dry season ~~2050s~~ average streamflow ~~climatology~~ predictions for 2050s ~~of wet and dry seasons~~ are derived from the basin outlet and pooled calibrations for Dakah. Uncertainties are evaluated by coefficient of variation (CV) of average season streamflow predictions. Three uncertainty sources are considered: ~~parameter~~ calibration uncertainty across 50 calibration trials (Par), climate uncertainty across GCM projections (Clim), and combined uncertainty (Joint).

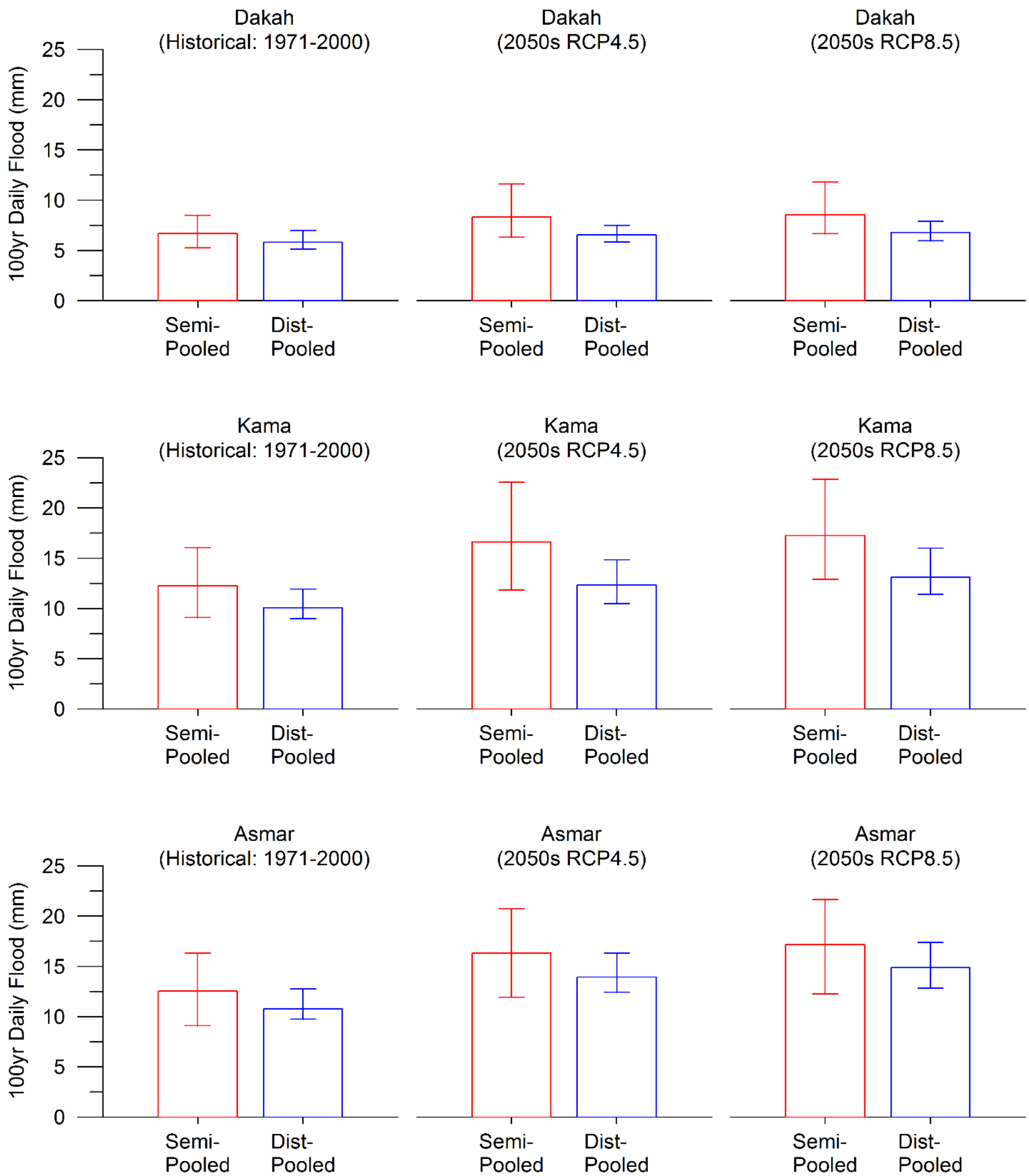
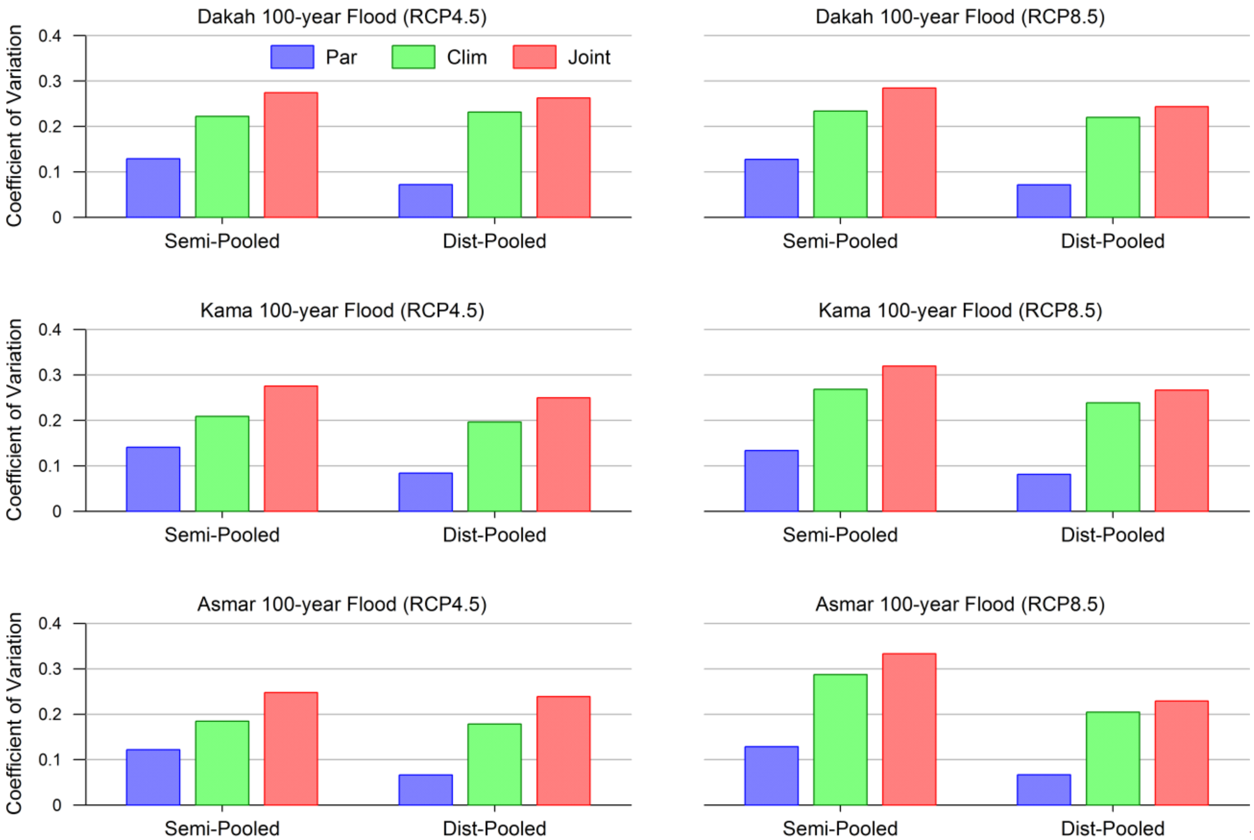


Figure ~~11~~12. Comparison of GCM average ~~100-year flood~~100-year daily flood events derived from the semi-distributed and distributed pooled calibrations. The uncertainty range is from 50 trials of the model calibration.


Dakah 100-year Flood (RCP4.5)
Dakah 100-year Flood (RCP8.5)
Kama 100-year Flood (RCP4.5)
Kama 100-year Flood (RCP8.5)
Asmar 100-year Flood (RCP4.5)
Asmar 100-year Flood (RCP8.5)
Par
Clim
Joint
Coefficient of Variation
0.4
0.3
0.2
0.1
0
Semi-Pooled
Dist-Pooled

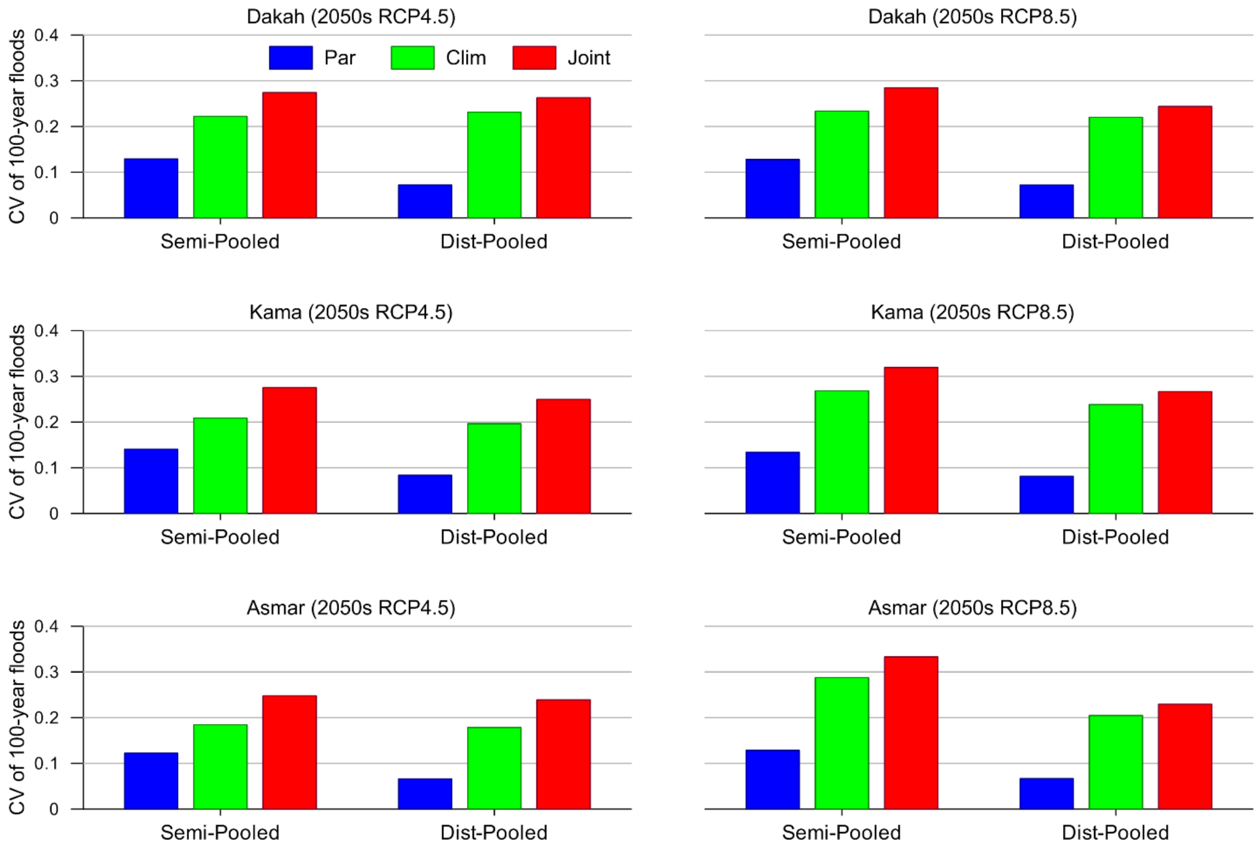
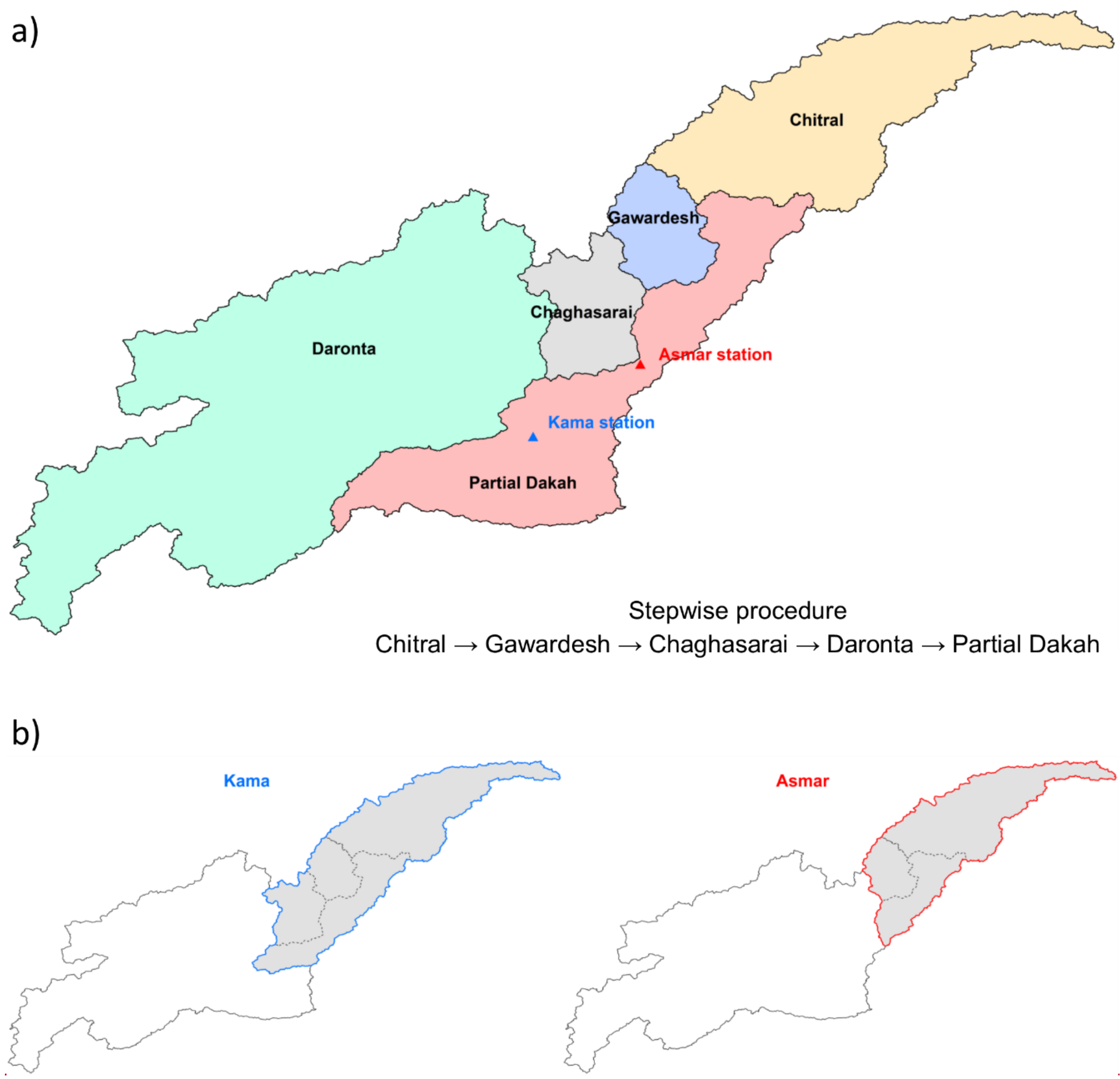


Figure ~~12~~13. Uncertainties in 100-year daily flood estimates for 2050s ~~Impact of three uncertainties on 100-year flood events~~ are assessed using ~~derived from~~ the Semi-Pooled and Dist-Pooled calibrations. Uncertainties are evaluated by calculating coefficient of variation (CV) of 2050s 100-year flood estimates under three uncertainty sources: calibration uncertainty across 50 calibration trials (Par), climate uncertainty across GCM projections (Clim), and combined uncertainty (Joint).

Supplementary materials



~~Figure S1. (a) Sub-basins corresponding to five gaging stations are used for the multisite calibrations. (b) Two sub-basins (Kama and Asmar) are assumed to be ungaged and used for evaluating the calibration approaches.~~

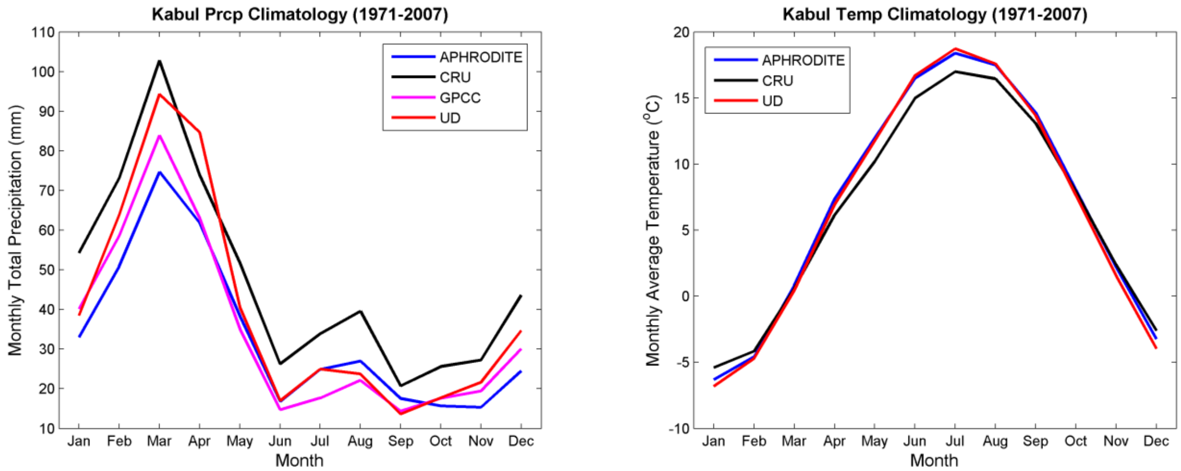


Figure ~~S2~~S1. Comparison of ~~climatology of~~ basin-wise average monthly precipitation and temperature for the Kabul River basin. Sources of data sets: APHRODITE (Asian Precipitation High-Resolved Observational Data Integration Towards Evaluation), CRU (Climatic Research Unit), GPCC (Global Precipitation Climatology Centre), UD (University of Delaware).

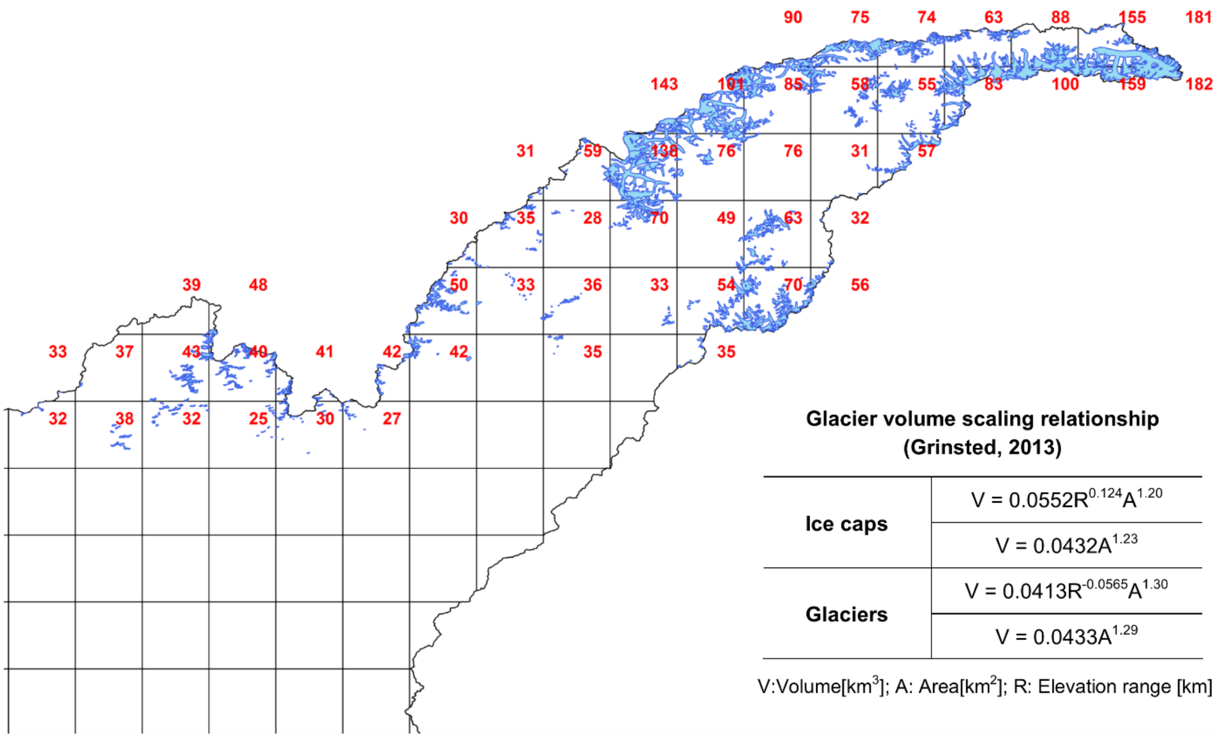


Figure ~~S3~~S2. Glacial coverage in the Kabul River basin based on the Randolph Glacier Inventory version 3.2. Glacier volume scaling relationship proposed by Grinsted (2013) is applied to derive glacier volume. Numbers in red represent glacier depths in meter of water for grid cells containing glaciers.

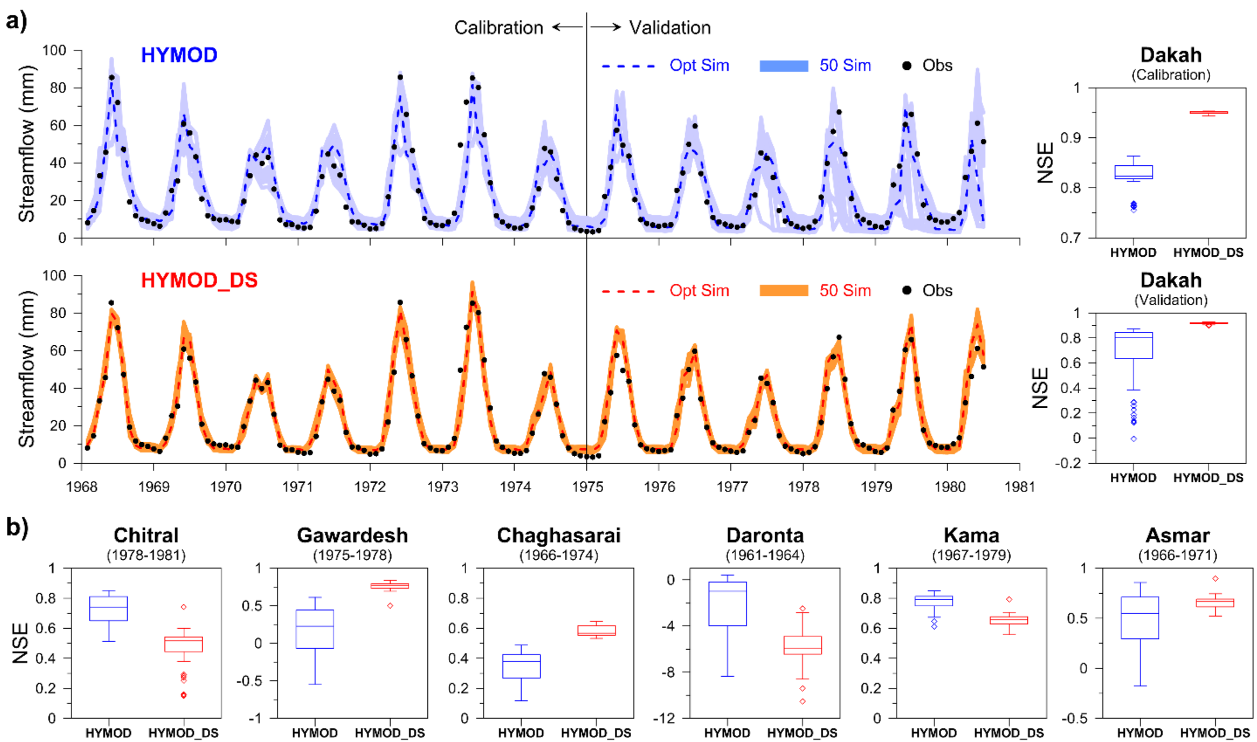


Figure S3. (a) Basin outlet (Dakah) simulations of HYMOD and MYMOD_DS (with the lumped parameterization) from 50 trials of calibration. The Box plots provide the performance evaluation on 50 simulations of both models for both calibration and validation periods. (b) Performances of the models at the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

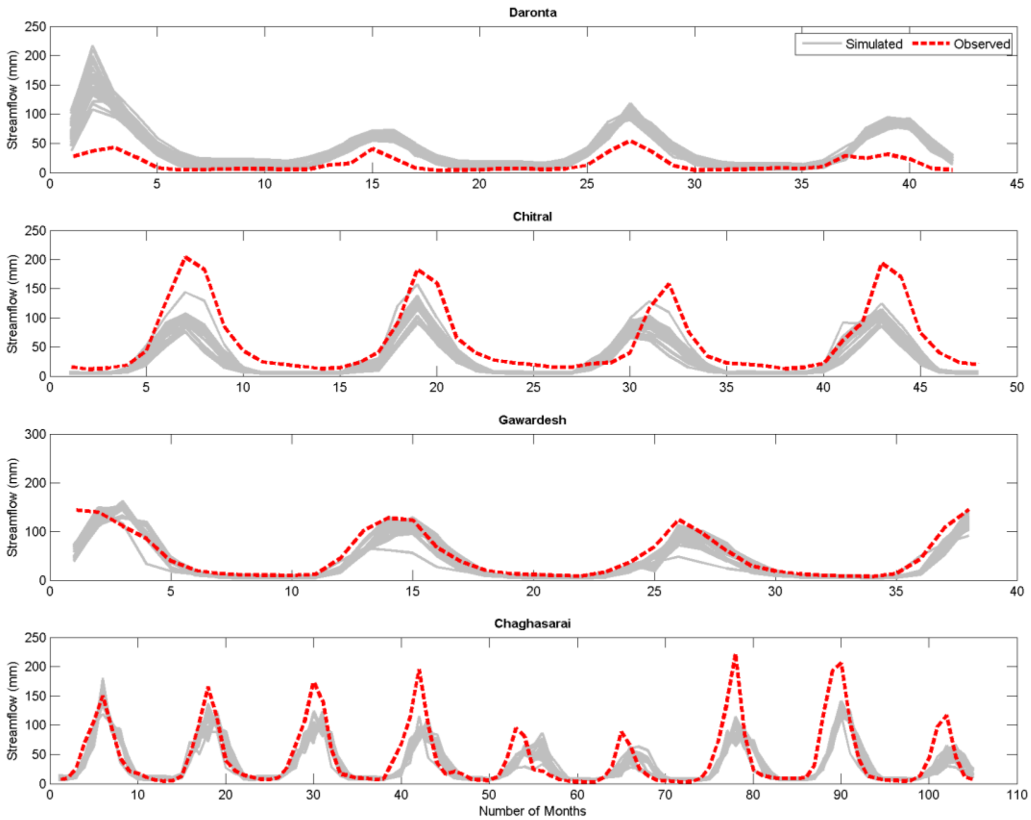


Figure S4. HYMOD_DS streamflow simulations at sub-basins from 50 trials of the basin outlet calibration under the lumped parameterization.

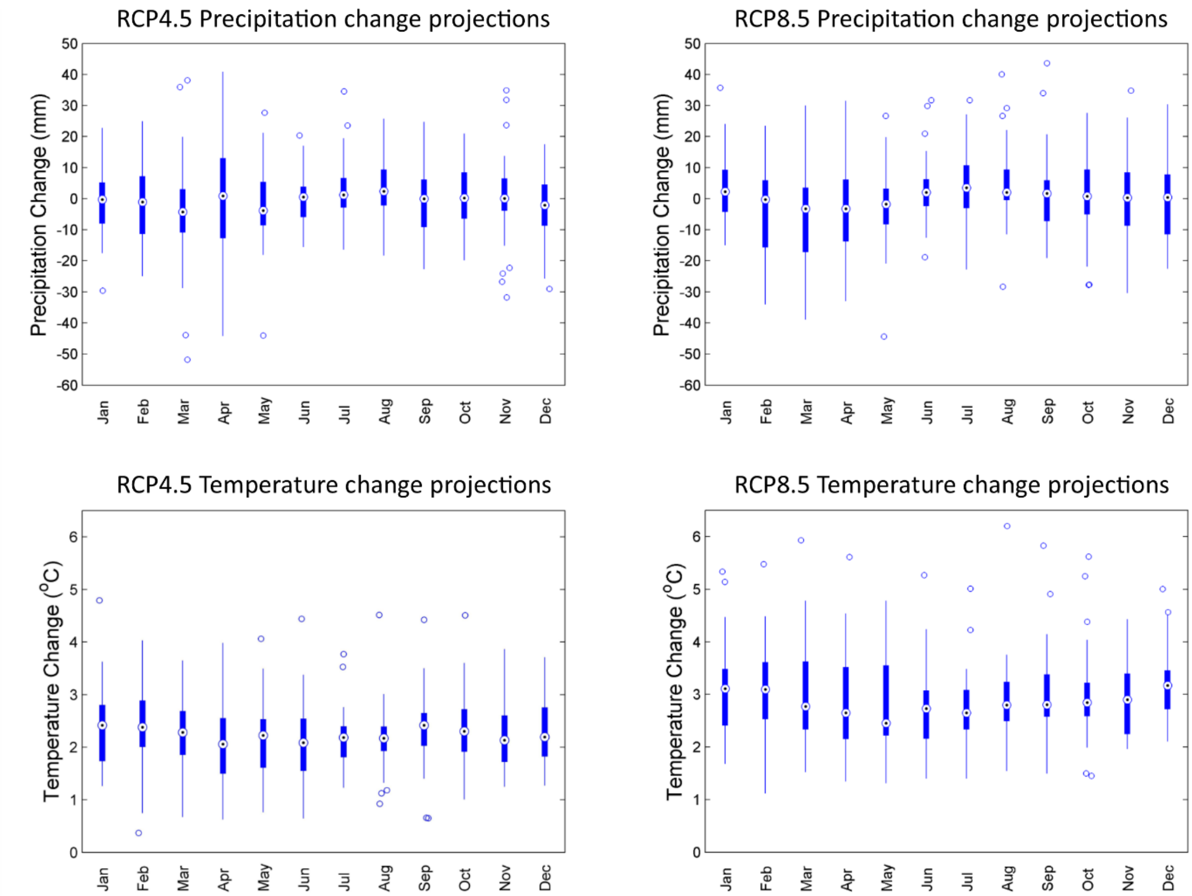


Figure S5. CMIP5 climate change projections of precipitation and temperature for the Kabul basin. The changes in ~~climatology of~~ average monthly total precipitation and mean temperature for the future period 2050s (2036-2065) were calculated from the comparison with the historical period (1976-2005). 36 GCMs were employed in this analysis.

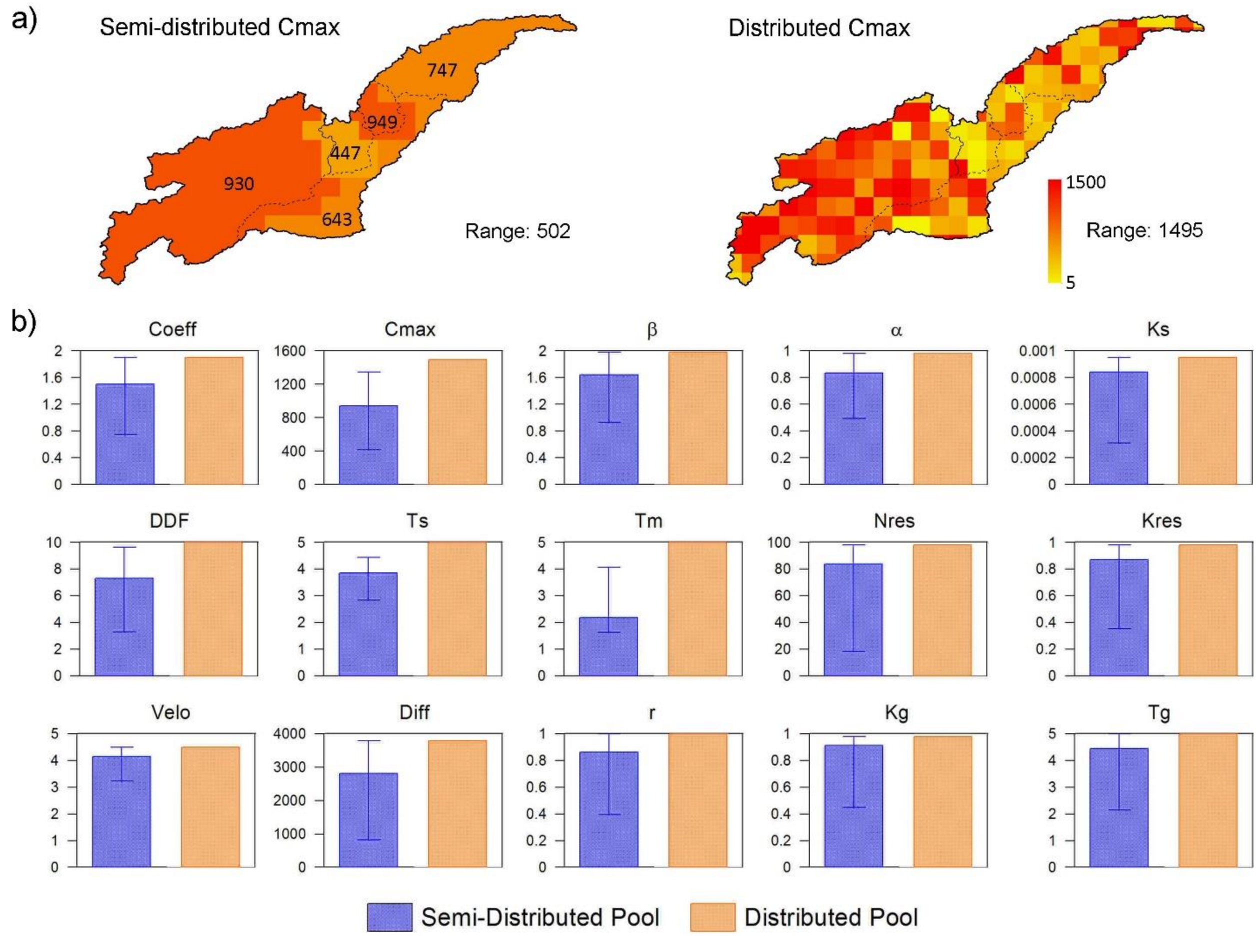


Figure S6. Spatial variability of the HYMOD_DS parameters. a) An example with C_{max} showing parameter ranges resulting from the single trail of Semi-Pooled and Dist-Pooled. b) Average spatial variability across 50 trials of calibration for all 15 parameters. Error bar in b) represents the range of parameter spatial variability from the 50 trails.

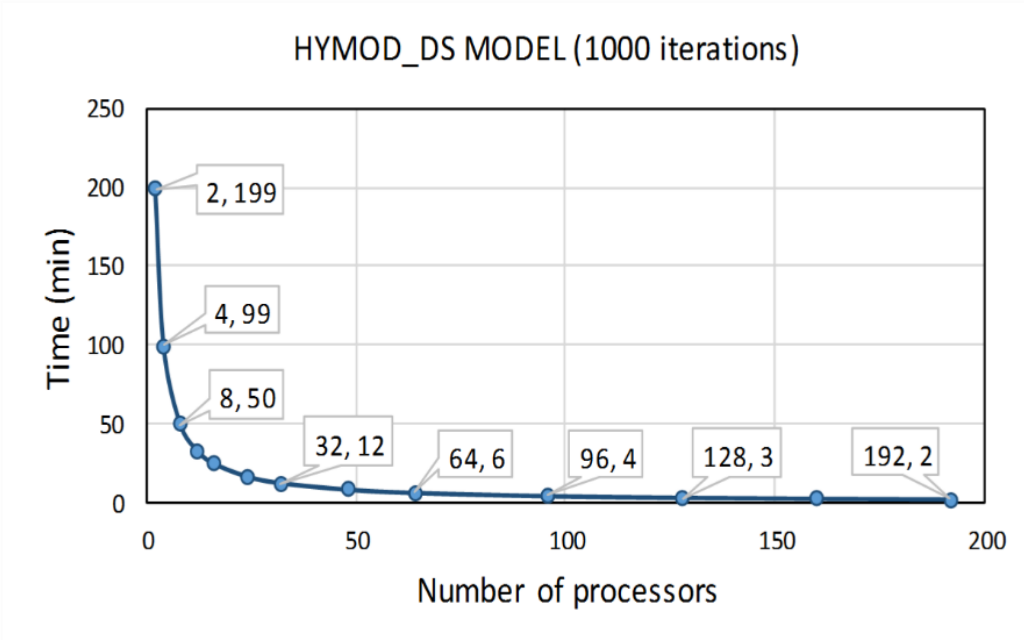


Figure S7. HYMOD_DS run time on parallel computing system.