

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

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

S. Wi¹, Y. C. E. Yang¹, S. Steinschneider¹, A. Khalil² and C. M. Brown¹

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

24 **Abstract**

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

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

48 1. Introduction

49 In an effort to advance hydrologic modelling and forecasting capabilities, the development
50 and implementation of physically-based, spatially distributed hydrologic models has proliferated
51 in the hydrologic literature, supported by readily available geographic information system (GIS)
52 data and rapidly increasing computational power. Distributed hydrologic models can account for
53 spatially variable physiographic properties and meteorological forcing (Beven, 2012), improving
54 simulations compared to conceptual, lumped models for basins where spatial rainfall variability
55 effects are significant (Ajami, et al., 2004; Koren, et al., 2004; Reed, et al., 2004; Khakbaz, et al.,
56 2012; Smith, et al., 2012) and for nested basins (Bandaragoda, et al., 2004; Brath, et al., 2004;
57 Koren, et al., 2004; Safari, et al., 2012; Smith, et al., 2012). The benefits of distributed modeling
58 have been recognized by the U. S. National Oceanic and Atmospheric Administration's National
59 Weather Service (NOAA/NWS) and demonstrated in the Distributed Model Intercomparison
60 Project (DMIP) (Reed, et al., 2004; Smith, et al., 2004; Smith, et al., 2012; Smith, et al., 2013).
61 Importantly, distributed hydrologic models can evaluate hydrological response at interior ungaged
62 sites, a benefit not afforded by ~~conceptual~~-lumped models. The use of distributed hydrologic
63 modelling for interior point streamflow estimation is particularly relevant for poorly gaged river
64 basins in developing countries, where reliable predictions at interior sites are often required to
65 inform water infrastructure investments. As international development agencies begin to integrate
66 climate change considerations into their decision-making processes (e.g., Yu et al., 2013), these
67 investments need to be robust under both current climate conditions and alternative-possible future
68 climate regimes.

69 Despite their roots in physical realism, distributed hydrologic models can suffer from
70 substantial uncertainty. A major source of uncertainty originates from the proper identification of

71 parameter values that vary across the watershed, especially when observed streamflow data is only
72 available at one or a few points ([Exbrayat et al., 2014](#)). Parameters can be discretized across the
73 watershed in several ways ([Flugel, 1995](#); [Efstratiadis et al., 2008](#); [Khakbaz, et al., 2012](#)): uniquely
74 for each grid cell or hydrologic response unit (fully distributed), based on ~~hydrologic response~~
75 ~~units~~sub-basins whose boundaries do not necessarily ensure homogenous characteristics (semi-
76 distributed), or in the simplest case, a single parameter set for all model grid cells (lumped). With
77 limited data, the parameter identification problem, particularly for the fully distributed case, can
78 be impractical or infeasible (Beven, 2001). The parameterization challenge has spurred substantial
79 advances in understanding appropriate calibration techniques for distributed hydrologic models.
80 Many studies have attempted to reduce the dimensionality of the calibration problem to alleviate
81 the issue of equifinality (Beven & Freer, 2001), which is the phenomenon whereby multiple
82 parameter sets produce indistinguishable model performance. This work has found favorable
83 results when the parametric complexity of the distributed model is aligned with the data available
84 for calibration (Leavesley, et al., 2003; Ajami, et al., 2004; Eckhardt, et al., 2005; Frances, et al.,
85 2007; Zhu & Lettenmaier, 2007; Cole & Moore, 2008; Pokhrel & Gupta, 2010; Khakbaz, et al.,
86 2012). There has also been extensive research exploring the use of multiple objectives and different
87 operational procedures to understand parameter estimation tradeoffs and identifiability for
88 distributed model calibration, with great success (Madsen, 2003; Efstratiadis & Koutsoyiannis,
89 2010; Li, et al., 2010; Kumar, et al., 2013).

90 Despite these advances, important questions still persist. It still remains difficult to
91 compare the uncertainty that emerges from different operational calibration procedures for
92 multisite applications (i.e. whether gages in series should be used sequentially or simultaneously
93 for calibration) and under different levels of parametric complexity. Due to the computational

94 burden required to calibrate distributed models, this uncertainty is problematic to explore. Further,
95 in poorly gaged basins, it is challenging to quantify the lost accuracy and increased uncertainty for
96 interior flow estimation when a distributed model is calibrated only at an outlet gage (which is
97 often all that is available in developing country river basins). In the case of significant spatial
98 variability in the basin properties that influence runoff generation (e.g., permeability, vegetation,
99 slope, etc.), accurate runoff predictions are unlikely at interior locations based only on the lumped
100 information obtained at the basin outlet ~~Many studies have reported that distributed models~~
101 ~~calibrated at the basin outlet are less accurate at interior locations~~ (Anderson et al., 2001; Cao, et
102 al., 2006; Breuer et al., 2009; Lerat et al., 2012; Simith et al., 2012; Wang, et al., 2012). ~~but~~ The
103 extent of ~~the~~ this error and uncertainty is not well understood for heterogeneous basins ~~unknown~~
104 due to the computational expense required ~~needed~~ to explore this issue. Finally, rarely have the
105 implications of these calibration issues been explicitly examined for ~~an alternative climate~~ possible
106 future climate conditions, which is required in climate change impact studies. This question has
107 been explored for lumped, conceptual models (Wilby, 2005; Steinschneider, et al., 2012), but has
108 been difficult to evaluate for computationally expensive distributed models.

109 This study addresses the above research challenges by focusing on the following four
110 questions: 1) How does calibration procedure for using multisite data effect the accuracy and
111 uncertainty of distributed models used for streamflow predictions at ungaged sites, 2) what effects
112 do increased parameter complexity have on distributed model calibration and prediction, 3) how
113 much degradation in model accuracy and uncertainty can be expected for interior flow estimation
114 based on a calibration procedure using only the basin outlet, and 4) how do different calibration
115 formulations for a distributed model alter projections of streamflow at ungaged sites under climate
116 change conditions? These questions are considered in an application of a distributed version of the

117 daily HYMOD hydrologic model to the Kabul River basin in Afghanistan and Pakistan. To address
118 these research questions, high performance computing is utilized to manage the computational
119 burden that often hinders such explorations, ~~a relatively recent technique employed in hydrological~~
120 ~~modeling research~~ (Laloy & Vrugt, 2012; Zhang, et al., 2013).

121

122 2. Study area

123 The Kabul River basin (67,370km²) is a plateau surrounded by mountains located in the
124 eastern central part of Afghanistan (Figure 1). It is the most important river basin of Afghanistan,
125 containing 35 percent of the country's population. While it encompasses just 12 percent of the area
126 of Afghanistan, the basin's average annual streamflow (about 24 billion cubic meters) is about 26
127 percent of the country's total streamflow volume (World Bank, 2010).

128 Water resources from the basin are shared by Afghanistan and Pakistan and serve as a water
129 supply source for more than 20 million people. The shared use of transboundary water between
130 these two countries is central in establishing regional water resources development for this area
131 (Ahmad, 2010). It is crucial to develop tools that can support engineering plans for existing and
132 potential water infrastructure to take full advantage of the water resources in the basin. The
133 Government of Afghanistan has developed comprehensive plans for new hydropower projects on
134 the Kabul River owing to its advantageous topography for the development of water storage and
135 hydropower (IUCN, 2010), and recently reached an agreement with the Pakistan government to
136 work on a 1,500MW hydropower project on the Kunar River (one of major tributary in the Kabul
137 River basin) as part of the joint management of common rivers between the two countries (DAWN,
138 2013).

139 The streamflow regime of the Kabul River can be classified as glacial with maximum
140 streamflow in June or July and minimum streamflow during the winter season. Approximately
141 70% of annual precipitation (475mm) falls during the winter season (November to April). While
142 the dominant source of streamflow in winter is baseflow and winter rainfall, Glaciers-glaciers and
143 snow cover are the most important long-term forms of water storage and, hence, the main source
144 of runoff during the ablation period for the basin (Shakir et al., 2010). In total 5.72.9%
145 (3813km²1954km²) of the basin is glacierized based on the Randolph Glacier Inventory version
146 3.2 (Pfeffer, et al., 2014). The melt water from glaciers and snow produce the majority (75%) of
147 the total streamflow (Hewitt, et al., 1989). Table 1 provides the climates and geophysical properties
148 of each sub-watershed delineated by the stations located inside the Kabul Basin (Figure 1). Two
149 different climate patterns are distinguishable across the sub-basins. The sub-basins on the Kunar
150 River tributary (Kama, Asmar, Chitral, Gawardesh, and Chaghasarai) receive moderate annual
151 precipitation and are highly affected by snow and glacier covers. All of these sub-basins have high
152 ratios of mean annual flow to mean annual precipitation, with the ratios for the Kama, Asmar,
153 Chitral, and Chaghasarai sub-basins larger than 1. Conversely, the Daronta sub-basin contains only
154 minimal glacial cover, and is relatively dry. Daronta is also much less productive, with annual
155 streamflow far below the other sub-basins with an average of only 165 mm/year.

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

162 [2014](#)) highlights the need to assess the impact of climate change on ~~water resources~~ [future water](#)
163 [availability](#) in this area (~~Immerzeel, et al., 2010; Immerzeel, et al., 2013; Molg, et al., 2014; Radie,~~
164 ~~et al., 2014~~).

165

166 **3. Data and Models**

167

168 **3.1. Data**

169 Gridded daily precipitation and temperature products with a spatial resolution of 0.25° were
170 gathered between calendar years 1961-2007 from the Asian Precipitation Highly Resolved
171 Observational Data Integration Towards Evaluation (APHRODITE) dataset (Yatagai, et al., 2012).
172 There has been some concern regarding underestimation of precipitation in APHRODITE for some
173 regions of Asia (Palazzi, et al., 2013); our preliminary data analysis (intercomparison of
174 precipitation products between 5 different databases) confirmed this for the Kabul River basin
175 (shown in Figure [S2S1](#)). Thus, the APHRODITE precipitation was bias-corrected by the
176 precipitation product from the University of Delaware global terrestrial precipitation (UD) dataset
177 (Legates & Willmott, 1990). Daily series of bias-corrected APHRODITE precipitation were
178 coupled with APHRODITE temperature for 160 0.25° grid cells to produce a climate forcing
179 dataset for the distributed domain of the Kabul River basin model.

180 This study used the set of global climate change simulations from the [World Climate](#)
181 [Research Programme's Coupled Model Intercomparison Project Phase 5 \(CMIP5\)](#) multi-model
182 ensemble (Talyor, et al., 2012). Monthly climate outputs of GCMs were downscaled to a daily

183 temporal resolution and 0.25° spatial resolution based on the bias-correction spatial disaggregation
184 (BCSD) statistical downscaling method introduced by Wood et al. (2004).

185 Monthly streamflow observations for seven locations in the Kabul River basin (Figure 1)
186 were gathered between calendar years ~~1961-1960-1980-1981~~ from two data sources: the Global
187 Runoff Data Centre (GRDC) database and the United States Geological Survey (USGS) database
188 (Table 1). Streamflow data were not collected in Afghanistan after September 1980 until recently
189 because streamgaging was discontinued soon after the Soviet invasion of Afghanistan in 1979
190 (Olson and Williams-Sether, 2010). Though measurements were taken at a daily time step, data
191 are only made available for public use at monthly aggregated levels, calculated using the mean of
192 the daily values. The available monthly -streamflow observations at each station were used for
193 calibrating and validating the distributed hydrologic model (Figure 32). Kama and Asmar stations
194 are treated as ungaged sites because they align with the potential dam project on the Kunar River
195 tributary. and The two gage stations are left out of the processes of multisite calibrations in order
196 to evaluate the model's ability to predict streamflow at interior ungaged sites. Furthermore, half of
197 the record at the Dakah station, located at the basin outlet, is also used for validation purposes.

198 The Randolph Glacier Inventory version 3.2 (RGI 3.2) dataset (Pfeffer, et al., 2014) was
199 used to extract glacial coverage in the Kabul River basin, which totaled 5.7% of the basin area
200 (Figure S3S2). In the hydrological modeling process, the model needs to be informed by reliable
201 estimates on volume of water retained in glaciers, especially for future simulations under warming
202 conditions. We followed the method proposed in Grinsted (2013), which uses multivariate scaling
203 relationships to estimate glacier and ice cap volume based on elevation range and area.
204 Specifically, the scaling law including area and elevation range factors was applied to estimate
205 glacier/ice cap volume when the glacier depth exceeded 10m. Otherwise, glacier/ice cap volume

206 was estimated with the area-volume scaling law. The elevation range spanned by each individual
207 glacier is estimated using the global digital elevation model (DEM) from the shuttle radar
208 topography mission (SRTMv4) in 250m resolution (Jarvis, et al., 2008). Density of ice (0.9167
209 g/cm³) is applied to calculate glacier/ice cap volume in meters of water equivalent.

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

214

215 **3.2. Distributed Hydrologic Model (HYMOD_DS)**

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

225 The HYMOD conceptual watershed model has been extensively used in studies on
226 streamflow forecasting and model calibration (Wagener, et al., 2004; Vrugt, et al., 2008; Kollat,
227 et al., 2012; Gharari, et al., 2013; Remesan, et al., 2013). The HYMOD is a soil moisture

228 accounting model based on the probability-distributed storage capacity concept proposed by
 229 Moore (1985). This conceptualization represents a cumulative distribution of varying storage
 230 capacities (C) with the following function:

$$231 \quad F(C) = 1 - \left(1 - \frac{C}{C_{\max}}\right)^B \quad 0 \leq C \leq C_{\max} \quad (1)$$

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

$$237 \quad S(t) = \frac{C_{\max}}{B+1} \cdot \left(1 - \left(1 - \frac{C^*(t)}{C_{\max}}\right)^{B+1}\right) \quad (2)$$

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

$$240 \quad Runoff_1 = \begin{cases} P(t) + C^*(t-1) - C_{\max} & \text{if } P(t) + C^*(t-1) \geq C_{\max} \\ 0 & \text{if } P(t) + C^*(t-1) < C_{\max} \end{cases} \quad (3)$$

$$241 \quad Runoff_2 = \begin{cases} (P(t) - Runoff_1) - (S(t) - S(t-1)) & \text{if } P(t) - Runoff_1 \geq S(t) - S(t-1) \\ 0 & \text{if } P(t) - Runoff_1 < S(t) - S(t-1) \end{cases} \quad (4)$$

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

245 $Q_{\text{fast}} = \text{Runoff}_1 + \alpha \cdot \text{Runoff}_2$ (5)

246 $Q_{\text{slow}} = (1 - \alpha) \cdot \text{Runoff}_2$ (6)

247 The potential evapotranspiration (PET) is derived based on the Hamon method (Hamon,
 248 1961), in which daily PET in mm is computed as a function of daily mean temperature and hours
 249 of daylight:

250
$$\text{PET} = \text{Coeff} \cdot 29.8 \cdot L_d \cdot \frac{0.611 \times \exp(17.27 \cdot T / (T + 273.3))}{T + 273.3}$$
 (7)

251 where, L_d is the daylight hours per day, T is the daily mean air temperature ($^{\circ}\text{C}$), and Coeff is a
 252 bias correction factor. The hours of daylight is calculated as a function of latitude and day of year
 253 based on the daylight length estimation model (CBM model) suggested by Forsythe et al. (1995).

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

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

$$265 \quad M_s = DDF_s \times (T - T_s) \quad (78)$$

$$266 \quad M_g = DDF_g \times (T - T_g) \quad (89)$$

267 with DDF_s (T_s) and DDF_g (T_g) applied separately for snow and glacier modules, respectively. To
 268 account for the higher melting rate of glacier than snow owing to the low albedo (Konz & Seibert,
 269 2010; Kinouchi, et al., 2013), we introduced a parameter $r > 1$ to constrain DDF_g to be larger than
 270 DDF_s (i.e. $DDF_g = r \times DDF_s$). For the rain that falls on the glacierized area, the glacier parameter
 271 K_g determines the portion of rain becoming surface runoff as a multiplier for the rainfall. The
 272 remaining rainfall is assumed to be accumulated to the glacier store.

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

$$277 \quad u(t) = \frac{K_q}{\Gamma(N)} (K_q t)^{N-1} \exp(-K_q t) \quad (910)$$

278 where, Γ is the gamma function. The within-grid groundwater routing process is simplified as a
 279 lumped linear reservoir with the storage recession coefficient of K_s .

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

$$282 \quad \frac{\partial Q}{\partial t} + C \frac{\partial Q}{\partial x} - D \frac{\partial^2 Q}{\partial x^2} = 0 \quad (911)$$

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

284 Similar to most other hydrological models (Efstratisdis et al., 2008), HYMOD DS is not
285 designed to model water abstractions for agricultural lands and dam operations within the basin.
286 According to World Bank (2010), water demand for agricultural use is about 2,000 MCM (million
287 cubic meters), or about 8.3% of the total annual flow. The Naglu dam (Figure 1) upstream of the
288 Daronta streamflow gage forms the largest and most important reservoir in the basin, with an active
289 storage of 379 MCM. In our hydrologic modelling process, the water consumed by irrigated
290 croplands is implicitly accounted for by the evapotranspiration module. We note that the degree
291 of irrigation impact during the time frame used for calibration (1960-1981) is likely much smaller
292 than the current level. We also expect that using monthly data for calibration somewhat reduces
293 the bias from human interference, particularly the daily operations of Naglu dam. Nevertheless,
294 the calibration results for the gage below this dam (Daronta), and to a lesser extent the basin outlet
295 (Dakah), should be approached with caution. Given that a majority of the gages examined in this
296 study are on an underdeveloped branch of the Kabul River, issues of human interference on
297 calibration are somewhat mitigated.

299 **4. Methods**

300 The purpose of this study is to explore the implications of different calibration strategies
301 and choices for a computationally expensive distributed hydrologic model. A variety of calibration
302 experiments are conducted, with the results from preceding experiments informing choices made
303 for subsequent ones. All calibration approaches are tested in terms of their ability to predict flows
304 at interior site gages that were left out of the calibration process. In all cases, the Genetic Algorithm
305 (GA) introduced by Wang (1991) is used as an optimization method for model parameter
306 calibration ~~(Wang, 1991; Zhang, et al., 2008; Kollat, et al., 2012)~~, and the objective function is

307 based simply on the Nash Sutcliffe efficiency (NSE) (Nash & Sutcliff, 1970), which is by far the
308 most utilized performance metric in hydrological model applications (Biondi et al., 2012). A
309 multisite average of the NSE is used when evaluating performance across multiple sites. We fully
310 recognize that the use of one objective, such as the NSE, is inferior compared to multi-objective
311 approaches that can identify Pareto optimal solutions that provide good model performance across
312 different components of the flow regime (Madsen, 2003; Efstratiadis & Koutsoyiannis, 2010; Li,
313 et al., 2010; Kumar, et al., 2013). However, in this particular study daily hydrologic model
314 simulations can only be compared against available monthly streamflow records, reducing the
315 number of viable objectives against which to calibration. That is, statistics representing peak flows,
316 extreme low flows, and other daily flow regime characteristics often used in multi-objective
317 optimization approaches are unavailable. We believe that the use of a monthly NSE value as a
318 single objective, while coarse, does not inhibit our ability to provide insight into the research
319 questions posed. In addition to the NSE, the Kling-Gupta efficiency (KGE) (Gupta et al., 2009) is
320 adopted as an alternative model performance metric, which equally weights model mean bias,
321 variance bias, and correlation with observations.

322 In this study, three levels of parameter complexity are considered: lumped, semi-
323 distributed, and fully distributed formulations (Figure 24). The different levels of parameter
324 complexity are defined according to the spatial distribution of unique hydrologic model
325 parameters. In the lumped formulation a single parameter set is applied to the entire basin. In the
326 semi-distributed formulation, a unique parameter set is assigned to each sub-basin, defined based
327 on the location of available streamflow gaging sites. The fully distributed parameter structure
328 follows the spatial discretization of climate input grids, allowing a unique parameter set for each
329 grid cell. No matter the parameterization scheme, the model structure follows the climate input

330 grids, i.e. the hydrological water cycle within each grid cell is modelled separately. We note that
331 a lumped model structure (i.e., no gridded or sub-unit structure) has often been considered as a
332 baseline model formulation in the assessment of distributed modelling frameworks (e.g., see Smith
333 et al., 2013). However, the focus of our study is on ungaged interior site streamflow estimation,
334 making this formation somewhat inappropriate. Further, preliminary tests comparing streamflow
335 simulations at the basin outlet (Dakah) between a gridded and basin-averaged structure, both with
336 a lumped parameter formulation, support the use of the distributed grid structure (Figure S3).

337 The parameter complexity will vary depending on the calibration experiment being
338 conducted, but for each experiment regardless of the parameterization, the optimization is
339 implemented 50 times using the GA algorithm to explore parameter-calibration uncertainty. The
340 considerably high computational cost required to perform a large number of calibrations is
341 managed using the parallel computing power provided by the Massachusetts Green High
342 Performance Computing Center (MGHPCC), from which several thousands of processors are
343 available.

344 In the first modeling experiment, we explore two calibration strategies for using multisite
345 streamflow data, a stepwise and pooled approach. In the stepwise calibration, parameters are
346 calibrated for upstream gaged sub-catchments and subsequently fixed during calibration of
347 downstream points, while for the pooled approach, parameters are calibrated for multiple sub-
348 catchments simultaneously. Both approaches are assessed for the semi-distributed formulation.
349 The better of the two methods is identified for use in the second experiment, where the effects of
350 increased parameter complexity are tested in terms of streamflow prediction accuracy and
351 uncertainty. In the third experiment, we consider the situation where there is only gaged
352 location data at the basin outlet for calibration. Here, the model is calibrated against the outlet gage

353 under all levels of parameter complexity and is compared against the best combination of
354 calibration strategy (step-wise or pooled) and parameter complexity (lumped, semi-distributed, or
355 fully distributed) identified in the previous experiments. Finally, a subset of the calibration
356 approaches deemed worthy of further investigation are compared in terms of their projections of
357 future streamflow under climate change to highlight how model calibration differences can alter
358 the results of a climate change assessment for water resources applications. These experiments are
359 described in further detail below.

360

361 **4.1. Multisite Calibration: Stepwise and Pooled Approaches**

362 In the first experiment, the semi-distributed parameterization concept is compared under
363 alternative multisite calibration strategies, the stepwise and pooled calibration approaches. To
364 conduct the stepwise calibration, a nested class of sub-basins is defined corresponding to multiple
365 gaging stations. In the first step of the stepwise calibration, the optimization process is carried out
366 with nested sub-basins at the lowest level (i.e., the most upstream sites). Once parameters of nested
367 sub-basins are determined, the parameters are fixed, and the calibration procedure proceeds with
368 nested basins at upper levels until parameters for the entire basin are determined. In this particular
369 application to the Kabul River basin, 5 gaged sub-basins were selected and the stepwise calibration
370 procedure for those sub-basins followed this direction: Chitral → Gawardesh → Chaghasarai →
371 Daronta → Dakah (Figure S45). The stepwise calibration approach involves a number of GA
372 implementations corresponding to the number of gaging sites. The GA optimization was carried
373 out a total of 250 times in this application, with 50 optimization runs containing GA
374 implementations for 5 sub-basin regions.

375 The pooled calibration strategy involves calibrating all parameters of the model domain
376 simultaneously against multiple streamflow gages within the watershed. This approach aims at
377 looking for suitable parameters that are able to produce satisfactory model results at all gaging
378 stations in a single implementation of GA optimization. That is, the GA searches the entire
379 parameter space at once to maximize the average NSE across all sites. This operational feature
380 reduces the processing time spent on the GA implementation compared to the stepwise calibration
381 strategy. To identify the better of the two multisite calibration approaches, the comparison focused
382 on their ability to predict streamflow and calibration uncertainties at two interior site gages (Kama
383 and Asmar) that were assumed to be ungaged (Figure S15), as well as for validation data at the
384 basin outlet.

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

390

391 **4.2. Increased Parameter Complexity**

392 In the second experiment, the better of the two approaches (step-wise or pooled) identified
393 in the first experiment is further tested with respect to the three different levels of parameter
394 complexity. In addition to the semi-distributed parameter formulation considered in the first
395 experiment, lumped and fully-distributed parameter formulations are calibrated for the selected
396 approach to investigate the gain or loss arising from different levels of parameter complexity. Since
397 the hydrologic model HYMOD employed in this study involves 15 parameters, the lumped version

398 of the HYMOD_DS contains a single, 15-member parameter set applied to all model grid cells.
399 The semi-distributed conceptualization of HYMOD_DS contains a single parameter set for each
400 sub-basin, totaling 75 parameters. In the distributed parameterization the number of parameters
401 increases dramatically. With 160 0.25° grid cells, the number of parameters requiring calibration
402 reaches 2,400. As the number of parameters increase across the parameterization schemes,
403 calibration becomes increasingly computationally expensive. The number of model runs used in
404 the GA optimization algorithm for the lumped, semi-distributed, and distributed parameterization
405 schemes are 15,000 (150 populations \times 100 generations), 75,000 (750 \times 100), and 480,000 (2400
406 \times 200), respectively. These population/generation sizes were supported using convergence tests
407 for each calibration. Again, 50 separate GA optimizations were used to explore calibration
408 uncertainties for each parameterization scheme. To give a sense of the computational burden of
409 this experiment, we note that 50 trials of the HYMOD_DS calibration under the distributed
410 conceptualization required 1,000 processors over 7 days on the MGHPCC system.

411

412 **4.3. Basin Outlet Calibration**

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

421 settings for each level of parameter complexity are same as the settings used in the second
422 experiment.

423

424 **4.4. Climate Change Projections of Streamflow**

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

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

443

444 **5. Results and Discussion**

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

452

453 **5.1. Pooled Calibration vs. Stepwise Calibration**

454 This section reports the results from the first experiment comparing the stepwise and
455 pooled calibration approaches for the semi-distributed model parameterization. Figure ~~5-6~~ shows
456 the comparison between the Semi-Stepwise and Semi-Pooled with boxplots representing the 50
457 trials of calibration. Under the stepwise calibration the results for 4 sub-basins (Chitral,
458 Gawardesh, Chaghasarai, and Daronta) are optimal because there is no interaction between those
459 sub-basins. However, the calibrated parameter sets of each sub-basin act as constraints in the last
460 step of the Semi-Stepwise resulting in the degradation of model skill at the basin outlet (Dakah)
461 and two left-out gages (Asmar and Kama). This becomes apparent when comparing the Semi-
462 Stepwise to the Semi-Pooled results. The model skill under the Semi-Pooled is similar to that from
463 the Semi-Stepwise with respect to the 4 upstream sub-basins, but it outperforms at the verification
464 gages. This is particularly true for the Asmar gage, which exhibits a downward bias and substantial

465 variability in performance under the Semi-Stepwise. The Semi-Pooled results suggest that small
466 sacrifices of model performance at certain sites can improve and stabilize basin-wide performance.
467 Expected values of KGE from 50 calibrations are also provided (values in parenthesis in the bottom
468 of Figure 56) and this performance metric also leads to the same conclusion. Therefore, the Semi-
469 Pooled was selected as the better multisite calibration strategy and is considered for further
470 analyses in the following sections.

471

472 **5.2. Pooled Calibration with Alternative Parameterizations**

473 Here we examine results for the three levels of parameter complexity applied to the pooled
474 calibration approach. Figure 67 shows the comparison of the pooled calibrations. Unsurprisingly,
475 streamflow predictions from the Lump-Pooled have the lowest accuracy and largest uncertainty at
476 the calibration sites, particularly for the Chaghasarai and Daronta sites. This demonstrates the well-
477 known difficulty in representing flow characteristics of a spatially variable system with a
478 homogenous parameter set (Beven, 2012). The pooled calibration substantially improves with
479 increasing parameter complexity at the calibration sites. Both the Semi-Pooled and Dist-Pooled
480 produce NSE values above 0.8 for all calibration sites, with the Dist-Pooled showing somewhat
481 higher performance, undoubtedly from its greater freedom to over-fit to the calibration data.
482 However, the advantage of the Dist-Pooled with respect to streamflow predictions at validation
483 sites becomes less clear. Only the Dist-Pooled at Kama shows marginally better predictions, while
484 the results are ambiguous at Dakah and Asmar. Overall, this likely suggests that the fully
485 distributed conceptualization leads to over-fitting of the model as compared to the Semi-Dist
486 conceptualization. We reached the same conclusion when examining the KGE values, which rise

487 with greater parameter complexity at calibration sites but no longer follow this pattern strictly at
488 validation sites.

489 Interestingly, the Lump-Pooled performs well at the verification sites despite its poor
490 performance at calibration sites. The Lump-Pooled does not show significant degradation in skill
491 at Kama compared to the more complex parameterizations, and the flow prediction at Asmar
492 actually exhibits the best performance of all three model variants. A partial reason for this
493 unexpected result arises from different overlapping periods in the calibration and validation data
494 (see Figure 32). The periods used for the calibration for Chitral (1978-1981) and Gawardesh (1975-
495 1978) have no overlapping periods with the one for Asmar (1966-1971), which encompasses those
496 two sub-basins. Instead, the validation at Asmar is mostly affected by the calibration to Dakah
497 because of the overlapping 4 years (1968-1971) between those two sites. This explains the reason
498 why the Lump-Pooled shows high skill at Asmar despite the low skill at its sub-basins. However,
499 the low model skill at Chaghasarai from the Lump-Pooled propagates to the validation result at
500 Kama, as these two sites have a relatively long overlapping period (8 years from 1967-1974).

501

502 **5.3. Limitations of the Basin Outlet Calibration**

503 In the third experiment the HYMODS_DS was calibrated only to data at the basin outlet
504 under all levels of parameter complexity, and streamflow records for all 6 sub-basins, as well as
505 flows at Dakah not used during calibration, are used for model validation. First, we consider the
506 flows at Dakah. During the calibration period, all three parameterization schemes produce very
507 accurate streamflow predictions with NSE (KGE) values above 0.95 (0.96) (Figure 78). High
508 accuracy holds even under the Lump_Outlet, ~~which is somewhat surprising givendespite~~ the
509 spatial heterogeneity of the basin. While NSE and KGE values at Dakah rise marginally with

510 greater parameter complexity during calibration, this no longer holds during the validation period,
511 suggesting no benefit with an increase in parameter complexity.

512 The validation results for the 6 sub-basins demonstrate the danger in relying on outlet data
513 alone when calibrating a distributed model for flow prediction at interior points. Streamflow
514 predictions at interior sites exhibit low accuracy and high uncertainty, with the worst performance
515 at the Daronta site (all NSEs and KGEs are negative). We note that the poor performance at
516 Daronta is likely due in part to the impacts of water abstraction and the operation of Naglu dam.
517 Further examination (Figure S4) showed that the HYMOD_DS significantly overestimated
518 streamflow at Daronta and underestimated flow at three sites in the eastern part of the basin
519 (Chitral, Gawardesh, and Chaghasarai). Model performance at Kama and Asmar is somewhat
520 better than the other validation sites, although improvements are not the same across all
521 parameterizations. The Lump-Outlet predictions at these sites still have low average accuracy
522 (average NSE < 0.7 and average KGE < 0.6), while the Semi-Outlet exhibits large uncertainty in
523 performance across the 50 optimization trials. Surprisingly, the over-parameterized Dist-Outlet
524 shows promising results with high expected accuracy at Kama and Asmar (mean NSE (KGE) of
525 0.84 (0.71) and 0.90 (0.88), respectively) and comparable performance at many of the other sites.
526 One exception is Gawardesh, where the Lump-Outlet outperforms the other model variants,
527 although the reason for this is not immediately clear. Overall, the results indicate that any
528 calibration based on basin outlet data should be used with substantial caution when predicting
529 flows at interior basin sites.

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

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

546

547 **5.4. Climate Change Projections of Streamflow with Uncertainty**

548 Here we explore how projections of future water availability and flood risk under climate
549 change are influenced by the choice of calibration approach. For the Kabul River basin, the CMIP5
550 GCM projections of monthly total precipitation and mean temperature are shown in Figure S5.
551 According to the CMIP5 ensemble, precipitation projections show no clear trend; the average
552 precipitation change in monthly total precipitation fluctuates between -10mm and 10mm. On the
553 other hand, temperature clearly shows an upward trend for both radiative forcing scenarios. The
554 average changes in annual temperature are +2.2°C and +2.8°C for RCP4.5 and RCP8.5, which,

555 using the Hamon method, correspond to an increase in annual PET by approximately 100mm and
556 150mm, respectively.

557 We first examine average monthly streamflow estimates ~~climatology~~—across four
558 calibration strategies: the Semi-Pooled and Dist-Pooled (most promising calibration strategies), as
559 well as the Lump-Outlet (as a baseline) and Dist-Outlet (the best outlet calibration strategy). Figure
560 9-10 shows the monthly streamflow ~~predictions~~ estimates for the historical period with and the
561 2050s under the RCP 4.5 and 8.5 scenarios. The ~~the~~ whisker bars indicating ~~the~~ uncertainty range
562 across the 50 calibration trials; ~~-. The monthly streamflow predictions are also provided~~ for the
563 2050s under the RCP 4.5 and 8.5 scenarios. For the future scenarios, the whisker bars are derived
564 by averaging over the 36 different climate projections for each of the 50 trials. For the historical
565 time period, all calibration schemes match the observed monthly streamflow ~~climatology~~ at Dakah
566 well, but monthly streamflow is underestimated in most of months at Kama and Asmar under the
567 basin outlet calibrations, particularly by the Lump-Outlet. The historical monthly streamflow
568 ~~estimates~~ ~~streamflow~~ ~~climatology~~ from the outlet calibration strategies also tends to be highly
569 uncertain for the months of June, July, August, and September, especially compared to the
570 SemiPool and DistPool.

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

578 of good performance at this site for both calibration and validation periods, it is notable that they
579 project future streamflow ~~climatology~~ somewhat differently.

580 Future monthly streamflow predictions ~~climatology~~ at Kama and Asmar vary widely
581 between the four calibration schemes, mostly an artifact of their historic differences (Figure 910).
582 Streamflow projections under the outlet calibration strategies tend to show large uncertainties at
583 these two sites, particularly the Lump-Outlet calibration. For three months, July through
584 September, the outlet calibration and pooled calibration strategies provide substantially different
585 insights about future water availability at Kama and Asmar. The outlet calibrations suggest less
586 water with large uncertainties for those months as compared to the pooled calibrations. At Kama,
587 the pooled calibrations suggest significant changes in the pattern of peak monthly flow timing
588 under both RCP scenarios; instead of having a clear peak in July, streamflow from May to August
589 show similar amounts of water.

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

601 Up to this point, there has been little difference between the Semi-Pooled and Dist-Pooled
602 model variants. These two versions were further analyzed with respect to extreme streamflow to
603 see if distinguishing characteristics emerge. It has been demonstrated that clear gains in predicting
604 peak flows from distributed models are noticeable (Reed et al., 2004) and spatial variability in
605 model parameters significantly influence the runoff behavior (Brath and Montanari, 2000; Pokhrel
606 and Gupta, 2011). The spatial variability of optimal parameters derived from the Semi-Pooled and
607 Dist-Pooled ~~was is~~ shown in Figure S6, with larger variability across all parameters for the Dist-
608 Pooled than for the Semi-Pooled. To understand the effects of ~~parameter spatial~~ variability and
609 ~~calibration~~ uncertainty ~~of parameters~~ on extreme event estimation, the ~~100-year flood~~100-year
610 daily flood event was calculated under the Semi-Pooled and Dist-Pooled for each of the 50 historic
611 simulations and 1800 future simulations across both RCP scenarios. Although the inter-model
612 comparison is intended to be a useful addition that provides a distinction between the
613 parameterization schemes in the pooled calibration approach, results from this analysis should be
614 viewed in the context of a theoretical calibration exercise, not for decision-making purposes,
615 because no observed daily streamflow is available against which to compare the estimated 100-
616 year daily flood events. While no observed data is available against which to compare the results,
617 an inter-model comparison is useful to distinguish the differences between the parameterization
618 schemes. Projections of the ~~100-year flood~~100-year daily flood, estimated using a Log-Pearson
619 type III distribution fit to annual peaks of 30 years, differ somewhat between the Semi-Pooled and
620 Dist-Pooled (Figure ~~4+12~~). At 3 validation sites, extreme floods are consistently larger under the
621 Semi-Pooled than the Dist-Pooled, and the mean difference in the ~~100-year flood~~100-year daily
622 flood estimate between the two calibration approaches grows between the historic runs and the
623 RCP 4.5 and 8.5 scenarios. This suggests that the flood-generation process is fundamentally

624 different between the two parameterizations, with the Semi-Pooled formalization magnifying the
625 effect of climate change on extremes. Furthermore, there is substantially more uncertainty in the
626 ~~100-year flood~~100-year daily flood estimate under the Semi-Pooled. Figure ~~11-12~~ shows the
627 combined uncertainty across both climate projections and calibrations, but this uncertainty is
628 broken down further in Figure ~~1213~~. Similar to Figure ~~1011~~, 3 sources of uncertainty are evaluated
629 for the ~~100-year flood~~100-year daily flood, including ~~parameter calibration~~ uncertainty alone,
630 climate projection uncertainty alone, and their combined effect. For both the Semi-Pooled and
631 Dist-Pooled, ~~parameter calibration~~ uncertainty has a smaller influence than projection
632 uncertainties, and for all sites, the Dist-Pooled has a smaller uncertainty range than the Semi-
633 Pooled, even for ~~parameter calibration~~ uncertainty alone. This was a truly surprising result, given
634 the parametric freedom in the Dist-Pooled model and the fact that no daily data was ever used in
635 the calibration of either model. It appears that a lack of model parsimony does not necessarily lead
636 to greater uncertainty in model simulations under different climate conditions, somewhat counter
637 to what would be expected of over-fit models. One possible reason for this result would be if
638 increased parametric freedom somehow offset the effects of structural deficiencies in the model.
639 However, further research is needed to investigate this issue.

640

641 **6. Discussion and Conclusion**

642 In this study we examined a variety of calibration experiments to better understand the
643 benefits and costs associated with different calibration choices for a complex, distributed
644 hydrologic model in a data-scarce region. The goal of these experiments was to provide insight
645 regarding the use of multisite data in calibration, the effects of parameter complexity, and the

646 challenges of using limited data for distributed model calibration, all in the context of projecting
647 future streamflow under climate change.

648 This study tested two multi-site calibration strategies, the stepwise and pooled approaches,
649 finding that the pooled approach using all data simultaneously provides improved calibration
650 results. This suggests that small sacrifices of model performance at certain sites can improve and
651 stabilize basin-wide performance. The pooled calibration substantially improves with increasing
652 parameter complexity at the calibration sites, but ~~the~~ similar streamflow predictions at the
653 validation sites between the semi-distributed and distributed pooled calibrations were found,
654 suggesting over-fitting of the model from the fully distributed conceptualization. It is worth noting
655 that for the transformation of rainfall to runoff, up to five or six parameters can be identified on
656 the basis of a single hydrograph (Wagner et al., 2001). Under this premise, the number of the
657 HYMOD_DS parameters being calibrated in the semi-distributed approach remains realistic, but
658 the fully distributed parameterization scheme likely causes poor identifiability of the parameters.
659 Thus, pursuing a parsimonious configuration (e.g. optimization for a small portion of the
660 parameters) with an effort to increase the amount of information (e.g. multivariable/multisite) is
661 critical in the calibration of watershed system models (Gupta et al., 1998; Efstratiadis et al., 2008).
662 We also note the important role of experienced hydrologists in designing a parsimonious
663 hydrologic calibration (e.g. Boyle et al., 2000). In this study, the feasible ranges of the
664 HYMOD_DS parameters were kept wide (as is often done in automatic hydrologic calibrations)
665 without consideration of the physical properties of the basin; the judgment of local hydrologic
666 experts could help reduce the feasible ranges used during the calibration and thus contribute to a
667 reduction of calibration uncertainty.

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

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

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

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

692 The semi-distributed pooled and distributed pooled calibrations are very similar for
693 monthly streamflow projections, yet differ in their projections of extreme flows in part due to their
694 differences in the spatial variability of optimal parameters, with the distributed pooled calibration
695 showing less uncertainty for ~~100-year flood~~100-year daily flood events. We evaluated the separate
696 and joint influence of uncertainties in parameter estimation and future climate on projections of
697 seasonal streamflow and ~~100-year flood~~100-year daily flood across calibration schemes and found
698 that the uncertainty resulting from variations in projected climate between the CMIP5 GCMs
699 substantially outweighs the calibration uncertainty. These results agree with other studies showing
700 the dominance of GCM uncertainty in future hydrologic projections (Chen et al., 2011; Exbrayat
701 et al., 2014). While the GCM-based simulations still have widespread use in assessing the impacts
702 of climate change on water resources availability, the bounds of uncertainty resulting from an
703 ensemble of GCMs cannot be well-defined because of the low credibility with which GCMs are
704 able to produce timeseries of future climate (Koutsoyiannis et al., 2008). This issue hinders a
705 straightforward appraisal of future water availability under climate change and has motivated other
706 efforts; e.g. performance-based selection of GCMs (Perez et al., 2014).

707 In addition to the uncertainties surrounding model parameters and future climate explored
708 in this study, there is also significant uncertainty in streamflow projections stemming from
709 structural differences between applied hydrologic models, which can be especially pertinent where
710 robust calibration is hampered by the scarcity of data (Exbrayat et al., 2014). Further, the residual
711 error variance of hydrologic model simulations would increase the effects of hydrologic model
712 uncertainty as compared to that of the climate projections (Steinschneider et al., 2014). These

713 issues need to be addressed in future work for exploring a comprehensive uncertainty assessment
714 of climate change risk for poorly monitored hydrologic systems.

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

733

734 **Acknowledgements**

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

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

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

742

743 **References**

- 744 Ahmad, S.: Towards Kabul Water Treaty: Managing Shared Water Resources - Policy Issues and
745 Options, Karachi, Pakistan, 2010.
- 746 Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed
747 hydrologic model for streamflow estimation along a river system, J HYDROL, 298, 112-135,
748 2004.
- 749 Anderson, J., Refsgaard, J. C., and Jensen, K. H.: Distributed hydrological modeling of the Senegal
750 river basin - model construction and validation, J HYDROL, 247(3-4), 200-214, 2001.
- 751 Bandaragoda, C., Tarboton, D. G., and Woods, R.: Application of TOPNET in the distributed
752 model intercomparison project, J HYDROL, 298, 178-201, 2004.
- 753 Beven, J. K.: Rainfall-Runoff Modelling: The Primer, 2nd Edition, Wiley-Blackwell, Chichester,
754 2012.
- 755 Beven, K.: How far can we go in distributed hydrological modelling?, HYDROL EARTH SYST
756 SC, 5(1), 1-12, 2001.
- 757 Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic
758 modelling of complex environmental systems using the GLUE methodology, J HYDROL, 249,
759 11-29, 2001.
- 760 [Boscarello, L., Ravazzani, G., and Mancini, M.: Catchment multisite discharge measurements for](#)
761 [hydrological model calibration, Procedia Environmental Sciences, 19, 158-167, 2013.](#)
- 762 Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., and Montanari, A.: Validation of hydrological
763 models: conceptual basis, methodological approaches and a proposal for a code of practice, Phys.
764 Chem. Earth, 42-44, 70-76, 2012
- 765 ~~[Boleh, T., Kulkarni, A., Kaab, A., Hyggel, C., Paul, F., Cogley, J. G., Frey, H., Kargel, J. S., Fujita,](#)~~
766 ~~[K., Scheel, M., Bajracharya, S., and Stoffel, M.: The State and Fate of Himalayan Glaciers,](#)~~
767 ~~[SCIENCE, 336, 310-314, 2012](#)~~

768 Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic
769 models: Combining the strengths of manual and automatic methods, Water Resources research,
770 36(12), 3663-3674, 2000.

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

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

775 Brath, A., Montanari, A., and Toth, E.: Analysis of the effects of different scenarios of historical
776 data availability on the calibration of a spatially-distributed hydrological model, J HYDROL, 291,
777 232-253, 2004.

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

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

786 Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological
787 impacts of climate change for a Canadian watershed, Water Resources Research, 47, W12509,
788 2011.

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

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

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

795 Eckhardt, K., Fohrer, N., and Frede, H. G.: Automatic model calibration, *HYDROL PROCESS*,
796 19, 651-658, 2005.

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

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

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

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

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

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

812 Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time
813 consistent model parameters: sub-period calibration, *HYDROL EARTH SYST SC*, 17, 149-161,
814 2013.

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

816 [Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Towards improved calibration of hydrologic
817 models: Multiple and noncommensurable measures of information, *Water Resources Research*,
818 34, 751-763, 1998.](#)

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

822 Hamon, W. R.: Estimating potential evapotranspiration, J HYDR ENG DIV-ASCE, 87, 107-120,
823 1961.

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

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

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

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

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

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

839 Khakbaz, B., Imam, B., Hsu, K., and Sorooshian, S.: From lumped to distributed via semi-
840 distributed: Calibration strategies for semi-distributed hydrologic models, J HYDROL, 418-419,
841 61-77, 2012.

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

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

848 Kinouchi, T., Liu, T., Mendoza, J., and Asaoka, Y.: Modeling glacier melt and runoff in a high-
849 altitude headwater catchment in the Cordillera Real, Andes, *HYDROL EARTH SYST SC*, 10,
850 13093-13144, 2013.

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

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

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

857 Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model
858 parameterization on water fluxes at multiple scales and locations, *WATER RESOUR RES*, 49(1),
859 360-379, 2013.

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

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

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

868 Leavesley, G. H., Hay, L. E., Viger, R. J., and Markstrom, S. L.: Use of Priori Paramter-Estimation
869 Methods to Constrain Calibration of Distributed-Parameter Models, *WATER SCI APPL*, 6, 255-
870 266, 2003.

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

873 [Lerat, J., Andreassian V., Perrin, C., Vaze, J., Perraud J. M., Ribstein, P., and Loumagne C.: Do](#)
874 [internal flow measurements improve the calibration of rainfall-runoff models?, WATER RESOUR](#)
875 [RES, 48, W02511, 2012.](#)

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

878 Lohmann, D., Raschke, R., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology: I.
879 Formulation of the VIC-2L model coupled to a routing model, HYDROLOG SCI J, 43(1), 131-
880 141, 1998.

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

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

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

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

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

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

893 [Olson, S. A., and Williams-Sether, T.: Streamflow characteristics at streamgages in Northern](#)
894 [Afghanistan and selected locations, U. S. Geological Survey, Reston, Virginia, 2010.](#)

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

897 [Perez, J., Menendez, M., Mendez, F. J., and Losada, I. J.: Evaluating the performance of CMIP3](#)
898 [and CMIP5 global climate models over the north-east Atlantic region, Climate Dynamics, 43,](#)
899 [2663-2680, 2014.](#)

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

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

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

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

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

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

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

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

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

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

925 Smith, M., Koren, V., Zhang, Z., Moreda, F., Cui, Z., Cosgrove, B., Mizukami, N., Kitzmiller, D.,
926 Ding, F., Reed, S., Anderson, E., Schaake, J., Zhang, Y., Andreassian, V., Perrin, C., Coron, L.,

927 Valery, A., Khakbaz, b., Sorooshian, S., Behrangi, A., Imam, B., Hsu, K. L., Todini, E., Coccia,
928 G., Mazzetti, C., Andres, E. O., Frances, F., Orozco, I., Hartman, R., Henkel, a., Fickenscher, P.,
929 and Staggs, S.: The distributed model intercomparison project - Phase 2: Experiment design and
930 summary results of the western basin experiments, J HYDROL, 507, 300-329, 2013.

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

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

937 [Steinschneider, S., Wi, S., and Brown, C.: The integrated effects of climate and hydrologic](#)
938 [uncertainty on future flood risk assessments, Hydrological Processes, DOI: 10.1002/hyp.10409,](#)
939 [2014.](#)

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

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

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

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

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

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

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

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

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

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

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

967 [World Bank: Afghanistan – Scoping strategic options for development of the Kabul River Basin:](#)
968 [a multisectoral decision support system approach, World Bank, Washington, D. C., 2010.](#)

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

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

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

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

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

983

984 **Tables**

985

986

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

987

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

988

<u>USGS/ GRDC</u>	<u>Daronta</u>	<u>Kabul</u>	<u>1959/10</u>	<u>1964/9</u>	<u>34.375</u>	<u>0.3</u>	<u>2.722</u>	<u>350</u>	<u>8.0</u>	<u>165</u>
-----------------------	----------------	--------------	----------------	---------------	---------------	------------	--------------	------------	------------	------------

989

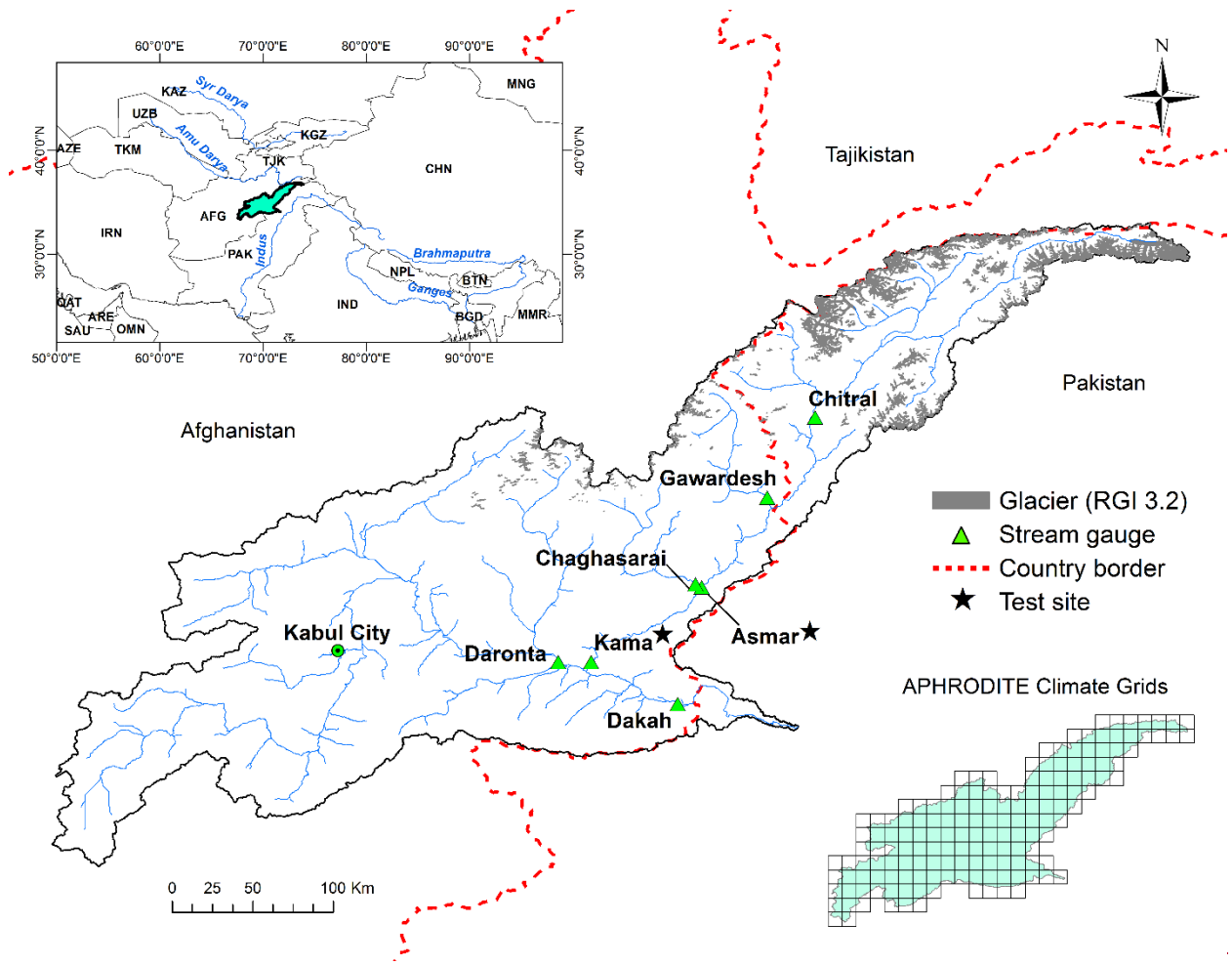
990

Table 2 HYMOD_DS parameters.

Parameter Name	Description	Feasible Range	
		Lower Bound	Upper Bound
<i>Coeff</i>	Hamon potential evapotranspiration coefficient	0.1	2
C_{\max}	Maximum soil moisture capacity [mm]	5	1500
<i>B</i>	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 [$\text{mm}\cdot^{\circ}\text{C}\cdot\text{day}^{-1}$]	0.001	10
T_{th}	Snow melt temperature threshold [$^{\circ}\text{C}$]	0	5
T_s	Snow/rain temperature threshold [$^{\circ}\text{C}$]	0	5
<i>r</i>	Glacier melt rate factor	1	2
K_g	Glacier storage release coefficient	0.01	0.99
T_g	Glacier melt temperature threshold [$^{\circ}\text{C}$]	0	5
<i>N</i>	Unit hydrograph shape parameter	1	99
K_q	Unit hydrograph scale parameter	0.01	0.99
<i>Velo</i>	Wave velocity in the channel routing [$\text{m}\cdot\text{s}^{-1}$]	0.5	5
<i>Diff</i>	Diffusivity in the channel routing [$\text{m}^2\cdot\text{s}^{-1}$]	200	4000

991

992



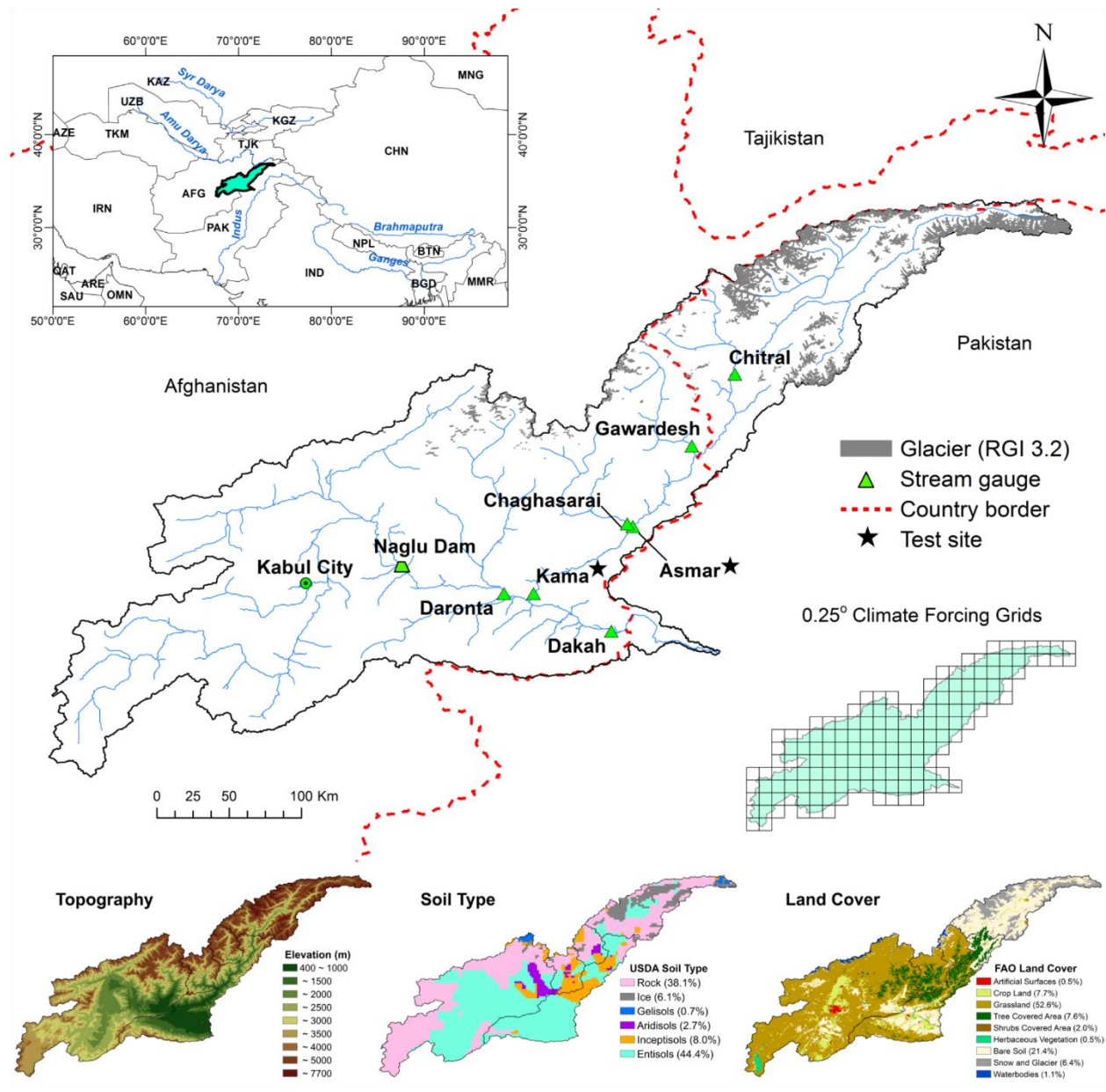
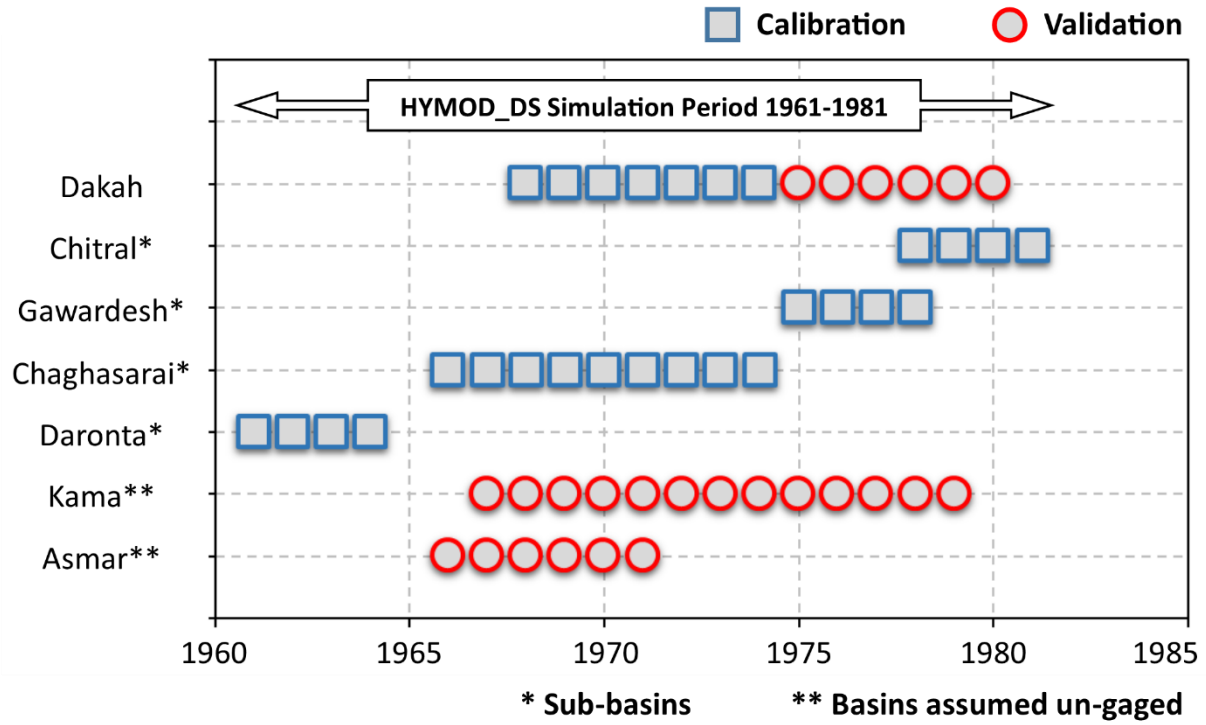


Figure 1. Kabul River basin.

995

996

997

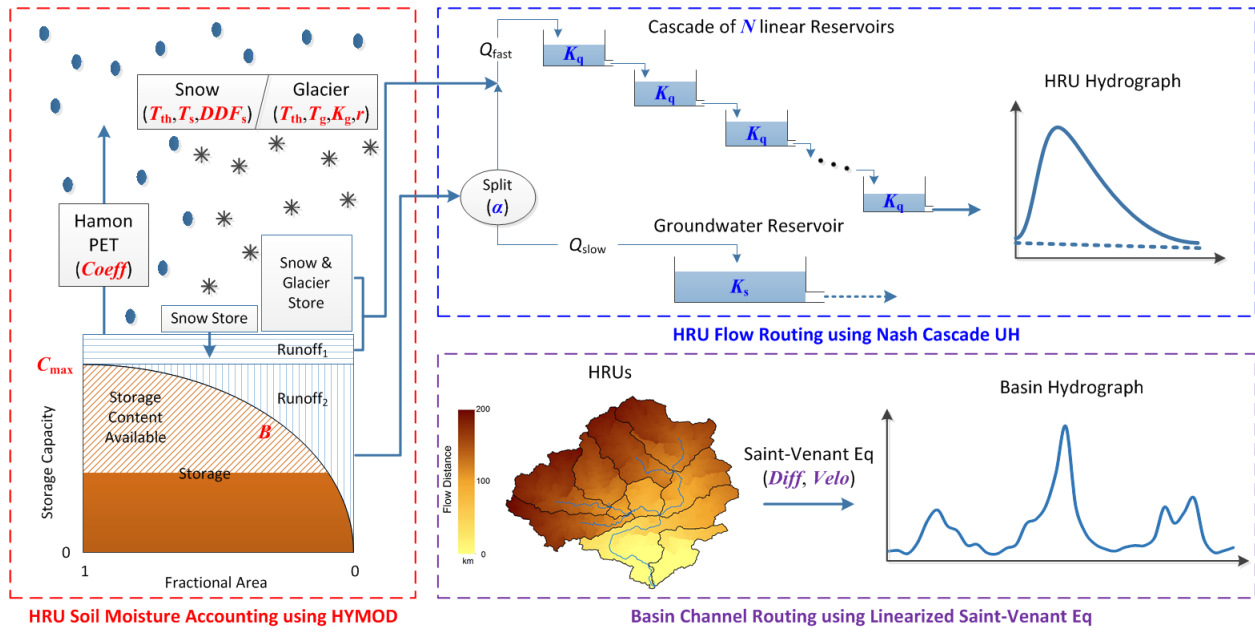


998

999

1000

Figure 32. Streamflow data usage for the model calibration and validation.


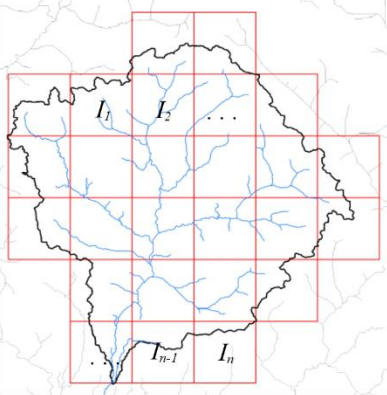

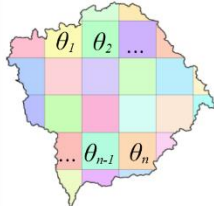


1001

1002

1003

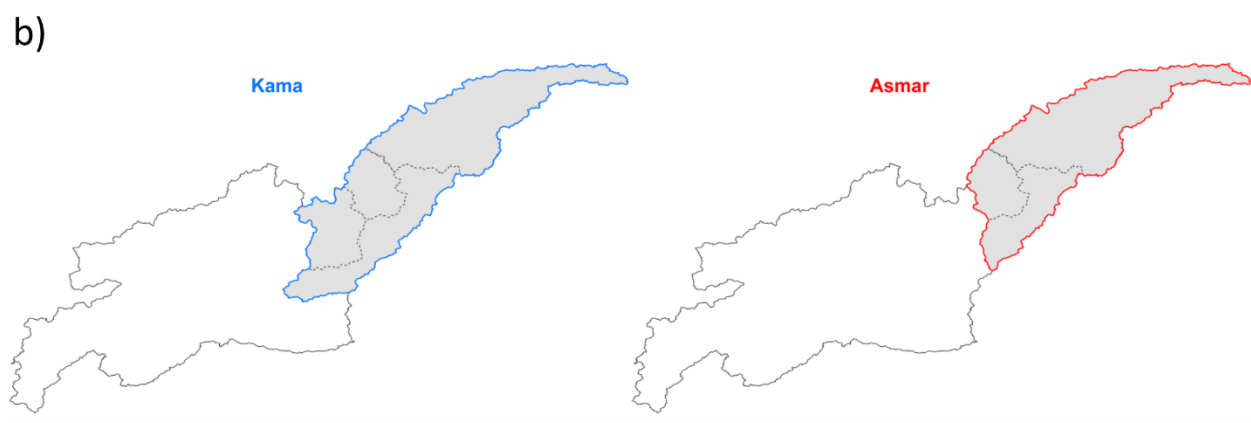
Figure 43. Distributed version of HYMOD model (HYMOD_DS).

	Model Structure	Parameter Structure
Lumped	I_i : Grid Input Set $I_1 \neq I_2 \neq \dots \neq I_{n-1} \neq I_n$ n : Number of Grids	 θ : Single Parameter Set
Semi-Distributed		 θ_i : Sub-Basin Parameter Set $\theta_1 \neq \theta_2 \neq \dots \neq \theta_{n-1} \neq \theta_n$ n : Number of Sub-Basins
Distributed		 θ_i : Grid Parameter Set $\theta_1 \neq \theta_2 \neq \dots \neq \theta_{n-1} \neq \theta_n$ n : Number of Grids

1004

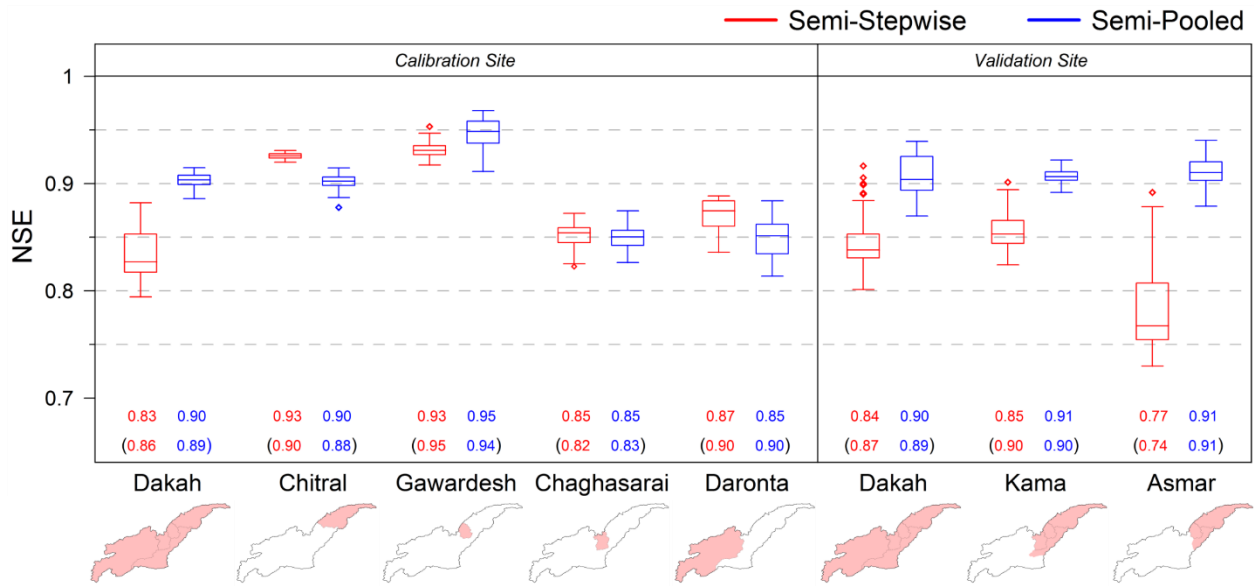
1005 Figure 24. Model structure based on climate input grids and three different parameterization
1006 concepts.

1007



1008
1009
1010
1011

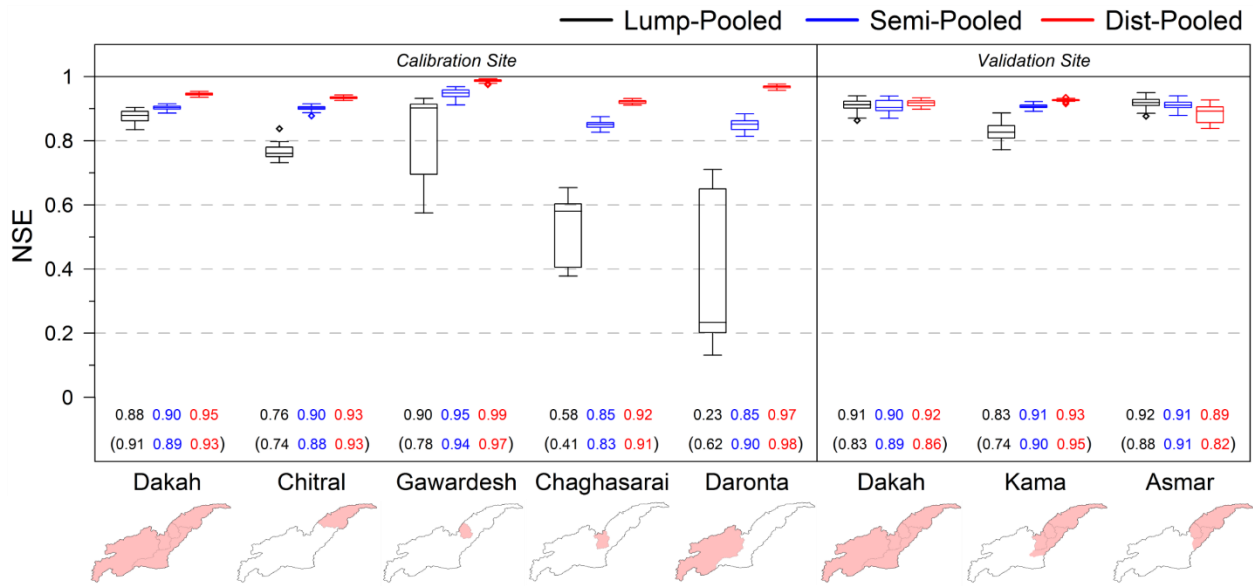
Figure 5. (a) Sub-basins corresponding to five gaging stations are used for the multisite calibrations. (b) Two sub-basins (Kama and Asmar) are assumed to be unaged and used for evaluating the calibration approaches.



1012

1013 Figure 56. Comparison of the stepwise and pooled calibrations under the semi-distributed
 1014 parameterization. Each calibration is conducted 50 times. Values on the bottom represent expected
 1015 values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

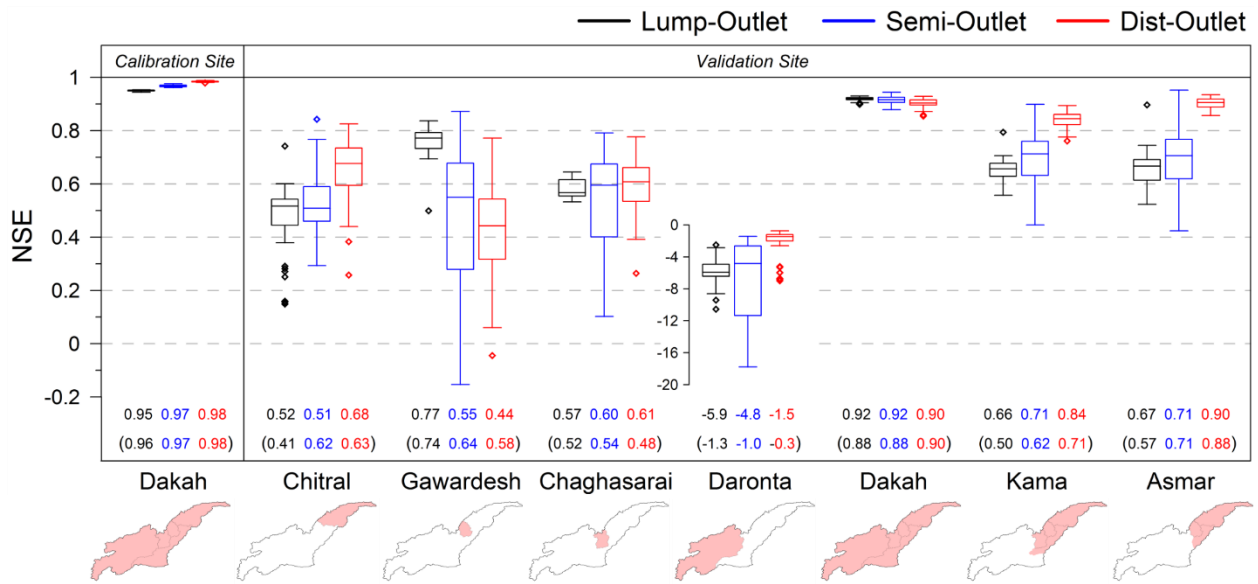
1016



1017

1018 Figure 67. Comparison of the pooled calibrations for the 3 parameterizations of lumped, semi-
 1019 distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent
 1020 expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50
 1021 calibrations.

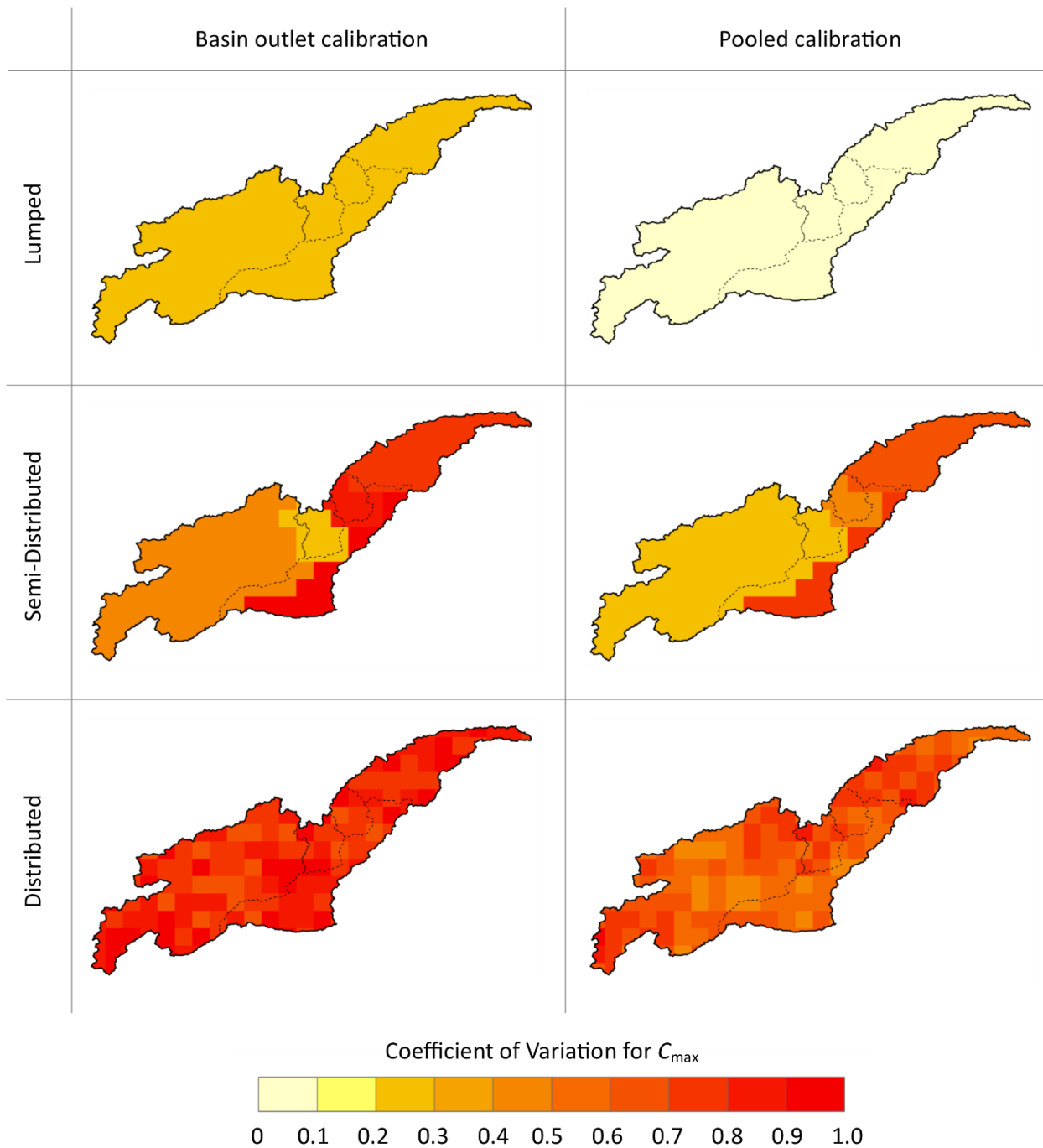
1022



1023

1024 Figure 78. Comparison of the basin outlet calibrations for the 3 parameterizations of lumped, semi-
 1025 distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent
 1026 expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50
 1027 calibrations.

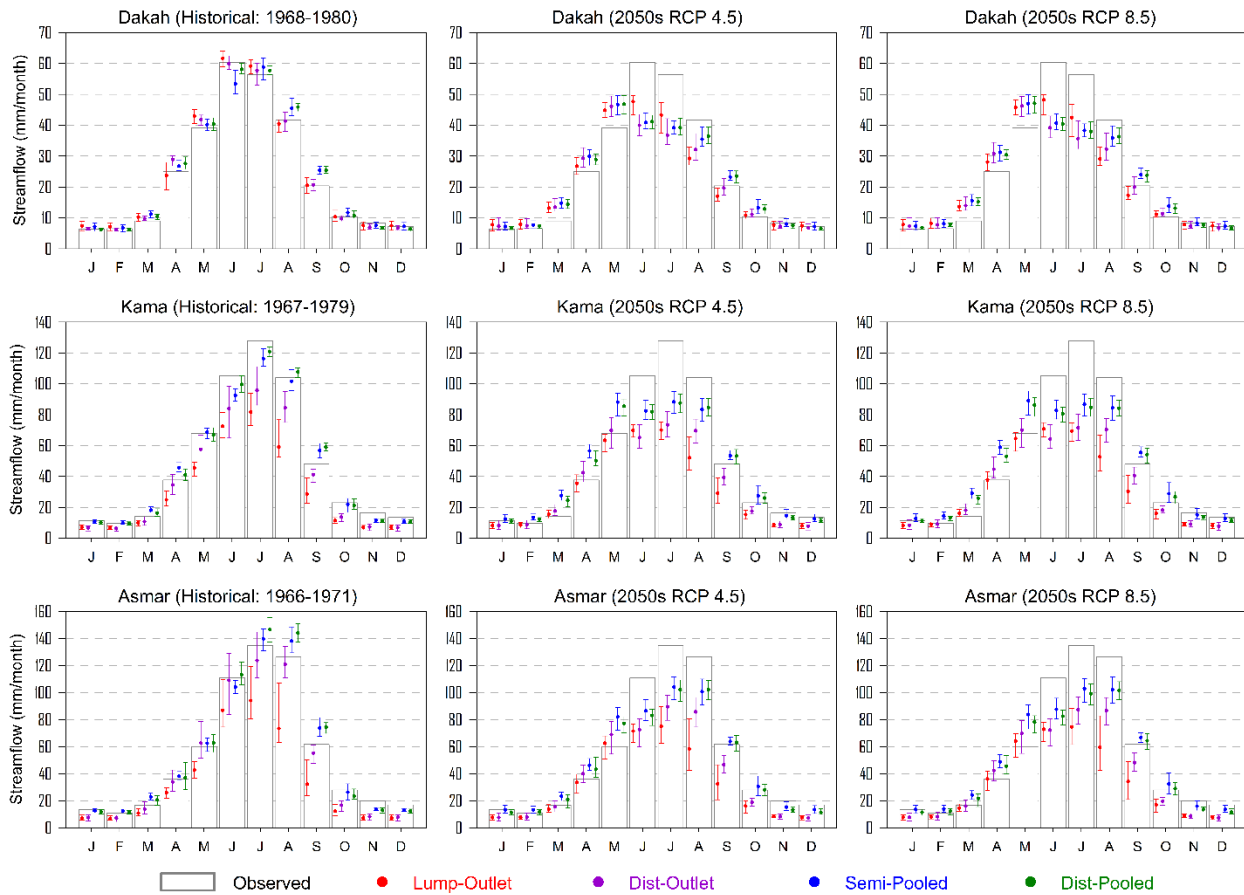
1028



1029

1030 Figure 89. Coefficient of variation (CV) of 50 optimal values of C_{max} (parameter for the soil
 1031 moisture accounting module in the HYMOD_DS) from the basin outlet calibrations (left panel)
 1032 and the pooled calibrations (right panel).

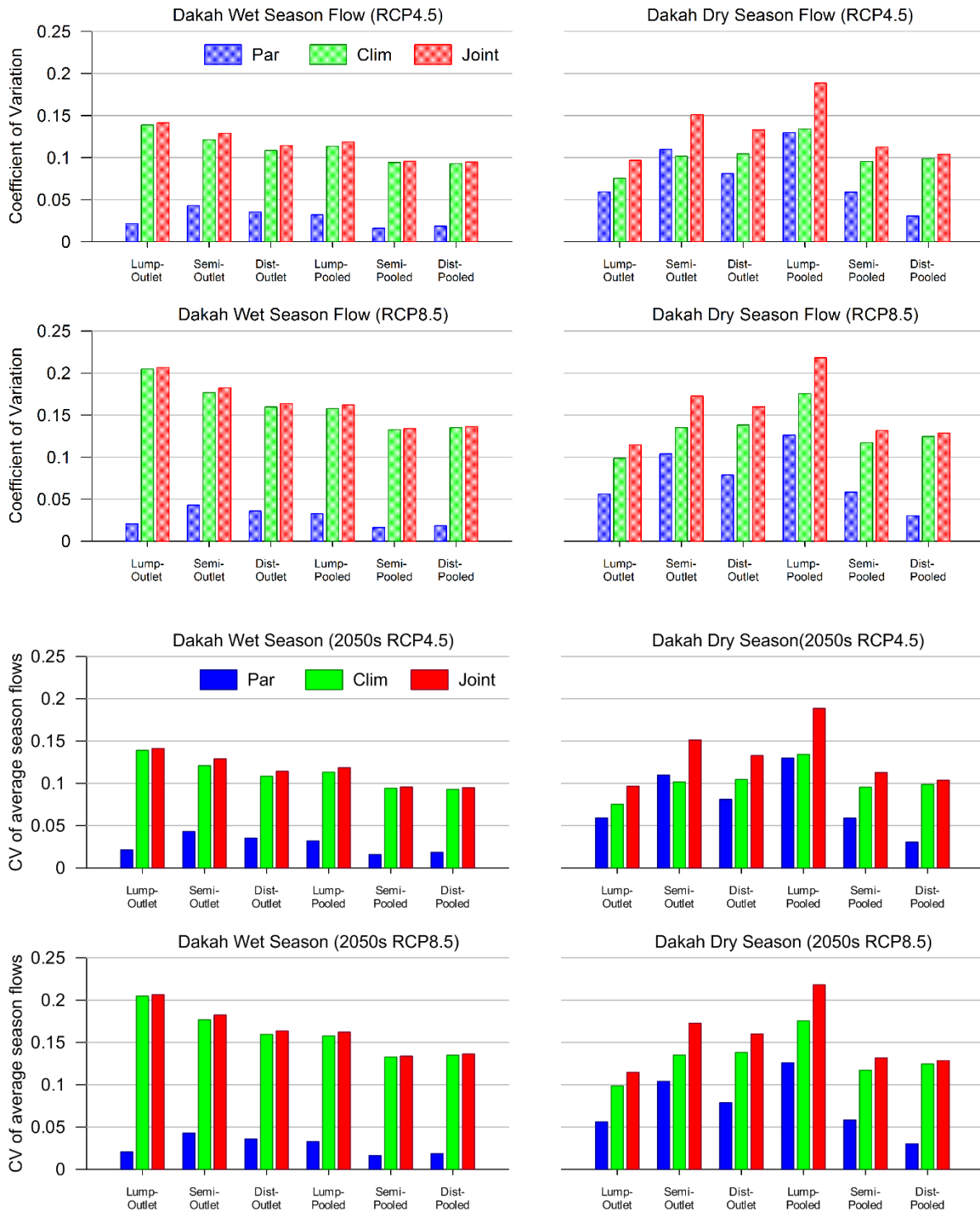
1033



1034

1035 Figure 910. Historical and 2050s average monthly streamflow ~~climatology~~ predictions at Dakah,
 1036 Kama, and Asmar under 4 calibration strategies: Lump-Outlet, Dist-Outlet, Semi-Pooled, and Dist-
 1037 Pooled. The error bars represent the streamflow ranges resulting from 50 trials of the HYMOD_DS
 1038 calibration. For each of the 50 trials, the 2050s streamflow predictions are averaged over 36 GCM
 1039 climate projections.

1040



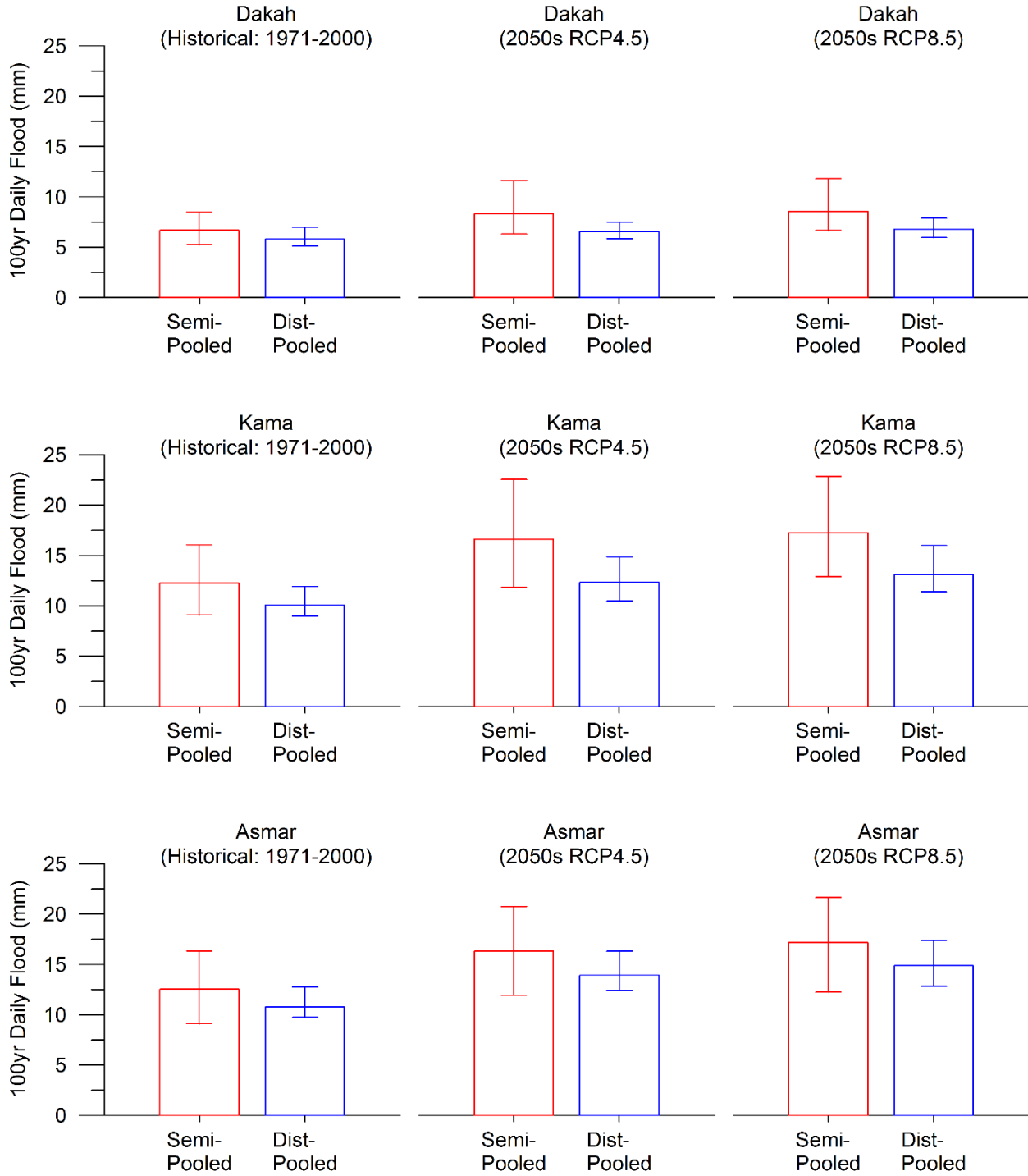
1041

1042

1043 Figure 1011. Uncertainties in wet and dry season 2050s average streamflow climatology
 1044 predictions for 2050s of wet and dry seasons are derived from the basin outlet and pooled
 1045 calibrations for Dakah. Uncertainties are evaluated by coefficient of variation (CV) of average

1046 season streamflow predictions. Three uncertainty sources are considered: parameter calibration
1047 uncertainty across 50 calibration trials (Par), climate uncertainty across GCM projections (Clim),
1048 and combined uncertainty (Joint).

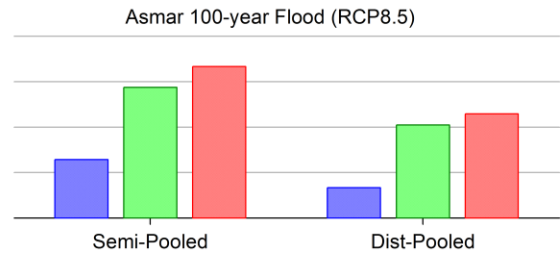
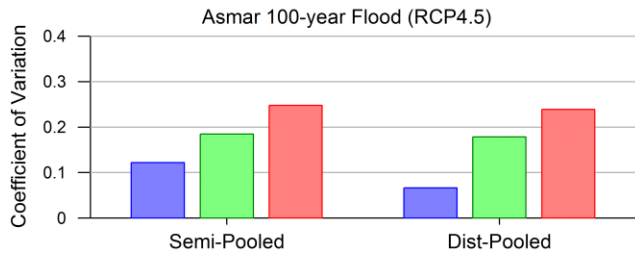
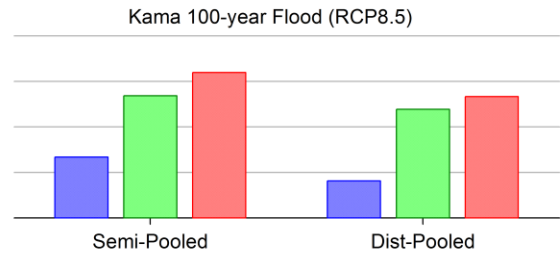
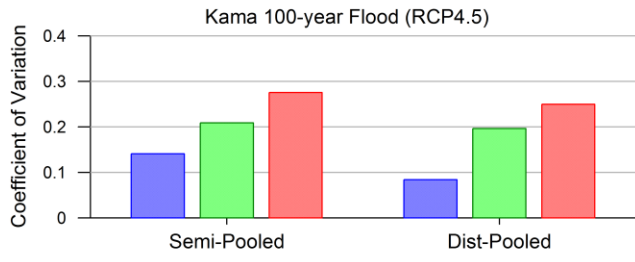
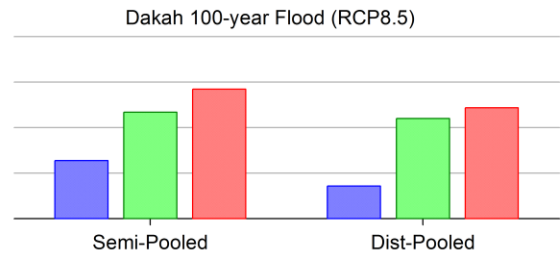
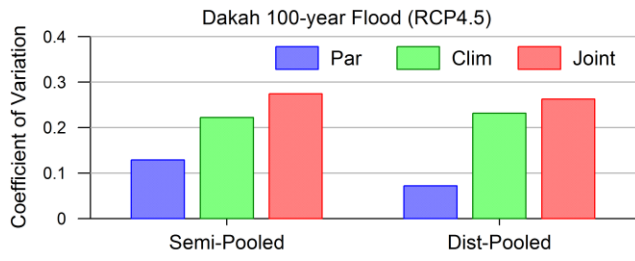
1049

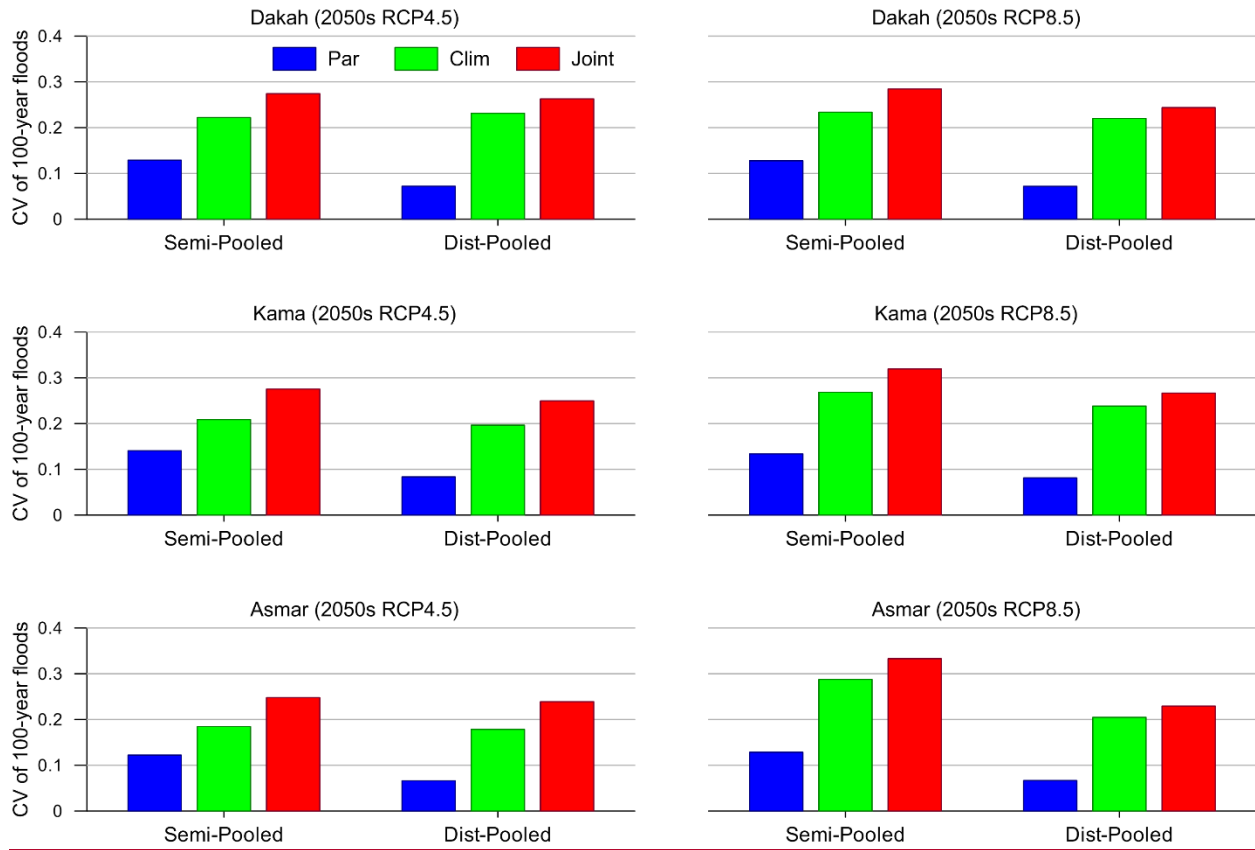


1050

1051 Figure 4.12. Comparison of GCM average 100-year flood events derived
 1052 from the semi-distributed and distributed pooled calibrations. The uncertainty range is from 50
 1053 trials of the model calibration.

1054

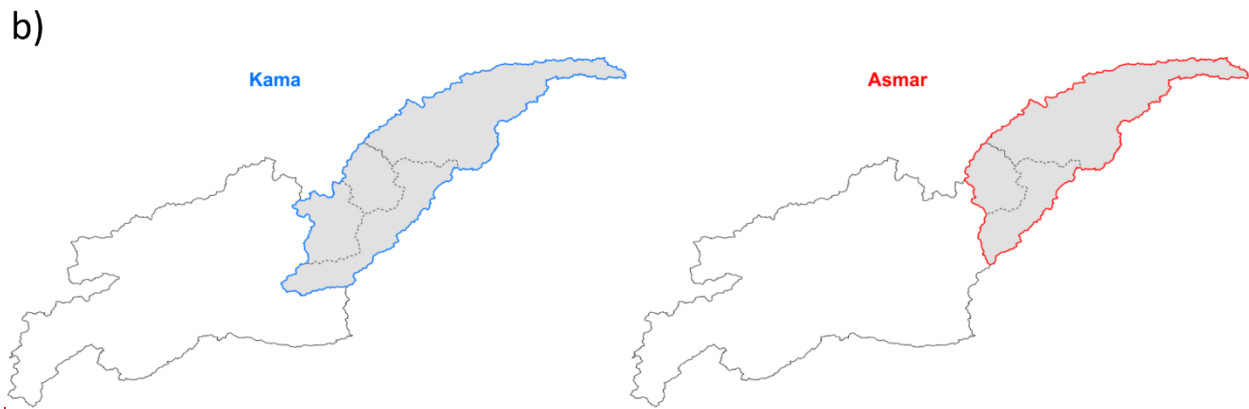




1056

1057 Figure 4213. Uncertainties in 100-year daily flood estimates for 2050s Impact of three
 1058 uncertainties on 100-year flood events are assessed using derived from the Semi-Pooled and Dist-
 1059 Pooled calibrations. Uncertainties are evaluated by calculating coefficient of variation (CV) of
 1060 2050s 100-year flood estimates under three uncertainty sources: calibration uncertainty across 50
 1061 calibration trials (Par), climate uncertainty across GCM projections (Clim), and combined
 1062 uncertainty (Joint).

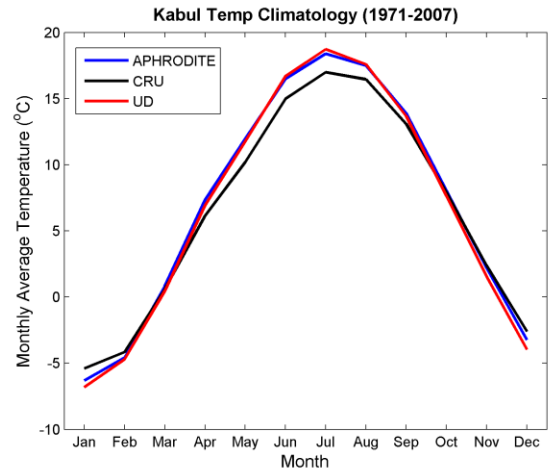
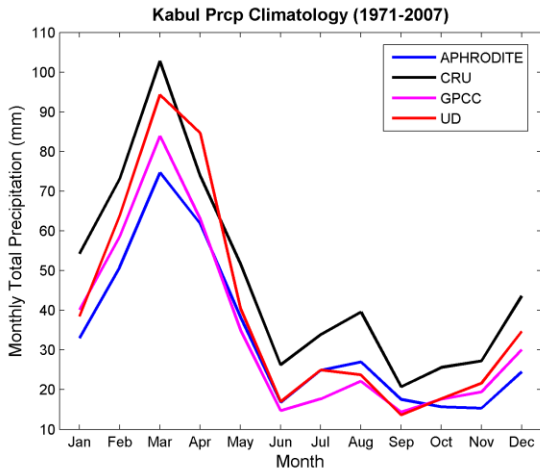
1063



1065

1066 ~~Figure S1. (a) Sub basins corresponding to five gaging stations are used for the multisite~~
1067 ~~calibrations. (b) Two sub-basins (Kama and Asmar) are assumed to be ungaged and used for~~
1068 ~~evaluating the calibration approaches.~~

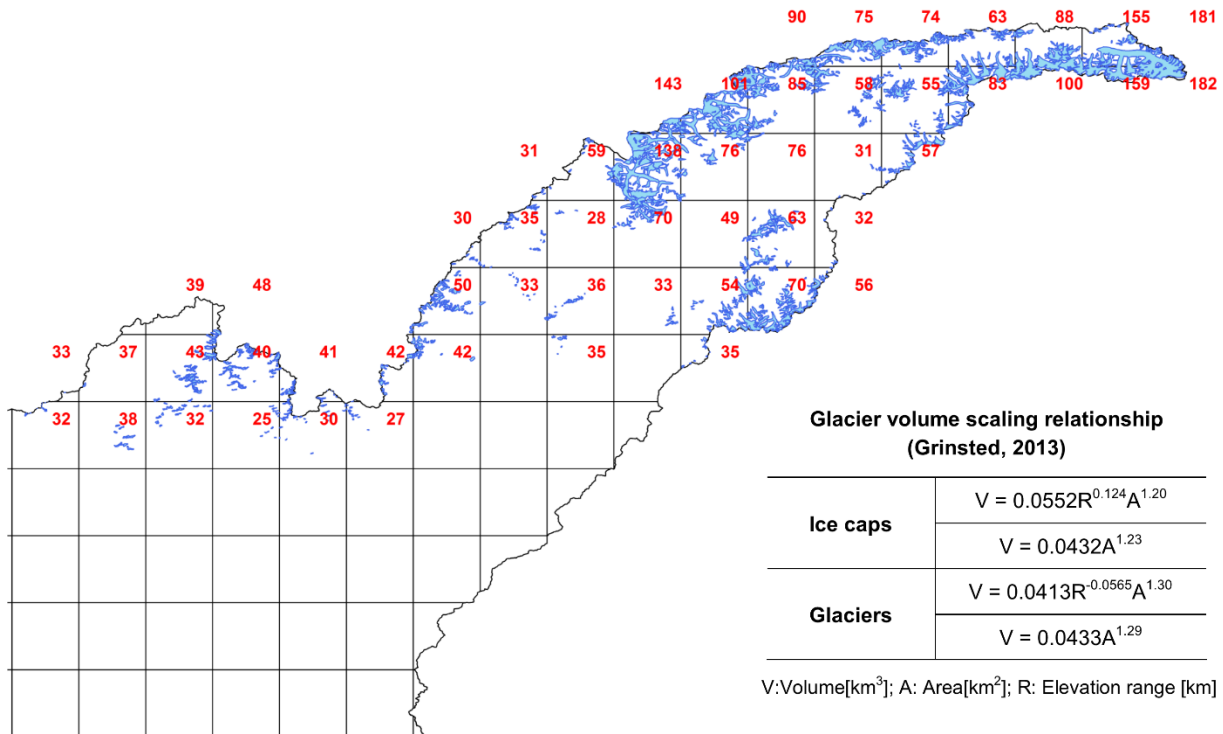
1069



1070

1071 Figure S2S1. Comparison of ~~climatology of~~ basin-wise average monthly precipitation and
 1072 temperature for the Kabul River basin. Sources of data sets: APHRODITE (Asian Precipitation
 1073 High-Resolved Observational Data Integration Towards Evaluation), CRU (Climatic Research
 1074 Unit), GPCC (Global Precipitation Climatology Centre), UD (University of Delaware).

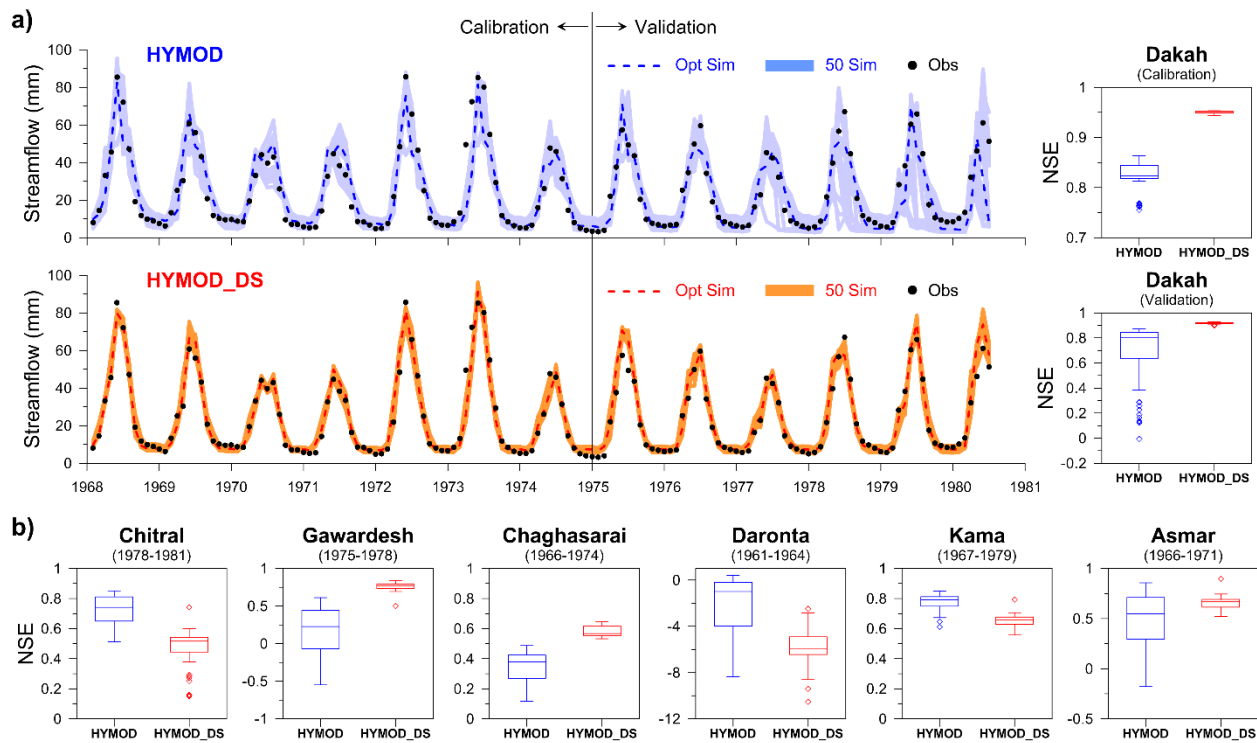
1075



1076

1077 Figure S3S2. Glacial coverage in the Kabul River basin based on the Randolph Glacier Inventory
 1078 version 3.2. Glacier volume scaling relationship proposed by Grinsted (2013) is applied to derive
 1079 glacier volume. Numbers in red represent glacier depths in meter of water for grid cells containing
 1080 glaciers.

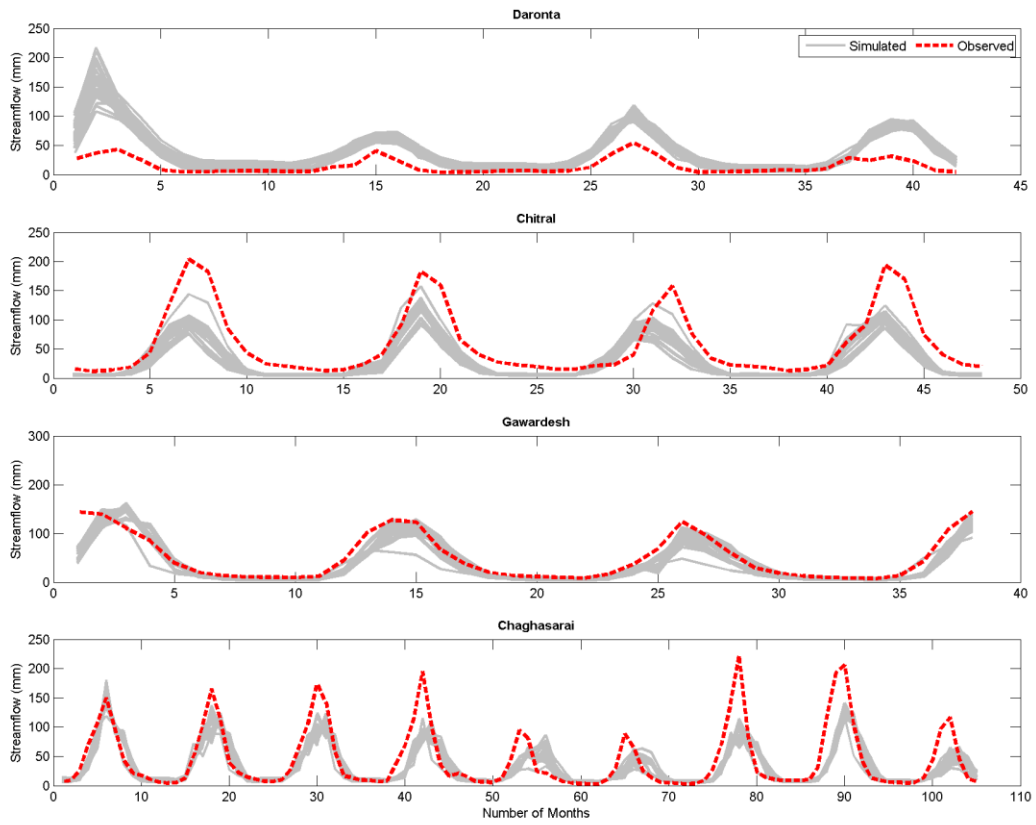
1081



1082
1083
1084
1085
1086
1087

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.

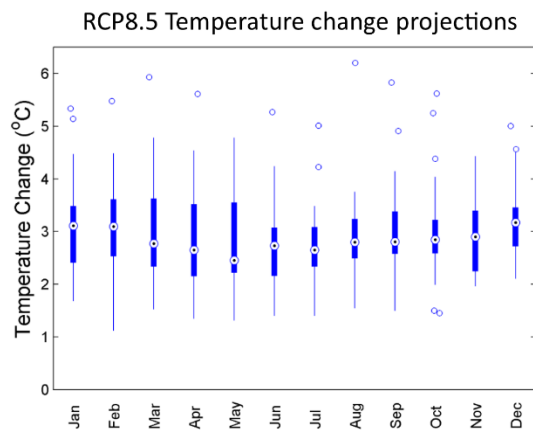
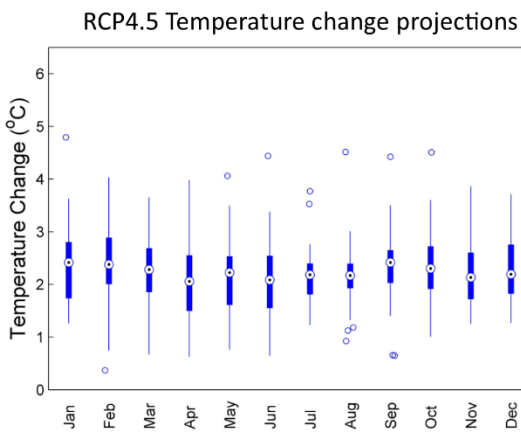
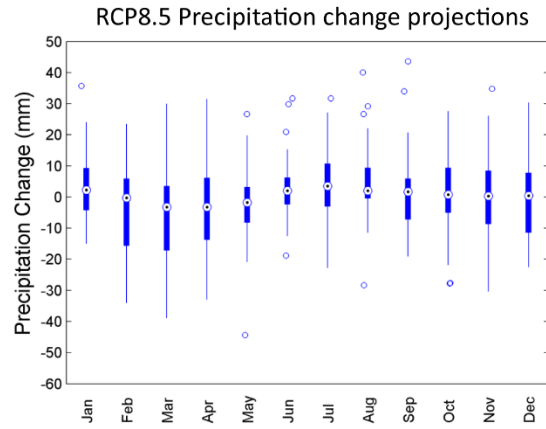
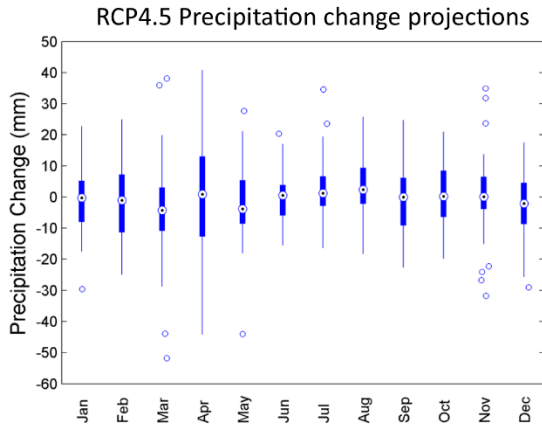
1088



1089

1090 Figure S4. HYMOD_DS streamflow simulations at sub-basins from 50 trials of the basin outlet
1091 calibration under the lumped parameterization.

1092

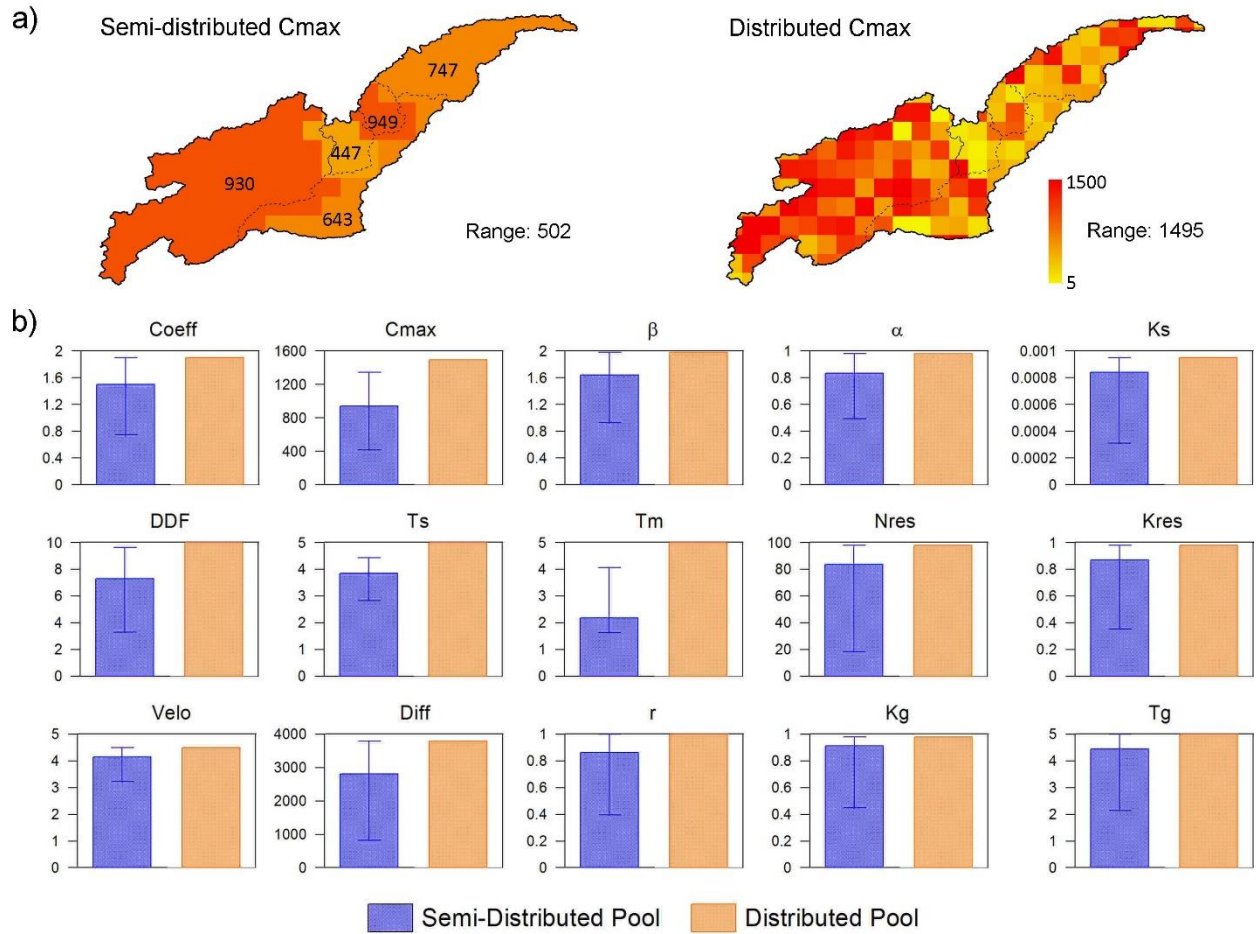


1093

1094 Figure S5. CMIP5 climate change projections of precipitation and temperature for the Kabul basin.

1095 The changes in climatology of average monthly total precipitation and mean temperature for the
 1096 future period 2050s (2036-2065) were calculated from the comparison with the historical period
 1097 (1976-2005). 36 GCMs were employed in this analysis.

1098

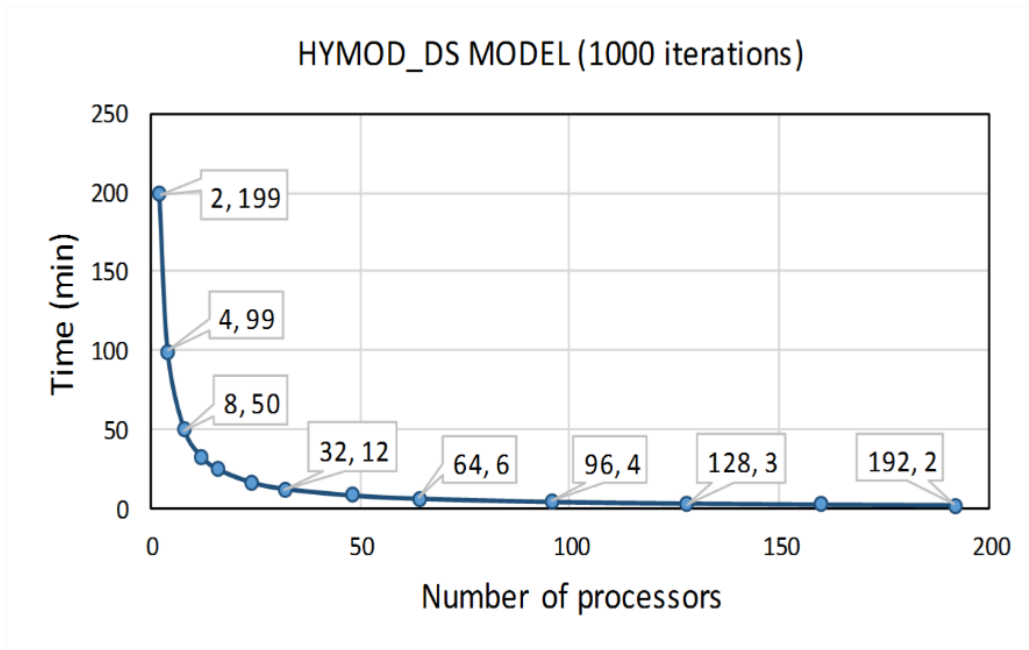


1099

1100 Figure S6. Spatial variability of the HYMOD_DS parameters. a) An example with C_{max} showing
 1101 parameter ranges resulting from the single trail of Semi-Pooled and Dist-Pooled. b) Average
 1102 spatial variability across 50 trials of calibration for all 15 parameters. Error bar in b) represents the
 1103 range of parameter spatial variability from the 50 trails.

1104

1105



1106

1107

Figure S7. HYMOD_DS run time on parallel computing system.

1108