1

Calibration approaches for distributed hydrologic models in poorly gaged basins:

Implication for streamflow projections under climate change

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

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

[2] The World Bank, USA

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

11
12
13
14
15
16
17
18
19
20

Abstract

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

45 variations in projected climate between climate models, which substantially outweighs the

46 calibration uncertainty.

1. Introduction

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

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

70   parameter values that vary across the watershed, especially when observed streamflow data is only

71   available at one or a few points (Exbrayat et al., 2014). Parameters can be discretized across the

72   watershed in several ways (Flugel, 1995; Efstratiadis et al., 2008; Khakbaz, et al., 2012): uniquely

73   for each grid cell or hydrologic response unit (fully distributed), based on sub-basins whose

74   boundaries do not necessarily ensure homogenous characteristics (semi-distributed), or in the

75   simplest case, a single parameter set for all model grid cells (lumped). With limited data, the

76   parameter identification problem, particularly for the fully distributed case, can be impractical or

77   infeasible (Beven, 2001). The parameterization challenge has spurred substantial advances in

78   understanding appropriate calibration techniques for distributed hydrologic models. Many studies

79   have attempted to reduce the dimensionality of the calibration problem to alleviate the issue of

80   equifinality (Beven & Freer, 2001), which is the phenomenon whereby multiple parameter sets

81   produce indistinguishable model performance. This work has found favorable results when the

82   parametric complexity of the distributed model is aligned with the data available for calibration

83   (Leavesley, et al., 2003; Ajami, et al., 2004; Eckhardt, et al., 2005; Frances, et al., 2007; Zhu &

84   Lettenmaier, 2007; Cole & Moore, 2008; Pokhrel & Gupta, 2010; Khakbaz, et al., 2012). There

85   has also been extensive research exploring the use of multiple objectives and different operational

86   procedures to understand parameter estimation tradeoffs and identifiability for distributed model

87   calibration, with great success (Madsen, 2003; Efstratiadis & Koutsoyiannis, 2010; Li, et al., 2010;

88   Kumar, et al., 2013).

89   Despite these advances, important questions still persist. It still remains difficult to

90   compare the uncertainty that emerges from different operational calibration procedures for

91   multisite applications (i.e. whether gages in series should be used sequentially or simultaneously

92   for calibration) and under different levels of parametric complexity. Due to the computational

93   burden required to calibrate distributed models, this uncertainty is problematic to explore. Further,

94   in poorly gaged basins, it is challenging to quantify the lost accuracy and increased uncertainty for

95   interior flow estimation when a distributed model is calibrated only at an outlet gage (which is

96   often all that is available in developing country river basins). In the case of significant spatial

97   variability in the basin properties that influence runoff generation (e.g., permeability, vegetation,

98   slope, etc.), accurate runoff predictions are unlikely at interior locations based only on the lumped

99   information obtained at the basin outlet (Anderson et al., 2001; Cao, et al., 2006; Breuer et al.,

100 2009; Lerat et al., 2012; Simith et al., 2012; Wang, et al., 2012). The extent of this error and

101 uncertainty is not well understood for heterogeneous basins due to the computational expense

102 required to explore this issue. Finally, rarely have the implications of these calibration issues been

103 explicitly examined for possible future climate conditions, which is required in climate change

104 impact studies. This question has been explored for lumped, conceptual models (Wilby, 2005;

105 Steinschneider, et al., 2012), but has been difficult to evaluate for computationally expensive

106 distributed models.

107   This study addresses the above research challenges by focusing on the following four

108 questions: 1) How does calibration procedure for using multisite data effect the accuracy and

109 uncertainty of distributed models used for streamflow predictions at ungaged sites, 2) what effects

110 do increased parameter complexity have on distributed model calibration and prediction, 3) how

111 much degradation in model accuracy and uncertainty can be expected for interior flow estimation

112 based on a calibration procedure using only the basin outlet, and 4) how do different calibration

113 formulations for a distributed model alter projections of streamflow at ungaged sites under climate

114 change conditions? These questions are considered in an application of a distributed version of the

115 daily HYMOD hydrologic model to the Kabul River basin in Afghanistan and Pakistan. To address

116   these research questions, high performance computing is utilized to manage the computational

117   burden that often hinders such explorations (Laloy & Vrugt, 2012; Zhang, et al., 2013).

118

2. Study area

120   The Kabul River basin (67,370km^2) is a plateau surrounded by mountains located in the

121   eastern central part of Afghanistan (Figure 1). It is the most important river basin of Afghanistan,

122   containing 35 percent of the country's population. While it encompasses just 12 percent of the area

123   of Afghanistan, the basin's average annual streamflow (about 24 billion cubic meters) is about 26

124   percent of the country's total streamflow volume (World Bank, 2010).

125   Water resources from the basin are shared by Afghanistan and Pakistan and serve as a water

126   supply source for more than 20 million people. The shared use of transboundary water between

127   these two countries is central in establishing regional water resources development for this area

128   (Ahmad, 2010). It is crucial to develop tools that can support engineering plans for existing and

129   potential water infrastructure to take full advantage of the water resources in the basin. The

130   Government of Afghanistan has developed comprehensive plans for new hydropower projects on

131   the Kabul River owing to its advantageous topography for the development of water storage and

132   hydropower (IUCN, 2010), and recently reached an agreement with the Pakistan government to

133   work on a 1,500MW hydropower project on the Kunar River (one of major tributary in the Kabul

134   River basin) as part of the joint management of common rivers between the two countries (DAWN,

135   2013). The streamflow regime of the Kabul River can be classified as glacial with maximum

136   streamflow in June or July and minimum streamflow during the winter season. Approximately

137   70% of annual precipitation (475mm) falls during the winter season (November to April). While

138 the dominant source of streamflow in winter is baseflow and winter rainfall, glaciers and snow

139 cover are the most important long-term forms of water storage and, hence, the main source of

140 runoff during the ablation period for the basin (Shakir et al., 2010). In total 2.9% (1954km^2) of the

141 basin is glacierized based on the Randolph Glacier Inventory version 3.2 (Pfeffer, et al., 2014).

142 The melt water from glaciers and snow produce the majority (75%) of the total streamflow (Hewitt,

143 et al., 1989). Table 1 provides the climates and geophysical properties of each sub-watershed

144 delineated by the stations located inside the Kabul Basin (Figure 1). Two different climate patterns

145 are distinguishable across the sub-basins. The sub-basins on the Kunar River tributary (Kama,

146 Asmar, Chitral, Gawardesh, and Chaghasarai) receive moderate annual precipitation and are

147 highly affected by snow and glacier covers. All of these sub-basins have high ratios of mean annual

148 flow to mean annual precipitation, with the ratios for the Kama, Asmar, Chitral, and Chaghasarai

149 sub-basins larger than 1. Conversely, the Daronta sub-basin contains only minimal glacial cover,

150 and is relatively dry. Daronta is also much less productive, with annual streamflow far below the

151 other sub-basins with an average of only 165 mm/year.

152   Issues of shared water resources between Afghanistan and Pakistan in the Kabul River

153 basin are becoming complex due to the impacts of climatic variability and change (IUCN, 2010).

154 The vulnerability of glacial streamflow regimes to changes in temperature and precipitation (Stahl,

155 et al., 2008; Immerzeel, et al., 2012; Radic et al., 2014) highlights the need to assess the impact of

156 climate change on future water availability in this area.

157

158 **3. Data and Models**

159

3.1. Data

160

161   Gridded daily precipitation and temperature products with a spatial resolution of 0.25^o were

162   gathered between calendar years 1961-2007 from the Asian Precipitation Highly Resolved

163   Observational Data Integration Towards Evaluation *(*APHRODITE) dataset (Yatagai, et al., 2012).

164   There has been some concern regarding underestimation of precipitation in APHRODITE for some

165   regions of Asia (Palazzi, et al., 2013); our preliminarily data analysis (intercomparison of

166   precipitation products between 5 different databases) confirmed this for the Kabul River basin

167   (shown in Figure S1). Thus, the APHRODITE precipitation was bias-corrected by the precipitation

168   product from the University of Delaware global terrestrial precipitation (UD) dataset (Legates &

169   Willmott, 1990). Daily series of bias-corrected APHRODITE precipitation were coupled with

170   APHRODITE temperature for 160 0.25^o grid cells to produce a climate forcing dataset for the

171   distributed domain of the Kabul River basin model.

172   This study used the set of global climate change simulations from the World Climate

173   Research Programme's Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model

174   ensemble (Talyor, et al., 2012). Monthly climate outputs of GCMs were downscaled to a daily

175   temporal resolution and 0.25^o spatial resolution based on the bias-correction spatial disaggregation

176   (BCSD) statistical downscaling method introduced by Wood et al. (2004).

177   Monthly streamflow observations for seven locations in the Kabul River basin (Figure 1)

178   were gathered between calendar years 1960-1981 from two data sources: the Global Runoff Data

179   Centre (GRDC) database and the United States Geological Survey (USGS) database (Table 1).

180   Streamflow data were not collected in Afghanistan after September 1980 until recently because

181   streamgaging was discontinued soon after the Soviet invasion of Afghanistan in 1979 (Olson and

182   Williams-Sether, 2010). Though measurements were taken at a daily time step, data are only made

183 available for public use at monthly aggregated levels, calculated using the mean of the daily values.

184 The available monthly streamflow observations at each station were used for calibrating and

185 validating the distributed hydrologic model (Figure 2). Kama and Asmar stations are treated as

186 ungaged sites because they align with the potential dam project on the Kunar River tributary. The

187 two gage stations are left out of the processes of multisite calibrations in order to evaluate the

188 model's ability to predict streamflow at interior ungaged sites. Furthermore, half of the record at

189 the Dakah station, located at the basin outlet, is also used for validation purposes.

190 The Randolph Glacier Inventory version 3.2 (RGI 3.2) dataset (Pfeffer, et al., 2014) was

191 used to extract glacial coverage in the Kabul River basin, which totaled 5.7% of the basin area

192 (Figure S2). In the hydrological modeling process, the model needs to be informed by reliable

193 estimates on volume of water retained in glaciers, especially for future simulations under warming

194 conditions. We followed the method proposed in Grinsted (2013), which uses multivariate scaling

195 relationships to estimate glacier and ice cap volume based on elevation range and area.

196 Specifically, the scaling law including area and elevation range factors was applied to estimate

197 glacier/ice cap volume when the glacier depth exceeded 10m. Otherwise, glacier/ice cap volume

198 was estimated with the area-volume scaling law. The elevation range spanned by each individual

199 glacier is estimated using the global digital elevation model (DEM) from the shuttle radar

200 topography mission (SRTMv4) in 250m resolution (Jarvis, et al., 2008). Density of ice (0.9167

201 g/cm^3) is applied to calculate glacier/ice cap volume in meters of water equivalent.

202 The database for land covers and soil types of the Kabul River basin (Figure 1) are provided

203 by the Food and Agriculture Organization of the United Nations (Latham, 2014) and United States

204 Department of Agriculture-Natural Resources Conservation Service Soils (USDA-NRCS, 2005),

205 respectively.

206

3.2. Distributed Hydrologic Model (HYMOD_DS)

207

208   In this study the lumped conceptual hydrological model HYMOD (Boyle, 2001) is coupled

209 with a river routing model to be suitable for modelling a distributed watershed system. We name

210 it HYMOD_DS denoting the distributed version of HYMOD. Snow and glacier modules have

211 been introduced to enhance the modelling process for glacier and snow covered areas within the

212 Kabul River basin. The HYMOD_DS is composed of hydrological process modules that represent

213 soil moisture accounting, evapotranspiration, snow processes, glacier processes and flow routing.

214 The model operates on a daily time step and requires daily precipitation and mean temperature as

215 input variables. The overall model structure of the HYMOD_DS and its 15 parameters are

216 described in Figure 3 and Table 2 respectively. Further details are provided below.

217   The HYMOD conceptual watershed model has been extensively used in studies on

218 streamflow forecasting and model calibration (Wagener, et al., 2004; Vrugt, et al., 2008; Kollat,

219 et al., 2012; Gharari, et al., 2013; Remesan, et al., 2013). The HYMOD is a soil moisture

220 accounting model based on the probability-distributed storage capacity concept proposed by

221 Moore (1985). This conceptualization represents a cumulative distribution of varying storage

222 capacities (C) with the following function:

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

224 where the exponent B is a parameter controlling the degree of spatial variability of storage capacity

225 over the basin and C_{max} is the maximum storage capacity. The model assumes that all storages

226 within the basin are filled up to the same critical level ($C^{*}(t)$), unless this amount exceeds the

227   storage capacity of that particular location. With this assumption, the total water storage S(t)

228   contained in the basin corresponds to

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

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

231   components:

232    $\text{Runoff}_1 = \begin{cases} P(t) + C^*(t-1) - C_{max} & \text{if } P(t) + C^*(t-1) \geq C_{max} \\ 0 & \text{if } P(t) + C^*(t-1) < C_{max} \end{cases}$ (3)

233    $\text{Runoff}_2 = \begin{cases} (P(t) - \text{Runoff}_1) + (S(t) - S(t-1)) & \text{if } P(t) - \text{Runoff}_1 \geq S(t) - S(t-1) \\ 0 & \text{if } P(t) - \text{Runoff}_1 < S(t) - S(t-1) \end{cases}$ (4)

234   where P(t) is precipitation, Runoff_1 is surface runoff, and Runoff_2 is subsurface runoff. A

235   parameter (α) is introduced to represent how much of the subsurface runoff is routed over the fast

236   (Q_{fast}) and slow (Q_{slow}) pathway:

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

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

239   The potential evapotranspiration (PET) is derived based on the Hamon method (Hamon,

240   1961), in which daily PET in mm is computed as a function of daily mean temperature and hours

241   of daylight:

242    $\text{PET} = Coeff \cdot 29.8 \cdot L_d \cdot \frac{0.611 \cdot exp\left(17.27 \cdot \frac{T}{(T+273.3)}\right)}{T+273.3}$ (7)

243   where, L_d is the daylight hours per day, T is the daily mean air temperature (ºC), and Coeff is a

244   bias correction factor. The hours of daylight is calculated as a function of latitude and day of year

245   based on the daylight length estimation model (CBM model) suggested by Forsythe et al. (1995).

246   The HYMOD_DS includes snow and glacier modules with separate runoff processes, i.e.,

247   the runoff from the glacierized area is calculated separately and added to runoff generated from

248   the soil moisture accounting module coupled with the snow module. The implicit assumption here

249   is that there is no interchange of water between soil layers and glacial area and runoff from glacial

250   areas is regarded as surface flow. The runoff from each area is weighted by its area fraction within

251   the basin to obtain total runoff.

252   The time rate of change in snow and glacier volume governed by ice accumulation and

253   ablation (melting and sublimation) is expressed by the Degree Day Factor (DDF) mass balance

254   model (Moore, 1993; Stahl, et al., 2008). The dominant phase of precipitation (snow vs. rain) is

255   determined by a temperature threshold (T_{th}). The snow melt M_s and glacier melt M_g is calculated

256   as:

257      $M_s = DDF_s \times (T - T_s)$ (8)

258      $M_g = DDF_g \times (T - T_g)$ (9)

259   with DDF_s (T_s) and DDF_g (T_g) applied separately for snow and glacier modules, respectively. To

260   account for the higher melting rate of glacier than snow owing to the low albedo (Konz & Seibert,

261   2010; Kinouchi, et al., 2013), we introduced a parameter r > 1 to constrain DDF_g to be larger than

262   DDF_s (i.e. $DDF_g = r \times DDF_s$). For the rain that falls on the glacierized area, the glacier parameter

263   K_g determines the portion of rain becoming surface runoff as a multiplier for the rainfall. The

264   remaining rainfall is assumed to be accumulated to the glacier store.

265   The within-grid routing process for direct runoff is represented by an instantaneous unit

266   hydrograph (IUH) (Nash, 1957), in which a catchment is depicted as a series of N reservoirs each

267   having a linear relationship between storage and outflow with the storage coefficient of K_q.

268   Mathematically, the IUH is expressed by a gamma probability distribution:

269    $u(t) = \frac{K_q}{\Gamma(N)}(K_q t)^{N-1} \exp(-K_q t)$ (10)

270   where, Γ is the gamma function. The within-grid groundwater routing process is simplified as a

271   lumped linear reservoir with the storage recession coefficient of K_s.

272   The transport of water in the channel system is described using the diffusive wave

273   approximation of the Saint-Venant equation (Lohmann, et al., 1998):

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

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

276   Similar to most other hydrological models (Efstratisdis et al., 2008), HYMOD_DS is not

277   designed to model water abstractions for agricultural lands and dam operations within the basin.

278   According to World Bank (2010), water demand for agricultural use is about 2,000 MCM (million

279   cubic meters), or about 8.3% of the total annual flow. The Naglu dam (Figure 1) upstream of the

280   Daronta streamflow gage forms the largest and most important reservoir in the basin, with an active

281   storage of 379 MCM. In our hydrologic modelling process, the water consumed by irrigated

282   croplands is implicitly accounted for by the evapotranspiration module. We note that the degree

283   of irrigation impact during the time frame used for calibration (1960-1981) is likely much smaller

284   than the current level. We also expect that using monthly data for calibration somewhat reduces

285   the bias from human interference, particularly the daily operations of Naglu dam. Nevertheless,

14

286   the calibration results for the gage below this dam (Daronta), and to a lesser extent the basin outlet

287   (Dakah), should be approached with caution. Given that a majority of the gages examined in this

288   study are on an underdeveloped branch of the Kabul River, issues of human interference on

289   calibration are somewhat mitigated.

290

4. Methods

292   The purpose of this study is to explore the implications of different calibration strategies

293   and choices for a computationally expensive distributed hydrologic model. A variety of calibration

294   experiments are conducted, with the results from preceding experiments informing choices made

295   for subsequent ones. All calibration approaches are tested in terms of their ability to predict flows

296   at interior site gages that were left out of the calibration process. In all cases, the Genetic Algorithm

297   (GA) introduced by Wang (1991) is used as an optimization method for model parameter

298   calibration, and the objective function is based simply on the Nash Sutcliffe efficiency (NSE)

299   (Nash & Sutcliff, 1970), which is by far the most utilized performance metric in hydrological

300   model applications (Biondi et al., 2012). A multisite average of the NSE is used when evaluating

301   performance across multiple sites. We fully recognize that the use of one objective, such as the

302   NSE, is inferior compared to multi-objective approaches that can identify Pareto optimal solutions

303   that provide good model performance across different components of the flow regime (Madsen,

304   2003; Efstratiadis & Koutsoyiannis, 2010; Li, et al., 2010; Kumar, et al., 2013). However, in this

305   particular study daily hydrologic model simulations can only be compared against available

306   monthly streamflow records, reducing the number of viable objectives against which to calibration.

307   That is, statistics representing peak flows, extreme low flows, and other daily flow regime

308   characteristics often used in multi-objective optimization approaches are unavailable. We believe

309   that the use of a monthly NSE value as a single objective, while coarse, does not inhibit our ability

310   to provide insight into the research questions posed. In addition to the NSE, the Kling-Gupta

311   efficiency (KGE) (Gupta et al., 2009) is adopted as an alternative model performance metric,

312   which equally weights model mean bias, variance bias, and correlation with observations.

313   In this study, three levels of parameter complexity are considered: lumped, semi-

314   distributed, and fully distributed formulations (Figure 4). The different levels of parameter

315   complexity are defined according to the spatial distribution of unique hydrologic model

316   parameters. In the lumped formulation a single parameter set is applied to the entire basin. In the

317   semi-distributed formulation, a unique parameter set is assigned to each sub-basin, defined based

318   on the location of available streamflow gaging sites. The fully distributed parameter structure

319   follows the spatial discretization of climate input grids, allowing a unique parameter set for each

320   grid cell. No matter the parameterization scheme, the model structure follows the climate input

321   grids, i.e. the hydrological water cycle within each grid cell is modelled separately. We note that

322   a lumped model structure (i.e., no gridded or sub-unit structure) has often been considered as a

323   baseline model formulation in the assessment of distributed modelling frameworks (e.g., see Smith

324   et al., 2013). However, the focus of our study is on ungaged interior site streamflow estimation,

325   making this formation somewhat inappropriate. Further, preliminary tests comparing streamflow

326   simulations at the basin outlet (Dakah) between a gridded and basin-averaged structure, both with

327   a lumped parameter formulation, support the use of the distributed grid structure (Figure S3).

328   The parameter complexity will vary depending on the calibration experiment being

329   conducted, but for each experiment regardless of the parameterization, the optimization is

330   implemented 50 times using the GA algorithm to explore calibration uncertainty. The considerably

331   high computational cost required to perform a large number of calibrations is managed using the

332   parallel computing power provided by the Massachusetts Green High Performance Computing

333   Center (MGHPCC), from which several thousands of processors are available.

334   In the first modeling experiment, we explore two calibration strategies for using multisite

335   streamflow data, a stepwise and pooled approach. In the stepwise calibration, parameters are

336   calibrated for upstream gaged sub-catchments and subsequently fixed during calibration of

337   downstream points, while for the pooled approach, parameters are calibrated for multiple sub-

338   catchments simultaneously. Both approaches are assessed for the semi-distributed formulation.

339   The better of the two methods is identified for use in the second experiment, where the effects of

340   increased parameter complexity are tested in terms of streamflow prediction accuracy and

341   uncertainty. In the third experiment, we consider the situation where there is only data at the basin

342   outlet for calibration. Here, the model is calibrated against the outlet gage under all levels of

343   parameter complexity and is compared against the best combination of calibration strategy (step-

344   wise or pooled) and parameter complexity (lumped, semi-distributed, or fully distributed)

345   identified in the previous experiments. Finally, a subset of the calibration approaches deemed

346   worthy of further investigation are compared in terms of their projections of future streamflow

347   under climate change to highlight how model calibration differences can alter the results of a

348   climate change assessment for water resources applications. These experiments are described in

349   further detail below.

350

351   **4.1. Multisite Calibration: Stepwise and Pooled Approaches**

352   In the first experiment, the semi-distributed parameterization concept is compared under

353   alternative multisite calibration strategies, the stepwise and pooled calibration approaches. To

354   conduct the stepwise calibration, a nested class of sub-basins is defined corresponding to multiple

355   gaging stations. In the first step of the stepwise calibration, the optimization process is carried out

356   with nested sub-basins at the lowest level (i.e., the most upstream sites). Once parameters of nested

357   sub-basins are determined, the parameters are fixed, and the calibration procedure proceeds with

358   nested basins at upper levels until parameters for the entire basin are determined. In this particular

359   application to the Kabul River basin, 5 gaged sub-basins were selected and the stepwise calibration

360   procedure for those sub-basins followed this direction: Chitral → Gawardesh → Chaghasarai →

361   Daronta → Dakah (Figure 5). The stepwise calibration approach involves a number of GA

362   implementations corresponding to the number of gaging sites. The GA optimization was carried

363   out a total of 250 times in this application, with 50 optimization runs containing GA

364   implementations for 5 sub-basin regions.

365   The pooled calibration strategy involves calibrating all parameters of the model domain

366   simultaneously against multiple streamflow gages within the watershed. This approach aims at

367   looking for suitable parameters that are able to produce satisfactory model results at all gaging

368   stations in a single implementation of GA optimization. That is, the GA searches the entire

369   parameter space at once to maximize the average NSE across all sites. This operational feature

370   reduces the processing time spent on the GA implementation compared to the stepwise calibration

371   strategy. To identify the better of the two multisite calibration approaches, the comparison focused

372   on their ability to predict streamflow and calibration uncertainties at two interior site gages (Kama

373   and Asmar) that were assumed to be ungaged (Figure 5), as well as for validation data at the basin

374   outlet.

375   It is important to note that the evaluation of these multisite calibration strategies is

376   somewhat weakened because of the lack of overlapping data periods among most of the stations

377   (Figure 2). This drawback prevents the calibration methods from accounting for simultaneous

378   information from different tributaries, which, if available, would better enable the calibration

379   methods to account for heterogeneity of hydrological processes across the sub-basins.

380

381   **4.2. Increased Parameter Complexity**

382   In the second experiment, the better of the two approaches (step-wise or pooled) identified

383   in the first experiment is further tested with respect to the three different levels of parameter

384   complexity. In addition to the semi-distributed parameter formulation considered in the first

385   experiment, lumped and fully-distributed parameter formulations are calibrated for the selected

386   approach to investigate the gain or loss arising from different levels of parameter complexity. Since

387   the hydrologic model HYMOD employed in this study involves 15 parameters, the lumped version

388   of the HYMOD_DS contains a single, 15-member parameter set applied to all model grid cells.

389   The semi-distributed conceptualization of HYMOD_DS contains a single parameter set for each

390   sub-basin, totaling 75 parameters. In the distributed parameterization the number of parameters

391   increases dramatically. With 160 0.25° grid cells, the number of parameters requiring calibration

392   reaches 2,400. As the number of parameters increase across the parameterization schemes,

393   calibration becomes increasingly computationally expensive. The number of model runs used in

394   the GA optimization algorithm for the lumped, semi-distributed, and distributed parameterization

395   schemes are 15,000 (150 populations $\times$ 100 generations), 75,000 (750 $\times$ 100), and 480,000 (2400

396    $\times$ 200), respectively. These population/generation sizes were supported using convergence tests

397   for each calibration. Again, 50 separate GA optimizations were used to explore calibration

398   uncertainties for each parameterization scheme. To give a sense of the computational burden of

399   this experiment, we note that 50 trials of the HYMOD_DS calibration under the distributed

400   conceptualization required 1,000 processors over 7 days on the MGHPCC system.

19

401

4.3. Basin Outlet Calibration

403   The third experiment considers the situation where there is only gaged data at the basin

404   outlet (Dakah) for calibration, a common situation when calibrating hydrologic models in data-

405   scarce river basins. Here, we evaluate the potential of the basin outlet calibration to estimate

406   interior watershed flows in terms of both accuracy and precision at all gaging stations. All levels

407   of parameter complexity are considered for this calibration. The main purpose of this experiment

408   is to compare the veracity of a distributed hydrologic model calibrated only using basin outlet data

409   with results from multisite calibrations to better understand the degradation in model performance

410   under data scarcity. Other than the use of an NSE objective only at the basin outlet, all other GA

411   settings for each level of parameter complexity are same as the settings used in the second

412   experiment.

413

4.4. Climate Change Projections of Streamflow

415   The fourth experiment investigates how the choice of calibration approach can alter the

416   projections of future streamflow under climate change. To explore this question, streamflow

417   simulations for the 2050s, defined as the 30-year period spanning from 2036 to 2065, are carried

418   out using climate projections from the CMIP5 (Talyor, et al., 2012). A total of 36 different climate

419   models run under two future conditions of radiative forcing (RCP 4.5 and 8.5) are used.

420   Streamflow projections are developed for the basin outlet (Dakah) and two interior gages left out

421   of the calibration (Kama and Asmar). By using 36 different General Circulation Models (GCMs)

422   and 50 optimization trials for each calibration scheme, this analysis compares the uncertainty in

423   future streamflow projections originating from uncertainty in different hydrologic model

424   parameterization schemes and under alternative future climates.

425   Streamflow projections are considered under all three parameterization schemes (lumped,

426   semi-distributed, and fully distributed) for both the basin outlet model and the best multi-site

427   calibration approach (step-wise or pooled). Multiple streamflow characteristics are evaluated,

428   including monthly streamflow, wet (April-September) and dry (October-March) season flows, and

429   daily peak flow response. The differences and uncertainty in these metrics across calibration

430   approaches will highlight the importance of calibration strategy for evaluating future water

431   availability and flood risk.

432

5. Results

434   For the remaining part of the paper, we introduce the following shorthand: Lump, Semi,

435   and Dist indicate the lumped, semi-distributed, and fully distributed parameterization schemes,

436   and Outlet, Stepwise, and Pooled correspond to basin outlet, stepwise, and pooled calibrations.

437   The comparison between different calibration strategies is based on the model performance

438   evaluated with the NSE, as well as an alternative metric, the KGE.

439

5.1. Pooled Calibration vs. Stepwise Calibration

441   This section reports the results from the first experiment comparing the stepwise and

442   pooled calibration approaches for the semi-distributed model parameterization. Figure 6 shows the

443   comparison between the Semi-Stepwise and Semi-Pooled with boxplots representing the 50 trials

444   of calibration. Under the stepwise calibration the results for 4 sub-basins (Chitral, Gawardesh,

445   Chaghasarai, and Daronta) are optimal because there is no interaction between those sub-basins.

446   However, the calibrated parameter sets of each sub-basin act as constraints in the last step of the

447   Semi-Stepwise resulting in the degradation of model skill at the basin outlet (Dakah) and two left-

448   out gages (Asmar and Kama). This becomes apparent when comparing the Semi-Stepwise to the

449   Semi-Pooled results. The model skill under the Semi-Pooled is similar to that from the Semi-

450   Stepwise with respect to the 4 upstream sub-basins, but it outperforms at the verification gages.

451   This is particularly true for the Asmar gage, which exhibits a downward bias and substantial

452   variability in performance under the Semi-Stepwise. The Semi-Pooled results suggest that small

453   sacrifices of model performance at certain sites can improve and stabilize basin-wide performance.

454   Expected values of KGE from 50 calibrations are also provided (values in parenthesis in the bottom

455   of Figure 6) and this performance metric also leads to the same conclusion. Therefore, the Semi-

456   Pooled was selected as the better multisite calibration strategy and is considered for further

457   analyses in the following sections.

458

5.2. Pooled Calibration with Alternative Parameterizations

460   Here we examine results for the three levels of parameter complexity applied to the pooled

461   calibration approach. Figure 7 shows the comparison of the pooled calibrations. Unsurprisingly,

462   streamflow predictions from the Lump-Pooled have the lowest accuracy and largest uncertainty at

463   the calibration sites, particularly for the Chaghasarai and Daronta sites. This demonstrates the well-

464   known difficulty in representing flow characteristics of a spatially variable system with a

465   homogenous parameter set (Beven, 2012). The pooled calibration substantially improves with

466   increasing parameter complexity at the calibration sites. Both the Semi-Pooled and Dist-Pooled

467   produce NSE values above 0.8 for all calibration sites, with the Dist-Pooled showing somewhat

468   higher performance, undoubtedly from its greater freedom to over-fit to the calibration data.

469   However, the advantage of the Dist-Pooled with respect to streamflow predictions at validation

470   sites becomes less clear. Only the Dist-Pooled at Kama shows marginally better predictions, while

471   the results are ambiguous at Dakah and Asmar. Overall, this likely suggests that the fully

472   distributed conceptualization leads to over-fitting of the model as compared to the Semi-Dist

473   conceptualization. We reached the same conclusion when examining the KGE values, which rise

474   with greater parameter complexity at calibration sites but no longer follow this pattern strictly at

475   validation sites.

476   Interestingly, the Lump-Pooled performs well at the verification sites despite its poor

477   performance at calibration sites. The Lump-Pooled does not show significant degradation in skill

478   at Kama compared to the more complex parameterizations, and the flow prediction at Asmar

479   actually exhibits the best performance of all three model variants. A partial reason for this

480   unexpected result arises from different overlapping periods in the calibration and validation data

481   (see Figure 2). The periods used for the calibration for Chitral (1978-1981) and Gawardesh (1975-

482   1978) have no overlapping periods with the one for Asmar (1966-1971), which encompasses those

483   two sub-basins. Instead, the validation at Asmar is mostly affected by the calibration to Dakah

484   because of the overlapping 4 years (1968-1971) between those two sites. This explains the reason

485   why the Lump-Pooled shows high skill at Asmar despite the low skill at its sub-basins. However,

486   the low model skill at Chaghasarai from the Lump-Pooled propagates to the validation result at

487   Kama, as these two sites have a relatively long overlapping period (8 years from 1967-1974).

488

5.3. Limitations of the Basin Outlet Calibration

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

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

512 0.84 (0.71) and 0.90 (0.88), respectively) and comparable performance at many of the other sites.

513 One exception is Gawardesh, where the Lump-Outlet outperforms the other model variants,

514 although the reason for this is not immediately clear. Overall, the results indicate that any

515 calibration based on basin outlet data should be used with substantial caution when predicting

516 flows at interior basin sites.

517   After reviewing all of the calibration experiments, it becomes clear that the Semi-Pooled

518 and Dist-Pooled calibrations provide more robust performance compared to the basin outlet

519 calibrations due to their improved representation of internal hydrologic processes across the basin.

520 To further compare these calibration strategies against one another, we evaluate the variability in

521 optimal parameters resulting from the 50 trials of the GA algorithm. Figure 9 shows the coefficient

522 of variation (CV) of Cmax (a parameter for the soil moisture account module) over the basin from

523 all combinations of calibration approaches (the outlet and pooled) and 3 parameterization schemes.

524 A clear pattern of increasing variability (higher uncertainty in Cmax) emerges as parameter

525 complexity increases for both the outlet and pooled calibration strategies. That is, the semi- and

526 fully-distributed parameterizations lead to significantly variable parameter sets that produce

527 similar representations of the observed basin response. Figure 9 also suggests that the equifinality

528 can be alleviated to an extent by pooling data across sites. The pooled calibration approaches

529 consistently show lower variability in Cmax compared to the outlet calibration at the same level

530 of parameter complexity. These results are relatively consistent across the remaining 14

531 HYMOD_DS parameters. The implications of parameter stability on streamflow projections under

532 climate change is addressed in the next section.

533

5.4. Climate Change Projections of Streamflow with Uncertainty

534

535   Here we explore how projections of future water availability and flood risk under climate

536   change are influenced by the choice of calibration approach. For the Kabul River basin, the CMIP5

537   GCM projections of monthly total precipitation and mean temperature are shown in Figure S5.

538   According to the CMIP5 ensemble, precipitation projections show no clear trend; the average

539   precipitation change in monthly total precipitation fluctuates between -10mm and 10mm. On the

540   other hand, temperature clearly shows an upward trend for both radiative forcing scenarios. The

541   average changes in annual temperature are +2.2°C and +2.8°C for RCP4.5 and RCP8.5, which,

542   using the Hamon method, correspond to an increase in annual PET by approximately 100mm and

543   150mm, respectively.

544   We first examine average monthly streamflow estimates across four calibration strategies:

545   the Semi-Pooled and Dist-Pooled (most promising calibration strategies), as well as the Lump-

546   Outlet (as a baseline) and Dist-Outlet (the best outlet calibration strategy). Figure 10 shows the

547   monthly streamflow estimates for the historical period with the whisker bars indicating the

548   uncertainty range across the 50 calibration trials. The monthly streamflow predictions are also

549   provided for the 2050s under the RCP 4.5 and 8.5 scenarios. For the future scenarios, the whisker

550   bars are derived by averaging over the 36 different climate projections for each of the 50 trials.

551   For the historical time period, all calibration schemes match the observed monthly streamflow at

552   Dakah well, but monthly streamflow is underestimated in most of months at Kama and Asmar

553   under the basin outlet calibrations, particularly by the Lump-Outlet. The historical monthly

554   streamflow estimates from the outlet calibration strategies also tends to be highly uncertain for the

555   months of June, July, August, and September, especially compared to the SemiPool and DistPool.

26

556   Under future climate projections for the 2050s, the four calibration strategies show similar

557   changes in monthly streamflow at Dakah, but the magnitudes of change are somewhat different.

558   All calibration strategies suggest reduction in streamflow for June, July, and August under both

559   RCP4.5 and RCP8.5 scenarios. Also, the peak monthly flow, which occurred in June or July in the

560   historical period, is shifted to May at Dakah. However, the Lump-Outlet predicts less reduction of

561   flow in June and July and a greater reduction in August and September as compared to the other

562   three calibrations. Considering that all calibration schemes had similar levels of good performance

563   at this site for both calibration and validation periods, it is notable that they project future

564   streamflow somewhat differently.

565   Future monthly streamflow predictions at Kama and Asmar vary widely between the four

566   calibration schemes, mostly an artifact of their historic differences (Figure 10). Streamflow

567   projections under the outlet calibration strategies tend to show large uncertainties at these two sites,

568   particularly the Lump-Outlet calibration. For three months, July through September, the outlet

569   calibration and pooled calibration strategies provide substantially different insights about future

570   water availability at Kama and Asmar. The outlet calibrations suggest less water with large

571   uncertainties for those months as compared to the pooled calibrations. At Kama, the pooled

572   calibrations suggest significant changes in the pattern of peak monthly flow timing under both

573   RCP scenarios; instead of having a clear peak in July, streamflow from May to August show

574   similar amounts of water.

575   To further understand the sources of uncertainty in future water availability, we evaluate

576   the separate and joint influence of uncertainties in parameter estimation and future climate on

577   seasonal streamflow projections across all calibration schemes. Figure 11 represents the

578   uncertainty of wet and dry seasonal streamflow at Dakah from three sources: 1) calibration

579   uncertainty across the 50 trials, with future climate uncertainty averaged out for each trial, 2) future

580   climate uncertainty across the 36 projections, with calibration uncertainty averaged out across the

581   50 trials, and 3) the combined uncertainty across all 1800 (50×36) simulations. The results suggest

582   somewhat surprisingly that uncertainty reduction can be expected as parameter complexity

583   increases, and less surprisingly, by applying pooled calibration approaches. Another clear point is

584   that the uncertainty resulting from different climate change scenarios substantially outweighs that

585   from calibration uncertainty.

586   Up to this point, there has been little difference between the Semi-Pooled and Dist-Pooled

587   model variants. These two versions were further analyzed with respect to extreme streamflow to

588   see if distinguishing characteristics emerge. It has been demonstrated that clear gains in predicting

589   peak flows from distributed models are noticeable (Reed et al., 2004) and spatial variability in

590   model parameters significantly influence the runoff behavior (Brath and Montanari, 2000; Pokhrel

591   and Gupta, 2011). The spatial variability of optimal parameters derived from the Semi-Pooled and

592   Dist-Pooled is shown in Figure S6, with larger variability across all parameters for the Dist-Pooled

593   than for the Semi-Pooled. To understand the effects of spatial variability and calibration

594   uncertainty of parameters on extreme event estimation, the 100-year daily flood event was

595   calculated under the Semi-Pooled and Dist-Pooled for each of the 50 historic simulations and 1800

596   future simulations across both RCP scenarios. Although the inter-model comparison is intended to

597   be a useful addition that provides a distinction between the parameterization schemes in the pooled

598   calibration approach, results from this analysis should be viewed in the context of a theoretical

599   calibration exercise, not for decision-making purposes, because no observed daily streamflow is

600   available against which to compare the estimated 100-year daily flood events. Projections of the

601   100-year daily flood, estimated using a Log-Pearson type III distribution fit to annual peaks of 30

28

602   years, differ somewhat between the Semi-Pooled and Dist-Pooled (Figure 12). At 3 validation

603   sites, extreme floods are consistently larger under the Semi-Pooled than the Dist-Pooled, and the

604   mean difference in the 100-year daily flood estimate between the two calibration approaches grows

605   between the historic runs and the RCP 4.5 and 8.5 scenarios. This suggests that the flood-

606   generation process is fundamentally different between the two parameterizations, with the Semi-

607   Pooled formalization magnifying the effect of climate change on extremes. Furthermore, there is

608   substantially more uncertainty in the 100-year daily flood estimate under the Semi-Pooled. Figure

609   12 shows the combined uncertainty across both climate projections and calibrations, but this

610   uncertainty is broken down further in Figure 13. Similar to Figure 11, 3 sources of uncertainty are

611   evaluated for the 100-year daily flood, including calibration uncertainty alone, climate projection

612   uncertainty alone, and their combined effect. For both the Semi-Pooled and Dist-Pooled,

613   calibration uncertainty has a smaller influence than projection uncertainties, and for all sites, the

614   Dist-Pooled has a smaller uncertainty range than the Semi-Pooled, even for calibration uncertainty

615   alone. This was a truly surprising result, given the parametric freedom in the Dist-Pooled model

616   and the fact that no daily data was ever used in the calibration of either model. It appears that a

617   lack of model parsimony does not necessarily lead to greater uncertainty in model simulations

618   under different climate conditions, somewhat counter to what would be expected of over-fit

619   models. One possible reason for this result would be if increased parametric freedom somehow

620   offset the effects of structural deficiencies in the model. However, further research is needed to

621   investigate this issue.

622

6. Discussion and Conclusion

623

624   In this study we examined a variety of calibration experiments to better understand the

625   benefits and costs associated with different calibration choices for a complex, distributed

626   hydrologic model in a data-scarce region. The goal of these experiments was to provide insight

627   regarding the use of multisite data in calibration, the effects of parameter complexity, and the

628   challenges of using limited data for distributed model calibration, all in the context of projecting

629   future streamflow under climate change.

630   This study tested two multi-site calibration strategies, the stepwise and pooled approaches,

631   finding that the pooled approach using all data simultaneously provides improved calibration

632   results. This suggests that small sacrifices of model performance at certain sites can improve and

633   stabilize basin-wide performance. The pooled calibration substantially improves with increasing

634   parameter complexity at the calibration sites, but similar streamflow predictions at the validation

635   sites between the semi-distributed and distributed pooled calibrations were found, suggesting over-

636   fitting of the model from the fully distributed conceptualization. It is worth noting that for the

637   transformation of rainfall to runoff, up to five or six parameters can be identified on the basis of a

638   single hydrograph (Wagner et al., 2001). Under this premise, the number of the HYMOD_DS

639   parameters being calibrated in the semi-distributed approach remains realistic, but the fully

640   distributed parameterization scheme likely causes poor identifiability of the parameters. Thus,

641   pursuing a parsimonious configuration (e.g. optimization for a small portion of the parameters)

642   with an effort to increase the amount of information (e.g. multivariable/multisite) is critical in the

643   calibration of watershed system models (Gupta et al., 1998; Efstratiadis et al., 2008). We also note

644   the important role of experienced hydrologists in designing a parsimonious hydrologic calibration

645   (e.g. Boyle et al., 2000). In this study, the feasible ranges of the HYMOD_DS parameters were

646   kept wide (as is often done in automatic hydrologic calibrations) without consideration of the

647   physical properties of the basin; the judgment of local hydrologic experts could help reduce the

648   feasible ranges used during the calibration and thus contribute to a reduction of calibration

649   uncertainty.

650   Calibration only based on data at the basin outlet is all too common in hydrologic model

651   applications and is sometimes considered comparable to multisite calibrations even for predictions

652   at interior gauges (Lerat et al., 2012). In contrast, others have reported improvements in interior

653   flow predictions by using internal flow measurements (Anderson et al., 2001; Wang et al., 2012;

654   Boscarello et al., 2013). This is in agreement with the finding from this study, demonstrating the

655   superiority of the pooled calibration approach to the basin outlet calibration in terms of its ability

656   to represent interior hydrologic response correctly. This study shows the danger in relying on an

657   outlet calibration for interior flow prediction.

658   It was shown that caution is needed when using an outlet calibration approach for

659   streamflow predictions under future climate conditions. This study showed that the basin outlet

660   calibration can lead to projections of mid-21st century streamflow that deviate substantially from

661   projections under multisite calibration strategies. From the test of implications of the pooled

662   calibration in the context of climate change, it was found that applying the pooled calibration with

663   semi-distributed and distributed parameter formulations showed clear gains in reducing

664   uncertainties in predictions of monthly and seasonal water availability as compared to the basin

665   outlet calibrations. Surprisingly, increased parameter complexity in the calibration strategies did

666   not increase the uncertainty in streamflow projections, even though parameter equifinality did

667   emerge. The results suggest that increased (excessive) parameter complexity does not always lead

668   to increased uncertainty if structural uncertainties in the model are present.

669　　　　The semi-distributed pooled and distributed pooled calibrations are very similar for

670　　monthly streamflow projections, yet differ in their projections of extreme flows in part due to their

671　　differences in the spatial variability of optimal parameters, with the distributed pooled calibration

672　　showing less uncertainty for 100-year daily flood events. We evaluated the separate and joint

673　　influence of uncertainties in parameter estimation and future climate on projections of seasonal

674　　streamflow and 100-year daily flood across calibration schemes and found that the uncertainty

675　　resulting from variations in projected climate between the CMIP5 GCMs substantially outweighs

676　　the calibration uncertainty. These results agree with other studies showing the dominance of GCM

677　　uncertainty in future hydrologic projections (Chen et al., 2011; Exbrayat et al., 2014). While the

678　　GCM-based simulations still have widespread use in assessing the impacts of climate change on

679　　water resources availability, the bounds of uncertainty resulting from an ensemble of GCMs cannot

680　　be well-defined because of the low credibility with which GCMs are able to produce timeseries of

681　　future climate (Koutsoyiannis et al., 2008). This issue hinders a straightforward appraisal of future

682　　water availability under climate change and has motivated other efforts; e.g. performance-based

683　　selection of GCMs (Perez et al., 2014).

684　　　　In addition to the uncertainties surrounding model parameters and future climate explored

685　　in this study, there is also significant uncertainty in streamflow projections stemming from

686　　structural differences between applied hydrologic models, which can be especially pertinent where

687　　robust calibration is hampered by the scarcity of data (Exbrayat et al., 2014). Further, the residual

688　　error variance of hydrologic model simulations would increase the effects of hydrologic model

689　　uncertainty as compared to that of the climate projections (Steinschneider et al., 2014). These

690　　issues need to be addressed in future work for exploring a comprehensive uncertainty assessment

691　　of climate change risk for poorly monitored hydrologic systems.

692   Successful automatic calibration algorithms for hydrologic models are based primarily on

693   global optimization algorithms that are computationally expensive and require a large number of

694   function evaluations (Kuzmin et al., 2008). Although the speed and capacity of computers have

695   increased multi-fold in the past several decades, the time consumed by running hydrological

696   models (especially complex, physically based, distributed hydrological models) is still a concern

697   for hydrology practitioners. A single trial of parameter optimization of HYMOD_DS associated

698   with 100,000 runs can take 28 days on a single processor (Figure S7). Accordingly, the use of high

699   performance computing power was essential in this study to better understand the implications of

700   different calibration choices and their associated uncertainty for streamflow projections. Enhanced

701   data with high spatial and temporal resolution are increasingly available from remote sensing and

702   satellite products. In the future, remote sensing and satellite information can be integrated into

703   calibration approaches to develop more robust estimates of spatially distributed parameter values,

704   enabling internal consistency of distributed hydrological modeling. Significant progress has been

705   made toward this end (Tang et al., 2009; Khan et al., 2011; Thirel et al., 2013). Future work will

706   consider using high performance computing power (e.g. Laloy and Vrugt, 2012; Zhang et al.,

707   2013) to understand how such information can enhance the hydrologic simulation at ungaged sites

708   and reduce the calibration uncertainty of distributed hydrologic models in data-scarce regions.

709

Acknowledgements

719 **References**

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

722 Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed
723 hydrologic model for streamflow estimation along a river system, J HYDROL, 298, 112-135,
724 2004.

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

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

729 Beven, J. K.: Rainfall-Runoff Modelling: The Primer, 2nd Edition, Wiley-Blackwell, Chichester,
730 2012.

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

733 Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic
734 modelling of complex environmental systems using the GLUE methodology, J HYDROL, 249,
735 11-29, 2001.

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

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

741 Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic
742 models: Combining the stregths of manual and automatic methods, Water Resources research,
743 36(12), 3663-3674, 2000.

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

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

748   Brath, A., Montanari, A., and Toth, E.:. Analysis of the effects of different scenarios of historical
749   data availability on the calibration of a spatially-distributed hydrological model, J HYDROL, 291,
750   232-253, 2004.

751   Breuer, L., Huisman J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H. G.,
752   Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P.,
753   Lindström, G., Seibert, J., Sivapalan, M., and Viney, N. R.: Assessing the impact of land use
754   change on hydrology by ensemble modeling (LUChEM). I: Model intercomparison with current
755   land use, Advances in Water Resources, 32, 129-146, 2009Cao, W., Bowden, W. B., Davie, T.,
756   and Fenemor, A.: Multi-variable and multi-site calibration and validation of SWAT in a large
757   mountainous catchment with high spatial variability, HYDROL PROCESS, 20, 1057-1073, 2006.

758   Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological
759   impacts of climate change for a Canadian watershed, Water Resources Research, 47, W12509,
760   2011.

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

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

765   Eckhardt, K., Fohrer, N., and Frede, H. G.: Automatic model calibration, HYDROL PROCESS,
766   19, 651-658, 2005.

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

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

772 Exbrayat, J. F., Buytaert, W., Timbe, E., Windhorst, D., and Breuer, L.: Addressing sources of

773 uncertainty in runoff projections for a data scarce catchment in the Ecuadorian Andes, Climatic

774 Change, 125, 221-235, 2014.

775 Flugel, W. A.: Delineating Hydrological Response Units (HRU's) by GIS analysis for regional

776 hydrological modelling using PRMS/MMS in the drainage basin of the River Brol, Germany,

777 Hydrol. Processes, 9, 423-436, 1995.

778 Forsythe, W. C., Rykiel Jr., E. J., Stahl, R. S., Wu, H., Schoolfield, R. M.: A model comparison

779 for daylength as a function of latitude and day of year, Ecological Modelling, 80, 87-95, 1995.

780 Frances, F., Velez, J. I., and Velez, J. J.: Split-parameter structure for the automatic calibration of

781 distributed hydrological models, J HYDROL, 332(1-2), 226-240, 2007.

782 Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time

783 consistent model parameters: sub-period calibration, HYDROL EARTH SYST SC, 17, 149-161,

784 2013.

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

786 Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Towards improved calibration of hydrologic

787 models: Multiple and noncommensurable measures of information, Water Resources Research,

788 34, 751-763, 1998.

789 Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared

790 error and NSE performance criteria: Implications for improving hydrological modelling, J

791 HYDROL, 377, 80-91, 2009.

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

793 1961.

794 Hewitt, K., Wake, C. P., Young, G. J., and David, C.: Hydrologicl investigations at Biafo glacier,

795 Karakoram Himalaya, Pakistan: An important source of water for the Indue River, ANN

796 GLACIOL, 13, 103-108, 1989.

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

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

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

805   Khakbaz, B., Imam, B., Hsu, K., and Sorooshian, S.: From lumped to distributed via semi-
806   distributed: Calibration strategies for semi-distributed hydrologic models, J HYDROL, 418-419,
807   61-77, 2012.

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

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

814   Kinouchi, T., Liu, T., Mendoza, J., and Asaoka, Y.: Modeling glacier melt and runoff in a high-
815   altitude headwater catchment in the Cordillera Real, Andes, HYDROL EARTH SYST SC, 10,
816   13093-13144, 2013.

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

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

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

823   Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model

824   parameterization on water fluxes at multiple scales and locations, WATER RESOUR RES, 49(1),

825   360-379, 2013.

826   Kuzmin, V., Seo D., and Koren V.: Fast and efficient optimization of hydrologic model parameters

827   using a priori estimates and stepwise line search, J HYDROL, 353, 109-128, 2008.

828   Laloy, E., and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using

829   multiple-try DREAM(ZS) and high-performance computing, WATER RESOUR RES, 48(1),

830   W01526, 2012.

831   Latham, J., Cumani, R., Rosati, I., and Bloise, M.: Global Land Cover SHARE (GLC-SHARE)

832   database   Beta-Release   Version   1.0,   available   at:

833   http://www.glcn.org/databases/lc_glcshare_en.jsp, last access: 2 January 2015, 2014.

834   Leavesley, G. H., Hay, L. E., Viger, R. J., and Markstrom, S. L.: Use of Priori Paramter-Estimation

835   Methods to Constrain Calibration of Distributed-Parameter Models, WATER SCI APPL, 6, 255-

836   266, 2003.

837   Legates, D. R., and Willmott, C. J.: Mean seasonal and spatial variability in gauge-corrected,

838   global precipitation, INT J CLIMATOL, 10(2), 111-127, 1990.

839   Lerat, J., Andreassian V., Perrin, C., Vaze, J., Perraud J. M., Ribstein, P., and Loumagne C.: Do

840   internal flow measurements improve the calibration of rainfall-runoff models?, WATER RESOUR

841   RES, 48, W02511, 2012.

842   Li, X., Weller, D. E., and Jordan, T. E.: Watershed model calibration using multi-objective

843   optimization and multi-site averaging, J HYDROL, 380(3-4), 277-288, 2010.

844   Lohmann, D., Raschke, R., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology: I.

845   Formulation of the VIC-2L model coupled to a routing model, HYDROLOG SCI J, 43(1), 131-

846   141, 1998.

847   Madsen, H.: Parameter estimation in distributed hydrologicl catchment modelling using automatic

848   calibration with multiple objectives, ADV WATER RESOUR, 26(2), 205-216, 2003.

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

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

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

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

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

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

861   Perez, J., Menendez, M., Mendez, F. J., and Losada, I. J.: Evaluating the performance of CMIP3
862   and CMIP5 global climate models over the north-east Atlantic region, Climate Dynamics, 43,
863   2663-2680, 2014.

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

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

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

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

875   Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. J., and DMIP Participants: Overall

876   distributed model intercomparison project results, J HYDROL, 298, 27-60, 2004.

877   Remesan, R., Bellerby, T., and Frostick, L.: Hydrological modelling using data from monthly

878   GCMs in a regional catchment, HYDROL PROCESS, 28(8), 3241-3263, 2013.

879   Safari, A., De Smedt, F., and Moreda, F.: WetSpa model application in the Distributed Model

880   Intercomparison Project (DMIP2), J HYDROL, 418-419, 77-89, 2012.

881   Shakir, A. S., Rehman, H., and Ehsan, S.: Climate change impact on river flows in Chitral

882   watershed, Pakistan Journal of Engineering and Applied Sciences, 7, 12-23, 2010.

883   Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N.,

884   Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project - Phase 2:

885   Motivation and design of the Oklahoma experiments, J HYDROL, 418-419, 3-16, 2012.

886   Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong,

887   S.: The distributed model intercomparison project (DMIP): motivation and experiment design, J

888   HYDROL, 298, 4-26, 2004.

889   Smith, M., Koren, V., Zhang, Z., Moreda, F., Cui, Z., Cosgrove, B., Mizukami, N., Kitzmiller, D.,

890   Ding, F., Reed, S., Anderson, E., Schaake, J., Zhang, Y., Andreassian, V., Perrin, C., Coron, L.,

891   Valery, A., Khakbaz, b., Sorooshian, S., Behrangi, A., Imam, B., Hsu, K. L., Todini, E., Coccia,

892   G., Mazzetti, C., Andres, E. O., Frances, F., Orozco, I., Hartman, R., Henkel, a., Fickenscher, P.,

893   and Staggs, S.: The distributed model intercomparison project - Phase 2: Experiment design and

894   summary results of the western basin experiments, J HYDROL, 507, 300-329, 2013.

895   Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., and Cannon, A. J.: Coupled modelling of

896   glacier and streamflow response to future climate scenarios, WATER RESOUR RES, 44(2),

897   W02422, 2008.

898   Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H.: Toward a statistical framework

899   to quantify the uncertainties of hydrologic response under climate change, WATER RESOUR

900   RES, 48(11), W11525, 2012.

901 Steinschneider, S., Wi, S., and Brown, C.: The integrated effects of climate and hydrologic
902 uncertainty on future flood risk assessments, Hydrological Processes, DOI: 10.1002/hyp.10409,
903 2014.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

947

948 **Tables**

949

950   Table 1 Streamflow gaging stations in the Kabul River basin.

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 (ºC)	Mean Annual Flow (mm)
USGS/ GRDC	Dakah	Kabul	1968/2	1980/7	67,370	2.9	2,883	418	7.7	282
USGS/ GRDC	Pul-i-Kama	Kunar	1967/1	1979/9	26,005	7.3	3,446	446	5.6	573
USGS	Asmar	Kunar	1960/3	1971/9	19,960	9.4	3,716	483	4.1	651
GRDC	Chitral	Kunar	1978/1	1981/12	11,396	14.4	4,126	518	2.1	698
USGS	Gawardesh	Landaisin	1975/5	1978/6	3,130	2.1	3,707	555	4.5	521
USGS/ GRDC	Chaghasarai	Pech	1960/2	1979/2	3,855	0.4	3,141	482	7.4	535
USGS/ GRDC	Daronta	Kabul	1959/10	1964/9	34,375	0.3	2,722	350	8.0	165

951

952

953 Table 2 HYMOD_DS parameters.

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

954

955
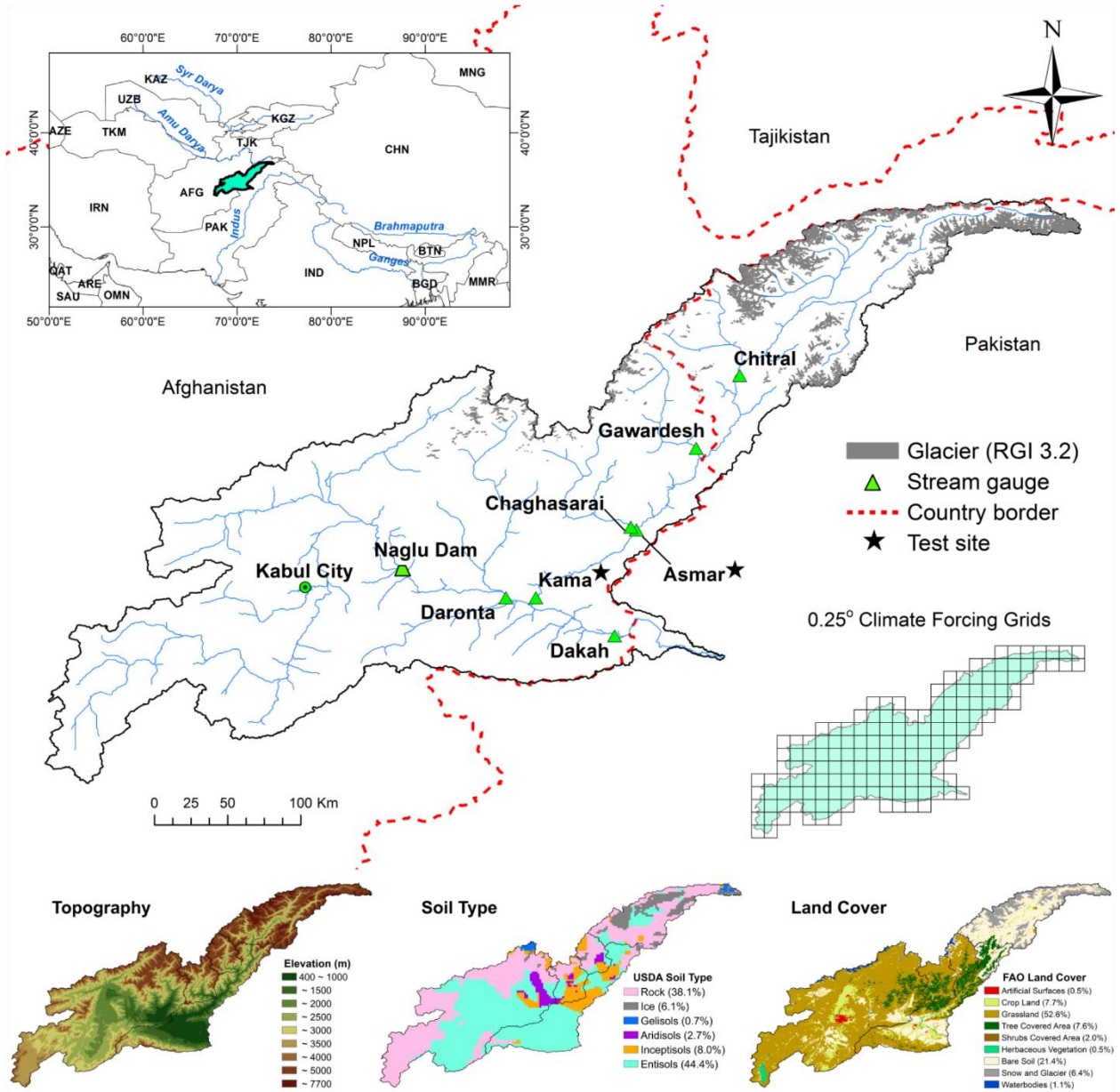
Figures

Figure 1. Kabul River basin.

961
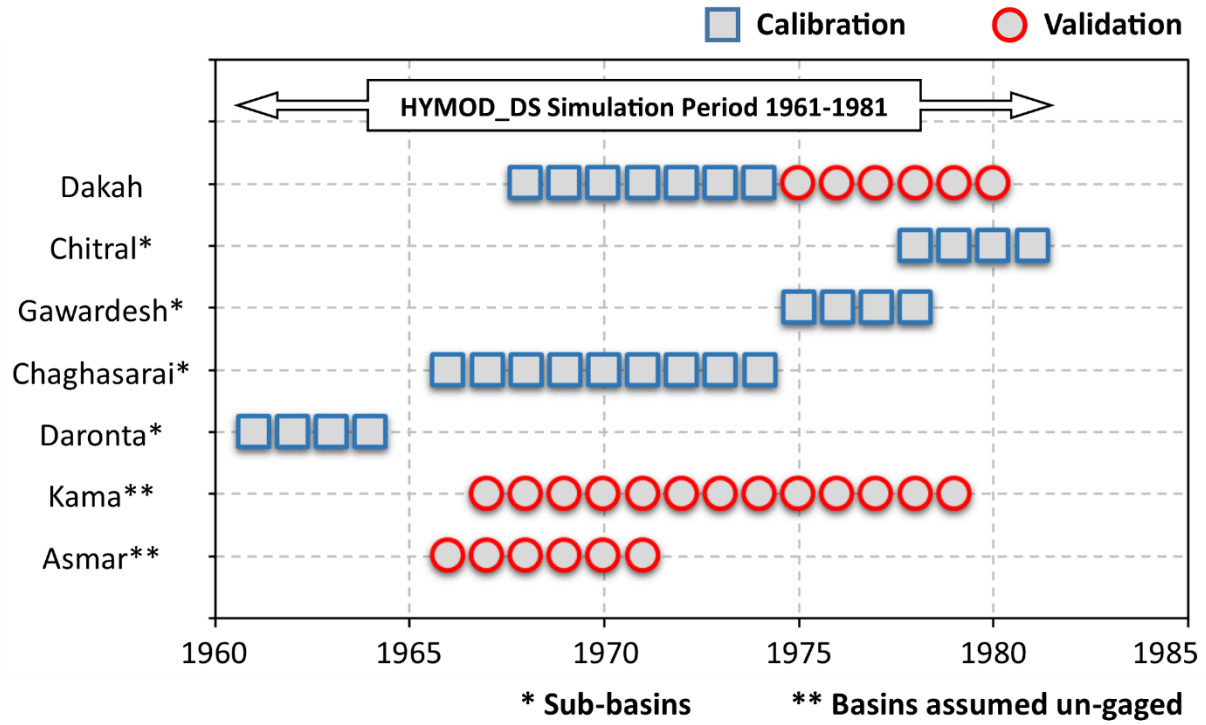
Figure 2. Streamflow data usage for the model calibration and validation.

963

964   **HRU Soil Moisture Accounting using HYMOD**   **Basin Channel Routing using Linearized Saint-Venant Eq**
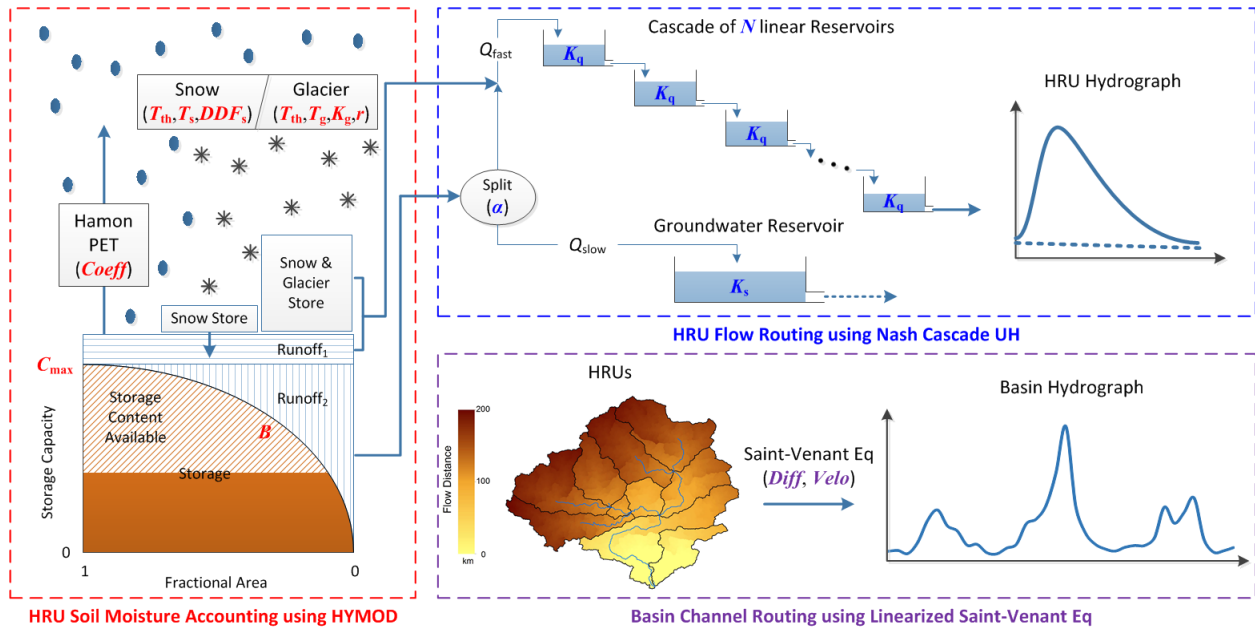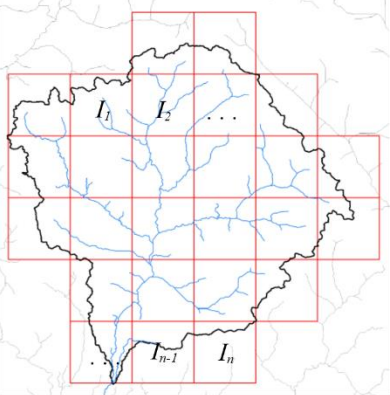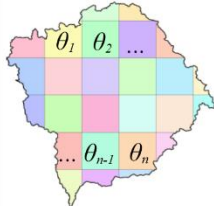
965   Figure 3. Distributed version of HYMOD model (HYMOD_DS).

966

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

967

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

968
969

970

a)

Chitral

Gawardesh

Chaghasarai

Daronta

Asmar station

Kama station

Partial Dakah

Stepwise procedure
Chitral → Gawardesh → Chaghasarai → Daronta → Partial Dakah

b)

Kama

Asmar

971
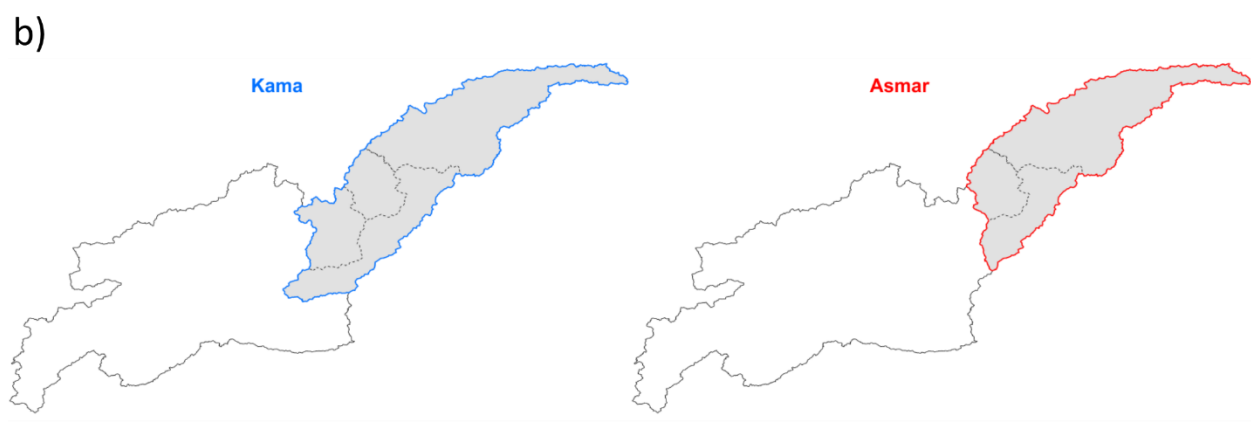
972   Figure 5. (a) Sub-basins corresponding to five gaging stations are used for the multisite
973   calibrations. (b) Two sub-basins (Kama and Asmar) are assumed to be ungaged and used for
974   evaluating the calibration approaches.
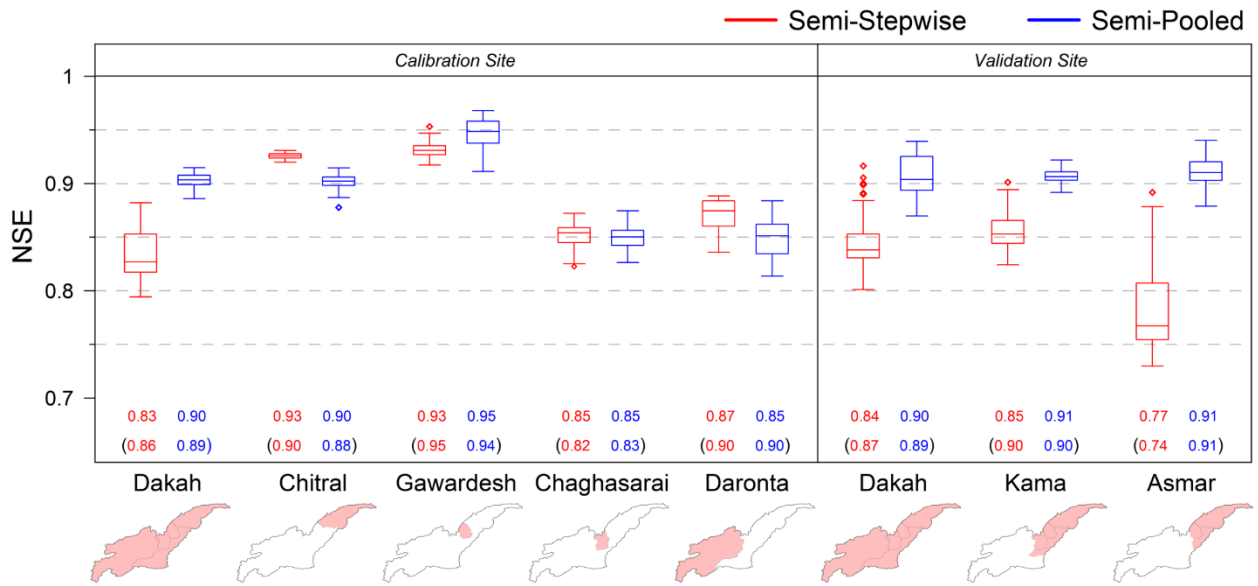
Figure 6. Comparison of the stepwise and pooled calibrations under the semi-distributed parameterization. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.
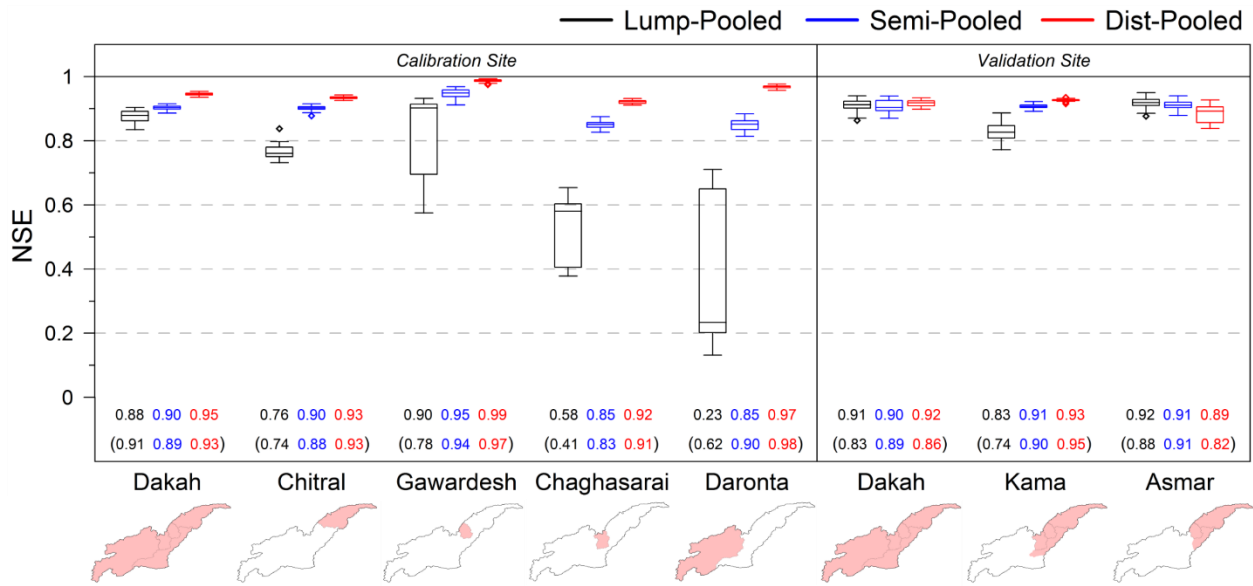
Figure 7. Comparison of the pooled calibrations for the 3 parameterizations of lumped, semi-distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

986
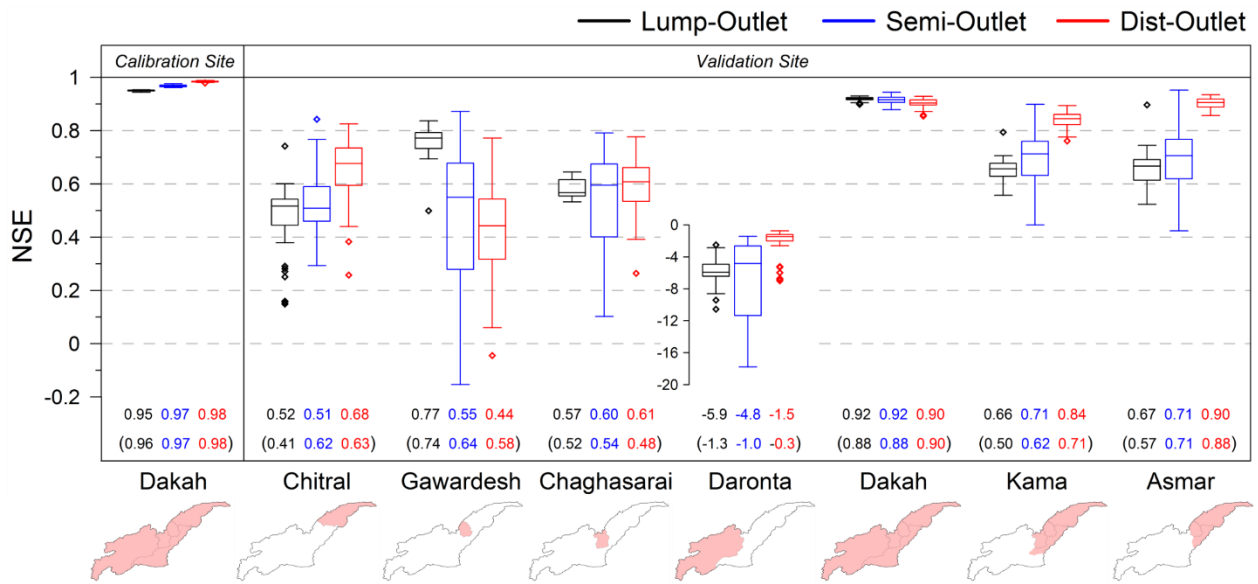
Figure 8. Comparison of the basin outlet calibrations for the 3 parameterizations of lumped, semi-distributed, and distributed. Each calibration is conducted 50 times. Values on the bottom represent expected values of NSE (in upper row) and KGE (within parenthesis in lower row) from 50 calibrations.

991
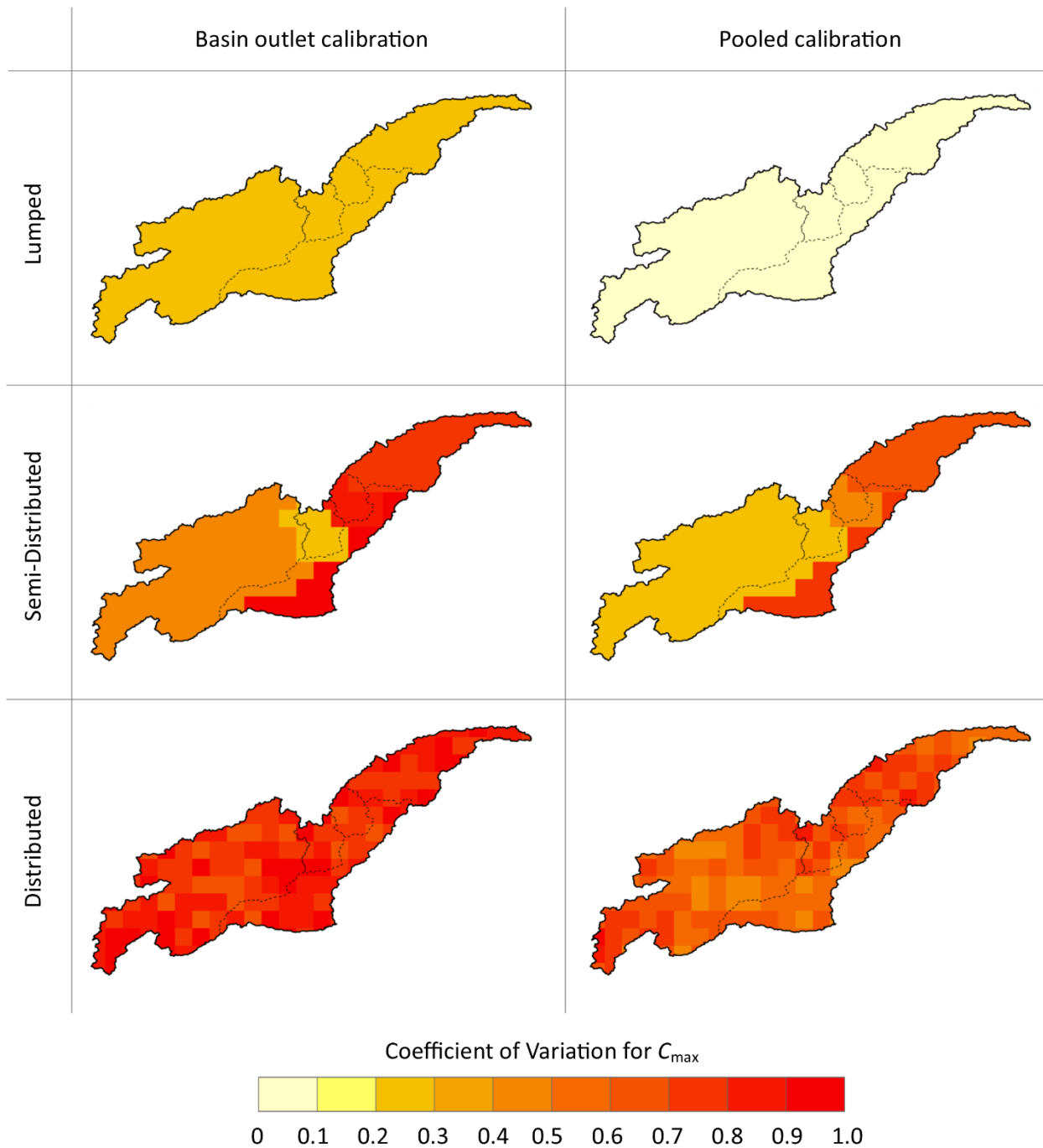
Figure 9. Coefficient of variation (CV) of 50 optimal values of C_{max} (parameter for the soil moisture accounting module in the HYMOD_DS) from the basin outlet calibrations (left panel) and the pooled calibrations (right panel).

992

993
994
995

996

997
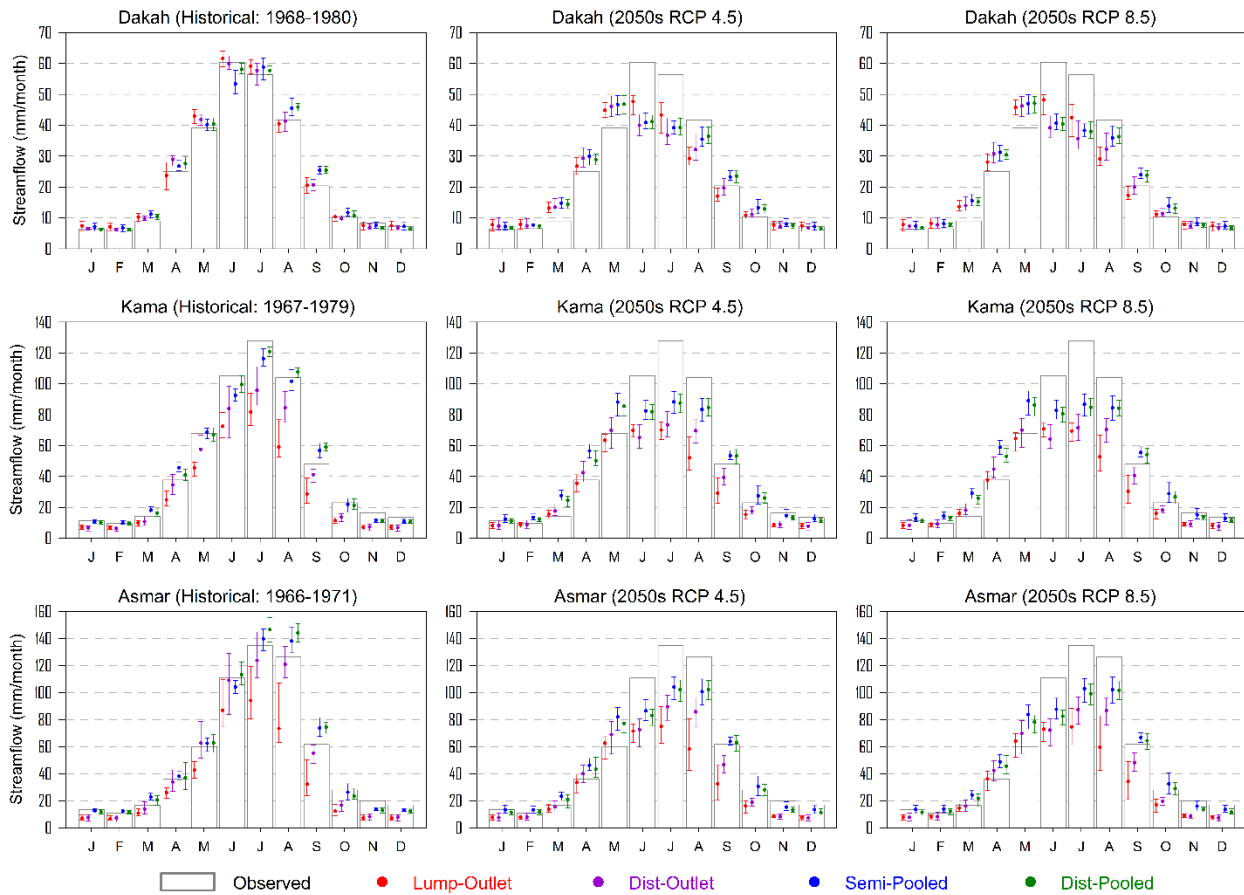
Figure 10. Historical and 2050s average monthly streamflow predictions at Dakah, Kama, and Asmar under 4 calibration strategies: Lump-Outlet, Dist-Outlet, Semi-Pooled, and Dist-Pooled. The error bars represent the streamflow ranges resulting from 50 trails of the HYMOD_DS calibration. For each of the 50 trials, the 2050s streamflow predictions are averaged over 36 GCM climate projections.

1004
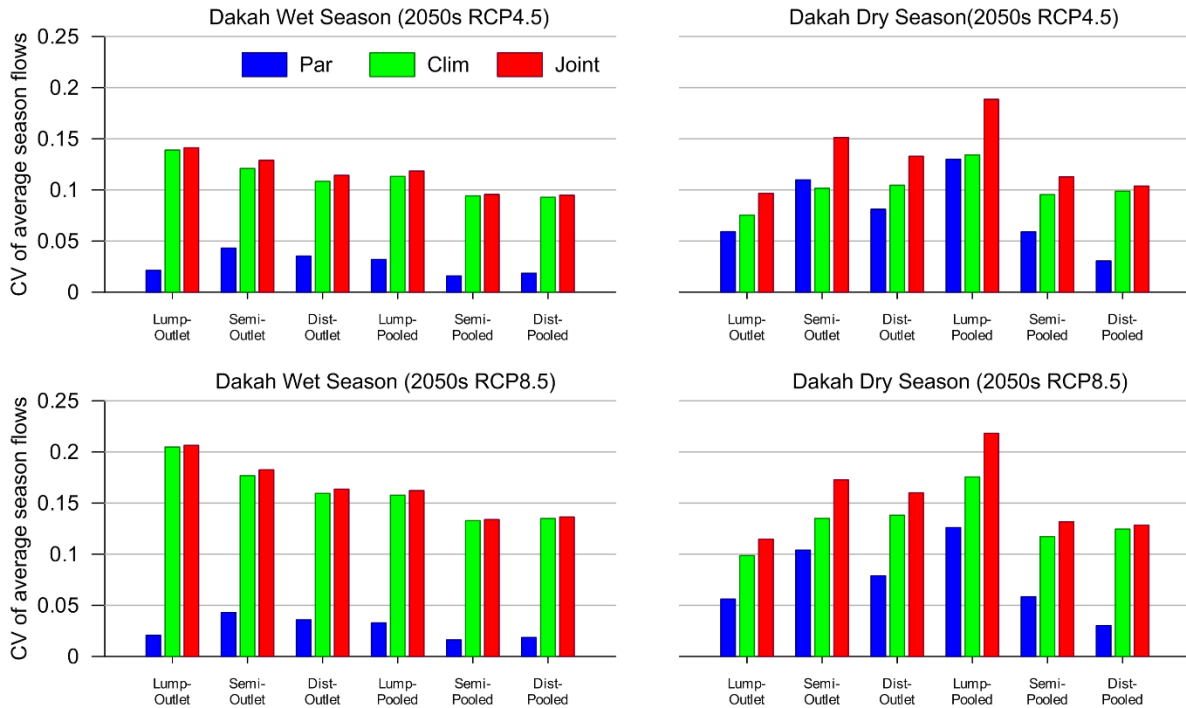
1006   Figure 11. Uncertainties in wet and dry season average streamflow predictions for 2050s are

1007   derived from the basin outlet and pooled calibrations for Dakah. Uncertainties are evaluated by

1008   coefficient of variation (CV) of average season streamflow predictions. Three uncertainty sources

1009   are considered: calibration uncertainty across 50 calibration trials (Par), climate uncertainty across

1010   GCM projections (Clim), and combined uncertainty (Joint).

1011
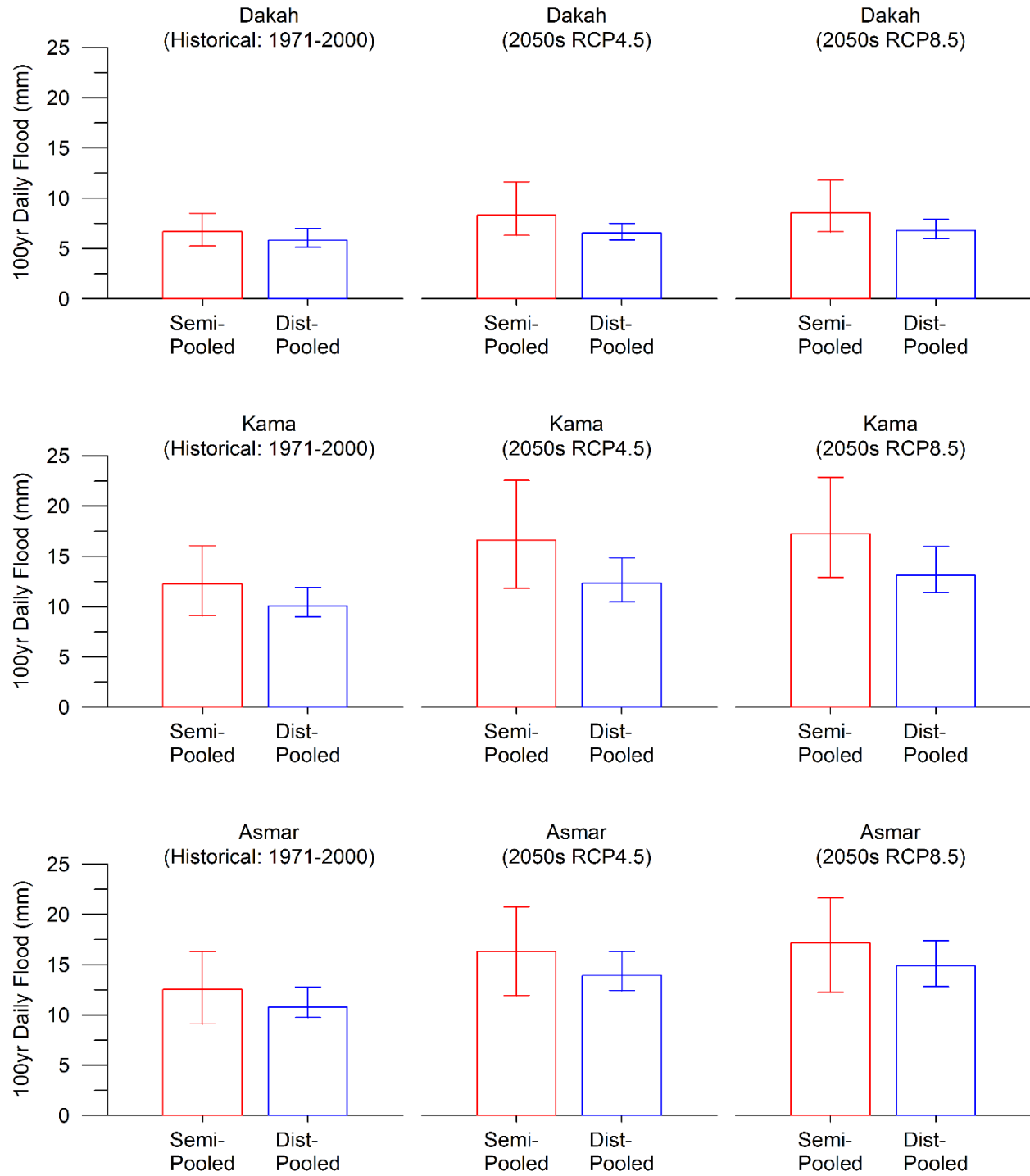
Figure 12. Comparison of GCM average 100-year daily flood events derived from the semi-distributed and distributed pooled calibrations. The uncertainty range is from 50 trials of the model calibration.

1017
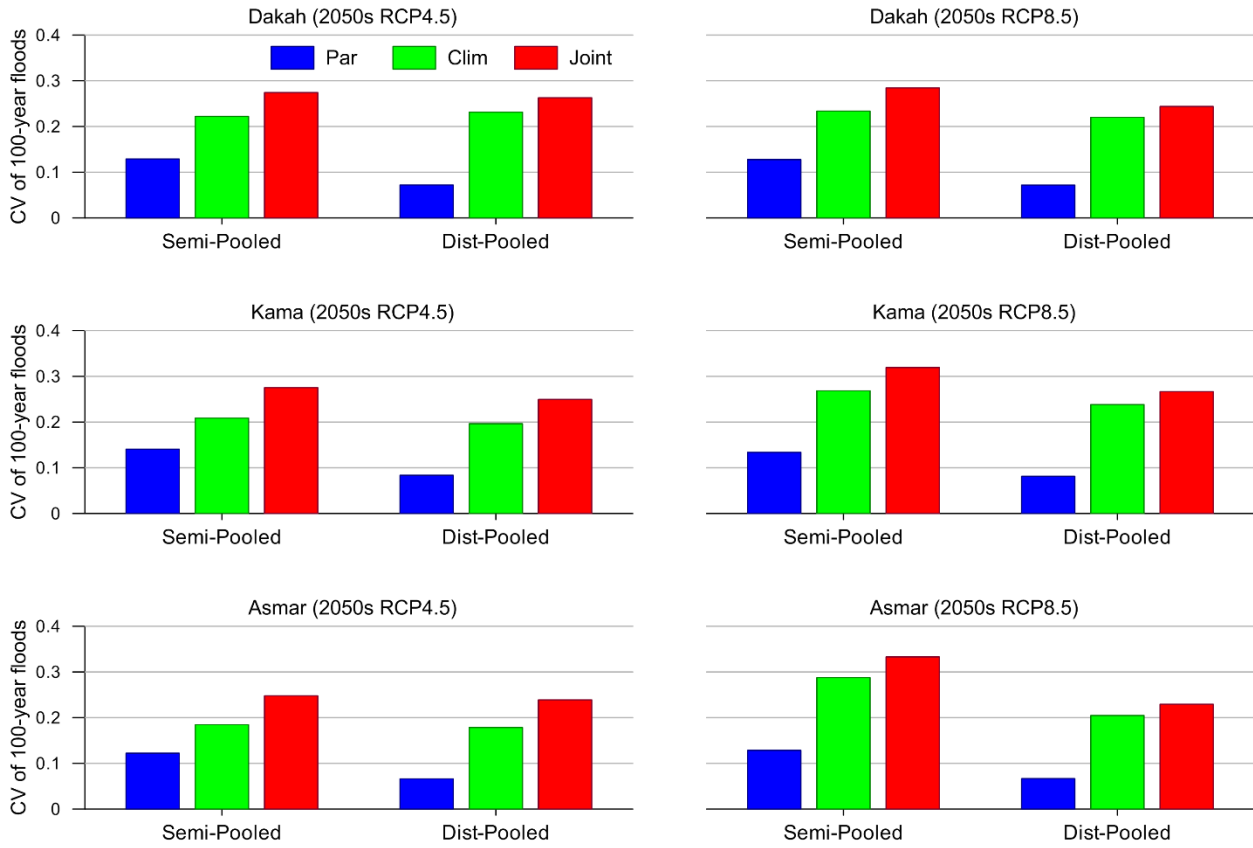
1018

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

1024

1025 **Supplementary materials**

1026
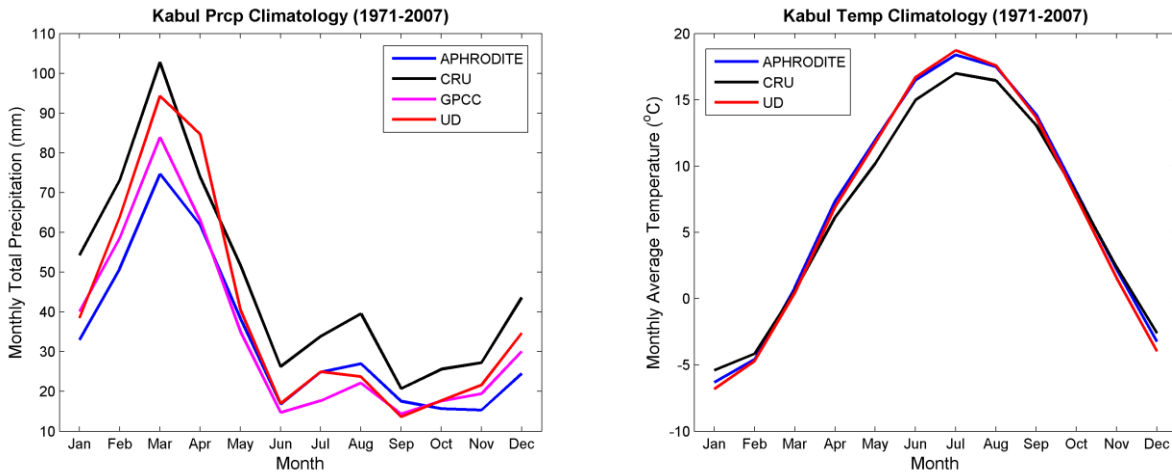
1027

1028   Figure S1. Comparison of basin-wise average monthly precipitation and temperature for the Kabul
1029   River basin. Sources of data sets: APHRODITE (Asian Precipitation High-Resolved
1030   Observational Data Integration Towards Evaluation), CRU (Climatic Research Unit), GPCC
1031   (Global Precipitation Climatology Centre), UD (University of Delaware).

1032

Glacier volume scaling relationship (Grinsted, 2013)

Ice caps	$V = 0.0552R^{0.124}A^{1.20}$
	$V = 0.0432A^{1.23}$
Glaciers	$V = 0.0413R^{-0.0565}A^{1.30}$
	$V = 0.0433A^{1.29}$

V:Volume[km^3]; A: Area[km^2]; R: Elevation range [km]
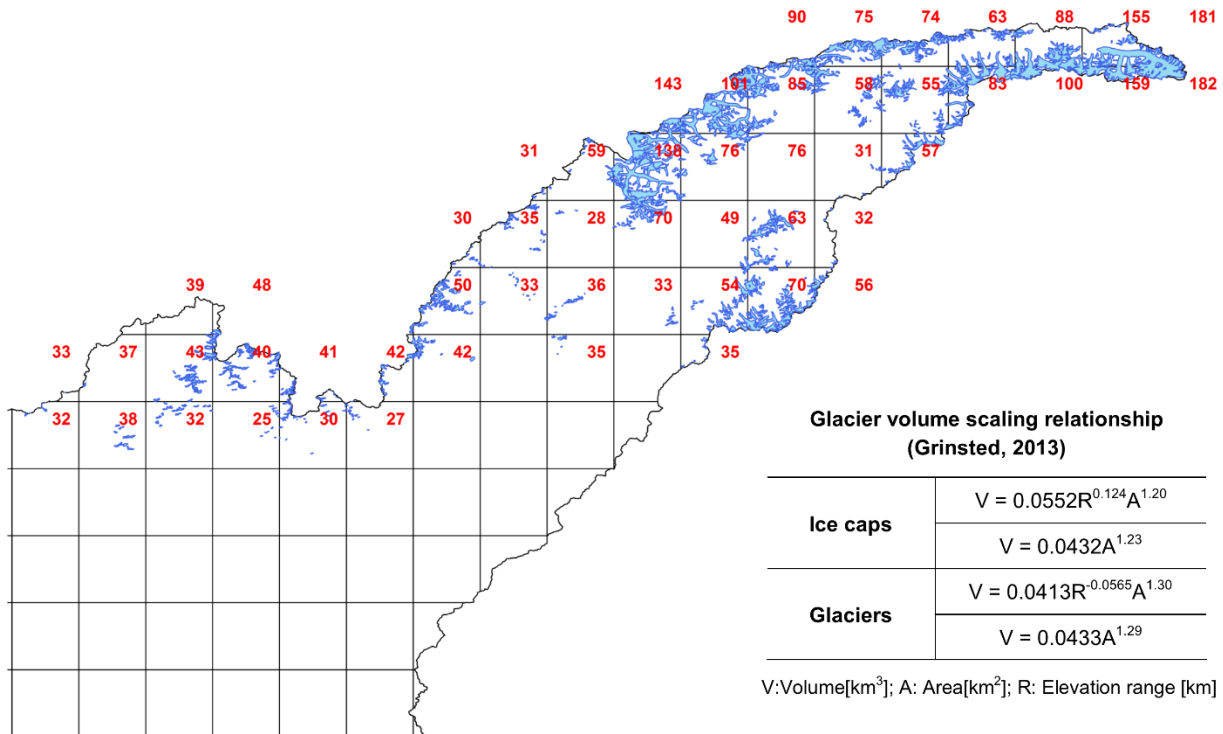
1033

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

1034
1035
1036
1037

1038
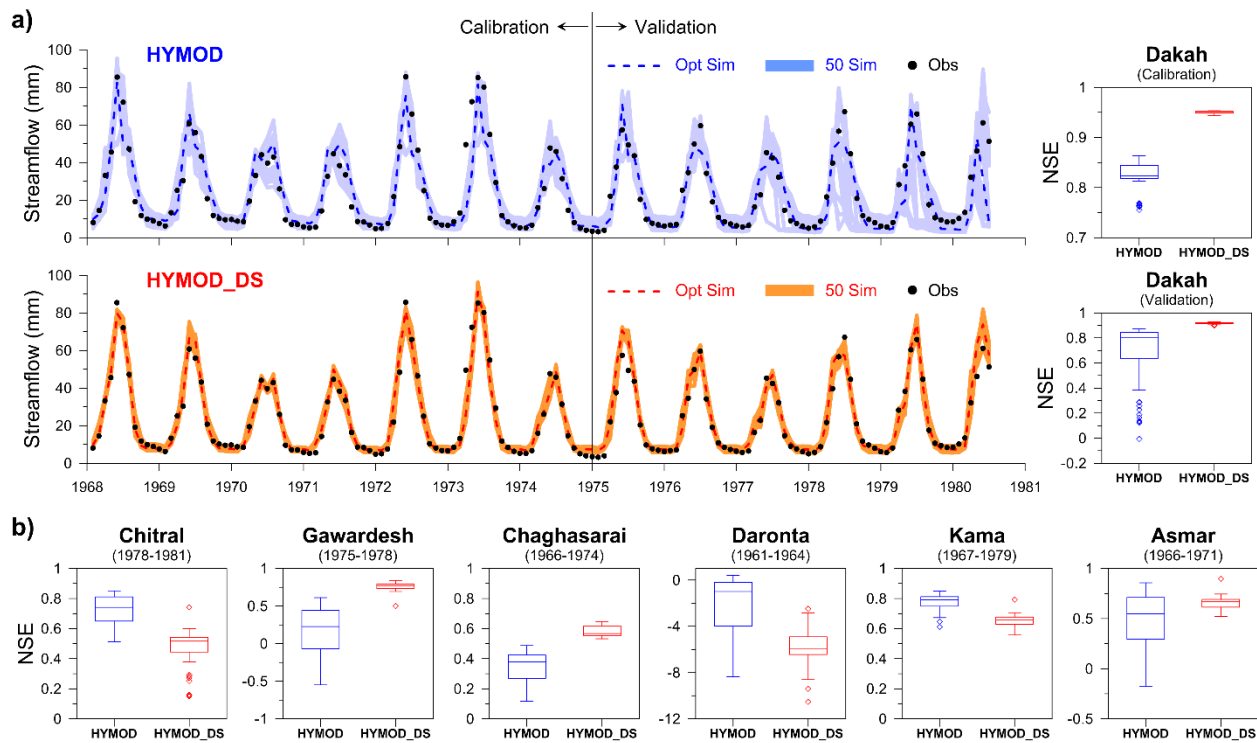
Figure S3. (a) Basin outlet (Dakah) simulations of HYMOD and MYMOD_DS (with the lumped parameterization) from 50 trials of calibration. The Box plots provide the performance evaluation on 50 simulations of both models for both calibration and validation periods. (b) Performances of the models at the interior points of the watershed are assessed.
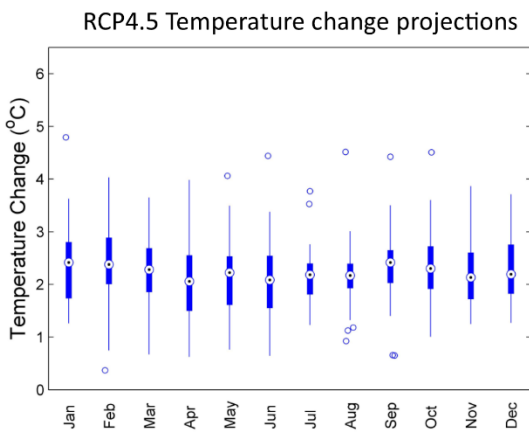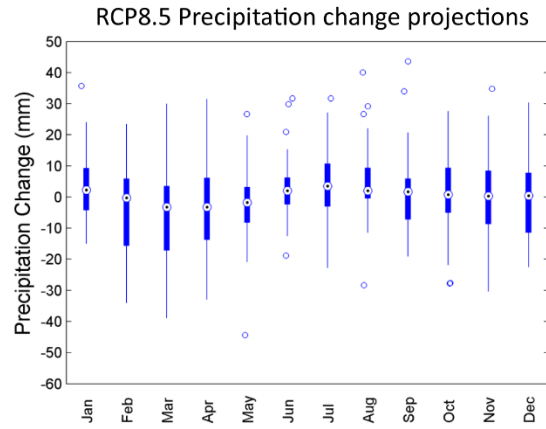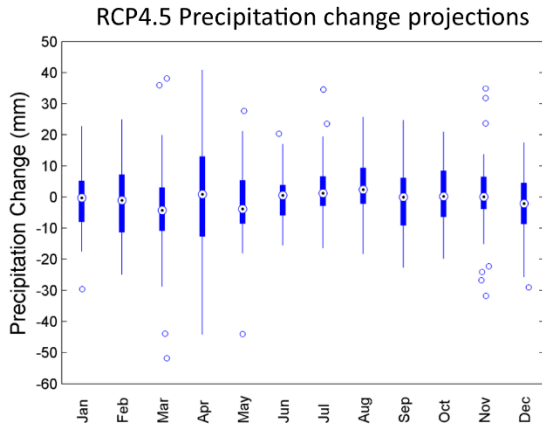
1045
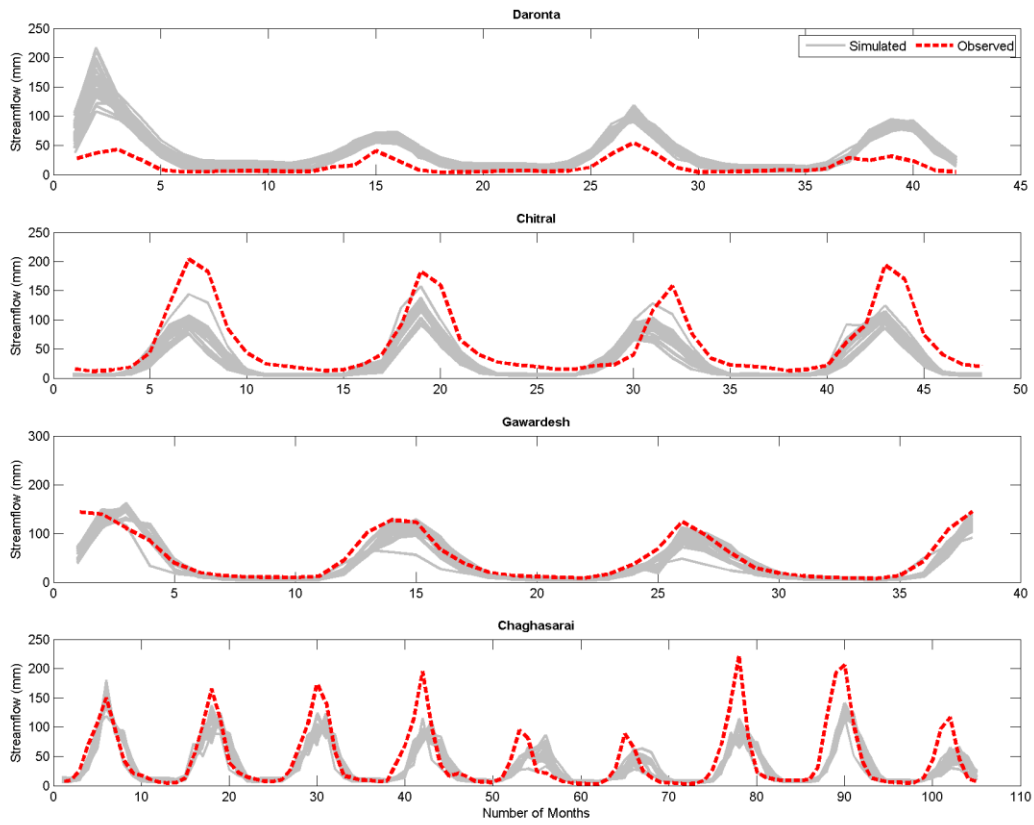
1046

1047 Figure S4. HYMOD_DS streamflow simulations at sub-basins from 50 trials of the basin outlet

1048 calibration under the lumped parameterization.

1049
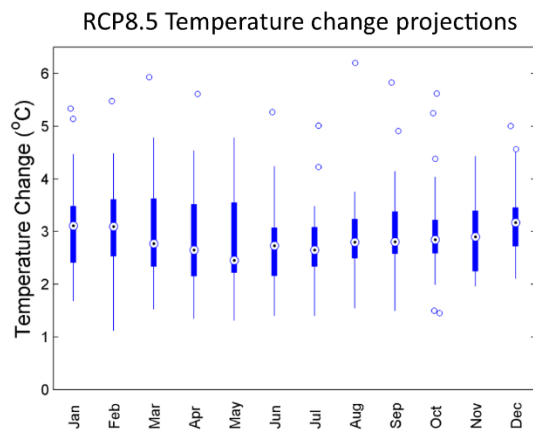
1050

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

1055
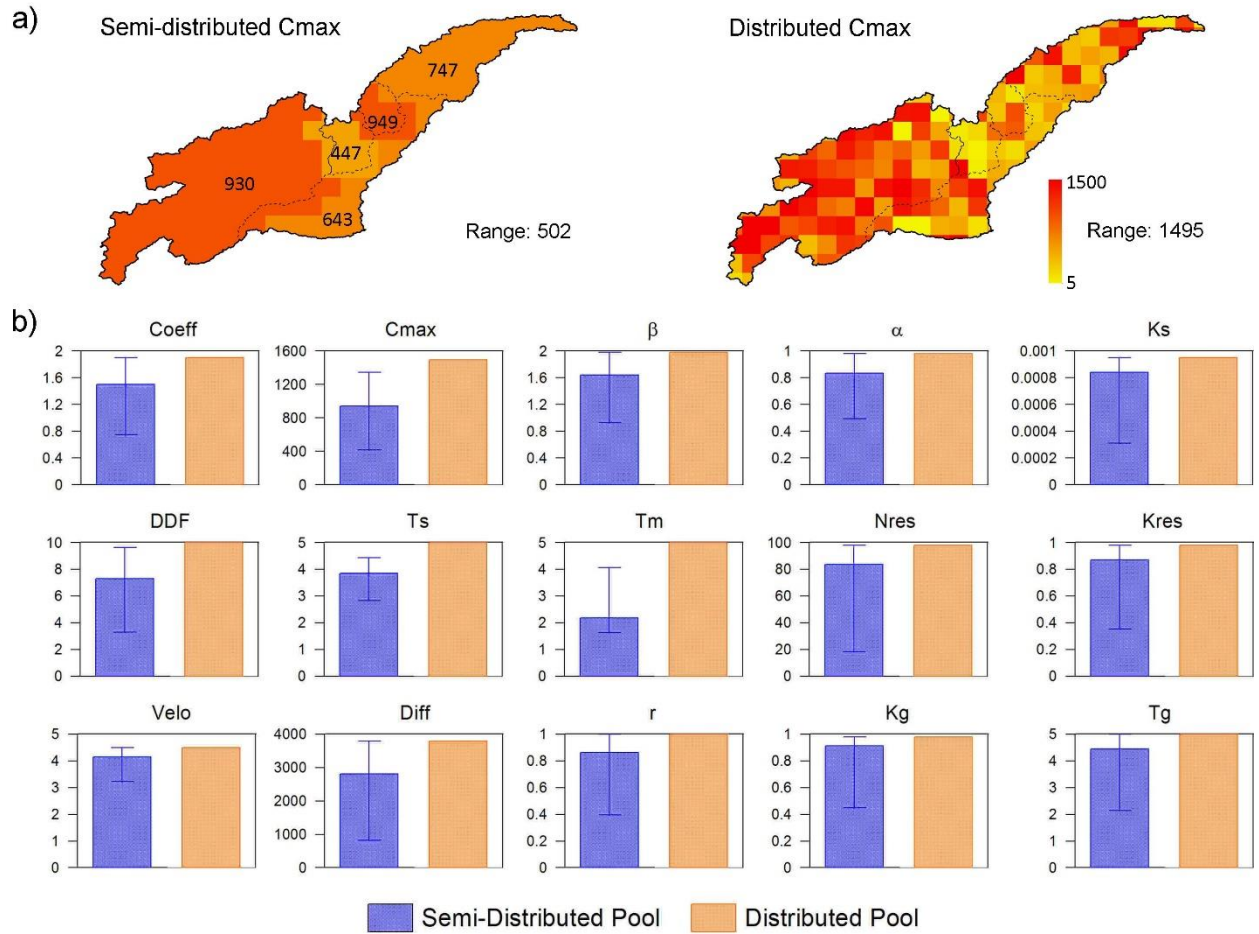
Figure S6. Spatial variability of the HYMOD_DS parameters. a) An example with C_{max} showing parameter ranges resulting from the single trail of Semi-Pooled and Dist-Pooled. b) Average spatial variability across 50 trials of calibration for all 15 parameters. Error bar in b) represents the range of parameter spatial variability from the 50 trails.

1062

HYMOD_DS MODEL (1000 iterations)
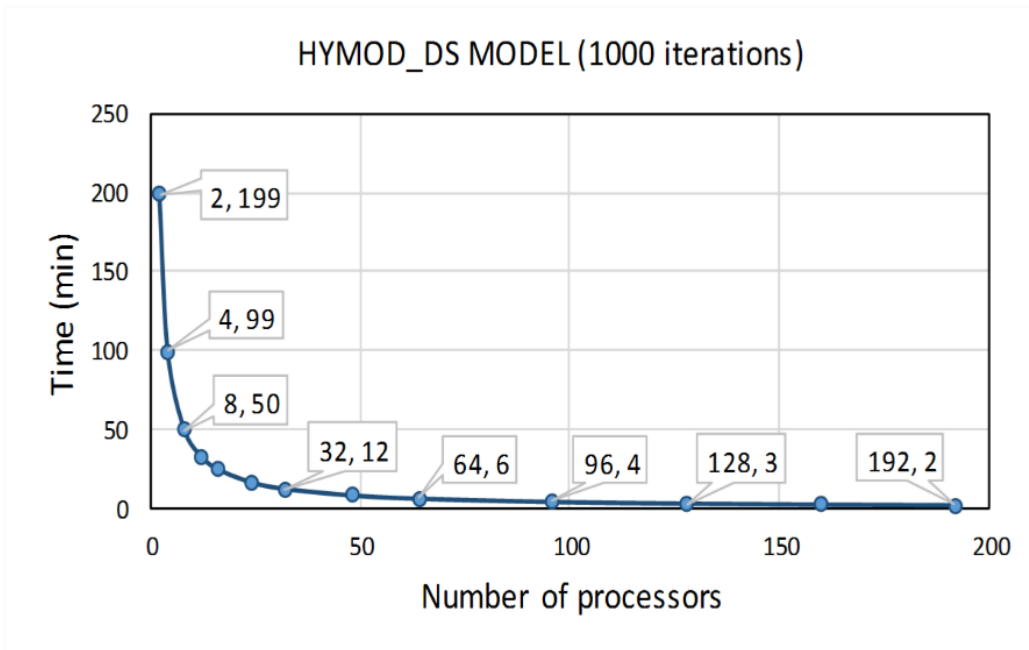
1063

1064   Figure S7. HYMOD_DS run time on parallel computing system.

1065