

1      **Evaluation of atmospheric models using parsimonious  
2      network routing model and streamflow observations  
3      over mountainous regions: A case study of the Yarlung  
4      Zangbo River on the Tibetan Plateau**

5      Heng Yang<sup>1</sup>, Shuanglong Chen<sup>2</sup>, Qingyun Bian<sup>3</sup>, and Hui Zheng<sup>3</sup>

6      <sup>1</sup>Science and Technology Research Institute, China Three Gorges Corporation, Beijing 100038, China

7      <sup>2</sup>Baihetan Hydropower Plant, China Three Gorges Corporation, Sichuan 615421, China

8      <sup>3</sup>Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

9      **Key Points:**

- 10     • A network routing-based method is proposed for atmospheric model evaluation,  
11     leveraging streamflow's superior spatial representativeness  
12     • Parsimonious routing method shows robustness in mountain basins, as celerity cal-  
13     ibration is insensitive to the runoff source  
14     • Uncertainty of parsimonious routing model in mountainous region is sufficiently  
15     small to discriminate among atmospheric model configurations

**16 Abstract**

Evaluating kilometer-scale atmospheric models in data-sparse mountains is challenging because in-situ meteorological observations are scarce and remote-sensing products are uncertain. Using hydrological models to link atmospheric outputs to streamflow is equally problematic, as those models carry large structural and parameter uncertainties in mountain terrain. We therefore propose a simple, low-uncertainty alternative: route the runoff generated by the atmospheric model through a network routing model whose parameters are calibrated against observed streamflow. We apply this approach to assess thirteen 3-kilometer Weather Research and Forecasting (WRF) model experiments over the Yarlung Zangbo River basin on the Tibetan Plateau, configured with systematically varied parameterizations for radiation, microphysics, planetary boundary layer, and orographic drag. In mountainous basins, routing can be substantially simplified by applying the Muskingum model. This parsimonious model proves effective and robust: hourly streamflow simulations achieve a median Pearson correlation of 0.75 across all examined gauging stations, and the calibrated wave celerity shows little sensitivity to the driving WRF experiment. Statistical analysis confirms that routing-model uncertainty is small enough to distinguish performance differences among the WRF configurations. Compared with precipitation-based evaluation, the routing-based approach provides complementary and more hydrologically relevant measures of atmospheric model performance, especially when basin-mean precipitation is already realistic. The method offers a discriminative tool that leverages the superior spatial representativeness of streamflow observations to evaluate atmospheric models in data-sparse mountainous regions.

**38 1 Introduction**

Atmospheric models are essential for hydrological forecasting (Najafi et al., 2024; Cosgrove et al., 2024; Thielen et al., 2009) and hydroclimatic projections (François et al., 2019; Wagner et al., 2016; Fowler et al., 2007; Maraun et al., 2010; Dougherty et al., 2020). Recent advances have steadily increased the spatial resolution of atmospheric models to five kilometers or less (P. Clark et al., 2016; Prein et al., 2015; Stevens et al., 2019; Tang et al., 2023; L. Li et al., 2024). These high-resolution simulations explicitly resolve deep convection, thereby obviating the parameterization that introduces substantial uncertainty in coarser-resolution models. Consequently, they are designated convection-permitting models (Prein et al., 2015; P. Clark et al., 2016; Mooney et al., 2017; Lucas-Picher et al.,

48 2021). Kilometer-scale simulations also capture fine-scale topographic influences on at-  
49 mospheric circulation (C. Lin et al., 2018; Zhou et al., 2021; Yuan et al., 2023; Sugimoto  
50 et al., 2021; Ma et al., 2023; G. Li et al., 2022). The explicit representation of deep con-  
51 vention and detailed topography enhances water cycle simulation accuracy (Jiang, Yang,  
52 Yang, et al., 2022; Sugimoto et al., 2021; Ma et al., 2023). Evidence indicates that kilometer-  
53 scale atmospheric models can outperform in-situ observations (Lundquist et al., 2019)  
54 and satellite-based products (Jiang, Yang, Li, et al., 2022) in mountainous regions, mak-  
55 ing their integration into hydrological applications particularly valuable (Xie et al., 2025;  
56 B. Zhang et al., 2021; Reszler et al., 2018; Rudisill et al., 2023).

57 However, kilometer-scale atmospheric models do not always perform optimally. Achiev-  
58 ing skillful simulations requires careful configuration of modeling domains, initial and  
59 boundary conditions, and process parameterizations (Bauer et al., 2015; Prein et al., 2015;  
60 Lucas-Picher et al., 2021). Atmospheric model intercomparison projects, including the  
61 Atmospheric Model Intercomparison Project (AMIP) (Gates et al., 1999), the Interna-  
62 tional Grand Global Ensemble (TIGGE) of the Observing System Research and Predictabil-  
63 ity Experiment (Swinbank et al., 2016; Parsons et al., 2017), and the Coordinated Re-  
64 gional Climate Downscaling Experiment (CORDEX) (Giorgi et al., 2009), represent sus-  
65 tained efforts to systematically compare atmospheric models across diverse scientific ob-  
66 jectives. In the Tibetan Plateau, a prominent mountainous region, the CORDEX Convection-  
67 Permitting Third Pole (CPTP) project (Prein et al., 2023) intercompares kilometer-scale  
68 atmospheric models through numerous simulations across three selected periods, each  
69 lasting several days (Prein et al., 2023). These projects yield critical insights into atmo-  
70 spheric model uncertainties and performance variations (Gates et al., 1999; Parsons et  
71 al., 2017; Giorgi et al., 2009; Prein et al., 2015, 2023). The optimal configurations iden-  
72 tified through these comparisons guide applications in climate projections, weather pre-  
73 dictions, and regional downscaling.

74 Evaluation of atmospheric models is fundamental to these intercomparison efforts (Prein  
75 et al., 2015; Lucas-Picher et al., 2021), yet remains particularly challenging in data-scarce  
76 mountainous regions. Current evaluation approaches predominantly depend on in-situ  
77 meteorological observations and remote-sensing products (Prein et al., 2023; Collier et  
78 al., 2024; Kukulies et al., 2023; Zou & Zhou, 2024; Ma et al., 2023; Zhou et al., 2021; Karki  
79 et al., 2017). In mountainous terrain, however, in-situ observations are typically sparse  
80 and unevenly distributed, introducing substantial biases (Miao et al., 2024; Lundquist

et al., 2019). Remote-sensing products, including satellite-based precipitation and snow water equivalent estimates, provide enhanced spatial coverage but are subject to considerable uncertainties arising from complex terrain effects and sensor limitations (Henn et al., 2018; Behrangi et al., 2014; Bian et al., 2019). The combination of sparse in-situ observations and uncertainties in remote-sensing data compromises the reliability of kilometer-scale atmospheric model assessments. Collier et al. (2024) conducted the first ensemble of kilometer-scale atmospheric simulations spanning a complete hydrological year (October 2019 to September 2020) over the Tibetan Plateau, finding that substantial discrepancies among observational products impeded model evaluation. This challenge is not unique; the CORDEX-CPTP project similarly considers the identification of suitable observational data for atmospheric model evaluation as a critical priority (Prein et al., 2023).

Streamflow observations constitute a promising basis for evaluating atmospheric models. As the integrated response of a catchment to meteorological variables such as precipitation, temperature, and evapotranspiration, streamflow encapsulates the cumulative effects of these variables across temporal and spatial dimensions (Beven, 2006; Henn et al., 2015, 2016). This characteristic provides an inherent advantage over in-situ meteorological observations in spatial representativeness. Additionally, as direct in-situ measurements, streamflow observations are generally more reliable than remote-sensing products.

Several methodologies have been developed to evaluate atmospheric models using streamflow observations. The most direct approach compares atmospheric model-simulated runoff directly with streamflow observations (Jiang, Yang, Li, et al., 2022). This method assumes that the routing process from runoff generation to streamflow gauges is negligible. This assumption is valid only for small basins or longer time scales where flow wave travel time constitutes a negligible fraction of the analysis period. S. Chen et al. (2025) conducted control experiments with and without routing processes in the Upper Jinsha River basin at the eastern edge of the Tibetan Plateau. They demonstrated that routing effects can be neglected at monthly time scales but significantly influence streamflow variations at daily and sub-daily resolutions. These findings align with Allen et al. (2018), who reported typical flow wave travel times on the order of days. Consequently, direct runoff–streamflow comparison is primarily suitable for headwater catchments or evaluations at monthly and longer time scales. Applying this method, Jiang, Yang, Li,

114 et al. (2022) evaluated basin-wide, multi-year average runoff from atmospheric models.  
115 However, this approach is inadequate for daily or sub-daily evaluations or for medium-  
116 to-large watersheds.

117 Another common approach uses hydrological models to evaluate atmospheric mod-  
118 els (Pang et al., 2020; Z. Zhang et al., 2022; Xie et al., 2025). This methodology uses pre-  
119 cipitation and other surface meteorological variables simulated from atmospheric mod-  
120 els to drive hydrological models, followed by comparison of simulated and observed stream-  
121 flow (Krier et al., 2012; Pang et al., 2020). However, hydrological models are built upon  
122 numerous scientific assumptions about model structure and parameters (M. P. Clark et  
123 al., 2011; Kirchner, 2009; Henn et al., 2016). While these assumptions can be adequately  
124 tested in data-rich regions (M. P. Clark et al., 2011; Zheng et al., 2020), they remain largely  
125 unconstrained in data-sparse mountainous areas such as the Tibetan Plateau. Lei et al.  
126 (2025) investigated streamflow uncertainties stemming from meteorological inputs and  
127 hydrological model structures in the Yarlung Zangbo River basin at the southern edge  
128 of the Tibetan Plateau. They found that model structural uncertainty exceeds meteo-  
129 rological input uncertainty. Similar conclusions emerge from (Kennedy et al., 2025), who  
130 examined a parameter-perturbed ensemble and reported that model parameter uncer-  
131 tainty can surpass meteorological input uncertainty. The substantial uncertainties in-  
132 herent in model structures and parameters render the distinction of different atmospheric  
133 models' outputs with streamflow observations in mountainous regions an ill-posed prob-  
134 lem (Renard et al., 2010): a given atmospheric model output corresponds to a wide range  
135 of feasible streamflow simulations due to hydrological model uncertainties, while mul-  
136 tiple atmospheric models can produce similar streamflow outputs if their uncertainty ranges  
137 overlap. When streamflow observations fall within these overlapping uncertainty ranges,  
138 distinguishing between atmospheric models becomes infeasible. This challenge parallels  
139 that encountered when using precipitation for atmospheric model evaluation over the Ti-  
140 betan Plateau (Collier et al., 2024).

141 Several studies have attempted to mitigate hydrological model uncertainties through  
142 parameter calibration. However, we contend that such efforts are largely ineffective for  
143 atmospheric model evaluation in mountainous regions where hydrological model struc-  
144 tural uncertainty is substantial. While calibration can reduce parameter uncertainty, it  
145 fails to address structural uncertainty (M. P. Clark et al., 2011, 2016). More significantly,  
146 when structural uncertainty is prominent, calibration can render hydrological model pa-

147 rameeters adaptive to model inputs, producing similar streamflow simulations despite dif-  
148 ferent meteorological inputs. S. Chen et al. (2025) compared multiple publicly available  
149 streamflow reanalyses for the Upper Jinsha River basin, revealing that despite substan-  
150 tial differences in meteorological inputs and hydrological models (Alfieri et al., 2020; Har-  
151 rigan et al., 2020; Y. Yang et al., 2021), the reanalyses yield similar streamflow estimates  
152 due to extensive calibration, reasonably reproducing observed streamflow. These results  
153 prove that streamflow observations alone are insufficient to distinguish between differ-  
154 ent meteorological inputs to hydrological models following calibration in mountainous  
155 regions.

156 Recognizing that the challenge primarily arises from hydrological model structural  
157 uncertainty, several studies have sought to reduce this uncertainty. We propose that these  
158 efforts can be classified into two principal approaches. The first approach leverages ad-  
159 ditional observations to impose stronger constraints on model structure. Henn et al. (2015,  
160 2016) employed snow water equivalent observations to constrain model structural un-  
161 certainties and developed a Bayesian inference framework for estimating basin-mean pre-  
162 cipitation from streamflow observations. Their results demonstrated that structural con-  
163 straints render year-to-year variations in inferred precipitation robust, enabling evalua-  
164 tion of long-term atmospheric model simulations (Rudisill et al., 2023). The second ap-  
165 proach utilizes simpler models with fewer parameters to reduce uncertainty. Adam et  
166 al. (2006) applied the Budyko framework—a simple hydrological model—to infer long-  
167 term basin-averaged precipitation from streamflow observations. This framework char-  
168 acterizes the relationship between long-term averaged precipitation and runoff using min-  
169 imal parameters, thereby reducing uncertainty relative to complex hydrological models.  
170 Barkhordari et al. (2025); S. Wang et al. (2019) enhanced this method by integrating  
171 remote-sensing evapotranspiration data into the Budyko framework, further reducing re-  
172 liance on calibratable parameters. However, both approaches are limited to evaluating  
173 long-term basin-averaged precipitation, and their reliability for short-term variations (within  
174 a year), non-precipitation meteorological variables, and scenarios with unclosed water  
175 budget observations (Zheng et al., 2020; Tan et al., 2022) remains largely unverified.

176 Hydrological models typically represent two fundamental processes: runoff gener-  
177 ation and flow routing. Although runoff generation processes exhibit considerable com-  
178 plexity in mountainous regions (van Tiel et al., 2024), flow routing is generally less in-  
179 tricate than in flat terrain (Getirana & Paiva, 2013; Moussa & Bocquillon, 1996). Moun-

tain rivers and terrain, characterized by steep slopes, facilitate water flow representation primarily through the kinematic wave approximation of the Saint–Venant equations (Moussa & Bocquillon, 1996) (channel routing) or the shallow water equations (terrain routing). This representation demonstrates that water flow dynamics are predominantly governed by the interaction between friction and terrain slope (Getirana & Paiva, 2013; Moussa & Bocquillon, 1996). Since terrain slopes can be derived globally from high-resolution digital elevation models (Yamazaki et al., 2017, 2019), water flow routing in mountainous basins depends substantially on friction assumptions. Friction represents a relatively minor component in comprehensive hydrological models. Characterizing friction uncertainties is typically more straightforward than addressing the structural and parameter uncertainties inherent in hydrological models. The relative simplicity of routing processes in mountainous basins may result in lower uncertainty in streamflow simulations. Consequently, flow routing models may be more suitable for evaluating atmospheric models in these regions.

Building upon this rationale, we propose an alternative approach for evaluating atmospheric models that exclusively utilizes flow network routing. This methodology directly routes atmospheric model-simulated runoff through the flow network to generate streamflow estimates at gauge locations. We hypothesize that due to the relative simplicity of routing processes in mountainous basins, the uncertainty contributed by the routing model is small enough to allow statistically significant discrimination among different atmospheric model configurations.

This study aims at demonstrate that a parsimonious network routing model can serve as a robust evaluation tool for kilometer-scale atmospheric simulations in data-scarce mountain regions. We applied the proposed method to the Yarlung Zangbo River basin. The basin is the largest river basin on the Tibetan Plateau and is now the site of the world's largest hydropower development project. Accurate atmospheric simulations are essential for operating and managing hydropower facilities along the river. Yet their evaluation is hindered by sparse in-situ observations and large uncertainties in satellite retrievals. The river extends approximately 2,000 kilometers, with an average elevation exceeding 4,000 meters above sea level. Its catchment area spans approximately 2.5 million square kilometers. The basin's scale precludes direct runoff–streamflow comparison at sub-monthly time scales, while its steep, high-altitude terrain introduces large structural uncertainty

212 in hydrological models. These characteristics make it an ideal testbed for routing network-  
213 based evaluation method.

214 The structure of this paper is as follows: Section 2 delineates the methodologies  
215 and datasets employed in our study. Section 3 presents the analytical outcomes. Finally,  
216 Section 4 synthesizes the principal findings and provides concluding remarks.

## 217 **2 Methods and Data**

### 218 **2.1 Experimental Design**

219 We implemented the proposed methodology to evaluate different parameterization  
220 schemes of the Weather Research and Forecasting (WRF) model (Powers et al., 2017).  
221 Figure 1 illustrates the experimental workflow. A series of thirteen WRF experiments  
222 with 3-kilometer grid spacing were conducted from May 1 to October 1, 2013, encom-  
223 passing the wet season. We selected 2013 because its streamflow closely approximates  
224 long-term climatology, and the wet season accounts for the majority of annual stream-  
225 flow (Zhou et al., 2021). The initial 50 days served as a spin-up period (May 1 to June  
226 19), sufficiently long to mitigate the effects of potentially inaccurate initial conditions,  
227 including snowpack. Simulations from June 20 to October 1 were analyzed.

228 The WRF experiments were initialized and driven by data from the European Cen-  
229 tre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) (Hersbach et al.,  
230 2020). WRF-simulated runoff was spatially re-mapped to represent lateral inflow into  
231 a pre-delineated flow network. Water flow within this network was subsequently com-  
232 puted using the Muskingum method, a well-established approach for routing mountain-  
233 ous river flows with minimal calibratable parameters.

234 Routing parameters and associated uncertainties were estimated through a three-  
235 step procedure. First, optimal parameter values were determined for each WRF exper-  
236 iment by calibrating against streamflow observations. Calibration followed an upstream-  
237 to-downstream sequence, ensuring that routing parameters for upstream gauges were es-  
238 tablished prior to those for downstream gauges. This sequential approach captures spa-  
239 tial dependencies of routing parameters along the river network. Second, optimal param-  
240 eter values were used to fit log-normal distributions at each river gauge. Third, these fit-  
241 ted distributions were subsequently used to generate ensembles of 100 random param-

242 parameter sets. Flow routing ensembles were executed using these parameter sets for each WRF  
243 experiment, enabling estimation of routing uncertainty and statistical assessments.

244 The WRF experiments were intercompared according to their skill in reproducing  
245 observed streamflow. Statistical significance of performance differences between them  
246 was assessed. Statistically significant disparities thereby support the study's central hy-  
247 pothesis.

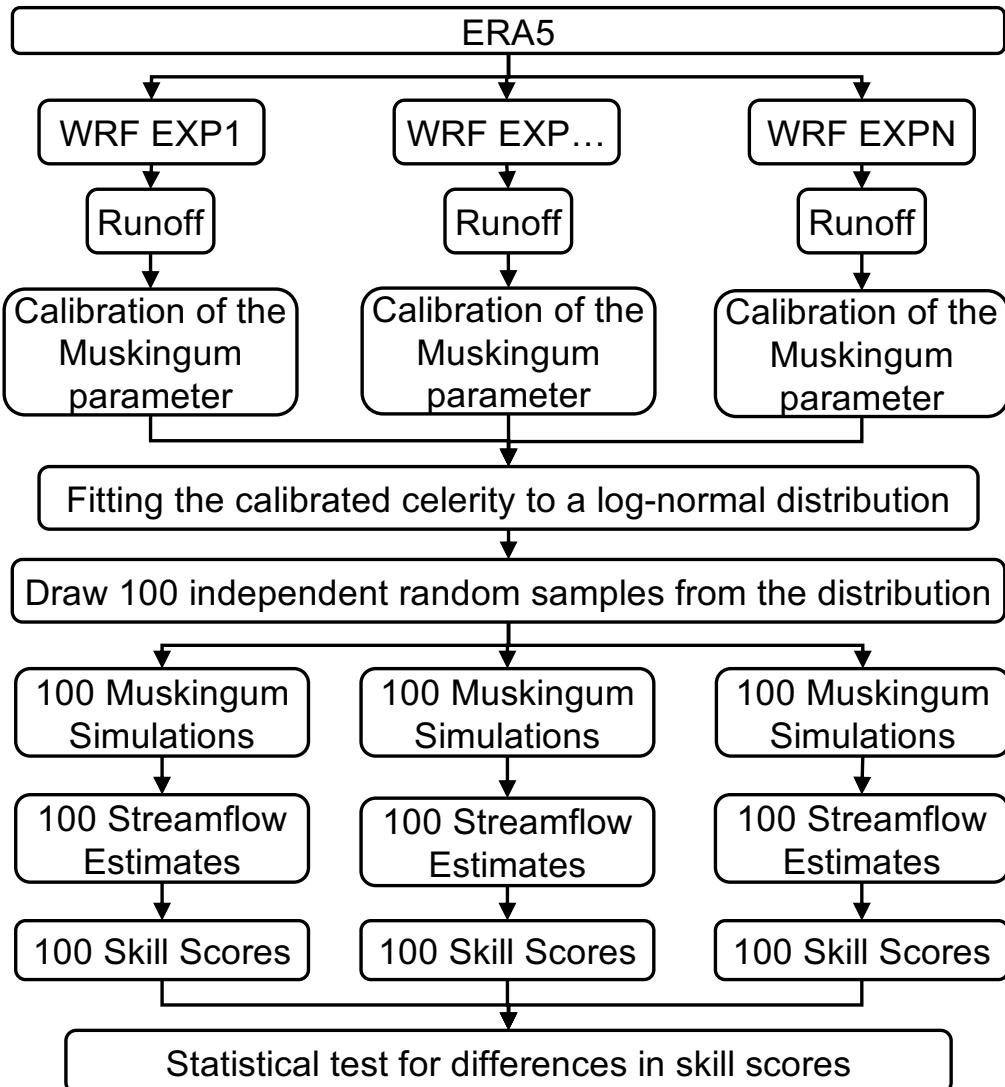
## 248 **2.2 Observational Data**

249 Streamflow data were obtained from the China Three Gorges Corporation and Min-  
250 istry of Water Resources of China, collected at six river gauges: Lazi, Nugesha, Lhasa,  
251 Yangcun, Gengzhang, and Nuxia. Figure 2 illustrates the spatial distribution of these  
252 gauge locations. Lazi, Nugesha, Yangcun, and Nuxia are situated along the mainstream  
253 in upstream-to-downstream sequence. Lhasa is located on a tributary between Nugesha  
254 and Yangcun, while Gengzhang lies between Yangcun and Nuxia. Instantaneous stream-  
255 flow observations were recorded hourly from June 20 to October 1, 2013.

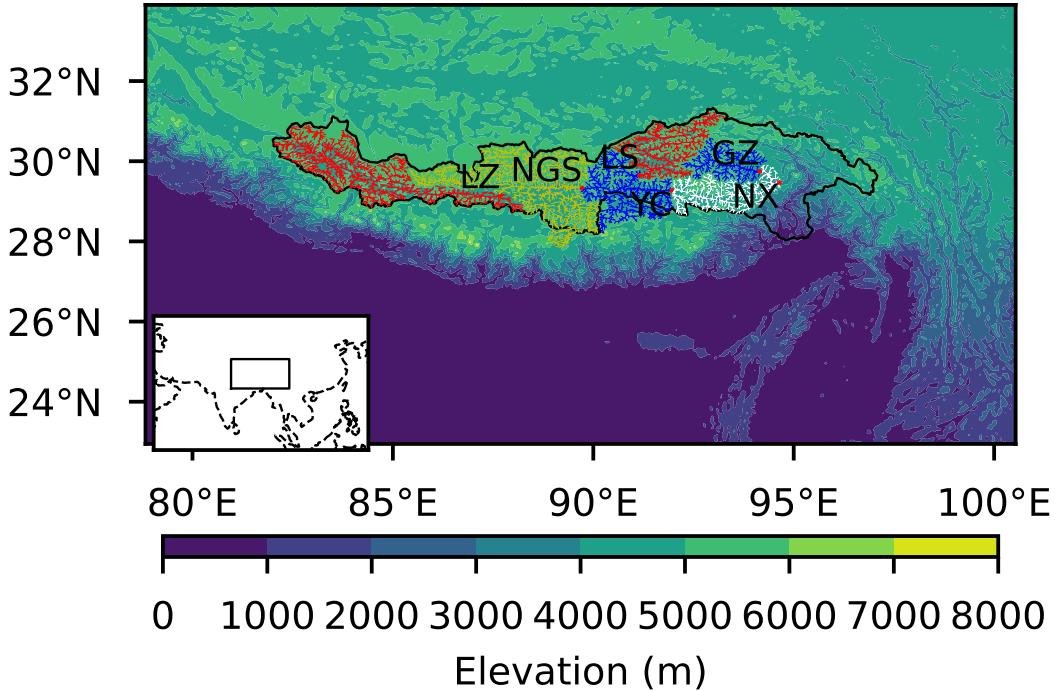
256 To contextualise the proposed evaluation method, we examined how tightly the eval-  
257 uation results track independent evaluations of the same WRF experiments. We used  
258 the Global Precipitation Measurement (GPM) Multi-satellitE Retrievals for GPM (IMERG)  
259 product (Huffman et al., 2023) as an illustrative benchmark. The GPM IMERG prod-  
260 uct provides data at  $0.1^\circ$  spatial resolution and 30-minute temporal resolution. The data  
261 were bilinearly interpolated to the 3-kilometer WRF grid to permit pixel-wise compar-  
262 ison.

## 263 **2.3 Study Area and Flow Routing Network**

264 Figure 2 displays the WRF model domain, which spans 380 by 660 grid cells, each  
265 measuring 3 by 3 kilometers. The domain covers the entire Yarlung Zangbo River basin  
266 and extends to include a buffer zone surrounding the basin. This buffer zone, which is  
267 over 200 kilometers wide, is strategically designed to allow the development of small-scale  
268 weather systems before they interact with the river basin (Denis et al., 2002). This de-  
269 sign helps to mitigate adverse effects stemming from inaccurate or low-resolution bound-  
270 ary conditions.



**Figure 1.** Schematic diagram of the workflow of this study. The parameterization schemes used in the WRF experiments are listed in Table 1.



**Figure 2.** WRF domain, boundary of the Yarlung Zangbo River, and the delineated flow network. The colormap provides a representation of the terrain elevation within the WRF domain, showcasing the basin's topographical characteristics. Black lines denote the basin boundary. Red dots indicate the locations of the four river gauges along the river's course, labeled as follows: LZ for Lazi, NGS for Nugesha, LS for Lhasa, YC for Yangcun, GZ for Gengzhang, and NX for Nuxia. Colored lines correspond to the flow paths that lie between consecutive gauges, offering a visual guide to their spatial distribution across the basin. The inset illustrates the study domain, depicted as a rectangle, along with the coastlines, which are represented by dashed lines.

271       The routing was performed on the network as illustrated in Figure 2. The rout-  
 272       ing network was delineated from the Multi-Error Removed Improved-Terrain Hydrog-  
 273       raphy (MERIT-Hydro) dataset (Yamazaki et al., 2017, 2019), which provides flow di-  
 274       rections and accumulative upstream area data at 3 by 3 arcseconds. The delineation pro-  
 275       cess proceeds in three sequential steps: Initially, the accumulative upstream area is em-  
 276       ployed to identify flow paths, with a grid cell being classified as such if its accumulative  
 277       upstream area exceeds  $10 \text{ km}^2$ . Subsequently, these flow paths are segmented from up-  
 278       stream to downstream, defining a flow path segment by an increase in the accumulative  
 279       upstream area of at least  $20 \text{ km}^2$ . Finally, the flow direction data are utilized to deter-  
 280       mine the catchment area of each flow path segment. The thresholds for defining flow paths

281 aligns with that used in previous large-domain river routing studies (P. Lin et al., 2021,  
 282 2019). This delineation process results in a fully connected network consisting of 5,800  
 283 flow path segments, with an average catchment area of 33 km<sup>2</sup>.

## 284 2.4 WRF Parameterization Schemes

285 Table 1 presents the parameterization schemes selected from WRF version 4.3.3.  
 286 Thirteen experiments were conducted, a number that satisfies the minimum requirement  
 287 for statistical estimation of celerity distribution while maintaining computational fea-  
 288 sibility. Experiments were configured by systematically modifying one parameterization  
 289 scheme at a time while maintaining consistency across other schemes. This design min-  
 290 imizes inter-experiment differences, thereby increasing the challenge of statistical signif-  
 291 icance tests and enhancing the robustness of study findings.

292 The experiments are organized into four groups based on the process being mod-  
 293 ified: radiation, cloud microphysics, planetary boundary layer, and orographic drag. These  
 294 parameterizations have demonstrated significance in previous investigations (Lv et al.,  
 295 2020; Prein et al., 2023).

296 The first group comprises five experiments (E01 to E05) examining radiation pa-  
 297 rameterization impacts. E03 was configured similarly to the High Asian Refined Anal-  
 298 ysis version 2 (X. Wang et al., 2021), with modifications to radiation and land surface  
 299 processes. The Rapid Radiative Transfer Model for GCMs (RRTMG) (Iacono et al., 2008)  
 300 replaced the Rapid Radiative Transfer Model (RRTM) scheme (Mlawer et al., 1997) for  
 301 both shortwave and longwave radiation transfer. RRTMG provides comparable radi-  
 302 ative forcing modeling to RRTM while offering enhanced computational efficiency (Iacono  
 303 et al., 2008). For land surface processes, the Noah land surface model with multiparam-  
 304 eterization options (Noah-MP) (Niu et al., 2011; Z.-L. Yang et al., 2011) was selected  
 305 instead of the Noah model used in HARR version 2. Noah-MP incorporates improve-  
 306 ments over the original Noah model, particularly in representing snow and runoff pro-  
 307 cesses (Niu et al., 2011). These enhancements have yielded superior runoff modeling per-  
 308 formance (Liang et al., 2019; Zheng et al., 2023), leading to Noah-MP's widespread adop-  
 309 tion in hydrological applications (Cosgrove et al., 2024; P. Lin, Hopper, et al., 2018). Among  
 310 Noah-MP's various runoff generation parameterization options, this study employed the  
 311 Noah runoff scheme, which utilizes exponentially distributed infiltration capacity for runoff

312 generation (Schaake et al., 1996) and assumes free drainage at the soil column base to  
 313 simulate subsurface runoff. Experiments E01 to E05 systematically replaced shortwave  
 314 and longwave radiation schemes individually, with the widely used Dudhia and Goddard  
 315 schemes selected for these perturbations (Dudhia, 1989; Matsui et al., 2020).

316 The second group includes three experiments (E03, E06, and E07) investigating  
 317 cloud microphysics scheme effects. E03 employs the Thompson scheme (Thompson et  
 318 al., 2008), while E06 and E07 utilize the Purdue Lin scheme (S.-H. Chen & Sun, 2002)  
 319 and the WRF Single-Moment 6-Class Microphysics (WSM6) scheme (Hong & Lim, 2006),  
 320 respectively. These three schemes represent the most commonly used single-moment mi-  
 321 crophysics parameterizations in WRF applications (Hong & Lim, 2006; Thompson et al.,  
 322 2008; S.-H. Chen & Sun, 2002).

323 The third group contains five experiments (E03, E08 to E12) examining planetary  
 324 boundary layer scheme impacts. Given the tight coupling between surface layer and shal-  
 325 low convection schemes with planetary boundary layer schemes, these were also exam-  
 326 ined within this group. E03 uses the Mellor–Yamada–Janjić scheme (Janjić, 1994), while  
 327 E08 to E12 employ: the Mellor–Yamada–Nakanishi–Niino level 2.5 scheme (Nakanishi  
 328 & Niino, 2006, 2009), the Yonsei University scheme (Hong et al., 2006), the Asymmet-  
 329 ric Convective Model version 2 (ACM2) scheme (Pleim, 2007a, 2007b), the Quasi-Normal  
 330 Scale Elimination (QNSE) scheme (Sukoriansky et al., 2005), and the Yonsei University  
 331 scheme coupled with the Global/Regional Integrated Modeling System shallow convec-  
 332 tion scheme (Hong & Jang, 2018). These schemes are extensively used in WRF appli-  
 333 cations and have proven effective for simulating boundary layer processes across diverse  
 334 climatic regions.

335 The fourth group encompasses two experiments (E08 and E13) exploring orographic  
 336 drag parameterization effects. Building upon E08, E13 incorporates small-scale (Tsiringakis  
 337 et al., 2017) and turbulent orographic drag (Beljaars et al., 2004).

## 338 2.5 Flow Routing Method

339 The WRF model includes a routing module known as WRF-Hydro (P. Lin, Yang,  
 340 et al., 2018; Givati et al., 2016). WRF-Hydro comprehensively integrates subsurface runoff,  
 341 soil moisture redistribution, diffusive terrain routing, and Muskingum–Cunge channel  
 342 routing. However, its complexity introduces a large number of parameterization options

**Table 1.** WRF experiments and corresponding parameterization schemes. Radiation column shows shortwave and longwave schemes, respectively. Planetary boundary layer column indicates planetary boundary layer, surface layer, and shallow convection parameterizations, respectively. For orographic drag, L and S denote large- and small-scale gravity waves; B and T denote flow blocking and turbulent orographic form drag, respectively. Parameterization abbreviations and full references are provided in Section 2.4.

Experiment	Radiation	Cloud Microphysics	Planetary Boundary Layer	Orographic Drag
E01	Dudhia & RRTMG	Thompson	MYJ & Eta & -	LB
E02	Dudhia & RRTMG	Thompson	MYJ & Eta & -	LB
E03	RRTMG & RRTMG	Thompson	MYJ & Eta & -	LB
E04	RRTMG & Goddard	Thompson	MYJ & Eta & -	LB
E05	Goddard & Goddard	Thompson	MYJ & Eta & -	LB
E06	RRTMG & RRTMG	Prudue Lin	MYJ & Eta & -	LB
E07	RRTMG & RRTMG	WSM6	MYJ & Eta & -	LB
E08	RRTMG & RRTMG	Thompson	MYNN2 & MYNN & EDMF	LB
E09	RRTMG & RRTMG	Thompson	YSU & MM5 & -	LB
E10	RRTMG & RRTMG	Thompson	ACM2 & MM5 & -	LB
E11	RRTMG & RRTMG	Thompson	QNSE & QNSE & QNSE	LB
E12	RRTMG & RRTMG	Thompson	YSU & MM5 & GRIMS	LB
E13	RRTMG & RRTMG	Thompson	MYNN2 & MYNN & EDMF	LBST

and parameters, making thorough calibration and uncertainty quantification challenging (RafieeiNasab et al., 2025), particularly in data-sparse regions such as the Yarlung Zangbo River basin (Lei et al., 2025).

As described in the Introduction, routing processes in mountainous regions can be substantially simplified. In alignment with our study purpose, we opted for Muskingum routing over a dense flow network instead of WRF-Hydro. A dense flow routing network functions not only as a prescribed water flow path for channel routing but also, to some extent, as a surrogate for terrain routing. When sufficiently dense, terrain routing within small catchments with steep slopes can be neglected.

The WRF-simulated runoff was re-mapped to represent lateral inflow into the flow routing network, following the remapping method detailed in (P. Lin, Yang, et al., 2018; S. Wang et al., 2019). Unlike P. Lin, Yang, et al. (2018), we identified all WRF grid cells that intersect with the catchment of a water flow path segment, rather than using a single grid cell at the centroid of the catchment. Runoff volume from these grid cells was then calculated by multiplying the runoff depth by the intersection area for each cell. These individual values were subsequently summed to determine the total lateral flow volume entering the flow path segment. This method conserves the total runoff volume despite variations in network geometry.

The Muskingum method is well-suited to characterize flow's kinematic wave propagation driven by the topographic gradient (Ponce et al., 1978). The formulation of the Muskingum method (Cunge, 1969; Fenton, 2019), when incorporating lateral inflows, can be expressed as follows:

$$Q_i^t = \frac{k - x}{1 - x + k} Q_{i-1}^t + \frac{1 - x - k}{1 - x + k} Q_i^{t-1} + \frac{x + k}{1 - x + k} Q_{i-1}^{t-1} + \frac{2k}{1 - x + k} Q_l^t, \quad (1)$$

$$k = \frac{c\Delta t}{2\Delta l}, \quad (2)$$

where  $Q_i^t$  ( $\text{m}^3 \text{s}^{-1}$ ) is the streamflow at current time step,  $Q_{i-1}^t$  ( $\text{m}^3 \text{s}^{-1}$ ) is the streamflow at the upstream position at current time step,  $Q_i^{t-1}$  ( $\text{m}^3 \text{s}^{-1}$ ) is the streamflow at previous time step,  $Q_{i-1}^{t-1}$  ( $\text{m}^3 \text{s}^{-1}$ ) is the streamflow at the upstream position at the previous time step,  $Q_l^t$  ( $\text{m}^3 \text{s}^{-1}$ ) is the lateral inflow at current time step.  $\Delta t$  is the time step (s), and  $\Delta l$  is the flow path length (m).  $x$  is the weighting factor (unitless), and  $c$  is the wave celerity ( $\text{ms}^{-1}$ ).

The application of the Muskingum method to a fully connected flow network must be executed in the correct order. Streamflow at the upstream position must be available before routing. We sorted all flow paths within the network according to stream order as introduced in (X. Yang et al., 2024). This sorting ensures upstream flow paths are always prioritized over their downstream counterparts. We then apply the Muskingum method to each flow path segment in the established sequence. This sequence guarantees that routing occurs first on upstream segments, followed by downstream segments.

The Muskingum method introduces two adjustable routing parameters: the weighting factor ( $x$ ; unitless) and wave celerity ( $c$ ;  $\text{ms}^{-1}$ ). The weighting factor  $x$  regulates the proportional contributions of the right-hand-side terms in Equation 1. Previous studies demonstrate that the simulated streamflow is relatively insensitive to variations in

382 the weighting factor (Koussis, 1978). Values ranging from 0.1 to 0.3 typically prove ef-  
 383 fective for most streams. Guided by the experiments reported by David et al. (2011),  
 384 we have chosen a parameter value of 0.3 for this study.

385 The wave celerity  $c$  represents the speed at which the flow wave propagates down-  
 386 stream. This parameter is influenced by the physical characteristics of the flow path, in-  
 387 cluding slope, roughness, and width. Given that direct observations of  $c$  are unavailable  
 388 for the Yarlung Zangbo River basin, we calibrated  $c$  using streamflow observations. The  
 389 calibration is performed for each river gauge sequentially from upstream to downstream.  
 390 For each gauge, a set of optimal celerity values are obtained by driving the routing model  
 391 with WRF-simulated runoff. The set of optimal celerity values is then fitted to a log-  
 392 normal distribution. This fitted distribution aims to capture the variations in wave celer-  
 393 ity calibration arising from different routing model inputs.

## 394 2.6 Evaluation Metrics

395 We utilized the Pearson correlation coefficient to identify the optimal wave celer-  
 396 ity value. The Pearson correlation coefficient ( $r$ ) is defined as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3)$$

397 where  $x_i$  and  $y_i$  are the simulated and observed streamflow at time step  $i$ , respectively.  
 398  $\bar{x}$  and  $\bar{y}$  are the mean of the simulated and observed streamflow, respectively. The Pear-  
 399 son correlation coefficient ranges from  $-1$  to  $1$ , with a value of  $1$  indicating a perfect pos-  
 400 itive linear relationship between the simulated and observed streamflow. This coefficient  
 401 is insensitive to biases in the streamflow estimates, making us to focus on daily scale stream-  
 402 flow variations in this study.

403 We used the Kling–Gupta efficiency (KGE) (Gupta et al., 2009) to intercompare  
 404 the WRF parameterization schemes. KGE systematically summarizes how a hydrolog-  
 405 ical simulation matches observations in correlation, standard deviation, and bias. The  
 406 KGE is defined as follows:

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}, \quad (4)$$

$$\alpha = \frac{\sigma}{\sigma_o}, \quad (5)$$

$$\beta = \frac{\mu_s}{\mu_o}, \quad (6)$$

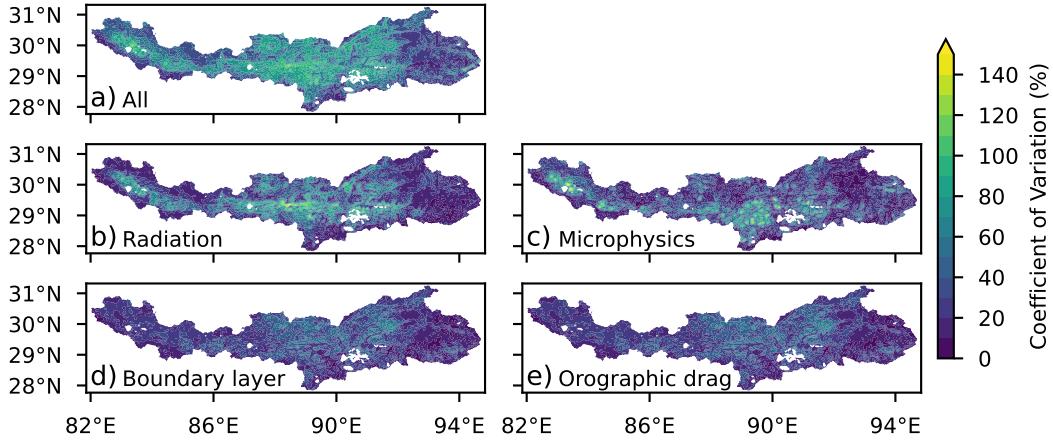
407 where  $r$  is the Pearson correlation coefficient between the simulation and observation,  
 408  $\sigma$  is the standard deviation of the simulation, and  $\mu$  is the mean. The subscripts  $s$  and  
 409  $o$  denote the simulated and observed values, respectively. The KGE ranges from  $-\infty$  to  
 410 1. A KGE value of 1 indicates a perfect match between the simulated and observed stream-  
 411 flow.

### 412 3 Results and Dicussion

413 Our analysis commences with intercomparing runoff simulated by the WRF exper-  
 414 iments. We then estimate routing parameters and their associated uncertainties. Finally,  
 415 we evaluate WRF experiment performance in streamflow estimation and assess whether  
 416 the experiments can be distinguished given uncertainty introduced by the routing pro-  
 417 cess.

418 Substantial differences in WRF-simulated runoff are evident across the thirteen ex-  
 419 periments (Figure 3). The inter-experiment spread, quantified by standard deviation,  
 420 generally exceeds the climatological mean, particularly in the middle and upper reaches  
 421 of the Yarlung Zangbo River. The runoff spread substantially exceeds that observed for  
 422 precipitation (Figure S1). Figure 3b suggests that runoff variations primarily correspond  
 423 to radiation parameterization differences, a pattern not evident for precipitation. Com-  
 424 parison between accumulated runoff and snow water equivalent (Figure S2) reveals neg-  
 425 ligible snow water equivalent relative to runoff, suggesting that runoff spread is not at-  
 426 tributable to radiation-modulated snow accumulation and ablation processes. The spread  
 427 in precipitation minus evapotranspiration (Figure S3) exhibits spatial patterns and mag-  
 428 nitudes similar to runoff, indicating that runoff spread likely originates from covariation  
 429 between precipitation and evapotranspiration. Given that radiation parameterization sig-  
 430 nificantly influences surface temperature (Liu et al., 2023; Lv et al., 2020) while precip-  
 431 itation shows marginal yet detectable responses (Hui et al., 2019), radiation emerges as  
 432 a substantial source of uncertainty in runoff simulation. Beyond radiation, parameter-  
 433 izations of cloud microphysics, planetary boundary layer, and small-scale orographic drag  
 434 also contribute to runoff variability (Figure 3c–e).

435 Despite substantial differences in simulated runoff, optimal celerity values yield-  
 436 ing the highest correlation coefficients exhibit consistency across WRF experiments at  
 437 each gauge (Figure 4). The median optimal correlation coefficient across examined gauges

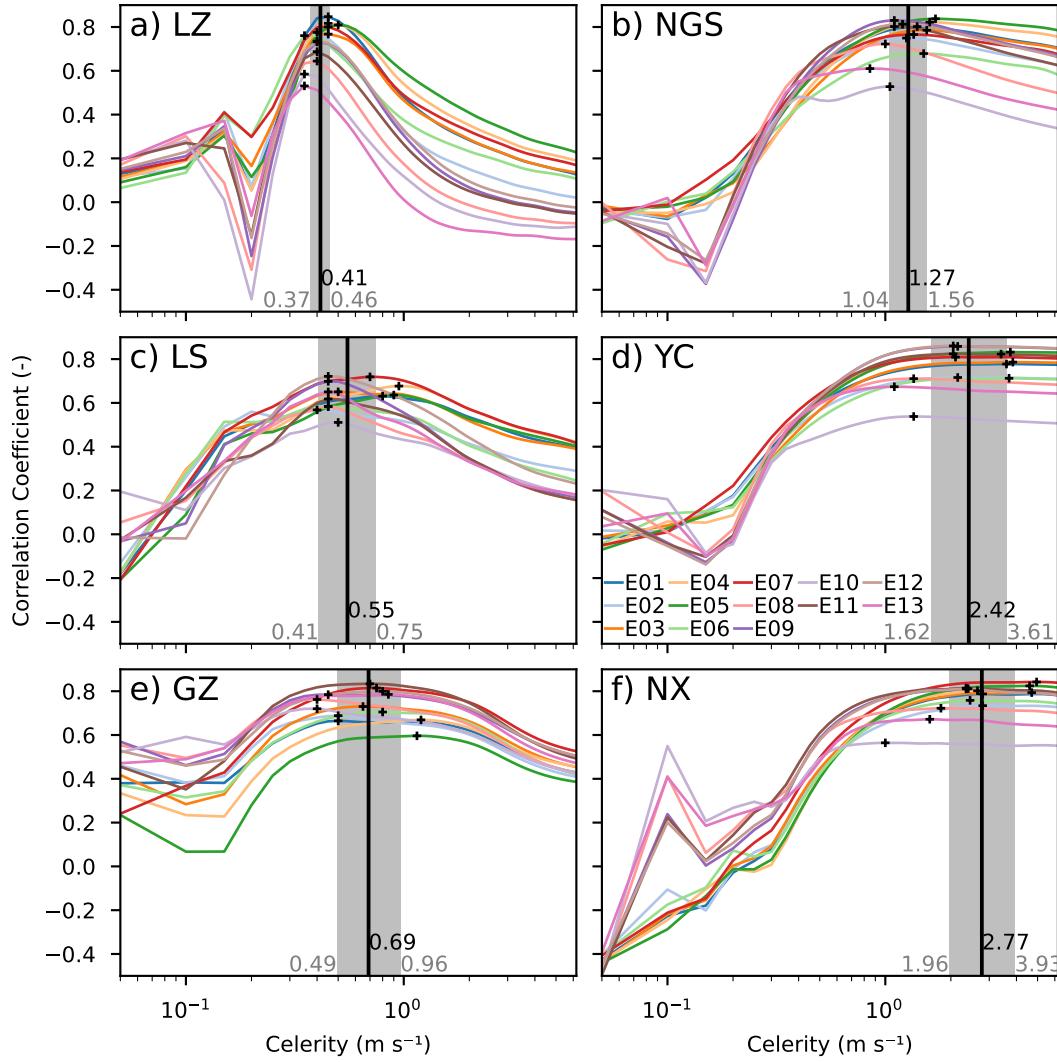


**Figure 3.** Coefficient of variation of runoff averaged from June 20 to October 1, 2013. (a) Standard deviation of runoff from 12 WRF simulations (excluding E10) divided by mean runoff. (b) Standard deviation from the radiation group (E01–E05) divided by the mean runoff from the 12 WRF experiments. (c) Standard deviation from the cloud microphysics group (E03, E06, and E07) divided by mean runoff. (d) Standard deviation from the planetary boundary layer group (E03, E08, E09, E11, and E12) divided by mean runoff. (e) Standard deviation from E08 and E13 divided by mean runoff, representing orographic drag uncertainty.

and WRF experiments is 0.75, demonstrating that the Muskingum routing model effectively captures water wave propagation dynamics. The Muskingum method is simple. As hypothesized, this simple method demonstrates robustness, rendering celerity calibration insensitive to model input variations. The consistent optimal celerity values among the WRF experiments reflect intrinsic flow path characteristics in mountainous basins. Specifically, optimal celerity is generally slower in upstream regions compared to downstream areas. As illustrated in Figure 2, downstream regions at the Tibetan Plateau margin feature broader river channels and steeper slopes, corresponding to elevated celerity values.

Optimal celerity values from WRF experiments were fitted to log-normal distributions, with expectations and standard deviations presented in Figure 4. Experiment E10 was excluded from this fitting procedure, as it consistently yielded correlation coefficients lower than other experiments (Figure 4). Given that this underperformance is statistically significant (Figure S4), the exclusion does not compromise assessment of flow routing uncertainty or the validity of the routing-based evaluation method. The log-normal

453 distributions generated 100 sets of random celerity values. Ensemble streamflow simu-  
 454 lations using these celerity value sets were conducted to estimate uncertainty associated  
 455 with the routing process.



**Figure 4.** Celerity calibration at river gauges: (a) Lazi, (b) Nugesha, (c) Lhasa, (d) Yangcun, (e) Gengzhang, and (f) Nuxia. Colored lines show correlation coefficient variations with celerity. Black plus signs indicate optimal celerity values yielding highest correlation coefficients for each WRF experiment. Optimal values are fitted to log-normal distributions. Vertical black lines and gray zones denote expectation and standard deviation of log-transformed optimal celerity values, with corresponding values indicated.

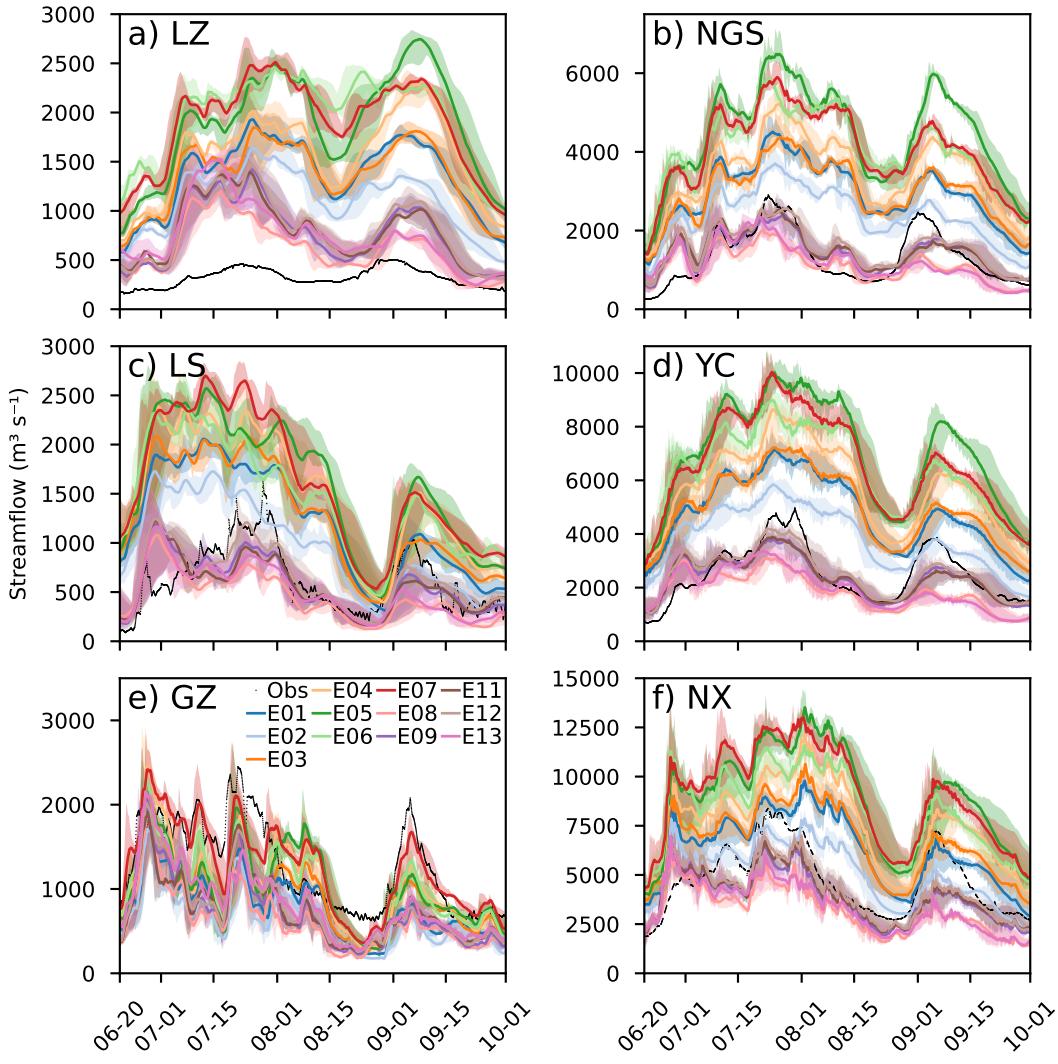
456 Figure 5 depicts ensemble means and uncertainty ranges of streamflow simulated  
 457 for each WRF experiment at the six river gauges. The WRF experiments exhibit sub-

stantial variability in streamflow estimates, reflecting pronounced differences in simulated runoff. Ensemble spread associated with the routing process is relatively small compared to inter-experiment differences. This routing uncertainty is also substantially smaller than uncertainty associated with hydrological process parameterization estimated in (Lei et al., 2025). Visual inspection indicates that simulations from E01 to E07 consistently overestimate streamflow, with overestimation magnitudes significantly exceeding the ensemble range. In contrast, experiments E08, E09, E11, and E12 produce visually similar streamflow estimates across all gauges. These four experiments more closely reproduce observed streamflow. We proceed to examine how flow routing-based evaluation distinguish different WRF experiments using quantitative performance metrics.

Figure 6 intercompares the WRF experiments by their median KGE. The most skillful experiments appear leftmost. The interquartile range of the KGE values from the ensemble simulations are displayed for comparison. The WRF experiment performance varies with location. No single experiment consistently outperforms all others at all gauges. However, when considering frequency of ranking among the top three performers across the six gauges, E11 demonstrates slight superiority, achieving this position five times. E09 follows, being the top three four times. E12 ranks third with three top-three placements. Experiment rankings exhibit minor variations depending on the performance metric employed. Figure S5 presents results using correlation coefficient as the evaluation metric. E12 emerges as the most skillful experiment with five top-three rankings, while E09 and E11 both achieve three top-three rankings.

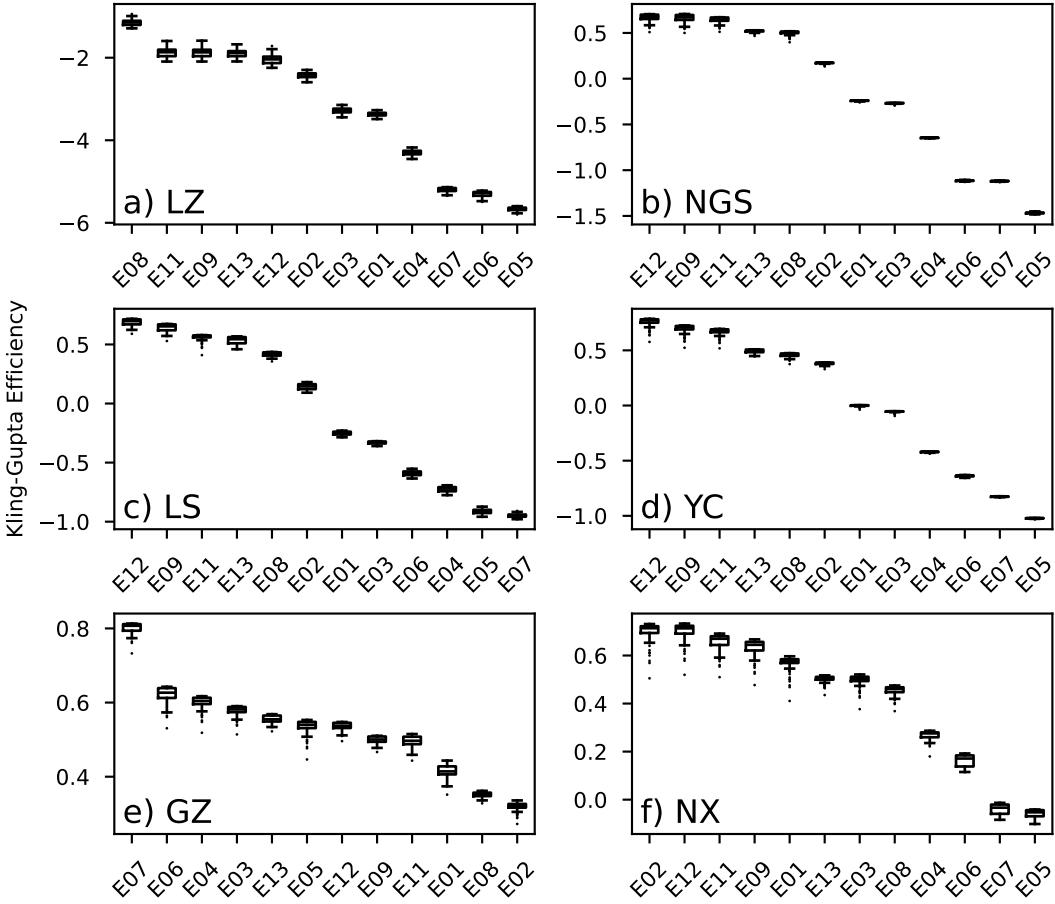
Statistical significance of performance differences between all WRF experiment pairs was assessed using ensemble KGE values at a 0.05 significance level. Figure 7 displays results for KGE, while Figure S6 presents corresponding analyses for correlation coefficient. Both analyses demonstrate that differences among WRF experiments remain statistically significant relative to routing process uncertainty. This finding indicates that the routing model exhibits sufficiently low uncertainty to distinguish performance among different WRF configurations, thereby supporting the hypothesis articulated in the Introduction.

The procedure of fitting optimal celerity values to log-normal distributions requires a number of WRF experiments, which is often computationally impractical given the substantial demands of kilometer-scale atmospheric simulations. An alternative approach



**Figure 5.** Ensemble mean and uncertainty range of simulated streamflow at (a) Lazi, (b) Nugesha, (c) Lhasa, (d) Yangcun, (e) Gengzhang, and (f) Nuxia. Solid lines represent ensemble averages for each WRF experiment, with shaded areas showing routing process uncertainty. Black dots indicate observations.

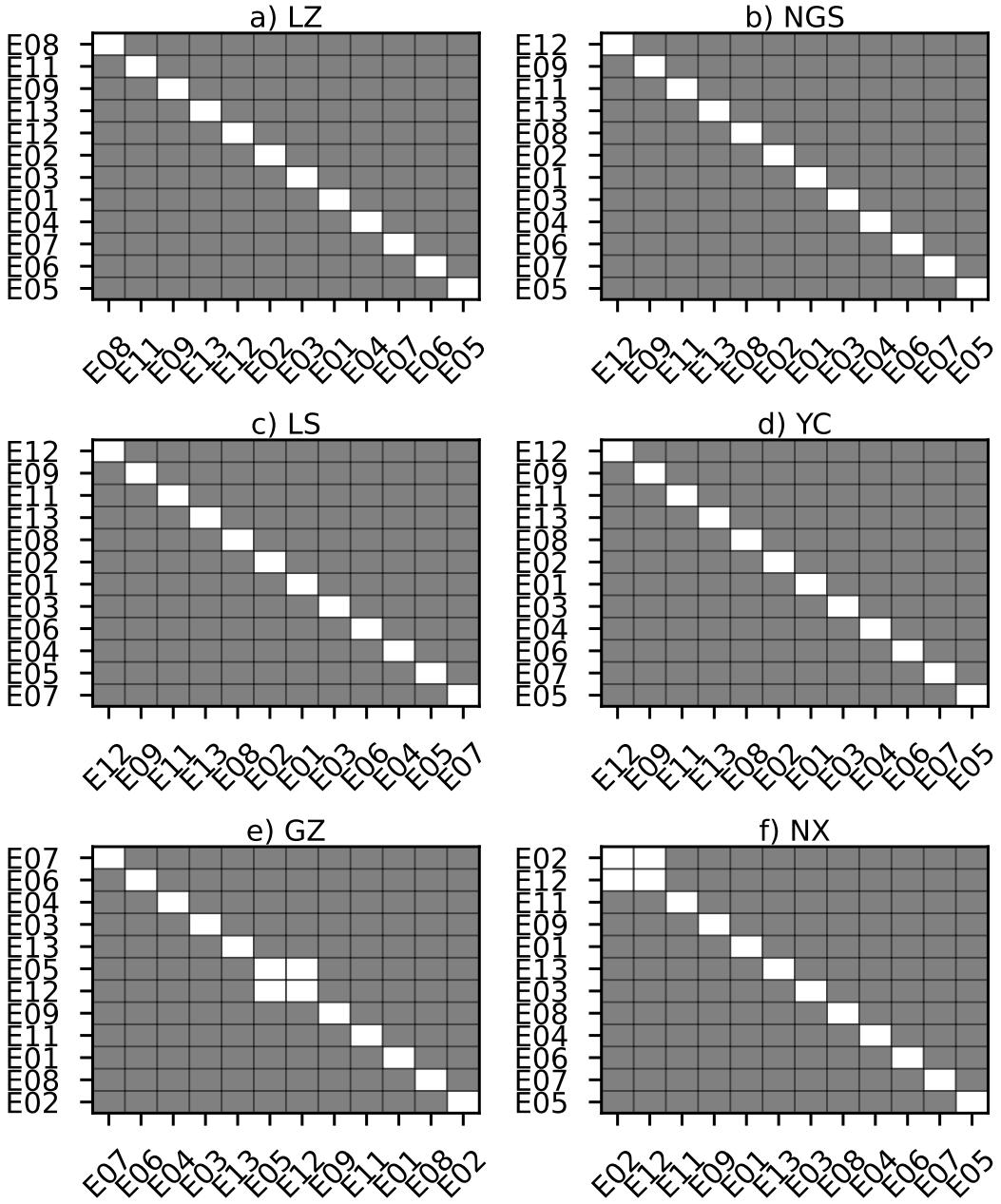
utilizes calibrated celerity values for individual WRF experiments without distribution fitting, though this simplified method does not incorporate routing process uncertainty. We examined whether this alternative approach yields rankings comparable to the uncertainty-incorporating methodology. Figure 8 compares WRF experiment rankings based on median KGE values from ensemble simulations with rankings derived from KGE values using calibrated celerity for each experiment. Table 2 presents Spearman correlation coefficients between these ranking approaches. Rankings exhibit strong consistency, with



**Figure 6.** Kling–Gupta Efficiency (KGE) at gauges: (a) Lazi, (b) Nugesha, (c) Lhasa, (d) Yangcun, (e) Gengzhang, and (f) Nuxia. Black boxes show interquartile ranges of KGE values across ensemble simulations. Horizontal lines within boxes denote median KGE values. Black dots indicate outliers. Experiments are sorted by median KGE values, with leftmost experiments being most skillful.

497 Spearman correlation coefficients exceeding 0.99 across all gauges. These findings demon-  
 498 strate that calibrated celerity values provide reliable experiment rankings even without  
 499 explicit uncertainty quantification. This consistency remains robust across various per-  
 500 formance metrics. Comparable consistency is observed for correlation coefficient rank-  
 501 ings, with Spearman correlation coefficients surpassing 0.90 at all river gauges (Figure S1  
 502 and Figure S8). This high consistency arises from routing uncertainty being relatively  
 503 minor compared to inter-experiment performance differences, as illustrated in Figure 6.

504 Figure 8 and Table 2 further compare WRF experiment rankings based on KGE  
 505 of basin-averaged precipitation and runoff with rankings derived from streamflow KGE.



**Figure 7.** Statistical significance of Kling-Gupta Efficiency (KGE) differences across WRF experiments at gauges: (a) Lazi, (b) Nugesha, (c) Lhasa, (d) Yangcun, (e) Gengzhang, and (f) Nuxia. Differences between experiment pairs are tested using paired t-tests. Gray color indicates significance at 0.05 level; white indicates non-significant differences.

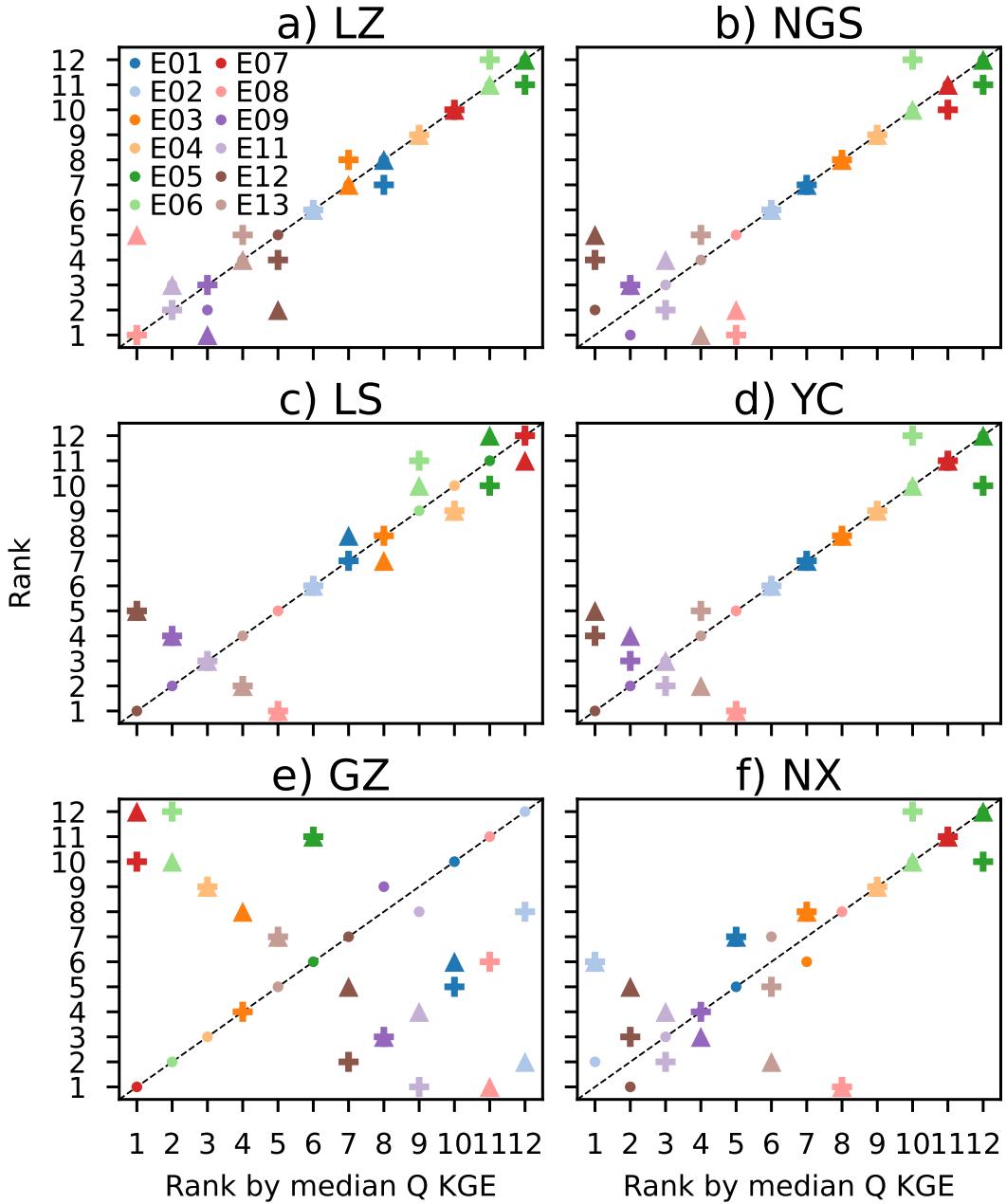
506 Basin-averaged precipitation KGE is calculated against the GPM IMERG precipitation  
 507 product, while runoff KGE is computed by directly comparing area-weighted sums of WRF-

508 simulated runoff with observed streamflow, thereby neglecting routing processes. These  
509 approaches represent commonly employed evaluation methods for atmospheric models.

510 For runoff evaluation, Table 2 hints that correspondence exists between higher rank-  
511 ings based on streamflow KGE and higher rankings based on runoff KGE. However, this  
512 correspondence lacks robustness across river gauges. Detailed analysis reveals that cor-  
513 respondence between these evaluation methods varies considerably with gauge location.  
514 At the upstream Lazi gauge, significant streamflow overestimation (Figure 5) results in  
515 bias dominating KGE contributions. High bias in streamflow correlates strongly with  
516 high bias in runoff, producing robust correspondence between runoff KGE and stream-  
517 flow KGE rankings. At Nugesha, Lhasa, Yangcun, and Nuxia, correspondence primar-  
518 ily emerges from low-performing WRF experiments (ranked beyond position 7 in Fig-  
519 ure 8), which exhibit substantial biases. For high-performing experiments, correspon-  
520 dence remains weak. At Gengzhang, all WRF experiments demonstrate similar bias mag-  
521 nitudes (Figure 5), resulting in consistently weak correspondence between runoff KGE  
522 and streamflow KGE rankings. If the bias is excluded from the skill measure, the cor-  
523 respondence is consistently weak, as shown for correlation coefficient in Figure S7. Crit-  
524 ically, visual comparison between runoff and streamflow time series indicates that ne-  
525 glecting routing processes is problematic at hourly time scales (Figure S8), as they ex-  
526 hibit substantial differences. Based on these findings, we conclude that runoff-based eval-  
527 uation methods neglecting routing processes lack reliability for atmospheric model eval-  
528 uation at hourly time scales.

529 For precipitation evaluation, correspondence with streamflow KGE-based rankings  
530 generally mirrors the pattern observed for runoff. We contend that this similarity stems  
531 from the fundamental relationship where higher precipitation typically generates higher  
532 runoff under consistent land surface conditions. These findings suggest that when atmo-  
533 spheric models exhibit substantial precipitation biases, basin-averaged precipitation eval-  
534 uation can provide indicative insights for streamflow assessment. However, when atmo-  
535 spheric models demonstrate sound precipitation simulation in basin-averaged temporal  
536 means, flow routing-based evaluation becomes essential for hydrologically relevant as-  
537 sessment. The two evaluation approaches may yield divergent performance rankings of  
538 atmospheric models. This discrepancy reflects the nonlinear response of streamflow to  
539 the spatiotemporal distribution of precipitation,(Stephens et al., 2015) and the influence  
540 of non-precipitation variables (e.g., the contrasting variability of runoff and precipita-

541 tion in Figure 3 and Figure S1). Consequently, precipitation- and routing-based eval-  
 542 uations are complementary rather than redundant.



**Figure 8.** Relationship between the rank of the median Kling–Gupta Efficiency (KGE) measured by streamflow and the rank measured by other observations across the WRF experiments at the gauge of (a) Lazi, (b) Nugesha, (c) Lhasa, (d) Yangcun, (e) Gengzhang, and (f) Nuxia. Dashed lines represent 1:1 lines indicating perfect agreement. Dots show optimal KGE ranks, plus signs indicate runoff-based ranks, and triangles denote basin-averaged precipitation ranks.

**Table 2.** Spearman's correlation coefficient between the rank of the median Kling–Gupta Efficiency measured by streamflow and the rank measured by other observations or skill measures across the WRF experiments.

Gauge	Calibration	Runoff	Precipitation
Lazi	0.99	0.98	0.90
Nugesha	0.99	0.88	0.87
Lhasa	1.00	0.84	0.84
Yangcun	1.00	0.87	0.86
Gengzhang	0.99	-0.49	-0.87
Nuxia	0.99	0.69	0.63

## 4 Conclusions

This study develops and validates a routing network-based methodology for evaluating atmospheric models using streamflow observations in data-sparse mountainous regions. The proposed approach distinguishes itself from conventional hydrological model-based evaluations by requiring substantially fewer assumptions regarding model structures and parameters. We hypothesize that by propagating runoff through a parsimonious network routing model, the additional uncertainty is kept small enough to discriminate among atmospheric model configurations.

The application of this methodology to thirteen WRF experiments in the Yarlung Zangbo River basin with varying parameterization schemes demonstrates its effectiveness in distinguishing atmospheric model performance. The Muskingum routing model proves robust and effective, with optimal celerity values exhibiting consistency across different WRF configurations and producing median correlation coefficients of 0.75 across gauges for hourly streamflow. Statistical significance testing confirms that routing model uncertainty remains sufficiently low to distinguish performance among different WRF configurations, supporting our central hypothesis.

The routing-based evaluation provides complementary insights to precipitation assessment using the GPM IMERG product. Specifically, when atmospheric models demonstrate sound precipitation simulation in basin-averaged temporal means, routing-based

562 evaluation becomes essential for hydrologically relevant atmospheric model assessment.  
563 This approach captures aspects of model performance that precipitation-based evalua-  
564 tion alone may overlook.

565 The methodology's practical implementation is facilitated by the finding that cal-  
566 ibrated celerity values provide reliable experiment rankings even without explicit uncer-  
567 tainty quantification, with Spearman correlation coefficients exceeding 0.99 between uncertainty-  
568 incorporating and simplified approaches. This computational efficiency makes the method  
569 particularly valuable given the substantial demands of kilometer-scale atmospheric sim-  
570 ulations.

### 571 **Conflict of Interest**

572 The authors declare no conflicts of interest relevant to this study.

### 573 **Open Research Section**

574 The GPM IMERG final run precipitation product version 7 was used to evaluate  
575 the WRF-simulated precipitation (Huffman et al., 2023). The ERA5 reanalysis data (Hersbach  
576 et al., 2020) were used to drive the WRF model version 4.3.3 (Powers et al., 2017) to sim-  
577 ulate precipitation and runoff. The MERIT-Hydro flow direction and cumulative upstream  
578 area data (Yamazaki et al., 2019) were used to delineate the routing network. The de-  
579 lineated routing network, WRF-simulated precipitation and runoff, and the code for rout-  
580 ing and parameter optimization are available at [https://doi.org/10.57760/sciencedb  
.11618](https://doi.org/10.57760/sciencedb.11618) (Zheng et al., 2024). The streamflow observations for the Yarlung Zangbo River  
582 were obtained from China Three Gorges Corporation and Ministry of Water Resources  
583 of China; however, they are not shareable due to licensing restrictions.

### 584 **Acknowledgments**

585 This study is supported by the National Key Research and Development Program  
586 of China (2023YFF0805501), China Yangtze Power Co., Ltd. (contract Z532302035), and  
587 the Natural Science Foundation of China (grants 42075165 and 42275178).

### 588 **References**

589 Adam, J. C., Clark, E. A., Lettenmaier, D. P., & Wood, E. F. (2006). Correction of

- 590 global precipitation products for orographic effects. *Journal of Climate*, 19(1),  
591 15–38. doi: 10.1175/JCLI3604.1
- 592 Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., &  
593 Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of  
Hydrology X*, 6, 100049. doi: 10.1016/j.hydroa.2019.100049
- 594 Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., & Famiglietti, J. S.  
595 (2018). Global estimates of river flow wave travel times and implications  
596 for low-latency satellite data. *Geophysical Research Letters*, 45(15), 7551–7560.  
597 doi: 10.1029/2018GL077914
- 598 Barkhordari, H., Asgari Dastjerdi, P., & Nasseri, M. (2025). Development of a  
599 framework estimating regional gridded streamflow and actual evapotranspi-  
600 ration datasets: Fusing Budyko and water balance closure methods using  
601 remotely sensed ancillary data. *Journal of Hydrology*, 660, 133456. doi:  
602 10.1016/j.jhydrol.2025.133456
- 603 Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical  
604 weather prediction. *Nature*, 525(7567), 47–55. doi: 10.1038/nature14956
- 605 Behrangi, A., Andreadis, K., Fisher, J. B., Turk, F. J., Granger, S., Painter, T.,  
606 & Das, N. (2014). Satellite-based precipitation estimation and its ap-  
607 plication for streamflow prediction over mountainous western U.S. basins.  
608 *Journal of Applied Meteorology and Climatology*, 53(12), 2823–2842. doi:  
609 10.1175/JAMC-D-14-0056.1
- 610 Beljaars, A. C. M., Brown, A. R., & Wood, N. (2004). A new parametrization of  
611 turbulent orographic form drag. *Quarterly Journal of the Royal Meteorological  
612 Society*, 130(599), 1327–1347. doi: 10.1256/qj.03.73
- 613 Beven, K. J. (2006). Searching for the Holy Grail of scientific hydrology:  $Qt = H(S,$   
614  $R, \delta t)A$  as closure. *Hydrology and Earth System Sciences*, 10(5), 609–618. doi:  
615 10.5194/hess-10-609-2006
- 616 Bian, Q., Xu, Z., Zhao, L., Zhang, Y.-F., Zheng, H., Shi, C., ... Yang, Z.-L. (2019).  
617 Evaluation and intercomparison of multiple snow water equivalent products  
618 over the tibetan plateau. *Journal of Hydrometeorology*, 20(10), 2043–2055. doi:  
619 10.1175/JHM-D-19-0011.1
- 620 Chen, S., Yang, H., & Zheng, H. (2025). Intercomparison of runoff and river  
621 discharge reanalysis datasets at the Upper Jinsha River, an alpine river

- 623       on the eastern edge of the Tibetan Plateau.             *Water*, 17(6), 871.       doi:  
624       10.3390/w17060871
- 625       Chen, S.-H., & Sun, W.-Y. (2002). A one-dimensional time dependent cloud model.  
626       *Journal of the Meteorological Society of Japan. Ser. II*, 80(1), 99–118. doi: 10  
627       .2151/jmsj.80.99
- 628       Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multi-  
629       ple working hypotheses for hydrological modeling. *Water Resources Research*,  
630       47(9), 1–16. doi: 10.1029/2010WR009827
- 631       Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson,  
632       B. M., ... Ceola, S. (2016). Improving the theoretical underpinnings of  
633       process-based hydrologic models. *Water Resources Research*, 52(3), 2350–2365.  
634       doi: 10.1002/2015WR017910
- 635       Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016).  
636       Convection-permitting models: a step-change in rainfall forecasting. *Meteo-*  
637       *logical Applications*, 23(2), 165–181. doi: 10.1002/met.1538
- 638       Collier, E., Ban, N., Richter, N., Ahrens, B., Chen, D., Chen, X., ... Ziska, C.  
639       (2024). The first ensemble of kilometer-scale simulations of a hydrologi-  
640       cal year over the third pole. *Climate Dynamics*, 62(8), 7501–7518. doi:  
641       10.1007/s00382-024-07291-2
- 642       Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., ...  
643       Zhang, Y. (2024). NOAA's National Water Model: Advancing operational  
644       hydrology through continental-scale modeling. *Journal of the American Water  
645       Resources Association*, 60(2), 247–272. doi: 10.1111/1752-1688.13184
- 646       Cunge, J. A. (1969). On the subject of a flood propagation computation method  
647       (Muskingum method). *Journal of Hydraulic Research*, 7(2), 205–230. doi: 10  
648       .1080/00221686909500264
- 649       David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., & Eikhout, V.  
650       (2011). River network routing on the NHDPlus dataset. *Journal of Hydrome-  
651       teorology*, 12(5), 913–934. doi: 10.1175/2011JHM1345.1
- 652       Denis, B., Laprise, R., Caya, D., & Côté, J. (2002). Downscaling ability of one-way  
653       nested regional climate models: the Big-Brother Experiment. *Climate Dynam-  
654       ics*, 18(8), 627–646. doi: 10.1007/s00382-001-0201-0
- 655       Dougherty, E., Sherman, E., & Rasmussen, K. L. (2020). Future changes in the hy-

- 656 drologic cycle associated with flood-producing storms in California. *Journal of*  
657 *Hydrometeorology*, 21(11), 2607–2621. doi: 10.1175/JHM-D-20-0067.1
- 658 Dudhia, J. (1989). Numerical study of convection observed during the winter  
659 monsoon experiment using a mesoscale two-dimensional model. *Journal of*  
660 *the Atmospheric Sciences*, 46(20), 3077–3107. doi: 10.1175/1520-0469(1989)  
661 046<3077:NSOCOD>2.0.CO;2
- 662 Fenton, J. D. (2019). Flood routing methods. *Journal of Hydrology*, 570, 251–264.  
663 doi: 10.1016/j.jhydrol.2019.01.006
- 664 Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change mod-  
665 elling to impacts studies: recent advances in downscaling techniques for hydro-  
666 logical modelling. *International Journal of Climatology*, 27(12), 1547–1578.  
667 doi: 10.1002/joc.1556
- 668 François, B., Schlef, K., Wi, S., & Brown, C. (2019). Design considerations for river-  
669 ine floods in a changing climate – A review. *Journal of Hydrology*, 574, 557–  
670 573. doi: 10.1016/j.jhydrol.2019.04.068
- 671 Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S.,  
672 ... Williams, D. N. (1999). An overview of the results of the Atmospheric  
673 Model Intercomparison Project (AMIP I). *Bulletin of the American Me-  
674 teorological Society*, 80(1), 29–56. doi: 10.1175/1520-0477(1999)080<0029:  
675 AOOTRO>2.0.CO;2
- 676 Getirana, A. C., & Paiva, R. C. (2013). Mapping large-scale river flow hydraulics  
677 in the Amazon Basin. *Water Resources Research*, 49(5), 2437–2445. doi: 10  
678 .1002/wrcr.20212
- 679 Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs  
680 at the regional level: the CORDEX framework. *World Meteorological Organi-  
681 zation Bulletin*, 58(3), 175–183.
- 682 Givati, A., Gochis, D., Rummler, T., & Kunstmann, H. (2016). Comparing  
683 one-way and two-way coupled hydrometeorological forecasting systems for  
684 flood forecasting in the Mediterranean region. *Hydrology*, 3(2), 19. doi:  
685 10.3390/hydrology3020019
- 686 Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition  
687 of the mean squared error and NSE performance criteria: Implications for im-  
688 proving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80–91. doi:

- 689 10.1016/j.jhydrol.2009.08.003
- 690 Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., ...  
691 Pappenberger, F. (2020). GloFAS-ERA5 operational global river discharge  
692 reanalysis 1979–present. *Earth System Science Data*, 12(3), 2043–2060. doi:  
693 10.5194/essd-12-2043-2020
- 694 Henn, B., Clark, M. P., Kavetski, D., & Lundquist, J. D. (2015). Estimating moun-  
695 tain basin-mean precipitation from streamflow using Bayesian inference. *Water*  
696 *Resources Research*, 51(10), 8012–8033. doi: 10.1002/2014WR016736
- 697 Henn, B., Clark, M. P., Kavetski, D., McGurk, B., Painter, T. H., & Lundquist,  
698 J. D. (2016). Combining snow, streamflow, and precipitation gauge observa-  
699 tions to infer basin-mean precipitation. *Water Resources Research*, 52(11),  
700 8700–8723. doi: 10.1002/2015WR018564
- 701 Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2018). An  
702 assessment of differences in gridded precipitation datasets in complex terrain.  
703 *Journal of Hydrology*, 556, 1205–1219. doi: 10.1016/j.jhydrol.2017.03.008
- 704 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,  
705 ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of*  
706 *the Royal Meteorological Society*, 146(730), 1999–2049. doi: 10.1002/qj.3803
- 707 Hong, S.-Y., & Jang, J. (2018). Impacts of shallow convection processes on  
708 a simulated boreal summer climatology in a global atmospheric model.  
709 *Asia-Pacific Journal of Atmospheric Sciences*, 54(1), 361–370. doi:  
710 10.1007/s13143-018-0013-3
- 711 Hong, S.-Y., & Lim, J.-O. J. (2006). The WRF Single-Moment 6-Class Microphysics  
712 Scheme (WSM6). *Asia-Pacific Journal of Atmospheric Sciences*, 42(2), 129–  
713 151.
- 714 Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an  
715 explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9),  
716 2318–2341. doi: 10.1175/MWR3199.1
- 717 Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2023).  
718 *GPM IMERG final precipitation L3 half hourly 0.1 degree x 0.1 degree V07*  
719 (*GPM\_3IMERGHH*). NASA/GSFC/SED/ESD/GCDC/GESDISC. Retrieved  
720 2025-09-21, from <https://doi.org/10.5067/GPM/IMERG/3B-HH/07> doi:  
721 10.5067/GPM/IMERG/3B-HH/07

- 722 Hui, P., Li, Y., Chen, Y., Zhang, L., Wei, F., Wang, S., & Tang, J. (2019). The  
 723 impact of radiation parameterization schemes on the regional climate simula-  
 724 tions over the CORDEX-EA domain. *Atmospheric Research*, 224, 81–98. doi:  
 725 10.1016/j.atmosres.2019.03.020
- 726 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., &  
 727 Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Cal-  
 728 culations with the AER radiative transfer models. *Journal of Geophysical  
 729 Research: Atmospheres*, 113(D13), D13103. doi: 10.1029/2008JD009944
- 730 Janjić, Z. I. (1994). The step-mountain Eta coordinate model: further developments  
 731 of the convection, viscous sublayer, and turbulence closure schemes. *Monthly  
 732 Weather Review*, 122(5), 927–945. doi: 10.1175/1520-0493(1994)122<0927:  
 733 TSMECM>2.0.CO;2
- 734 Jiang, Y., Yang, K., Li, X., Zhang, W., Shen, Y., Chen, Y., & Li, X. (2022).  
 735 Atmospheric simulation-based precipitation datasets outperform satellite-  
 736 based products in closing basin-wide water budget in the eastern Tibetan  
 737 Plateau. *International Journal of Climatology*, 42(14), 7252–7268. doi:  
 738 10.1002/joc.7642
- 739 Jiang, Y., Yang, K., Yang, H., Lu, H., Chen, Y., Zhou, X., ... Wang, Y. (2022).  
 740 Characterizing basin-scale precipitation gradients in the Third Pole region  
 741 using a high-resolution atmospheric simulation-based dataset. *Hydrology and  
 742 Earth System Sciences*, 26(17), 4587–4601. doi: 10.5194/hess-26-4587-2022
- 743 Karki, R., ul Hasson, S., Gerlitz, L., Schickhoff, U., Scholten, T., & Böhner, J.  
 744 (2017). Quantifying the added value of convection-permitting climate simula-  
 745 tions in complex terrain: a systematic evaluation of WRF over the Himalayas.  
 746 *Earth System Dynamics*, 8(3), 507–528. doi: 10.5194/esd-8-507-2017
- 747 Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier,  
 748 N., ... Wood, A. W. (2025). One-at-a-time parameter perturbation ensemble  
 749 of the Community Land Model, version 5.1. *Journal of Advances in Modeling  
 750 Earth Systems*, 17, e2024MS004715. doi: 10.1029/2024MS004715
- 751 Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment char-  
 752 acterization, rainfall-runoff modeling, and doing hydrology backward. *Water  
 753 Resources Research*, 45(2), W02429. doi: 10.1029/2008WR006912
- 754 Koussis, A. D. (1978). Theoretical estimation of flood routing parameters. *Journal*

- 755                  *of the Hydraulics Division*, 104(1), 109–115. doi: 10.1061/JYCEAJ.0004909
- 756   Krier, R., Matgen, P., Goergen, K., Pfister, L., Hoffmann, L., Kirchner, J. W., ...  
757                  Savenije, H. H. G. (2012). Inferring catchment precipitation by doing hydrology backward: A test in 24 small and mesoscale catchments in Luxembourg.  
758                  *Water Resources Research*, 48(10), W10525. doi: 10.1029/2011WR010657
- 760   Kukulies, J., Prein, A. F., Curio, J., Yu, H., & Chen, D. (2023). Kilometer-scale  
761                  multimodel and multiphysics ensemble simulations of a mesoscale convective  
762                  system in the lee of the Tibetan Plateau: implications for climate simulations.  
763                  *Journal of Climate*, 36(17), 5963–5987. doi: 10.1175/JCLI-D-22-0240.1
- 764   Lei, X., Lin, P., Zheng, H., Fei, W., Yin, Z., & Ren, H. (2025). Systematic analyses  
765                  of the meteorological forcing and process parameterization uncertainties  
766                  in modeling runoff with Noah-MP for the Upper Brahmaputra river basin.  
767                  *Journal of Hydrology*, 653, 132686. doi: 10.1016/j.jhydrol.2025.132686
- 768   Li, G., Chen, H., Xu, M., Zhao, C., Zhong, L., Li, R., ... Gao, Y. (2022). Impacts of  
769                  topographic complexity on modeling moisture transport and precipitation over  
770                  the Tibetan Plateau in summer. *Advances in Atmospheric Sciences*, 39(7),  
771                  1151–1166. doi: 10.1007/s00376-022-1409-7
- 772   Li, L., Bisht, G., Hao, D., & Leung, L. R. (2024). Global 1 km land surface parameters  
773                  for kilometer-scale Earth system modeling. *Earth System Science Data*,  
774                  16(4), 2007–2032. doi: 10.5194/essd-16-2007-2024
- 775   Liang, J., Yang, Z.-L., & Lin, P. (2019). Systematic hydrological evaluation of the  
776                  Noah-MP land surface model over China. *Advances in Atmospheric Sciences*,  
777                  36(11), 1171–1187. doi: 10.1007/s00376-019-9016-y
- 778   Lin, C., Chen, D., Yang, K., & Ou, T. (2018). Impact of model resolution on sim-  
779                  ulating the water vapor transport through the central Himalayas: implication  
780                  for models' wet bias over the Tibetan Plateau. *Climate Dynamics*, 51(9),  
781                  3195–3207. doi: 10.1007/s00382-018-4074-x
- 782   Lin, P., Hopper, L. J., Yang, Z.-L., Lenz, M., & Zeitler, J. W. (2018). Insights into  
783                  hydrometeorological factors constraining flood prediction skill during the May  
784                  and October 2015 Texas Hill Country flood events. *Journal of Hydrometeorol-  
785                  ogy*, 19(8), 1339–1361. doi: 10.1175/JHM-D-18-0038.1
- 786   Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., ... Wood, E. F.  
787                  (2019). Global reconstruction of naturalized river flows at 2.94 million reaches.

- 788        *Water Resources Research*, 55(8), 6499–6516. doi: 10.1029/2019WR025287
- 789     Lin, P., Pan, M., Wood, E. F., Yamazaki, D., & Allen, G. H. (2021). A new vector-  
790     based global river network dataset accounting for variable drainage density.  
791     *Scientific Data*, 8(1), 28. doi: 10.1038/s41597-021-00819-9
- 792     Lin, P., Yang, Z.-L., Gochis, D. J., Yu, W., Maidment, D. R., Somos-Valenzuela,  
793     M. A., & David, C. H. (2018). Implementation of a vector-based river network  
794     routing scheme in the community WRF-Hydro modeling framework for flood  
795     discharge simulation. *Environmental Modelling & Software*, 107, 1–11. doi:  
796     10.1016/j.envsoft.2018.05.018
- 797     Liu, L., Zhang, X., Han, C., & Ma, Y. (2023). A performance evaluation of various  
798     physics schemes on the predictions of precipitation and temperature over the  
799     Tibet Autonomous Region of China. *Atmospheric Research*, 292, 106878. doi:  
800     10.1016/j.atmosres.2023.106878
- 801     Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., ...  
802     Caillaud, C. (2021). Convection-permitting modeling with regional climate  
803     models: Latest developments and next steps. *WIREs Climate Change*, 12(6),  
804     e731. doi: 10.1002/wcc.731
- 805     Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our skill in mod-  
806     eling mountain rain and snow is bypassing the skill of our observational net-  
807     works. *Bulletin of the American Meteorological Society*, 100(12), 2473–2490.  
808     doi: 10.1175/bams-d-19-0001.1
- 809     Lv, M., Xu, Z., & Yang, Z. (2020). Cloud resolving WRF simulations of precipi-  
810     tation and soil moisture over the central Tibetan Plateau: an assessment of  
811     various physics options. *Earth and Space Science*, 7(2), e2019EA000865. doi:  
812     10.1029/2019EA000865
- 813     Ma, M., Ou, T., Liu, D., Wang, S., Fang, J., & Tang, J. (2023). Summer re-  
814     gional climate simulations over Tibetan Plateau: from gray zone to con-  
815     vection permitting scale. *Climate Dynamics*, 60(1), 301–322. doi:  
816     10.1007/s00382-022-06314-0
- 817     Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Wid-  
818     mann, M., ... Thiele-Eich, I. (2010). Precipitation downscaling under  
819     climate change: Recent developments to bridge the gap between dynami-  
820     cal models and the end user. *Reviews of Geophysics*, 48(3), RG3003. doi:

- 821 10.1029/2009RG000314
- 822 Matsui, T., Zhang, S. Q., Lang, S. E., Tao, W.-K., Ichoku, C., & Peters-Lidard,  
823 C. D. (2020). Impact of radiation frequency, precipitation radiative  
824 forcing, and radiation column aggregation on convection-permitting West  
825 African monsoon simulations. *Climate Dynamics*, 55(1), 193–213. doi:  
826 10.1007/s00382-018-4187-2
- 827 Miao, C., Immerzeel, W. W., Xu, B., Yang, K., Duan, Q., & Li, X. (2024). Un-  
828 derstanding the Asian water tower requires a redesigned precipitation obser-  
829 vation strategy. *Proceedings of the National Academy of Sciences*, 121(23),  
830 e2403557121. doi: 10.1073/pnas.2403557121
- 831 Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997).  
832 Radiative transfer for inhomogeneous atmospheres: RRTM, a validated  
833 correlated-k model for the longwave. *Journal of Geophysical Research: At-*  
834 *mospheres*, 102(D14), 16663–16682. doi: 10.1029/97JD00237
- 835 Mooney, P. A., Broderick, C., Bruyère, C. L., Mulligan, F. J., & Prein, A. F. (2017).  
836 Clustering of observed diurnal cycles of precipitation over the United States  
837 for evaluation of a WRF multiphysics regional climate ensemble. *Journal of*  
838 *Climate*, 30(22), 9267–9286. doi: 10.1175/JCLI-D-16-0851.1
- 839 Moussa, R., & Bocquillon, C. (1996). Criteria for the choice of flood-routing meth-  
840 ods in natural channels. *Journal of Hydrology*, 186(1-4), 1–30. doi: 10.1016/  
841 S0022-1694(96)03045-4
- 842 Najafi, H., Shrestha, P. K., Rakovec, O., Apel, H., Vorogushyn, S., Kumar, R.,  
843 ... Samaniego, L. (2024). High-resolution impact-based early warning  
844 system for riverine flooding. *Nature Communications*, 15(1), 3726. doi:  
845 10.1038/s41467-024-48065-y
- 846 Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3  
847 model: Its numerical stability and application to a regional prediction  
848 of advection fog. *Boundary-Layer Meteorology*, 119(2), 397–407. doi:  
849 10.1007/s10546-005-9030-8
- 850 Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure  
851 model for the atmospheric boundary layer. *Journal of the Meteorological Soci-  
852 ety of Japan. Ser. II*, 87(5), 895–912. doi: 10.2151/jmsj.87.895
- 853 Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., ... Xia,

- 854 Y. (2011). The community Noah land surface model with multiparameteriza-  
855 tion options (Noah-MP): 1. Model description and evaluation with local-scale  
856 measurements. *Journal of Geophysical Research: Atmospheres*, 116(D12),  
857 D12109. doi: 10.1029/2010JD015139
- 858 Pang, J., Zhang, H., Xu, Q., Wang, Y., Wang, Y., Zhang, O., & Hao, J. (2020).  
859 Hydrological evaluation of open-access precipitation data using SWAT at mul-  
860 tiple temporal and spatial scales. *Hydrology and Earth System Sciences*, 24(7),  
861 3603–3626. doi: 10.5194/hess-24-3603-2020
- 862 Parsons, D. B., Beland, M., Burridge, D., Bougeault, P., Brunet, G., Caughey,  
863 J., ... Toth, Z. (2017). THORPEX research and the science of predic-  
864 tion. *Bulletin of the American Meteorological Society*, 98(4), 807–830. doi:  
865 10.1175/BAMS-D-14-00025.1
- 866 Pleim, J. E. (2007a). A combined local and nonlocal closure model for the atmo-  
867 spheric boundary layer. Part II: Application and evaluation in a mesoscale  
868 meteorological model. *Journal of Applied Meteorology and Climatology*, 46(9),  
869 1396–1409. doi: 10.1175/JAM2534.1
- 870 Pleim, J. E. (2007b). A combined local and nonlocal closure model for the atmo-  
871 spheric boundary layer. Part I: Model description and testing. *Journal of Ap-  
872 plied Meteorology and Climatology*, 46(9), 1383–1395. doi: 10.1175/JAM2539  
873 .1
- 874 Ponce, V. M., Simons, D. B., & Li, R.-M. (1978). Applicability of kinematic and dif-  
875 fusion models. *Journal of the Hydraulics Division*, 104(3), 353–360. doi: 10  
876 .1061/JYCEAJ.0004958
- 877 Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill,  
878 D. O., ... Duda, M. G. (2017). The Weather Research and Forecasting model:  
879 overview, system efforts, and future directions. *Bulletin of the American Mete-  
880 orological Society*, 98(8), 1717–1737. doi: 10.1175/bams-d-15-00308.1
- 881 Prein, A. F., Ban, N., Ou, T., Tang, J., Sakaguchi, K., Collier, E., ... Chen,  
882 D. (2023). Towards ensemble-based kilometer-scale climate simulations  
883 over the Third Pole region. *Climate Dynamics*, 60(11), 4055–4081. doi:  
884 10.1007/s00382-022-06543-3
- 885 Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., ... Le-  
886 ung, R. (2015). A review on regional convection-permitting climate modeling:

- 887 Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2),  
888 323–361. doi: 10.1002/2014RG000475
- 889 RafieeiNasab, A., Fienen, M. N., Omani, N., Srivastava, I., & Dugger, A. L. (2025).  
890 Ensemble methods for parameter estimation of WRF-hydro. *Water Resources  
891 Research*, 61(1), e2024WR038048. doi: 10.1029/2024WR038048
- 892 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Un-  
893 derstanding predictive uncertainty in hydrologic modeling: The challenge of  
894 identifying input and structural errors. *Water Resources Research*, 46(5),  
895 W05521. doi: 10.1029/2009WR008328
- 896 Reszler, C., Switanek, M. B., & Truhetz, H. (2018). Convection-permitting re-  
897 gional climate simulations for representing floods in small- and medium-sized  
898 catchments in the Eastern Alps. *Natural Hazards and Earth System Sciences*,  
899 18(10), 2653–2674. doi: 10.5194/nhess-18-2653-2018
- 900 Rudisill, W., Flores, A., & Carroll, R. (2023). Evaluating 3 decades of precip-  
901 itation in the Upper Colorado River basin from a high-resolution regional  
902 climate model. *Geoscientific Model Development*, 16(22), 6531–6552. doi:  
903 10.5194/gmd-16-6531-2023
- 904 Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., & Chen, F. (1996). Simple  
905 water balance model for estimating runoff at different spatial and temporal  
906 scales. *Journal of Geophysical Research: Atmospheres*, 101(D3), 7461–7475.  
907 doi: 10.1029/95JD02892
- 908 Stephens, E., Day, J. J., Pappenberger, F., & Cloke, H. (2015). Precipitation and  
909 floodiness. *Geophysical Research Letters*, 42(23), 10316–10323. doi: 10.1002/  
910 2015GL066779
- 911 Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ...  
912 Zhou, L. (2019). DYAMOND: the DYnamics of the Atmospheric general circu-  
913 lation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary  
914 Science*, 6(1), 61. doi: 10.1186/s40645-019-0304-z
- 915 Sugimoto, S., Ueno, K., Fujinami, H., Nasuno, T., Sato, T., & Takahashi, H. G.  
916 (2021). Cloud-resolving-model simulations of nocturnal precipitation over  
917 the Himalayan slopes and foothills. *Journal of Hydrometeorology*, 22(12),  
918 3171–3188. doi: 10.1175/JHM-D-21-0103.1
- 919 Sukoriansky, S., Galperin, B., & Perov, V. (2005). Application of a new spec-

- 920 tral theory of stably stratified turbulence to the atmospheric boundary  
921 layer over sea ice. *Boundary-Layer Meteorology*, 117(2), 231–257. doi:  
922 10.1007/s10546-004-6848-4
- 923 Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson,  
924 T. D., … Yamaguchi, M. (2016). The TIGGE project and its achieve-  
925 ments. *Bulletin of the American Meteorological Society*, 97(1), 49–67. doi:  
926 10.1175/BAMS-D-13-00191.1
- 927 Tan, X., Liu, B., Tan, X., & Chen, X. (2022). Long-term water imbalances of  
928 watersheds resulting from biases in hydroclimatic data sets for water bud-  
929 get analyses. *Water Resources Research*, 58(3), e2021WR031209. doi:  
930 10.1029/2021WR031209
- 931 Tang, Q., Golaz, J.-C., Van Roekel, L. P., Taylor, M. A., Lin, W., Hillman, B. R.,  
932 … Bader, D. C. (2023). The fully coupled regionally refined model of E3SM  
933 version 2: overview of the atmosphere, land, and river results. *Geoscientific  
934 Model Development*, 16(13), 3953–3995. doi: 10.5194/gmd-16-3953-2023
- 935 Thielen, J., Bartholmes, J., Ramos, M.-H., & de Roo, A. (2009). The European  
936 Flood Alert System – Part 1: Concept and development. *Hydrology and Earth  
937 System Sciences*, 13(2), 125–140. doi: 10.5194/hess-13-125-2009
- 938 Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit fore-  
939 casts of winter precipitation using an improved bulk microphysics scheme. part  
940 II: implementation of a new snow parameterization. *Monthly Weather Review*,  
941 136(12), 5095–5115. doi: 10.1175/2008MWR2387.1
- 942 Tsiringakis, A., Steeneveld, G. J., & Holtslag, A. a. M. (2017). Small-scale orographic  
943 gravity wave drag in stable boundary layers and its impact on synoptic  
944 systems and near-surface meteorology. *Quarterly Journal of the Royal Meteo-  
945 rological Society*, 143(704), 1504–1516. doi: 10.1002/qj.3021
- 946 van Tiel, M., Aubry-Wake, C., Somers, L., Andermann, C., Avanzi, F., Baraer, M.,  
947 … Yapiyev, V. (2024). Cryosphere–groundwater connectivity is a miss-  
948 ing link in the mountain water cycle. *Nature Water*, 2(7), 624–637. doi:  
949 10.1038/s44221-024-00277-8
- 950 Wagner, S., Fersch, B., Yuan, F., Yu, Z., & Kunstmann, H. (2016). Fully coupled  
951 atmospheric-hydrological modeling at regional and long-term scales: Devel-  
952 opment, application, and analysis of WRF-HMS. *Water Resources Research*,

- 953                   52(4), 3187–3211. doi: 10.1002/2015wr018185
- 954                   Wang, S., Zheng, H., Lin, P., Zheng, X., & Yang, Z.-L. (2019). Evaluation and un-  
955                   certainty attribution of the simulated streamflow from NoahMP-RAPID over a  
956                   high-altitude mountainous basin. *Chinese Science Bulletin (in Chinese)*, 64(4),  
957                   444–455. doi: 10.1360/N972018-00598
- 958                   Wang, X., Tolksdorf, V., Otto, M., & Scherer, D. (2021). WRF-based dynamical  
959                   downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new  
960                   version of the High Asia Refined analysis. *International Journal of Climatol-*  
961                   *ogy*, 41(1), 743–762. doi: 10.1002/joc.6686
- 962                   Xie, K., Li, L., Chen, H., & Xu, C.-Y. (2025). Assessing the performance of  
963                   convection-permitting climate model in reproducing basin-scale hydrological  
964                   extremes: A western Norway case study. *Journal of Hydrology*, 656, 132989.  
965                   doi: 10.1016/j.jhydrol.2025.132989
- 966                   Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M.  
967                   (2019). MERIT Hydro: A high-resolution global hydrography map based on  
968                   latest topography dataset. *Water Resources Research*, 55(6), 5053–5073. doi:  
969                   10.1029/2019wr024873
- 970                   Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal,  
971                   J. C., ... Bates, P. D. (2017). A high-accuracy map of global ter-  
972                   rain elevations. *Geophysical Research Letters*, 44(11), 5844–5853. doi:  
973                   10.1002/2017GL072874
- 974                   Yang, X., Wei, C., Li, Z., Yang, H., & Zheng, H. (2024). A stream-order family and  
975                   order-based parallel river network routing method. *Water*, 16(14), 1965. doi:  
976                   10.3390/w16141965
- 977                   Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., ... Wood, E. F.  
978                   (2021). Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bul-*  
979                   *letin of the American Meteorological Society*, 102(11), E2086–E2105. doi:  
980                   10.1175/BAMS-D-20-0057.1
- 981                   Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., ...  
982                   Xia, Y. (2011). The community Noah land surface model with multipa-  
983                   rameterization options (Noah-MP): 2. Evaluation over global river basins.  
984                   *Journal of Geophysical Research: Atmospheres*, 116(D12), D12110. doi:  
985                   10.1029/2010JD015140

- 986 Yuan, X., Yang, K., Lu, H., Wang, Y., & Ma, X. (2023). Impacts of moisture trans-  
987 port through and over the Yarlung Tsangpo Grand Canyon on precipitation  
988 in the eastern Tibetan Plateau. *Atmospheric Research*, 282, 106533. doi:  
989 10.1016/j.atmosres.2022.106533
- 990 Zhang, B., Wang, S., & Wang, Y. (2021). Probabilistic projections of multidimen-  
991 sional flood risks at a convection-permitting scale. *Water Resources Research*,  
992 57(1), e2020WR028582. doi: 10.1029/2020WR028582
- 993 Zhang, Z., Arnault, J., Laux, P., Ma, N., Wei, J., Shang, S., & Kunstmann, H.  
994 (2022). Convection-permitting fully coupled WRF-Hydro ensemble simulations  
995 in high mountain environment: impact of boundary layer- and lateral flow pa-  
996 rameterizations on land-atmosphere interactions. *Climate Dynamics*, 59(5),  
997 1355–1376. doi: 10.1007/s00382-021-06044-9
- 998 Zheng, H., Fei, W., Yang, Z.-L., Wei, J., Zhao, L., Li, L., & Wang, S. (2023). An  
999 ensemble of 48 physically perturbed model estimates of the 1/8° terrestrial  
1000 water budget over the conterminous United States, 1980–2015. *Earth System  
1001 Science Data*, 15(7), 2755–2780. doi: 10.5194/essd-15-2755-2023
- 1002 Zheng, H., Yang, H., & Chen, S. (2024). *Evaluating atmospheric model pa-  
1003 rameterization schemes with river network routing and streamflow obser-  
1004 vations: A case study of the Yarlung Zangbo River on the Tibetan Plateau.*  
1005 <https://www.scidb.cn>: Science Data Bank. doi: 10.57760/sciencedb.11618
- 1006 Zheng, H., Yang, Z.-L., Lin, P., Wu, W.-Y., Li, L., Xu, Z., ... Wang, S. (2020).  
1007 Falsification-oriented signature-based evaluation for guiding the develop-  
1008 ment of land surface models and the enhancement of observations. *Jour-  
1009 nal of Advances in Modeling Earth Systems*, 12(12), e2020MS002132. doi:  
1010 10.1029/2020MS002132
- 1011 Zhou, X., Yang, K., Ouyang, L., Wang, Y., Jiang, Y., Li, X., ... Prein, A. (2021).  
1012 Added value of kilometer-scale modeling over the third pole region: a  
1013 CORDEX-CPTP pilot study. *Climate Dynamics*, 57(7), 1673–1687. doi:  
1014 10.1007/s00382-021-05653-8
- 1015 Zou, L., & Zhou, T. (2024). Convection-permitting simulations of current and future  
1016 climates over the Tibetan Plateau. *Advances in Atmospheric Sciences*, 41(10),  
1017 1901–1916. doi: 10.1007/s00376-024-3277-9