Revision notes for "Potential application of hydrological ensemble prediction in forecasting flood and its components over the Yarlung Zangbo River Basin, China" (hess-2018-179)

Dear Editor and referees,

Thanks a lot for your great efforts to review this manuscript and give very valuable comments. We agree with your suggestions which will be of great help to improve the quality of our manuscript. Here we have addressed the comments from you and the detailed description is attached in this document.

Best regards,
Li Liu, Yue-Ping Xu,Suli Pan, Zhixu Bai

To Referee #2

1. Page 9 line 10: the out-performance during evaluation period could also be related to the shorter length of the timespan.

Response: Thanks for your suggestion. We have added this reason in the revised manuscript.

"Generally speaking, the model performance during evaluation is more satisfactory than that during calibration. It is probably caused by the existence of considerable extraordinary flood events during the calibration period. The relatively shorter timespan in evaluation is also one of the reasons."

2. Page 9 line 15: please add some references to support the opinion.

Response: We have added some previous studies to justify the opinion. Please see Page 9, Line 15-17:

"We also guess that within downstream regions the hydrological process becomes too complicated due to human activities to be simulated by models (Li et al., 2013; Liu et al., 2014)."

Li, F., Xu, Z., Feng, Y., Liu, M. and Liu, W.: Changes of land cover in the Yarlung Tsangpo River basin from 1985 to 2005, Environ. Earth Sci., 68, 181-188, ,2013.

Liu, Z., Yao, Z., Huang, H., Wu, S., and Liu, G.: Land use and climate changes and their impacts on runoff in the Yarlung Zangbo river basin, China, Land Degrad. Dev., 25, 203-215, doi:10.1002/ldr.1159, 2014.

3. Page 9 line 31: "for the output snow depth from VIC is actually the sum of snow and glacier/ice" is redundant, delete it.

Response: We have deleted this sentence in the revised manuscript.

4. Page 13 line 31: change "compared" into "comparing".

Response: We have changed "compared" with "comparing".

5. Page 14 line 10: since you mentioned other two methods calculating the glacier melt, and meanwhile in line 13 stated "Overly complicated methods probably bring out more uncertainties….more observations are available with the development of technologies in the future, more elaborate separation method is expected" what's the observed input for these two models, are they overly complicated and with more uncertainty? If not, please modify the deduction since it cannot be reached either based on your study or the studies you referred to.

In the discussion section, please add some references on the similar studies using the proxy method and make a comparison with this study.

Response: Thank you for your suggestion.

(1) What's the observed input for these two models, are they overly complicated and with more uncertainty? If not, please modify the deduction since it cannot be reached either based on your study or the studies you referred to.

The simple degree-day glacier algorithm requires only the temperature data and the distribution of glacier in study area. The significant defect of this method is without considering water balance. The energy balance method requires incoming longwave radiation, emitted longwave radiation, outgoing longwave radiation and other data to calculate related glacier surface-energy balance and mass balance. If all those data are calculated from air temperature rather than observation, the sparse distributed meteorological network is highly likely to bring in additional uncertainties.

The deduction here is somewhat deviated from our conclusion, and we have modified the related part based on our study.

(2) In the discussion section, please add some references on the similar studies using the proxy method and make a comparison with this study.

The practice using proxy as observations is common when observed streamflow is absent. Similar studies can be found in Arnal et al. (2018) and Harrigan et al. (2018). We have added the relevant studies in the discussion section.

"For streamflow components forecast, the biggest challenge is the absence of data series of in-situ streamflow components. Therefore, in this study the simulation driven by observed forcing becomes an alternative to act as proxy and thus the error stemming from hydrological model is avoided. This is a common practice when observation is absent (Arnal et al., 2018; Harrigan et al., 2018)."

Arnal, L., Cloke, H.L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B. and Pappenberger, F.,: Skilful seasonal forecasts of streamflow over Europe? Hydrol. Earth Syst. Sci., 22(4), 2057-2072, 2018.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K. and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrol. Earth Syst. Sci., 22(3), 2023-2039, 2018.

Page 18, line 25: please add the year for the reference.

Response: We have added the year for the reference.

To Referee #3

1. How the "ability" of ensemble predictions to forecast can be defined? It could be expected that some weather projections predict observed flood peak and other not. At which point the hypothesis can be falsified? What is a level of CRPS, CRPSS or MAE at which authors would agree that ensemble predictions are unable to forecast a flood peak? In my opinion it always is possible to insist that the approach is "able" - better or worse.

The problem should be rather analysed in terms of suitability, where the ensemble predictions are compared with other methods. Then it would be possible quantifying the performance of different methods using i.e. CRPSS. Such an approach can be found i.e. in Pappenberger et al. (2005) who analyzed ensemble along with deterministic predictions.

Response: Thank you very much.

(1) We agree that it is always difficult to interpret CRPS, CRPSS and MAE. However, it is still the most commonly used metrics to quantify ensemble forecasts. We have read the recommended study by Pappenberger et al. (2005). It is a good example to analyze ensemble with deterministic predictions, but, in our opinion, it is not suitable for our study. Firstly, in the study of Pappenberger et al. (2005) (hereinafter referred as "Pappenberger' study"), only one case study is used, and thus the plot of hydrograph against lead time is feasible. In our study, we choose more than ten case studies for each station, making it redundant to adopt similar plots. Secondly, the deterministic predictions in Pappenberger' study consisted of 6 ensembles, and it is convenient to graph all the hydrographs, but in our study the S-simulation prediction is still composed by 51 ensembles.

(2) We totally agree to classify the problem as suitability, and we have revised the expression in the manuscript. The CRPSS has been used to quantify the model performance in original paper as shown in Fig. 6 and Fig. 10-13.

"The two purposes of this study are therefore to investigate the suitability of HEPS in forecasting flood volume and its components over cold and mountainous area, and the impact of an ensemble of selected pareto optimal solutions on model simulation and forecasting compared to a single parameter set."

2. Hydrograph separation is probably the weakest part of the study. According to the text, glaciers are important part of the basin system. However, the applied model does not account for this component. In Page 5, line 16-18 authors assume that it can be described as snow accumulation. This is a very rough assumption and its limitations can be seen in Fig 3 (authors are aware of this fact, page 8, lines 25-26), where flows, probably shaped by glacier outflow, are not explained. What is the point analysing forecasted streamflow components section 4.3), if the essential element of the base flow is missing? The conclusion to this remark could be the sentence from the Discussion section, referring to the forecasted components (Page 14, lines 18-19):

"Therefore, in this study the simulation driven by observations becomes an alternative to act as proxy although it is difficult to determine whether such proxy is believable or not." If it is hard to determine that we can believe the method or not, it is not a scientific approach.

Response: It is our mistake to unclearly describe the glacier part. Actually, the glacier only takes up about 2% of the area in the YZR (Zhang et al., 2013), and of streamflow less than 10% (Chen et al., 2017). The underestimation of baseflow in Fig.3 is not, to a great degree, caused by lack of glacier modeling. For the low flow period is the time when the glacier should be accumulated, the absence of glacier is supposed to overestimate baseflows. We have compared our results with a previous study (Zhang et al., 2013, Fig.3(e), the study period is somewhat different, from 1961 to 1999 in that study), similar underestimation exits in low flow period, even though a glacier is embedded in the hydrological model. The authors thought the errors were mainly from the observed inputs. However, how much the inputs might be underestimated for the YZR is actually unknown. In this study, one of the reasons for this phenomenon is that the objective functions used to calibrate hydrological model emphasize more on high flows. The underestimation is, in the meanwhile, caused by the errors of meteorological measurements in the study area, which has been documented by previous studies like Tong et al. (2014) and Zhang et al. (2013).

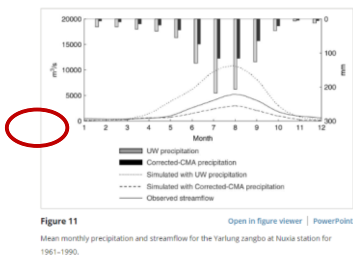


Fig.11 in Tong et al. (2014)

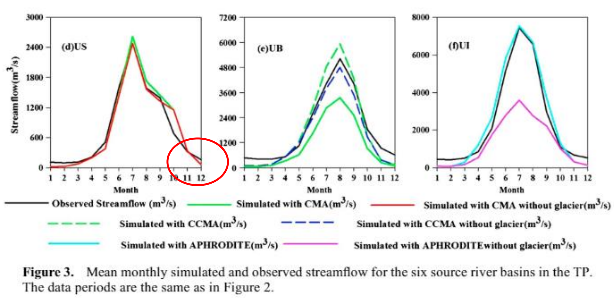


Fig.3 in Zhang et al. (2013)

Second, the snow is truly an important contributor to total runoff, especially during spring. It is reported that near 25% of the total runoff is derived from snowmelt water (Zhang et al., 2013). Most of previous studies (Li et al., 2014; Liu et al., 2014; Sun et al., 2013) considered glacier together with snow when simulating streamflow components in this study area, as the glacier is neglectable if compared to snow. Another reason to modelling glacier as snow is that based on current hydrological model structure and available observations, it is hard to separate glacier from snow. Finally, our study is more interested in meltwater induced streamflow rather than snow and glacier meltwater streamflow respectively.

As for “What is the point analyzing forecasted streamflow components section 4.3”, it is because the model simulation is possibly biased and using simulation driven by observed forcing data as observation proxy is an effective way to remove hydrological errors and is suitable to compare the N-simulations and S-simulation in forecasts components.

All of abovementioned contents has been added or rewrote in the revised manuscript.

“The observed and simulated hydrographs during the evaluation period at Nuxia are presented in Fig. 3. An obvious underestimation can be observed in low flow periods, which is similar to previous studies by Zhang et al. (2013) and Tong et al. (2014). The absence of glacier module in VIC is believed to has limited influence on this underestimation, for similarly underestimated low flow was found when glacier modelling was embedded in VIC (Zhang et al., 2013). For our study, the underestimation is, in the meanwhile, caused by the fact that the objective functions used for calibration have the tendency to give more attention to high flows, as the flood is the focus of our investigation. As noticed in Fig. 3, the flood peaks are well captured by S-simulation in most cases. N-simulations are able to cover all the extreme values while sometimes slight overestimation exists.”

Li, F., Xu, Z., Liu, W. and Zhang, Y.: The impact of climate change on runoff in the Yarlung Tsangpo River basin in the Tibetan Plateau. Stoch. Env. Res. Risk A., 28(3), 517-526, 2014

Liu, Z., Yao, Z., Huang, H., Wu, S. and Liu, G.: Land use and climate changes and their impacts on runoff in the Yarlung Zangbo river basin, Land Degrad. Dev., 25(3), 203-215, DOI: 10.1002/ldr.1159, 2014

Chen, X., Long, D., Hong, Y., Zeng, C. and Yan, D.: Improved modeling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin? Water Resour. Res., 53, 2431-2466, doi:10.1002/2016WR019656, 2017.

Tong, K., Su, F., Yang, D., Zhang, L., and Hao, Z.: Tibetan Plateau precipitation as depicted by gauge observations, reanalyses and satellite retrievals, Int. J. Remote Sens., 34, 265-285, doi:10.1002/joc.3682, 2014.

Sun, R., Zhang, X., Sun, Y., Zheng, D. and Fraedrich, K.: SWAT-based streamflow estimation and its responses to climate change in the Kadongjia River watershed, southern Tibet, J. Hydrometeorol., 14(5), 1571-1586, 2013

Zhang, L., Su, F., Yang, D., Hao, Z., and Tong, K.: Discharge regime and simulation for the upstream of major rivers over Tibetan Plateau, J. Geophys. Res., 118, 8500-8518, doi:10.1002/jgrd.50665, 2013.

3) The concept of the ensemble of Pareto optimal parameters is the most interesting, but not addressed properly. Authors have identified a N-set of parameters that provides the tradeoffs between four fit measures: the Nash–Sutcliffe efficiency and relative bias for whole flow time series and limited to 10% of highest flows. However, the performance of the forecast system is analyzed only in the respect of a single measure: accumulated flood volumes. For such a single criterion it is quite obvious

that a single compromise simulation should be better. The advantage of the parameter ensemble could be found, if other measures, similar to those in calibration, were also included.

Response: Thanks for the referee being interested in our study. We agree that accumulated flood volumes only present the function of 10% of highest flows, but this sentence is right only when we are talking about the annual maximum flood. For annual first flood (spring flood) and the peak time simulating, it is more related to the whole hydrograph. This means actually both flood volumes and peak time are considered in evaluating the performance of the forecast system.

4) In the reviewer's opinion, authors should revise their research design, focusing maybe on the forecast quality, in the respect of identified parameters (Pareto-sets), instead of stream flow components.

Response: As a matter of fact, the focus of this study is both forecasts in total runoff and streamflow components. This is the reason why we constructed the manuscript as the order of "streamflow simulating-total runoff forecasting-streamflow components forecasting". We found the problem in our previous study which made the streamflow components forecasting an ultimate goal. We have revised the structure of the manuscript. Please check the revised submission.

A) The description of Snowmelt Tracking Algorithm, provided in Section 3.3 is unclear and not supported with references to Li et al. (2017). The original Snowmelt Tracking Algorithm Li et al. 2017) acquires internal variables of a runoff model:

"In this study we develop explicit quantification of the historical and future fQ , snow over the western U.S. We quantify fQ , snow by tracking the fate of snowmelt in modeled hydrologic fluxes" - Li et al. (2017).

In my opinion all assumptions used in eq. (5-7) should be justified and explained by authors. It is unclear, why it is possible to assume, that infiltration (f_{i,snow}) and runoff ratios (f_{R,snow}) are equal (Eq. 5). Also Eq. 7 is unclear. What is a definition of fi,snow? Equiation 5 suggests that it is a ratio: fi,snow = \frac{I_snow}{R}. Why in Eq. 7 it is multiplied by i (infiltration?)? The same applies to F_{w, snow}.

Please, provide also dimensions for variables: R, B, W and ETt.

Response: (1) Thank you very much for this comment. We don't assume the infiltration and runoff ratios to be equal. As a matter of fact, we only assume meltwater and rainfall exhibit identical infiltration and runoff ratios. But we did make some mistakes when writing Eq. (5) in the original manuscript. We have checked our note and code, and make sure that this is just a writing error. We are so sorry to overlook this mistake in the last revision.

(2) The definition of $f_{i,snow,t}$ is the ratio of snow-induced infiltration in total infiltration. Thus, it can be multiplied by total infiltration (i_t).

(3) The dimensions for variables R_t, B_t, W_t and ET_t are millimeter and have been added in the manuscript. We show the results at the basin average.

B) Page 3, Lines 10-15 are unclear and please consider rewriting. Parameters ensembles of a hydrological model and parameter sets refer here to a wider problem of the hydrological model uncertainty. Mentioned techniques are just possible solutions. I suggest defining this problem as an uncertainty analysis and provided these references as examples. Please note, that in the present form the text is confusing, as it is hard to distinguish between "forecast ensemble" (meteorological model uncertainty) and "parameter ensemble" (for hydrological models' uncertainty).

Response: Thank you very much for your suggestion. We are sorry to mix the concept of meteorological uncertainty and hydrological uncertainty. In the revised manuscript, we have defined the uncertainty from parameter of hydrological model as hydrological uncertainty and added the relevant references. Please see Page 3, Lines 10-15:

"<u>Due to this limitation, utilizing an ensemble of parameter sets to represent uncertainty from hydrological model is referential. Pappenberger et al. (2005) used six different parameter sets to identify uncertainty from hydrological model. Teutschbein and Seibert (2012) employed 100 different optimized parameter sets in HBV to simulate streamflow in order to consider parameter uncertainty. The basic principle in ensemble forecasts is using ensemble spread to quantify forecast uncertainty and thus provide essential information to users (Bauer et al., 2015). Analogous to this concept, the benefit of adopting an ensemble of parameter sets from Pareto optimal front by multi-objective optimization algorithm for flood forecasting with consideration of hydrological parameter uncertainty remains unresolved and is noteworthy to investigate. Especially for streamflow components modelling and forecasting, with limited or unavailable observations, it is impossible to achieve rigorous calibration, and thus accounting for hydrological parameter uncertainty is necessary (Pappenberger et al., 2005).</u>"

<u>Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J. and De Roo, A.P.J., Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrol. Earth Syst. Sci., 9(4), 381-393, 2005.</u>

C) Page 4, Lines 16-17: Please note, that the supporting references are more than 4 years old.

Response: Thanks for your reminder. We have replaced those supporting references with the latest ones. Please see Page 4, Lines:

"<u>ECMWF is selected in this study due to the well-known fact that forecasts from ECMWF are more skilful than other Ensemble Prediction Systems in TIGGE database (Aminyavari et al., 2018; Louvet et al., 2016; Hamill and Scheuerer, 2018).</u>"

<u>Louvet, S., Sultan, B., Janicot, S., Kamsu-Tamo, P.H. and Ndiaye, O.: Evaluation of TIGGE precipitation forecasts over West Africa at intraseasonal time scale,Clim.Dynam., 47(1-2), 31-47, DOI 10.1007/s00382-015-2820-x, 2016.</u>

<u>Aminyavari, S., Saghafian, B. and Delavar, M.: Evaluation of TIGGE Ensemble Forecasts of Precipitation in Distinct Climate Regions in Iran, Adv. Atmos. Sci., 35(4), 457-468, 2018.</u>

<u>Hamill, T.M. and Scheuerer, M.: Probabilistic Precipitation Forecast Postprocessing Using Quantile Mapping and Rank-Weighted Best-Member Dressing, Mon. Weather Rev., 146(12), 4079-4098, 2018.</u>

D) Page 5, Line 7, what data was used in cross-validation - from these 27 meteorological stations? If so, it should not be written, that lapse rates were estimated using available meteorological data?

Response: Yes, records from these 27 meteorological stations were used in a leave-one-out-cross-validation. The lapse rate is defined by performing the least square fitting considering elevation.

E) Section 4.2. It is unclear for how a flood forecasting system was tested. Did authors use a receding horizon or forecast was issued for chosen events? Terms "first floods" maximum floods etc. should be defined.

Response: We didn't use the receding horizon. The forecast was issued for each chosen event. We have added the relevant information in the revised manuscript. The terms of "first floods" and "maximum floods" were defined in the original manuscript in Section 3.5.

"The annual maximum flood is picked out as typical flood events. Meanwhile, the first flood event in each year is also selected. Maximum flood is determined by the maximum daily streamflow in a year. For first flood, the definition seems to be slightly subjective. Nevertheless, first flood is just introduced as an example to verify the skill of VIC/ECMWF system to forecast the meltwater components. There are three criterions for us to define the first flood: (1) the peak flow should be more than twice of the average daily streamflow during dry period (November to March); (2) the duration for the flood event should be longer than 7 days; (3) the observed snowpack is present. Forecasts were issued for each chosen event. Considering that maximum flood events in YZR usually last for several months, flood volume over the entire flood event is impossible to be covered by medium-range weather forecasts. Four typical flood volumes are therefore chosen to represent the volume performance…"

F) Page 13, Line 21: "Firstly, N-simulations generally behave better when the trade-offs in multi-objectives are significant." does the study supports this finding?

Response: This conclusion was drawn based on the results that N-simulations generally behave poor at Nuxia station, while for Nugesha and Yangcun N-simulations are better than S-simulation (Table 1 and 2). The difference for these three stations is that the trade-off between two objective functions is arc-shaped at Nugesha and Yangcun stations, indicating when moving on the front the improvement in one objective function will result in deterioration in the other objective function. While at Nuxia station, the compromise point is almost the best parameter set with highest score in most of the objective functions. From this perspective, we think that N-simulations generally behave better when the trade-offs in multi-objectives are significant. We possibly didn't explain this clearly in the previous study, and we have added this in the revised manuscript. Please see Section 4.1 and Conclusion:

"Those different behaviors at three stations imply that N-simulations is preferable when the trade-offs between multi-objective functions are significant (no single parameter set behaves well in most of the objectives like Nuxia Station)…

"Firstly, N-simulations generally behave better when the trade-offs in multi-objectives are significant. In this case, the N-simulations can synthesize advantages from different components. This is why N-simulations can provide more desirable skill at Nugesha than Yangcun…"

Potential application of hydrological ensemble prediction in forecasting flood and its components over the Yarlung Zangbo River Basin, China

Li Liu[1], Yue P. Xu[1], Su L. Pan[1], Zhi X. Bai[1]

[1]Institute of Hydrology and Water Resources, Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, China

Correspondence to: Yue-P. Xu (yuepingxu@zju.edu.cn)

Abstract. In recent year, flood becomes a serious issue in Tibetan Plateau (TP) due to climate change. Many studies have shown that ensemble flood forecasting based on numerical weather predictions can provide early warning with extended lead time. However, the role of hydrological ensemble prediction in forecasting flood volume and its components over the Yarlung Zangbo River (YZR) Basin, China, hasn't been ~~systematically~~ investigated. This study adopts Variable Infiltration Capacity (VIC) model to forecast annual maximum floods and annual first floods in YZR based on precipitation, maximum and minimum temperature from European Centre for Medium-Range Weather Forecasts (ECMWF). N-simulations is proposed to account for ~~more scenarios of~~ parameter~~s~~ uncertainty in VIC. When trade-offs between multiple objectives are significant, N-simulations is recommended for better simulation and forecasting. This is why better results are obtained for Nugesha and Yangcun stations. Ensemble flood forecasting system can skilfully predict maximum floods with a lead time of more than10 days, and about 7 days ahead for melt-water related components. The accuracy of forecasts for first floods is inferior with a lead time of only 5 days. The baseflow components for first floods are insensitive to lead time except at Nuxia Station, whilst for maximum floods an obvious deterioration in performance with lead time can be perceived. The meltwater-induced surface runoff is the most poorly captured component by the forecast system, and the well-predicted rainfall-related components are the major contributor for good performance. The performance in 7-day accumulated flood volumes is better than the peak flows.

1. Introduction

The Tibetan Plateau (TP) as the source of many major rivers is known as "the world water tower" (Xu et al., 2008). Due to its special geological, topographic and meteorological conditions, the ecosystem in this area is vulnerable and susceptible to climate changes (Zhao et al., 2006). According to previous studies, it is confirmed that the atmospheric and hydrological cycle in TP have undergone significant changes. Evident climate warming (Guo and Wang, 2012; Wang et al., 2014; Yang et al., 2014), increased precipitation (Kuang and Jiao, 2016, Wang et al., 2017), glacier retreat and permafrost degradation (Cheng and Wu, 2007) can be perceived, and these impacts are expected to be exacerbated by future climate change (Su et al., 2013). As a result, frequent natural disasters, such as flooding and debris flow, take place, with an estimated direct economic loss amounting to 100 million RMB per year (Zhang et al., 2001). Thus, seeking advanced techniques to improve the accuracy of flood forecasts plays a critical important role in enhancing disaster resilience (Kalra et al., 2013; Yucel et al., 2015; Girons et al., 2017).

It is now a routine practice to introduce the Numerical Weather Prediction (NWP) products into research ~~operational~~ and operational ~~research~~ flood forecasting system to generate ensemble streamflow forecasts (Cloke and Pappenberger, 2009). Compared with traditional single-value deterministic flood forecasts, ~~it has been verified that the~~ forecasts based on Hydrological Ensemble Prediction System (HEPS) outperform the traditional deterministic ones with higher accuracy and longer lead time (Bartholmes et al., 2009; Cloke et al., 2013; 2017; Li et al., 2017; Pappenberger et al., 2015; Todini, 2017). Flood forecasting is one of the most important topics applying HEPS (Arheimer et al., 2011; Shi et al., 2015), but most of the studies only focus on peak flows (Alvarez-Garreton et al., 2015; Valeriano et al., 2010; Dittmann et al., 2009), and few studies have investigated the suitability~~ability~~ of HEPS forecasts in typical accumulated flood volumes and the respective components contributed to the flood volumes, especially the snowmelt-induced component. It is shown that snow water availability and snow dynamics are issues of fundamental importance in high mountain hydrology (Bavera et al., 2012). Investigating the components constituting the total runoff facilitates the understanding of runoff generation mechanism and furtherly improving flood forecasting in high mountains where our study area is located.

Investigating the skill of HEPS in streamflow component simulation requires effective methods to separate total runoff into different components of interest. Numerous researchers have studied the methods to achieve hydrograph separation. Some researchers are interested in separating baseflow or groundwater component from total runoff. For example, Partington et al. (2011) developed a hydraulic mixing-cell method to determine the groundwater component and Luo et al. (2012) utilized the digital filter program to separate baseflow from streamflow. However, many of the hydrological models per se have the ability to separate streamflow into baseflow and surface runoff, like SWAT (Luo et al., 2012) and VIC ~~model~~ (Liang et al., 1994), thus the separation of snow/glacier-~~driven~~ induced component from rainfall-induced component gains increasing interests. The most common and historical practice to separate snowmelt and glaciermelt components is to conduct stable isotope analysis (Isotopic hydrograph separation, HIS) (Laudon et al., 2002). Sun et al. (2016) applied HIS in the Aksu River and successfully calculated the relative contribution of the glacier and snow meltwater to total runoff. Besides the experimental approaches, considerable studies obtain snowmelt component via a simple ratio of rainfall and snowmelt from hydrological model simulation (Cuo et al., 2013a; Siderius et al., 2013), whereas these methods are often primitive and neglect the physical processes that affect the transformation from snow to runoff, such as evapotranspiration, sublimation, and infiltration. Li et al. (2017) developed a new snowmelt tracking algorithm in the VIC model to compute the ratio of the snow-derived runoff to the total runoff with consideration of systematic analyses, demonstrating promising performance in applications over western United States.

Generally, evaluating model performance should be performed based on in-situ observations. However, observed streamflow components are usually unavailable, making the evaluation of streamflow component simulations/forecasts intractable. Meanwhile, with limited or unavailable observations, it is impossible to achieve rigorous calibration, and thus accounting for hydrological parameter uncertainty is necessary (Pappenberger et al., 2005). ~~Generally, evaluating model performance should~~

Formatted: No underline, Font color: Auto

~~be performed based on in-situ observations. However, observed streamflow components are usually unavailable, making the evaluation of streamflow component simulations/forecastsintractable. The alternative solution is to verify forecasts on model simulationsassuming that simulations driven by meteorological observations present the actual hydrological components. However, the successof this practice highly depends on how well the hydrological model iscalibrated.~~ Yapo et al. (1998) showed that there is no single objective function that can represent all the features of runoff hydrographs such as time to peak, peak flow and runoff volume. Increasing investigators have realized that multiple objectives optimization can bring out better results than single ones, and currently majority of the hydrological models are calibrated based on multi-objective optimization algorithms (Kamali et al., 2013; Troy et al., 2008; Voisin et al., 2011; Yuan et al., 2013). Multi-objective formulation will result in a set of Pareto optimal solutions that represent trade-offs among different objectives (Wöhling et al., 2013). Thus, compromise is necessary (Gong et al., 2015). Most of the studies eventually select only one value from the Pareto front to represent the model parameter set for their simulation (Troy et al., 2008; Voisin et al., 2011; Yuan et al., 2013; Liu et al., 2017). This value is usually the compromise point that balances the diverse and sometimes conflicting requirements. However, these solutions provided by multi-objective optimization algorithms have the feature that moving from one objective to another along the trade-off surface results in the improvement of one objective while causing deterioration in at least one other objective. ~~In some cases~~Additionally, as mentioned by Kollat et al. (2012), it is difficult, in some cases, to cause the two-objective trade-off to collapse to one single point. Due to this limitation~~,~~, ~~a few studies tend to~~utilizinge an ensemble of parameter sets to ~~cover more probabilities in hydrological model state~~represent uncertainty from hydrological model is necessary. ~~Wöhling and Vrugt (2008) employed Bayesian model averaging to generate forecast ensembles of soil hydraulic models, showing similar skills to the best ones.~~ Pappenberger et al. (2005) used six different parameter sets to identify uncertainty from hydrological model. Teutschbein and Seibert (2012) employed 100 different optimized parameter sets in HBV to simulate streamflow ~~with a wide range of potential sources of variability~~in order to consider parameter uncertainty. The basic principle in ensemble forecasts is using ensemble spread to quantify forecast uncertainty and thus provide essential information to users (Bauer et al., 2015). Analogous to this concept, the benefit of adopting an ensemble of parameter sets from Pareto optimal front by multi-objective optimization algorithm ~~of hydrological models~~for flood ~~forecastingand components forecasting to considerwith~~forecasting with consideration ~~ofmore possible hydrological initial conditionshydrological~~of hydrological parameter uncertainty remains ~~an~~ unresolved and is noteworthy ~~question~~to investigate. ~~parameter~~

The two purposes of this study are therefore to investigate the ~~ability~~ suitability of HEPS in forecasting flood volume and its components (rainfall-induced and meltwater-induced streamflow) over cold and mountainous area, and the ~~impact~~ impact of an ensemble of selected pareto optimal solutions on model simulation and forecasting compared to a single parameter set. To this end, the paper is structured as follows: Section 2 describes the information of study area and data used. Methodology description is in Section 3. Section 4 provides the result analysis, and Section 5 discusses the main findings and points for future research directions, and conclusion is presented in Section 6.

2. Study area and data

2.1 Study area

We focus our analysis on the Yarlung Zangbo River (YZR) Basin, located at the upper reaches of Brahmaputra River basin, which stretches across the southern part of TP from the west to the east, with a drainage area of $2.1 \times 10^5\ km^2$ controlled by Nuxia hydrological station in China. The basin is selected for the greatest population density in TP, increasing runoff and glacier~~-~~/snow melt (Wang et al., 2009; Liu et al., 2014), making it an ideal region to investigate flood forecasting and its components. YZR is one of the highest great rivers in the world with a mean elevation exceeding $4000\ m$ a.s.l. The climate from upstream to downstream regions of the basin exhibits an obvious difference due to the location and topographical feature of TP (Liu et al., 2014). The downstream area has a warm and humid subtropical climate; the midstream area has a temperate forest-grassland climate and the upstream valley has a cold and dry temperate steppe climate (Liu et al., 2007; Shen et al., 2012). The average annual temperature in this basin is about 6.27 ℃. The average annual precipitation is about $560\ mm$, most of which occurs during the wet season from May to September (Li et al., 2014). Approximately $1/3$ of the basin area is covered by snow and glacier, resulting in a significant glacier-snow melt induced floods in late spring and early summer.

[Figure 1]

2.2 Data

The gauged meteorological data, including daily precipitation, minimum and maximum temperature, wind speed and relative humidity, from 1998 to 2015 are collected from 27 National Meteorological Observatory stations located in and around the YZR basin as shown in Fig.1. Daily streamflow from three ~~control~~ hydrological stations are utilized in this study, i.e. Nugesha Station, Yangcun Station and Nuxia Station from the most upstream to downstream region. Except data missing in 2009, the record period of observed streamflow at Nugesha and Nuxia is consistent with that of the meteorological data~~ and~~. T~~t~~he period of observed streamflow at Yangcun is shorter, spanning from 1998 to 2012. The first year is used as warm-up period. Periods from 1999 to 2005, 2006 to 2008 and 2010 to 2012/2015 are adopted for calibration, validation and evaluation purpose respectively.

The daily Quantitative Precipitation Forecasts (QPF) and Maximum/Minimum Temperature (MXT/MNT) from 2007 to 2015 are obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) with lead time from 24h to 360h. To be consistent with the observations, the data issued at 0000 UTC is downloaded. ECMWF is selected in this study due to the well-known fact that forecasts from ECMWF are more skilful than other Ensemble Prediction Systems in TIGGE database (Aminyavari et al., 2018; Louvet et al., 2016; Hamill and Scheuerer, 2018~~Buizza et al., 2005; Froude, 2010; Tao et al., 2014~~). Snow depth data provided by Cold and Arid Regions Science Data Center at Lanzhou, China (http://westdc.westgis.ac.cn/) are used to evaluate snow depth simulations. The data is derived from passive microwave remote sensing at a resolution of 0.25° ×0.25° (Che et al., 2008; Dai et al., 2015). The digital elevation model (DEM) data used in the hydrological model is downloaded from Geospatial Data Cloud (http://www.gscloud.cn) at the resolution of 90 m×90 m. The vegetation and soil parameters in the model are defined according to 1 km China soil map based on Harmonized World Soil Database (Fischer et al., 2008) and 1 km land cover products of China (Ran et al., 2010).

3. Methodology

3.1 Hydrological model

The Variable Infiltration Capacity model (VIC, Liang et al., 1994; 1996) is employed in this study to investigate the ~~ability~~ suitability of ensemble flood forecasting in YZR. VIC is a well-established and extensively used rainfall-runoff model, especially in areas with existence of snowmelt and frozen soil (Tang and Lettenmaier, 2010; Cuo et al., 2013a; Su et al., 2016). A two-layer snow model is embodied in VIC, which considers snow accumulation and ablation in a ground pack and an overlying forest canopy based on energy balance (Andreadis et al., 2009). The frozen soil algorithm makes it possible to represent the effects of seasonally frozen ground on surface water and energy fluxes (Cherkauer and Lettenmaier, 1999; 2003). These are two of the critical elements in VIC that are particularly relevant to our research.

In this study, VIC is operated at a six-hourly time step in both water and energy balance model with a spatial resolution of $0.125^{\circ} \times 0.125^{\circ}$. The snow and frozen soil algorithms are active. Gauged and forecasted meteorological data are interpolated into the required resolution using the Inverse Distance Weighted (IDW) method coupled with an elevation-based lapse rate. The lapse rate in this study is set as $0.6\ mmkm^{-1}$ for precipitation and $-6.5\ ^{\circ}Ckm^{-1}$ for temperature. These two lapse rates are determined by a cross-validation process based on records from the 27 meteorological stations. Our results ~~and~~ are roughly consistent with the findings in Cuo et al. (2013b), who performed the least squares fitting on daily temperature and precipitation over the TP area to gain the best lapse rate for interpolation.

Model calibration is conducted by a parallel-programmed Epsilon-Dominance Non-Dominated Sorted Genetic Algorithm II (ε-NSGA II) as proposed by the authors (Liu et al. 2017). The ε-NSGA II is coupled with Message Passing Interface (MPI) to achieve parallel autocalibration with high efficiency. As snow and frozen soil algorithms are activated, two additional parameters related to snow modelling, namely the maximum temperature at which snow can fall (T_{snow}) and the minimum temperature at which rain can fall (T_{rain}), are optimized together with other seven conventional calibration parameters (Detailed descriptions about the calibration of these seven typical parameters can be found in our previous studies (Liu et al., 2017)). The roles of those two temperature parameters in VIC are to determine what fraction of incoming precipitation is solid (snow) and liquid (rain). T_{snow} and T_{rain} are originally fixed for a given vegetation type. Considering glacier ablation and accumulation are simulated as snow in this study due to the absence of glacier module in the current VIC model, the ratio of solid and liquid precipitation is different from the original value. We tend to adjust them via ~~calibration.The~~calibration. The parameter ranges are defined as [-5,5] according to Chen et al. (2017), who used similar parameters in the CREST model for snow and glacier melting simulation.

As flood peaks and volumes are our focuses in this study, more weights are given to high flows during calibration. Four objective functions are used for model calibration at three hydrological stations: the Nash–Sutcliffe efficiency and relative bias for all flows and for the top 10% flows. Detailed formulas are defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_{obs}(i)-Q_{sim}(i))^2}{\sum_{i=1}^{N}(Q_{obs}(i)-\overline{Q_{obs}(i)})^2} \quad (1)$$

$$Bias = \frac{\sum_{i=1}^{N}[Q_{sim}(i)-Q_{obs}(i)]}{\sum_{i=1}^{N}Q_{obs}(i)} 100\% \quad (2)$$

$$NSE_{10\%} = 1 - \frac{\sum_{i=1}^{M}(Q_{obs,10\%}(i)-Q_{sim,10\%}(i))^2}{\sum_{i=1}^{M}(Q_{obs,10\%}(i)-\overline{Q_{obs,10\%}(i)})^2} \quad (3)$$

$$Bias_{10\%} = \frac{\sum_{i=1}^{M}[Q_{sim,10\%}(i)-Q_{obs,10\%}(i)]}{\sum_{i=1}^{M}Q_{obs,10\%}(i)} 100\% \quad (4)$$

in which N and M are the number of all daily flows and top 10% flows, respectively; Q_{obs} and Q_{sim} are the observed and simulated daily flows; and $Q_{obs,10\%}$ and $Q_{sim,10\%}$ are the observed and corresponding simulated top 10% flows, respectively.

3.2 N-Pareto-optimal parameter sets

After calibration, a series of feasible solutions are produced by ε-NSGA II. An inevitable challenge for users of automatic calibration routines is to face the task of selecting a set of suitable model parameters (preferred solution set) from numerous Pareto-optimal ~~sets.The~~sets. The method of Preference Ordering Routine (POR), developed by Khu (2005), is exactly designed to solve this kind of problem by sorting out a small number of preferred solutions. POR has been successfully applied for calibration of MIKE11/NAM rainfall-runoff model and is able to provide the best estimated parameter sets with good overall model performance. Therefore, POR is selected in this study to pick out the desired N-Pareto-optimal parameter sets.

There are two key attributes for this method. The first is the efficiency of k order (or k-Pareto-optimal points). Considering all the possible k-dimensional subspace of the original m-dimensional objective functions provided by ε-NSGA II ($1 \le k \le m, m = 4$ in this study), a point is defined as being efficient of order k if this point is not dominated by any other points in any of the k-dimensional subspaces. The second attribute is the efficiency of order k with degree p (or $[k,p]$-Pareto-optimal points). A point is defined as being efficient of order k with degree p if it is not dominated by any other points for exactly p out of the possible k-dimensional subspaces. By reducing the efficiency of order k and increasing the degree of order p in a sequential manner, POR is able to achieve the reduction of the number of possible solutions and to short-list the most relevant ones for retention as preferred parameters. The essence of POR is to tighten the criteria of Pareto optimality, and thus enables to determine the limited preferred ~~solutions.Detailed~~solutions. Detailed procedures and examples to apply POR are omitted here, and interested readers can refer to Khu (2005).

In this study, the POR is performed throughout all possible subspaces, and the parameter which is not dominated by any of the subspaces is retained. Additionally, some other points on the Pareto front are also retained: the extreme value for each objective function (indicated by filled circles in Fig. 2) and the compromise value in the two-objective trade-off (indicated by filled star in Fig. 2). In this way, limited number of parameter sets is picked out to represent different scenarios of model state. For convenience, the simulations driven by the N-Pareto-optimal parameter sets are referred as N-simulations, and the simulation by ~~only one~~single parameter set (the compromise point) is indicated by S-simulation thereafter.

3.3 Hydrograph separation

In this study, we are more interested in meltwater ~~(including snow and glacier melting~~(including snow and glacier melting)~~,~~ ~~as glacier only contributesabout2% to the area in the YZR (Zhang et al., 2013) and to total runoff less than 10% (Chen et al., 2017))~~ and rainfall induced components. Thus, ~~G~~glacier is modelled together with snow ~~(including snow and glacier melting,~~ as glacier only contributes about 2% to the area in the YZR basin (Zhang et al., 2013) and to total runoff less than 10% (Chen et al., 2017)~~)~~. The Snowmelt Tracking Algorithm (STA), proposed by Li et al. (2017), is ~~adopted in this study to separate the hydrograph~~thus an appropriate method to achieve the needed hydrograph separation. ~~As mentioned before, due to the lack of glacier algorithm in the VIC model, in this study the glaciermelting is considered together with snowmelting, and the term of “meltwater” is used for representation.~~ In order to obtain the streamflow derived from meltwater~~in total runoff~~, $Q_{snow,t}$, STA calculates the meltwater-induced streamflow in surface runoff (R_t ~~R~~) and baseflow (B_t ~~B~~) separately. For surface runoff derived from meltwater, ~~(~~$R_{snow,t}$), STA assumes that meltwater and rainfall exhibit identical infiltration ~~($f_{i,snow,t}$)~~ and runoff ratio. In ~~($f_{R,snow,t}$) ratios. In t~~this way, $R_{snow,t}$ is computed as a function of the ratio of meltwater M_t to meltwater + rainfall, $M_t + Rain_t$:

$$R_{snow,t} = M_t - i_{snow,t} = M_t - i_t f_{i,snow,t} = M_t - i_t \frac{M_t}{M_t+Rain_t} \cancel{R_t f_{R,snow,t} = R_t f_{i,snow,t} = R_t \frac{M_t}{M_t+Rain_t}} \qquad (5)$$

in which ~~the~~ i_t is the infiltration, and it is calculated by mass balance on the ground surface: $Rain_t + M_t = i_t + R_t$; $f_{i,snow,t}$ is the ratio of meltwater induced infiltration in total infiltration.

The fraction of ~~baseflow induced by~~ meltwater induced baseflow ($f_{B,snow,t}$) is assumed to be equal to the proportion of soil moisture that originated from meltwater in all soil moisture layers ($f_{W,snow,t}$). Thus:

$$B_{snow,t} = B_t f_{W,snow,t} \qquad (6)$$

Then, $f_{W,snow,t}$ is obtained by an iteration process. The formula used to obtain $f_{W,snow,t}$ is defined as follows:

$$f_{W,snow,t} W_t = f_{W,snow,t-1} W_{t-1} + f_{i,snow,t-1} i_t \Delta t - f_{W,snow,t-1}(ET_t - Sub_t)\Delta t - f_{W,snow,t-1} B_t \Delta t \qquad (7)$$

where W_t and ET_t are soil moisture and evapotranspiration outputs from VIC, respectively. Sublimation Sub_t is calculated from the evolution of the snow water equivalent (SWE).

A similar equation to Eq. (7) can be written for rain ($f_{W,rain,t}$). At each time step, $f_{W,snow,t} + f_{W,rain,t} + f_{W,unkown,t} = 1$. At step time $t = 1$, $f_{W,unkown,t} = 1$, indicating that the source of runoff (meltwater or rainfall) is unknown at initial time step. After the tracking system performed, $f_{W,unkown,t}$ decreases to 0, and sum of $f_{W,snow,t}$ and $f_{W,rain,t}$ is equal to 1 with fully explained soil moisture sources.

Unlike Li et al. (2017), all the aforementioned variables are integrated values over the entire basin in units of millimetre. When performing hydrograph separation, one-year warming up is used to achieve fully explained soil moisture sources. Total runoff

Formatted: Font: (Asian) +Body Asian (宋体), (Asian) Chinese (PRC), (Other) English (United States) is separated into four components, that is, the surface runoff derived from meltwater ($R_{snow,t}$) and from rainfall ($R_{rain,t}$); the baseflow derived from meltwater ($B_{snow,t}$) and from rainfall ($B_{rain,t}$).

3.4 Post-processing of forecasts from ECMWF

In order to improve the raw forecasts from ECMWF, we propose a post-processing method by coupling parameterized Quantile Mapping (QM) with Schaake Shuffle (hereafter referred to QM-SS). QM is adopted in this study for it is a simple yet effective statistical bias correction method in hydrological applications (Li et al., 2010; Xu et al., 2014; Salathé et al., 2015). In most cases, the empirical cumulative distribution function is used to present the data distribution in QM. However, many studies (Viste et al., 2013; Stauffer et al., 2017; Tao et al., 2014) have demonstrated that it is more appropriate to use fitted parametric distributions as no frequent interpolation or extrapolation would be requested (Li et al., 2010). For QPFs, due to the strongly positively skewed distribution in rainfall (Stauffer et al., 2017), QM based on single gamma distribution is recommended and utilized for bias correction in this study, although some studies found that a combination of double-gamma (Yang et al., 2010) or gamma-GEV distribution (Smith et al., 2014) can be more effective. There are two reasons for our choice here. Firstly, we compared the single gamma with double gamma and gamma-GEV distributions, and obtained almost similar performance scores according to Mean Squared Error. Secondly, the bias correction in this study is performed for each grid, each lead time and each variable. Given the heavy computation ~~labor~~labour, the single gamma distribution is selected here for timesaving and efficiency. For MXT and MNT, four-~~-~~parameter beta distribution is utilized as suggested by Li et al. (2010). Owing to the limited record of ECMWF ~~forecast,the~~forecast, the data excluded the forecast year is used as training data to determine the parameters of QM.

Since forecasts are post-processed for individual lead time, grid and variable, the forecast ensembles therefore tend to be inappropriately space-~~-~~time correlated. To generate ensemble members with appropriate space-~~-~~time correlations, Schaake shuffle (Clark et al., 2004) is applied to link historical data to ensemble members and to create sequences with realistic temporal-spatial patterns. 38 years of historical data from 1978onward is used to apply the Schaake Shuffle procedure. Details to conduct Schaake Shuffle can be found in Clark et al. (2004) and Schepen et al. (2017).

3.5 Evaluation indicators

The annual maximum flood is picked out as typical flood events. Meanwhile, the first flood event in each year is also selected. Maximum flood is determined by the maximum daily streamflow in a year~~one-day peak flow~~. For first flood, the definition seems to be slightly subjective. ~~Nevertheless,first~~Nevertheless, first flood is just introduced as an example to verify the skill of VIC/ECMWF system to forecast the meltwater components. There are three criterions for us to define the first flood: (1) the peak flow should be more than twice of the average daily streamflow during dry period (November to March); (2) the duration for the flood event should be longer than 7 days; (3) the observed snowpack is present. Forecasts are issued for each chosen ~~Considering~~event. Considering that maximum flood events in YZR usually last for several months, flood volume over the entire flood event is impossible to be covered by medium-range weather forecasts. Four typical flood volumes are therefore chosen to represent the volume performance, i.e. the peak flow (Q1), the accumulative 3-days flows ~~centered~~centred on peak flow (Q3), the accumulative 5-days flows ~~centered~~centred on peak flow(Q5~~),the~~), the accumulative 7-days flows ~~centered~~centred on peak flow (Q7). Term “duration” is adopted to represent the number of days used to generate flood volumes. The Continuous Ranked Probability Skill Score (CRPSS) (Hersbach, 2000) is adopted to indicate the overall performance of the forecasts as a comprehensive evaluation metric, which is calculated via normalizing the Continuous Ranked Probability Score (CRPS) by a reference forecast. The reference forecast in this study is an ensemble of hydrological forecasts simulated by the VIC model using sampled historical meteorological observations at the same calendar day as input to the model (Bennett et al., 2014). For deterministic forecasts, the CRPS score reduces to Mean Absolute Error (MAE), and can be directly compared. CRPS and MAE are negatively oriented and tend to increase with forecasts bias or poor reliability (Shrestha et al., 2015). The value of CRPSS ranges from -∞ to 1, with best score equal to 1.

Two specialized indicators for flood events are utilized according to works by Smith et al. (2004), i.e., the percent absolute flood volume error E_q and percent absolute peak time error E_t. The definitions are in formulas (8)~~—~~-(9):

$$E_q = \frac{\sum_{i=1}^{N} |B_i|}{N Y_{avg}} \times 100 \qquad (8)$$

$$E_t = \frac{\sum_{i=1}^{N} |T_{pi} - T_{psi}|}{N} \times 100 \qquad (9)$$

where B_i is the volume bias for ith flood event; Y_{avg} is the average observed flood volume for N selected flood events. T_{pi} and T_{psi} are the observed and simulated time to ith peak.

4. Results

In this study, the performance of N-simulations and S-simulation in simulating and forecasting floods is analysed. Moreover, the results for forecasting different streamflow components are also shown.

Formatted: Font: (Asian) +Body Asian (宋体), (Asian) Chinese (PRC)

Formatted: Normal

4.1 Hydrological model performance

~~As a result of unavailability of observed streamflow components, the evaluation of streamflow components has to be donebased on the facts that the total runoffis accurately simulated and forecasted, and that the ratio of meltwater-induced streamflow is similar to records in previous studies. Hence, a general assessment on total runoff simulations and forecasts is crucial and thus done first in Subsection 4.1 and 4.2.~~

Fig. 2 shows an example of two-dimensional Pareto plots for Bias and NSE at Nugesha Station. The performance of the selected N-Pareto-optimal parameter sets and single compromise parameter set during calibration and evaluation periods for three hydrological stations are listed in Table 1. ~~Generallyspeaking~~Generally speaking, the model performance during evaluation is more satisfactory than that during calibration. It is probably caused by the existence of considerable extraordinary flood events during the calibration period. The relatively shorter timespan during evaluation may be also one of the reasons. It is noticeable that simulation at Nugesha is better than that at other two stations with NSE greater than 0.77 for daily streamflow and NSE near 0.5 for top 10% streamflow. Performance at Nuxia is inferior with bias greater than 30%, which is similar ~~as~~ to the previous studies by Tong et al. (2014) and Zhang et al. (~~2012~~2013). They both claimed that the underestimation in streamflow simulation at Nuxia ~~is~~ was highly likely to be caused by the largely underestimated CMA observations in this area~~, and w~~. We also guess ~~it is also a reason~~ that within downstream regions the hydrological process becomes ~~more~~ too complicated due to human activities to be simulated by models (Li et al., 2013; Liu et al., 2014). ~~The bolded values in the table are the cases where S-simulation behaves better than N-simulations according to the selected objective functions.~~ It is obvious that ~~in these cases~~ S-simulation generally performs well during calibration, whilst during the evaluation period S-simulation loses the advantage to some degree. ~~The NSE of top 10% streamflow at Nugesha is the only case where S-simulation consistently outperforms the N-simulations in either calibrationor evaluation period.~~

[Figure 2]

The observed and simulated hydrographs during the evaluation period at ~~Nugesha~~Nuxia are presented in Fig. 3. An obvious underestimation can be observed in low flow periods~~.~~, which is similar to previous stud~~y~~ies by Tong et al. (2014) and Zhang et al. (2013). The absence of glacier module in VIC is believed to has limited influence on this underestimation, for similarly underestimated low flow was found when glacier modelling was embedded in VIC (Zhang et al., 2013). ~~There are two possible reasons for this phenomenon. Firstly,~~For our study, the underestimation is, ~~at~~in the meanwhile, caused by the ~~reason~~ fact that the objective functions used for calibration have the tendency to give more attention to high flows, as the flood is the focus of our investigation. ~~Secondly, the absence of glacier module in VIC is believed todeteriorate model performance in some way, in particular in low flow periods.~~ As ~~noticed~~ revealed in Fig. 3, the flood peaks are well captured by S-simulation in most cases. N-simulations are able to cover all the extreme values while sometimes slight overestimation exists.

[Figure 3]

The indicators for typical flood volumes simulated by VIC for first floods and maximum floods during the whole study period are listed in Table 2. Two statistical indictors are adopted here, i.e., CRPS for N-simulations and MAE for S-simulation. For Nugesha and Yangcun, CRPS is consistently smaller than MAE, indicating better simulation by N-simulations. The improvement is about 10% compared with S-simulation and tends to be greater for longer durations. On the contrary, S-simulation at Nuxia consistently provides better performance than N-simulations, for the selected single parameter set for S-simulation at this station is actually the best parameter set for three of the objective functions, which can be viewed in Table 1. ~~Those~~ The different behaviours at three stations imply that N-simulations is preferable when the trade-offs between multi-objective functions are significant (no single parameter set behaves well in most of the objectives like Nuxia Station~~).In~~). In order to present more detailed performance of flood volume, Fig. 4 exhibits simulated flood volumes versus observations for maximum floods during the evaluation period at three stations. It is noticeable that the majority of the flood events can be captured by N-simulations, and volumes tend to be better covered with duration increasing. The flood volume at Yangcun is better simulated than that at the other two stations. It is consistent to the highest NSE for top 10% streamflow at this station as shown in Table 1. The floods at Nuxia are obviously underestimated. In most cases, the N-simulations even fail to cover the observations. Similar but better behaviours exist for first floods and thus omit here.

[Table 1] [Table 2] [Figure 4]

VIC simulated snow cover is compared with snow depth derived from passive microwave remote-sensing data. Fig. 5 shows the spatial distribution of observed and simulated daily average snow depths during evaluation. For simplicity, only the results at Nuxia is displayed. An acceptable agreement (Correlation Coefficient = 0.63) can be found over the entire domain, especially for the middle reaches. Some overestimation exists in the upstream and downstream regions. Explanation for these errors in snow depth will be furtherly described in Section 5. We also compare the fraction of meltwater-induced components to total runoff with previous studies (Liu, 1999; Cuo et al., 2014) as shown in Table 3. It is noticeable that the results by S-simulation are quite close to the records, except Yangcun with about 5% overestimation. Most of the records are covered by N-simulations. However, all the parameter sets slightly underestimate the meltwater streamflow at Nuxia.

[Figure 5] ~~–~~[Table 3]

4.2 Flood volumes forecasts

Streamflow forecasts are driven by QM-SS post-processed QPF and temperature data. A preliminary analysis of raw and post-processed ECMWF forecasts reveals that QM-SS is effective for reducing errors and the post-processed forecasts are skilful enough for streamflow forecasting (seeing S.1 in supporting information). Fig.6 displays the CRPSS values of different flood volumes at three hydrological stations. Lead times of day 3, 5, 7, 10, 12 and 14 are chosen as representative to trace the forecast ~~quality.Generally~~quality. Generally, flood volumes tend to be better captured with the increase of duration. One reason is that there are often larger errors in simulated flood ~~peak,making~~peak, making the single day flood volume more prone to bias. Another reason is that when the duration increases, the bias ~~from~~ in streamflow for this relatively long period can offset with each other. Performance of the VIC/ECMWF system deteriorates with increasing lead time as expected. The lead time of skilful forecasts for first floods is shorter than maximum floods. This can be explained by the generation mechanism of first floods. First floods are usually dominated by baseflow and meltwater. Compared with maximum floods, first floods normally occur in the same period within one year, so historical meteorological observations on the same calendar day can provide skilful input. This fact results in a reference forecast which is hard to beat. As for maximum floods, streamflow can be predicted at least 10 days ahead. Similar to Table 2, forecasts driven by S-simulation gain higher CRPSS at Nuxia, while for the other two stations, performance of S-simulation and N-simulations varies with lead time and duration. It seems that N-simulations gradually lose the advantage with increasing lead times, which may have something to do with the superposition of interaction of model parameter errors and meteorological forecasts uncertainty.

[Figure 6]

Another statistical indicator computed from forecasted flood volumes driven by S-simulation and N-simulations is illustrated by boxplots in Fig. 7 for first floods and Fig. 8 for maximum floods. For simplicity, only Q1 and Q7 are displayed, and an overall progressive improvement can be found from Q1 to Q7. As can be perceived, E_q increases with lead time, but for longer lead times a decrease exists. The decrease begins from day +10for first floods and day +12 for maximum floods. Meanwhile, whiskers of boxplots become wider and wider with the increase of lead time, indicating larger degree of variability over lead times. For Nugesha, the percent absolute flood volume error is found to be 40% on average. Greater E_q in Yangcun is highly likely to be caused by insufficient modelling ability of VIC at this station for the NSE value for entire streamflow is only up to 0.65. Lower variances can be found for Nuxia Station. Regarding comparisons between S-simulation and N-simulations, we can observe that for first floods (Fig. 7) S-simulation outperforms N-simulations with smaller median and narrower whiskers, and in terms of Q7, the difference becomes minor especially for longer lead times.

[Figure 7]

As demonstrated in Fig. 8, E_q for maximum floods is smaller than that for first floods, with majority of the streamflow errors confined within 40%. Unlike first floods, maximum floods are dominated by the precipitation inputs during a relevant period. Accordingly, the influence from hydrological errors becomes minor. Additionally, maximum floods as high flow events are intensively calibrated. For Nugesha (Fig.8 (a)-(b)), the most upstream station, E_q is greater in the beginning which is possibly caused by the shorter response time and thus greater influence from hydrological errors. Undeniably, model with NSE<=0.48 for high flows indeed impairs skilful forecasts. The smallest E_q at Yangcun in short lead time is attributed to the minor hydrological errors for VIC with NSE up to 0.73 for top 10% streamflow. Although on average performance level, forecasts derived from S-simulation certainly have smaller errors, certain cases exist where part of the members in N-simulations have the ability to provide forecasts with the smallest errors. Moreover, the differences for two simulation modes become smaller compared to first floods, for the errors in streamflow forecasts are dominated by errors in ECMWF forecasts for maximum flood events.

[Figure 8]

The errors in peak time prediction are displayed in Fig. 9. The left sides are subplots for first floods, and the results for maximum floods are shown in right-hand subplots. Similar to E_q, $E_t$deteriorates with lead time and peaks at lead time of 10 day. The peak time errors at three stations are about 1-5 days for both first floods and maximum floods, yet errors in maximum floods are larger than that of first floods. An average of 2 days in E_t is found for first floods at Yangcun and larger errors are present in other two stations.E_t of first floods at Nugesha is the largest, and the cause is similar to E_q. As for maximum flood events, an obvious increase in E_t from day +7 can be perceived. Performance of S-simulation and N-simulations in this round varies with flood categories and stations, but generally smaller errors are found in peak times forecasted by N-simulations, especially for maximum floods.

[Figure 9]

4.3Streamflow components forecasts

~~From analysis above, anencouragingly accurate VIC simulationand forecastingsystem is established in YZR. This is an important precondition for subsequent evaluation of streamflow components. In some cases, for example first floods at~~

~~Yangcun, VIC fails to produce accurate enough simulations and thus poor forecasts (Fig. 6b and Fig. 7c). When discussing streamflow components in these circumstances, we should evaluate model performance more carefully with consideration of errors possibly stemming from hydrological model, although the comparison of streamflow component forecasts is performed based on simulated components andto some degree, the errors in forecasts are mainly subject to meteorological inputs and therefore the hydrological error becomes negligible.~~This subsection presents results of N-simulations and S-simulation in forecasting streamflow components~~ of interest~~.

Figures 10-12 show CRPSS of meltwater-induced and rainfall-induced volumes at three hydrological stations. The reference forecasts used to compute CRPSS are forecasts driven by the same parameter set with inputs of historical observations at the same calendar day. Thus, CRPSS here is ~~just~~ an indicator to show the forecast skill against lead time and to present the errors only from meteorological data. Only the results for Q1 are presented, as the results show no obvious correlations with duration. From Fig. 10, it is noticeable that for first floods at Nugesha, errors in forecasting surface runoff components is the main source contributing to errors in forecasting total runoff. Forecast skill for baseflow components seems to be insensitive to lead time (Figs. 10 a-b). On one hand, ~~thereason may be that~~these components are mainly generated by available water storage in the catchment. On the other hand, the baseflow process often evolves slowly, possibly making the forecast lead time not able to cover the base flow variability. As for maximum floods, the errors derived from surface runoff forecasts are similarly the main contributor to errors in total runoff forecasts, but the baseflow exhibits a similar tendency with surface runoff and total runoff, deteriorating with lead times as shown in Figs. 10 c-d. This means during the period of maximum floods the infiltration is substantial in VIC and makes the moisture in bottom soil layer vary with the rainfall and meltwater inputs. The information in Figs. 10c-d is in good agreement with results displayed in Fig.6. Fluctuating CRPSS in Q_{snow} and Q_{rain} results in similarly fluctuating CRPSS in Q.The well-predicted Q_{rain} component is the critical factor for high CRPSS for total runoff. The meltwater-induced components can be predicted with 7 days in advance for first floods, and the lead time is much shorter for maximum floods. The rainfall-induced components can be ~~skillfully~~skilfully forecasted up to day +14 compared with reference forecasts.

[Figure 10]

Similar performance can be found at Yangcun as shown in Fig. 11. Baseflow components for first floods are consistently well reproduced by the system with CRPSS greater than 0.8 for all the lead times. The variation in total runoff is fairly consistent with surface runoff. However, higher CRPSS in both Q_{snow} and Q_{rain} fails to give birth to higher CRPSS in Q (shown in Fig.6b). According to Table 1, the MAE value for S-simulation is $258.64\ m^3/s$ for Q1, and the average observed peaks during this period is about $630\ m^3/s$. Hence, the errors in hydrological model are too large to capture the actual flood process. The high CRPSS here is caused by the exclusion of hydrological errors. With regard to maximum floods, errors in surface runoff is still the main contributor to errors in total runoff. The meltwater-related components are forecasted with short lead time as Nugesha. Results from S-simulation totally fall out of the 95% confidence ~~interval,while~~interval, while for rainfall-induced components, S-simulation produces higher CRPSS for lead time longer than 10 days.

The most noticeable phenomenon at Nuxia is that baseflow components for first floods at this station exhibit an obvious deterioration with lead times (Fig. 12a-b). Nuxia is located in the most downstream reaches, and concentrates water from hundreds of tributaries. Some tributaries are fairly small with rapid response of baseflow and surface runoff and some tributaries may have intensive interactions between entire soil layer, causing the baseflow in the outlet to vary with lead time. CRPSS of all the flood components has similar changes to scores of total ~~runoff~~runoffs in Fig.6. Generally, the Q_{snow} and Q_{rain} forecasts are skillful in lead time of 7 day and 10 day, respectively. Surface runoff remains the toughest part for forecasts, in which the meltwater-induced components can be predicted in only 5 days ahead.

[Figure 11] [Figure 12]

5. Discussion

In this study, N-Pareto-Optimal parameter sets were adopted to solve the multiple feasible solutions by multi-objective optimization. Before NWP was introduced into the flood forecasting system, the streamflow driven by N-simulations was better simulated than that by S-simulation as shown in Table 2, although the NSE and Bias value are more ~~favorable~~favourable for S-simulation during calibration. When it comes to flood forecasting, neither of the outputs by these two simulation modes has overwhelming advantages over every aspect of forecasting, which coincides with the conclusion from a previous study by Zhu et al. (2016b). Three preliminary findings were made for N-simulations. Firstly, N-simulations generally behave better when the trade-offs in multi-objectives are significant. In this case, the N-simulations can synthesize advantages from different components. This is why N-simulations can provide more desirable skill at Nugesha than Yangcun. Secondly, N-simulations indeed improves the streamflow simulation as shown in Table 1 and 2, but when it comes to forecasting, the interaction of errors in hydrological model parameters and meteorological forecast may degrade the forecast skill at longer lead time (Fig. 6). Last, N-simulations may fail to provide better results on average model performance level, but individual member in N-Pareto-Optimal parameter sets can capture the events with the lowest errors.

As there is no glacier module in the current VIC model, similar to previous studies (Li et al., 2014; Liu et al., 2014; Sun et al., 2013), the glacier-related process was considered together with the snow in this study. In other words, the rainfall input into VIC is separated into only two components, the liquid(rainfall) and solid parts (snow), and the portion of rainfall which is supposed to turn to glacier/ice is treated as snow instead. That is why the snow depth simulated by VIC is somewhat ~~obviously~~higher than that of remote sensing data shown in Fig. 5 while the melt water proportion is close to the records (Table 3)~~, for the output snow depth from VIC is actually the sum of snow and glacier/ice~~. Additionally, ~~compared~~ comparing with the distribution of used meteorological stations shown in Fig. 1, we can infer that these positive biases were also induced by the interpolation using data from stations at which there are more snow/glacier present. To verify our conclusion, we plot the VIC-simulated snow depth together with the distribution of glacier in the YZR basin. The glacier data is downloaded from The Second Glacier Inventory Dataset of China (http://westdc.westgis.ac.cn/data/f92a4346-a33f-497d-9470-2b357ccb4246). From Fig. 13, it is noticeable that the locations of overestimation do coincide with the locations of glacier. For Zone 1 and

Zone 2, the overestimation is exacerbated by interpolating with gauges at which more snow and glacier exist. To relieve this problem, there are generally two ways to consider glacier-melt separately: temperature-index models to quantify an empirical relationship between air temperature and melt rate (Su et al., 2016; Zhang et al., 2013)~~energy balance models to calculate melt as residual in the heat balance equation~~, and energy balance models to calculate melt as residual in the heat balance equation (Zhao et al., 2013) ~~temperature-index models to quantify an empirical relationship between air temperature and melt rate (Zhang et al., 2013)~~. However, the error, as a result of overly sparse meteorological network, will consistently and largely hamper the ~~applicationof~~application of those complicated methods (Tong et al., 2014).~~Some studies have successfully coupled VIC with these two kinds of glacier-melt models. Zhang et al. (2013) and Su et al. (2016) embodied a simple degree-day glacier algorithm into VIC, and Zhao et al. (2013) coupled an energy balance-based glacier model with VIC, showing acceptable performance with efficiency coefficient greater than 0.8 for the complete simulation period. However, one thing we should bear in mind is that with the limited observed data in this special area, it is difficult to accurately separate the snow from glacier. Overly complicated methods probably bring out more uncertainties. In that sense, simply dividing the input rainfall into two parts can be an acceptable way.When more observations are available with the development of technologies in the future, more elaborate separation method is expected.~~

[Figure 13]

For streamflow components forecast, the biggest challenge is the absence of data series of in-situ streamflow components. Therefore, in this study the simulation driven by ~~observations~~ observed forcing becomes an alternative to act as proxy ~~although it is difficult to determine whether such proxy is believable or not~~and thus the error stemming from hydrological model is avoided~~, which~~. This is a common practice when observation is absent (Arnal et al., 2018; Harrigan et al., 2018). ~~By using proxy~~Without calibration of specific streamflow components, conclusion simply based on simulation of single parameter set ~~is~~ may be risky~~. Similar to hydrological ensemble prediction,, ensemble~~and ensemble from multi-parameter sets is believed to be more confident with consideration of hydrological uncertainty. From our results, different parameter sets behave ~~consistently~~ similarly in streamflow components forecast, i.e. deteriorating with increasing lead time. However, when it comes to specific skill score, slight differences can be viewed from Figs. 10-12. Sometimes, S-simulation provides skil~~l~~ful forecasts with longer lead time, while in some other cases, performance of S-simulation becomes inferior and falls out of the 95% CI. Single parameter set may present overestimation or underestimation to some degree. ~~From the view of ensemble, the phenomenon captured by more parameter sets is regarded as the most possible occurrence. Single parameter set may present overestimation or underestimation to some degree.~~

The meltwater-induced components in streamflow are found to be difficult for the system to forecast, in which those in surface runoff are the toughest part. This is reasonable since the surface runoff is the most susceptible variable to various hydrometeorological factors. Specifically, R_{snow} in the study area is mainly determined by the amount of snowfall and the temperature at which the snowpack begins to melt. In VIC, the input precipitation is separated into snowfall and rainfall according to a predefined temperature. In consequence, errors from all the ECMWF forecasts would affect the R_{snow} forecasts whilst R_{rain} is merely influenced by one meteorological input, QPF. This is also the reason why rainfall-induced streamflow forecasts are the major contributor to satisfactory forecasting. This illustrates the importance of components study.

6. Conclusion

In this study, a hydrological ensemble prediction system composed by VIC and ECMWF medium-range precipitation and temperature forecasts was developed and applied in the YZR Basin to investigate the forecasting performance of flood volumes and streamflow components. Two different simulation modes were adopted. One is S-simulation which is driven by conventional single parameter set, and the other one is N-simulations which is driven by an ensemble of parameter sets selected from the Pareto front using the Preference Ordering Routine method. A newly published hydrograph separation algorithm was employed to separate the streamflow into four individual components: the surface runoff and baseflow induced by rainfall and meltwater respectively. The findings are summarized as following:

(1) N-simulations was proven to be superior in model simulation. For flood forecasting, the performance of N-simulations and S-simulation varies with lead time and basin scale, and N-simulations is recommended when the multi-objective trade-offs are significant. When lead time extends, the differences between N-simulations and S-simulation become minor.

(2) Flood forecast skill deteriorates with lead time. The forecast skill of flood volume increases with duration. Q7 can be better captured than Q1. The forecasting system provides better forecasts for maximum floods than first floods. The flood volume of first floods can be predicted in 7-14 days in advance. The lead time for maximum floods is 10-14 days.

(3) At Nugesha and Yangcun stations, base flow components tend to be insensitive to increase of lead time due to the slowly-evolved baseflow process. At Nuxia Station, baseflow exhibits similar patterns to total runoff.

(4) Meltwater-induced component in surface runoff is the most difficult part for the proposed system to forecast compared with reference forecasts, which can only be captured in 4-7 days. Well-forecasted rainfall-induced streamflow is the main contributor for successful flood forecasting.

Author contribution

Suli Pan provides methodology used to bias correct the raw ECMWF forecasts. Zhixu Bai helped to develop the model code. Yue-Ping Xu guided and supervised the ~~research~~study. Li Liu performed the simulation and prepared the manuscript with contributions from all co-authors.

Acknowledgement

This study is financially supported by National Natural Science Foundation of China (91547106) and National Key Research and Development Plan "Inter-governmental Cooperation in International Scientific and Technological

Innovation"(2016YFE0122100). National Climate Center of China Meteorological Administration and Hydrology and Water Resource Bureau of Tibet are greatly acknowledged for providing meteorological and hydrological data used in the study area. QPFs and temperature forecasts were obtained from ECMWF's TIGGE data portal. ~~Thanks~~Thanks are also given to ECMWF for the development of this portal software and for the archives of this immense ~~dataset.We~~dataset. We would like to acknowledge the editors and reviewers for their reviews and very constructive feedback.

Reference

Ahl, R. S., Woods, S. W., and Zuuring, H. R.: Hydrologic calibration and validation of SWAT in a snow-dominated Rocky Mountain watershed, Montana, USA, J. Am. Water Resour., 44, 1411-1430, doi:10.1111/j.1752-1688.2008.00233.x, 2008.

Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., and Leahy, C.: Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes, Hydrol. Earth Syst. Sci., 19, 1659-1676, doi:10.5194/hess-19-1659-2015, 2015.

Aminyavari, S., Saghafian, B. and Delavar, M.: Evaluation of TIGGE Ensemble Forecasts of Precipitation in Distinct Climate Regions in Iran, Adv. Atmos. Sci., 35(4), 457-468, 2018.

Andreadis, K. M., Storck, P., and Lettenmaier, D. P.: Modeling snow accumulation and ablation processes in forested environments, Water Resour. Res., 45, W05429, doi:10.1029/2008WR007042, 2009.

Arheimer, B., Lindström, G., and Olsson, J.: A systematic review of sensitivities in the Swedish flood-forecasting system, Atmos. Res.,100, 275-284, doi.org/10.1016/j.atmosres.2010.09.013, 2011.

Arnal, L., Cloke, H.L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B. and Pappenberger, F.,: Skilful seasonal forecasts of streamflow over Europe? Hydrol. Earth Syst. Sci., 22(4), 2057-2072, 2018.

Bartholmes, J. C., Thielen, J., Ramos, M. H., and Gentilini, S.: The european flood alert system EFAS–Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrol. Earth Syst. Sci., 13, 141-153, https://doi.org/10.5194/hess-13-141-2009, 2009.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47-55, doi:10.1038/nature14956, 2015.

Bavera, D., De Michele, C., Pepe, M., and Rampini, A.: Melted snow volume control in the snowmelt runoff model using a snow water equivalent statistically based model, Hydrol. Process., 26, 3405-3415, DOI: 10.1002/hyp.8376, 2012.

Bennett, J. C., Robertson, D. E., Shrestha, D. L., Wang, Q. J., Enever, D., Hapuarachchi, P., and Tuteja, N. K.: A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days, J. Hydrol., 519, 2832-2846, https://doi.org/10.1016/j.jhydrol.2014.08.010, 2014.

~~Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y., and Wei, M.: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, Mon. Weather Rev., 133, 1076-1097, https://doi.org/10.1175/MWR2905.1, 2005.~~

Formatted: No underline, Font color: Auto

Che, T., Xin, L., Jin, R., Armstrong, R., and Zhang, T.: Snow depth derived from passive microwave remote-sensing data in China, Annals of Glaciology, 49, 145-154, DOI:10.3189/172756408787814690, 2008.

Chen, X., Long, D., Hong, Y., Zeng, C. and Yan, D.: Improved modelling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin? Water Resour. Res., 53, 2431-2466, doi:10.1002/2016WR019656, 2017.

Cheng, G., and Wu, T.: Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau, J. Geophys. Res.,112, F02S03, doi:10.1029/2006JF000631, 2007.

Cherkauer, K. A., and Lettenmaier, D. P.: Simulation of spatial variability in snow and frozen soil, J. Geophys. Res., 108, 8858, DOI:10.1029/2003JD003575, 2003.

Cherkauer, K. A., and Lettenmaier, D. P.: Hydrologic effects of frozen soils in the upper Mississippi River basin, J. Geophys. Res., 104, 19599-19610, DOI: 10.1029/1999JD900337, 1999.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields, J. Hydrometeorol., 5, 243-262, 2004.

Cloke, H. L., Pappenberger, F., van Andel, S. J., Schaake, J., Thielen, J., and Ramos, M. H.: Hydrological Ensemble Prediction Systems (HEPS). Preface to special issue, Hydrol. Process., 27, 1-4, DOI: 10.1002/hyp.9679, 2013.

Cloke, H. L., Pappenberger, F., Smith, P. J., and Wetterhall, F.: How do I know if I've improved my continental scale flood early warning system? Environ. Res. Lett., 12, 044006, https://doi.org/10.1088/1748-9326/aa625a, 2017.

Cuo, L., Zhang, Y., Gao, Y., Hao, Z., and Cairang, L.: The impacts of climate change and land cover/use transition on the hydrology in the upper Yellow River Basin, China, J. Hydrol., 502, 37-52, https://doi.org/10.1016/j.jhydrol.2013.08.003, 2013a.

Cuo, L., Zhang, Y., Wang, Q., Zhang, L., Zhou, B., Hao, Z., and Su, F.: Climate change on the northern Tibetan Plateau during 1957–2009: Spatial patterns and possible mechanisms, J. Clim., 26, 85-109, https://doi.org/10.1175/JCLI-D-11-00738.1, 2013b.

Cuo, L., Zhang, Y., Zhu, F., and Liang, L.: Characteristics and changes of streamflow on the Tibetan Plateau: A review, J. Hydrol., 2, 49-68, https://doi.org/10.1016/j.ejrh.2014.08.004, 2014.

Dai, L., Che, T., and Ding, Y.: Inter-calibrating SMMR, SSM/I and SSMI/S data to improve the consistency of snow-depth products in China, Remote Sens., 7, 7212-7230, doi:10.3390/rs70607212, 2015.

Dan, L., Ji, J., Xie, Z., Chen, F., Wen, G. and Richey, J.E.: Hydrological projections of climate change scenarios over the 3H region of China: A VIC model assessment, J. Geophys. Res., 117, D11102, doi:10.1029/2011JD017131, 2012.

Dittmann, R., Froehlich, F., Pohl, R., and Ostrowski, M.: Optimum multi-objective reservoir operation with emphasis on flood control and ecology, Nat. Hazard Earth Sys., 9, 1973-1980, https://doi.org/10.5194/nhess-9-1973-2009, 2009.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., and Burlando, P.: The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, Water Resour. Res., 47, W07519, DOI: 10.1029/2010WR009824, 2011.

~~Froude, L. S.: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. Weather Forecast., 25, 819-836, https://doi.org/10.1175/2010WAF2222326.1, 2010.~~

Girons Lopez, M., Di Baldassarre, G., and Seibert, J.: Impact of social preparedness on flood early warning systems, Water Resour. Res., 53, 522-534, DOI: 10.1002/2016WR019387, 2017.

Guo, D., and Wang, H.: The significant climate warming in the northern Tibetan Plateau and its possible causes, Int J Climatol., 32, 1775-1781, doi:10.1002/joc.2388, 2012.

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., Ye, A. and Miao, C.: Multi-objective parameter optimization of common land model using adaptive surrogate modeling. Hydrol. Earth Syst. Sci., 19, 2409-2425, doi:10.5194/hess-19-2409-2015, 2015.

Hamill, T. M.: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States, Mon. Weather Rev., 140, 2232-2252, https://doi.org/10.1175/MWR-D-11-00220.1, 2012.

Hamill, T.M. and Scheuerer, M.: Probabilistic Precipitation Forecast Postprocessing Using Quantile Mapping and Rank-Weighted Best-Member Dressing, Mon. Weather Rev., 146(12), 4079-4098, 2018.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K. and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrol. Earth Syst. Sci., 22(3), 2023-2039, 2018.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559-570, 2000.

Kalra, A., Li, L., Li, X., and Ahmad, S.: Improving streamflow forecast lead time using oceanic-atmospheric oscillations for Kaidu River Basin, Xinjiang, China, J. Hydrol. Eng., 18, 1031-1040, 2012.

Kamali, B., Mousavi, S. J., and Abbaspour, K. C.: Automatic calibration of HEC-HMS using single-objective and multi-objective PSO algorithms, Hydrol. Process., 27, 4028-4042, DOI: 10.1002/hyp.9510, 2013.

Khu, S. T., and Madsen, H.: Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration, Water Resour. Res., 41, W03004, DOI: 10.1029/2004WR003041, 2005.

Kuang, X. and Jiao, J.J.: Review on climate change on the Tibetan Plateau during the last half century, J. Geophys. Res., 121, 3979-4007, doi:10.1002/2015JD024728, 2016.

Laudon, H., Hemond, H. F., Krouse, R., and Bishop, K. H.: Oxygen 18 fractionation during snowmelt: Implications for spring flood hydrograph separation, Water Resour. Res., 38, 1258, DOI: 10.1029/2002WR001510, 2002.

Li, D., Wrzesien, M. L., Durand, M., Adam, J., and Lettenmaier, D. P.: How much runoff originates as snow in the western United States, and how will that change in the future? Geophys. Res. Lett., 44, 6163–6172, doi:10.1002/2017GL073551, 2017.

Li, F., Xu, Z., Feng, Y., Liu, M. and Liu, W.: Changes of land cover in the Yarlung Tsangpo River basin from 1985 to 2005, Environ.Earth Sci., 68, 181-188, ,2013.

Li, F., Xu, Z., Liu, W., and Zhang, Y.: The impact of climate change on runoff in the Yarlung Tsangpo River basin in the Tibetan Plateau, Stoch. Env. Res. Risk A., 28, 517-526, DOI: 0.1007/s00477-013-0769-z, 2014.

Li, H., Sheffield, J., and Wood, E. F.: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, J. Geophys. Res., 115, D10101, doi:10.1029/2009JD012882, 2010.

Liu, L., Gao, C., Xuan, W., and Xu, Y. P.: Evaluation of medium-range ensemble flood forecasting based on calibration strategies and ensemble methods in Lanjiang Basin, Southeast China, J. Hydrol., 554, 233-250, 2017.~~.~~ ~~https://doi.org/10.1016/j.jhydrol.2017.08.032, 2017.~~

Liu, T. C.: Hydrological characteristics of Yarlung Zangbo River, Acta Geographica Sinica, 54(Suppl 1), 157-164, 1999 (In Chinese).

Liu, Z., Tian, L., Yao, T., Gong, T., Yin, C., and Yu, W.: Temporal and spatial variations of δ 18 O in precipitation of the YarlungZangbo River Basin, J. Geogr. Sci., 17, 317-326, DOI: 10.1007/s11442-007-0317-1, 2007.

Liu, Z., Yao, Z., Huang, H., Wu, S., and Liu, G.: Land use and climate changes and their impacts on runoff in the Yarlung Zangbo river basin, China, Land Degrad. Dev., 25, 203-215, doi:10.1002/ldr.1159, 2014.

Liang, X., Lettenmaier, D. P., and Wood, E. F.: One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, J. Geophys. Res., 101, 21403-21422, doi:10.1029/96JD01448, 1996.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99, 14415-14428, doi:10.1029/94JD00483, 1994.

Louvet, S., Sultan, B., Janicot, S., Kamsu-Tamo, P.H. and Ndiaye, O.: Evaluation of TIGGE precipitation forecasts over West Africa at intraseasonal timescale, Clim. Dynam., 47(1-2), 31-47, DOI 10.1007/s00382-015-2820-x, 2016.

Luo, Y., Arnold, J., Allen, P., and Chen, X.: Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China, Hydrol. Earth Syst. Sci., 16, 1259-1267, https://doi.org/10.5194/hess-16-1259-2012, 2012.

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., and Thielen, J.: The monetary benefit of early flood warnings in Europe, Environ. Sci. Policy, 51, 278-291, https://doi.org/10.1016/j.envsci.2015.04.016, 2015.

Parajka, J., and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, J. Hydrol., 358, 240-258, https://doi.org/10.1016/j.jhydrol.2008.06.006, 2008.

Partington, D., Brunner, P., Simmons, C. T., Therrien, R., Werner, A. D., Dandy, G. C., and Maier, H. R.: A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water–groundwater flow models, Environ. Model Softw., 26, 886-898, https://doi.org/10.1016/j.envsoft.2011.02.007, 2011.

Salathé Jr, E. P., Hamlet, A. F., Mass, C. F., Lee, S. Y., Stumbaugh, M., and Steed, R.: Estimates of twenty-first-century flood risk in the Pacific Northwest based on regional climate model simulations, J. Hydrometeorol., 15, 1881-1899, https://doi.org/10.1175/JHM-D-13-0137.1, 2014.

Schepen, A., Zhao, T., Wang, Q. J., and Robertson, D. E.: A new method for post-processing daily sub-seasonal to seasonal rainfall forecasts from GCMs and evaluation for 12 Australian catchments, Hydrol. Earth Syst. Sci. Discuss., 1-27, https://doi.org/10.5194/hess-2017-380, 2017.

Formatted: Font: Not Bold, No underline, Font color: Auto

Formatted

Shen, W., Li, H., Sun, M., and Jiang, J.: Dynamics of aeolian sandy land in the Yarlung Zangbo River basin of Tibet, China from 1975 to 2008, Glob. Planet. Change, 86, 37-44, https://doi.org/10.1016/j.gloplacha.2012.01.012, 2012.

Shi, H., Li, T., Liu, R., Chen, J., Li, J., Zhang, A., and Wang, G.: A service-oriented architecture for ensemble flood forecast from numerical weather prediction, J. Hydrol., 527, 933-942, https://doi.org/10.1016/j.jhydrol.2015.05.056, 2015.

Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., and Hapuarachchi, H. A. P.: Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose, Hydrol. Earth Syst. Sci., 17, 1913-1931, https://doi.org/10.5194/hess-17-1913-2013, 2013.

Shrestha, M., Koike, T., Hirabayashi, Y., Xue, Y., Wang, L., Rasul, G., and Ahmad, B.: Integrated simulation of snow and glacier melt in water and energy balance-based, distributed hydrological ~~modeling~~modelling framework at Hunza River Basin of Pakistan Karakoram region, J. Geophys. Res., 120, 4889-4919, doi: 10.1002/2014JD022666, 2015.

Siderius, C., Biemans, H., Wiltshire, A., Rao, S., Franssen, W. H. P., Kumar, P., et al.: Snowmelt contributions to discharge of the Ganges, Sci. Total Environ., 468, S93-S101, https://doi.org/10.1016/j.scitotenv.2013.05.084, 2013.

Smith, A., Freer, J., Bates, P., and Sampson, C.: Comparing ensemble projections of flooding against flood estimation by continuous simulation, J. Hydrol., 511, 205-219, https://doi.org/10.1016/j.jhydrol.2014.01.045, 2014.

Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., et al: The distributed model intercomparison project (DMIP): motivation and experiment design, J. Hydrol., 298, 4-26, https://doi.org/10.1016/j.jhydrol.2004.03.040, 2004.

Stauffer, R., Mayr, G. J., Messner, J. W., Umlauf, N., and Zeileis, A.: Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model, Int. J. Climatol., 37, 3264-3275, doi:10.1002/joc.4913, 2017.

Su, F., Duan, X., Chen, D., Hao, Z., and Cuo, L.: Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau, J. Clim., 26, 3187-3208, https://doi.org/10.1175/JCLI-D-12-00321.1, 2013.

Su, F., Zhang, L., Ou, T., Chen, D., Yao, T., Tong, K., and Qi, Y.: Hydrological response to future climate changes for the major upstream river basins in the Tibetan Plateau, Glob. Planet. Change, 136, 82-95, https://doi.org/10.1016/j.gloplacha.2015.10.012, 2015.

Sun, C., Chen, Y., Li, X., and Li, W.: Analysis on the streamflow components of the typical inland river, Northwest China, Hydrolog. Sci. J., 61, 970-981, https://doi.org/10.1080/02626667.2014.1000914, 2016.

Sun, R., Zhang, X., Sun, Y., Zheng, D. and Fraedrich, K.: SWAT-based streamflow estimation and its responses to climate change in the Kadongjia River watershed, southern Tibet, J. Hydrometeorol., 14(5), 1571-1586, 2013

Tang, Q., and Lettenmaier, D. P.: Use of satellite snow-cover data for streamflow prediction in the Feather River Basin, California, Int. J. Remote Sens., 31, 3745-3762, https://doi.org/10.1080/01431161.2010.483493, 2010.

Tao, Y., Duan, Q., Ye, A., Gong, W., Di, Z., Xiao, M., and Hsu, K.: An evaluation of post-processed TIGGE multimodel ensemble precipitation forecast in the Huai river basin, J. Hydrol., 519, 2890-2905, https://doi.org/10.1016/j.jhydrol.2014.04.040, 2014.

Teutschbein, C., and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, J. Hydrol., 456, 12-29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Todini, E.: Flood Forecasting and Decision Making in the new Millennium. Where are We? Water Resour. Man., 31, 3111-3119, DOI: 10.1007/s11269-017-1693-7, 2017.

Tong, K., Su, F., Yang, D., Zhang, L., and Hao, Z.: Tibetan Plateau precipitation as depicted by gauge observations, reanalyses and satellite retrievals, Int. J. Remote Sens., 34, 265-285, doi:10.1002/joc.3682, 2014.

Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modelling, Water Resour. Res., 44, W09411, doi:10.1029/2007WR006513, 2008.

Viste, E., Korecha, D., and Sorteberg, A.: Recent drought and precipitation tendencies in Ethiopia, Theor. Appl. Climatol., 112, 535-551, DOI: 10.1007/s00704-012-0746-3, 2013.

Valeriano, S., Oliver, C., Koike, T., Yang, K., Graf, T., Li, X., Wang, L. and Han, X.: Decision support for dam release during floods using a distributed biosphere hydrological model driven by quantitative precipitation forecasts, Water Resour. Res., 46, W10544, doi:10.1029/2010WR009502, 2010.

Voisin, N., Pappenberger, F., Lettenmaier, D. P., Buizza, R., and Schaake, J. C.: Application of a medium-range global hydrologic probabilistic forecast scheme to the Ohio River basin, Weather Forecast., 26, 425-446, https://doi.org/10.1175/WAF-D-10-05032.1, 2011.

Wang, X., Pang, G., and Yang, M.: Precipitation over the Tibetan Plateau during recent decades: A review based on observations and simulations, Int. J. Remote Sens., 38, 1116-1131, doi:10.1002/joc.5246, 2017.

Wang, X., Yang, M., Liang, X., Pang, G., Wan, G., Chen, X., and Luo, X.: The dramatic climate warming in the Qaidam Basin, northeastern Tibetan Plateau, during 1961–2010, Int. J. Remote Sens., 34, 1524-1537, doi:10.1002/joc.3781, 2014.

Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H.D., Högy, P., Cuntz, M., Attinger, S., Wulfmeyer, V. and Streck, T.: Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3. 5, Water Resour. Res., 49, 8200-8221, doi:10.1002/2013WR014536, 2013.

~~Wöhling, T., and Vrugt, J. A.: Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models,Water Resour. Res., 44, W12432, doi:10.1029/2008WR007154, 2008.~~

Xu, X., Lu, C., Shi, X., and Gao, S.: World water tower: An atmospheric perspective, Geophys. Res. Lett., 35, L20815, doi:10.1029/2008GL035867, 2008.

Xu, Y. P., Gao, X., Zhu, Q., Zhang, Y., and Kang, L.: Coupling a regional climate model and a distributed hydrological model to assess future water resources in Jinhua river basin, east China, J. Hydrol. Eng., 20, 04014054, DOI: 10.1061/(ASCE)HE.1943-5584.0001007, 2014.

Yang, K., Wu, H., Qin, J., Lin, C., Tang, W., and Chen, Y.: Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: A review, Glob. Planet. Change, 112, 79-91, https://doi.org/10.1016/j.gloplacha.2013.12.001, 2014.

Yang, W., Andréasson, J., Graham, L. P., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrol. Res., 41, 211-229, DOI: 10.2166/nh.2010.004, 2010.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83-97, https://doi.org/10.1016/S0022-1694(97)00107-8, 1998.

Yuan, X., Wood, E. F., Roundy, J. K., and Pan, M.: CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States, J. Clim., 26, 4828-4847, https://doi.org/10.1175/JCLI-D-12-00683.1, 2013.

Yucel, I., Onen, A., Yilmaz, K. K., and Gochis, D. J.: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall, J. Hydrol., 523, 49-66, https://doi.org/10.1016/j.jhydrol.2015.01.042, 2015.

Zhang, J. P., Chen, X. H., and Zou, X. Y.: The eco-environmental problems and its countermeasures in Tibet, J. Mt. Sci., 19, 81-86, 2001.

Zhang, L., Su, F., Yang, D., Hao, Z., and Tong, K.: Discharge regime and simulation for the upstream of major rivers over Tibetan Plateau, J. Geophys. Res., 118, 8500-8518, doi:10.1002/jgrd.50665, 2013.

Zhang, X., Booij, M. J., and Xu, Y. P.: Improved Simulation of Peak Flows under Climate Change: Postprocessing or Composite Objective Calibration? J. Hydrometeorol., 16, 2187-2208, https://doi.org/10.1175/JHM-D-14-0218.1, 2015.

Zhao, Q., Ye, B., Ding, Y., Zhang, S., Yi, S., Wang, J., Shangguan, D., Zhao, C. and Han, H.: Coupling a glacier melt model to the Variable Infiltration Capacity (VIC) model for hydrological modeling in north-western China, Environ. Earth Sci., 68, 87-101, DOI: 10.1007/s12665-012-1718-8, 2013.

Zhao, Y.Z., Zou, X.Y., Cheng, H., Jia, H.K., Wu, Y.Q., Wang, G.Y., Zhang, C.L. and Gao, S.Y.: Assessing the ecological security of the Tibetan plateau: Methodology and a case study for Lhaze County, J. Environ. Manage., 80, 120-131, https://doi.org/10.1016/j.jenvman.2005.08.019, 2005.

Zhu, Q., Xuan, W., Liu, L., and Xu, Y. P.: Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China, Hydrol. Process., 30, 3061-3083, doi: 10.1002/hyp.10846, 2016a.

Zhu, Q., Zhang, X., Ma, C., Gao, C., and Xu, Y. P.: Investigating the uncertainty and transferability of parameters in SWAT model under climate change, Hydrolog. Sci. J., 61, 914-930, https://doi.org/10.1080/02626667.2014.1000915, 2016b.

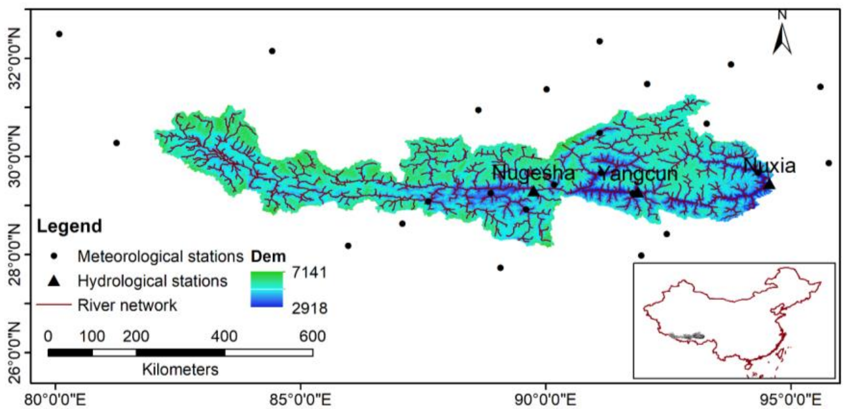


Figure1. Location of the study area, and distribution of hydrological and meteorological stations used in this study.

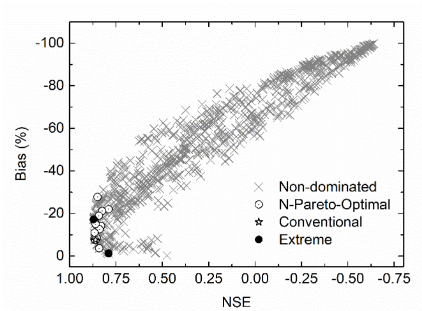


Figure 2. Two-dimensional Pareto plots for Bias and NSE at Nugesha. The cross markers indicate all the non-dominated solutions and the circle ones are selected N-Pareto-optimal parameter sets. The conventional parameter set is denoted as star markers.

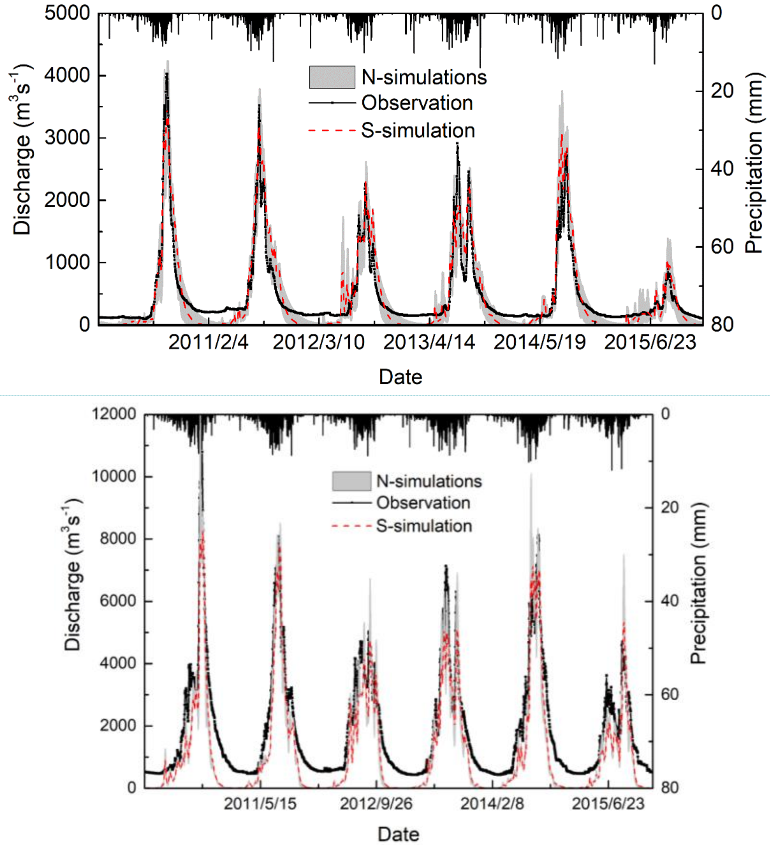


Figure 3. Daily time series of simulated and observed streamflow at ~~Nugesha~~Nuxia Station. The upper bar is the areal precipitation.

Formatted: Font: 小四, Not Bold, Font color: Auto

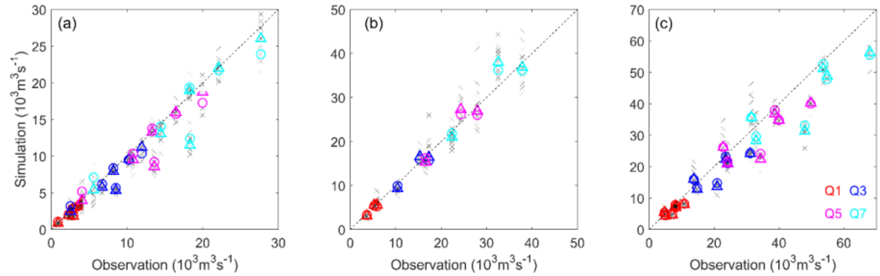


Figure 4. Typical simulated flood volumes versus observed ones. The crosses in the figures are results by N-simulations and triangles are median values of N-simulations. Circles are results from S-~~simulation.The~~simulation. The different colors are volumes for different durations, red for Q1, blue for Q3, magenta and cerulean for Q5 and Q7. (a) Nugesha, (b) Yangcun, and (c) Nuxia.

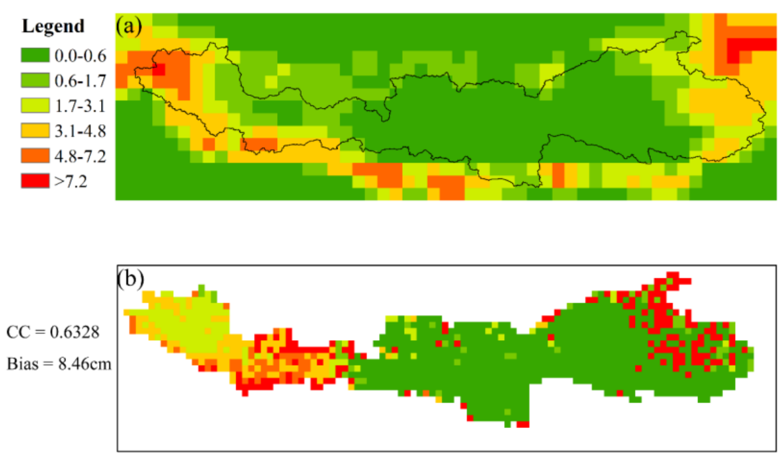


Figure 5. Spatial distribution of daily average snow depths derived from remote-sensing (a) and simulation by S-simulation (b) at Nuxia.

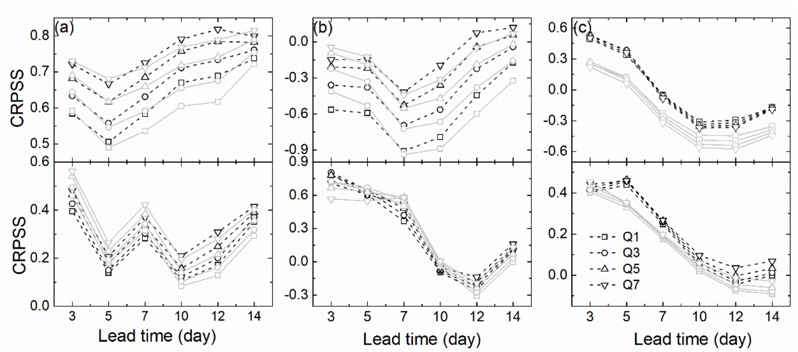


Figure 6. CRPSS for different typical accumulated flood volumes against lead time. The upper panels are results for first floods and the lower ones are for maximum floods. Scores derived from S-simulation sets are marked in black while results for N-simulations are in grey. (a) Nugesha, (b) Yangcun, (c) Nuxia.

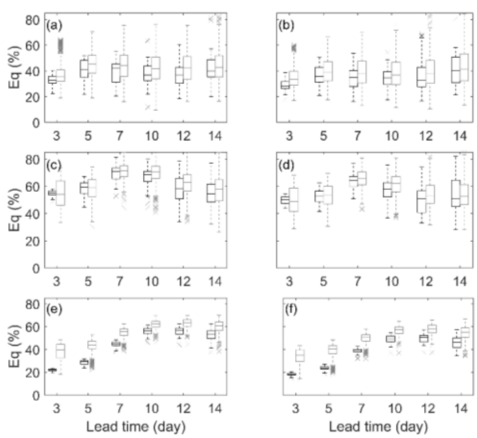


Figure 7. E_q of first floods for Q1 and Q7 at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.

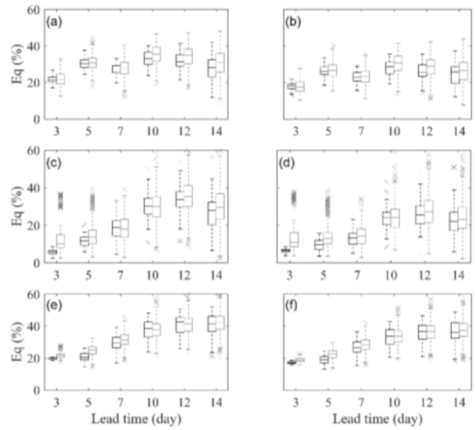


Figure 8. E_q of maximum floods for Q1 and Q7 at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.

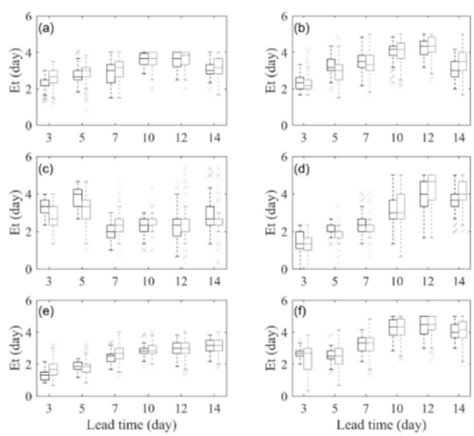


Figure 9. E_t for first flood and maximum flood at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.

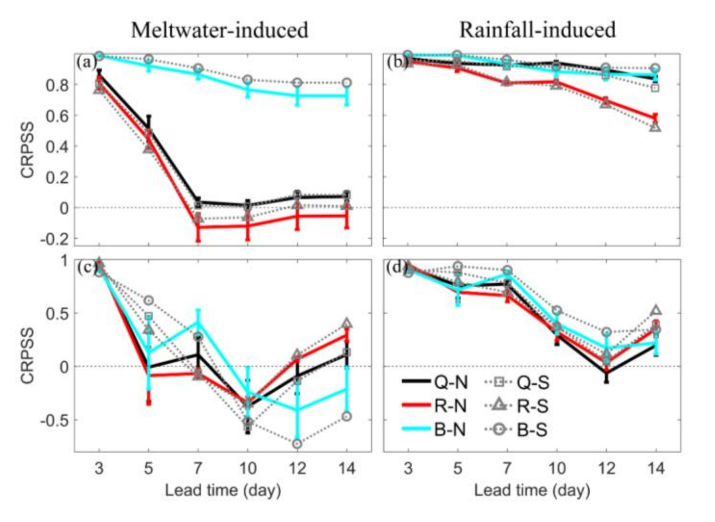


Figure 10. CRPSS of four different streamflow components against lead time at Nugesha. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars showing 95% confidence intervals and the dashed lines with different markers are CRPSS by S-simulation. Black lines are meltwater/rainfall components in total runoff (Q). Red lines are CRPSS for components in surface runoff (R) and blue ones are in base flow (B).

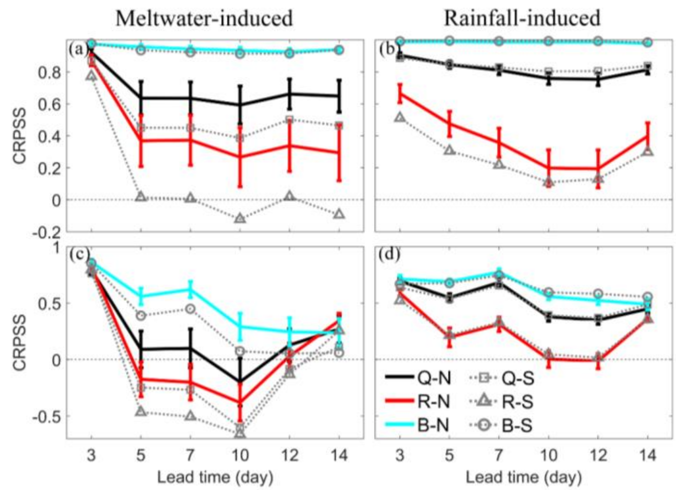


Figure 11. CRPSS of four different streamflow components against lead time at Yangcun. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars showing 95% confidence interval and the dashed lines with different markers are CRPSS by S-simulation. Black lines are snowmelt/rainfall components in total runoff (Q). Red lines are CRPSS for components in surface runoff (R) and blue ones are in base flow (B).

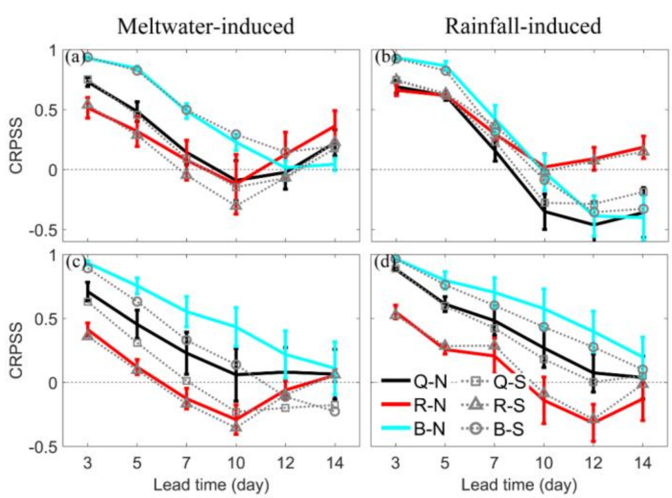


Figure 12. CRPSS of four different streamflow components against lead time at Nuxia. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars showing 95% confidence intervals and the dashed lines with different markers are CRPSS by S-simulation. Black lines are snowmelt/rainfall components in total runoff (Q). Red lines are CRPSS for components in surface runoff (R) and blue ones are in base flow (B).

Figure 13. Spatial distribution of VIC snow depth and glacier in YZR basin.

Table1. Information of N-simulations and S-simulation During Calibration and Evaluation at Three Hydrological Stations.

Station	Numbers	Mode	Calibration/Evaluation NSE	Bias(%)	NSE$_{10\%}$	Bias$_{10\%}$(%)
Nugesha	16	N-simulations	0.77~0.87/ 0.77~0.88	-27.75~-1.30/ -21.72~6.37	0.06~0.51/ -0.05~0.48	-8.83~8.05/ -9.29~16.38
		S-simulation	0.86/0.86	-7.55/-2.74	**0.51***/**0.48**	-3.02/1.29
Yangcun	15	N-simulations	0.71~0.88/ -0.07~0.65	-34.03~-10.52/ -17.72~6.37	-1.11~0.34/ -1.41~0.73	-14.21~2.60/ -4.29~19.64
		S-simulation	**0.88**/0.56	-13.54/-8.81	0.32/**0.73**	-7.43/5.21
Nuxia	11	N-simulations	0.65~0.77/ 0.58~0.79	-44.33~-34.82/ -46.53~-34.45	-1.27~-0.45/ -0.87~0.23	-27.83~-20.06/ -16.36~-4.17
		S-simulation	**0.77**/0.74	-35.03/-35.51	**-0.45**/0.06	**-20.06**/-5.33

*The bolded values in the table are the cases where S-simulation behaves better than N-simulations according to the selected objective functions.

Formatted: Indent: Left: 0.63 cm

Formatted: Font: 六号

Formatted: Font: (Asian) +Body Asian (宋体), (Asian) Chinese (PRC)

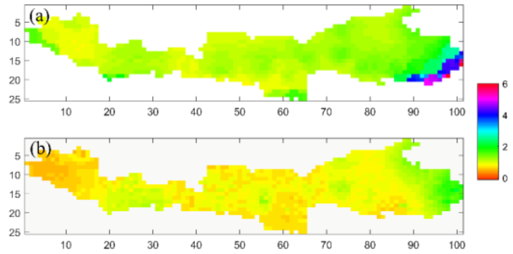
Table 2. CRPS and MAE for N-simulations and S-simulation on four typical flood volumes during the whole period. The results are displayed as MAE/CRPS. CRPS is the indicator used for N-simulations. MAE is used for S-simulation.

Events	Volumes	MAE/CRPS		
		Nugesha	Yangcun	Nuxia
First floods	Q1	107.65/96.42	258.64/230.82	315.74/379.21
	Q3	297.30/266.81	714.26/636.85	795.62/998.83
	Q5	461.82/409.22	1089.45/976.46	1181.44/1517.56
	Q7	611.13/530.84	1412.74/1274.65	1524.84/2010.17
Maximum floods	Q1	537.88/467.14	818.24/731.23	1824.27/2025.75
	Q3	1497.96/1267.92	2280.90/2021.00	5125.15/5608.94
	Q5	2304.14/1919.31	3471.46/3081.09	7820.15/8514.79
	Q7	3016.17/2514.06	4438.17/3975.66	10091.79/10940.98

Table 3. Fractions of meltwater-induced streamflow to total runoff during the evaluation period for three stations.

Station	Recorded	Simulated	
		N-simulation	S-simulation
Nugesha	18%	14%~25%	16%
Yangcun	20%	11%~30%	25%
Nuxia	38%	20%~37%	35%

Supporting information



S1. Spatial patterns of CRPS for ECMWF QPF for lead time of 3 day during wet season (May to September). (a) CRPS for raw forecasts and (b) CRPS for post-processed QPF by QM-SS.