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 read through 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 1

Response to main comments:

1. The application of flow separation in forecasting and analysis of forecast skills of different flow components, including flood volumes, base flow, first flood in a year and annual maximum flood are the main novelty of the study. The second aim of the paper is to study "the impact of an ensemble of Pareto optimal solutions on model simulations". I feel that the authors have failed in combining those aims together and it is the main drawback of the paper.

Response: Thank you very much for your comments. We agree that the original manuscript fails to combine the two aims properly. As the streamflow components are unknown, a plausible total runoff doesn't mean accurate streamflow components. Under this circumstance, we attempt to capture the most possible flow components by applying an ensemble of parameters to take account of more scenarios. The accuracy of multiple parameters in total runoff is the precondition for application in further flow components forecasts. Thus, the evaluation of N-simulations (simulation from ensemble of multiple parameters) in Subsection 4.1 and 4.2 is to demonstrate the efficacy of N-simulations and to prove the feasibility for flow components simulation. However, we didn't mention it in the original manuscript and make the paper confused and fragmented. Our solution is to clarify the purpose to adopt an ensemble of Pareto optimal parameter for flow components forecasting to make the paper more logical and integral. Thus, we added some clarifications in our revised paper.

Page 2, Lines 28-32: "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. The alternative solution is to verify forecasts on model simulations assuming that simulations driven by meteorological observations present the actual hydrological components. However, the success of this practice highly depends on how well the hydrological model is calibrated…"

Page 7, Lines 3-6: "As a result of unavailability of observed streamflow components, the evaluation of streamflow components has to be done based on the facts that the total runoff is 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…"

Page 12, Lines 5-11: "…From analysis above, an encouragingly accurate VIC simulation and forecasting system 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 and to some degree, the errors in forecasts are mainly subject to meteorological inputs and therefore the hydrological error becomes negligible…"

2. The available observations were divided into calibration, verification and evaluation periods. The authors refer to their previous paper published in Journal of Hydrology (Liu et al., 2017) when listing the additional datasets used in the present paper. I advise them to repeat that information in the present paper to help the reader. Also, the authors refer to that paper while describing the calibration stage and as a result the description has become not very transparent. It is not clear if the snowmelt component was previously used and the number and names of parameters optimised in the present paper are missing. The authors should state clearly which parameters they optimize and how the separation into snow-induced runoff component is calibrated. The authors mention some validation but the description is not clear. In summary, the Data section should be extended and the Methodology section re-organised.

Response: We are so sorry to omit the relevant data description and methodology introduction in the original manuscript and we agree that this part should be rephrased and reorganized. Different with the previously published paper, in current study the snow model related parameters are calibrated with the other normally calibrated parameter and detailed calibrated parameter will be clarified in Subsection 3.1 in the revised manuscript. The separation into snow-induced runoff components is calibrated based on two model parameters: the maximum temperature at which snow can fall and the minimum temperature at which rain can fall. These two parameters can separate the input precipitation into two parts: the liquid and the iced portion, and the total runoff is separated following the hydrological separation method described in Subsection 3.2. Seeing Page 5, Lines 9-20:

"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

<u>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 description 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 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.</u>**”**

<u>References:</u>

<u>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(3), 2431-2466, doi:10.1002/2016WR019656, 2017.</u>

<u>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(D11). doi:10.1029/2011JD017131, 2012.</u>

<u>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. https://doi.org/10.1016/j.jhydrol.2017.08.032, 2017.</u>

3. The authors compare an ensemble of multiple objective Pareto simulations and single best in their ability to forecast different flood components. The authors apply Preference Ordering Routine (POR) to choose the N-Pareto-optimal sets. There is no explanation of why that particular method was used nor what is the physical meaning of the applied numerical procedure. The authors set the number of ensemble members to ten, but it is not explained why that number was chosen in statistical terms.

Response: The explanation for why the POR method was adopted and why the number of ten was set will be added in the revised manuscript.

In the practical sense, users of automatic calibration routines have to face the task of selecting a set of suitable model parameters (preferred solution set) from the numerous Pareto-optimal sets. This is also the challenge for our study. The POR is proposed exactly to solve this kind of problem by sorting out a small number of preferred solutions. This method has been demonstrated for calibration of MIKE11/NAM rainfall-runoff model and is able to provide the best estimated parameter sets with good overall model performance (Khu, 2005). These are the reasons why POR was adopted in this study.

The number of 10 is set mainly for the consideration of computational capacity. The number of parameter sets would be more than 10 when superadded the extreme value and the compromise value in any two-objective trade-off. Given the large number of ensemble meteorological forecasts and different lead times, the parameter sets less than 20 would be manageable and achievable. In the work of Khu (2005), two solutions with efficiency of order 3 with degree 4 are able to provide still good performance, so in this study, the number of 10 is enough for representative of model parameter. As a matter of fact, when applying POR for all possible combinations of the four objective functions in this study, the final number of preferred parameter set is less than 10. Though we presupposed the upper limit, the final results didn't reach it. We have revised the paper as follows. Please see Section 3.2:

Page 6, Lines 2-7: "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 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…"

Page 6, Lines 15-16: "The essence of POR is to tighten the criteria of Pareto optimality, and thus enables to determine the limited preferred solutions…"

Page 6, Lines 18-23: "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 parameter set (the compromise point) is indicated by S-simulation thereafter.

…"

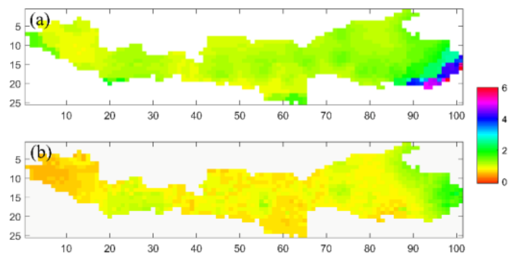
4. In the hydrograph separation subsection 3.2, the authors do not explain which data were used for the calibration/validation of the separation parameters.

Response: There is no parameters needed to be calibrated in Subsection 3.2. All the variables in the formulas are either known (model outputs) or unknown but able to be obtained by an iteration process like $f_{W,snow,t}$. The only related parameters are the Maximum Temperature at which snow can fall and the Minimum Temperature at which rain can fall. These two parameters determine the amount of rainfall $Rain_t$ and snowmdelt M_t in Eq. (5) ($R_{snow,t} = R_t f_{R,snow,t} = R_t f_{i,snow,t} = R_t \frac{M_t}{M_t+Rain_t}$) and are calibrated with other soil parameters in VIC based on streamflow objective functions (Eq. (1)-(4)).

5. The post-processing of ECMWF forecasts is performed in an arbitrary way, without checking if it is necessary and provides better forecasts. Bias correction does not always give positive results regarding forecasting (Kiczko et al., 2015, Benninga et al., 2018).

Response: As a matter of fact, we conducted a preliminary analysis of the performance of parameterized QM on ECMWF forecasts and we found that the CRPS from bias-corrected forecasts is much smaller than that from the raw forecasts especially for temperature forecasts. Given the already redundant figures in the original manuscript, we left them out in the Results section, but have added them in the revised manuscript as supporting information. At the same time, we added a simple conclusion in the manuscript to show readers that the post-processed forecasts are skillful enough for streamflow forecasting. Please see Page 10, Lines 21-23:

"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)…"



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.

6. The Results section includes hydrological model performance and an assessment of flood volume and flood component forecasts. This section I find very confusing. It does not help that the authors use abbreviations that the reader needs to be acquainted with.

Response: We thought the abbreviations would be beneficial for the readers to read and understand the manuscript. According to the suggestion from referee, we reduced the abbreviations in the revised manuscript. The hydrological station names and the flood events were indicated by full name which would be helpful for reading and make less abbreviations that the reader needs to be acquainted with. At the same time, we improved our expression and wording to make the paper more readable.

7. The authors conclude that 7-day accumulated flood volumes are easier to forecast than the peak flows. The snow-induced flood component is not well captured whilst the rainfall-induced floods are forecast well. Taking into account the fact that the snow and glacier melt forecasts were not available that conclusion is not surprising.

Response: We agree with you that since snow and glacier melt forecasts are not available, the performance in these streamflow components is unsurprisingly inferior due to various uncertainties. We know that the snow/glacier melting is influenced by not only the input rainfall and temperature but also the ability of model to capture melting process. However, considering the rainfall-induced components (observation) are also unavailable and the portion of rainfall used to generate this component is determined by the same procedure used to determine the snowfall to generate snowmelting, the comparison between these components can tell some relatively valid conclusion. We will add the reason you suggested in the revised manuscript and it would make the conclusion more strongly justified.

8. The authors find that the base flow component forecast is insensitive to the forecast lead. As the base flow dynamics is slower, the forecast lead may not cover the base flow variability. However, on page 12, the authors state that for NX, the base flow forecasts show a deterioration with a lead time. A synthesis of the overall results is missing.

Response: We totally agree that the insensitivity of base flow to lead time is also caused by the slower flow dynamics and the lead time doesn't cover the flow variability. We have added this possible reason in the revised manuscript to make the study strongly justified. Seeing Page 12, Lines 17-20:

"Forecast skill for baseflow components seems to be insensitive to lead time (Figs. 10a-b). On one hand, the reason 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."

Thanks for the reminding. Due to our negligence, the unique behavior of base flow at Nuxia was missing in the conclusion. We have added the contents related to Nuxia into the final section and make the conclusion more accurate and thorough.

"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.."

9. The language requires correcting by a native English speaker. The text is rather difficult to follow.

Response: We are so sorry and the manuscript was carefully checked and polished by the native English speaker.

Response to Editorial comments:

1. Page 1: The Abstract conclusions are not transparent. It is not mentioned that the forecasting performance varies within the catchments studied.

Response: The forecasting performance does vary for different sub-catchments and the detailed difference in three sub-area has been added in the Abstract in the revised manuscript. The main difference for three sub-catchments is that baseflow components at Nuxia tends to change with the lead time. Seeing abstract:

"N-simulations is proposed to account for more scenarios of parameters 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."

2. Page 2, lines 5-9 Style should be corrected.
Response: We have corrected the style for lines 5-9 in Page 2.

3. Page 2., lines 25-27: Is this snow tracking model used in the present paper?
Response: Yes, this snow tracking model is used in this study.

4. Page 3, lines 111-13: style should be corrected.
Response: We have corrected the style for lines 111-113 in Page 3.

5. Page 5, lines 17 and 20: instead of the word "theorem" I would use "attribute".
Response: We have replaced the word "theorem" with "attribute" in the revised manuscript.

6. Pages 7- to the end: there are language problems in nearly all pages and language editing by a native English speaker is needed.
Response: Native English speaker was asked to check and polish the entire manuscript.

7. Figure 5 – there should be some quantitative assessment of the differences between simulated and observed snow cover. The comparison does not look well!
Response: We have added the quantitative assessment for the simulated and observed snow cover in the revised manuscript. The special correlation coefficient and the overall bias in the entire study area were calculated and attached in the figure to show more direct evaluation. Seeing Fig. 5.

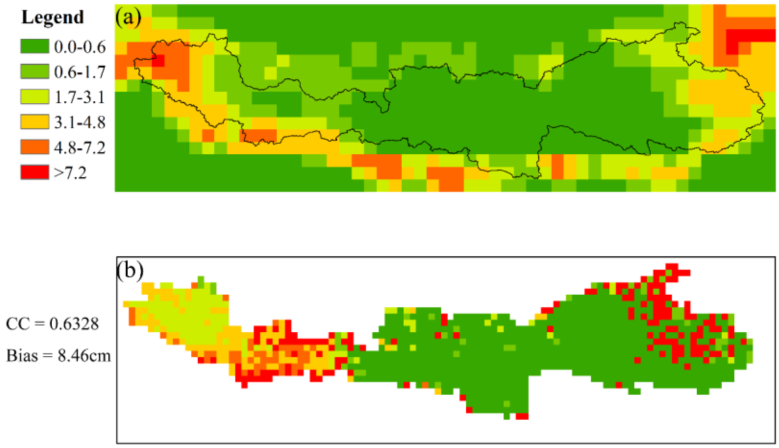


Figure 5. Spatial distribution of daily average snow depths derived from remote-sensing (a) and simulation by S‒simulation (b) at Nuxia.

8. Figures 10-12 - it would help if the columns were named (snow-melt -induced and rainfall-induced components).
Response: We have named the columns in Figures 10-12 to make the figure more accessible.

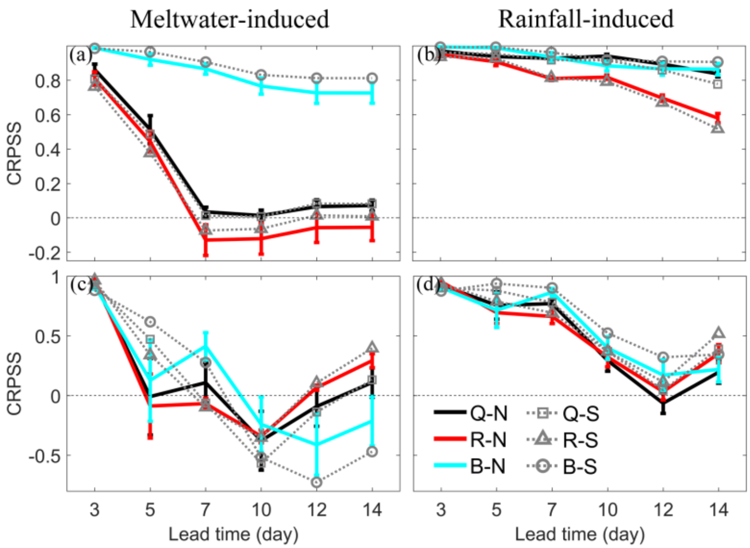


Figure 10. CRPSS of four different streamflow components against lead time at Nugesha. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars 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).

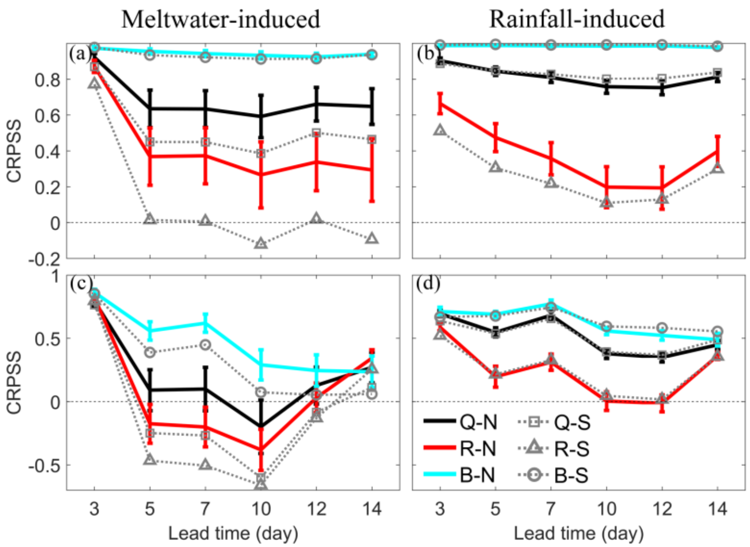


Figure 11. CRPSS of four different streamflow components against lead time at Yangcun. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars 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).

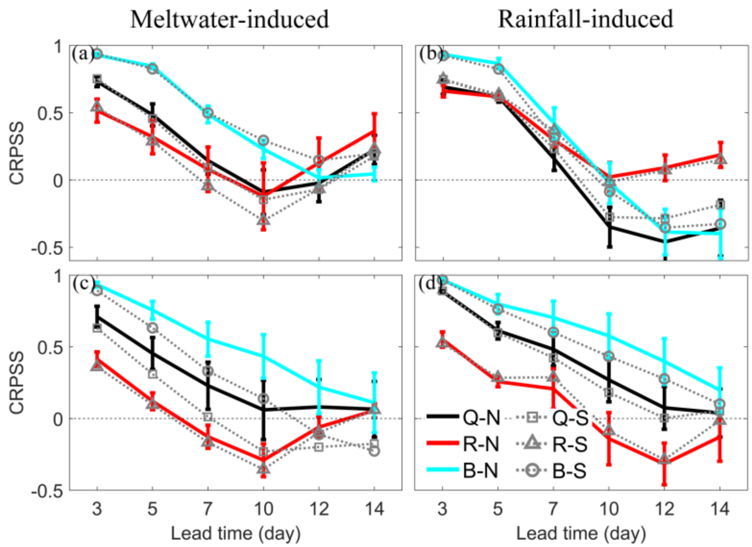


Figure 12. CRPSS of four different streamflow components against lead time at Nuxia. Meltwater-induced components for first floods (a) and maximum floods (c), rainfall-induced components in first floods (b) and maximum floods (d). The thick and solid lines are CRPSS by N-simulations with vertical bars 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).

9. Benninga et al. (2018) is not referred to in the text.

Response: The reference for Benninga et al. (2018) was removed in the revised manuscript.

To Referee 2

Response to main comments:

1. The main issue in this manuscript is that the authors need more efforts to outline the advantages and disadvantages by using multiple parameter sets (N-simulations in the manuscript). It seems that more attentions are focused on the influence of multiple parameter sets on entire streamflow. I would suggest the authors added more contents in discussion to clarify the merits of N-simulation for streamflow components forecasting and evaluating.

Response: Thank you very much for your comments. We agree that more attention should be paid to clarify the advantages and disadvantages for total runoff and component flow by using multiple parameter sets. In the original manuscript, we seem to give excessive description for entire streamflow assessment. Actually, the evaluation of multiple parameter sets on entire streamflow is to verify the feasibility and applicability of N-simulations for streamflow components. As the observed component flows are unavailable, the ability in simulating entire flow to some degree indicates the ability for component flows. According to the suggestion from referee, we have discussed our results for streamflow components in more detail. Please see Subsection 4.3.

2. The main part in current manuscript is the evaluation of N-simulations for entire streamflow. In my opinion, the evaluation of entire streamflow is the foundation of evaluation of streamflow components. For there is no direct reference data for evaluation of streamflow components. The accuracy in entire streamflow is to some degree the evidence of model ability. This is the reason for usage of simulated streamflow components driven by observation as reference in components evaluation. I suggest the authors clarify this in the manuscript.

Response: Thanks for the suggestion. The evaluation of N-simulations for entire streamflow is exactly the base of evaluation for component flows, which is not mentioned in the original manuscript. To make the paper more logical, we have added this interpretation in the subsection 4.1.

<u>Page 2, Lines 28-32: "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. The alternative solution is to verify forecasts on model simulations assuming that simulations driven by meteorological observations present the actual hydrological components. However, the success of this practice highly depends on how well the hydrological model is calibrated…"</u>

<u>Page 7, Lines 3-6: "As a result of unavailability of observed streamflow components, the evaluation of streamflow components has to be done based on the facts that the total runoff is 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…"</u>

Page 12, Lines 5-11: "…From analysis above, an encouragingly accurate VIC simulation and forecasting system 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 and to some degree, the errors in forecasts are mainly subject to meteorological inputs and therefore the hydrological error becomes negligible…"

3. A brief data description used for hydrological modelling should be supplemented though it is similar for the previous publication.

Response: As mentioned in Response to Referee #1, snow model and frozen algorithm are used in the current study, which is different to the previous study. In this way, the description of related snow depth data and additional calibration parameters were added in Subsection 2.2 and 3.1 respectively.

Page 4, Lines 18-23: "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)."

Page 5, Lines 11-20: 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 description 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 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.

References:

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(3), 2431-2466, doi:10.1002/2016WR019656, 2017.

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(D11). doi:10.1029/2011JD017131, 2012.

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. https://doi.org/10.1016/j.jhydrol.2017.08.032, 2017.

4. I would suggest the authors to simplify the contents about hydrograph separation. Detailed description can be found in Li et al. (2017), and it is better for the authors to just list the differences for the method used in this manuscript with the original one.

Response: We agree with this comment. In the revised manuscript, we have tried our best to simplify the introduction of hydrological separation. In the research of Li et al. (2017), only the snow induced components in total runoff is calculated, while in this study the streamflow is separated into four different parts. This difference is depicted in detail in the revised manuscript.

5. There is no need to verify the forecasts with two reference data. Evaluation based on simulations is irrelevant to the objectives mentioned in introduction. I suggest omitting the related parts. If it remains, I would like to see it more strongly justified

Response: We agree that the verification based on two reference data is to some degree irrelevant to the two aims mentioned in introduction. We have revised the manuscript by using only the observed streamflow as verification data which is supposed to make the paper clearer. Seeing Figs. 7-9.

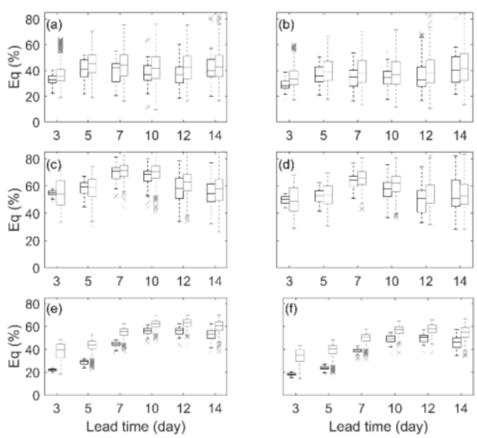


Figure 7. E_q of first floods for Q1 and Q7 at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.

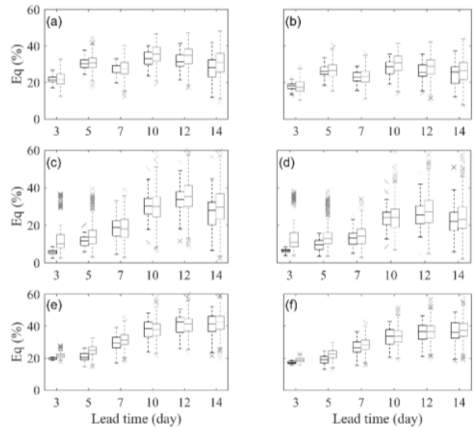


Figure 8. E_q **of maximum floods for Q1 and Q7 at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.**

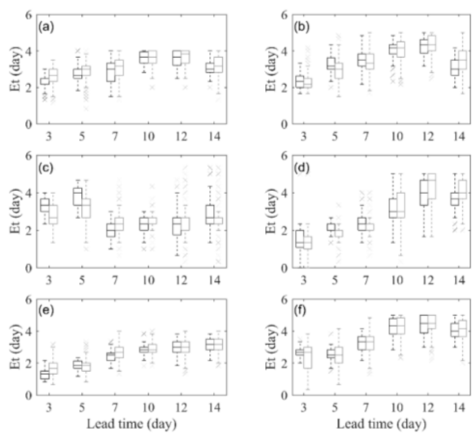


Figure 9. E_t **for first flood and maximum flood at (a)-(b) Nugesha, (c)-(d) Yangcun, and (e)-(f) Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey.**

6. I could understand most of the manuscript without difficulty, and the methods are well documented. However, there are many minor errors in spelling, grammar and English style that I have not corrected. I recommend that the authors have the manuscript proof-read by a native English speaker before publication.

Response: The manuscript has been carefully checked and polished by a native English speaker.

Response to specific comments:

7. Page 6, Line 4: as snow and glacier melt water is considered together, term "meltwater" is better representative than "snowmelt".

Response: The term "snowmelt" has been replaced by "meltwater" in the revised manuscript.

8. Page 9, Line 31: "Lead times of 3, 5, 7, 10, 12 and 14 days" make it "Lead times of day 3, 5, 7, 10, 12 and 14"

Response: "Lead times of 3,5,7,10,14 days" has been changed into "Lead times of day 3, 5,7,10,12,14".

9. Page 9, Line 32-33: "Generally, flood volumes tend to be better captured with the increase of "duration", especially for lead times from 7 day to 12 day" should be explained.

Response: The detailed explanation was added in the revised manuscript. Seeing Page 10, Lines 25-27:

<u>"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 the single day flood volume more prone to bias. Another reason is that when the duration increases, the bias from streamflow for this relatively long period can offset with each other."</u>

10. Page 10, Line 32: "It seems that N-simulations scheme works in poorly-calibrated regions." from what results the authors draw this conclusion.

Response: When we compared Figures 7-9, we found that Nuxia is the only case that N-simulations consistently provide comparable or even better results than S-simulation when verified on simulated streamflow. Since the contents related to evaluation on simulated streamflow has been left out, this sentence was deleted in the revised manuscript.

11. Page 13, Lines 2: "but" make it "while"

Response: "but" is changed into "while".

12. Page 13, Lines 27: "We believe that the phenomenon captured by most of the parameter sets would be the most possible truth." Change it to "From the view of ensemble, the phenomenon captured by more parameter sets is regarded as the most possible occurrence."

Response: Thanks for your suggestion. The sentence was revised into "From the view of ensemble, the phenomenon captured by more parameter sets is regarded as the most possible occurrence."

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]~~, Yue P. Xu[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 ~~(YZR)~~, China has ~~not~~n't been systematically investigated. This study adopts Variable Infiltration Capacity (VIC) model to forecast annual maximum floods ~~(MF)~~ and annual first flood~~s~~s ~~(FF)~~ 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 parameters in VIC ~~and~~. When trade-offs between multiple objectives are significant, N-simulations is recommended for better simulat~~in~~iong and forecasting. ~~shows improved flood simulation~~This is why better results are obtained for Nugesha and Yangcun stations. Ensemble flood forecasting system can skilfully predict ~~MF~~maximum flood~~-~~s with a lead time of more than10 days, and ~~has skill in forecasting the snowmelt-related components in~~ about 7 days ahead for melt-water related components. The accuracy of forecasts for ~~FF~~first flood~~-~~s is inferior with a lead time of only 5 days. ~~The performance in 7-day accumulated flood volumes is better than the peak flows.~~ The baseflow components ~~in baseflow~~ for ~~FF~~first floods are ~~irrelevant~~ insensitive to lead time except at Nuxia ~~s~~Station, whilst for ~~MF~~maximum floods an obvious deterioration in performance with lead time can be perceived. The ~~snowmelt~~meltwater-induced surface runoff is the most poorly captured component by the 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. ~~From this study, it is concluded that snowmelt-induced flood volume plays an important role in YZR Basin especially in FF.~~

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 ~~the~~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 ~~researches~~studies, it is confirmed that the atmospheric and hydrological cycle in TP have ~~underwent~~ 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

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

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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 operational and 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 ~~seldom~~ few studies have investigated the 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 ~~components~~ component simulation ~~requests~~ 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 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 ~~corresponding~~in-situ observations. However, observed streamflow components are usually unavailable~~ at present~~, making the evaluation of ~~forecasting~~ streamflow component simulations/forecast~~ingss~~ intractable. ~~One~~ The alternative solution is to verify forecasts on model simulations assuming that simulations driven by meteorological observations present the actual hydrological components. However, ~~T~~the

~~successful application of hydrologic models~~success of this practice highly depends on how well the hydrological model~~s~~ ~~are~~ is calibrated. 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 ~~property~~ 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, as mentioned by Kollat et al. (2012), it is difficult to cause the two-objective trade-off to collapse to one single point. Due to this limitation, a few studies tend to utilize an ensemble of parameter sets to cover more probabilities in hydrological model state. Wöhling and Vrugt (2008) employed Bayesian model averaging to generate forecast ensembles of soil hydraulic models, showing ~~as~~ similar skills ~~as~~to the best ones. Teutschbein and Seibert (2012) employed 100 different optimized parameter sets in HBV to simulate streamflow with a wide range of potential sources of variability. 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 of hydrological ~~model~~models for flood and components forecasting to consider more possible hydrological initial conditions remains an unresolved and noteworthy question.

The two purposes of this study are therefore to investigate the ~~forecast~~ ability of HEPS in forecasting flood volume and its components over cold and mountainous area, and the impact of an ensemble of pareto optimal solutions on model ~~simulations~~~~simulatingn~~simulation~~g~~ 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 ~~to~~ for future research directions, and conclusion is presented in Section 6.

2. Study area and data

2.1 Study area

We ~~focuse~~focus~~d~~ our analysis on the Yarlung Zangbo River (YZR) Basin ~~(YZR)~~, 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 ~~the~~ topographical feature of ~~the~~ 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 ~~were~~ 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 ~~were~~ are utilized in this study, i.e. Nuges Station ~~(NGS)~~, Yangcun Station ~~(YC)~~ and Nuxia Station ~~(NX)~~ from the most upstream to downstream region. Except data missing in 2009, the record period of observed streamflow at ~~NGS~~ Nugesha and ~~NX~~ Nuxia is consistent with that of the meteorological data and the period of observed streamflow at ~~YC~~ Yangcun is shorter, spanning from 1998 to 2012. The first year ~~was~~ is used as warm-up period. Periods from 1999 to 2005, 2006 to 2008 and 2010 to 2012/2015 ~~were~~ are adopted for calibration, validation and evaluation purpose respectively.

Formatted: Font: (Asian) Times New Roman

Formatted: Font: (Asian) Times New Roman

The daily Quantitative Precipitation Forecasts (QPF) and Maximum/Minimum Temperature (MXT/MNT) from 2007 to 2015 ~~were~~ 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 ~~was~~ is downloaded. ECMWF ~~was~~ is selected in this study ~~as~~ due to the well-known fact that ~~it is well-known that the~~ forecasts from ECMWF are more ~~skillful~~skilful than other Ensemble Prediction Systems in TIGGE database (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/) ~~was~~are used to evaluate snow depth simulations. The data ~~was~~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 ~~was~~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 ~~were~~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).

~~The additional data used for hydrological modeling is similar to our previous work (Liu et al., 2017), thus omitted in this paper.~~

3. Methodology

3.1 Hydrological model ~~and N-Pareto-optimal parameter sets~~

The Variable Infiltration Capacity model (VIC, Liang et al., 1994; 1996) ~~was~~ is employed in this study to investigate the ability 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 ~~was~~ 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 ~~were~~ are active. Gauged and forecasted meteorological data ~~were~~ are interpolated into the ~~requested~~ required resolution using the Inverse Distance Weighted (IDW) method coupled with an elevation-based lapse rate. The lapse rate in this study ~~was~~ is set as $0.6\ mm\ km^{-1}$ for precipitation and $-6.5\ {}^{\circ}C\ km^{-1}$ for temperature. These two lapse rates ~~values~~ ~~were~~ are determined by a cross-validation process and ~~were~~are roughly consistent with the findings in Cuo et al. (2013b) ~~which~~ who performed the least squares fitting on daily temperature and precipitation over the ~~Northern~~ TP area to gain the best lapse rate for interpolation.

Model calibration ~~was~~ 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 ~~was~~ is coupled with Message Passing Interface (MPI) to achieve parallel autocalibration with high efficiency. As snow and frozen soil algorithms ~~were~~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}), ~~were~~are optimized together with other seven conventional calibration parameters (Detailed description about the calibration of these seven typical parameters can be found in our previous studies (Liu et al., 2017). The roles~~s~~ 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 parameter ranges ~~were~~are defined as [-5,5] according to Chen et al. (2017), who ~~have~~ used similar parameters in the CREST model for snow and glacier melting simulation.

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

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

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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

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

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

in which N and M are the number of ~~the daily~~all daily flows and top 10% flows, respectively; Q_{obs} and Q_{sim} are the observed and simulated daily ~~streamflow~~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 ~~were~~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 ~~these~~ numerous Pareto-optimal 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 ~~was~~is selected in this study to pick out the desired N-Pareto-optimal parameter sets~~sets were selected according to the method of Preference Ordering Routine (POR) developed by Khu (2005)~~. There are two key ~~theorems~~ 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~~k~~-dimensional subspaces. The second ~~theorem~~ 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~~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~~calibration~~ parameters. The essence of POR is to tighten the criteria of Pareto optimality, and thus enables to determine the limited preferred 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 ~~until less than 10 points are obtained~~. 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 parameter set (the compromise point) is indicated by S-simulation thereafter.

Formatted: Font: (Asian) Times New Roman

Formatted: Heading 2

Formatted: Font: (Asian) Times New Roman

3.~~2~~ 3 Hydrograph separation

The Snowmelt Tracking Algorithm (STA) proposed by Li et al. (2017) is adopted in this study to separate the hydrograph. As mentioned before, ~~Due~~ due to the lack of glacier algorithm in the VIC model, in this study~~,~~ the glacier melting is considered together with snow melting, and the term of "~~snowmelt~~meltwater" is used for representation. ~~Based on VIC modeling, i~~In order to obtain the streamflow derived from ~~snowmelt~~ meltwater in total runoff $Q_{snow,t}$, STA calculates the ~~snow~~meltwater-induced streamflow in surface runoff (R) and baseflow (B) separately. For surface runoff derived from ~~snowmelt~~meltwater ($R_{snow,t}$), STA assumes that ~~snowmelt~~meltwater and rainfall exhibit identical infiltration ($f_{i,snow,t}$) and runoff ($f_{R,snow,t}$) ratios. In this way, $R_{snow,t}$ is computed as a function of the ratio of ~~snowmelt~~meltwater M_t to ~~snowmelt~~meltwater + rainfall, $M_t + Rain_t$:

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

The fraction of baseflow induced by ~~snowmelt~~meltwater ($f_{B,snow,t}$) is assumed to be equal to the proportion of soil moisture that originated ~~as~~ from ~~snowmelt~~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)~~ on the ground~~.

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 (~~snowmelt~~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 ~~watershed~~basin. When performing hydrograph separation, one-year warming up ~~was~~is used to achieve fully explained soil moisture sources. Total ~~In this way, the streamflow~~ runoff is separated into four components, that is, the surface runoff derived from ~~snowmelt~~meltwater ($R_{snow,t}$) and from rainfall ($R_{rain,t}$); the baseflow derived from ~~snowmelt~~meltwater ($B_{snow,t}$) and from rainfall ($B_{rain,t}$).

3.~~3~~ 4 Post-processing of forecasts from ECMWF

In order to improve the raw forecasts from ECMWF, we propose~~d~~ a post-processing method by coupling parameterized Quantile Mapping (QM) with Schaake Shuffle (hereafter ~~refered~~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 ~~was~~ 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 ~~were~~ 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 ~~was~~ is performed for each grid, each lead time and each variable. ~~Considering~~ Given the heavy computation labor, the single gamma distribution ~~was~~ is selected here for timesaving and efficiency. For MXT and MNT, four -parameter beta distribution ~~was~~ is utilized as suggested by Li et al. (2010). Owing to the limited record of ECMWF forecast, the data ~~from 2007 to 2009~~excluded the forecast year ~~was~~ is used as training data to determine the parameters of QM ~~for distribution~~.

~~As mentioned before, the~~Since forecasts are~~were~~ post-processed for ~~different~~ individual lead time, grids and variable~~s were post-processed independently. T~~, the forecast ensembles therefore ~~do not have~~tend to be inappropriately space–time ~~correlations~~correlated. To generate ensemble members with appropriate space–time correlations, Schaake shuffle (Clark et al., 2004) ~~was~~ is applied to link historical data to ensemble members and to create sequences with realistic temporal-spatial patterns. 38 years of historical data from 1978 onward 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.~~4~~ 5 Evaluation indicators

The annual maximum flood ~~(MF)~~ is picked out as typical flood events. Meanwhile, the ~~first flooding~~first flood ~~(FF)~~ event in each year is also selected. Maximum flood ~~MF~~ is determined by one-day peak flow~~, while~~. F~~f~~or first flood, ~~FF~~ the definition seems to be ~~kind of~~slightly subjective~~subjective.~~ . ~~Nevertheless~~Nevertheless~~,~~ ~~FF~~first flood is just introduced as an example to verif~~f~~y the skill of VIC/ECMWF system to ~~predict~~ forecast the ~~snowmelt~~meltwater components~~,. from this point of view,~~. ~~From this point of view, it doesn't matter whether it is the exact and objective first flood or not~~. There are ~~four~~ three criterions for us to define the ~~FF~~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~~; (4) more than half of the N-Pareto-Optimal sets simulate the occurrence of snowmelt~~. Considering that ~~MF~~maximum flood ~~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 ~~indexes~~performance, i.e. the peak flow (Q1), the accumulative 3-days flows centered on peak flow (Q3), the accumulative 5-days flows centered on peak flow(Q5), the accumulative 7-days flows centered 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 ~~research~~ 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):

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

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

~~herein~~ 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.

~~Additionally, two types of verification data are utilized in this study to compute the aforementioned statistical indicators if applicable. One is the observation derived from hydrological station measurement, and the other is the simulation outputs from VIC driven by observed meteorological data. Indicators verified on observations show errors from both hydrological model and meteorological forecasts, whereas these verified on simulation outputs show errors only from meteorological forecasts.~~

4. Results

4.1 Hydrological model performance

As a result of unavailability of observed streamflow components, the evaluation of streamflow components has to be done based on the facts that the total runoff is 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 ~~were~~ ~~are~~is crucial and thus ~~analysed~~done first in ~~s~~Subsection 4.1 and 4.2.

Fig. 2 shows an example of two-dimensional Pareto plots for Bias and NSE at ~~NGS~~Nugesha Station. The p~~P~~erformance of the selected N-Pareto-optimal parameter sets and ~~traditional~~ single compromise parameter set during calibration and evaluation periods for three hydrological stations are listed in Table 1. ~~Broadly~~ 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. It is noticeable that simulation at ~~NGS~~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 ~~NX~~Nuxia is inferior with bias greater than 30%, which is similar as the previous studies by Tong et al. (2014) and Zhang et al. (2012). They both claimed that the underestimation in streamflow simulation at ~~NX~~Nuxia is highly likely to be caused by the largely underestimated CMA observations in this area, and we guess it is also a reason that within downstream regions the hydrological process becomes more complicated due to human activities. The bolded values in the table are the cases ~~that~~ where S-

Formatted: Font: (Asian) +Body Asian (宋体), (Asian) Chinese (PRC), (Other) English (United States) 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~~it~~ loses the advantage to some degree. The NSE of top 10% streamflow at ~~NGS~~Nugesha is the only ~~one~~ case where S-simulation consistently outperforms the N-simulations in either calibration ~~period~~ or evaluation period.

[Figure 2]

The observed and simulated hydrographs during the evaluation period at ~~NGS~~Nugesha are presented in Fig. 3. An obvious underestimation can be observed in low flow periods. There are two possible reasons for this phenomenon. Firstly, 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 to deteriorate model performance in some way, in particular in low flow periods. 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.

[Figure 3]

The indicators ~~performance indices of~~ for typical flood volumes simulated by VIC for ~~FF~~first floods and ~~MF~~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 ~~NGS~~Nugesha and ~~YC~~Yangcun, CRPS is consistently smaller than MAE, ~~implying~~ indicating ~~better~~ 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 ~~NX~~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 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 order to present more detailed performance of flood volume, Fig~~ure~~. 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]

~~In order to present more detailed performance of flood volume simulation by VIC, Figure~~ Fig. ~~4 exhibits typical flood volumes for MF from 2010 to 2015 respectively. It is noticeable that the majority of the flood events can be captured by N-simulations, and volumes tend to be better covered as with the “duration” increasing. The flood volume at YC 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 NX are obviously underestimated. In most of cases, the N-simulations even fail to cover the observations. Similar but better behaviors exist for FF and thus omit here.~~

[Figure 4]

VIC simulated snow cover ~~,was~~is compared with snow depth derived from passive microwave remote-sensing data~~ by Che et al. (2008) and Dai et al. (2015)~~. ~~Figure Fig. 5~~Figure. 5 shows the spatial distribution of observed and simulated daily average snow depths during evaluation. For simplicity, only the results at ~~NX~~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 ~~depicted~~described in Section 5. We also compare~~d~~ the fraction of ~~snowmelt~~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 ~~YC~~Yangcun with about 5% overestimation. Most of the records are covered by N-simulations. However,~~, and at NX,~~ all the parameter sets slightly underestimate the ~~snowmelt~~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).

~~Lead times of 3, 5, 7, 10, 12 and 14 days are chosen as representative to trace the forecast quality. Figure~~ Figure. 6 displays the CRPSS values of different flood volumes at three hydrological stations. Lead times of day 3, 5, 7, 10, 12 and 14 ~~days~~ are chosen as representative to trace the forecast quality. Generally, flood volumes tend to be better captured with the increase of ~~"~~duration~~"., especially for lead times from 7 day to 12 day.~~ One reason is that there are often larger errors in simulated flood peak, making the single day flood volume more prone to bias. Another reason is that when the duration increases, the bias from 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 ~~effective~~ lead time of skilful forecasts for ~~FF~~first floods is shorter than ~~MF~~maximum floods. This can be explained by the generation mechanism of ~~FF~~first floods. ~~FF~~First floods ~~is~~ are usually dominated by baseflow and ~~snowmelt~~meltwater. Compared with ~~MF~~maximum floods, ~~FF~~first floods ~~as representative ofas low-flow flood events,~~ normally occur~~s~~ in the same period within one year, so historical meteorological observations on the same calendar day can provide ~~plausible~~ skilful ~~inputs. , and~~~~This~~input. This fact results in a reference forecast which is hard to beat. As for ~~MF~~maximum floods, ~~flood~~ streamflow can be predicted ~~in~~ at least 10 days ahead. Similar to Table 2, forecasts driven by S-simulation gain higher CRPSS at ~~NX~~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~~ over several parameter sets~~.

[Figure 6]

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

Another statistical indicator computed from forecasted flood volumes driven by S-simulation and N-simulations is illustrated by boxplots in Fig. 7 for ~~FF~~first floods and Fig. 8 for ~~MF~~maximum floods. For simplicity, only Q1 and Q7 are displayed, and ~~the~~ an overall ~~tendency is~~ progressive improvement can be found from Q1 to Q7. As can be perceived, E_q increases ~~against~~ with lead time, but for longer lead times a decrease exists. The decrease begins from day +10 for ~~FF~~first floods and day +12 for ~~MF~~maximum floods. Meanwhile, whiskers of boxplots become wider and wider with the increase of lead time, ~~signaling~~ indicating larger degree of variability ~~in streamflow errors~~ 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~~ly accurate~~ 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 ~~FF~~first floods (Fig. 7) S-simulation outperforms N-simulations ~~for forecasts verified on observations~~with smaller median and narrower whiskers, and in terms of Q7, the difference becomes minor especially for longer lead times. ~~, whilst for results computed from simulations opposite performance emerges, which means that the errors from hydrological model degrade the efficacy of N-simulations. For YC and NX the hydrological model errors are dominant in the beginning and from lead time longer than 7 days errors from ECMWF forecasts prevail. However, at NGS errors from ECMWF forecasts become dominant from day +3. This difference is probably caused by the basin scale. The hydrological response time is shorter in NGS, and the streamflow is more sensitive to meteorological inputs.~~

[Figure 7]

As demonstrated in Fig. 8, E_q for ~~MF~~maximum floods is smaller than that for ~~FF~~first floods, with majority of the streamflow errors confined within 40%. Unlike ~~FF~~first floods, ~~MF~~maximum floods ~~is~~ are ~~usually~~ 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. ~~This is why there are smaller differences between computed from observations and from simulations for lead time longer than 3 days. But f~~For ~~NGS~~Nugesha (Fig. 8 ~~(12~~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~~the hydrological error impacts the forecasts up to day +14~~. 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 o~~Although o~~n 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~~ Fig. ~~8 also presents additional details about the ability of S-simulation and N-simulations. For NGS and YC, at which VIC was well calibrated (Table 1), behaviors of N-simulations are inferior, while for NX, forecasts of N-simulations exhibit comparable with S-simulation when verifying on observations and even preferable performance verified on simulations. It~~

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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

~~seems that N-simulations scheme works in poorly-calibrated regions. Moreover, bigger~~higher ~~Bias and smaller NSE are present at NX (Table 1), but in term of , the forecasts at this station are not inferior compared with the other two stations.~~

[Figure 8]

The errors in peak time prediction are displayed in Fig. 9. The left sides are subplots for ~~FF~~first floods, and the results for ~~MF~~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 ~~FF~~first floods and ~~MF~~maximum floods, yet errors in ~~MF~~maximum floods are larger than that of ~~FF~~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 ~~FF~~first floods at ~~NGS~~Nugesha is the largest, and the cause ~~may be the shorter response time~~is similar to E_q. As for maximum flood events, an obvious increase in E_t from day +7 can be perceived. ~~The differences between the peak time errors verified on observations and simulations are not that significant compared with streamflow errors, so errors from meteorological forecasts should take most of the responsibility for , especially for NGS and NX. In other words, the calibrated hydrological model has a satisfactory skill in capturing the peak times.~~ 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]

4.3 Streamflow components forecasts

From analysis ~~mentioned~~ above, an encouragingly accurate VIC simulat~~ing~~ion and forecasting system 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. However, the comparison of streamflow component forecasts is performed based on simulated components. To some degree, ~~-t~~~~T~~he errors in forecasts~~, they~~ are mainly subject to meteorological inputs~~,~~ and therefore the hydrological error becomes negligible.

~~In this section, performance of the proposed HEPS in forecasting flood components is analyzed.~~ Figures 10-12 show CRPSS of ~~snowmelt~~meltwater-induced and rainfall-induced volumes at three hydrological stations. ~~For comparison purpose, we assume that during the evaluation period the streamflow components simulated by corresponding observed meteorological inputs represent the actual components condition, and used as the observation proxy.~~ 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, ~~the~~ CRPSS here is just an indicator to show the forecast skill against lead time and to present the errors only from meteorolog~~y~~ical data. Only the results for Q1 are presented, as the results show no obvious correlations with "duration".

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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

Formatted: Font: (Asian) Times New Roman

Formatted: Normal

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

From Fig~~.~~ure. 10, it is noticeable that for ~~FF~~first floods at ~~NGS~~Nugesha, errors in forecasting surface runoff components is the main source contributing to errors in forecasting total runoff. Forecast skill for baseflow components ~~in baseflow~~ seems to be insensitive to lead time (Figs. 10a-b). On one hand, ~~After all,~~the reason may be that these components are mainly generated by available water storage in the catchment~~ and therefore hardly influenced by meteorological inputs~~. On the other hand, the ~~hydrological~~ baseflow process often evolves slowly~~,~~, possibly making the forecast lead time not able to cover the base flow variability~~.~~. ~~-~~As for ~~MF~~maximum floods, ~~similarly~~ 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. 10c-d. This means during the period of ~~MF~~maximum floods the infiltration is substantial in VIC ~~modeling~~ and makes the moisture in bottom soil layer vary with the ~~precipitation~~ rainfall and ~~snowmelt~~meltwater inputs. The information in Figs. 10c-d is in good agreement with results displayed in Fig. ~~10~~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 ~~snowmelt~~meltwater-induced components can be predicted with 7 days in advance for ~~FF~~first floods, and the lead time is much shorter for ~~MF~~maximum floods. The rainfall-induced components can be skillfully forecasted up to day +14 compared with reference forecasts.

[Figure 10]

Similar performance can be found at ~~YC~~Yangcun as shown in Fig. 11. Baseflow ~~Components in baseflow for~~components for ~~FF~~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. ~~10b~~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 ~~in this section~~here is caused by the exclusion of hydrological errors. With regard to ~~MF~~maximum floods, errors in surface runoff is still the main contributor to errors in total runoff. T~~t~~he ~~snowmelt~~meltwater-related components are forecasted with~~ as~~ short~~er~~ lead time ~~as well as~~as ~~that at NGS~~Nugesha~~, and it is difficult to distinguish the system skill in different flood components.~~. Results from S-simulation totally fall out of the 95% confidence interval, while for rainfall-induced components, S-simulation produces higher CRPSS for lead time longer than 10 days.

The most noticeable phenomenon at ~~NX~~Nuxia is that baseflow components~~ in baseflow~~ for ~~FF~~first floods at this station exhibit an obvious deterioration with lead times (Fig. 12a-b). ~~NX~~Nuxia is located in the most downstream reaches, and concentrates water from hundreds of tributaries. ~~In s~~Some tributaries~~, like Niyang River located~~ ~~at the near upper reaches of NX station, streamflow is dominated by glacier-snow me~~ are ~~lt, perhaps leading to~~fairly small with ~~the~~ rapid response of baseflow and surface runoff and some tributaries may have intensive interactions between entire soil layer~~ responses rapidly~~, 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 in Fig. ~~10~~6. Generally, the Q_{snow} and Q_{rain} forecasts are skillful in lead time of 7 day and 10 day, respectively. Surface runoff

Formatted: No underline

Formatted: No underline

Formatted: No underline

Formatted: No underline

Formatted: Not Highlight

remains the toughest part for forecasts, in which the ~~snowmelt~~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 ~~is~~ was better simulated than that by S-simulation as shown in Table 2, although the NSE and Bias value are more favorable 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 ~~is~~ are ~~stronger~~significant. In this case, the N-simulations can synthesize advantages from different components. This is why N-simulations can provide more desirable skill at ~~NGS~~Nugesha than ~~NX~~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 ~~would~~ 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 ~~mentioned before, t~~there is no glacier module in the current VIC model~~. The~~, the glacier-related process ~~is~~ 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~~liquid~~ (rainfall) and ~~solid~~ 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 obviously higher than that ~~by~~ of remote sensing data ~~showed~~ shown in Fig. 5 ~~but~~ 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 with the distribution of used meteorological stations shown in Fig. 1, we can infer that these positive biases ~~are~~ 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~~ing~~ glacier-melt separately: energy balance models to calculate melt as residual in the heat balance equation, and temperature-index models to quantify an empirical relationship between air temperature and melt rate (Zhang et al., 2013). 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

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman

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~~ to consider meltwater as used in this study~~. ~~While w~~When more observations are available with the development of technologies in the future, more elaborate separation method is expected.

[Figure 13]

For streamflow components ~~prediction~~forecast, the biggest challenge is the absence of data series of in-situ streamflow components. ~~In this way~~Therefore, in this study the simulation driven by observations becomes an alternative to act as proxy~~,~~ ~~but~~ although it is difficult to determine whether such proxy is believable or not. ~~In this way, c~~By using proxy, conclusion simply based on simulation of single parameter set is risky. Similar to hydrological ensemble prediction, ensemble from multi-parameter sets is believed to be more ~~plausible~~confident. From our results, different parameter sets behave consistently in streamflow components forecast~~prediction~~, i.e. ~~deterioration~~ deteriorating with increasing lead time. However, when it comes to specific skill score, ~~some~~ slight differences can be viewed from Figs. 10-12. Sometimes, S-simulation provides skillful forecasts with longer lead time, while in some other cases, performance of S-simulation becomes inferior and falls out of the 95% CI. From the view of ensemble, the phenomenon captured by more parameter sets is regarded as the most possible occurrence~~We believe that the phenomenon captured by most of the parameter sets would be the most possible truth~~. Single parameter set may present overestimation or underestimation to some degree.

The ~~snow~~meltwater-induced components in streamflow are found to be difficult for the system to ~~predict~~forecast, in which ~~these~~ those in surface runoff ~~is~~ 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~~ted~~ 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 ~~traditional~~ 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

Formatted: Font: (Default) +Body Asian (宋体), (Asian) +Body Asian (宋体), (Asian) Chinese (PRC) employed to separate the streamflow into four individual components: the surface runoff and baseflow induced by rainfall and ~~snowmelt~~meltwater respectively. The findings are summarized as following:

(1) N-simulations~~ mode~~ ~~has been~~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 ~~the~~ N-simulations is recommended when the multi-objective trade-offs ~~is~~ are ~~significant~~~~strong~~significant. When lead time extends, the differences between N-simulations and S-simulation become ~~mimor~~minor.

(2) Flood forecast~~s~~ 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 ~~MF~~maximum floods than first floods. The flood volume of ~~FF~~first floods can be predicted in 7-14 days in advance. The lead time for ~~MF~~maximum floods is 10-14 days~~.~~

~~(3) The meteorological error is the main source of errors in MF forecasts from lead time of +3 day, but for FF and smaller watershed (NGS) the influence of hydrological errors lasts for longer time. QPF for FF tends to be overestimated and for MF opposite bias trend is observed~~.

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

(4) ~~Snowmelt~~Meltwater-induced component in surface runoff is the most difficult part for the proposed system to ~~predict~~ 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.

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

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. 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~~Xizang Hydrological Bureau~~ 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 are also given to ECMWF for the development of this portal software and for the archives of this immense 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.

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.

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 9days, 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.

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.

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., 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, 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. 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 Yarlung Zangbo 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.

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.

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 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.

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.

~~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*(6), 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., & 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(4), 1659-1676, doi:10.5194/hess-19-1659-2015, 2015.~~

~~Andreadis, K. M., Storck, P., and Lettenmaier, D. P.: Modeling snow accumulation and ablation processes in forested environments, Water Resour. Res., 45(5), 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(2), 275-284, doi.org/10.1016/j.atmosres.2010.09.013, 2011.~~

~~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(2), 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*(7567), 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*(22), 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 9days, J. Hydrol., *519*, 2832-2846. https://doi.org/10.1016/j.jhydrol.2014.08.010, 2014.~~

~~Benninga, H. J. F., Booij, M. J., Romanowicz, R. J., and Rientjes, T. H.: Performance of ensemble streamflow forecasts under varied hydrometeorological conditions. Hydrol. Earth Syst. Sci.,21(10), 5273-5291. https://doi.org/10.5194/hess-21-5273-2017, 2017.~~

~~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(5), 1076-1097.https://doi.org/10.1175/MWR2905.1, 2005.~~

~~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(1), 145-154.DOI:10.3189/172756408787814690, 2008.~~

~~Cheng, G., and Wu, T.: Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau. Journal of Geophysical Research: Earth Surface, 112(F2). 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*(D22). 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*(D16), 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(1), 243-262.https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2, 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), 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(4), 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*(1), 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*(6), 7212-7230. doi:10.3390/rs70607212, 2015.~~

~~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(6), 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*(7), 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(3), 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*(1), 522-534. DOI: 10.1002/2016WR019387, 2017.~~

Formatted: Font: Not Italic

Formatted: Pattern: Clear (White)

~~Guo, D., and Wang, H.: The significant climate warming in the northern Tibetan Plateau and its possible causes, Int J Climatol., 32(12), 1775-1781.doi:10.1002/joc.2388, 2012.~~

~~Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., ... and Miao, C.: Multi-objective parameter optimization of common land model using adaptive surrogate modeling. Hydrol. Earth Syst. Sci., *19*(5), 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(7), 2232-2252. https://doi.org/10.1175/MWR-D-11-00220.1, 2012.~~

~~Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15(5), 559-570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 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(8), 1031-1040. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000707, 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(26), 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(3), 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(8), 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(11). 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., 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, 28(3), 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. 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 Yarlung Zangbo River Basin, J. Geogr. Sci., 17(3), 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(3), 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(D16), 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(D7), 14415-14428. doi:10.1029/94JD00483, 1994.~~

~~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(4), 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(3), 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(7), 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(5), 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.~~

~~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(5), 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 framework at Hunza River Basin of Pakistan Karakoram region, J. Geophys. Res., 120(10), 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(1), 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(7), 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(10), 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(5), 970-981. https://doi.org/10.1080/02626667.2014.1000914, 2016.~~

~~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(14), 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(2), 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(9). W09411, doi:10.1029/2007WR006513, 2008.~~

~~Viste, E., Korecha, D., and Sorteberg, A.: Recent drought and precipitation tendencies in Ethiopia, Theor. Appl. Climatol., 112(3-4), 535-551.DOI: 10.1007/s00704-012-0746-3, 2013.~~

~~Valeriano, S., Oliver, C., Koike, T., Yang, K., Graf, T., Li, X., ... 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(10),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(4), 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., 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(5), 1524-1537,doi:10.1002/joc.3781, 2014.~~

~~Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H. D., Högy, P., et al: 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(12), 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(12), 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(20).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(4), 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(3-4), 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(1-4), 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(13), 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(1), 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(15), 8500-8518. doi:10.1002/jgrd.50665, 2013.~~

~~Zhang, X., Booij, M. J., and Xu, Y. P. (2015). Improved Simulation of Peak Flows under Climate Change: Postprocessing or Composite Objective Calibration? J. Hydrometeorol., 16(5), 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., ... 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(1), 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., ... and Gao, S. Y.: Assessing the ecological security of the Tibetan plateau: Methodology and a case study for Lhaze County, J. Environ. Manage., 80(2), 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(17), 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(5), 914-930. https://doi.org/10.1080/02626667.2014.1000915, 2016b.~~

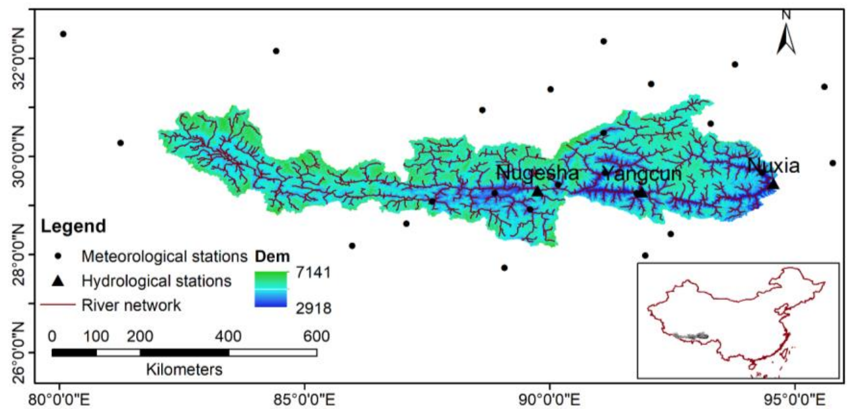


Figure 1. Location of the study area, and distribution of hydrological and meteorological stations used in this study.

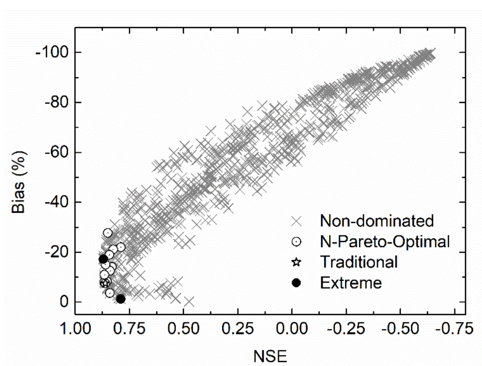


Figure 2. Two-dimensional Pareto plots for Bias and NSE at ~~NGS~~Nugesha. The cross markers indicate all the non-dominated solutions and the circle ones are selected N-Pareto-optimal parameter sets. The ~~traditional~~ conventional parameter set is denoted as star markers.

Formatted: Centered

Formatted: Caption, Left

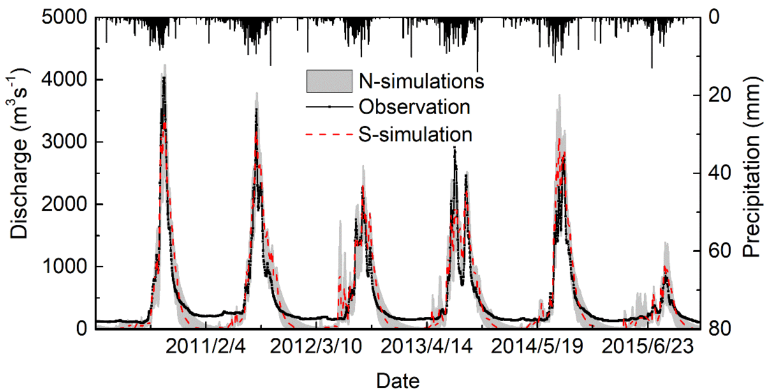


Figure 3. Daily time series of simulated and observed streamflow at ~~NGS~~Nugesha. The upper bar is the areal precipitation.

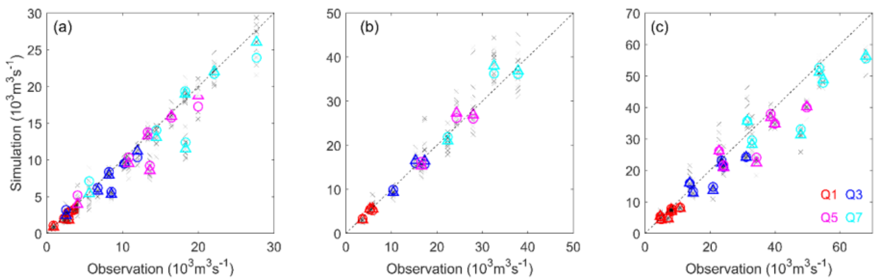


Figure 4. Typical simulated flood volumes versus observed ones. The crosses in the figures are results by N-simulations and~~ larger~~ triangles are median values of N-simulations. ~~c~~Circles are results from S-simulation. The different colors are volumes for different durations, red for Q1, blue for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

Formatted: Centered

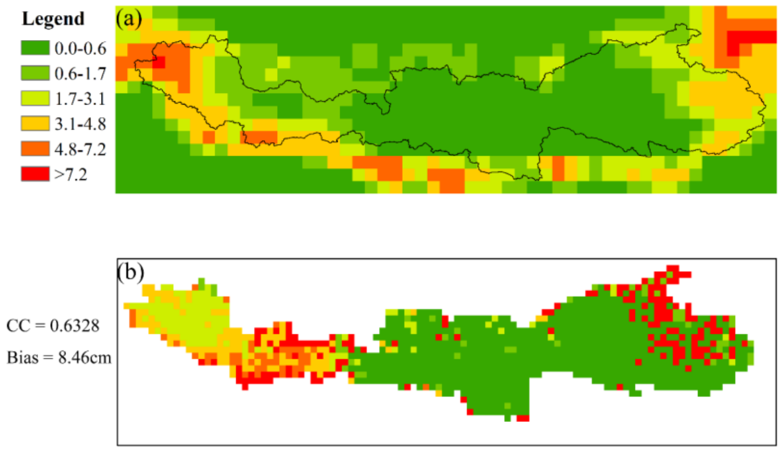


Figure 5. Spatial distribution of daily average snow depths derived from remote-sensing (a) and simulation by S-simulation (b) at ~~NX~~Nuxia.

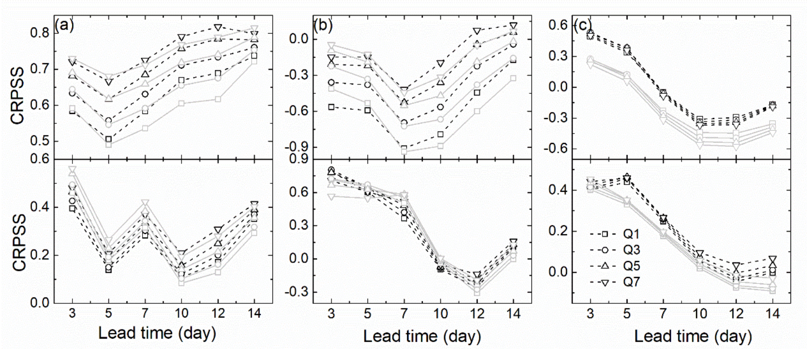


Figure 6. CRPSS for different typical accumulated flood volumes against lead times. The upper panels are results for ~~FF~~first floods and the lower ones are for ~~MF~~maximum floods. Scores derived from S-simulation sets are marked in black while results for N-simulations are in grey. (a) ~~NGS~~Nugesha,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

Formatted: Centered

Formatted: Centered

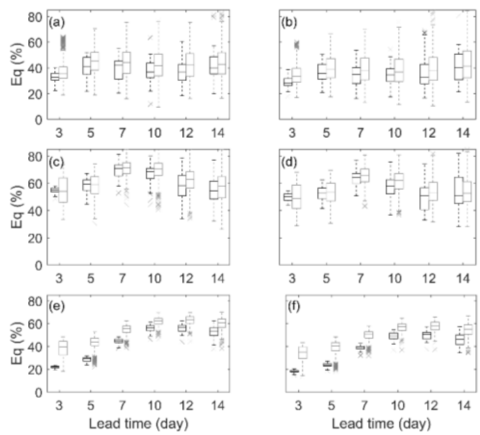


Figure 7. E_q of ~~FF~~first floods for Q1 and Q7 at (a)-(b) ~~NGS~~Nugesha, (c)-(d) ~~YC~~Yangcun, and (e)-(f) ~~NX~~Nuxia. ~~The first two of the four boxplots at each lead time in each subplot are results verified on observations and the remaining boxplots are verified on simulations.~~ The ~~unfilled~~ black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by ~~filled ones~~grey.

Formatted: Centered

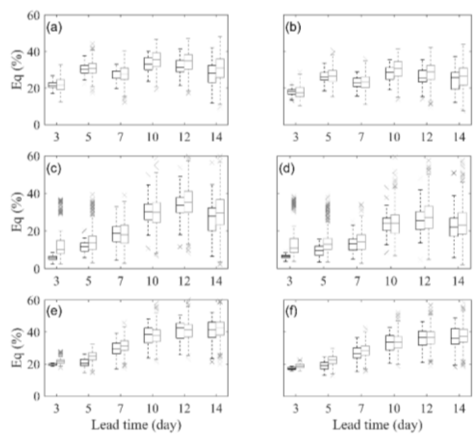
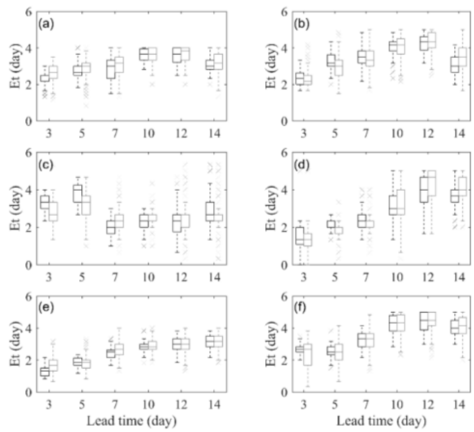


Figure 8. E_q of ~~MF~~maximum floods for Q1 and Q7 at (a)-(b) ~~NGS~~Nugesha, (c)-(d) ~~YC~~Yangcun, and (e)-(f) ~~NX~~Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey. ~~The first two of the four boxplots at each lead time in each subplot are results verified on observations and the remaining boxplots are verified on simulations. The unfilled boxplot are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by filled ones.~~

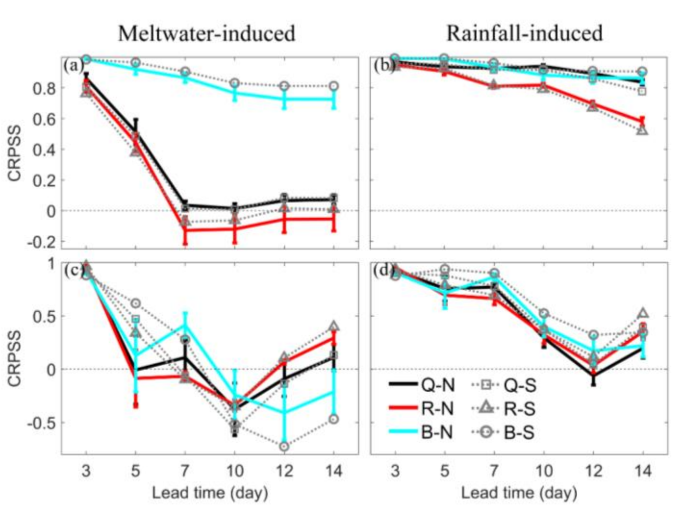
Formatted: Centered

Formatted: Font: (Default) Times New Roman, (Asian) Times New Roman



Formatted: Centered

Figure 9. E_t **for ~~FF~~first flood and ~~MF~~maximum flood at (a)-(b) ~~NGS~~Nugesha, (c)-(d) ~~YC~~Yangcun, and (e)-(f) ~~NX~~Nuxia. The black-coloured boxplots are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by grey. ~~The first two of the four boxplots at each lead time in each subplot are results verified on observations and the remaining boxplots are verified on simulations. The unfilled boxplot are forecasts driven by S-simulation and forecasts derived from N-simulations are denoted by filled ones.~~**

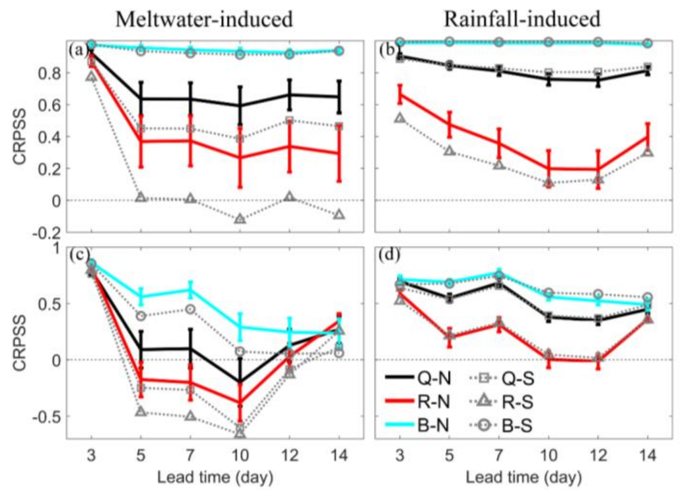


Formatted: Centered

Figure 10. CRPSS of four different streamflow components against lead time at ~~NGS~~Nugesha. ~~Snowmelt~~Meltwater-induced components for ~~FF~~first floods (a) and ~~MF~~maximum floods (c), rainfall-induced components in ~~FF~~first floods (b) and ~~MF~~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~~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).

Formatted: Centered



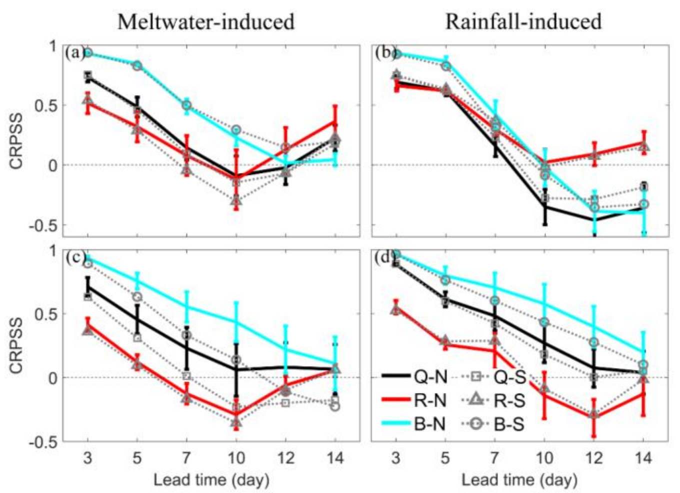
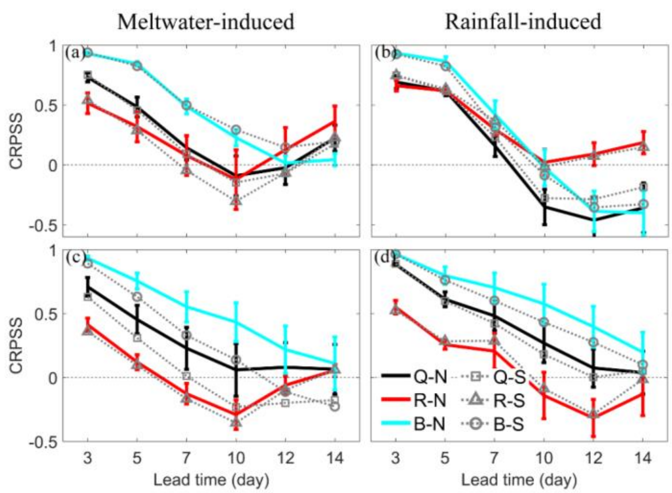


Figure 11. CRPSS of four different streamflow components against lead time at ~~YC~~Yangcun. ~~Snowmelt~~Meltwater-induced components for ~~FF~~first floods (a) and ~~MF~~maximum floods (c), rainfall-induced components in ~~FF~~first floods (b) and ~~MF~~maximum

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).



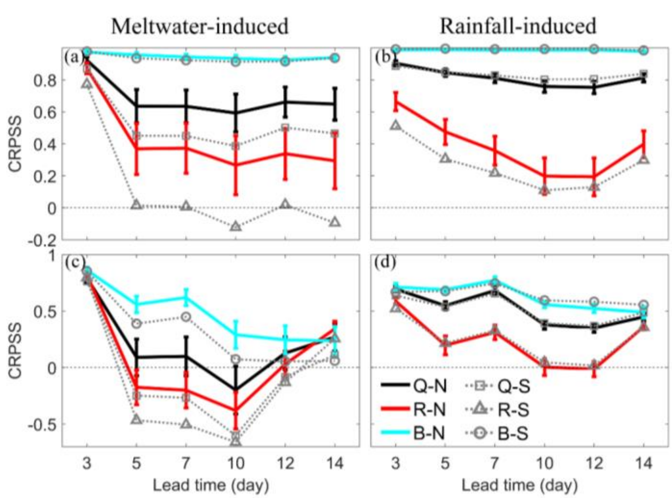


Figure 12. CRPSS of four different streamflow components against lead time at ~~NX~~Nuxia. ~~Snowmelt~~Meltwater-induced components for ~~FF~~first floods (a) and ~~MF~~maximum floods (c), rainfall-induced components in ~~FF~~first floods (b) and ~~MF~~maximum floods (d).

Formatted: Centered

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.

Formatted: Centered

Formatted: Normal

Table 1. Information of N-simulations and S-simulation During Calibration and Evaluation at Three Hydrological Stations.

Station	Numbers	Mode	Calibration/Evaluation			
			NSE	Bias(%)	$NSE_{10\%}$	$Bias_{10\%}$(%)
~~NGS~~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
~~YC~~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/ -~~9.29~~4.29~ ~~16.38~~19.64
		S-simulation	**0.88**/0.56	-13.54/-8.81	0.32/**0.73**	-7.43/~~-9.29~~5.21
~~NX~~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

Formatted: Normal

Formatted: Subscript

Formatted: Subscript

Table 2. CRPS and MAE for N-simulations and S-simulation on four typical flood volumes ~~indexes~~ 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		
		~~NGS~~Nugesha	~~YC~~Yangcun	~~NX~~Nuxia
~~FF~~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
~~MF~~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

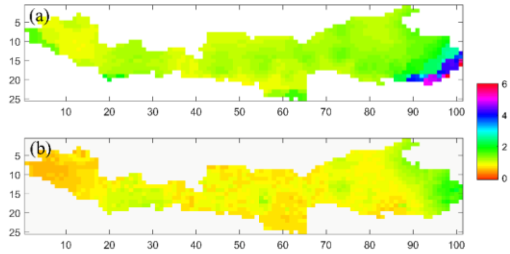
Formatted Table

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

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

Formatted: Normal

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.

Formatted: Centered

Formatted: Font: 小五, Bold, English (United Kingdom)

Formatted: Font: 小五, Bold, English (United Kingdom)

Formatted: Font: 小五, Bold, English (United Kingdom)

Formatted: Font: 小五, Bold, English (United Kingdom)

Formatted: Font: 小五, Bold, English (United Kingdom)

Formatted: English (United States)

Formatted: Normal