Response to the comments from Reviewer #1

We are grateful to the reviewer for the constructive and careful review. We have incorporated the comments to the extent possible. The reviewer's comments are italicized and our responses immediately follow.

The authors proposed an integrated ensemble prediction approach based on hydrometeorological modeling and machine learning for streamflow forecasting over a cascade reservoir catchment. The performance of the prediction with different model settings is compared. Results show the potential of the integrated hydro-meteorological and machine learning approach. Some comments are provided as follows and need to be addressed before the potential publication of this study.

Response: We would like to thank the reviewer for the positive comments. Please see our responses below.

1) Line 114: Please give the full name of CSSPv2

Response: We have revised as "... a newly developed high-resolution land surface model, named as the Conjunctive Surface-Subsurface Process model version 2 (CSSPv2, Yuan et al., 2018), to provide runoff and streamflow forecasts ..."

2) Lines 145-146: The authors calibrate the model based on the runoff at each grid (instead of using the streamflow at the control station). What is the motivation or advantage of this model calibration?

Response: Thanks for your comments. The calibration can be divided into two steps: the calibration of parameters for rainfall-runoff generation process for each grid, and the calibration of parameters for river routing process for the entire catchment. We have clarified the motivation in the revised manuscript as follows:

"The gridded runoff was used to calibrate the CSSPv2 model at each grid (Yuan et al., 2016), which would generate distributed model parameters that are different within the catchment to better represent the heterogeneity of the rainfall-runoff processes."

The detailed description of model calibration can be found in section 2.2.3.

3) Line 217, equation (2); Here "y>observation" is assigned "1". In other words, no matter how high the simulation is (if higher than the observation), it will result in a low CRPS, right? By the way, is this commonly used in previous studies?

Response: We used the CRPS definition according to Wilks (2005). A low CRPS value means the forecast distribution yields a better concentration of probability around the step function located at the observed value. If forecast members got extremely higher or lower values than the observation, the CRPS will increase because $[F(y)-F_o(y)]^2$ will increase. The CRPS is widely used in evaluating the accuracy of forecasts, and a skill score of CRPS (i.e., CRPSS) is used to compare with a benchmark ensemble forecast.

4) Lines 234-235: The simulation is better for the downstream. Is there any specific reason for this pattern?

Response: Thanks for your comments. We have clarified in the revised manuscript as follows:

"Higher NSE in the upstream part of Jiazhuan station (Figure 1) is due to more humid climate (not shown), where hydrological models usually have better performance over wetter areas. For the downstream areas with less precipitation, the higher NSE is related to the higher percentage of sand in the soil (not shown). Under the same meteorological conditions, there is higher hydraulic conductivity with higher sand content (Wang et al., 2016), and it yields less runoff under infiltration excess, which is more suitable for the saturation excess-based runoff generation for the CSSPv2 model (Yuan et al., 2018)."

Figures R1 and R2 show the distributions of precipitation and sand percentage.

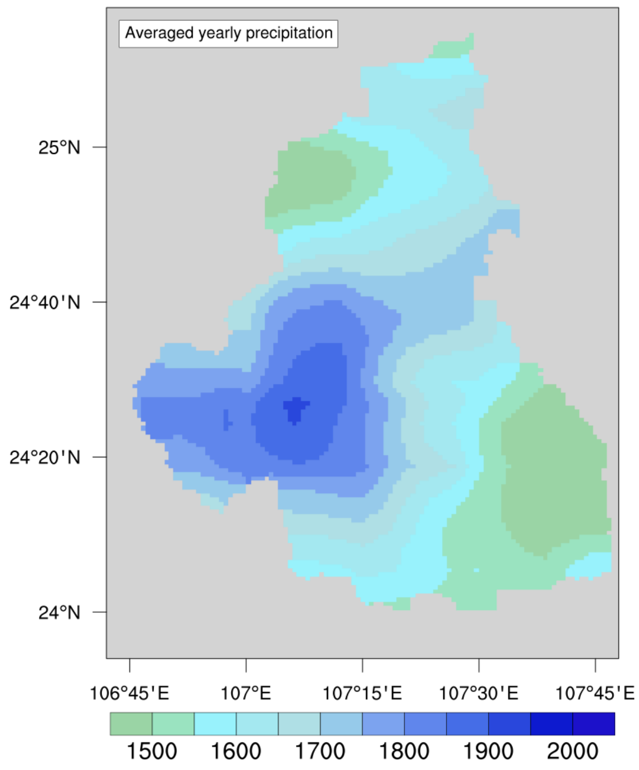

Figure R1. The spatial distribution of average yearly precipitation (mm).

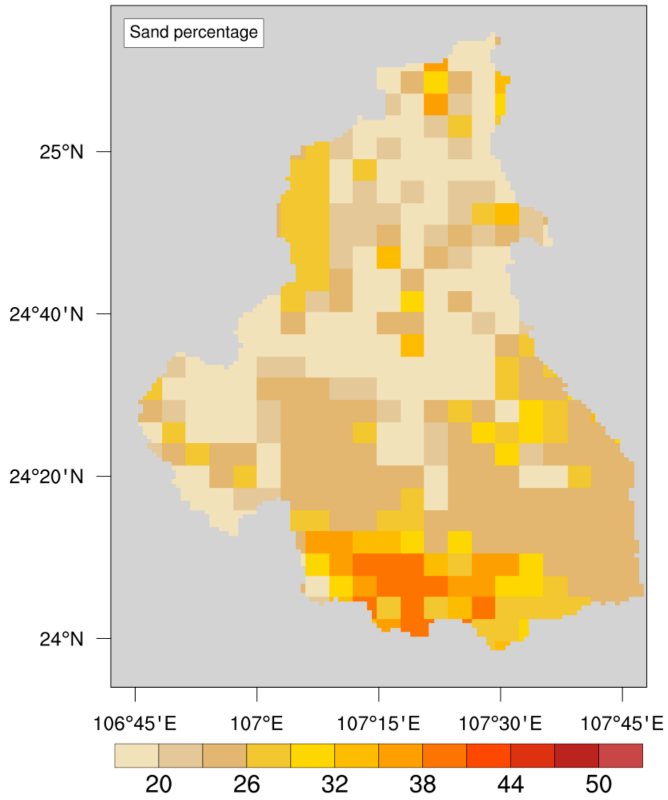

Figure R2. Spatial distribution of sand proportion.

5) Lines 276-278 (Figure 7): For the CRPS and RMSE for the lead time of seven days, there is a strong cycle in the performance. What is the reason for the strong cycle? In addition, from the RMSE, we see low RMSE values for the lead time within 1-day, and relatively high values after 1-day. However, for the CRPS, such a pattern does not exist. In addition, for the lead time beyond 1-day, the variation of CRPS and RMSE does not seem to depend on the lead time (prediction performance generally degraded for longer lead time, right?). Please clarity.

Response: Thanks for the comments. The diurnal cycles of CRPS and RMSE are associated with the diurnal cycle of upstream water release from Longtan station, and the diurnal cycle of catchment-averaged precipitation. Please see Figure 8 and related text.

About the CRPS pattern, we are sorry to inform that there were several mistakes during programming. The revised figure has been now portraited as below, which looks normal now for both CRPS and RMSE, where the error increases as the forecast lead increases.

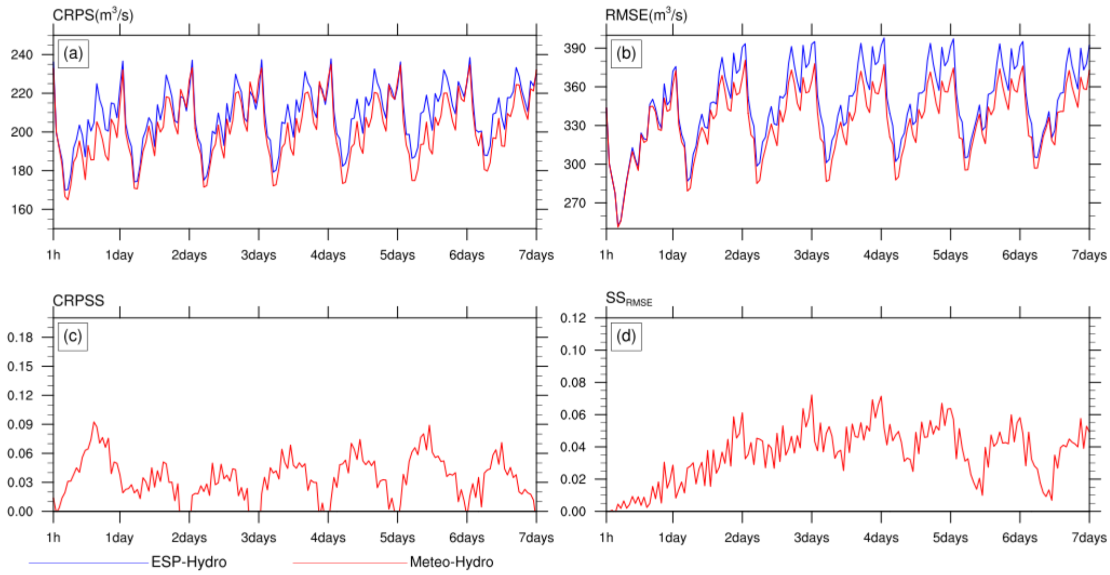

Figure 7. (a) Continuous Ranked Probability Score (CRPS) and (b) Root Mean Squared Error (RMSE) for daily streamflow ensemble forecasts at Yantan gauge. (c) and (d) are the skill score in terms of CRPS and RMSE for Meteo+Hydro, where ESP+Hydro is used as reference forecast.

6) It is hard to read the station name in this figure. Please improve it.

Response: Thanks for the comment. We have redrawn Figure 1 as follows:

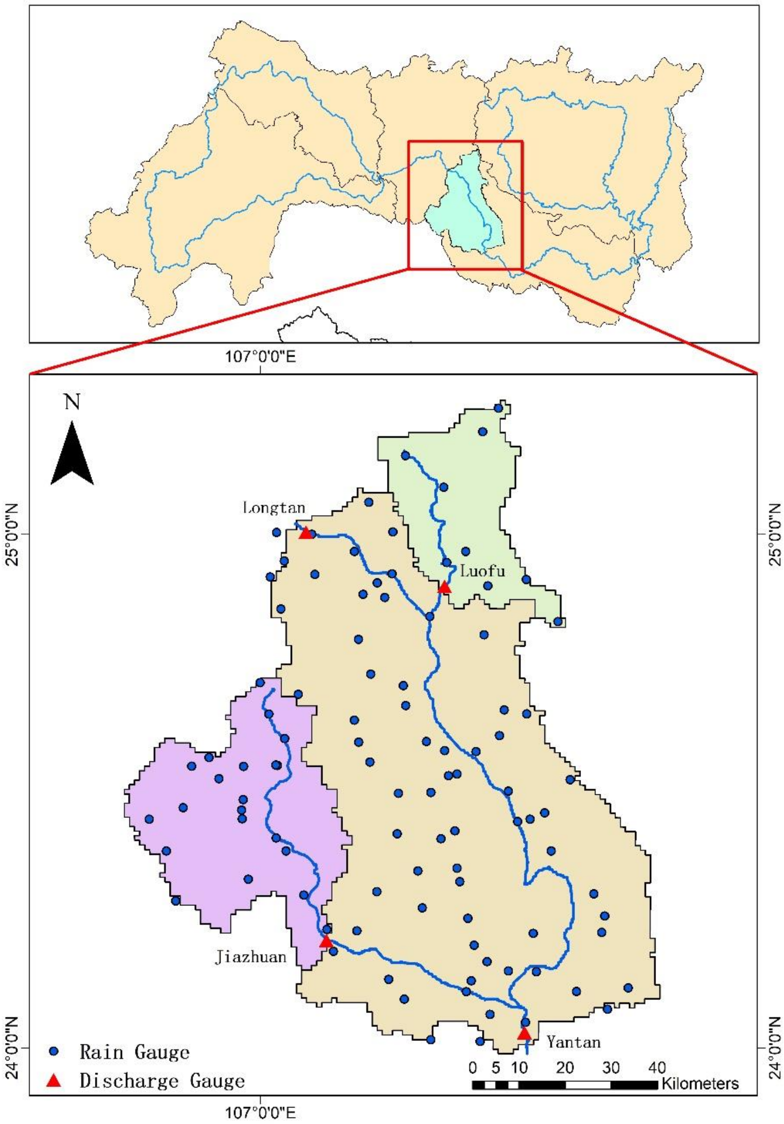

Figure 1. Locations of discharge gauges and rain gauges over the Yantan catchment.

7) Are reservoir data used in the simulation or incorporated in the model?

Response: The upstream reservoir (Longtan) discharge data was used as model's upstream inputs. We have clarified this caveat in the discussion as follows:

"This study mainly focused on exploring the added values of meteorology-hydrology coupled forecast and LSTM forecast in a non-closed catchment, so the forecast uncertainty from upstream outflow was ignored by using the observed outflow. In the future, the upstream outflow forecast is planned to include, but this requires the development of upstream hydrometeorological forecast capability, as well as the reservoir regulation forecast that is very challenging. The artificial intelligence (AI) techniques are expected to complement the physical model for reservoir regulation forecast."

Reference:

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition International Geophysics, 2005.

Wang, Y., Fan, J., Cao, L., et al.: Infiltration and Runoff Generation Under Various Cropping Patterns in the Red Soil Region of China. Land. Degrad. Dev. 27(1), 83-91. https://doi.org/10.1002/ldr.2460 , 2016.