Anonymous Referee #1

Referee comment on "Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning" by Lei Xu et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-432-RC1, 2021

In the presented manuscript, a joint uncertainty modeling method is proposed. The input data uncertainty, target data uncertainty and model uncertainty are jointly modeled in a deep learning precipitation forecasting framework to estimate the predictive uncertainty. The results show that the proposed method can improve precipitiation forecasting accuracy and reduce predictive uncertainty. Having said that I am lost and confused. Here are some of my major concern regarding the presented study:

Response: Thank you very much for the reviewing of our manuscript.

Precipitation forecasting datasets are usually daily. In this study, the authors use three datasets are all daily. Why the authors convert the data to weekly data? In Line 134, the historical three consecutive weeks are used to forecast the precipitation in the target week. How to determine the "three" weeks?

Response: The weekly precipitation forecasting is used as a case to demonstrate the efficiency of the uncertainty quantification method. The daily data is converted into weekly data because the weekly averaged datasets are used as predictors and the weekly total precipitation is regarded as the predictand in the forecasting experiment based on deep learning method. The historical three consecutive weeks are used to forecast the precipitation in the target week. For example, the predictors in the first week, the second week and the third week (P_1, P_2, P_3) are used to forecast the total precipitation in the fourth week.

Why the authors use NCEP R2, ERA-5 and MERRA-2 data? For exsample, NCEP CFSv2 also have weekly precipitation forecasting data. In this study, MERRA-2 is used as the reference data. Do different reference data in the uncertainty estimation?

Response: The three datasets, NCEP R2, ERA-5 and MERRA-2, are used to obtain the prior uncertainty estimation of predictors and predictands based on the three-cornered hat method. The uncertainty estimation by the three-cornered hat method is independent on the selection of the reference dataset. Therefore, any of the three datasets can be the reference data for uncertainty estimation.

How to determine the structure of the deep learning model in Figure 3? Besides the model parameters, the model structure can also generate uncertainty. Have the authors considered this part of uncertainty in this study?

Response: The model structure is constructed based on Convolutional Neural Networks (CNNs). We will plot the detailed model structure in the revised manuscript. We agree that the model structure can also generate uncertainty in forecasting. Since the model structure is constructed

by CNNs, the structural uncertainty comes from to what extent the model structure can represent the precipitation evolution process. In this case, the model structure uncertainty is not considered in this study. We use the dropout techniques in the deep learning model and the neural network cells are randomly dropped from the complete model by a certain probability. The model uncertainty is derived by constructing an ensemble of forecasts by dropout method.

Lines 423-424: In the places where the annual rainfall is abundant, the water cycle process is accelerated and the precipitation observations may suffer from large uncertainty. Why this uncertainty dosen't exhibit in Figure 7? There are no larger uncertainty observed in the southern China in Figure 7. Is it contradictory?

Response: The precipitation observations in southern China may suffer from larger uncertainty than that of northern China because of the accelerated water cycle for the former region. Figure 6 demonstrates the spatial patterns of the root mean square error (RMSE) for precipitation forecasting. The RMSE is larger for southern China than northern China, which is consistent with the underlying phenomenon that the observational uncertainty is possibly larger for the former than the latter. However, the predictive uncertainty is not necessarily dependent on the predictive RMSE and is not absolutely dependent on the observational uncertainty. The predictive uncertainty is dependent on data uncertainty, model uncertainty, training process, uncertainty modeling and so on. Therefore, the difference of the spatial patterns between predictive RMSE and predictive uncertainty is not contradictory. This difference is also similar with the difference between accuracy and precision. We will discuss the spatial patterns of predictive uncertainty in detail in the revised version.

How to calculate RMSE and uncertainty in Table 1? Are they the average of all the grid cells? How dose the uncertainty processing to improve forecasting accuracy and reduce predictive uncertainty? Why some methods have considered uncertainty processing but their RMSE increase compared with the no-uncertainty method? Loquercio's method also considers the data and model uncertainties. Why the uncertainty of Loquercio's method is so large?

Response: The RMSE and uncertainty in Table 1 are calculated based on the averaged values of all the grid cells. The proposed method exhibits comparable forecasting accuracy with existing methods as the differences of the RMSE values are very small between the used methods for precipitation forecasting. The predictive uncertainty in this study is smaller than the Kendall and Gal (2017)'s and Loquercio et al. (2020)'s methods, because the objective function is designed by jointly considering input data uncertainty, target data uncertainty and model uncertainty. The uncertainty modeling methods quantify the predictive uncertainty, but may not reduce the RMSE because the RMSE is calculated based on the difference between observations and predictions while the predictive uncertainty is dependent on the predictive spread.

There are two differences between Loquercio's method and our method. One is that the data uncertainty is estimated by the three-cornered hat method, while the data uncertainty is assumed as unknown parameters in Loquercio's method. Another difference is that the input data uncertainty and model uncertainty are considered in the objective function in Loquercio's method, while the target data uncertainty is not considered. However, the target data

uncertainty is included in our method besides the input data uncertainty and model uncertainty. The two differences lead to the different model parameters and data uncertainty estimation. We will discuss these differences in the revised manuscript in detail.

The section "results" is too brief. The authors may analyse how the proposed framework improve RMSE and uncertainties in detail, for example, the contribution of model uncertainty, input data uncertainty and target data uncertainty.

Response: Thanks for your suggestions. We will expand the results and discussion sections and analyze how the proposed framework influence the RMSE and uncertainties in the revised version. The contribution of model uncertainty, input data uncertainty and target data uncertainty to the predictive uncertainty will also be analyzed.

Minor comments:
The abstract is not complete in my point of view. There are four parts in the abstract usually, i.e., background, method, results and conclusion. The results and conclusion are missing in the abstract.

Response: Thank you for the instructive comments. We will include the results and conclusion parts in the abstract in the revised version.

Line 150. Is it seasonal or weekly?

Response: Thanks for pointing out this error. It's weekly.

Line 156. NECP or NCEP?

Response: It should be NCEP.

Lines 268-269. This sentence is unclear; what is the point the authors want to make with it?

Response: We would like to express that the sampling methods are used to sample from the data distribution to produce ensemble forecasts. The sampling process is conducted both for predictor data and predictand data. We will clarify the expression in the revised version.

There are many symbols without introduction in Figure 2, such as x_n^l, σ_n^l, etc. The x_n^l is also seen in Equation 16 without introduction.

Response: Sorry for the unclear clarification. The unclarified symbols will be explained in the revised manuscript.

Lines 319-320. Which deep learning network is used in this study? CNN, RNN, LSTM or all of them? Please make it clear.

Response: The CNN network is used in this study. This will be clarified in the revision.

Line 376. Typographical error.

Response: This error will be corrected in revision.

Lines 453-454. Please rephrase.

Response: This sentence is rephrased as “The higher the data and model uncertainties, the more divergent and less reliable the forecasting”.