Authors' responses (GMD-2022-276)

The authors would like to thank the editor for your precious time and invaluable comments. The corresponding changes and refinements are highlighted in yellow in the revised paper and are also summarized in our responses below. Authors' responses are in blue. Editor's comments are in black. When the manuscript in cited, it is shown in *italics*.

Response to RC1

The manuscript titled 'Key factors for quantitative precipitation nowcasting using ground weather radar data based on deep learning' presented a thorough analysis of different schemes to approach precipitation nowcasting problems using deep-learning techniques. The different schemes were tested using ground weather radar data over South Korea. In recent days, multiple works have explored the application of deep-learning algorithms for quantitative precipitation forecasting. Exploring endless options and schemes is necessary to understand better the feasibility of using these methods in an operational scenario. I appreciate the authors' effort in conducting a systematic analysis and discussions. Some parts of the manuscripts are still hard to understand and not very clear.

- ➔ The authors are very appreciative of your valuable time and effort in helping us improve our study. Based on your comments, we have updated our manuscript.

Major comments:

1. Data imbalance: The major problem in precipitation nowcasting is the lack of representation of intense precipitation due to data imbalance. Did the authors try to consider this problem in their analysis?

- ➔ In our original manuscript, data imbalance was not considered, as there were already many factors to compare. However, since data imbalance does matter in precipitation nowcasting, we addressed it in the revised manuscript. To investigate the effect of several trials to cope with data imbalance, we examined balanced loss functions from previous studies (Shi et al., 2017; Franch et al., 2020; Xiong et al., 2021; Kim and Hong). A balanced loss function is an approach that sets different weights for different intensities. Its definition and equation are described in Section 2.3 and Equation 3. Furthermore, we included the histogram of precipitation and corresponding weights for balanced loss in Figure 2 to illustrate the data imbalance and its anticipated impact.

2.3 Balanced loss function

Authors' responses (GMD-2022-276)

The loss function guides the direct optimization of DL models. The basic loss function in DL-QPN is MSE. By summing up the error of each pixel, it produces a single value for a given prediction image. As most valid precipitation pixels are severely skewed in weak rainfall intensity (about ≤ 5 mm/h), calculating MSE (Equation 1) with a uniform weight for all pixels might result in an underestimation problem. Shi et al. (2017) suggested the BMSE to mitigate the sample imbalance by using different weights for precipitation intensity (Equation 2).

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N} \quad (1)$$

$$BMSE = \frac{\sum_{i=1}^{N} w(y_i)(y_i - \hat{y}_i)^2}{N}, w(y_i) = \begin{cases} 1, & y_i < 2 \\ 2, & 2 < y_i < 5 \\ 5, & 5 < y_i < 10 \\ 10, & 10 < y_i < 30 \\ 30, & y_i > 30 \end{cases}, \quad (2)$$

where y is the reference value, and y ˆ represents the predicted value and N is the number of all valid pixels within the radar area. Figure 2 shows the distribution of rainfall intensity and weights for BMSE.

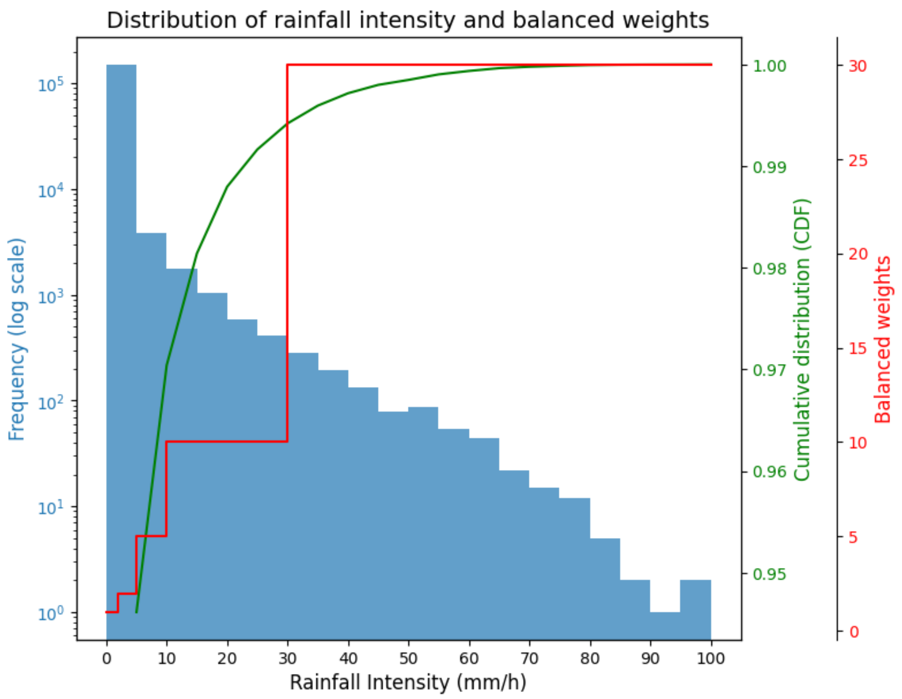

Figure 1. Mean distribution of rainfall intensity for the summers of 2020-2022 in a pixel window of 400×400. The blue bar represents the histogram of rainfall intensity. The green line shows the cumulative distribution function. The red line represents the balanced weights for mitigating data imbalances, as suggested by Shi et al. (2017).

➔ When comparing the BMSE, it appears to reduce errors in higher intensity compared to the original MSE. However, it also tends to overestimate at low levels, demonstrating that the overall estimation is generally higher than the original. To address the problem caused by data imbalance, we also tested an ensemble of original and Balanced MSEs. The ensemble results exhibit significant improvements in evaluation results, as summarized in Figure 5.

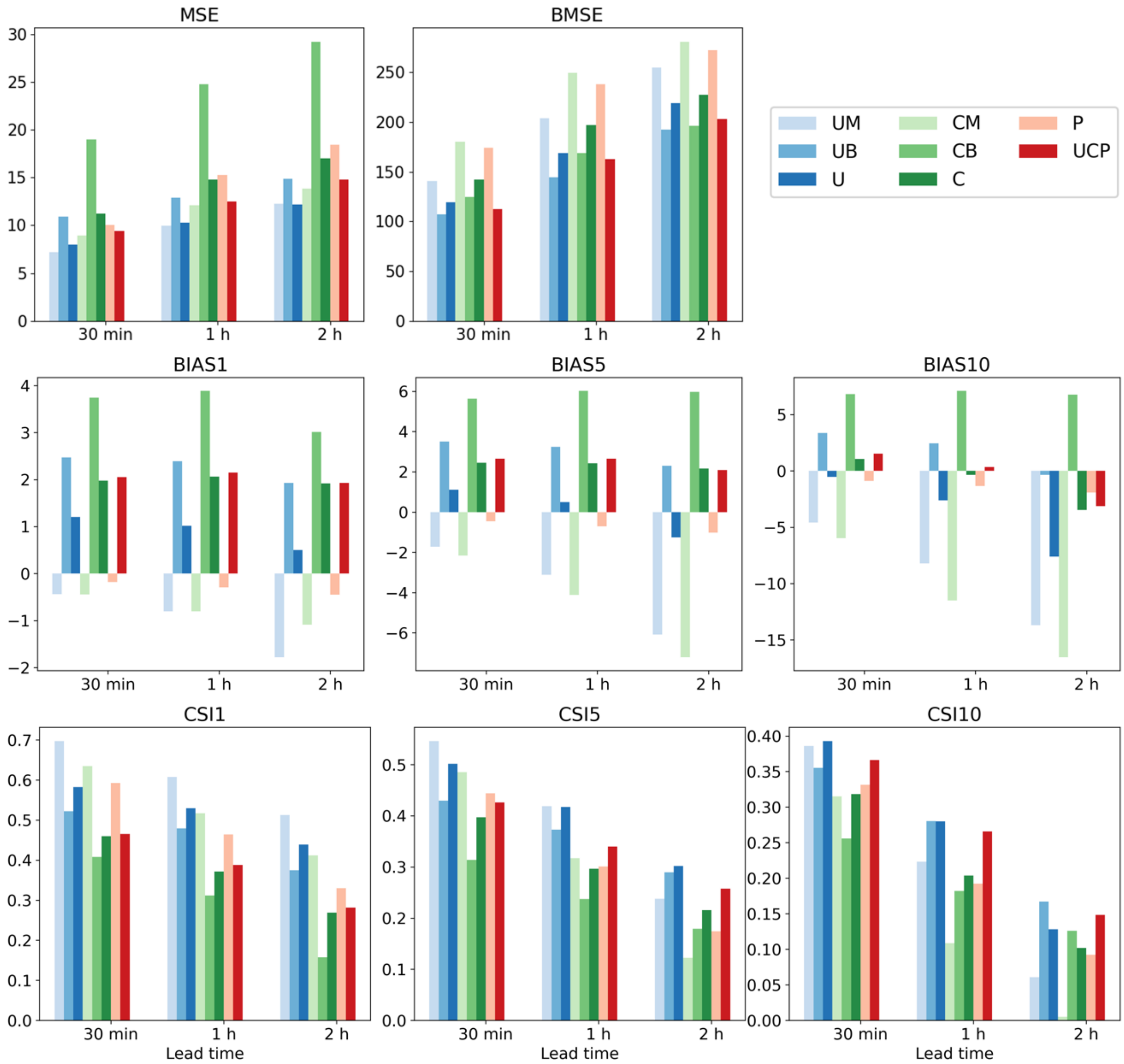

Figure 5. Quantitative performance over summers of 2020-2022 of lead times of 30 min, 1 h, and 2 h. Please refer to Table 2 for each scheme. The numbers after metrics indicate the thresholds of precipitation for evaluation.

2. Data: Do the authors consider datasets with overlap when training is done? If t1-tn is

used as input and tn+1 to tn+m is the forecast, is tn+1 - tn+n used as input for another sample?

- ➔ Yes, we noticed that there is overlap when the stride of the time step is shorter than the maximum lead time. In our literature review, we found lack of explicit discussion or explanation of how this overlap was handled. However, in response to your comment, we increased the stride time step from 10 minutes to 30 minutes to reduce excessive overlap among samples. Additionally, overlapped data in the same batch might lead to data collinearity and decrease model generalization. We believe that this approach will address the sampling issue and potential problems from data overlap.

3. Equations 2 and 3: The Mean Absolute error and mean bias equation are not normalized. Missing 1/n

- ➔ Thank you for pointing this out. We have corrected the errors in the Equations.

4. Line 271 and Section 4.2.3: Adding a dummy zero variable to input causes sparsity. Adding white noise is a better idea. But, I feel that the entire part (Section 4.2.3) does not add much to the paper. It just lengthens the paper. I will suggest the authors remove that part.

- ➔ Several ways exist to check model sensitivity, and our approach might not be the most optimal. Following your suggestion, we removed the sensitivity analysis and focused on other discussions.

5. Table 4 and others: Persistence is not explained previously in the manuscript.

- ➔ Thank you for your comment. We added it in lines 282-283.

The n-hour persistence model represents a straightforward approach in which the current precipitation is assumed to persist without any change for the next n hours.

Minor comments:

Figure 2: Why is dBZ converted to rain rate? Why not just train the model for reflectivity values?

- ➔ As the final goal of QPN is to determine the amount of precipitation, we used the unit of mm/h. However, forecasting with dBZ is also an active research area. We discuss the options for forecasting using reflectivity or precipitation intensity in Section 5.3.

(Line 459-464)

As precipitation is calculated from radar reflectivity, direct prediction of the original signal *can also be considered. Some previous studies utilized radar reflectivity in DL-QPN (Bonnet et al., 2020; Lepetit et al., 2022; Albu et al., 2022; Han et al., 2022). To our knowledge, there has been few studies comparing radar reflectivity and precipitation intensity directly in DL-QPN. In this study, we chose to forecast precipitation intensity because our final interest is in the strength of the precipitation. However, as the precipitation intensity can be converted from predicted reflectivity, further investigation is needed in the future to find a better skill score.*

Figure 3: It is better to mark the study region on the map.

➔ Thank you for your comment. We have updated Figure 3 to only display the study area with valid radar coverage and the position of each radar.

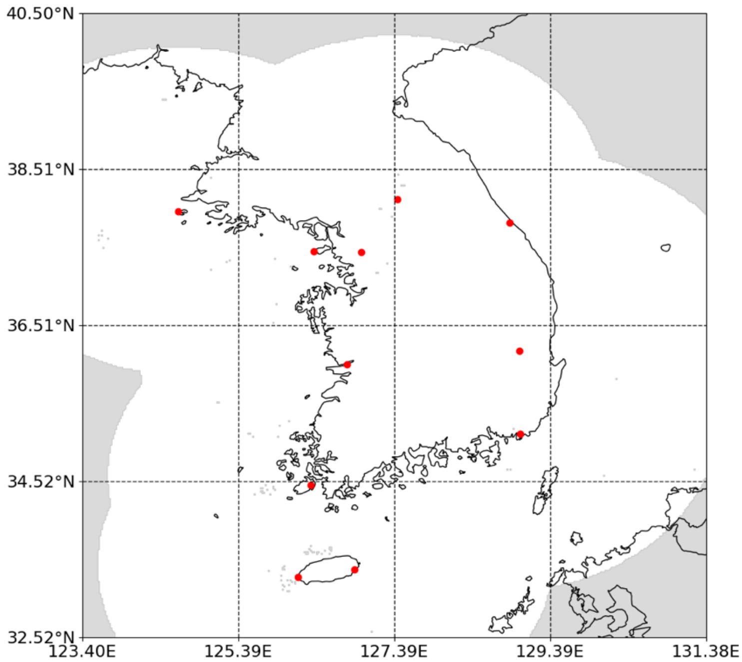

Figure 2 Weather radar over the Korean Peninsula used in this study. The grey shadow at the boundary indicates the area outside of valid radar coverage. The locations of the eleven weather radars are represented by red dots..

Line 229: Why is leaky relu not used for U-net and only used for ConvLSTM?

➔ As we employed the original models, we retained the model design, including their activation functions.

Authors' responses (GMD-2022-276)

Table 2: Why is SU-120-60 or RU-120-60 not considered for analysis? Please justify this in the text.

- ➔ Since SU-120-120 and RU-120-120 indicate the maximum lead time, they encompass SU-120-60 and RU-120-60, respectively. However, this part was omitted in the revised manuscript, as we excluded the model design from key factors.

Table 4: Is there a reason why the best bias value is not highlighted? Just curious.

- ➔ Mean bias can signal the overall tendency of underestimation or overestimation. When the magnitude of mean bias is near zero, it may indicate better results if other metrics are similar or improved. However, we cannot assert an optimal mean bias, as substantial positive or negative residuals may result in a zero-like mean bias. We have included a sentence to clarify this.

(Line 301)

Zero bias does not inherently signify superior performance

Figure 8 and 10: The thresholds should be 5 mm/h. Please check the captions.

- ➔ Thank you for your comment. We have updated the time-series figures, which can now be found in Figures 6 and 7 in the revised version.

Authors' responses (GMD-2022-276)

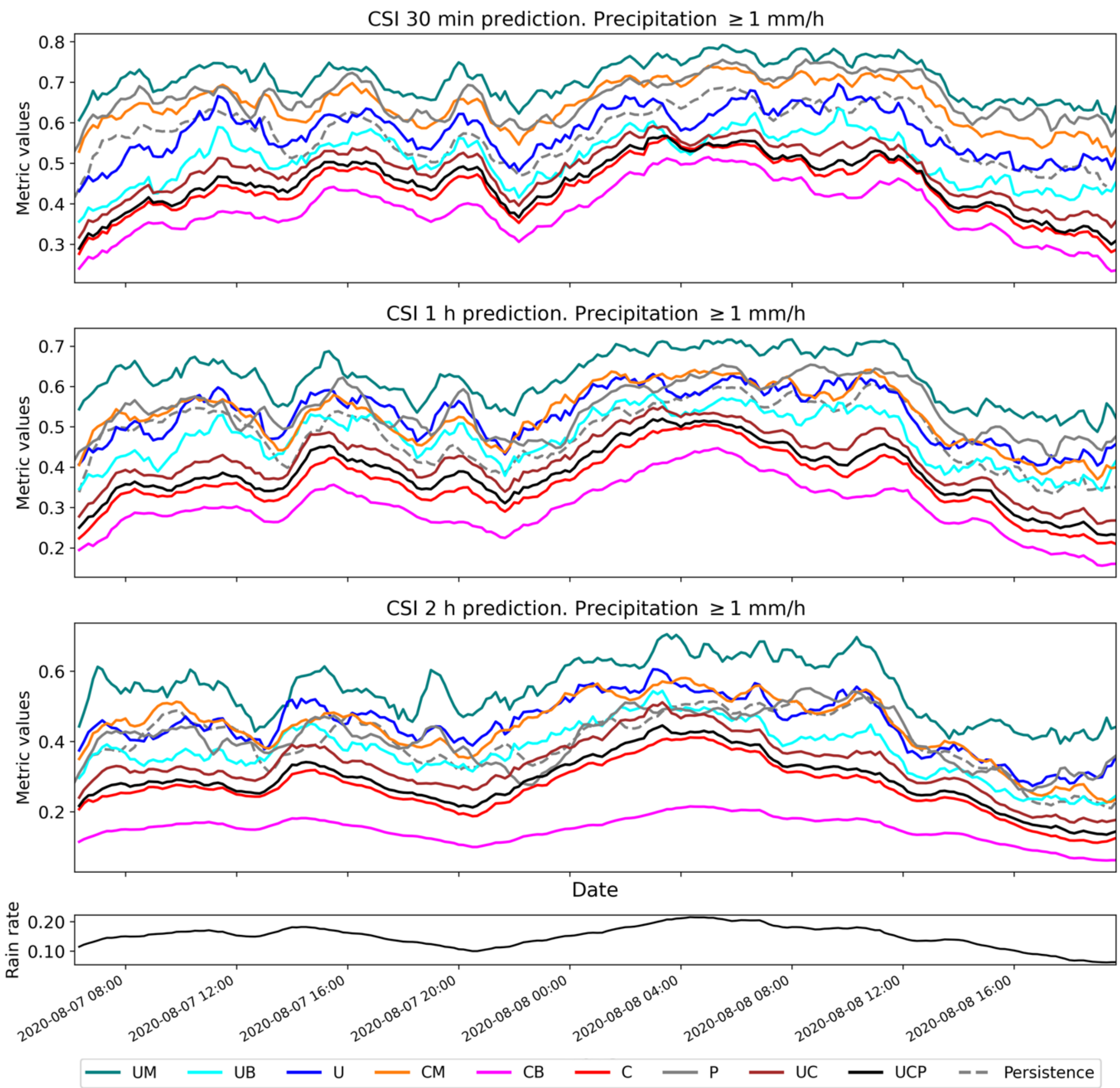

Figure 6. Comparison of CSI performance for the case of heavy rainfall over South Korea from 7th to 8th August 2020 with the 1 mm/h threshold. Refer to Table 2 for scheme names. The bottom black line represents the ratio of precipitation pixels > 1 mm/h for each radar scene.

Authors' responses (GMD-2022-276)

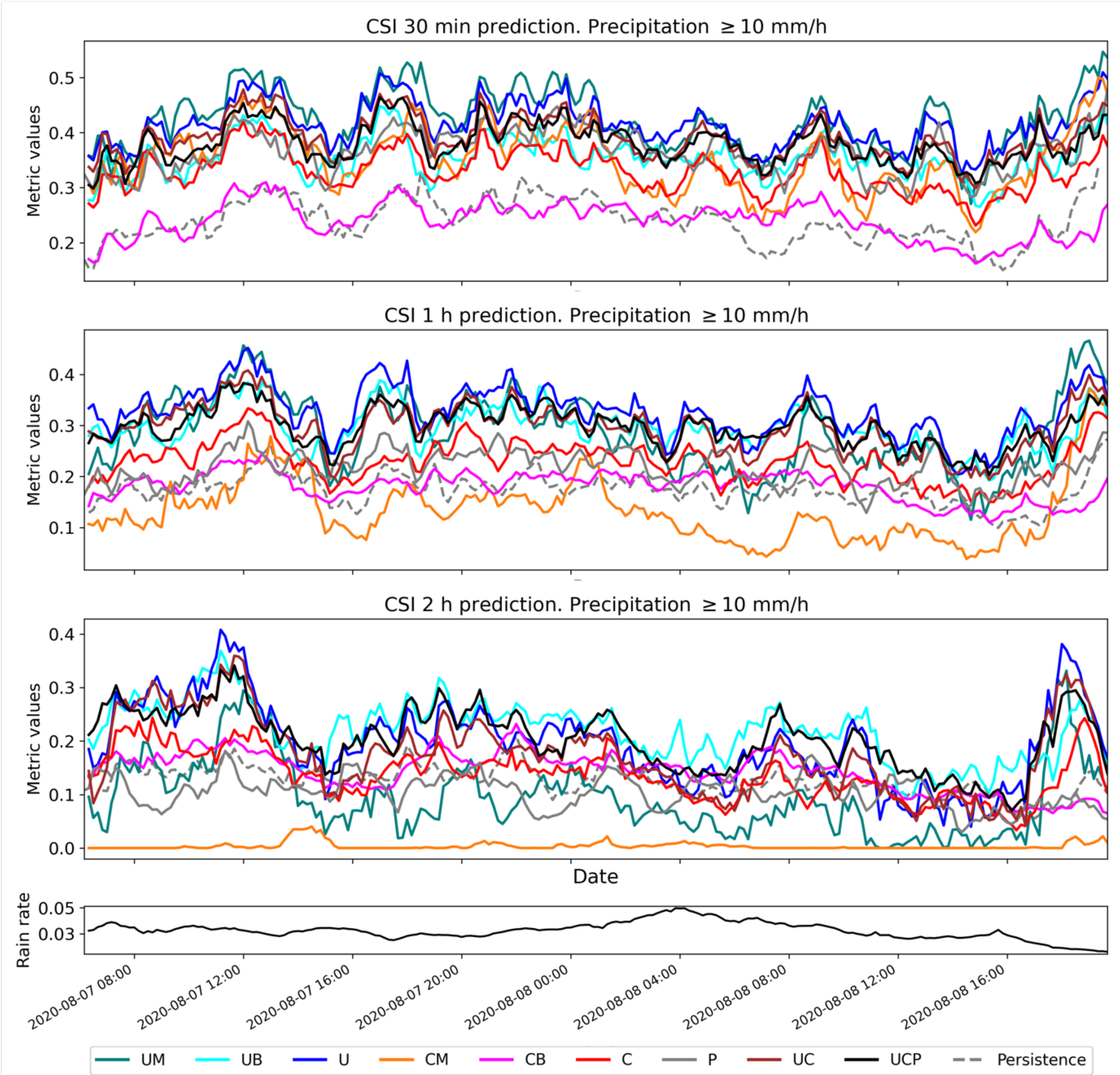

Figure 7. Comparison of CSI performance for the case of heavy rainfall over South Korea from 7th to 8th August 2020 with the 10 mm/h threshold. Refer to Table 2 for scheme names. The bottom black line represents the ratio of precipitation pixels > 10 mm/h for each radar scene.