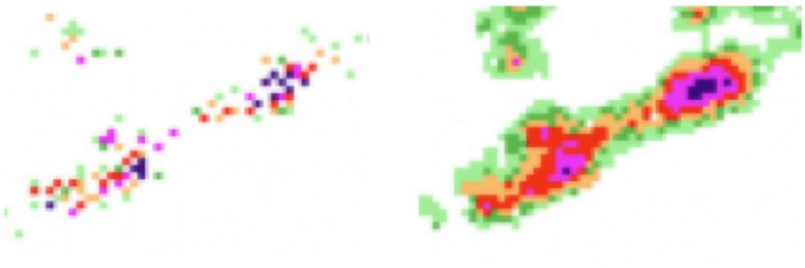
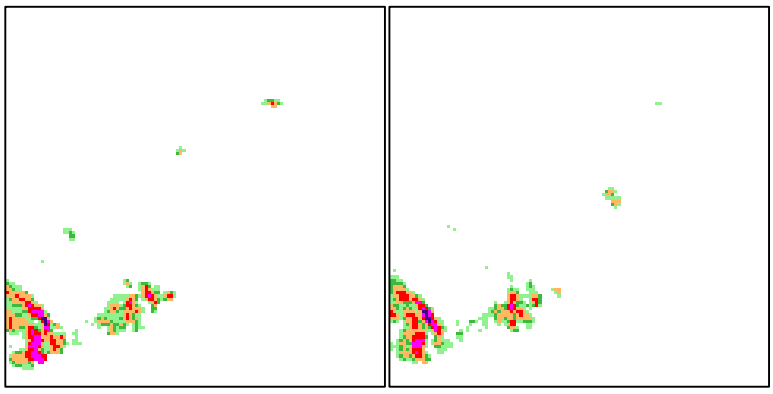
Response to Reviewer 2 Comments

Point 1: Lines 62-66 : *Existing rainfall nowcasting methods based on machine learning and deep learning are difficult to achieve rainfall nowcasting with high spatiotemporal resolution. To solve this problem, We set forecast target to the grid data interpolated from rainfall data in dense automatic weather stations. Since this data is directly measured by rain gauge, it reflects real-world rainfall distribution. In the grid data, the forecasting time resolution and spatial resolution are 30 minutes and 5 km, respectively, which can meet the high spatiotemporal resolution requirement.*

I do not agree with these sentences. If weather stations data is too sparse, interpolation with high resolution will produce smooth maps and therefore does not reflect ground truth. This is particularly true with precipitation which is characterized by strong spatio-temporal heterogeneities. For example, if the smallest distance between a point on the grid and the nearest rain gauge is 50 km, an interpolation with a resolution of 5 km does not make sense due to the high heterogeneity of the rain. Therefore, although the developed method provides high resolution maps, the latter do not necessarily reflect the ground truth at this resolution (see the figures below which are zooms of fig. 1.b and 1.c).



Response 1:



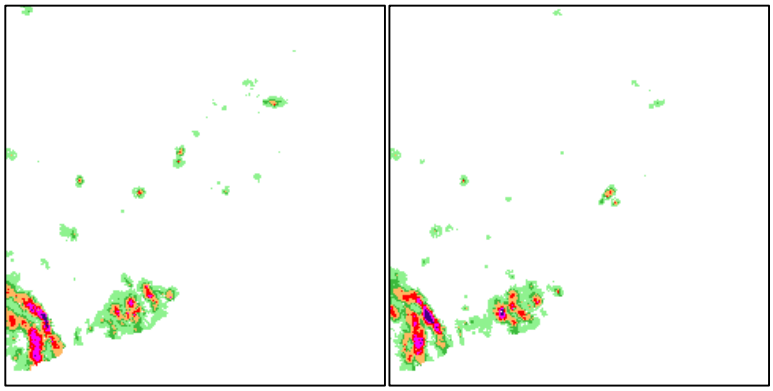


Figure 1 . The first line is the one-hour cumulative rainfall data of the China Meteorological Administration and its spatial resolution is 0.1°(about 10km). The second line is the rainfall data in our dataset. In order to compare the two, we interpolated and accumulated our rainfall data so that its time and spatial resolution is the same as the rainfall data of the China Meteorological Administration. The accumulation end time of rainfall in the first and second columns is 20180724 01:00 and 20180724 02:00 respectively.

Sorry, our description in the article is not reasonable enough.
In future research, we will use 1-hour resolution rainfall data from China Meteorological Administration. It is obtained by fusion of high-density automatic station data and CMORPH precipitation products through probability density function and optimal interpolation (PDF + OI). It can more accurately reflect the ground truth. As shown in Figure 1, by comparing the rainfall data in our dataset with the 1-hour resolution rainfall data released by the China Meteorological Administration, we found that although our data can roughly reflect the ground truth, its edges still be too smooth. We have modified part of the description in our article.

In the Section 1, the modifications are as follows:
Existing rainfall nowcasting methods based on machine learning and deep learning are difficult to achieve rainfall nowcasting with high spatiotemporal resolution. To solve this problem, We set forecast target to the grid data interpolated from rainfall data in dense automatic weather stations. Since the original data is directly measured by the rain gauge, the grid data will reflect the real-world rainfall distribution as much as possible. In the grid data, the forecasting time resolution and spatial resolution are 30 minutes and 5 km, respectively, which can meet the high spatiotemporal resolution requirement.

Point 2: Line 91 : "Besides, a high-pass filter is constructed to remove data below 15 dBZ"
Do you mean **:** A threshold is applied to remove data below 15 dBZ ?
Response 2: Yes. Precipitation generally does not occur in areas where the radar echo is lower than 15dBZ.

Point 3: Figure 1.a : Please add the spatial scale on the figure, I guess the area represents a part of Nord East of China. Provide also a color scale.

Response 3: We added the spatial scale and the color scale to Figure 1.

In the Figure 1, the modifications are as follows:

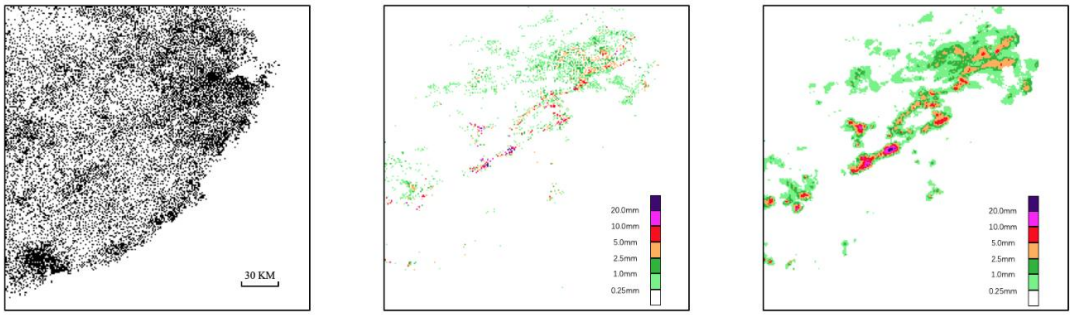


Figure 1. (a)Distribution map of automatic weather stations. Automatic weather station rainfall data of 09:00-09:30 on 23 May 2017 (b) and data after interpolation (c).

Point 4 : Line 98 : "*We use Inverse Distance Weight*"

Kriging technique could be more suitable

Response 4: Kriging interpolation method is better than IDW. We believe that the rainfall data of China Meteorological Administration in Response 1 will be better than the interpolation data of the above two methods. We also hope to directly use the rainfall data of the China Meteorological Administration in further research.

Point 5 : Line 99 : "*With such high-density data interpolation"*

Please provide an estimation of that density (number of rain gauges / km2)

Response 5: 0.02/km^2(0.5/grid).

Our original precipitation data is composed of data from 13,655 automatic weather stations. Half of the 240*240(1200km*1200km) grid points are oceans, which do not belong to the interpolation area. Calculated as follows:

$$\frac{13655}{1200 * \frac{1200}{2}} \approx 0.02\ /km^2$$

$$\frac{13655}{240 * \frac{240}{2}} \approx 0.5\ /grid$$

Point 6 : Figure 2 : What is the role of the first subsampling layer between inputs and the first TrajGRU layer in Encoder modules?

Response 6: It is used to extract low-level spatial features of the input data. In this layer, the size of input data is 1*240*240(Channels, Height, Width), and the size of output data is 8*80*80(Channels, Height, Width).

Point 7 : Line 173 : Implementation details
How do you train your model? Did you train each module separately or is it end-to-end learning? Please provide details.
Response 7: In this article, our deep learning models are all trained end-to-end.
In the Section 3.2, the modifications are as follows:
The proposed neural networks are implemented with Pytorch and are trained end-to-end. All network parameters are initialized with a normal distribution. All models are optimized with L2 loss, and they are trained using the Adam optimizer with a starting learning rate of 10^-4.

Point 8 : Section 4.1 : In fact, there are 2 sections dedicated to data processing (section 2.1 & section 4.1). It's quite confusing I propose to merge these sections
Response 8: Sorry, we have some descriptions of the data processing in section 4.1, which makes it easy to be considered as a section of data processing.
In the Section 4.1, the modifications are as follows:
Radar Echo Data: We use the data from the southeast coast of China. The data is stored in a 240*240 grid, and its spatial resolution is 5km. The time resolution is 6 minutes, and the time range includes May to September in 2018 and 2019. The original radar echo data is converted into pixel data. (delete)
Rainfall Data: The spatial range, spatial resolution, and time resolution of rainfall data from automatic weather stations are same as those of radar echo data. Its original time resolution is 10 minutes, due to the small value and sparse spatial distribution after interpolation, its time resolution is converted to 30 minutes. The original rainfall data is converted into pixel data. (delete)

In the Section 2.1, the modifications are as follows:
Therefore, we construct a singular point filter and a bilateral filter to filter the value domain and the spatial domain, which can effectively eliminate the pulsation and clutter while retaining the echo characteristics. Besides, a high-pass filter is constructed to remove data below 15 dBZ, and only data related to precipitation is retained. Since the data will be saved as image format, we convert the radar echo data into pixel data.

Z(. , .) is the value of this coordinate, and d_ij is the distance between the sample point and the interpolation point. lambda_ij is the weight of the sample point to the interpolation point. Finally, we convert the interpolated rainfall data into pixel data.

After modification, 2.1 and 4.1 describe data processing and dataset structure respectively. This paper structure is also used in some field-related papers, such as [1].

[1] Geng, Y.a.; Li, Q.; Lin, T.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. Lightnet: A dualspatiotemporal encoder network model for lightning prediction.

Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2439–2447.

Point 9 : Line 191 : what do you mean by "sparse spatial distribution after interpolation"
Response 9:

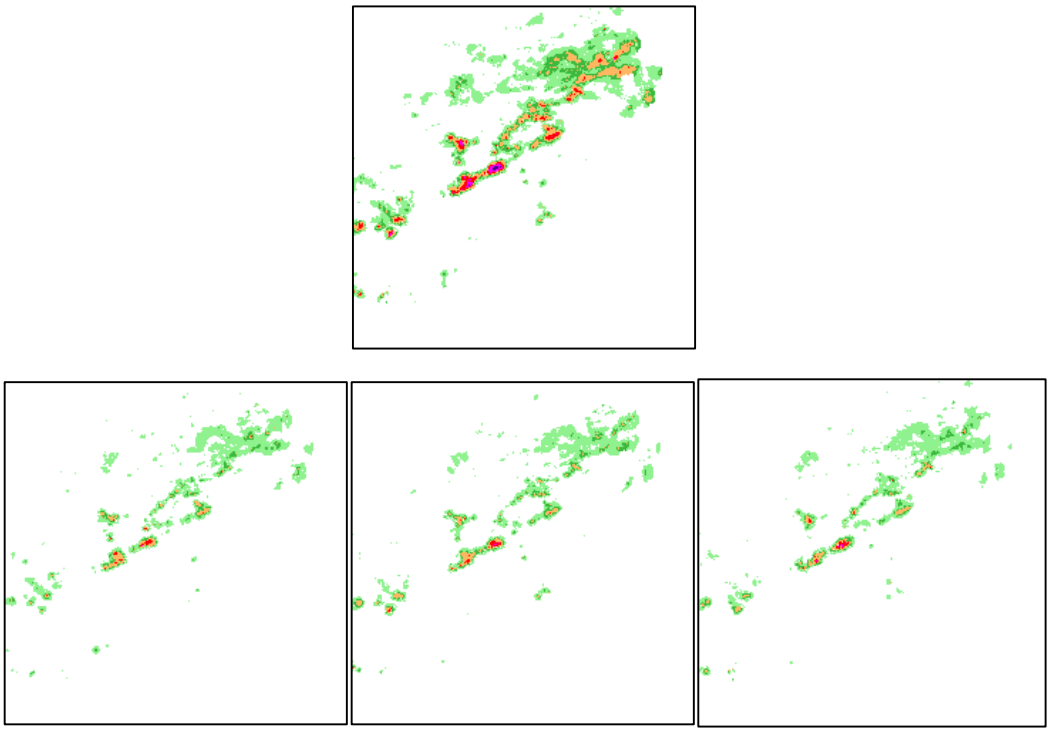


Figure 2 The first line is rainfall data with 30 minutes time resolution, and the second line is rainfall data with 10 minutes time resolution in the same time period.
As shown in Figure 2, compared with the precipitation interpolation data with time resolution of 30 minutes, the precipitation interpolation data with time resolution of 10 minutes has smaller values and more sparse spatial distribution.

Point 10 : Line 193 : "In the dataset, 207 days for training, 29 days for validation, and 57 days for testing."
How many images does it correspond to? Is it enough to train/validate and test ?
Response 10: In our dataset, the number of rainfall data in the training set, validation set and testing set are 6515, 628 and 1439 respectively, and the number of radar echo data in the training set, validation set and testing set are 48400, 7614 and 13467 respectively. Our dataset is large enough to complete the training, validation and testing.

Point 11 : Line 222 : *Finally, the performance metric includes 5 evaluation results of cumulative rainfall within 0.5 hours, 1 hour and 2 hours, and average values of the first two frames within 1 hour and four frames within 2 hours*

What is the difference between the cumulative rainfall within 1 hour and the average value within first two frames ?

Response 11: For the cumulative rainfall within one hour, we first add up the rainfall of the two frames in the first hour, and then evaluate. For the average of the first two frames, we first evaluate the first two frames separately, and then average the results.

The former is to evaluate the forecasting effect from the perspective of meteorological field. In the field of meteorology, the forecasting effect of evaluating rainfall within one hour is to evaluate the accumulated amount within one hour. The latter is to evaluate the forecasting effect from the perspective of spatiotemporal prediction. In the field of spatiotemporal prediction, the forecasting results are evaluated frame by frame. The evaluation result in a period of time is the average of the evaluation results of each frame in that period.

The research in this article belongs to the cross-field research, and verifying its effect needs to consider the characteristics of each field.

Point 12 : Line 232 : "*The dual-input method can make the rainfall nowcasting obtain the support of two kinds of data simultaneously*"

What does it means ?

Response 12: Sorry, our description is not accurate enough.

In the Section 4.3, the modifications are as follows:

We try three deep learning model input methods: rainfall data single input, radar echo data single input, and rainfall data and radar echo data dual input. The single input method of rainfall data is similar to the radar echo extrapolation method, which forecasts its future development by past data. The single input method of radar echo data is used to detect whether the radar echo data is effective for rainfall nowcasting. The dual-input model can simultaneously obtain the meteorological spatiotemporal feature of radar echo data and rainfall data. Three deep learning models are tried for each input method, including ConvLSTM, TrajGRU and PredRNN. The experiment contains 9 deep learning methods, and the best method is dual-input dual-encoder RNN which uses TrajGRU, namely RN-Net.

Point 13 : Lines 248-251 ; *Compared with WRF model rainfall nowcasting, RN-Net's TS within 0.5 hour, 1 hour and 2 hours of 0.25mm as the threshold are increased by nearly 4 times, 3 times, and 3 times respectively, and RN-Net's CSI within 1 hour and 2 hours of 0.25mm as the threshold are increased by nearly 4 times and 3 times respectively*

What does it means ?

Response 13: In the effect evaluation of rainfall nowcasting, CSI and TS are the two most important metrics. CSI and TS have the same calculation formula. TS evaluates accumulated rainfall in the period, and CSI is the average value of multiple frame evaluations in the period. "*RN-Net's TS within 0.5 hour, 1 hour and 2 hours"* means the TS of cumulative rainfall within 0.5 hours, 1 hour and 2 hours.

"*RN-Net's CSI within 1 hour and 2 hours*" means the CSI of the first two frames within 1 hour and four frames within 2 hours. In addition, their calculation needs to set a threshold, and points greater than the threshold are added to the area of the evaluation calculation.

Point 14 : Line 261 : *It is due to the fact that spatiotemporal features of the rainfall data are too small to support their forecasts*
Please could can you clarify what it means: *spatiotemporal features of the rainfall data are too small*
Response 14: Sorry, our description is not accurate enough.
The main factor that forms radar echo is the backscattering of electromagnetic waves by clouds in the atmosphere and various water vapor condensates in precipitation. The rainfall data is only determined by the actual surface rainfall. Compared with radar echo data, rainfall data contains fewer meteorological features.
In the Section 4.3, the modifications are as follows:
Comparison of Different Input Methods: The rainfall extrapolation method which similar to the radar echo extrapolation method is not good, and even worse than methods that use radar echo data as input. It is due to the fact that meteorological features of the rainfall data are too few to support their forecasts. The dual-input methods are better than the single-input methods.

Point 15 : Tables 3 – 7 : I wonder if it is really necessary to present all of Tables 3 to 7. They could be summarized, if not add sentences to comment on them.
Response 15: We think it is necessary to list all the data in the five tables.
The 5 tables respectively list 5 evaluation results of cumulative rainfall within 0.5hours, 1 hour and 2 hours, and average values of the first two frames within 1 hour and four frames within 2 hours. This helps to reflect the forecast effect from the two perspectives of spatiotemporal prediction and meteorology.
The items in the table also help to reflect the forecast effect comprehensively.We choose to use the thresholds 0.25mm, 1mm, 2.5mm to calculate these metrics. The threshold setting refers to the rainfall level, and the corresponding relationship is shown in Table 1.

Table 1 Correspondence between threshold and rainfall level.

Half-hour of rainfall(mm)	rainfall level
$r < 0.25$	No or hardly noticeable
$0.25 \leq r < 1$	Light
$1 \leq r < 2.5$	Light to moderate
$2.5 \leq r$	Moderate or greater

For the first time, we use spatiotemporal prediction method to realize short-term precipitation forecast. Sufficient experiments are needed to explore the effects of different RNNs and input methods on the prediction results.
We are the first to realize the rainfall nowcasting by the spatiotemporal prediction method. Full experiments are needed to explore the impact of different RNNs and input methods on the forecasting effect.

Point 16 : Figure 3 : It is difficult to exploit this figure. For example in fig3.b at t+3 does RN-Net better than RE-RFConvLSTM ? It would be better to zoom in so that only the bottom left part of the figure is displayed. It would also be useful in each figure to display the MAE or MSE
Response 17: Sorry, our figure showing the forecasting effect are not clear enough. We think it might be better to show the entire image. There is also precipitation in other parts of the real data and WRF model forecast results. We mark the CSI of 0.25mm as threshold for each frame in the figure, so that the forecast effect can be displayed more clearly.

In the Figure 3, the modifications are as follows:

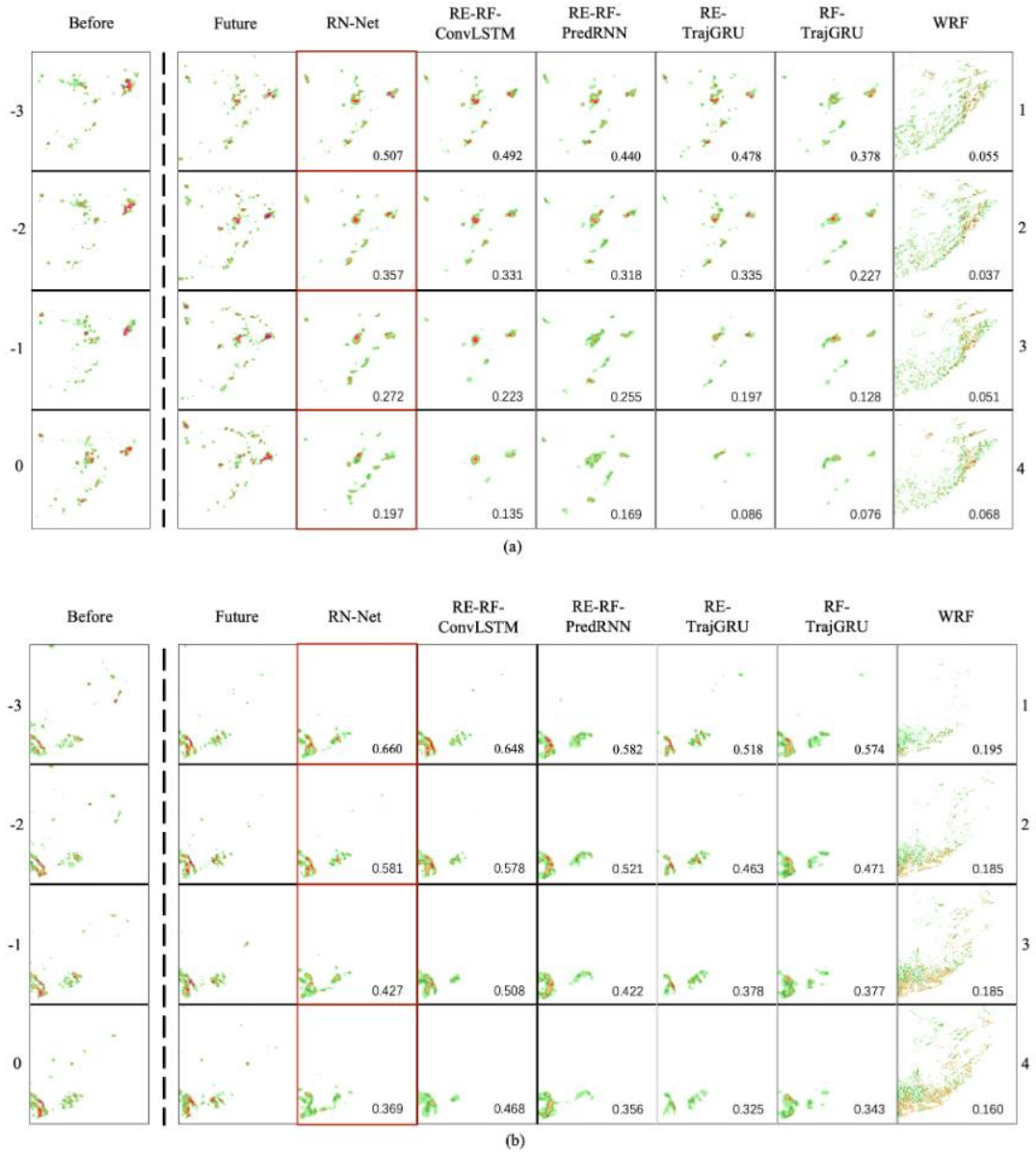


Figure 4. Visualize two representative rainfall nowcasting cases. In each example, from left to right are the actual rainfall data of the past two hours $[RF_t]_{t=-3}^{0}$, the actual rainfall data of the next two hours $[RF_t]_{t=1}^{4}$ and rainfall nowcasting $[\widehat{RF}_t]_{t=1}^{4}$ made by RN-Net, RE-RF-ConvLSTM, RE-RF-PredRNN, RE-TrajGR , RF-TrajGRU and WRF models.The value in each forecast frame is the CSI with 0.25mm as threshold for this frame .

Point 17 : Lines :280-281: Do you mean : WRF model prediction results often have large-scale false alarms, which is consistent with the performance provided by evaluation metrics

Response 17: Yes.

Point 18 : Comment : To assess whether high-resolution interpolation makes sense, it may be of interest to evaluate performance metrics on a subset of pixels corresponding

to the locations of the rain gauges. In this case, it is possible to use the true values of RF and not its interpolated value to calculate the error.

Response 18: Our rainfall nowcasting method uses the spatiotemporal prediction method in computer vision. The spatiotemporal prediction method uses past image data to predict future image data. We believe that using the interpolated grid data for evaluation can better reflect the effect of the deep learning prediction model. In addition, due to the denser weather stations, such evaluation results roughly include the evaluation results using raw point data.

Point 19 : Lines 282-299 : The conclusion should be more developed. For example, you can discuss the validity of your model: could your model be applied to other regions of the globe? What is its capacity to generalize?

Response 19: As rainfall is affected by topography, climate and other factors, we believe that RN-Net may have poor generalization ability and cannot get accurate forecast results in other regions. It may be necessary to use a dataset containing many regional data for training to make the model obtain strong generalization.

In the Section 5, the modifications are as follows:

In future research, we will extend our current work to four aspects. First, we will enhance the quality of data, such as improving the interpolation method of rainfall data at automatic weather stations and the denoising method of radar data. Secondly, in order to improve the generalization ability of the model, we will increase the scale of the dataset and expand the area included in the dataset. Third, we will add more input data to provide more meteorological spatiotemporal features for forecast, such as meteorological satellite data and various data related to precipitation in WRF model. Finally, we look forward to future work trying other deep learning networks to improve forecasting effects, such as novel RNN and GAN.