Thank you for your interest in our paper and for your helpful remarks. We have corrected structural and grammar issues. The model architecture, training and evaluating method, and discussion about relevance of our results have been added in the revised manuscript. The point-by-point responses to all comments are as follows:

1. L9 – Change to 'predict precisely'

Response: The change has been made.

Changes in the manuscript: 'in the imminent future the rainfall rate affected by which is difficult to precisely predict precisely'

2. L13 – insert 'is' between that and suitable

Response: The change has been made.

Changes in the manuscript: 'it is important to train a data-driven model from scratch that is suitable to…'

3. L13 – change collect to collected

Response: The change has been made.

Changes in the manuscript: 'We collected three kinds of data (radar, satellite, precipitation) in flood season…'

4. L22 – Define SSIM

Response: We develop the acronyms SSIM as Structural Similarity (SSIM) and the definition of which is in eq.4:

$$Loss = -1 \times SSIM(y_{pred}, y_{true}) = -1 \times \frac{\left(2\mu_{y_{pred}}\mu_{y_{true}}+C_1\right)\left(2\sigma_{y_{pred}y_{true}}+C_2\right)}{\left(\mu_{y_{pred}}^2+\mu_{y_{true}}^2+C_1\right)\left(\sigma_{y_{pred}}^2+\sigma_{y_{true}}^2+C_2\right)}$$

Changes in the manuscript: 'we applied a modified Structural Similarity (SSIM) index as a loss function.'

5. L23 – explain the traditional Z-R relationship

Response: The change has been made.

Changes in the manuscript: 'the results outperform those of the traditional Z-R relationships that use logarithmic function to describe the non-linear relationships between radar reflectivity and rainfall rate'

6. What is NOAA's HRRR? There is a need to go through the entire paper and define all the acronyms in its first use, even for something as common as AI.

Response: National Oceanic and Atmospheric Administration (NOAA) High Resolution Rapid Refresh (HRRR). Other acronyms have been defined.

Changes in the manuscript: 'which is superior to High Resolution Rapid Refresh (HRRR) numerical prediction from National Oceanic and Atmospheric Administration (NOAA) when the prediction time is within 6 hours.'

7. L35 – Remove repeated citation

Response: The change has been made.
Changes in the manuscript: ‘Sonderby et al.(2020) proposed’

8. L35 – What is MetNet? Do not assume the reader knows

Response: We have explained the ‘MetNet’ in the revised manuscript.
Changes in the manuscript: ‘Sonderby et al.(2020) proposed a Neural Weather Model(NWM) called MetNet that uses axis self-attention (Ho et al., 2019) to discover the weather pattern from radar and satellite data. MetNet can predict the next 8 hours precipitation with a resolution of 1 kilometer in 2-minute intervals.’

9. L37 – Remove repeated citation and do this for every other instance where this is the case

Response: The change has been made and the entire paper has been thoroughly edited for English writing and grammar.
Changes in the manuscript: ‘Shi et al.(2015)’

10. L37 – How were Shi.et al able to achieve this? Explain their method of prediction of spatio-temporal predictions

Response: We have explained their method in the revised manuscript.
Changes in the manuscript: ‘Shi.et al (2015) treated the precipitation nowcasting as a problem of predicting spatio-temporal sequence and modified the fully-connected Long Short-Term Memory (FC-LSTM) by replacing the hadamard product with convolution operation in the input-to-state and state-to-state transitions. They believe that cloud movement is highly uniform in some areas, and convolution can capture these local characteristics. Therefore, the convolution operation in the input transformations and recurrent transformations of their proposed Convolutional Long Short-Term Memory (ConvLSTM) helps to handle the spatial correlations.’

11. L41 – Avoid the use of contractions eg. Haven’t

Response: The change has been made.
Changes in the manuscript: ‘and have not been applied to the numerous meteorological data’

12. L43 – Avoid using etc. if you cannot name more items

Response: The change has been made.
Changes in the manuscript: ‘Computer vision techniques have long been used in object detection, video prediction, and human motion prediction.’

13. L44 – L45 – incorrect use of tense, past (used) and present (will mislead). Check for grammatical errors throughout the paper

Response: The use of tense has been corrected and the entire paper has been proof read by an English first language checker.
Changes in the manuscript: ‘Song(2019) used image quality assessment techniques as a new loss function instead of the common mean squared error(MSE), which misled

the process of training and generate the blurry image.'

14. L45 – How is optical flow method related to Trans method? Or is it not related? Why is it mentioned.

Response: It is related to Ayzel et al 's work and has been moved to the right part.
Changes in the manuscript: 'Ayzel et al. (2019) designed an advanced model based on the multiple optical flow algorithm for QPN, but it still performs badly in the prediction of onset and decay of precipitation systems because Optical flow methods simply calculate the position and velocity of the radar echo with a constant velocity rather than consider the changing intensity of radar echo.'

15. L51 – Remove et.al

Response: We have removed 'et.al'
Changes in the manuscript: 'Given this background, from the perspective of atmospheric science, we build a multi-source data model (MSDM) with the aim to fully use multi-source observation data (for example, radar reflectivity, infrared satellite data, and rain gauge data)'

16. L59 – Rephrase "to train the deep learning model to learn"

Response: It has been rephrased.
Changes in the manuscript: 'To train a Deep learning model that can capture the precipitation characteristics of East China'.

17. L66 – Provide links for datasets/sources

Response: The links have been provided in the revised manuscript.
Changes in the manuscript:
'Radar data: http://data.cma.cn/data/detail/dataCode/J.0012.0003.html , AWS data: http://data.cma.cn/data/detail/dataCode/A.0012.0001.html , Himawari 8 satellite data: http://www.cr.chiba-u.jp/databases/GEO/H8_9/FD/index.html '

18. L74 – change "we compared" to a comparison was made

Response: The change has been made.
Changes in the manuscript: 'To test our method, comparison was made…'

19. L75 – What exactly is being predicted first? Needs more clarity

Response: We have clarified in the revised manuscript.
Changes in the manuscript: 'Due to limits on computational resource, we use few frames to predict the results in half an hour. Then, the output results are used to iteratively predict the radar echo in the next half an hour to achieve a lead time of 2 hours (Fig 4). For the baseline sequence-to-sequence models (ConvLSTM, Optical flow), we use first 5 frames (T_{-4}~T_0) to predict a sequence of the next 5 frames(T_1~T_5), and use this result to iteratively predict the remaining three sequences (T_6~T_{10}, T_{11}~T_{15}, T_{16}~T_{20}). For image-to-image models (U-Net, MSDM), we use frame T_0 to predict frame T_5, and use this result as input to iteratively predict the following frames (T_{10},

T_{15}, T_{20}).'

20. L84 – change to "makes it better to predict". The entire paper needs to be thoroughly edited for English writing and grammar.

Response: The change has been made and the entire paper has been thoroughly edited for English writing and grammar.

Changes in manuscript: 'make it better to predict…'

21. L84 – Are you trying to say that the satellite data is more coarse ? Or has low spatial resolution? Or are you referring to the time resolution?

Response: Yes, the temporal resolution of satellite data is more coarse. Its interval is 30 minutes. We try to express that it only has four frames in the lead time of 2 hours. Therefore, we can get the sequence of the following 2 hours through one prediction rather than iterative prediction.

Changes in manuscript: 'Also, the temporal resolution of satellite data is more coarse (30 minutes), so we can directly get the sequence of four frames in the following 2 hours through one prediction rather than iterative prediction. Optical flow can predict such short sequence quickly and shows great advantages in saving computing resources and avoiding error accumulation.'

22. L87-88 – Explain more what you mean by "increases level through recursive application" and justify your use of satellite data.

Response: We have explained in the revised manuscript. The justification of our use of satellite data is in the response of comment 21.

Changes in manuscript: 'Besides, the biggest drawback of convolution is that it smooths the characteristics of image, and the level of smoothness increases when applying convolution recursively in deep learning models. Therefore, to ease the smoothing of radar echo and preserve more details of precipitating systems, we decide to use the results of predicted satellite data by Optical flow in our model.'

23. L120-124 –There is no mention of how the model was trained and the results were validated. Ideally this information should be in the methods

Response: We add a new part 'Model description' that shows the architecture of each of our model, including parameters such as kernel size, padding, drop out, learning rate, optimizer, loss function etc. Also, another part 'Reference models' shows how we compare with other baseline models. 'Training and evaluating method of Multi-source Data Model (MSDM)' shows how we train and evaluate MSDM.

Changes in manuscript: We add three parts to explain how the model was trained and the results were validated.

24. L120 – It almost appears that the aim of the paper has not been clearly stated. You are using multiple sources of data and at the same time creating a multi-source data model (MSDM). This is very difficult to follow throughout the paper. Make this distinction clear

Response: We will explain our aim in the introduction part of the revised manuscript.
Changes in manuscript: 'On one hand, the current massive amounts of data are underutilized, on the other hand, scientists in the field of machine learning focus on pursuing high accuracy by increasing the complexity of models based on a single source of data. Given this background, from the perspective of atmospheric science, we build a multi-source data model (MSDM) with the aim to fully use multi-source observation data and find suitable machine learning algorithms for each type of data that can ensure accuracy while saving computing resources. Besides, due to the high degree of freedom and non-linearity of neural network, it is hard to apply physical constraints to theses machine learning models. Hence, we hope multi-source data will function as a proxy of physical constraints to guide the model in the training process.'

25. L125 – Use other metrices to evaluate model performance
Response: We introduce more metrics to evaluate model performances: CSI, HSS, FAR, RMSE, SSIM.
Changes in the manuscript: We add a new part 'Performance evaluation' to introduce the metrics we use to evaluate model performance. In the 'Results' part, the evaluation are shown in the form of table and graph.

26. L148 – Who made that claim? Citation ?
Response: Yu et.al (2018) made the claim that 'recurrent networks for sequence learning require iterative training, which introduces error accumulation by steps.'
Changes in manuscript: We add the citation. 'The ConvLSTM is prone to error accumulation due to the iterative training and needs massive computing resources (Yu et al., 2018).'

27. L195-210 - Include a thorough discussion to examine the relevance of your results and how it relates to other studies, previous methods used etc.
Response: We discuss the relevance of our models with other studies in the revised manuscript.
Changes in manuscript: In the 'Conclusions and discussions' part, we discuss the background of existing study and explain our contribution to this problem. We copy a small part here and the whole part has been revised in the manuscripts:
'As a conventional precipitation nowcasting method, Optical Flow has played a certain role in the forecast of advective precipitation. However, it performs poor on the prediction of advective precipitation due to the simplicity of its algorithm and the lack of use of existing massive data(Woo and Wong, 2017). Meanwhile, deep learning shows great advantages in processing vast amount of data. By using convolution and LSTM structures, deep learning algorithms are better at capturing spatiotemporal correlations. Nevertheless, the recurrent network (represented by ConvLSTM) for predicting spatial-temporal sequence are widely known to be difficult to train and computationally expensive (Yu et.al, 2018). Compared with the traditional spatial-temporal sequence task in the field of machine learning, such as moving mnist prediction, human position prediction, traffic flow prediction, REE task has its specific background and physical

constraints. Therefore, merely getting predictions with higher scores does not reflect the quality of the results. Wang et.al (2018,2019) designed state-of-art models to capture comprehensive correlations between spatial-temporal sequences. But when we apply them to the physics-based tasks represented by REE and QPN, we must evaluate its prediction from the point of atmospheric science. The prediction is of reference significance only when it is physically reasonable rather than with high scores. However, it is difficult to apply physical constraints to neural network due to its high degree of freedom and non-linearity. Hence, we input more kinds of data as features to the network with the intention that it could get more information through feature interaction. Therefore, we collect multi-source data and design MSDM. There could be a situation that when the model goes to a wrong way and tries to predict low radar reflectivity, the incorporated satellite data will balance it. We hope the multi-source data function as a proxy of constraint to the model. Solving sequence-to-sequence problem is computationally expensive, so we treat the QPN as an image-to-image problem and design MSDM based on a CNN (U-Net) with high efficiency and few parameters. The biggest advantage of MSDM is its transferability. Apart from satellite data, any other data (wind speed, pressure, temperature, etc.) can be used as input to the model in the future. Wind speed data could add dynamic constraints, and temperature data could add thermodynamic constraint…'

28. L195 – Can it be conclusively stated that looking at the problem through image-image prediction is better than focusing on the problem as spatio-temporal sequence problem. If that is the aim of your paper, then what is the conclusion? Deliberate further.

Response: We evaluate the four models in terms of 12 aspects to show the advantages and drawbacks of the models and of their combinations.
Changes in manuscript: We use a table to evaluate these models and discuss their advantages and drawbacks. The aim of our paper also be concluded in the 'Conclusions and dicussions'.

Reference:
Ho, J., Kalchbrenner, N., Weissenborn, D., and Salimans, T.: Axial Attention in Multidimensional Transformers, 2019.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., Springer International Publishing, Cham, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Yu, B., Yin, H., Zhu, Z.:Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, 2018