We appreciate the reviewer's recognition for our work and the valuable comments. We have followed all the comments and revised manuscript accordingly. Please check the following point-by-point responses in attachment. The reviewer's comments are in black, our responses are in blue, and the quotes from our manuscript are in italics.

[Comment]: Concentrations at a point can be affected by emissions from local as well as can be transported through long range transport. How do the authors make sure that while updating the emissions using the difference in gradients between the predicted and observed concentration, not only the local emissions but emissions from other regions having possibility of transport of concentrations to the said point are also updated?

[Response]: We thank the reviewer for the valuable comment and apologize for the lack of clarity in our manuscript. The deep learning process has considered the impact of transmission between different grids. The NN-CTM, which refers to U-Net branch in particular, employs the convolution neural network to utilize neighbor information effectively. We visualize a demo case of 3×3 convolution and 5×5 convolution in Fig. 1. In U-Net, the stacked of convolution can get the neighbor information with a bigger receptive field (e.g. stacking 5×5 convolution and 5×5 convolution can get a 9×9 convolution), the non-linear function (P-RELU) is employed to improve model fitting with nearly zero extra computational cost and little overfitting risk, and the batch normalization and dropout are employed to enhance the robustness of the model. We calculate that the receptive field of our model is 38×38 grid. In other words, the predicted pollutant concentration is related to its surrounding 38×38 grid's information, which represents the transmission between different grids. Therefore, as shown in Fig. 2, if the red gird has observation data, the surrounding blue grids' emission inventory within the receptive field will also be updated. At the same time, the grids with a longer distance will have a lower update weight. In extreme circumstances, if we have no observation data, our method will not work as we have no more information to adjust the emission inventory. If the observation data is denser, the emission inventory estimation is more accurate as it can consider more observation data.
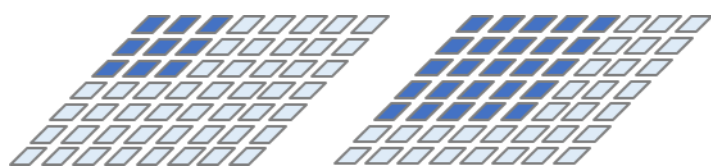

Figure 1: The visualization of convolution neural network (left: 3×3 kernel size, right:
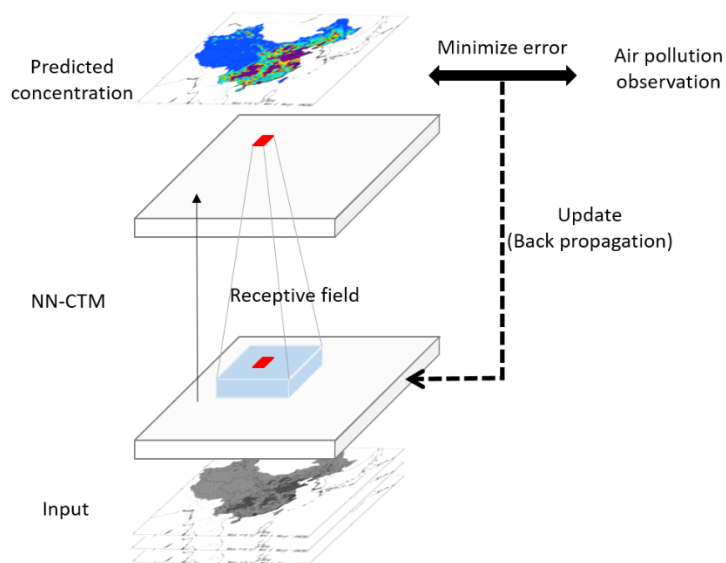
5×5 kernel size).

Figure 2: The visualization of neighbor emission update.

We have clarified the impact of transmission between different grids and the relation between observation data and emission inventory in the revised manuscript, as follows: (Section 2.2, Paragraph 3) *"In U-Net, the stacked of convolution can get the neighbor information with a bigger receptive field (e.g. stacking 5×5 convolution and 5×5 convolution can get a 9×9 convolution), the non-linear function (P-RELU) is employed to improve model fitting with nearly zero extra computational cost and little overfitting risk, and the batch normalization and dropout are employed to enhance the robustness of the model. We calculate that the receptive field of our model is 38×38 grid. In other words, the predicted pollutant concentration is related to its surrounding 38×38 grid's information, which represents the transmission between different grids. Meanwhile, the closer the distance, the greater the contribution."*

(Section 2.3, Paragraph 1) *"The observation data will help update the surrounding grids' emission inventory within the receptive field. However, in extreme circumstances, if we have no observation data, our method will not work as we have no more information to adjust the emission inventory. If the observation data is denser, the emission inventory estimation is more accurate as it can consider more observation data."*

[Comment]: The simulated concentrations will have some positive or negative bias based on a range of factors ranging from under/over prediction of meteorology,

chemical mechanism as well as emissions. The authors seem to add/subtract all this bias in predicted concentration to update the emission inventory. How correct is to then update the emission inventory based on biases in predicted concentration?

[Response]: We thank the reviewer for the valuable suggestion to improve the quality of the paper. There exist biases in meteorological conditions and chemical mechanism, which determines that we cannot attribute all the errors to the emission inventory. So we set constrain that the update rate of emission inventory to be a maximum of 200% compared with the prior emission for each grid when learning to ensure reasonableness. The update of the emission inventory is limited, thus leaving room for potential errors caused by other factors such as meteorology and chemical module. What's more, the updated emission inventory must be positive when learning.

On the other hand, since we are more concerned about the update of the emission inventory, we have adopted a series of methods to reduce the errors of meteorological and chemical modules as much as possible. For meteorology bias, data assimilation is adopted in WRF simulations based on observation data for the upper air and surface from National Centers for Environmental Prediction (NCEP) datasets, to ensure model performance within the benchmark range. As for chemical mechanism bias, the NN-CTM can well reproduce the CTM within an acceptable bias as introduced in Section 3.2. Such bias (<4%) is much smaller than that of simulation compared to observations which is normally more than 10% even 20%.

We have added more description for clarity in the revised manuscript, as follows:

(Section 2.3, Paragraph 3) *"Meanwhile, aiming at ensuring the reasonableness and effectiveness of estimated emission inventory, we set two constrains: 1) The update rate of emission inventory to be a maximum of 200% compared with the prior emission for each grid. There exist biases in meteorological conditions and chemical mechanism, which determines that we cannot attribute all the errors to the emission inventory. If the update ratio is very large, the NN-CTM cannot well reflect the correlation of the unseen data. Furthermore, the prior emission is accurate to a certain extent in terms of the spatial and temporal dimensions. 2）The updated emission inventory must be positive."*

[Comment]: Compared with the observed value, there are still errors in the simulation results using the new inventory, which is inevitable. Can the authors explain more about the constraints when inserting machine learning method into the emission update? In

other words, to what extent can we consider the improvement of model performance to be reasonable and acceptable.

[Response]: We thank the reviewer for the valuable comment and apologize for the lack of clarity in our manuscript. We have added more description about the constraints when learning. The update rate of emission inventory to be a maximum of 200% compared with the prior emission for each grid when learning to ensure reasonableness. The update of the emission inventory is limited, thus leaving room for potential errors caused by other factors such as meteorology and chemical module. What's more, the updated emission inventory must be positive when learning. Although we have set some constrains for emission inventory update, our method is affected by the observed data in terms of quality and sparsity. For quality, if the observed data is not accurate, the estimated emission inventory will be not meaningful. For sparsity, in extreme circumstances, if we have no observation data, our method will not work as we have no more information to adjust the emission inventory. If the observation data is denser, the emission inventory estimation is more accurate as it can consider more observation data. (Section 2.3, Paragraph 3) *"Meanwhile, aiming at ensuring the reasonableness and effectiveness of estimated emission inventory, we set two constrains: 1) The update rate of emission inventory to be a maximum of 200% compared with the prior emission for each grid. There exist biases in meteorological conditions and chemical mechanism, which determines that we cannot attribute all the errors to the emission inventory. If the update ratio is very large, the NN-CTM cannot well reflect the correlation of the unseen data. Furthermore, the prior emission is accurate to a certain extent in terms of the spatial and temporal dimensions. 2）The updated emission inventory must be positive."*

[Comment]: Page 5, Line 117. Change 'step' to 'steps'.
[Response]: We apologize for the typo, and we have modified.

[Comment]: Page 5, Line 127. Change 'the same to' to 'the same as'.
[Response]: We thank the reviewer for the suggestion and have modified.

[Comment]: Page 6, Line 140. Change 'sequence' to 'sequences'.
[Response]: We apologize for the typo and have corrected.

[Comment]: Page 6, Line 142. Remove 'a' in 'a supplementary …'.

[Response]: We apologize for the typo and have modified.

[Comment]: Page 8, Line 171. Should use 'based on the error …'.

[Response]: We apologize for the typo and have corrected.

[Comment]: Page 9, Line 199. Change 'has' to 'have'.

[Response]: We apologize for the typo and have corrected.

[Comment]: Page 9, Line 225. Notice the singular and plural forms.

[Response]: We thank the reviewer and have corrected. We also checked grammar carefully for the whole manuscript.

[Comment]: Page 11, Line 249. Change 'regions' to 'region'.

[Response]: We apologize for the typo and have modified.

[Comment]: Page 17, Line 341. Change 'achieve' to 'achieves'.

[Response]: We apologize for the typo, and we have modified.