



Supplement of

**Retrieving ground-level PM_{2.5} concentrations in China (2013–2021)
with a numerical-model-informed testbed to mitigate
sample-imbalance-induced biases**

Siwei Li et al.

Correspondence to: Siwei Li (siwei.li@whu.edu.cn) and Jia Xing (jxing3@utk.edu)

The copyright of individual parts of the supplement might differ from the article licence.

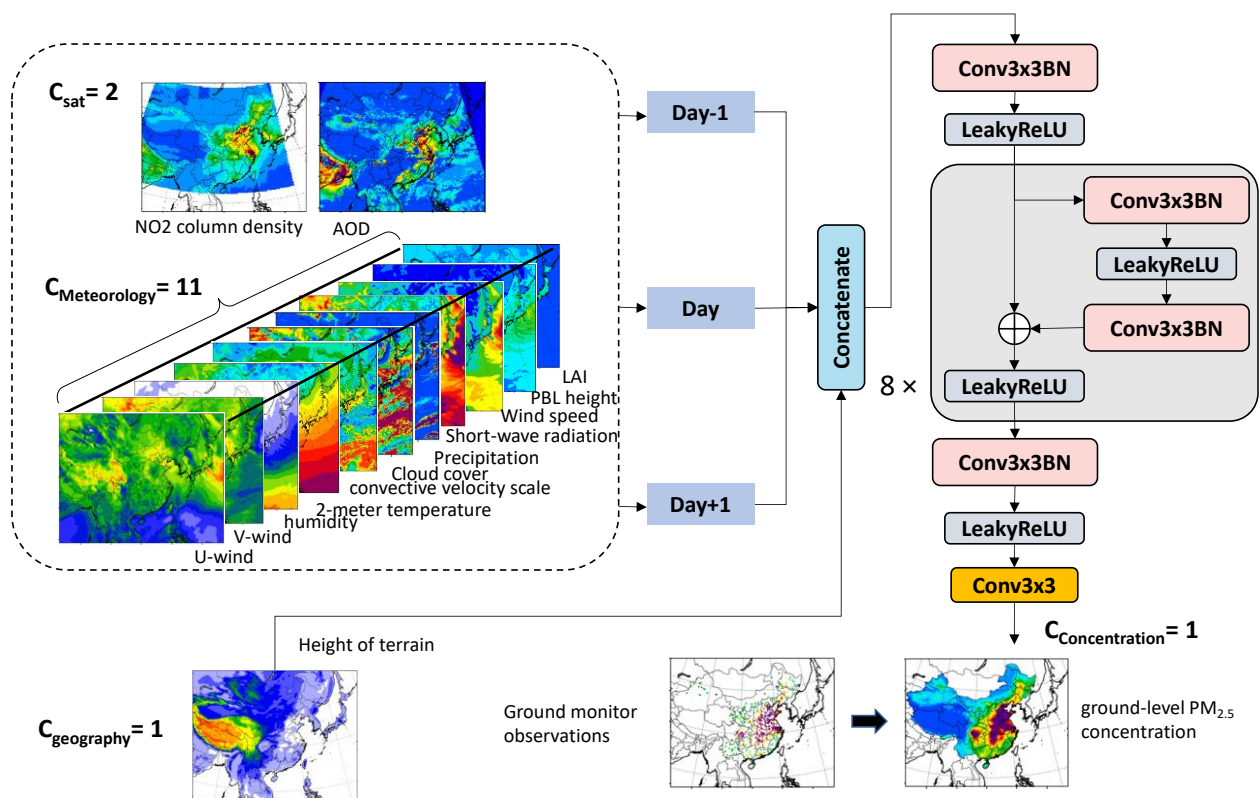


Figure S1. ResNet model structure

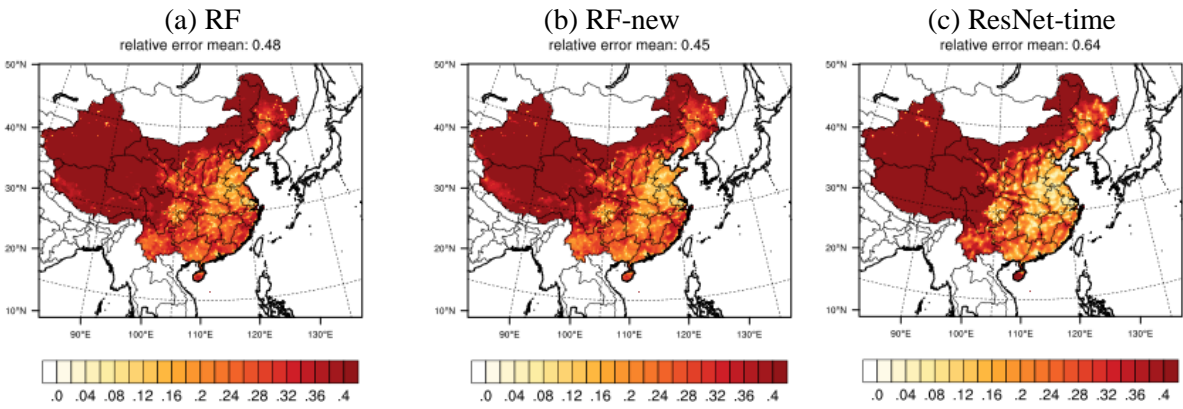


Figure S2. Comparison of relative errors in predicting ground-level PM_{2.5} using RF, RF-new, and ResNet-time models with the testbed.

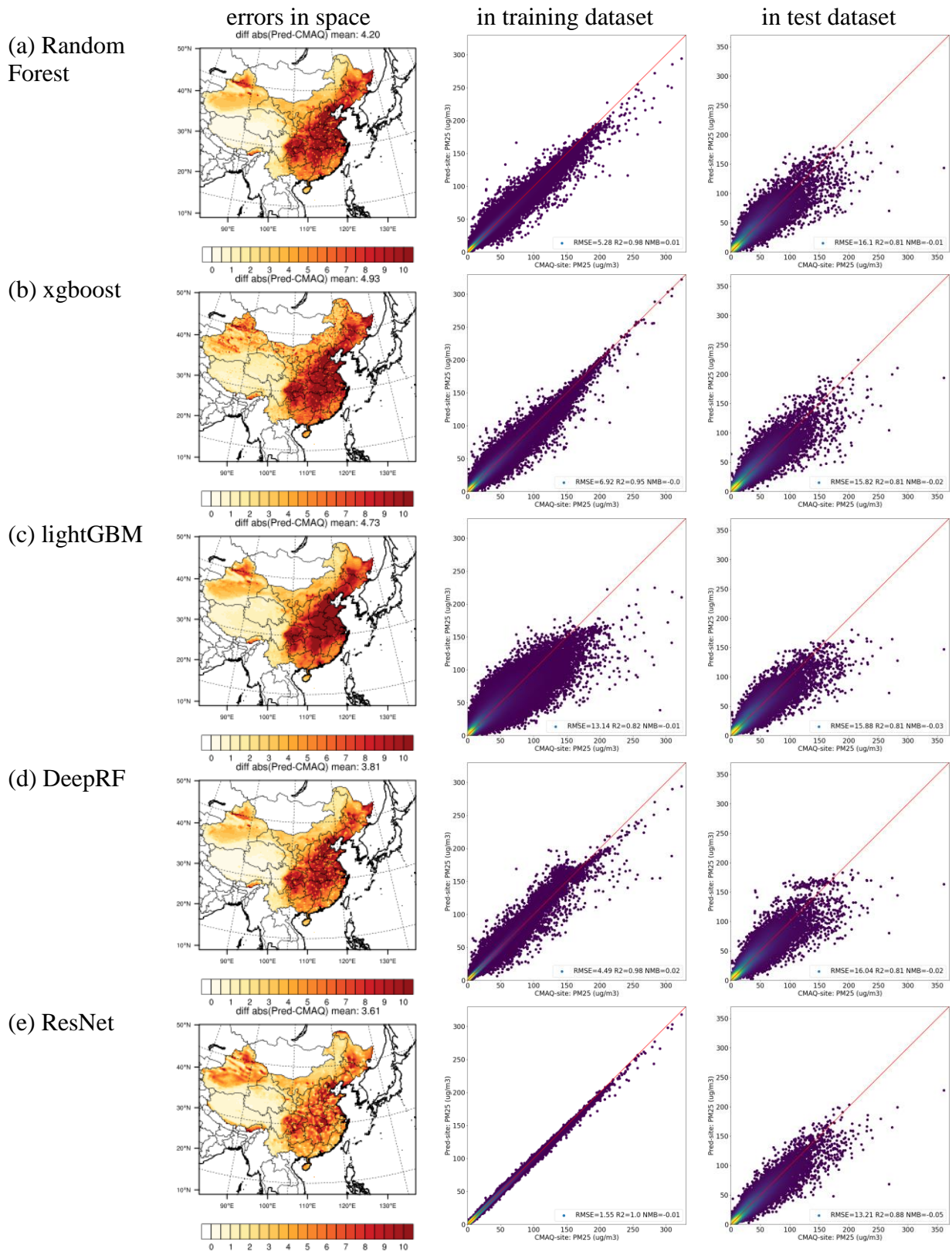


Figure S3. Errors in predicting surface PM_{2.5} with monitor-located grids training models

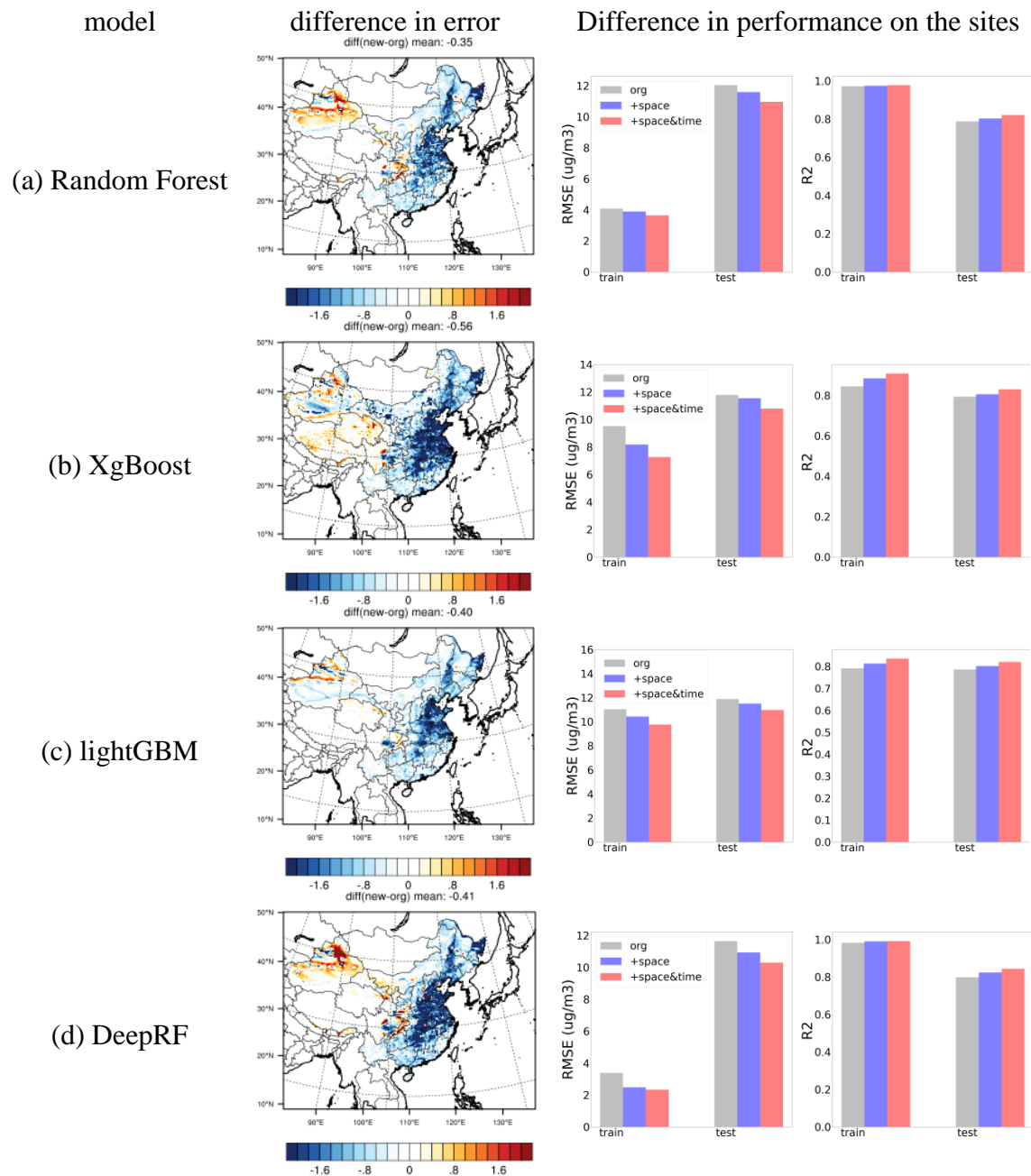
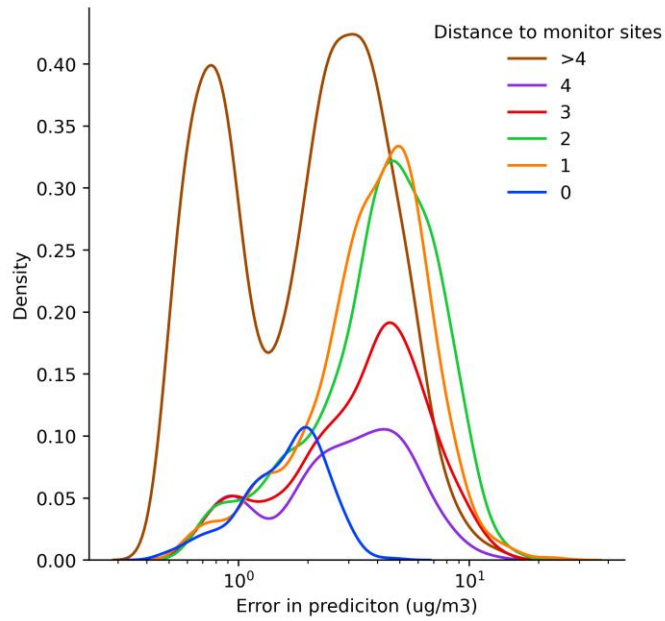
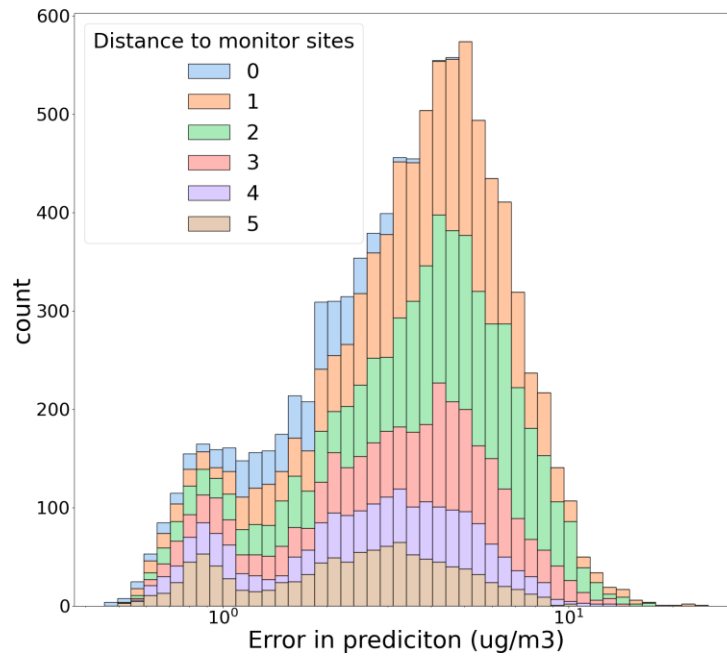


Figure S4. Improvement after implementing the features in surrounding grid cells (compared to each baseline model without spatiotemporal-neighbourhood features)



(a) by density



(b) by count

Figure S5. Error distribution across the distance to monitor sites (D-site) based on ResNet-time model

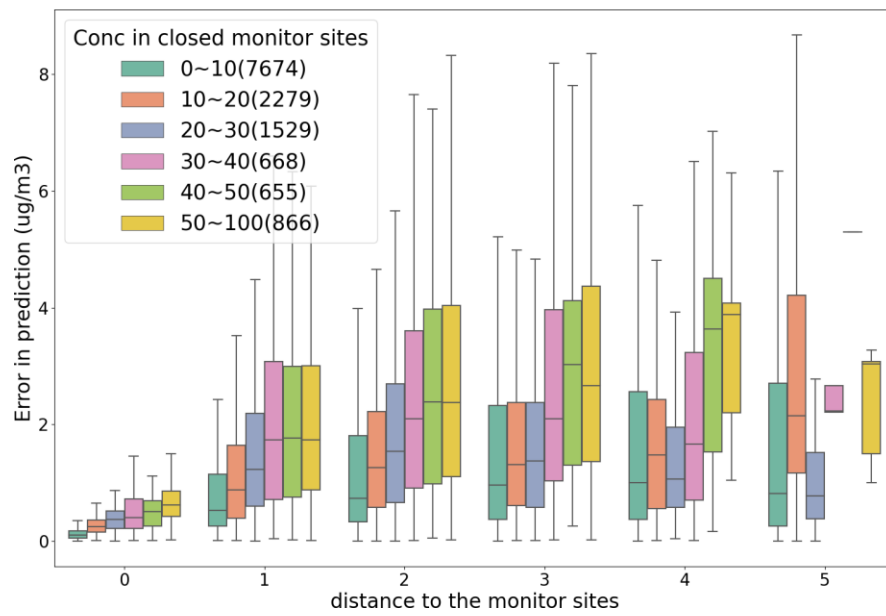


Figure S6. Error distribution across the monitor concentrations (B-conc) based on ResNet-time model

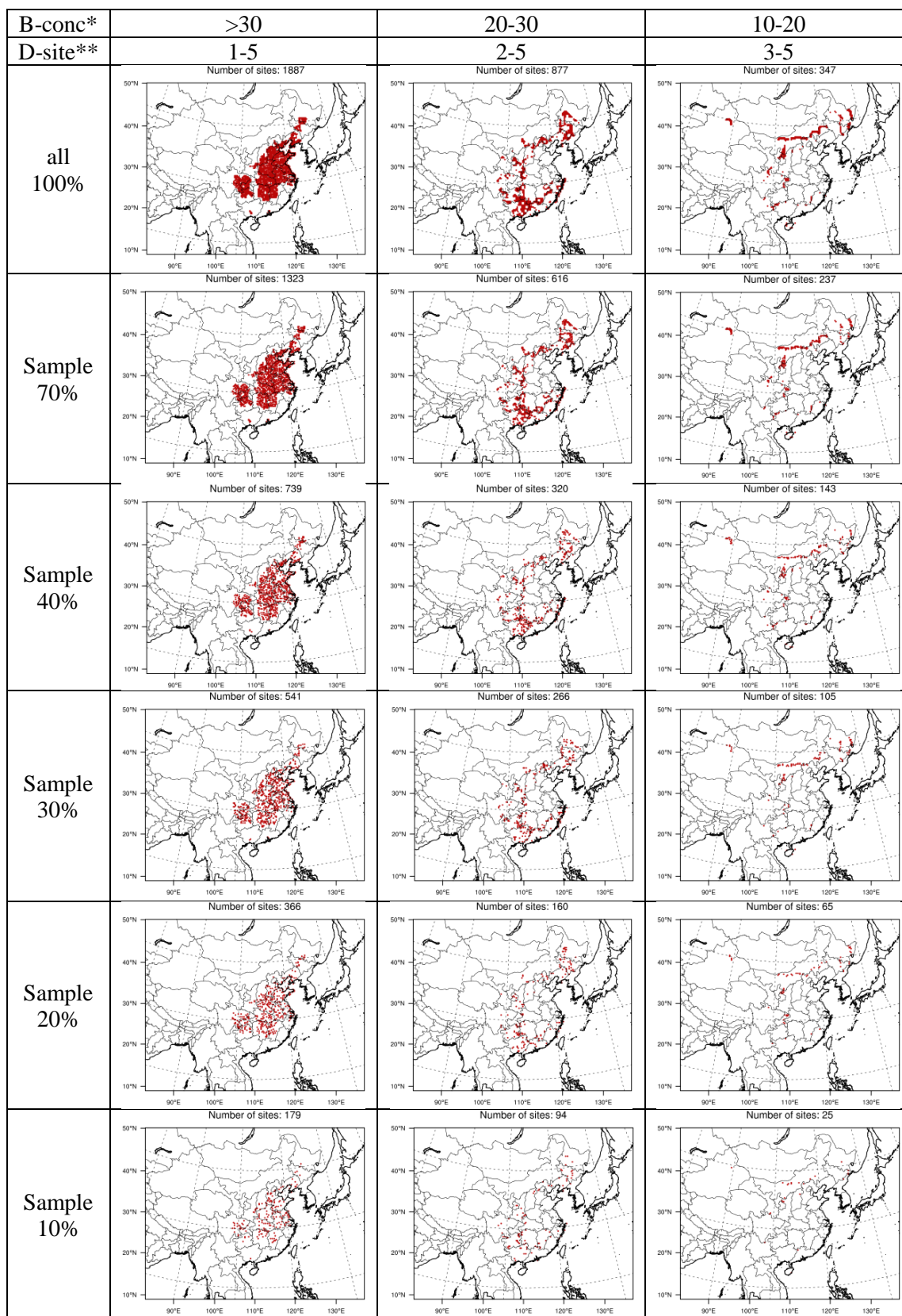
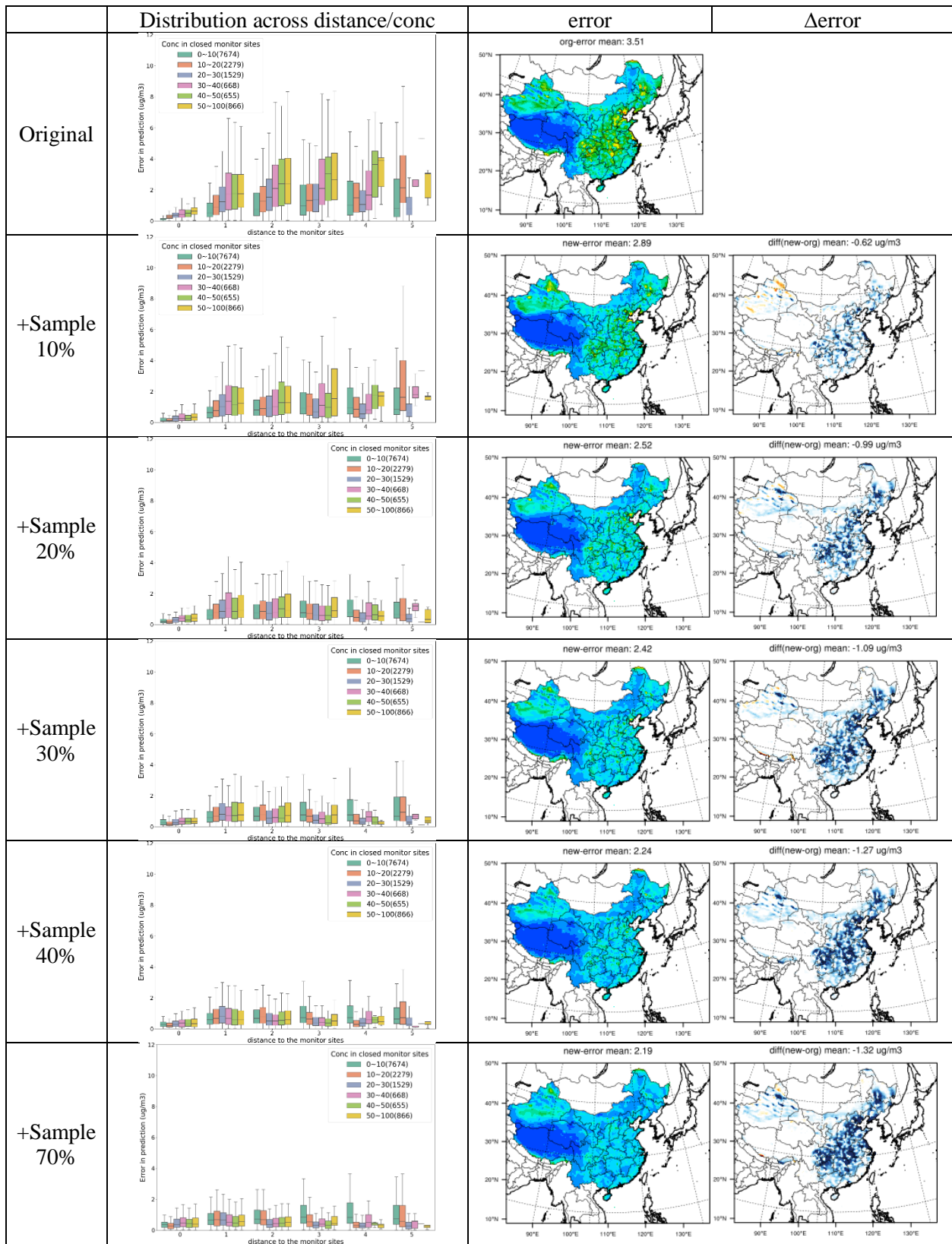


Figure S7. Spatial distribution of selected adding sites with certain levels of sampling (B-conc: conc in closed monitor sites; D-site: distance from monitor sites)



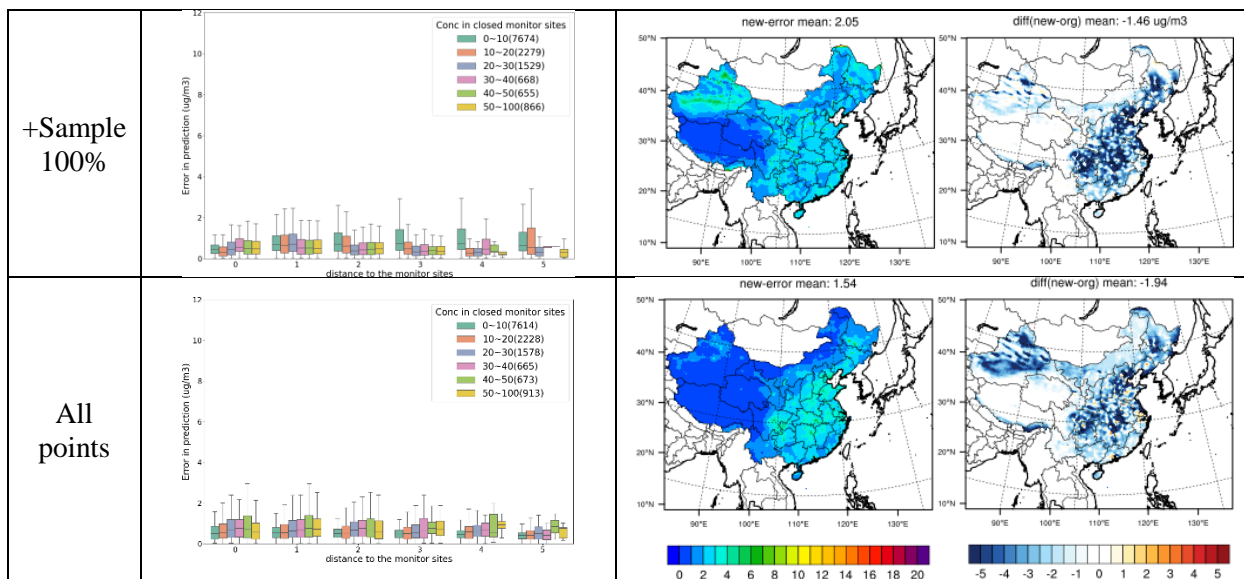


Figure S8. Improvement with selected adding sites with certain levels of sampling

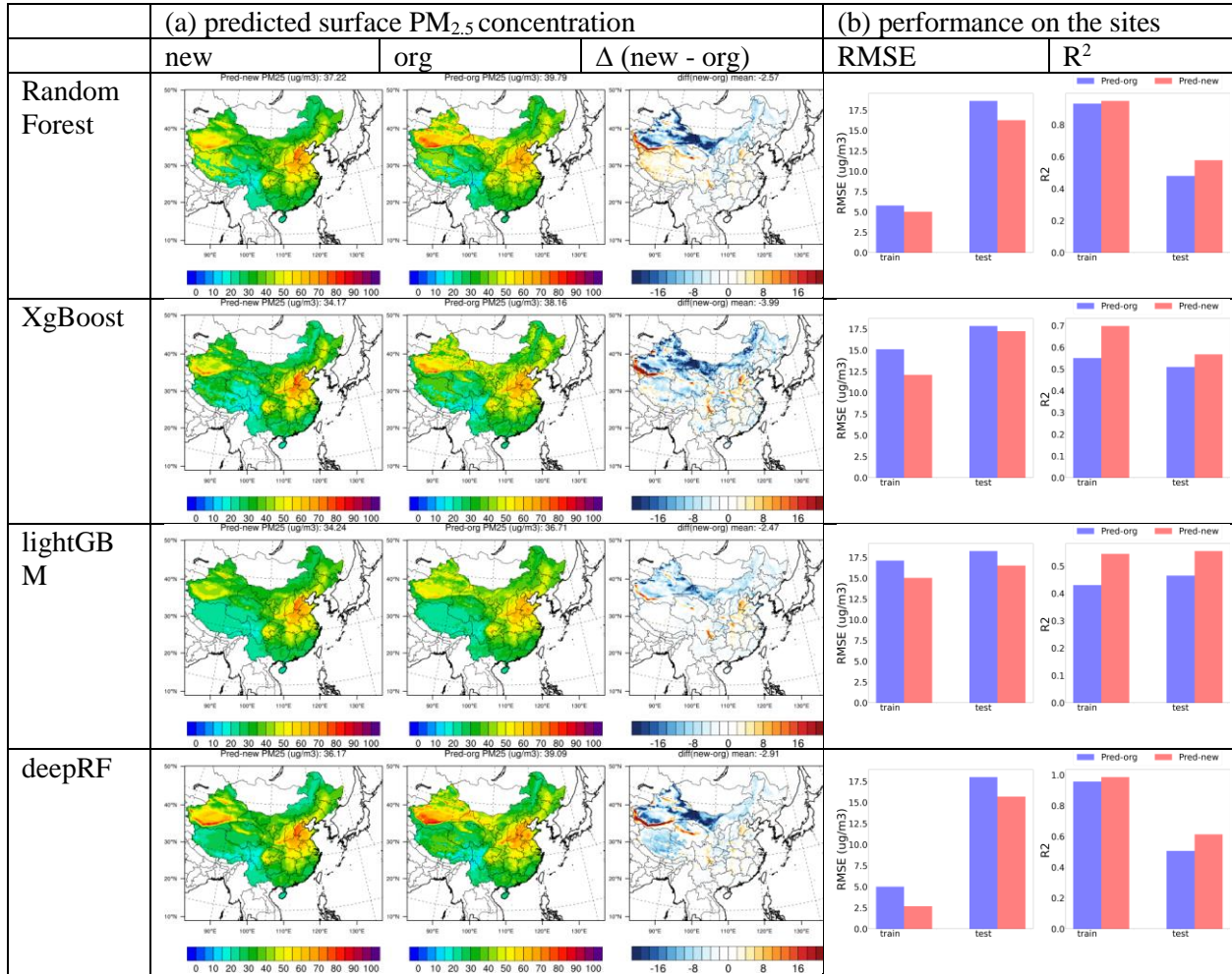
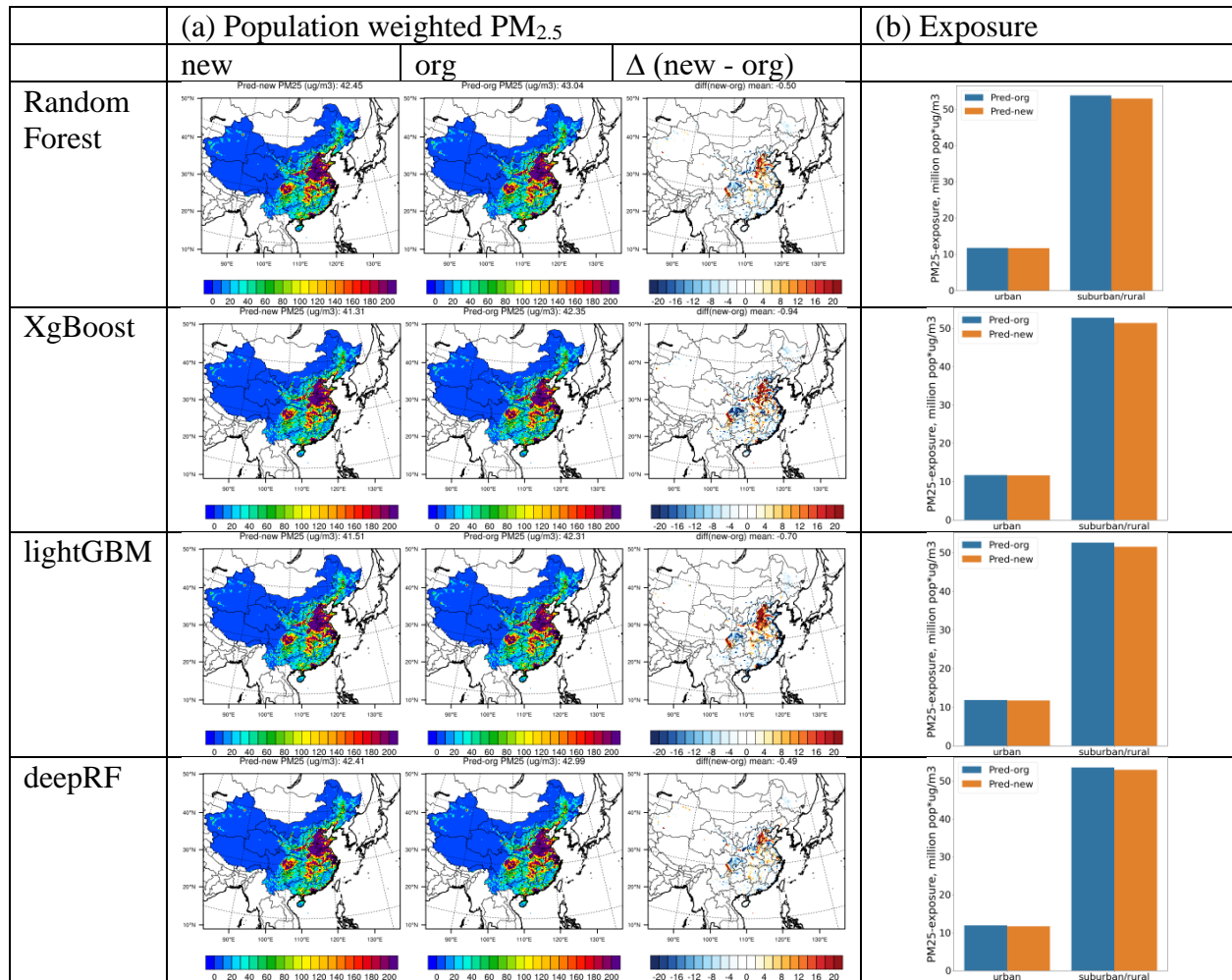


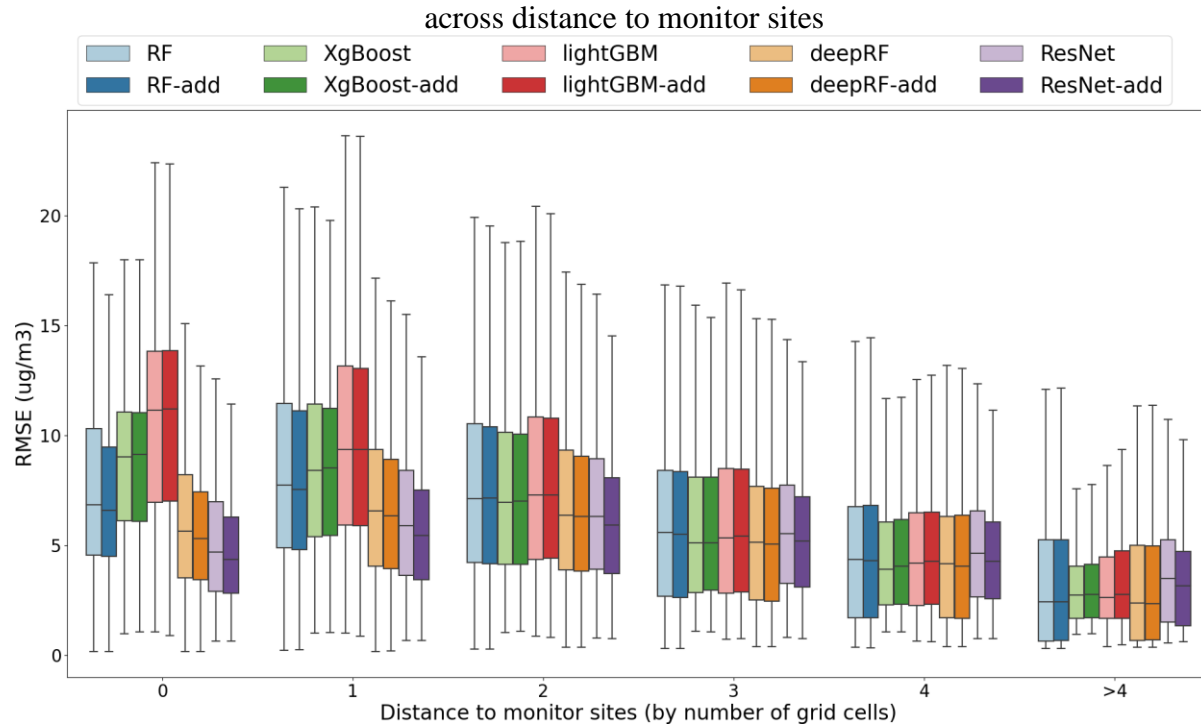
Figure S9. Improved performance with inclusion of spatiotemporal-neighbourhood features trained with real measurement dataset



*Note: within 5 grid cells

Figure S10. Uncertainties in estimation of PM_{2.5}-related exposure across China

(a) performance in scenarios with adding points



(b) +sample new sites during 2017-2021

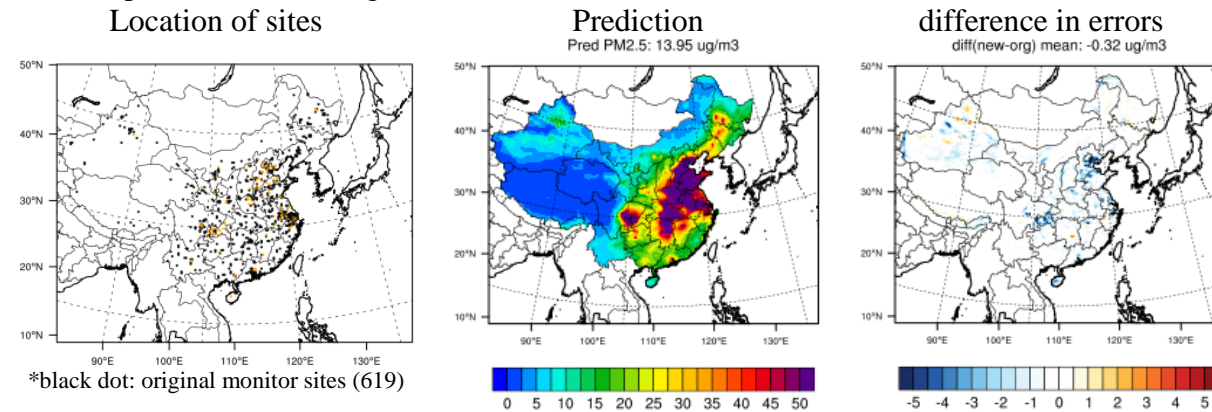


Figure S11. Improvement with the inclusion of new sites after 2017 in testing with CMAQ simulations

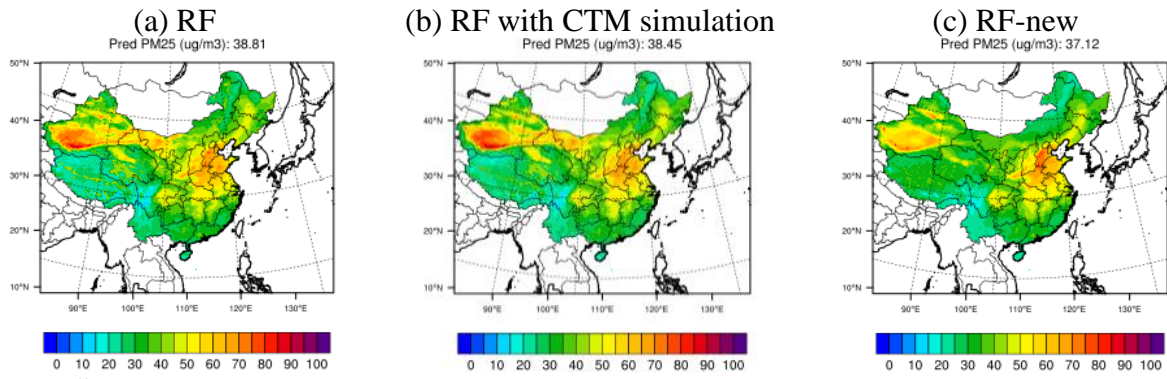
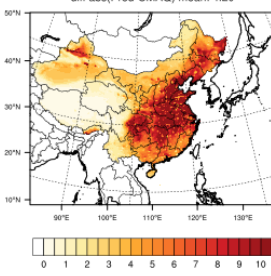
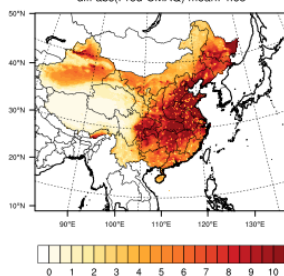


Figure S12. Comparison of predicted PM_{2.5} by adding simulation data and the proposed method in this study (2017 for example)

Error in original RF



Error in RF-noNO₂



Performance in training and testing dataset

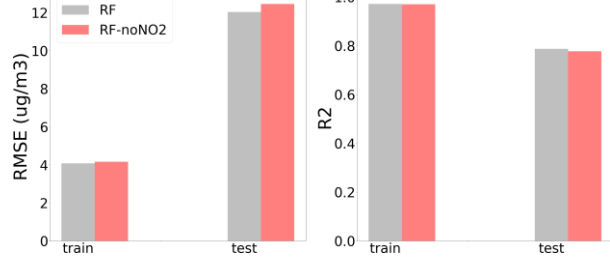


Figure 13. Comparison of model performance in predicting surface PM_{2.5} without NO₂ column feature