

## **Response to Comments on the Manuscript (egusphere-2023-2640)**

### **Diagnosing Ozone-NO<sub>x</sub>-VOCs-Aerosols Sensitivity to Uncover Urban-nonurban Discrepancies in Shandong, China using Transformer-based High-resolution Air Pollution Estimations**

Dear Referee,

We are grateful to the reviewer for their time and energy in providing helpful comments and suggestions which have significantly improved the manuscript. We have revised our manuscript according to all of the reviewer's comments to address these concerns in full.

The referee's comments and concerns are answered in detail point-by-point. The referee's comments are shown in black and the authors' responses are shown in blue.

#### **Reviewer #2 Comments to Author:**

This study developed a novel spatiotemporal deep learning model for concurrent prediction of three air pollutants (ozone, NO<sub>2</sub>, PM<sub>2.5</sub>). The authors used the generated fine-scale concentrations to assess urban-nonurban differences and ozone-NO<sub>x</sub>-VOCs-aerosols sensitivity in Shandong, China. To facilitate the analysis, interpretable machine learning was employed to handle nonlinearity and isolate impacts of drivers relating to ozone photochemistry. The methodology is solid, and the findings are important for the development of ozone control strategies, though a few issues remain.

1. Line 259: Please explain the possible reason why NO<sub>2</sub> has significantly lower out-of-site CV-R<sup>2</sup> (0.75) than ozone and PM<sub>2.5</sub> (>0.9), note that the out-of-sample CV results are comparable across all pollutants?

**Response:** The decreased  $R^2$  for  $\text{NO}_2$  in out-of-site cross-validation could result from the short atmospheric chemistry lifetime of  $\text{NO}_2$ , which leads to greater potential disparities in the relationship of satellite column density and surface  $\text{NO}_2$  between various monitoring stations. Meanwhile, previous studies also show the same problem. For example, Wei et al. (Wei et al., 2022) estimate the ground-level  $\text{NO}_2$  surveillance with an average out-of-city (out-of-sample) cross-validation  $R^2$  of 0.71 (0.93) using interpretable spatiotemporally weighted artificial intelligence. The same trouble of the underestimation of high values leads to the reduced evaluation metric.

References:

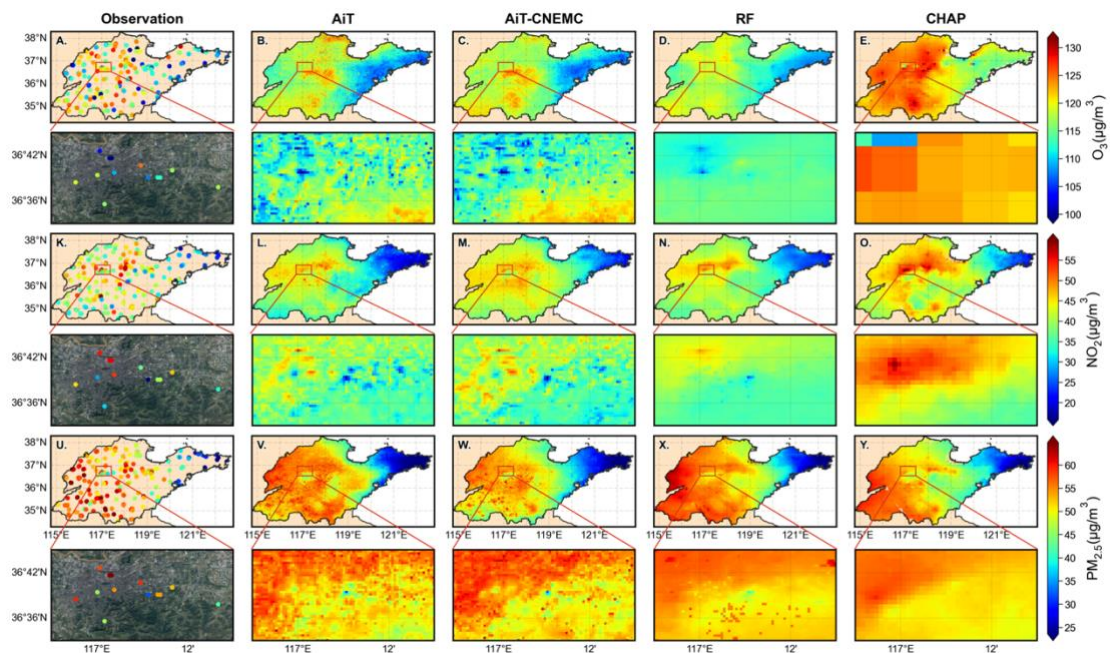
Wei, J., Liu, S., Li, Z., Liu, C., Qin, K., Liu, X., Pinker, R.T., Dickerson, R.R., Lin, J., Boersma, K.F., Sun, L., Li, R., Xue, W., Cui, Y., Zhang, C., Wang, J., 2022. Ground-Level  $\text{NO}_2$  Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence. *Environ. Sci. Technol.* [acs.est.2c03834](https://doi.org/10.1021/acs.est.2c03834). <https://doi.org/10.1021/acs.est.2c03834>

2. Lines 265-266: In evaluating stability and robustness of the model, it would be interesting to see if the CNEMC-trained model can obtain local concentration variations and interpretation outcomes similar to that from the CNEMC+SDEM-trained model.

**Response:** Thank you for your insightful comments. We have conducted the suggested comparison and incorporated the results into Figure 3, comparing multiple datasets. We apologize for the initial error in the verification of the CNEMC-trained model on the SDEM dataset, which has also now been rectified. Additionally, we have included a kernel density estimation result to visually represent these verification results. It is observed that the CNEMC-trained model exhibits only an acceptable degradation in predictive accuracy on the SDEM dataset compared to out-of-site cross-validation of AiT (Figure S6). The outcome also reveals a similar spatial gradient at the urban scale (Figure 3). Meanwhile, the

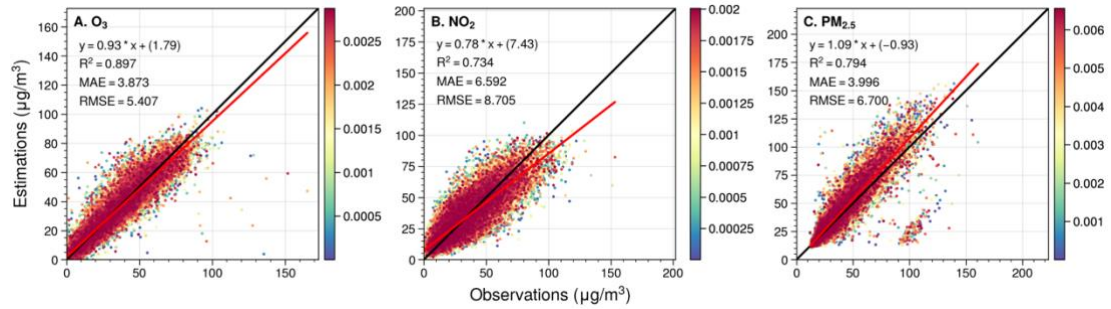
comparison results on a daily scale during sandstorms show that although the model trained with CNEMC data exhibits some overestimation or underestimation in certain areas, it demonstrates similar spatial distribution and temporal variation trends as the model trained with all data (Figure S9). These results reveal the reliability of my deep learning model and the promising prospect of continuously improving the model's generalization ability with more ground-level monitoring data.

The main revision is as follows in lines 294-296: "This spatial gradient is also captured by AiT trained with CNEMC data, revealing the reliability of the deep learning model structure", lines 327-329: "The model trained solely on CNEMC data is also capable of effectively capturing the drastic changes in air quality during the pollution episode (Figure S9)", and lines 637-641: "Meanwhile, the results between AiT trained with all data and that trained exclusively with CNEMC data across various spatiotemporal scales underscore the promising prospect for improving the model's generalization ability with more ground-level monitoring data and the growing space of methods."

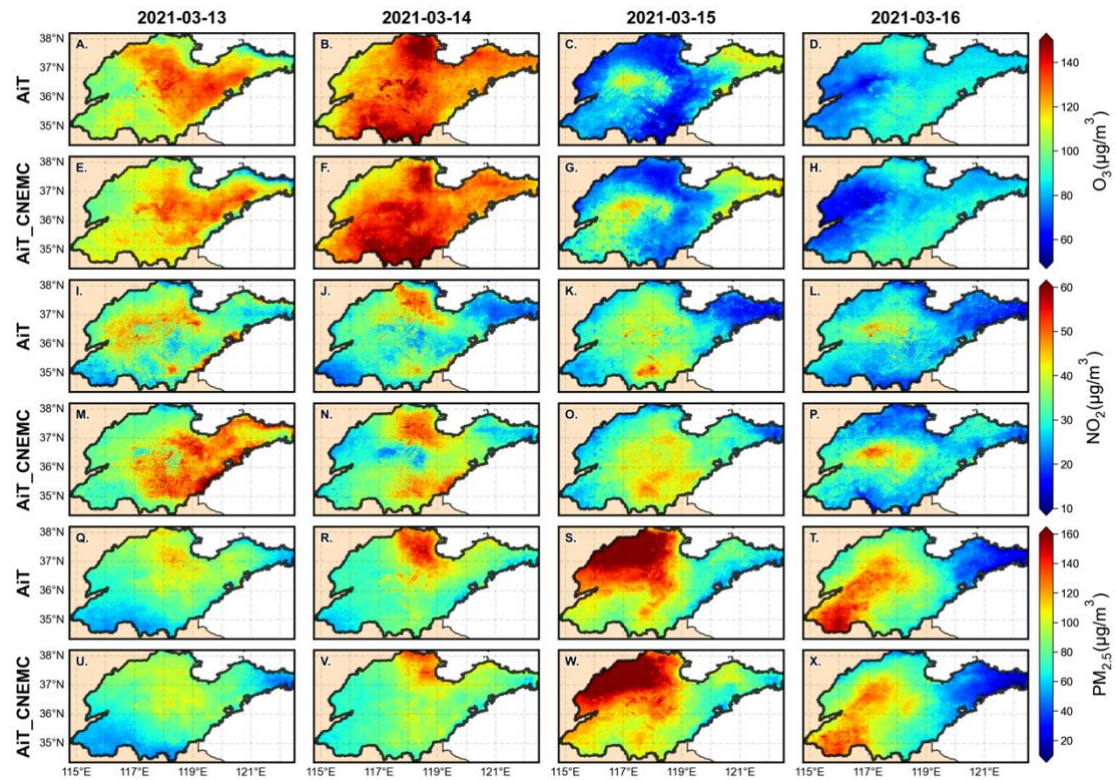


**Figure 3.** Spatial distribution of the annual mean (A-E) O<sub>3</sub>, (K-O) NO<sub>2</sub>, and (U-Y) PM<sub>2.5</sub> concentrations from observations, Air Transformer (AiT), CNEMC-trained

AiT, Random Forest (RF) and ChinaHighAirPollutants (CHAP), respectively, in 2019. The region enclosed by the red rectangular box corresponds to the zoomed-in maps of the satellite (© Tianditu: www.tianditu.gov.cn) and pollutant concentrations at a city scale for the capital city of Shandong Province, Jinan.



**Figure S6.** Validation for daily ground-level O<sub>3</sub>, NO<sub>2</sub>, and PM<sub>2.5</sub> concentration in the SDEM dataset based on the AiT model trained by monitoring data of CNEMC.



**Figure S9.** Comparison of spatial distribution between estimations from AiT trained with all data and AiT with CNEMC data during the dust storm.

3. Lines 322-323: The time span of the training data should be given, as that information is important to understand whether the good agreements between measurements and estimations reflect fitting or prediction performance.

**Response:** Thanks for your kind suggestions. We have added the time information of the training dataset in section 2.2 at lines 197-200: “The aggregated feature data from June 2019 to June 2021 were utilized to train and validate the model through cross-validation (CV), where the optimal model, trained based on out-of-sample CV, was used to estimate multiple pollutant concentrations during the study period, which was then employed for subsequent analysis.”.

4. How many monitoring stations are there in urban areas? A map highlighting the urban and nonurban areas is recommended for intuitive understanding.

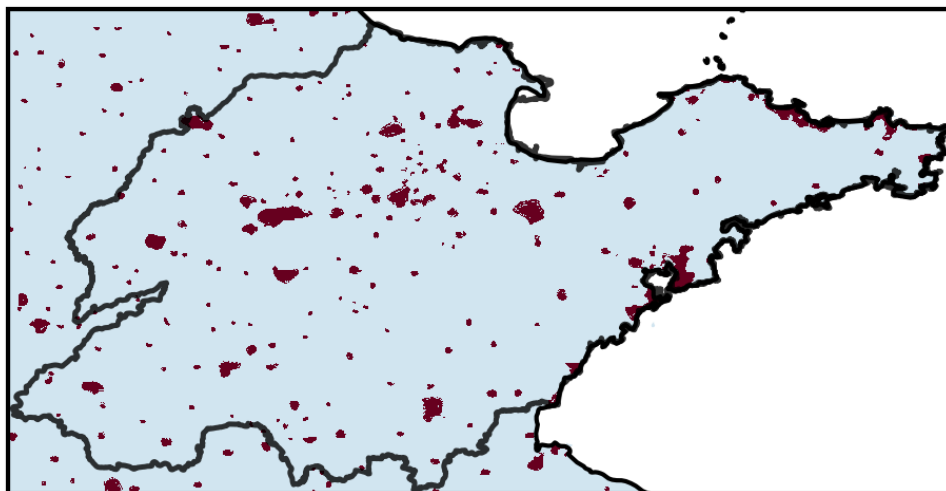
**Response:** We counted the number of monitoring sites in urban and non-urban areas in Table S4. The number of urban sites in 13 cities exceeds that in non-urban areas. Particularly in cities like JNA, LC, LY, QD, and YT, the disparity in the number of urban and non-urban sites is significant, leading to urban-nonurban differences that are contrary to those observed in AiT. We also added the map of urban extents in supporting information as Figure S11.

The main revision for the number of monitoring stations is as follows in lines 397-400: “The notable disparity between the number of urban and non-urban sites in cities such as JNA, LC, LY, QD, and YT results in a pattern of urban-nonurban differences that contrasts markedly with the observed in AiT (Table S4).”.

**Table S4.** The number of monitoring stations across urban and non-urban areas. (YT: Yantai, BZ: Binzhou, DY: Dongying, WH: Weihai, DZ: Dezhou, JNA: Jinan, QD: Qingdao, WF: Weifang, ZB: Zibo, LC: Liaocheng, LW: Laiwu, TA: Taian, LY: Linyi, RZ: Rizhao, JNI: Jining, HZ: Hezhe, ZZ: Zaozhuang)

City Name	BZ	DY	DZ	HZ	JNA	JNI	LC	LW	LY
Non-urban	9	2	10	6	2	6	7	2	8
Urban	7	11	14	14	17	15	15	1	14
City Name	QD	RZ	TA	WF	WH	YT	ZB	ZZ	
Non-urban	1	5	4	9	3	3	10	2	
Urban	11	5	7	15	7	18	6	8	

The main revision for the map of urban extents is as follows in lines 373-374: “The urban extents in Shandong Province in 2019 are depicted in Figure S11.”

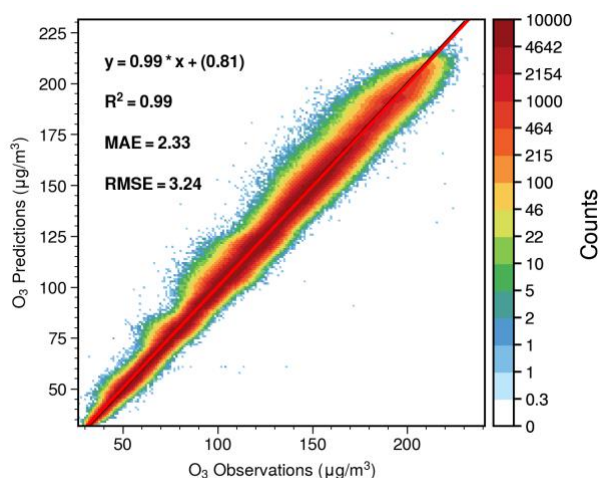


**Figure S11.** Urban extents (red) in Shandong province, China in 2019.

5. There is a lack of validation for the XGBoost model, given that reliability of interpretation outcomes should be based on the model with high accuracy.

**Response:** Thanks for your insightful comments. We added the results of 10-fold cross-validation as shown in Figure S16.

The main revision is as follows in lines 500-502: “As depicted in Figure S16, the performance of the XGBoost model is robust, evidenced by a high  $R^2$  value of 0.99 coupled with a low RMSE of  $3.24 \mu\text{g}/\text{m}^3$  and MAE of  $2.33 \mu\text{g}/\text{m}^3$ ”.



**Figure S13.** Results of 10-fold cross-validation in validation dataset based on XGBoost for modeling the nonlinear response of monthly O<sub>3</sub> variations to

meteorology and chemical indicators from 2019 to 2020.

6. Please provide more explanations for the SHAP interaction values. The statement “lower NO<sub>2</sub> ... could diminish the formation of ozone under high PM<sub>2.5</sub> concentrations” (Line 456) is difficult to follow. In Figure 7e, lower NO<sub>2</sub> and negative PM<sub>2.5</sub>-NO<sub>2</sub> SHAP interaction values are observed at lower PM<sub>2.5</sub> levels.

**Response:** Thanks for the reviewer’s suggestion. We have added more information in section 3.3 for easier understanding.

The main revision is as follows in lines 466-479: “The SHAP interaction plot in Figure 7e, f illustrates that the influence of NO<sub>2</sub> and HCHO on O<sub>3</sub> formation is not constant and is influenced by the levels of PM<sub>2.5</sub>. Typically, at a certain level of PM<sub>2.5</sub>, the lower NO<sub>2</sub> concentration, the stronger inhibition effect on O<sub>3</sub> production. This could be due to aerosols exerting stronger suppression through the HO<sub>2</sub> sink at lower NO<sub>x</sub> levels. As the concentration of PM<sub>2.5</sub> increases, often involving a concurrent increase in NO<sub>2</sub> as a key precursor, there is a greater need for higher levels of NO<sub>2</sub> to be converted into nitrous acid (HONO) through the heterogeneous uptake by aerosols. This process produces more OH radicals, which facilitate photochemical O<sub>3</sub> formation, offsetting the increased inhibitory effect of the HO<sub>2</sub> sink. Under high PM<sub>2.5</sub> concentrations, an increase in NO<sub>2</sub> along with a decrease in HCHO enhances their effect on the promotion of O<sub>3</sub> formation. This enhancement could be caused by increased titration of O<sub>3</sub> by NO, resulting from weaker conversion from NO to NO<sub>x</sub> through the RO<sub>x</sub> radical. Meanwhile, the impact of HCHO shifts from promotion to suppression as PM<sub>2.5</sub> pollution intensifies.”

7. Figure 8d shows that the NO<sub>x</sub>-limited regime dominates in urban areas. Please confirm.

**Response:** Thanks for your insightful comment. We sincerely apologize for any confusion caused by the inaccuracies in our previously stated conclusions regarding the ozone formation regimes. Upon closer examination, we acknowledge that our initial interpretation, suggesting an increased sensitivity of

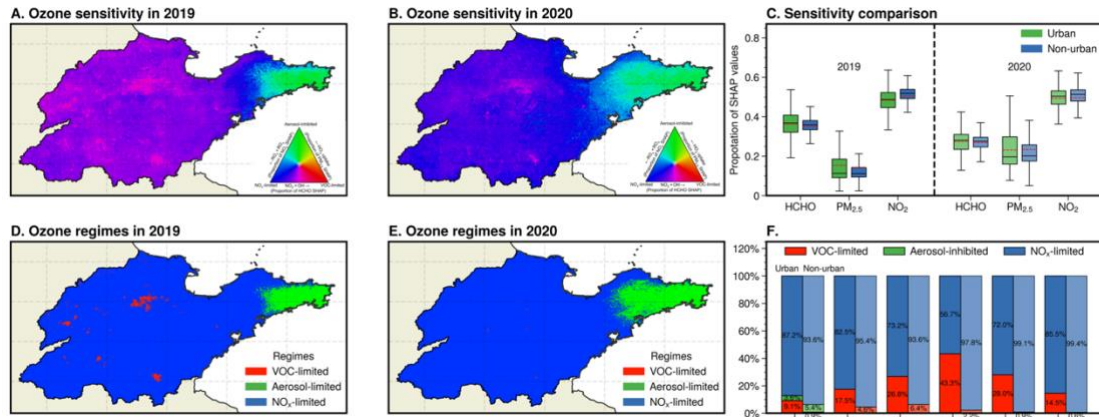
ozone to  $\text{NO}_x$  transitioning from urban to non-urban areas and erroneously concluding urban areas to be predominantly VOC-limited. In response to your comment, we have conducted a thorough reanalysis of the pertinent data and have updated our findings accordingly in Figure 8. This revised analysis provides a more precise and quantified insight into the distribution of ozone formation regimes across different urban and non-urban settings. Specifically, our updated results indicate that in certain cities, the prevalence of the VOC-limited regime within urban areas varies between 15% to 43%. This contrasts with non-urban areas, where the VOC-limited regime is significantly less common.

The main revision in the abstract is as follows in lines 32-35: “The ozone sensitivity in nonurban areas, dominated by nitrogen oxide ( $\text{NO}_x$ )-limited regime, was observed to shift towards increased sensitivity to volatile organic compounds (VOCs) when extended to urban areas.”

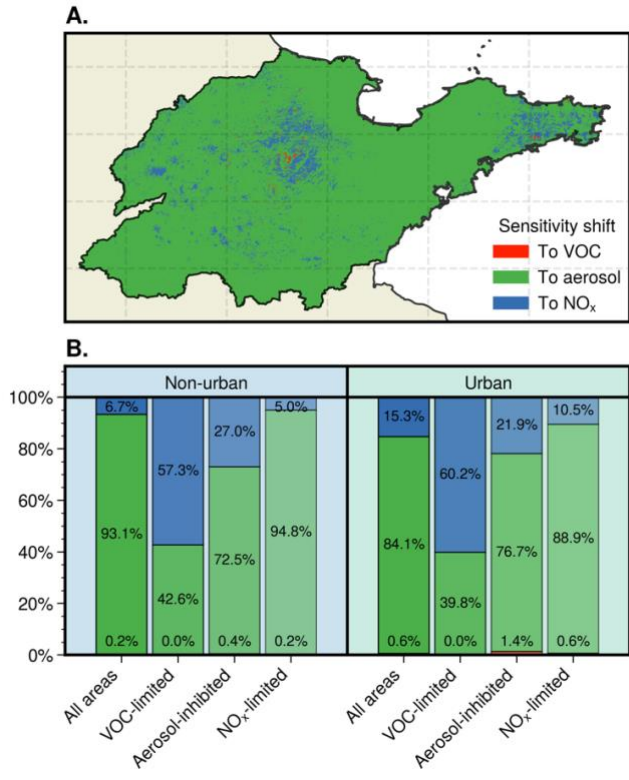
The main revision in section 3.3.1 is as follows in lines 536-639: “Moving along an urban-to-rural gradient, reactions dominated by  $\text{RO}_x$  radical self-reactions are continuously enhanced with increasing  $\text{NO}_x$  SHAP values, resulting in the majority of rural Shandong being situated in  $\text{NO}_x$ -limited regimes.”, and lines 555-572: “In several cities, including Binzhou, Zibo, Liaocheng, Linyi, and Jining, a greater proportion of urban areas, as compared to their nonurban counterparts, exhibited a VOC-limited regime in 2019, as indicated by the prevalence of red regions in Figure 8D. The percentage of urban areas in these cities under a VOC-limited regime ranges from 15% to 43%, in stark contrast to non-urban areas where such a regime is typically rare (Figure 8F). The comparison of  $\text{O}_3$  sensitivities from 2019 to 2020 shows a regional shift towards increased sensitivity to aerosol and  $\text{NO}_x$ , along with a decreased VOC sensitivity as a result of  $\text{NO}_x$  reduction (Figure 8A-C). This shift has led to the majority of areas in Shandong being dominated by a  $\text{NO}_x$ -limited regime in 2020, with an expanded aerosol-inhibited regime region in the Jiaodong Peninsula (Figure 8E). Additionally, the discrepancy in  $\text{O}_3$  formation sensitivity between urban and non-urban areas has been diminishing during this period (Figure 8C). As illustrated in Figure 9, while the ozone regime transitions



towards NO<sub>x</sub>-limited, there is a marked shift towards greater aerosol sensitivity across nearly 90% of areas, leading to a 1.6% increase in aerosol-inhibited grids. Compared to nonurban regions, a higher number of grids in urban areas demonstrate a shift towards NO<sub>x</sub> sensitivity. Conversely, urban areas that were predominantly aerosol-inhibited in 2019 showed a lower sensitivity shift towards NO<sub>x</sub>.”



**Figure 8.** Comparison of geographical distribution for ozone formation regimes between 2019 and 2020 in the summertime. All surface daily O<sub>3</sub>, PM<sub>2.5</sub>, and NO<sub>2</sub> estimations from Air Transformer (AiT) are averaged over each month from May to October 2019-2020 for matching monthly HCHO derived from TROPOMI (500 \* 500 m). (A, B) Geographical distribution of fractional contribution of chemical factors representing O<sub>3</sub> formation regimes. The ternary phase diagram in the legend depicts the normalized fraction of SHAP values for O<sub>3</sub> attributed to HCHO, NO<sub>2</sub>, and PM<sub>2.5</sub> at the surface, representing VOC-limited (red), aerosol-inhibited (green), and NO<sub>x</sub>-limited (blue) regimes, respectively. (C) Statistical Changes in the fractional contribution of chemical factors. (D, E) Geographical distribution of O<sub>3</sub> chemical regimes. (F) Proportion of three O<sub>3</sub> chemical regimes across urban and nonurban areas in 2019 in Shandong (SD), and individual cities (BZ: Binzhou, ZB: Zibo, LC: Liaocheng, LY: Linyi, JNI: Jining).



**Figure 9.** Geographical distribution of changes in ozone sensitivity from 2019 to 2020 in summertime (A). Comparison of ozone sensitivity changes across areas dominated by different chemical regimes in 2019 between urban and non-urban areas (B).

8. Minor typos and grammar errors need to be corrected. For example, Line 355: the upper right area of E, M, and U; Line 370: are shown in Figure 5, etc.

**Response:** Thank you for pointing out these typos and grammar errors. We have carefully modified them in the manuscript. We also conducted a thorough review of the entire text to ensure the accuracy and clarity.