Response to reviewer comments on “Global hotspots and mechanisms of extreme humid heat and air pollution co-occurrence” in review at Atmospheric Chemistry and Physics

S. Bartusek, Y. Wu, M. Ting, A. Fiore, and D. Westervelt

Anonymous Referee #1

Note: author responses are in blue, and all line numbers refer to those in the **non-tracked-changes** version of the manuscript.

General comments:

This manuscript investigates an important and increasingly relevant issue, the co-occurrence of extreme humid heat and air pollution. The analysis relies heavily on spatial and temporal processing steps that lack sufficient validation, and key interpretations are not supported by robust statistical evidence or a nuanced understanding of the underlying mechanisms.

We thank the reviewer for their careful attention and evaluation. We have introduced multiple changes throughout our analysis to address the reviewer’s concerns. Please see our responses to each comment below.

Specific comments:

1. The handling of different spatial resolutions between CAMSRA (0.75°) and ERA5 (0.25°) is problematic. Merely stating that "ERA5 grid cell at the center" of each CAMSRA cell is used is insufficient. This spatial mismatch introduces potential biases that need to be quantified and addressed more thoroughly. The method should be explained with consideration of how this choice affects results and whether it may bias the analysis toward certain regions.

We thank the reviewer for this comment. We agree that using ERA5 grid cells to define dry and humid heat days for the comparison of chemistry conditions averaged across the much larger CAMSRA cells introduced unnecessary inconsistency. We have now conducted the entire analysis based on the CAMSRA grid. All ERA5 variables have been (conservatively) regridded to CAMSRA gridcells, and all calculation of climatologies, anomalies, trends, and/or thresholds occurred after the regridding. Note that we still use ERA5 meteorology at CAMSRA resolution rather than CAMSRA meteorological variables, as some of the variables rely on daily maximum calculations (2m temperature and 2m wet-bulb temperature), which likely better represent peak temperatures that CAMSRA chemistry are affected by (its model timesteps are at sub-hourly resolution, but its output is only available at 3-hourly resolution). Additionally, some variables we analyze are not available in CAMSRA at all (see response to Reviewer 2’s comment on CAMSRA vs. ERA5 meteorology).

All of the figures and numerical results have been updated with the new data resolution, and no notable changes have been found.

2. The use of 95th percentile for defining extreme heat days is reasonable but requires more justification. Alternative thresholds should be tested or discussed to demonstrate robustness of findings.

We thank the reviewer for this comment, and agree that sensitivity to this threshold choice is important to consider. We have now repeated the ozone and PM2.5 difference calculations across three thresholds (the 93rd, 95th as used in the main analysis, and 97th percentiles of dry- and wet-bulb temperature) and three types of data (actual values, anomalies, and detrended anomalies as used in the main analysis), together creating 8

sensitivity tests that contextualize the main analysis case. We incorporate these sensitivity tests into Figure 2, introducing a robust significance test on top of the gridcell-wise bootstrap procedure. Specifically, gridcells that are significant in all 9 cases are denoted (with bootstrapping done consistently across all 9 cases; all corrected by the false discovery rate), and hotspot regions can only be defined from among those gridcells; a further mask is applied taking into account gridcell population. See Figure 2's caption and the Methods section for full descriptions.

3. The manuscript reports significant results based on data processing steps (e.g., compositing, anomaly calculations) that have not been adequately validated. The interpretation of statements like "increased humidity corresponds with worsened pollution" requires robust support. The authors should demonstrate the reliability of their methodology through independent validation or sensitivity testing.

We thank the reviewer for this comment, and have more carefully ensured and reported on the robustness of our results. In particular: (1) We have expanded sensitivity testing for the calculation of anomalies and extreme heat thresholds, as detailed above. (2) Robustness of the quoted statement has increased for hotspot definitions, because for a region to be defined a hotspot we require that the statement must be true across all three percentile thresholds, and not only for anomalies but also detrended anomalies and non-anomaly values of pollutants (and each of those within all three of the threshold cases). (3) Considering the global land area in total (i.e. the histograms in Figure 2), the robustness of the finding that the share of population experiencing worsened pollution during humid heat is larger than that experiencing reduced pollution is now also enhanced, as only significant differences are considered. (4) The composite anomaly differences are now more robustly validated against US Embassy in-situ measurements: we now directly test whether the dry vs. humid heat differences at Embassy sites are inconsistent with those from CAMSRA data, rather than comparing the mean differences, finding that none of them are (except for Islamabad, which only shares four humid heat days across its short record with CAMSRA after the data cleaning and processing is applied).

4. The use of a bootstrapping procedure for regional composites is positive, but its lack of application to the global-scale analysis is a significant oversight. Statistical significance should be consistently reported for all major findings to ensure the reliability and reproducibility of the results.

We thank the reviewer for this comment, and have taken care to include statistical significance testing throughout the analysis. We have now applied bootstrapping to the global-scale analysis, which is false-discovery-rate corrected (with α=0.1, as recommended in Wilks 2016). To visually communicate this (Figure 2), gridcells are denoted with a white dot where the difference is not significant, un-dotted if significant just in the main analysis case (i.e. using 95th percentile thresholds and non-detrended anomalies), and denoted with a gray or black dot where *robustly* (and positively) significant, i.e. significantly positive across all 8 sensitivity tests of thresholds and value types in addition to the main case. See our responses to Comment 2 and 3 above and see Figure 2 caption and Methods section for full descriptions. Significance testing is now additionally applied to the multiple linear regression analyses, taking into account spatial autocorrelation; see answer to Comment 6 below, and the correlations in Figure 6.

5. The claim that "humid heat is associated with stagnation" is presented as a factual statement rather than a correlation. The authors must acknowledge the complex and often bidirectional relationship between humid heat and stagnation, drawing on existing literature

on humid heat mechanisms. Stagnation creates humid heat, but humid heat also creates stagnation through various feedback processes.

We thank the reviewer for clarifying about this. If this quote refers to the abstract, "associated" here is intended to communicate a correlation without specifying any causal direction. We have edited this to read "correlation", though we note that "association" should not imply anything more than a correlation, and we use the word in various contexts throughout the manuscript as such. We absolutely agree the bidirectional relationship between humid heat and stagnation deserves dedicated mention, and we have added a sentence that explicitly highlights it (L39).

6. The suggestion of a causal link between background chemistry regimes and co-occurrence patterns lacks sufficient experimental design to support it. While the correlation between NO2/HCHO ratios and co-occurrence is interesting, it requires cautious interpretation. Correlation does not equal causation, and alternative explanations should be considered.

We thank the reviewer for this comment. We agree and do not intend to overstate any correlational findings as inappropriately causative. In the case of the relationship between background NO2/HCHO ratio and O3 and/or PM2.5 co-occurrence, we have taken care in this version of the manuscript to be precise in characterizing our findings. We agree that the association between background NO2/HCHO and co-occurrence is correlational, and we have revised the text to discuss it as such. Specifically, we have introduced additional tests using multiple linear regression modeling to better understand the details of this relationship, presented in Figure S22–S24 and discussed at the end of the Results section.

These tests are meant to determine whether the NO2/HCHO association is consistent with specific mechanisms, and to rule out alternative explanations, rather than to specifically establish causality. First, the model setup in Figures 8 and S22 explicitly controls for multiple meteorological conditions (stagnation, radiation, temperature) that simultaneously co-vary with both chemistry regime and humid heat—this ensures that associations found are not simply proxies for one another. This model setup also tests mechanism-specific predictions by including interaction terms (tracking how the predictors' coefficients change as the background regime changes), instead of only analyzing overall correlation. This set of analyses also includes robustness checks: Figure S23 checks the robustness of interaction coefficients from Figure S22 to the addition of HCHO and NO2 anomaly differences (the humidity~ratio interaction for PM2.5 survives this, but the BLH~ratio and WS~ratio interactions for O3 do not), and Figure S24 checks the robustness of associations in Figure 8 to the addition of background OH as a possible intermediate (OH absorbs the NO2 coefficient for O3, but does not affect any moisture interactions for PM2.5). Finally, to ensure caution is taken, all spatial regression analyses use standard errors clustered in spatial blocks, so that spatial autocorrelation is controlled for in the significance testing.

Together, these new analyses narrow down plausible explanations for the observed correlation and inform as to in what cases the evidence is stronger (stagnation and humidity effects for PM2.5) versus more suggestive (HOx chemistry related effects for O3). Ultimately, they remain observational tests and cannot establish causality. We now state this explicitly in the Discussion (L487), and state that determining causality would require more targeted model experiments (L495). When reporting on the association between background ratio and co-occurrence earlier in the manuscript we make clear we are not inferring causality ("raises the question of whether there may be a causal reason", L232, "Whether this association is causal —whether a NOx-saturated regime somehow predisposes a region to stronger humid heat and pollution co-occurrence—remains unclear and worthy of future investigation", L254)

7. Data and Method section requires substantial revision to address the issues outlined above, particularly regarding spatial resolution, threshold selection, and validation of

processing steps. The results section should be revised to reflect the limitations of the methodology and avoid overstating the strength of the findings.

We thank the reviewer for their attention to these issues. We have revised the Methods and Results sections in accordance with the manuscript edits and reviewer suggestions.

Anonymous Referee #2

Note: author responses are in blue, and all line numbers refer to those in the **non-tracked-changes** version of the manuscript.

General comments:

This study deals with intensive statistical investigation of how O3 and PM2.5 are associated with extreme dry/humid heat conditions at a global scale and for some selected regions/cities as foci. Research on joint air pollution and extreme climate is contemporary and active. The study also further looks to basic underlying meteorological and chemistry conditions and then point out certain overall pictures related to the issue reasonably. Although the approaches and tools applied seem basic or standard, the work is shown to take full advantage of them with centering on anomaly quantities. The findings and discussion are fine, some of which are potentially useful policy-wise regionally. Given the importance of the topic and the overall technical quality in and scientific merits gained from this study, its publication would be in my favor. Nonetheless, some concerns still exist and should be addressed adequately (in terms of justification, clarification, and improvement) properly, my comments and suggestions are listed below (with the major ones marked with asterisks ***).

We thank the reviewer for their careful attention and evaluation. We have introduced multiple changes throughout our analysis to address the reviewer's concerns. Please see our responses to each comment below.

Specific comments:

1. Word choices: It is sometimes confusing when reading the following terms:
   1.1. "extreme non-humid/dry heat day" vs "non-humid/dry day"
   1.2. "extreme humid heat day" and "humid heat day".
   1.3. "extreme humid/dry heat" vs "humid/dry extreme heat"
   1.4. In the text/figures, would "with or without "extreme" for humid/dry heat be meant the same?
   1.5. "hot-dry/humid days" appear in a number of subplots in Supplementary Material
   Given many terms defined in the study, try to state them clearly and later avoid confusion by keeping the consistency throughout the manuscript.

We thank the reviewer for this comment; we have increased the consistency of the terms in the revised manuscript. We now state explicitly what "dry" and "humid" heat days/conditions refer to in the Methods section, and that that all usages of these terms in the manuscript, with our without "extreme", refer to those definitions (L150). In the introduction, terms such as "non-humid" are sometimes used according to the context of the sentence, before our specific definitions have been laid out.

2. Writing style: Suggest
   2.1. "daily-maximum temperature" >> "daily maximum temperature"

2.2. "daily-mean 2-meter dewpoint temperature" >> "daily mean 2-m dewpoint temperature" or "daily mean dewpoint temperature"
2.3. "daily-maximum O3" >> "daily maximum O3" and so on.

We thank the reviewer for these suggestions. We have removed hyphens for daily-maximum and daily-mean throughout.

3. Check the consistency of abbreviations (of variables) throughout. For example, DEWPOINT vs T_d, BL Height vs BLH, etc.

We thank the reviewer for this point. We have ensured that abbreviations of variables are consistent throughout.

4. Clarify a bit more the term "global land area" whether it means the full global extent or most of it? As seen in many figures, the global maps are limited or bounded to certain latitudes.

We thank the reviewer for this comment; we have clarified that the scope of our analysis is land area between 60° S and N, and any usage of the term "global land area" refers to these extents (L119).

5. ABSTRACT and CONCLUSIONS: The term "makeup" is mentioned but a bit vague. Suggest to have it clarified or reworded.

We thank the reviewer for this suggestion and have replaced the word with "regime" where applicable.

6. Give the full names of chemical species at the first encounters in text (e.g., HCHO)

We thank the reviewer for this suggestion and have ensured all chemical species acronyms are spelled out on first usage.

7. The terms "background" meteorology and "background" chemistry" said throughout the manuscript are unclear or even misleading in the current context. I am not sure if the authors' intention is\ "underlying" or "driving". I would suggest define or elaborate on them more specifically or using an alternative term for clarity.

We thank the reviewer for this comment. The intent is to refer to underlying conditions, (but noting that these underlying mean-state conditions may also play an active role, as discussed throughout the manuscript). We have clarified that the term "background" refers to warm-season means of quantities, and added the term "underlying", in L240–242.

8. LINE 45: Suggest "particulate matter less than 2.5 μm in diameter (PM2.5)" >> "particulate matter less than or equal to 2.5 μm in diameter (PM2.5), to be consistent with WHO's.

We thank the reviewer for this suggestion and have adjusted this wording.

9. LINE 70: Recheck "Particulate matter may be increased by humidity through hygroscopic growth of aerosol particles". I am not sure whether it should be said as hygroscopic growth or heterogeneous (i.e., aqueous phase) reactions or both. Note that hygroscopic growth may affect aerosol size but not as directly on dry-basis PM mass.

We thank the reviewer for pointing this out. CAMSRA's PM2.5 is indeed dry-basis mass, so this is important to bring up in our paper. For this introductory sentence that is just process

based (not reporting on CAMSRA itself), both are true, and we have added a clause emphasizing aqueous-phase reactions to this sentence (L78).

10. LINE 88: Suggest "at 2 meter elevation" >> "at 2 m (above ground level)"

    We thank the reviewer for pointing this out and have adjusted this wording.

11. LINE 89: Suggest to separate this phrase as a new sentence from its previous text: "with the 1981–2010 period also analyzed for Tmax and TWmax to calculate extreme heat thresholds (see below)"

    We thank the reviewer for this suggestion and have adjusted this wording.

12. LINE 90: The terms "(see below)" is written but no calculation info is shown below. Please check.

    We thank the reviewer for this comment and have clarified this wording.

13. For the global maps presented throughout, perhaps the US state boundaries may be changed to country's for consistency, if possible.

    We thank the reviewer for this suggestion, and have implemented it.

14. As compliment to FIG 1, a map of cell-wise day count for dry heat days and that for humid heat day could be useful and given in Supplementary Information. Optionally, percentages (with respect to total days of the entire study period) may be used for cell-wise day counts.

    We thank the reviewer for this suggestion. We have now included Supplementary Fig. 4, reporting the day count for dry and humid heat days, as well as the fractional overlap (i.e. what proportion of extreme-Tmax days are also extreme-TWmax days; recall that the days extreme in *both Tmax and TWmax* are defined as humid heat days and the days extreme in *only Tmax* are defined as dry heat days), along with the average TWmax during humid heat days.

15. *** SEC 2: It is unclear about using meteorological variables from ERA5 (not directly from CAMSRA) given that the authors states that CAMSRSA also generates meteorological fields. State the reason(s). Are those from ERA5 and CAMSRA are comparable in values at least for the variables considered at a daily scale? If they are greatly different, would CAMSRA-simulated O3 and PM2.5 (driven by the model-internal meteorology) be well representative in view of their association with meteorology?

    We thank the reviewer for this comment. In response to both this and Reviewer 1's Comment 1, we have now conducted the analysis with ERA5 meteorological data regridded to CAMSRA gridcells (see our response to Reviewer 1 Comment 1). This removes a potential spatial bias between ERA5 heat conditions at a much smaller gridcell than CAMSRA's in relating to CAMSRA chemistry conditions across its larger gridcells. It is also likely that daily-maximum temperature and wet-bulb temperature calculated from ERA5's hourly temporal resolution are more accurate to the meteorology experienced by CAMSRA's chemistry model, which is run at sub-hourly timesteps (but is only available at 3-hourly resolution). Furthermore, some meteorological variables we analyze (boundary layer height, precipitation, solar radiation) are not available in CAMSRA; we have therefore chosen ERA5 as the more appropriate dataset and stayed consistent with it for all meteorology. We have also noted in the manuscript that the close similarity of the IFS used in both ERA5 and CAMSRA reduces potential biases between their meteorology (L105).

16. The authors review the literature and state solar radiation as a relevant meteorological variable for air pollution. Why is it not included in the analysis.? ERA5 and/or CAMSRA could have global radiation available.

We thank the reviewer for this suggestion. We have now included solar radiation across Figures 6–8 and all regional composites supplementary figures.

17. Precipitation is known for a key factor of wet scavenging and surface-wetness deposition. Should precipitation be treated in the analysis (or perhaps applied as flag somehow during identifying heat days)?

We thank the reviewer for this point. We have now included precipitation in the regional composites (Supplementary Figures 7–20), and briefly mention that it shows varying signals across hotspot regions (L302).

18. SEC 2.1: Would approximate thicknesses of the lowest model layer in CAMSRA be given and mentioned?

We thank the reviewer for this comment and have mentioned the approximate thickness (L117).

19. I admit that I have never been familiar with CAMSRA. As said in the manuscript, it incorporates the MACCity anthropogenic inventory. Is the MACCity inventory used in CAMSRA time-varying and corresponding to the entire study period, i.e., providing annual/ monthly emissions reasonably? Given this study's focus on the hotspot regions where cities are within, I think that this point is important and should be discussed; otherwise pointed out or recommended in the last section of the manuscript.

We thank the reviewer for this comment, and agree that it is important to include more details about the MACCity inventory. We have now included these lines in the Methods: “MACCity emissions are monthly and time-varying, but are based on RCP8.5 scenario extrapolation over our study period, so may not capture trends diverging strongly from this extrapolation; however, MACCity was found to agree well with other inventories (particularly for NOx; Granier et al., 2011), and CAMSRA's data assimilation corrects for such (Inness et al., 2019). Furthermore, we base our analysis on detrended anomalies to minimize remaining long-term trend biases.” (L110–L114)

20. SEC 2.3:
    20.1. LINE 124: Suggest "their 95th percentile value" >> "their corresponding 95th percentiles”
    20.2. LINE 136: Suggest as a new sentence" "with the linear trend in the anomalies’ daily time series over 2003–2021 removed" >> "The linear trend in the daily anomalies' time series over 2003–2021 is then removed to finally produce the daily anomalies for use.”

We thank the reviewer for these comments. We have clarified the wording of the first sentence (L146) and the second sentence (L164).

21. SEC 2.3 and FIG S7: The presented approach of identifying dry/humid heat days is quite good and logical, as explained in more detail in FIG S7 (and its caption). But I feel difficulty understanding the contents well given three cases shown and compared many variable I would suggest to organize the contents and, if possible, describe each case more clearly.

We thank the reviewer for this comment, and have clarified the details of these cases (L147).

22. *** SEC 2.3: Initially, I followed the idea of the linear-trend removal on daily anomalies. After some later thoughts, I think that doing so needs more objectivity. To me, daily anomalies without trends removed can be representative of actual driving conditions (emissions/ chemistry/weather). Trend removal is still reasonable for correlation because trends could affect or obscure underlying correlations. I think the authors may recheck the objectivity of each key data processing step, state it well, and re-compute when needed if any change.

We thank the reviewer for this comment, and agree it is important to consider both. We now incorporate both non-detrended anomaly and actual value differences into our objective hotspot definition (see Figure 2 caption).

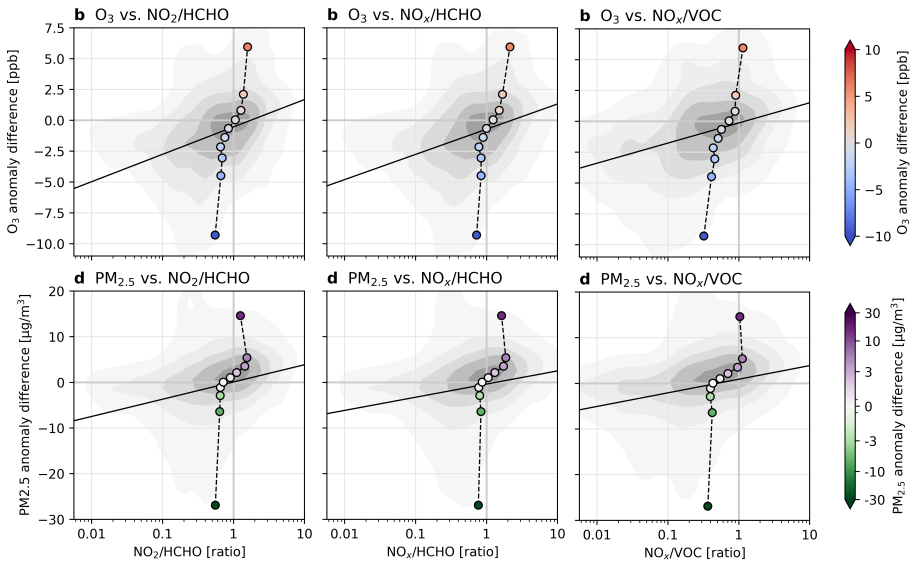
23. SEC 3.1 (from LINE 175) and FIG 6: Would it help if listing the hotspot regions considered, as well as their associated countries and city locations (central and bounding lon/lat values), as a table in the main text or Supplementary Information for clarity. Also see "All cities (means)" in FIG 6, I assume the term "mean" is of only across all the cities in question (i.e., 8 for O3 and 7 for PM2.5). Perhaps, state this briefly in the figure caption.

We thank the reviewer for this suggestion, and have included Table 1 providing this information (as well as details from the hotspot selection process). We have additionally clarified the mentioned information in the caption.

24. *** SEC 3.2: Give citation(s) to the sentence "HCHO and NO2 are often used as proxies for total VOC reactivity and all NOx species (NO + NO2), respectively"

We thank the reviewer for this suggestion, and have added citations to this sentence (L242).

25. I think CAMSRA also reportt NO (nitric oxide) as output, which can be summed with NO2 for NOX. Using NOx (as opposed to NO2 alone) may provide its chemical pool better than NO2 alone. Similarly, CAMSRA may produce other VOCs, some of which can also be used and combined with HCHO. It is not clear to me about the current proxy choices. I hope this could be more clarified.



We thank the reviewer for this comment. We have tested the effects of including NO and isoprene in the NOx/VOC ratio, as pictured in the attached figure. We find that adding NO to NO2 (first vs. second column below) and adding isoprene to HCHO (second vs. third column below) to make small impacts. Furthermore, we also now include isoprene in the multiple linear regression analysis section (it is not a significant predictor, so not shown in Figure 8, but does appear in Supplementary Figure 24).

26. SEC 2.2: The 2nd paragraph is quite important and may be more clarified, as the authors focus on daily O3 and PM2.5 anomalies:
    26.1. Climatological seasonal cycles. Also, this study set it to a 1-month or 30-day scale, correct?
    26.2. A total of 6 US-Embassy sites considered (as hotspot)? How many such sites with monitored data are available and the reason of having just the 6 sites compared?
    26.3. State briefly about the adequacy or necessary treatment of the US-embassy data used (percent missing, outliers, ...)
    26.4. Lack of the definitions of "humid heat" and "non-humid heat" therein. Thus, referenced to the section where these terms are described

We thank the reviewer for these comments, and have overhauled the relevant Methods section discussion (L132–144) and explained in more detail the procedure for treating Embassy data in Supplementary Figure 1 caption.

27. LINE 234: Suggest "higher during humid heat than dry heat days" >> "during humid heat days than dry heat days" for readability (also check other locations in the manuscript)

We thank the reviewer for this suggestion; we have clarified the wording of these comparisons throughout the manuscript.

28. It would be helpful for FIGs 4 and 5 fully showing all subplots (as seen in Supplementary Information).

We thank the reviewer for this comment. In this version of the manuscript, we have added multiple chemical and meteorological variables to all the regional composites, as seen in Supplementary Figures 7–20. For the figures in the main manuscript, these additional analyses would make the figures too large, so we have kept them as shown previously.

29. I am not sure, when mentioning figures (stored in Supplementary Information) from the main text, if one should say or attach the phrase "see Supplementary Information" next to where they are mentioned?

We thank the reviewer for this comment. At the first mention of a supplementary figure (in Methods), we included a note to see Supplementary Material.

30. FIG 2:
    30.1. The dashed boxes (hotspot regions) are still not easy to see. Use a more distinct color.
    30.2. Would each dashed box be possibly named or numbered (as legends) in the maps? Some readers like me may not be much familiar with global regions/cities.
    30.3. Is it correct that the maps in a) and c) are of average over 2003-2021 (19 years) and that the histograms in b) and d) are based on daily values? Anyhow, I would suggest having them stated clearly in the figure caption.
    30.4. Suggest: The caption should be rearranged as humid first and add the word "level", i.e., "Difference in pollution level between --humid-- heat and --dry-- heat …".

30.5. For the histogram plots in b), given the x-axis using a linear scale (as opposed to log scale). add more minor linear ticks on the axis for clarity.

We thank the reviewer for these comments. We have updated the appearance of hotspots regions, added numbers and labels to them, clarified the histogram data (they plot the same gridcell-based means shown in the map, where significant), and added more ticks to the linear histogram.

31. SEC 3.2, I feel that it comes a bit out of nowhere for the discussion about urban and remote areas. Describe how these urban and remote areas are assigned cell-wise, directly using LAND COVER data or visually? If visually, ensure reliability.

We thank the reviewer for this comment. We do not classify urban and remote areas in this analysis, we just analyze the NO2/HCHO ratio and discuss how that ratio is typically indicative of urban or remote types of background chemical regime. At the beginning of this section specifically, we have added a parenthetical that hopefully clarifies what we mean (L238).

32. FIG 6:
    32.1. Would the titles of b) and d) be "Across global O3/PM2.5 difference deciles" or "Across global O3/PM2.5 anomaly difference deciles"?
    32.2. Given many quantities shown in b) and d) (i.e., solid, dashed, and dotted lines), the subplots become too small to read. Suggest to rearrange them to be much larger or have them separate them as a new figure.

We thank the reviewer for this comment. We have adjusted this panel titles, and we have separated these figures into Figure 6 and 7.

33. For all figures, check if the font sizes used within the (sub)plots are too small, if graphics are blurred, and if the yellow lines when drawn are too hard to see. Suggest to improve the problems when found.

We thank the reviewer for this comment. We have increased the size of text in some figures and improved graphical aspects to improve readability.

34. *** SEC 3.5: Multiple linear regression include a large number of input variables (also including interaction terms). The regression framework presented is good but I am not sure whether collinearity and overfitting would arise and become a concern on the resulting models and interpretation (on regression coefficients and their statistical significance). In that case, stepwise regression and regression diagnostics may possibly be helpful to mitigate or remedy such concern.

We thank the reviewer for this comment, and have expanded and made more robust our multiple linear regression analysis section. In Figure S1, we provide the full characterization of the regression model from Figure 8, showing that differences between R-squared and adjusted R-squared (which accounts for the number of variables) is very minimal (differences of 0.001 for both O3 and PM2.5), which indicates that there is no concern of overfitting. We test the impact of collinearity be including and excluding certain variables (see Figures S22–24), and discuss how these differences inform our results. Additionally, statistical significance of coefficients is now reported more cautiously, by using spatial pooling of standard errors. See response to Reviewer 1 Comment 6 for more details on these points.

35. FIG 7c: Recheck the sentence "Tildes represent an interaction term between the left side and right side predictors; for example, the regression coefficient for the correlation of wind

speed with wet-bulb temperature overall is near-zero (WS ≈ 0), but becomes negative with increasing background NO2 (WS~NO2 < 0) or decreasing HCHO (WS~HCHO > 0).", it is not clear why saying "correlation of wind speed with wet-bulb temperature" for the term alone (ie., not interaction). Of note. the meanings of "joint variables (as interaction)" and "correlation" are typically different.

We thank the reviewer for this comment. For meteorological predictors, each term alone is a correlation between TWmax and a meteorological variable, while the tildes then represent the interaction between that correlation variable and either NO2 or HCHO. The description of these details is updated in the Figure 8 caption and the main text (L363).