Response to reviewer 2 comments

Reviewer comments are in red, and author responses are in blue.

Firstly, the authors have undertaken a detailed response to the review comments (from both reviewers). However, while they provide useful information justifying their methods and results, they infrequently make updates/clarifications to the manuscript to make this clear. If the reviewers find something unclear in the manuscript, it is very useful for the authors to provide good, detailed responses, but the manuscript needs updating to address this as future readers may have the same/similar questions.

We thank the reviewer for their additional comments. We agree that some further updates around clarity would be beneficial and have therefore updated the manuscript to address the comments below and those of the Editor.

Secondly, I am concerned by the new Table 1 in response to my comment #3. The authors clearly show that the downscaling approach improves the model (e.g. in terms of the absolute values), however, this is not the case for the surface ozone trends. In my comment #3, I wondered how important the sampling differences were between the model and observations (i.e. for the regional statistics, I believe the model/downscaling uses all regional pixels while only a few observation sites are used). To try and address this, the authors provided Table 1 which shows the trends for the model/downscaled data sub-sampled to the observations (i.e. closest pixel/grid box). In the manuscript, the authors discuss trends from the three data sources (Figures 7 and 8) but now I am not convinced they are comparable given the model/downscaled data is essentially representing a different quantity to that of the observations. From their Table 1 in the response document, there is a mixed response in whether the co-located model or downscaled trends more closely match the observations. I agree the observations have less spatial coverage than that of the model/downscaled data, so are not truly representative of county regional trends. However, if the observations are used in the ML approach to generate an improved higher resolution surface ozone dataset, then I would expect the downscaled data, when co-located to the observations, to be more representative of what the observations are showing (if only for several sites used to determine an observational trend). As a result, can the authors be confident in their downscaling approach to produce a higher spatial resolution product of surface ozone to investigate temporal evolution? I feel this needs to be addressed in detail in the main manuscript.

We agree with the reviewer that additional text on this was needed in the manuscript. We have therefore added the following text to section 4.2 and have included the table mentioned in the appendix.
"The analysis presented above provides valuable insights into the trends derived from downscaled and EMEP4UK data across a given domain. The downscaled and EMEP4UK trends encompass all pixels within a designated area. To delve deeper into the sensitivity of the trend analysis concerning sample size, we undertook a sub-sampling process for both the downscaled and EMEP4UK data specifically at measurement locations. The resulting annual mean trends are

given in Table A9, demonstrating the impact of sample size on trend outcomes. However, a note of caution is warranted against drawing excessive conclusions from small, largely non-significant trends observed across datasets. Both the downscaled and EMEP4UK products are susceptible to sampling errors due to the process of condensing a coarse grid model to specific point locations. As a result, over-interpreting these trends might lead to misleading assumptions. Therefore, the trends derived from the gridded products are anticipated to be the most regionally representative when considering the entire domain."

Response to editor comments

Editor comments are in red and author responses are in blue.

Thanks for submitting a revision of your manuscript in response to the two reviews. As both reviewers had major concerns about the study, I had asked them for a second review. Please respond to the comments by Reviewer#2 on the revised version of the manuscript.

While the reviewer remains sceptic about your study, I think that it could stimulate further discussion and could be a starting point for further studies. Therefore, I favour publication in ACP. However, I agree with reviwer#2 that when revising the manuscript you omitted a lot of the information from the response to the reviewer comments. I understand that you are concern about the length of the manuscript and I agree, that it should not become much longer, but at the same time I feel that some additional information might be useful for the readers.

We thank the Editor for their encouraging remarks and have addressed all comments by adding additional text in the manuscript, as outlined below.

Therefore, I ask you to prepare a new minor revision of the manuscript taking into account the new review and the following aspects.

Remarks on Response to Reviewer #1 (no 2nd review received)

- Add the explanation on the interpretation of SHAP value to the manuscript. I think is necessary as SHAP values are not a widely known metric.

The following text has been added to Section 3.3 to further explain the interpretation of the SHAP values:
"Instead, SHAP values display the difference between the average value of the response and the conditional average of the response given a specific value of the feature. Positive SHAP values can co-occur with either high (red) or low (blue) values of a feature, and similarly for negative SHAP values."

- Add a statement about not including emissions to text (3rd comment by Reviewer #1).

The following text has been added to Section 2.3:
"While this approach indirectly encompasses the influence of NO_X, a comprehensive treatment of NO_X within the ML model is beyond the current scope of our study. Additionally, the presence of sharp gradients in NO_X emissions introduces a potential risk of introducing spurious features during the downscaling process."

- The response argues with a citation of Ren et al. (2020); this reference along with the argument of Liu et al and Ren et al achieving similar correlation values should be added to the text.

The following text has been added to Section 3.2.2: "Our results are not inconsistent with other machine learning downscaling approaches for ozone. Liu et al., (2020) applied a similar method to produce a spatiotemporal surface of ozone concentrations in China from 2005 to 2017 and achieved a daily site cross-validation R^2 score of 0.64 and RMSE of 27.27 $\mu g/m^3$. Ren et al., (2020) investigated various machine learning models to predict ozone across the US and the highest spatial validation R^2 score was 0.68."

Remarks on Response to Reviewer #2 (numbers referring to major comments from reviewer's first review):

1. ok

2. The text in section 3.1. in version 5 of the manuscript remains less detailed than the response to the reviewer as subsection 3.1 remained largely unmodified – please change and add details to the manuscript.

The additional text below detailing the 10-fold cross validation test has been added to Section 3.2.1:

"To do this, we divided the measurement data into ten subsets by their location. The model was trained on nine subsets, while the remaining subset served as the evaluation set. We repeated this process for all subsets, ensuring that each subset was used for evaluation exactly once."

3. Please add the Table from your response to the appendix and summarise the results briefly in the main text also addressing the reviewer's concerns about the representativeness of model.

This table has been added to the appendix (Table A9), and additional text has been added to section 4.2 discussing the representativeness of the model.

4. ok