Authors' response to Reviewer 2

We sincerely thank the Reviewer for taking the time to thoroughly review our manuscript and provide relevant feedback to improve it further. Below we address their concerns and suggestions point by point.

Reviewer 2

General

The authors developed the tool CLAQC to quickly assess the impact of policy scenarios on Air Quality using 2 methods: elastic net modelling (EN) and an extreme gradient boosting regressor (ML). CLAQC can be used to attribute sectoral and country specific emissions changes to changes in $PM_{2.5}$ and O_3 concentrations without great computation burden. It is a useful too for policy makers and other stakeholders. The authors evaluate the performance of both models on a country level and find that the model performance differs depending on country and model used, while generally both models are better at predicting O_3 than $PM_{2.5}$. The paper is excellently written, the language is easy to understand and the paper well structured. The figures were unfortunately of low resolution and should be improved for publication. In several sections a more detailed explanation or discussion on top of the description of results would be useful.

Comments:

(Abbreviations used: PXX-LYY $->$ page xx, line YY; EQZ $->$ Equation Z)

Reviewer 2 Point 1 — *P2-L28f Did you mean "secondary O_3 formation and secondary PM formation"?*

Reply: Yes, we meant "secondary O_3 formation and secondary PM formation". Thank you very much for noticing this. We have updated the text as suggested.

Reviewer 2 Point 2 — *P2-L40ff The 2 sentences starting with "As new data" and "As new and better data" seem repetitive. Please consolidate these sentences.*

Reply: Thank you for this suggestion. We have changed the text as follows: "As new and better data come in every year, the emulator can be updated, and a higher detail level may be possible at lower trade-off costs".

Reviewer 2 Point 3 — *P5-L7 Why did you choose TerraClimate over ERA5 for the majority of variables used? Is TerraClimate's rain product more accurate than the one from ERA5? Since*

you're aggregating all data to 0.5° x 0.5°, you don't seem to make use of TerraClimate's higher horizontal resolution.

Reply: It is true that the ERA5 dataset is also a very complete product. Nevertheless, the TerraClimate product is a peer-reviewed dataset published in Scientific Data. It has a spatial resolution of about 4 km x 4 km and a monthly temporal resolution and was validated against weather measurement station data from several different networks, e.g., the Global Historical Climatology Network (GHCN) database, the Snow Telemetry network (SNOTEL), and the RAWS USA Climate Archive (Abatzoglou et al., 2018). Indeed, we could have used other reanalysis products such as the ERA5 ones for the weather variables included in the models. However, TerraClimate is both spatially and temporally suitable for our purpose and allows us to keep flexibility for further improvements in spatial and temporal resolution.

Reviewer 2 Point 4 — *P5-L23ff You are describing how a reanalysis again here when you had already described it in greater detail in the introduction to section 2 (P3-L4ff). This seems repetitive.*

Reply: We thank the Reviewer for this suggestion. We have removed the first sentence, "EAC4 reanalysis combines model data with observations, *in-situ* and satellite, from all over the world into a globally complete and consistent data set using a model of the atmosphere based on the laws of physics and chemistry", and moved the remaining part of the paragraph to the introductory part of Section 2.

Reviewer 2 Point 5 — *P7-L21 Please elaborate on the general purpose of monotonic constraints for the benefit of the reader not too familiar with that kind of modelling.*

Reply: We thank the Reviewer for this suggestion: we have briefly introduced the general purpose of monotonic constraints, explaining their purpose and providing an example. We have changed the text as follows: "We impose monotonic constraints on certain model coefficients to align with expected physicochemical relationships. These constraints specify how input variables should affect the target, ensuring interpretable and physically plausible results. For instance, a positive monotonic constraint enforces a non-negative relationship, ensuring that as an input variable increases, the predictor output does not decrease.

In the presence of noise, complex interactions in the data, or predictor cross-correlation, models may otherwise learn patterns that are not realistic or physically plausible. Additionally, monotonic constraints help prevent overfitting, enhancing robustness when input data are limited or uncertain. For example, it is not expected that an increase in BC emissions would lead to a decrease in $PM_{2.5}$ concentrations.

While at the local scale, reducing certain precursors of secondary inorganic aerosols might not always lead to a decrease in $PM_{2.5}$ levels — due to nonlinear atmospheric reactions noted by Thunis et al. (2019); Ding et al. (2021) — our national-scale models focus on broader trends. To avoid giving undue importance to cases where local emissions reductions might result in increased levels of inorganic $PM_{2.5}$, we apply monotonic constraints. In scenarios where secondary reactions substantially affect the overall

mass of $PM_{2.5}$, our models are designed to exclude such precursors from the predictor list, thereby not reflecting a decrease in $PM_{2.5}$ levels."

Reviewer 2 Point 6 — *EQ5 Please define β and β_0.*

Reply: We have included the definition of β and β_0 as follows: "It solves the following minimization problem for the model parameters β_0 and β, where β_0 is the model's intercept and β represents the coefficients of the input variables:"

Reviewer 2 Point 7 — *EQ7 Should this be n_{test} instead of n^{test}? In P9-L24 test is a sub- not a superscript.*

Reply: Thank you for pointing out this inconsistency. The superscripts in EQ6&7 are used as a reference to the test set. Regarding the n constant, we have updated it in EQ7 to n_{test}, as in P9-L24.

Reviewer 2 Point 8 — *EQ8&9 Several parameters are not defined, e.g. λ, β, γ_i, δ, μ, ν, ϵ, θ. Are the α and β the same as in EQ5?*

Reply: We thank the Reviewer for pointing out this. The mentioned parameters (λ, β, γ_i, δ, μ, ν, ϵ, θ) are the predictors' coefficients. The β parameters in EQ5, 8, and 9 have theoretically the same meaning. Instead, the α parameter in EQ5 is a regularization parameter that sets up the elastic net setting while in EQ8 and 9 it represents the linear regression intercept. In EQ5, α corresponds to a value of 0.5. If the α parameter was 1, then a LASSO shrinkage regression would be performed. If the α parameter was 0, a Ridge regression. We have added these explanations in the text.

Reviewer 2 Point 9 — *EQ8&9 In both equations emissions are used multiple times: with emissions depending on sector & pollutant (β-term), just sector (δ-term) and just pollutant (λ-term). Please clarify what the purpose of the multiple emission terms is. In the P11-L7ff you describe how the terms in EQ8&9 mimic secondary production, transport and dispersion. It would be useful for the reader to also understand what processes or dependencies the multiple emission terms are a proxy for.*

Reply: We thank the Reviewer for this comment. To better clarify, we have updated the text as follows: "In both Equations, we include multiple emission terms to increase the chances that models capture variations in emissions. In Equation 8, to model the formation of secondary inorganic aerosol, we interact emissions of NO_x and NH_3, SO_2 and NH_3, and NO_x and SO_3, respectively. This approach helps to represent the formation of secondary inorganic aerosols, such as ammonium salts, which result from reactions between these precursors. Similarly, in equation 9, we interact emissions of NO_x and NMVOC, SO_2 and NMVOC, and SO_2 and NO_x. As before, this attempts to capture the reactions between the precursors of O_3, since the presence of at least two of these precursors is necessary for its

formation. In both Equations, we also interact sectoral emissions with wind speed and direction to proxy transport and dispersion of pollutants. We include total sectoral emissions to reflect that sector-specific policies typically impact multiple pollutants through dedicated emission offset protocols. Additionally, we consider total emissions from individual pollutants because variations in total pollutant emissions may result not only from specific sectors but also from inter-sector changes, transported emissions, and chemical reactions."

Reviewer 2 Point 10 — *EQ9 The emission terms for the O_3 equations are slightly different to the ones for $PM_{2.5}$. For both pollutants, there are terms depending on sector & pollutant at the same time (β-term) and then just on the sector (δ-term). In EQ9, term depending on just the pollutant (λ) is for a different pollutant (p_3) than the β-term. Why do the β- and λ-term for $PM_{2.5}$ (EQ8) depend on the same pollutants, but the β- and λ-terms for O_3 (EQ9) do not: namely the λ-term depends on one additional pollutant (SO_2).*

Reply: Correct, the emission terms for the O_3 equations are slightly different with respect to the ones for $PM_{2.5}$ and the λ-term depends on one additional pollutant (SO_2), due to their different atmospheric reactions. Regarding emissions, the O_3 model includes sectoral emissions related to NMVOC and NO_x, which are the main precursors of O_3 (Baird and Cann, 2013; John H. Seinfeld, 2016). It also includes sector totals, total emissions of NMVOC, NO_x, and SO_2, and their interactions (specifically, $NO_x \times$ NMVOC, $SO_2 \times$ NMVOC, and $SO_2 \times NO_x$). Plus, interactions between sectoral emissions and wind speed and direction. While NMVOC and NO_x are O_3 main precursors, reacting in the presence of solar radiation, SO_2 plays an indirect role in O_3 formation. SO_2 is typically emitted by industrial sources. It is involved in secondary PM formation, which can reduce the radiative properties and oxidative capacity of the atmosphere, indirectly affecting O_3 formation. We have tried to clarify this further in the text under Subsection 3.3. Elastic net models:

"Note that the emission terms in the Equations differ due to their different atmospheric reactions. In Equation **??**, to model the secondary inorganic aerosol formation, we interact total emissions of NO_x and NH_3, SO_2 and NH_3, and NO_x and SO_2, respectively. Similarly, in equation **??**, we interact total emissions of NO_x and NMVOC, SO_2 and NMVOC, and SO_2 and NO_x. While NMVOC and NO_x are O_3 main precursors, reacting in the presence of solar radiation, SO_2 plays an indirect role in O_3 formation (Baird and Cann, 2013; John H. Seinfeld, 2016). SO_2 is typically emitted by industrial sources. It is involved in secondary PM formation, which can reduce the radiative properties and oxidative capacity of the atmosphere, indirectly affecting O_3 formation. In both Equations, we also interact sectoral emissions with wind speed and direction to proxy transport and dispersion of pollutants. Refer to section **??** for the EN model specifications with DACCIWA emissions."

Reviewer 2 Point 11 — *P11-L29 You say that randomisation occurs over the temporal dimension. Does that mean that the concentration fields calculated by the ML model do not depend on the*

previous time step (month in this case)? Is there an initialisation of the pollutant concentrations or is the assumption essentially that the ML model can estimate the pollutant concentration of the current month based on the emissions and meteorological conditions of the current month only, without knowledge of previous atmospheric conditions and pollutant concentrations?

Reply: We apologize for not being clear on this point. Your interpretation is correct. Regarding ML estimation through the XGBoost architecture, it takes place without initialisation and without imposing any functional form, except for monotonic constraints. The model estimates the pollutant concentration for the current month based on emissions, weather, and other conditions of that same month (pixel identifier and seasonal identifiers), without knowledge of previous atmospheric conditions. The model captures interactions between features as well, without explicitly defining them.

Reviewer 2 Point 12 — *P13-L12f The road and residential sectors are named as having the greatest impact in DEU, ITA and BRA. Is that referring to Fig 6a? I cannot see that in the figure. Italy seems to only have impact from Agriculture. The resolution of the plot is quite low so it's hard to see details.*

Reply: Correct, this statement is referring to Figure 6a. Unfortunately, due to the size constraints of the uploaded preprint document, the plots' resolution is low, sometimes not allowing to accurately see plot details. Though, all our plots have a resolution of 300 dpi. In Figure 6, both ranges and line plots are displayed. Given that the model algorithms may include multiple emission variables within a sector, e.g., both NO_x and BC emissions from the Agriculture sector, to account for the sectoral range variability we calculate the minimum and maximum annual percentage variation in predictions from perturbed emissions by perturbation and sectoral level. When the model chooses only one sectoral predictor, the minimum and maximum annual percentage variation in predictions from perturbed emissions is the same, and a simple line is displayed in the plot, as in the case of the Road and Residential sectors for DEU, ITA, and BRA.

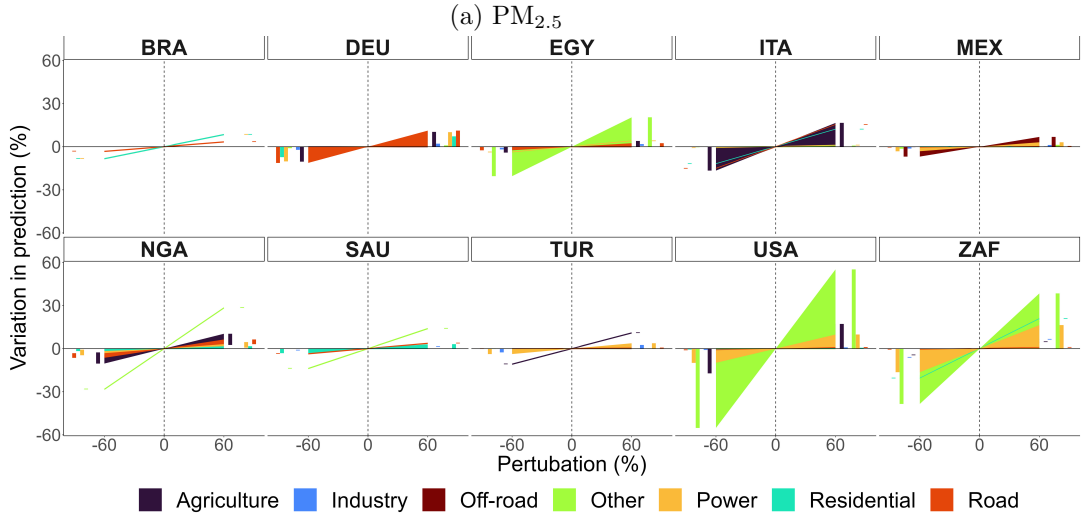


Figure 6: Percentage variation in predicted concentrations by sector and perturbation for selected countries in EN models for $PM_{2.5}$. Bar charts on the sides of each subplot help visualize overlapping variations.

Reviewer 2 Point 13 — *P13-L28ff Please discuss why the models perform so poorly in some countries. Is it inconsistencies in either emission or concentration data for that country? Are there important mechanisms occurring in this countries that are missed by the models? Are there pollutants missing from the emission data sets that are important in those countries? Does the model perform poorly because of some of the inputs or is it something in the model that you could change to improve the performance?*

Reply: It is true that results vary by prediction target ($PM_{2.5}$ vs O_3) and input type (e.g., CAMS vs DACCIWA emissions), which may reflect the inconsistencies in emission data mentioned. Generally, both modeling methods are better at predicting O_3, as O_3 concentrations are highly correlated with incoming radiation or temperature, while predicting $PM_{2.5}$ is more challenging due to its complex secondary chemistry, local sources and particle composition. Chemistry transport models predict better O_3 than PM as well, due to the more complex mixture of particles and local effects from more sources of the latter one (Guérette et al., 2020). Many of the potential issues you noted can contribute to country-level performance variations. Additionally, local factors such as unique orography and micro-meteorological conditions can significantly impact predictions in some areas, even country-level averages. This is suggested by the stronger performance of ML models incorporating pixel identifiers in Figure 10, highlighting the importance of local features in achieving better performance. Although we did not dive into the reasons for some countries performing better than others, Figures 7 to 9 can help us identify some sources of poor performance. For instance, results indicate that Europe and some African countries, plus some countries in Central and South America perform better with ML models, implying that their

variability is better explained by nonlinear relationships, since their R^2 is higher in ML models (Figure 9). However, we also observe an increase in error in some of these models. While EN models seem to outperform ML ones in North America for both pollutants.

In response to your suggestion, we have extended the discussion in subsection 4.2. Model internal validation results, as below:

"Figures 7 and 8 map the out-of-sample R^2 and RMSE for EN and ML models obtained from both CAMS and DACCIWA emissions. We do not advise using the models for countries with R^2 smaller than 0.5 or RMSE higher than 12.

Results vary by prediction target ($PM_{2.5}$ vs O_3) and input type (e.g., CAMS vs DACCIWA emissions), which may reflect inconsistencies in the emission or concentration data. Additionally, local factors such as unique orography and micro-meteorological conditions can significantly impact predictions in some areas, even country-level averages. Generally, both modeling methods are better at predicting O_3 than $PM_{2.5}$, as O_3 concentrations are highly correlated with incoming radiation or temperature, while predicting $PM_{2.5}$ is more challenging due to its complex secondary chemistry, local sources and particle composition. Chemistry transport models predict better O_3 than PM as well, due to the more complex mixture of particles and local effects from more sources of the latter one (Guérette et al., 2020)".

We have also added to the 5. "Limitations" Section the fact of not having analyzed the reasons behind countries' poor performance.

Reviewer 2 Point 14 — *P14-L1 In P13-L43f it sounds like DACCIWA is used everywhere where available, so in Africa CAMS is never used, correct? In fig 9c then, are the runs using DACCIWA actually being compared with runs using CAMS in Africa?*

Reply: We thank the Reviewer for this comment and apologize for not being clear on this aspect. We clarify a bit here and in the text. While CAMS-GLOB-ANT emissions are available for all countries worldwide, DACCIWA emissions are only available for Africa. Note that we construct and provide independent modeling versions with both CAMS-GLOB-ANT and DACCIWA emissions for Africa. In Figure 9, regarding the performance metrics RMSE and R^2, runs using DACCIWA are compared with runs using CAMS in Africa in subplots 9a and 9b, respectively. While, as the Reviewer correctly pointed out, in subplot 9c, we underline that we recommend DACCIWA over CAMS-GLOB-ANT as the preferred input source for emissions in Africa. Therefore, models built with DACCIWA emissions for African countries maximize the *Source* criterion, as shown in 9c. To clarify further, we discuss the case of South Africa. In figure 9, regarding $PM_{2.5}$, R^2 is maximized for South Africa with ML runs (pixel version) on DACCIWA emissions (9a), and regarding RMSE the model and input version that minimizes RMSE is elastic net with DACCIWA emissions (9b), while in terms of source DACCIWA is preferred (9c). Regarding O_3 instead, the EN run with CAMS-GLOB-ANT maximizes R^2 (9a), and the EN run with DACCIWA minimizes RMSE (9b), while DACCIWA is the preferred source (9c). We have improved clarity on this in the text.

Reviewer 2 Point 15 — *P14-L15 Is the CAMS reanalysis you mention here the EAC4 reanalysis*

you introduced in 2.3? If yes, it is confusing for the reader to refer to the same product with different names. If no, please introduce the CAMS reanalysis.

Reply: With "CAMS reanalysis" we meant both CAMS emission and concentration reanalysis products employed for CLAQC, so the data described under Subsections 2.1.2. CAMS emissions (CAMS GLOB-ANT reanalysis) and 2.3. Concentrations (EAC4 reanalysis), respectively. Although the EAC4 reanalysis product is operated by ECMWF, it is part of the services offered by the Copernicus Atmosphere Monitoring Service (CAMS). Therefore, it is also known as the CAMS global reanalysis EAC4 product. As this may create confusion, we have tried to make this clearer in the text, as follows: "However, given the high disparities in the available ground monitoring data across the globe, we believe that CAMS reanalysis products, such as CAMS-GLOB-ANT and EAC4, are the next state-of-the-art available solution for these regions".

Reviewer 2 Point 16 — *P14-L36 The sentence starting with "It is a complimentary model" is confusing:*

- *"A [...] model to the model [...] community" is repetitive. Maybe "A complimentary tool"?*
- *Which scenario community? The policy scenario community?*
- *Did you mean "providing empirically based estimates"?*

Reply: We thank the Reviewer for these suggestions. We have updated the text as follows: "The CLAQC framework lends itself to multiple developments. It is a complementary tool to the modeling and policy scenario community, providing empirically based estimates and added value for global scale sectoral and country-level analyses."

Reviewer 2 Point 17 — *P14-L39 Unless there is a second paper planned describing the CLAQC tool's functionality, it would be useful to have a short overview over the kind of scenarios that can be run. I.e. is the 60% perturbation fixed or can the user have some control over the scenario selection (apart from country, model, specification, etc used).*

Reply: We thank the Reviewer for this comment. Regarding the scenarios, we simulate perturbations of emissions from -60% to +60% at 20% steps based on the last 5 years of data (2017-2021). This means that emission perturbations are fixed to those levels. However, the user has control over the scenarios as they can select the following parameters: country, model, specification, sector, precursor pollutant, baseline concentration, baseline emissions, and perturbation level. We discussed scenarios under Subsection "Comparing two scenarios" in the Appendix Subsection A.4.2. Machine learning — Guide to the Excel spreadsheet.

Reviewer 2 Point 18 — *P14-L44 Link is broken. Is the code embargoed until the paper is published?*

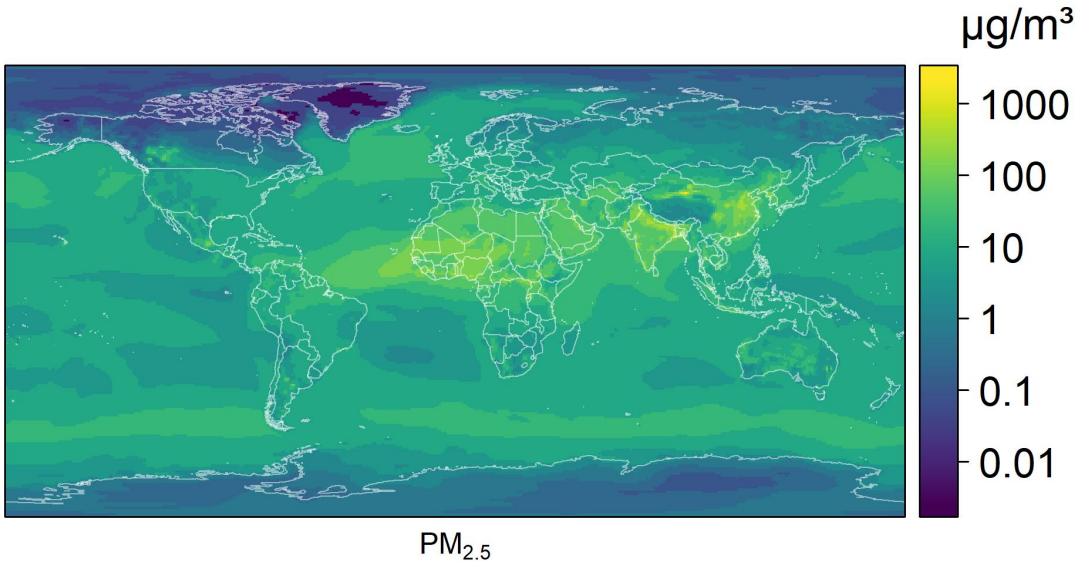
Reply: We had reserved a DOI (10.17632/wt25vt6ycr.1) through a Mendeley Data repository but unfortunately we are having problems accessing it again to make it public. This might be due to some changes in the Mendeley services that have occured recently. We have set up a new frozen repository with the original code through Zenodo at the following URL: `https://zenodo.org/records/14177055`.

Figures

Reviewer 2 Point 19 — *Fig 1 Please include more labels for the colour scale in b).*

Reply: We thank the Reviewer for this suggestion that we have implemented as in Figure 1. Note that concentration levels are displayed on a logarithmic scale as before.

(a) $PM_{2.5}$



(b) O_3

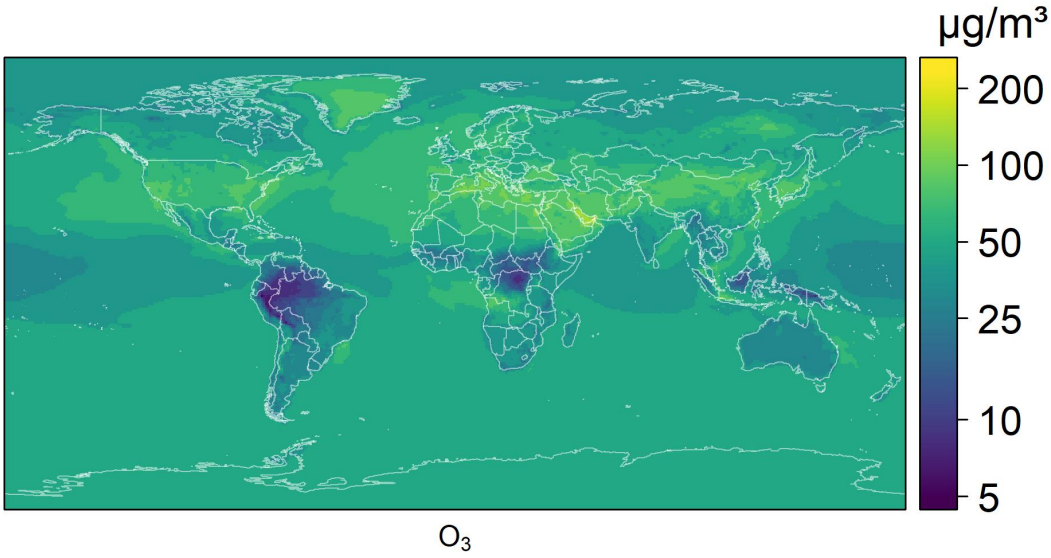


Figure 1: Level plots of EAC4 concentrations of $PM_{2.5}$ (January 2018) and O_3 (July 2018) in $\mu g/m^3$ with color bar in logarithmic scale.

Reviewer 2 Point 20 — *Fig 6a In the plots for BRA, NGA, SAU and TUR there are line plots instead of filled areas for some of the sectors. Is that a plotting error or does that signify something?*

Reply: Line plots in Figure 6 are not a plotting error. Given that the model algorithms may include multiple emission variables within a sector, e.g., both NO_x and BC emissions from the Road sector, to

account for the sectoral range variability we calculate the minimum and maximum annual percentage variation in predictions from perturbed emissions by perturbation and sectoral level. When the model chooses only one sectoral predictor, the minimum and maximum annual percentage variation in predictions from perturbed emissions is the same, and a simple line is displayed in the plot.

Reviewer 2 Point 21 — *Fig 7 It is difficult to see which countries are below 0.5 with a continuous colour scale. Maybe include a colour break at 0.5?*

Reply: We thank the Reviewer for this suggestion. We have included a colour break at 0.5 in the plots as proposed. See below Figure 7.

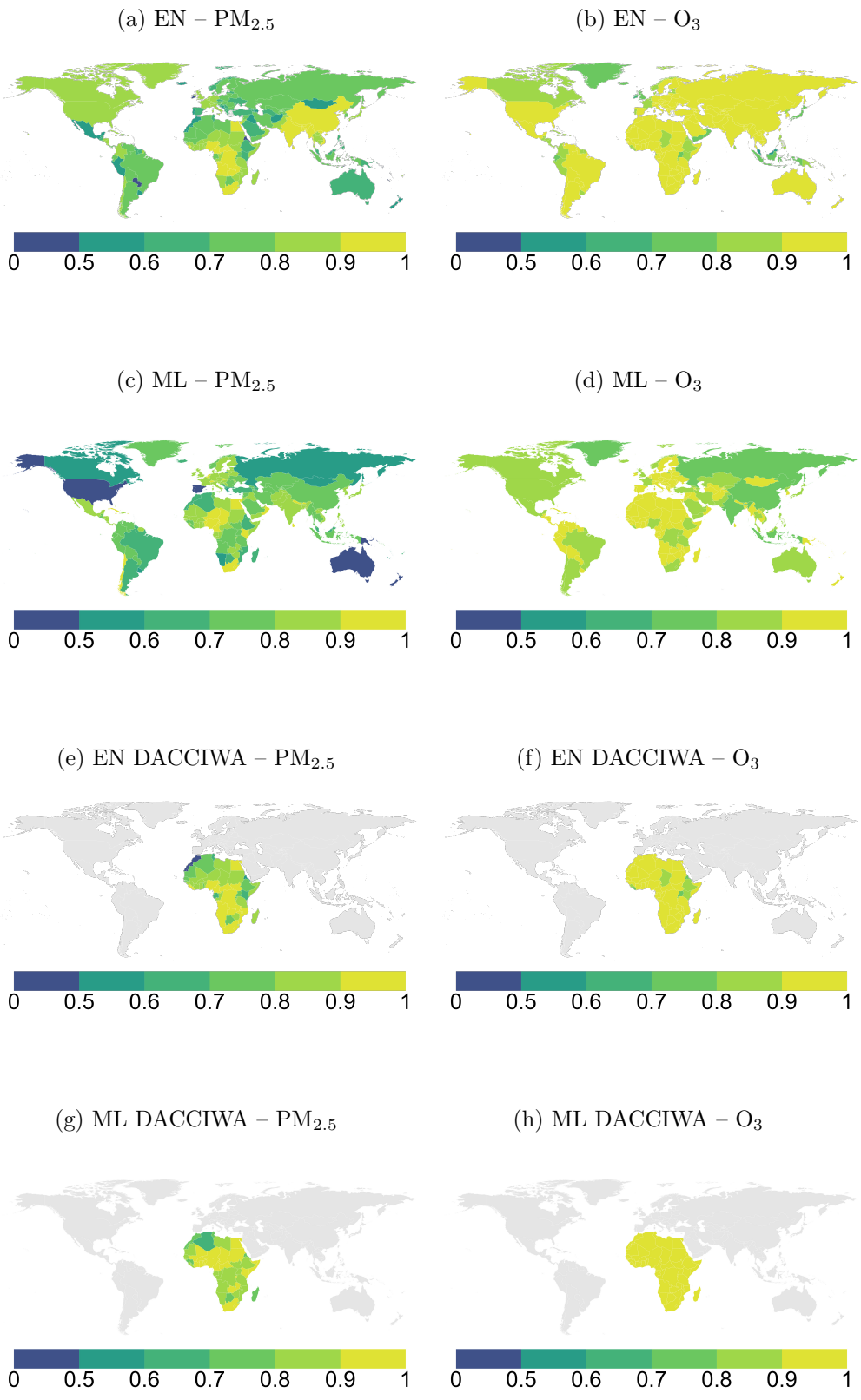


Figure 7: Out-of-sample performance metrics of ML (no pixel) and EN models (both from CAMS and DACCIWA data): R^2.

Reviewer 2 Point 22 — *Fig 8 Similar to Fig 7, it is not possible to see the cut-off of 12 with the colour scale used.*

Reply: We thank the Reviewer for this suggestion: we have updated the plots accordingly. See below Figure 8.

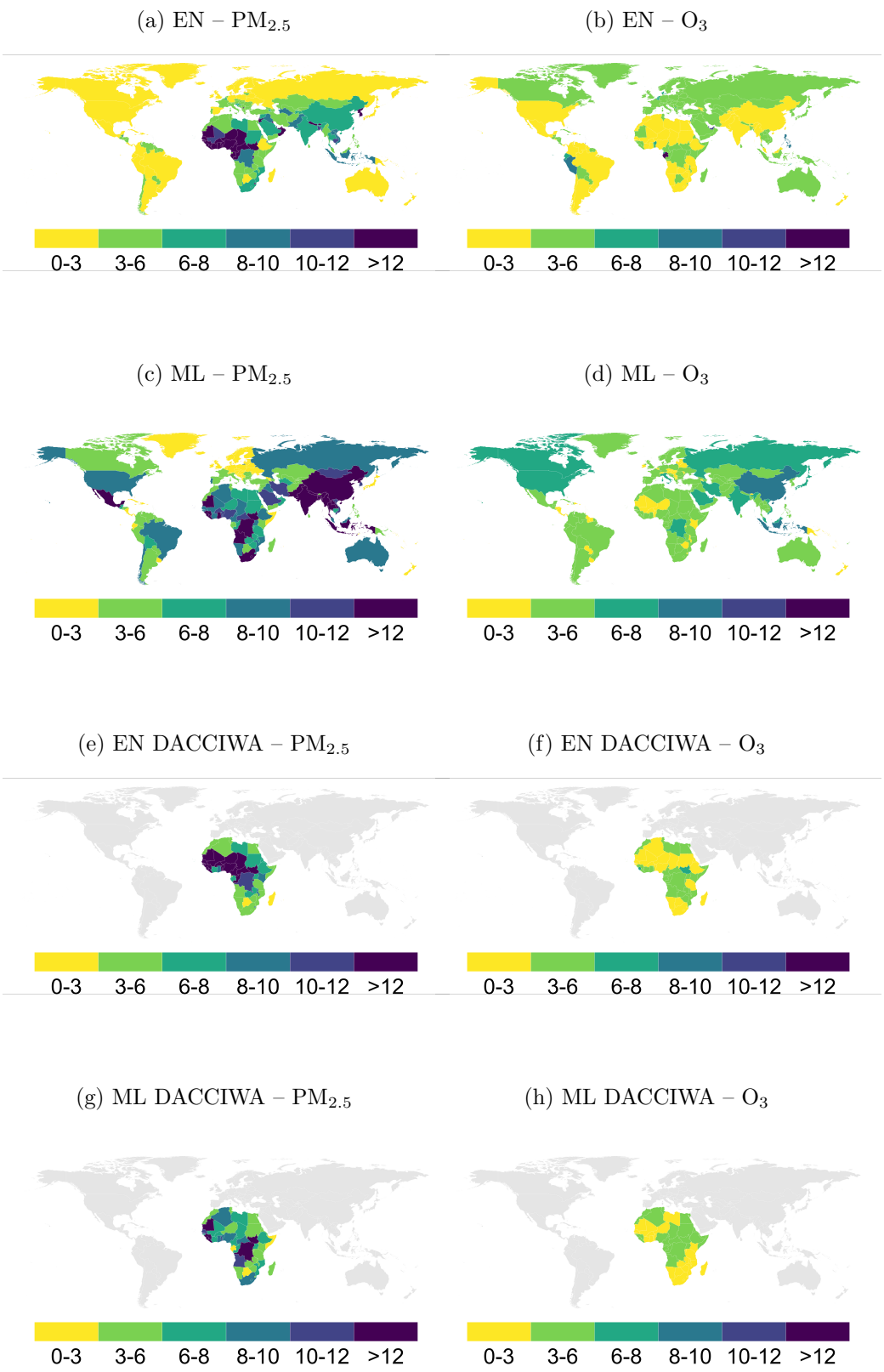


Figure 8: Out-of-sample performance metrics of ML and EN models (both from CAMS and DACCIWA data): RMSE.

Reviewer 2 Point 23 — *Fig 9 "with" should be abbreviated with "w" or "w/" not "w/t".*

Reply: Thank you for highlighting this imprecision. We have updated the plot legend as suggested. See Figure 9.

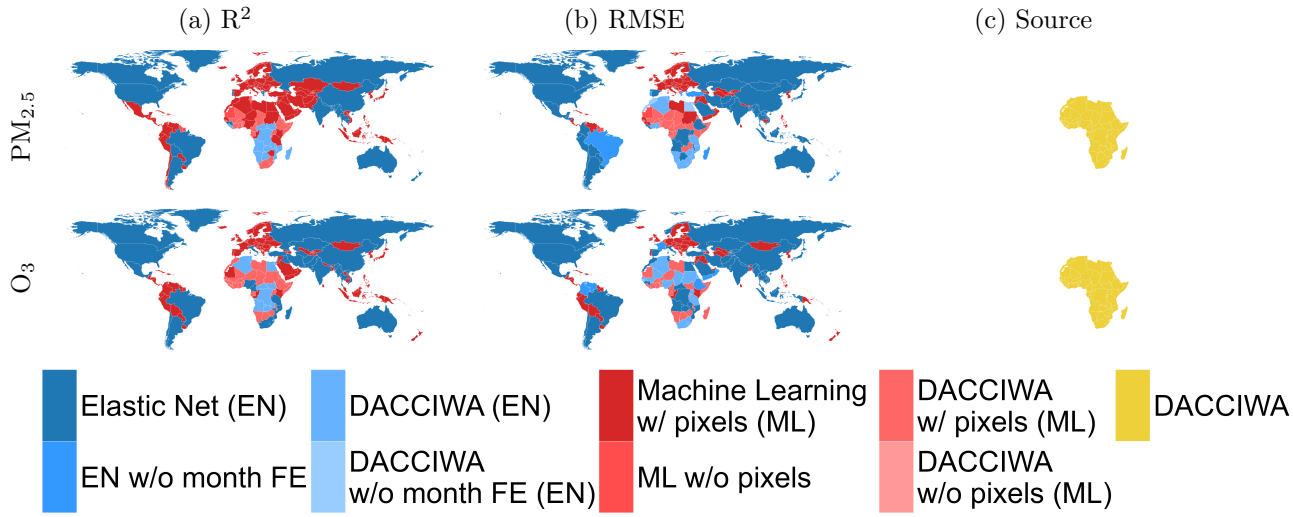


Figure 9: Best model score for each pollutant, country, and decision criterion.

Reviewer 2 Point 24 — *Fig 9 To make the best performing model "group" (EN vs ML) more obvious, you could use one hue per model group, i.e. all EN models in shades of blue and all ML models in shades of red.*

Reply: We thank the Reviewer for this suggestion that we have considered for improving the plot's readability. See Figure 9.

Reviewer 2 Point 25 — *There are some small inconsistencies in notation the authors may want to address, e.g.*

P7-L24ff PM2.5 not subscripted for some occurrences in this paragraph.

P2-L31 O_3 is cursive here but nowhere else.

L39 Remove gap between T and g for Tg.

P10 in description for TMINt and TMAXt, use °C, not degC just as elsewhere in the text

Reply: We thank the Reviewer for noticing these inconsistencies. We have corrected the $PM_{2.5}$ notation and the cursive O_3. Regarding the unit of measure on P3-L39, there is no gap between T and g for Tg: it may appear as a gap just due to the cursive font. We have harmonized the abbreviation for degrees Celsius to *°C* throughout the text.

References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Scientific Data 5, 170191. URL: https://www.nature.com/articles/sdata2017191, doi:10.1038/sdata.2017.191.

Baird, C., Cann, M., 2013. Chimica ambientale. Terza edizione italiana condotta sulla quinta edizione americana.

Ding, D., Xing, J., Wang, S., Dong, Z., Zhang, F., Liu, S., Hao, J., 2021. Optimization of a nox and voc cooperative control strategy based on clean air benefits. Environmental Science amp; Technology 56, 739–749. doi:10.1021/acs.est.1c04201.

Guérette, E.A., Chang, L.T.C., Cope, M.E., Duc, H.N., Emmerson, K.M., Monk, K., Rayner, P.J., Scorgie, Y., Silver, J.D., Simmons, J., Trieu, T., Utembe, S.R., Zhang, Y., Paton-Walsh, C., 2020. Evaluation of Regional Air Quality Models over Sydney, Australia: Part 2, Comparison of PM2.5 and Ozone. Atmosphere 11, 233. doi:10.3390/atmos11030233.

John H. Seinfeld, S.N.P., 2016. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. WILEY. URL: https://www.ebook.de/de/product/25599491/john_h_seinfeld_spyros_n_pandis_atmospheric_chemistry_and_physics_from_air_pollution_to_climate_change.html.

Thunis, P., Clappier, A., Tarrason, L., Cuvelier, C., Monteiro, A., Pisoni, E., Wesseling, J., Belis, C., Pirovano, G., Janssen, S., Guerreiro, C., Peduzzi, E., 2019. Source apportionment to support air quality planning: Strengths and weaknesses of existing approaches. Environment International 130, 104825. doi:10.1016/j.envint.2019.05.019.