Dear Reviewer,

We sincerely appreciate the valuable and constructive feedback provided by Reviewer #2. These insights have significantly contributed to improving our manuscript.

Below, we provide a detailed point-by-point response to all comments. The reviewer's comments are highlighted in *italic black*, our responses are marked in blue, and the corresponding revisions in the manuscript and Supporting Information are indicated in red.

Thank you for your time and consideration.

All line numbers given in our replies given below refer to the revised manuscript (color-tracked) and Supporting Information (color-tracked).

This paper summarizes the trends of O3 and its precursors over a 20 year period in Saxony and concludes that NOx reductions outpacing VOC reductions has led to increasing mean concentrations of ozone in urban regions. The paper fits the scope of ACP and is an important analysis with direct implications for designing effective future mitigation strategies in the region. Overall, the analysis is detailed despite the fact that there were a lot of assumptions that had to be made given data availability over this time period.

We thank the reviewer for his/her positive overall assessment of our manuscript.

General Comments:

R2-C1: *Overall, I found it difficult to read the paper given mis-matches between the text and figures/captions and I think particular attention needs to be paid to better align this information. In particular, the figure captions do not adequately describe was being shown in each figure and I was not able to determine what information was different in the subpanels without referencing the text. All figure captions should "stand alone" and be able to fully describe what is being shown and should be edited accordingly. Additionally, I found that there just seemed to be "missing" information where a figure was first referenced that was only mentioned much further down in the text. This made it difficult to understand what was in the figure the first time I encountered it, especially if that information was also not defined in the caption.*

(A) **A1:** We have carefully checked all Figure captions and revised the ones where we agree with the reviewer that more information needed to be given to make them stand alone. The substantially revised captions of Figures now read as follows: Caption changes in the main manuscript.

Figure 3: Boxplots of observed O_3 trends over three time periods until 2020 (given in columns 1-3) at the four station types. Transparent dots indicate statistically non-significant values. Black squares indicate means across all station values.

Figure 4: Mean yearly O_3 concentration differences, i.e. decrements, between different station types. Yellow point lines represent decrements from mean mountain to mean rural stations concentrations, blue point lines represent rural to urban, and red point lines represent urban to traffic stations decrements.

Figure 5: Theil-Sen trend values for different percentiles of O_3 concentrations in four station types (given in rows 1-4) and for three time periods until 2020 (given in columns A-C), respectively. Coloured dots show the values of individual stations, with solid dots for statistically significant trends and transparent dots otherwise. All dots are summarized by the boxplots. The black square reflects the mean trend value per percentile.

Figure 6: Theil-Sen trend values at different times of the year at the various station types (given in rows 1-4) and within each of the three time spans up to 2020 (given in columns A-C). Coloured dots show the values of individual stations, with solid dots for statistically significant trends and transparent dots otherwise. All dots are summarised by the boxplots. The black square shows the mean trend value per season.

Figure 7: Relationships between O_3 trends and NO_x trends across all stations for the 3 different time periods (given in rows 1-3). Coloured dots show the values of individual stations, with solid dots for statistically significant O_3 trends and transparent dots otherwise. Also shown are linear fits for O_3 trends and NO_x trends in different time periods.

Two newly updated Figs. 8 and 9, incorporating modifications in response to comments R2-C6 and R2-C7 have been added at lines 478 and 504 of the manuscript. See the revised captions of updated Figs. 8 and 9 below.

Figure 8: Diurnal profiles of hourly averaged modelled and observed NO and NO_2 in Saxony at rural background sites (A, C for summer case, and B, D for winter cases). Shaded areas indicate night-time, black lines indicate the modelled values, gray lines represent averaged observed concentrations and coloured lines refer to the observed concentrations at each station. It is noted that all observed NO concentrations below the DL of 1 µg m^{-3} are not known and cannot be shown.

Figure 9: Diurnal profiles of hourly averaged modelled and observed O_3 in Saxony at rural background sites (A and B for summer and winter cases, respectively). Shaded areas indicate night-time, black lines indicate the modelled O_3, gray lines represent averaged observed O_3 and other coloured lines refer to the observed O_3 at each station.

A newly update Fig. 10 with some modifications referring to R2-C8, has been added at line 524 of the manuscript. In addition, to avoid the "missing" information caused by referencing a figure before it is discussed much further down in the text, the relevant details have also been included in the caption of Fig. 10. See the revised caption of updated Fig. 10 below.

Figure 10: Isopleth plots for the net O_3 production rate ($NetPO_3$ in ppb h^{-1}) during 12:00 - 13:00 CET as a function of averaged NO_x and TNMVOC concentrations (A and B for summer and winter cases, respectively). $NetPO_3$ is shown using a rainbow color scale. The coloured points represent the conditions in different years with squares showing traffic stations and triangles showing mean urban stations conditions. The red dotted arrows represent hypothetical future scenarios with reductions in only one precursor (either TNMVOC or NO_x), while the yellow dotted arrows illustrate more plausible future pathways, where $NetPO_3$ slightly decreases through strong VOCs emission controls combined with moderate NO_x reductions. Note that based on the underlying modelling, these data in Fig. 10 are given as mixing ratios. For better comparability with mass-based concentrations elsewhere in the manuscript, conversion factors of NO_x (µg m^{-3}) ≈ 1.5 × NO_x (ppb) and O_3 (µg m^{-3}) ≈ 2 × O_3 (ppb) can be used.

The caption of Figure 11 has been revised based on the comment of R1-C25 and R1-C30. See the revised version below.

Figure 11: Anthropogenic emissions of NMVOCs in Saxony for five years from 2000-2019. Emission data were obtained by averaging values across the approximated area of the entire state of Saxony (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

(B) Caption changes in the Supporting Information (SI)

A newly updated Fig. S1, incorporating modifications in response to comments R1-C19 and R1-C25, have been added at line 18 of the Supporting Information. See the revised captions of updated Fig. S1 and Fig. S3-S5 below.

Figure S1: The NO_x emission (in kg m^{-2}) from road transport in the entire Saxony state region in 2019. A continuous red frame indicates the entire simulated Saxony area. Two specific rectangles in rose pink, A and B, represent the representative urban area (city of Leipzig) (0.13° × 0.20°) and traffic-dominated area (highway close to Dresden) (0.07° × 0.10°) used for the simulation, respectively.

Figure S3: Relationships between O_3 trends and NO, NO_2 trends across all stations for three different time periods (given in rows 1-3). Coloured dots show the values of individual stations, with solid dots for statistically significant O_3 trends and transparent dots otherwise. Also shown are linear fits for O_3 trends and NO, NO_2 trends in different time periods.

Figure S4: Linear correlations between measured averaged noon (12:00 - 13:00) NO_x and O_3 in ppb of four station types in summer over 5 years (2000, 2005, 2010, 2015 and 2019). For better comparability with mass-based concentrations elsewhere in the manuscript, conversion factors of NO_x ($\mu g\ m^{-3}$) ≈ 1.5 × NO_x (ppb) and O_3 ($\mu g\ m^{-3}$) ≈ 2 × O_3 (ppb) can be used.

Figure S5: Linear correlations between measured averaged noon (12:00 - 13:00) NO_x and O_3 in ppb of four station types in winter over 5 years (2000, 2005, 2010, 2015 and 2019). For better comparability with mass-based concentrations elsewhere in the manuscript, conversion factors of NO_x ($\mu g\ m^{-3}$) ≈ 1.5 × NO_x (ppb) and O_3 ($\mu g\ m^{-3}$) ≈ 2 × O_3 (ppb) can be used.

Two newly added figures as Fig. S6 and S7, in response to comments R2-C9, have been added at lines 98-107 of the Supporting Information. The numbering of subsequent figures in the Supporting Information (starting from Fig. S6) has been updated accordingly. The former Fig. S6 is revised to Fig. S8. See the revised captions of newly added Figs. S6 and S7 and Figs. S8-S10 below.

Figure S6: The distribution of resulting TNMVOC and NO_x concentrations (in ppb) based on a total of 800 and 1200 model runs (see Table S6 for details) for summer (A) and winter (B), respectively. Panels C (summer) and D (winter) are subsets of Panels A and B, respectively, and use the same concentration coordinate ranges as Fig. S7. The different colors indicate different batch runs.

Figure S7: The grid of modelled net O_3 production rate ($NetPO_3$ in ppb h^{-1}) during 12:00 - 13:00 CET as a function of both modelled and emission inventory-derived NO_x and TNMVOC concentrations (in ppb) (see Sect. 2.3) after bivariate linear interpolation. The X-axis and Y-axis in both summer (A) and winter (B) are interpolated using a very fine-resolved 1000 × 1000 grid. $NetPO_3$ is shown using a rainbow color scale.

Figure S8: Diurnal profiles of the modelled net O_3 production rate ($NetPO_3$ in ppb h^{-1}) and measured O_3 change rate (dO_3/dt, ppb h^{-1}) in Saxony at rural background sites (A and B for summer and winter cases, respectively). Shaded areas indicate night-time, black line indicates the modelled $NetPO_3$ and other coloured lines refer to the observed dO_3/dt at each station. Linear correlation coefficient (r) between modelled and measured rates is larger than 0.8 in both seasons.

Figure S9: Diurnal profiles of hourly averaged measured O_3 in summer over five years (2000, 2005, 2010, 2015, and 2019) across four station types.

Figure S10: Diurnal profiles of hourly averaged measured O_3 in winter over five years (2000, 2005, 2010, 2015, and 2019) across four station types.

Four newly updated Figs. S11 - S15, in response to comments R1-C30, have been added at lines 137-165 of the Supporting Information. See the revised captions of updated Figs. S11 - S15 below.

Figure S11: Anthropogenic emissions of NO_x in Saxony for five years from 2000-2019. Emission data were obtained by averaging values across the approximated area of the entire state of Saxony (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

Figure S12: Anthropogenic emissions of NMVOCs in traffic area for five years from 2000-2019. Emission data were obtained by averaging values from the selected traffic-dominated area (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

Figure S13: Anthropogenic emission of NO_x in traffic area for five years from 2000-2019. Emission data were obtained by averaging values from the selected traffic-dominated area (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

Figure S14: Anthropogenic emission of NMVOCs in urban area for five years from 2000-2019. Emission data were obtained by averaging values from the selected urban area (city of Leipzig) (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

Figure S15: Anthropogenic emission NO_x in urban area for five years from 2000-2019. Emission data were obtained by averaging values from the selected urban area (city of Leipzig) (see Fig. S1). Details of the emission categories and corresponding emission sectors can be found in Table S13.

The captions of Figs. S16 and S17 has been revised based on the comment of R1-C31. See the revised versions below.

Figure S16: Biogenic emissions of isoprene (ISO) and alpha-pinene (API) in Saxony for the year 2019, along with anthropogenic and biogenic emission data for limonene (LIM) in the same year. Emission data were obtained by averaging values across the approximated area of the entire state of Saxony (see Fig. S1). The emission data source is from the German Environment Agency (UBA) for Germany and Thürkow et al. (2024). Detailed data can be referred to Table S12.

Figure S17: Biogenic emissions of isoprene (ISO) and alpha-pinene (API) in urban area for the year 2019, along with anthropogenic and biogenic emission data for limonene (LIM) in the same year. Emission data were obtained by averaging values from the selected urban area (city of Leipzig) (see Fig. S1). The emission data source is from the German Environment Agency (UBA) for Germany and Thürkow et al. (2024). Detailed data can be referred to Table S12.

R2-C2: *In general, I found it odd that gas phase concentrations are being reported in ug/m3 rather than ppb. Moreover, the units used throughout the paper are broadly inconsistent with production rates being reported in ppb/h, but concentrations reported in ug/m3. I am more familiar with O3/ NOx/ VOC concentrations being reported in ppbv than ug/m3 given that U.S. standards are in those units. But, regardless of the authors preference for units, at minimum the units of the concentrations and production rates should be consistent throughout the paper. For example, in Figure 10, NOx concentrations are in ppb and O3 production is in ppb/h, but in Figure 8 & 9 NOx and O3 concentrations are ug/m3).*

A2: We agree the use of different units is confusing. In Germany and Europe, the standards of O_3 and NO_x are in µg m^{-3} and all observations are reported in this unit. We therefore chose to discuss the trends of observed concentrations in this unit. Figs. 8 & 9 compare the measured values with our modelled ones and to better link these Figures to the previous observation-based data, we preferred to keep the µg m^{-3} unit here.

For the isopleth plots as in Fig. 10, however, it would be very uncommon to present the production rates in µg h^{-1}. It would also be very impractical to convert TNMVOC to µg m^{-3}as all the modelling is done on the basis of number density (molec cm^{-3}). We therefore prefer to keep these Figures in ppb. To improve the comparability of Fig. 10 with the remaining ones of the manuscript, we now include the conversion factors in the Figure caption, which are also used in the reporting of measurements from the monitoring network.

Figure 10: Isopleth plots for the net O_3 production rate ($NetPO_3$ in ppb h^{-1}) during 12:00 - 13:00 CET as a function of averaged NO_x and TNMVOC concentrations (A and B for summer and winter cases, respectively). $NetPO_3$ is shown using a rainbow color scale. The coloured points represent the conditions in different years with squares showing traffic stations and triangles showing mean urban stations conditions. The red dotted arrows represent hypothetical future scenarios with reductions in only one precursor (either TNMVOC or NO_x), while the yellow dotted arrows illustrate more plausible future pathways, where $NetPO_3$ slightly decreases through strong VOCs emission controls combined with moderate NO_x reductions. Note that based on the underlying modelling, these data in Fig. 10 are given as mixing ratios. For better comparability with mass-based

concentrations elsewhere in the manuscript, conversion factors of NO_x (µg m^{-3}) ≈ 1.5 × NO_x (ppb) and O_3 (µg m^{-3}) ≈ 2 × O_3 (ppb) can be used.

***R2-C3: R/e other reviewer's comments**: Overall, I concur with most of their comments and technical notes, particularly in regards to their suggestions to improve the match between figure captions/the text and their detailed notes about things that are missing distinctions within the text. I only note I have one distinct difference in opinion from their comments. References to "p" in the figures is clearly meant to be a p-value indicating statistical significance and to me including it next to the r2 value makes this abundantly clear, as well as the fact that there would be no reason this should be "pressure". However, defining this the first time they reference the p-values in the text would be useful.*

A3: Please see our responses to R2-C1 for details on improving the match between figure captions/the text and their detailed notes in the text. To avoid the "missing" information caused by referencing a figure before it is discussed much further down in the text, the relevant details have also been included in the caption of Fig. 10.

p-values means probability. Now "(probability)" has been added in the line 127 and the sentence reads in the manuscript:

Theil-Sen function derives p-values (probability) and uncertainties by bootstrap simulations.

Specific Comments:

***R2-C4:** Lines 130-140: This section is missing some details that would be required to reproduce the simulations (e.g. Which version of CAPRAM is used in the work? What meteorology fields/versions are used? What other model options (e.g. specific deposition schemes, emissions inventories etc.) were assumed?) I'm not a CAPRAM user, so I don't know what details are typically provided, but it seems to me that versions/meteorological field names are certainly relevant/ required to reproduce the work.*

A4: In the present study, only the detailed near-explicit gas-phase chemistry mechanism MCMv3.3.1 (Master Chemical Mechanism) is applied (this information can be found in the lines 138-139 of the revised manuscript). No version of CAPRAM (Chemical Aqueous Phase Radical Mechanism) is used. Although the SPACCIM model framework allows for the coupling of MCM with CAPRAM to simulate multiphase chemistry, it also enables the use of each mechanism independently. Here, SPACCIM is employed solely with the gas-phase chemistry mechanism MCMv3.3.1.

To avoid misunderstanding, the revised paragraph describing the mechanisms used in lines 132-139 of the text is as follows: To understand the role of photochemistry for O_3 concentration evolution in Saxony, photochemical simulations were performed with the air parcel box model SPACCIM (SPectral Aerosol Cloud Chemistry Interaction Model). SPACCIM combines a size-resolved multiphase chemistry model with a microphysical model, enabling both to function independently while accounting for their interdependencies. Detailed descriptions of the SPACCIM model framework can be found in Wolke et al. (2005). In the present study, only the detailed near-explicit gas-phase chemistry mechanism, MCMv3.3.1, is used, which comprises 17224 reactions (http://mcm.york.ac.uk/MCM/) (Saunders et al., 2003).

SPACCIM operates as a box model (fixed at a certain location), where air properties such as temperature, pressure, and relative humidity etc., are defined using fixed parameters based on observed condition to simulate atmospheric chemical processes.

Some specific information on meteorology has been included in lines 149-152 of the article, and some sentences have been revised in lines 146-152. The sentences now read:

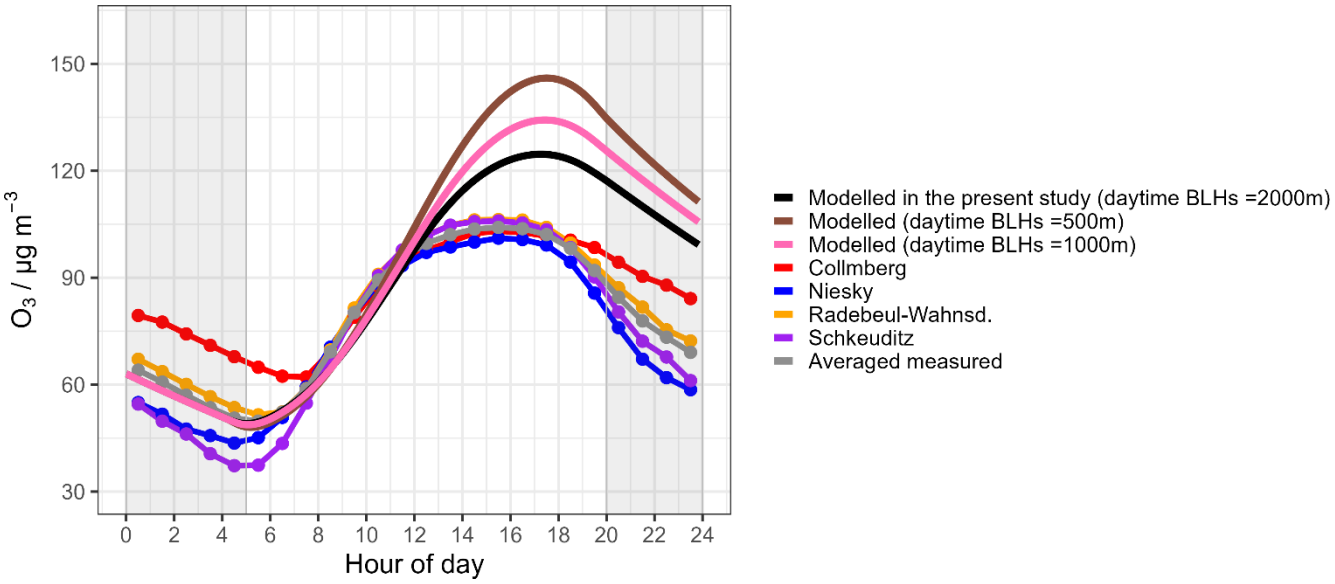
Besides emission values, other initial parameters had to be adjusted to their typical daytime and nighttime levels under rural conditions (see Table S4 for details). For meteorological parameters and trace gas concentrations (except SO_2, HONO and PAN), data was derived from measurements described in Sect. 2.1. The air temperature was set to 15°C in summer and 4°C in

winter. The pressure was kept constant at 1000 hPa for both seasons, while the relative humidity was maintained at 70%. Ratio of solar radiation, defined as the mean value between 10:00 and 13:00 divided by the maximum clear sky radiation value during the same period, was calculated to 0.7 in summer and 0.4 in winter.

The emission inventories have been described in the lines 143-146 of the manuscript. Also, the setting of dry deposition velocities has been illustrated in the lines of 158-164 of the text.

R2-C5: *Lines 156-157: The choices for the boundary layer heights ascribed need to be justified. These seem reasonable to me, but given the model-measurement discrepancies for NOx/O3 ascribed to this choice, I think its important they state why they were chosen (e.g. match available measurements decently well?). I have more typically seen BLH set in box modeling by toggling them until a secondary species with a long lifetime expected to be lost primarily to dilution is matched.*

A5: First, the selected boundary layer heights are based on previously available measurements in Germany during summer and winter (Wiegner et al., 2006; Brümmer et al., 2012; Kotthaus et al., 2023). Then, by scaling the daytime boundary layer heights, a more reasonable agreement is achieved between the modelled O_3 concentrations (with BLHs = 2000 m) and the measured values, as shown in the following figure.



The phrase of "based on previously available measurements in Germany (Wiegner et al., 2006; Brümmer et al., 2012; Kotthaus et al., 2023)" has been added at lines of 166-167 of the manuscript to support the selection of boundary layer heights. Now the sentence reads:

The BLHs in the simulations were set to 500 m at night and 2000 m during daytime in summer and to 250 m at night and 1000 m during daytime in winter, based on previously available measurements in Germany (Wiegner et al., 2006; Brümmer et al., 2012; Kotthaus et al., 2023).

Additionally, the references (Wiegner et al., 2006; Brümmer et al., 2012; Kotthaus et al., 2023) have been inserted in lines 911-913, 718-719, and 818-821 of the manuscript. The references now read:

Wiegner, M., Emeis, S., Freudenthaler, V., Heese, B., Junkermann, W., Münkel, C., Schäfer, K., Seefeldner, M., and Vogt, S.: Mixing layer height over Munich, Germany: Variability and comparisons of different methodologies, Journal of Geophysical Research: Atmospheres, 111, 2006.

Brümmer, B., Lange, I., and Konow, H.: Atmospheric boundary layer measurements at the 280 m high Hamburg weather mast 1995-2011. Mean annual and diurnal cycles, Meteorologische Zeitschrift (Berlin), 21, 2012.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmos. Meas. Tech., 16, 433-479, 10.5194/amt-16-433-2023, 2023.

R2-C6: *Section 3.3.1 - Were all modeled [NO] concentrations adjusted up by 1 ug/m3 or only nighttime values?? Seems like it's a bit misleading if all modeled [NO] is arbitrarily adjusted up when only the values < DL are adjusted in that manner for the observations.*

A6: We appreciate the reviewers' question as it made us realise that, indeed, all simulated NO values were accidentally increased by 1 µg m^{-3} for plotting the data in Fig. 8. As this is obviously wrong, we now provide an updated figure without such increase of simulated NO values and instead remove all observation data below the DL of 1 µg m^{-3}, as their concentration cannot be known and were just arbitrarily set to the DL of 1 in the previous figure.

Since the "average of all sites" line was added to the updated Fig. 8 in response to R2-C7, the revised figure and its caption are now as follows:

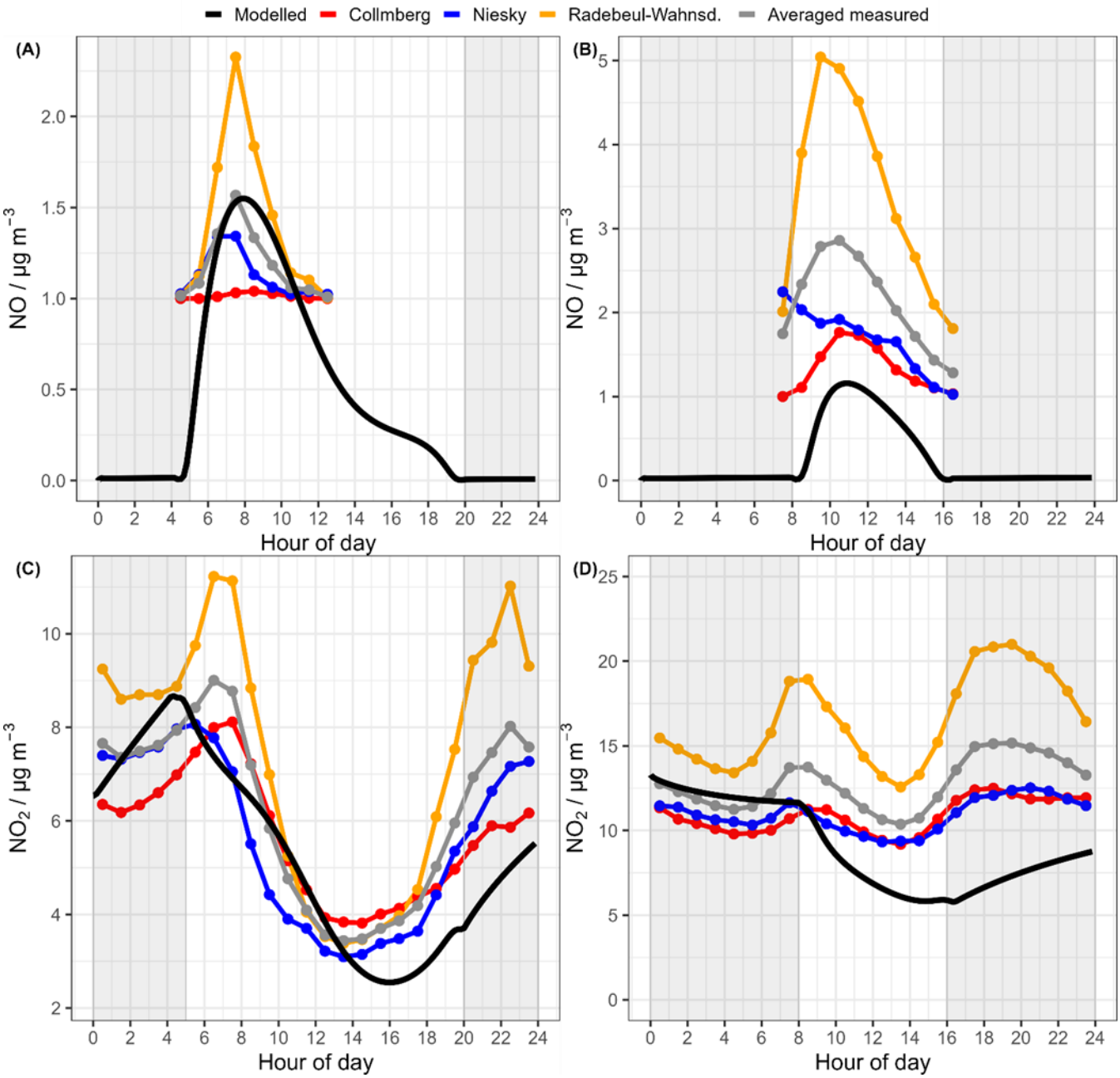


Figure 8: Diurnal profiles of hourly averaged modelled and observed NO and NO_2 in Saxony at rural background sites (A, C for summer case, and B, D for winter cases). Shaded areas indicate night-time, black lines indicate the modelled values, gray lines represent averaged observed concentrations and coloured lines refer to the observed concentrations at each station. It is noted that all observed NO concentrations below the DL of 1 µg m^{-3} are not known and cannot be shown.

To enhance clarity, the paragraph of comparing measured NO and modelled NO at lines 458-475 of the manuscript now reads as follows:

As a first step in model development, the performance of the base cases simulations for Saxony in 2019 was assessed. The year 2019 was chosen because of the best availability of the required input parameters. As a first set of results, the modelled NO and NO_2 diurnal patterns during summer and winter are compared with observed data at rural background stations in Fig. 8. It is noted that all measured NO concentrations below the detection limit (DL) of 1 µg m-3 is unknown and therefore not shown in the Figure. Accordingly, the comparison of the NO patterns is only possible for hours with measured NO above the DL. The modelled NO concentrations during the period from 06:00 to 11:00 (CET) demonstrate satisfactory agreement with the observed summer data (Fig. 8A), with the exception of the most polluted Radebeul-Wahnsdorf site. During winter (Fig.

8B), the model seems to underpredict the NO concentrations maybe because of a too high mixing layer being considered. However, the temporal pattern agrees rather well from 09:00-16:00 (CET). Therefore, the linear correlation coefficient (r) exceeds 0.8 from 06:00 to 11:00 in summer and remains above 0.7 from 09:00-16:00 (CET) in winter, indicating a robust correspondence between model outputs and measurements for both seasons. Additionally, the normalized mean bias factor (NMBF) (see its definition in the Supporting Information) (Jaidan et al., 2018) for model-measurement comparisons of NO is -0.03 during 06:00-11:00 in summer and -0.70 during 09:00-16:00 (CET) in winter, indicating satisfactory agreement of model with mean measurements.

R2-C7: *Figures 8/9: It would be nice to add an "average of all sites" line that's directly comparable to the single model line in Figures 8 and 9. And would also enable you to calculate the normalized mean bias factor (NMBF) summarizing the model-measurement average disagreement (in addition to r2) between average of all sites and modeled results to more robustly support the conclusions of this section.*

A7: We appreciate this suggestion and several lines of "measured averaged values of all sites" have now been added into the updated Figs. 8 and 9.

See the updated Fig. 8 (see R2-C6) and Fig. 9 and its caption below:

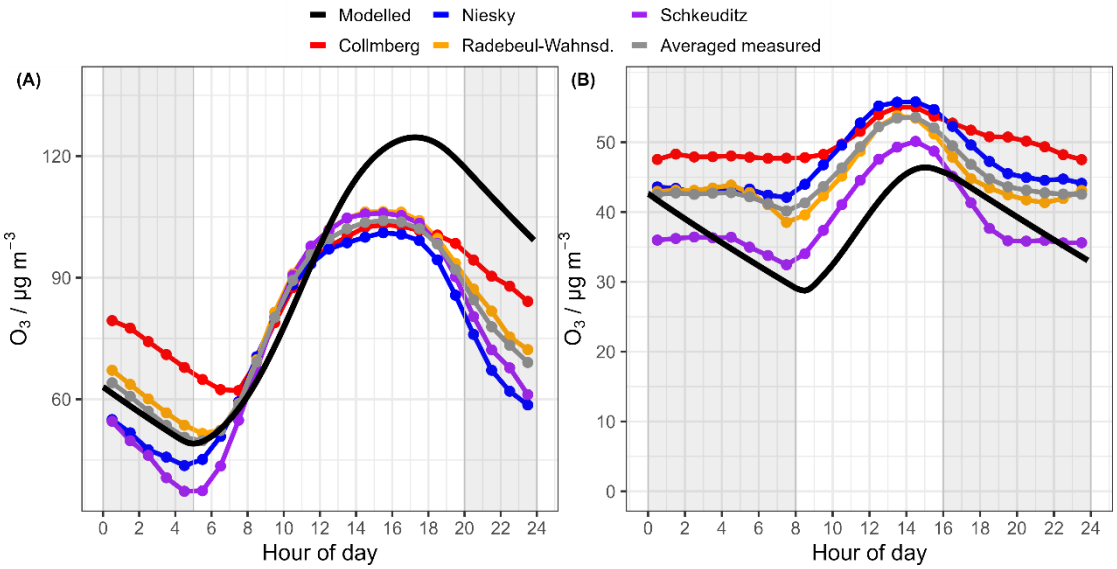


Figure 9: Diurnal profiles of hourly averaged modelled and observed O_3 in Saxony at rural background sites (A and B for summer and winter cases, respectively). Shaded areas indicate night-time, black lines indicate the modelled O_3, gray lines represent averaged observed O_3 and other coloured lines refer to the observed O_3 at each station.

In addition, we have now calculated the normalized mean bias factor (NMBF) summarizing the model-measurement average disagreement. The NMBF values for model-measurement comparisons of O_3, NO, and NO_2 are summarized in the table below.

	Normalized mean bias factor (NMBF)	
	Summer	**Winter**
O_3	0.12	-0.16
NO	-0.03 during 06:00-11:00	-0.70 during 09:00-16:00
NO_2	-0.15	-0.29

In summary, except for NO in winter, the model predictions for NO_2 and O_3, in both summer and winter, overpredict or underpredict the observed values by no more than ±30% on average, based on the normalized mean bias factor.

The definition of the normalized mean bias factor (NMBF) has been added in lines 87-93 of the Supporting Information, as shown below.

Definition of the normalized mean bias factor (NMBF) as a model performance evaluation metric

The hourly average concentrations of NO, O_3, and NO_2 observed (O) across rural background sites in Saxony are calculated in the same way as the modelled hourly averages (M). To evaluate the model performance for each pollutant, the NMBF is used as a statistical metric (Yu et al., 2006; Jaidan et al., 2018), defined as follows:

$$NMBF = \frac{\sum_{i=1}^{n}(M_i - O_i)}{\sum_{i=1}^{n} O_i}$$

where n is the number of hourly mean values for NO, O_3 and NO_2.

In addition, two references (Jaidan et al., 2018 and Yu et al., 2006) have been inserted in lines 194-196 and 215-216 of the Supporting Information. The reference now reads:

Jaidan, N., El Amraoui, L., Attié, J.-L., Ricaud, P., and Dulac, F.: Future changes in surface ozone over the Mediterranean Basin in the framework of the Chemistry-Aerosol Mediterranean Experiment (ChArMEx), Atmospheric Chemistry and Physics, 18, 9351-9373, 2018.

Yu, S., Eder, B., Dennis, R., Chu, S. H., and Schwartz, S. E.: New unbiased symmetric metrics for evaluation of air quality models, Atmospheric Science Letters, 7, 26-34, 2006.

In the manuscript, a sentence for illustrating model-measurement comparisons of NO has been added in lines 472-475. See the sentences below.

Additionally, the normalized mean bias factor (NMBF) (see its definition in the Supporting Information) (Jaidan et al., 2018) for model-measurement comparisons of NO is -0.03 during 06:00-11:00 in summer and -0.70 during 09:00-16:00 (CET) in winter, indicating satisfactory agreement of model with mean measurements.

A sentence for illustrating model-measurement comparisons of NO_2 has been added in lines 490-491 of manuscript. See the sentence below.

Besides, NMBF values for model-measurement comparisons of NO_2 is -0.15 in summer and -0.29 in winter, indicating good agreement of model with mean measurements.

A sentence for illustrating model-measurement comparisons of O_3 has been added in the lines 498-499 of manuscript. See the sentence below.

Additionally, NMBF values for model-measurement comparisons of O_3 is 0.12 in summer and -0.16 in winter, indicating good model performance.

R2-C8: *Figure 10- The markers depicting the years are hard to see on the figure because the lines marking the trends and future emission scenarios are on top of them. I'd suggest flipping the order these are plotted in so that the markers appear on top of the lines so that where each year appears on the figure is easier to see. Additionally. I can't tell the difference between the dark blue that 2010 is plotted in vs the black that 2019 is plotted in, particularly for the summer urban cases that are the focus of the paper. Additionally, I think it would be useful to lower the y-limit a bit on these figures so that it is easier to see what's happening on the lower end of the figures (e.g. whether the reductions in NOx vs. VOCs lead to more /less P(O3) in the future emission scenarios shown as dotted lines on the figure). Some simple edits to this figure would make it much more*

*readable. The caption needs to better describe what the future scenario lines are. These are only mentioned *much* further down in the text far after this figure is first referenced. Additionally, the lines don't all appear to be aligned properly to the last 2019 value… Finally, I wonder if its even necessary to "label" these future scenario lines on the figure itself, as these are pretty typical in the literature. Rather, a key describing if it's a NOx/VOC reduction or a reduction of both after 2019 would suffice and clean up the figure a bit.*

A8: The updated Fig. 10 has been incorporated into the manuscript. The modifications include changing the drawing order of lines and points, adjusting the color of the points, slightly lowering the y-axis limit, and removing "label" of future scenario to make the figure cleaner. In addition, the caption of Fig. 10 has been revised.

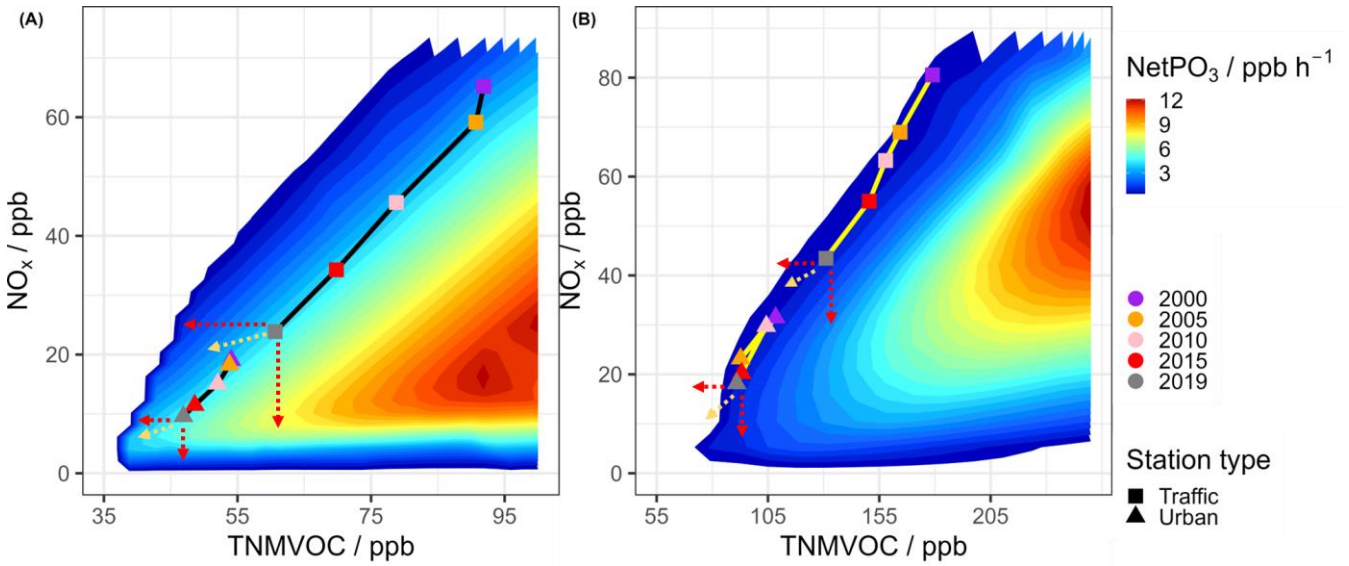


Figure 10: Isopleth plots for the net O_3 production rate ($NetPO_3$ in ppb h^{-1}) during 12:00 - 13:00 CET as a function of averaged NO_x and TNMVOC concentrations (A and B for summer and winter cases, respectively). $NetPO_3$ is shown using a rainbow color scale. The coloured points represent the conditions in different years with squares showing traffic stations and triangles showing mean urban stations conditions. The red dotted arrows represent hypothetical future scenarios with reductions in only one precursor (either TNMVOC or NO_x), while the yellow dotted arrows illustrate more plausible future pathways, where $NetPO_3$ slightly decreases through strong VOCs emission controls combined with moderate NO_x reductions. Note that based on the underlying modelling, these data in Fig. 10 are given as mixing ratios. For better comparability with mass-based concentrations elsewhere in the manuscript, conversion factors of NO_x (µg m^{-3}) ≈ 1.5 × NO_x (ppb) and O_3 (µg m^{-3}) ≈ 2 × O_3 (ppb) can be used.

In addition, to avoid the "missing" information caused by referencing a figure before it is discussed much further down in the text (see the comments in R2-C1 and R2-C3), the relevant details have also been included in the caption of Fig. 10 above. The future scenario pathways were also documented in lines 649-652 of the manuscript. The revised sentences now read as follows:

However, a scenario that only reduce VOCs (as indicated by red dotted arrows to the left in Fig. 10) is not realistic as it comes with the cost of constant NO_x. Therefore, a scenario (as indicated by yellow dotted arrows in Fig. 10) where $NetPO_3$ slightly decreases, achieved through strong VOCs emission control with moderate NO_x emission decrease, is more realistic.

R2-C9: *Lines 490-499: I can certainly appreciate the challenge of not having enough VOC data to create Figure 10. But, I'm generally confused about how TNMVOC is estimated here and further clarification is needed in this section especially. The main text states that "The grid of modelled NetPO3 as a function of modelled, inventory-derived NOx and TNMVOC concentrations (see Sect. 2.3), was therefore interpolated to derive TNMVOC concentrations for given measured NOx and dO3/dt. I read this to mean that they used the model with inventory estimates for NOx and TNMVOC which is what is shown in the column for Table S8? And because they see decent agreement between the model & measurements of NetPO3 and NOx*

that they are assuming TNMVOCs predicted by the model are "right" at each station. What is extremely unclear to me is this "interpolated to derive TNMVOC concentrations for given measured NOx and dO3/dt" statement. That implies that they preformed some sort of correlation/interpolation that is not shown anywhere in the manuscript or supplement to estimate "where to place the markers on Figure 10 on the TNMVOC axis" / get the values shown in Table S9. If that's what was done, I would like to see this figure and have it described with quantitative supporting statistics to convince us that where their TNMVOC estimates are accurate (e.g. show Table S8 as a figure and the equation used to generate the values in Table S9). Additionally, if this methodology has been used in the past for such analysis, it would benefit the paper to reference that such a methodology has been used before in this section to justify this methodology choice. I have seen prior papers use the correlation between CO or HCHO with TNMVOCs to estimate TNMVOC when only CO or HCHO is available in the past, but all of those showed this in the supplement with supporting statistics to show it was a reasonable way to estimate TNMVOCs. Regardless, I'm confused enough about this section that I really don't understand how the values in Table S8 correspond to that shown in Figure S6 or how what's shown in Table S8 is used to get the data in Table S9/ the values used on the TNMVOC axis in Figure 10 and the authors certainly need to clarify this.

A9: We appreciate the reviewer's valuable comment, which helped clarify the explanation for estimating TNMVOC.

First, the sensitivity simulations for drawing O_3 isopleths were done by scaling the base case emissions of TNMVOC and NO_x 20 times in each of three batch runs, i.e. each combination is considered (Table S6 (former Table S5)). Three batches were performed to achieve a sensible range of resulting TNMVOC and NO_x concentrations in the total of 800 and 1200 model runs for summer and winter, respectively. See the newly added Fig. S6 below showing points distribution of resulting TNMVOC and NO_x concentrations.

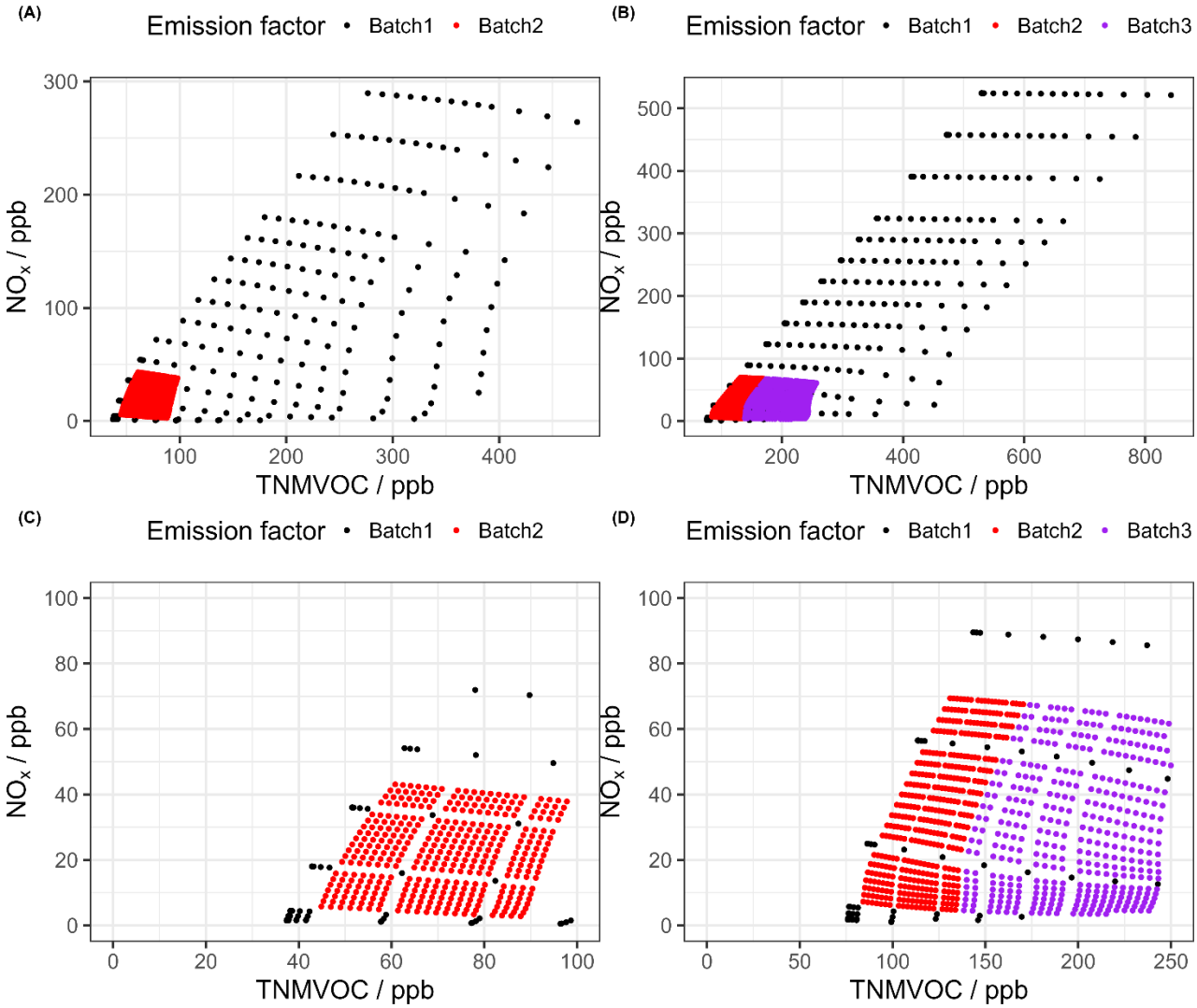


Figure S6: The distribution of resulting TNMVOC and NO_x concentrations (in ppb) based on a total of 800 and 1200 model runs (see Table S6 for details) for summer (A) and winter (B), respectively. Panels C (summer) and D (winter) are subsets of Panels A and B, respectively, and use the same concentration coordinate ranges as Fig. S7. The different colors indicate different batch runs.

Besides, the averaged instantaneous rate of net ozone production ($NetPO_3$) during noon time of 12:00 - 13:00 CET for each simulated scenario in both summer and winter conditions was obtained for each run. The resulting $NetPO_3$ was interpolated onto a regular 1000 × 1000 grid in the TNMVOC vs. NO_x space to generate Fig. S7 (see below). The O_3 isopleths (Fig. 10) were then fitted to this high-resolution grid from Fig. S7. At the same time, two tables present summer and winter data obtained after interpolation. From those, one can identify similar NO_x values along with their corresponding $NetPO_3$ and TNMVOC concentrations.

In our current study, we found in Fig. S8 (former Fig. S6) an excellent correlation between the measured dO_3/dt (especially from 06:00 to 12:00 in summer and 08:00 to 12:00 in winter) and the modeled $NetPO_3$. So, we regard measured dO_3/dt (from 06:00 to 12:00 in summer and 08:00 to 12:00 in winter) a good proxy to the value of modeled $NetPO_3$. By picking the known NO_x and dO_3/dt (finding the close values of $NetPO_3$), the TNMVOC concentration is then identified. The derived TNMVOC together with a comparison of measured and modelled NO_x and dO_3/dt for the station types are given in Table S9 (former S8) and Table S10 (former S9).

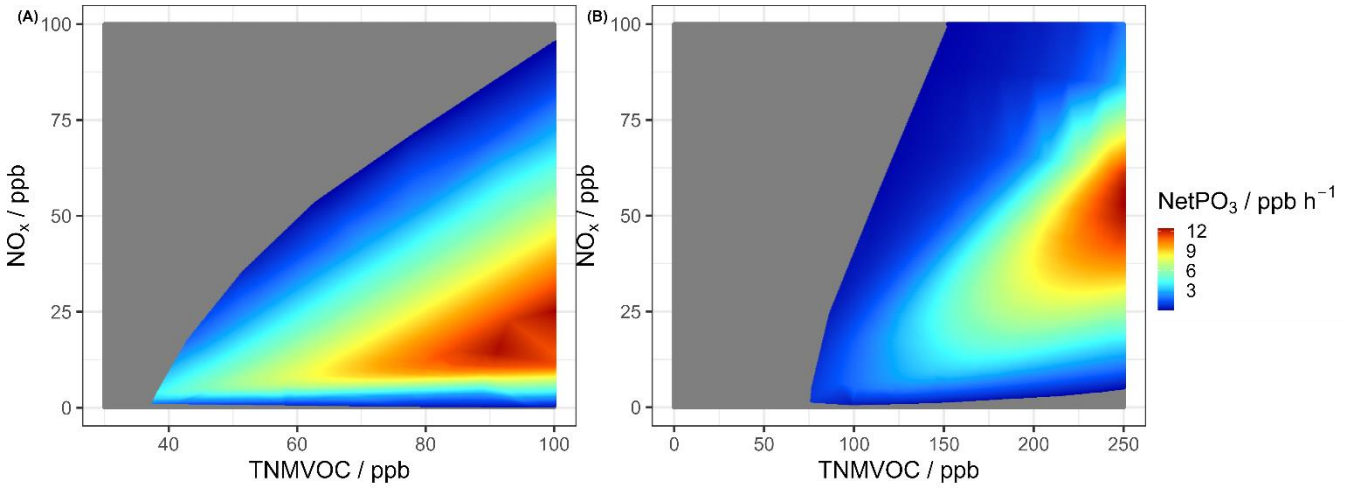


Figure S7: The grid of modelled net O_3 production rate ($NetPO_3$ in ppb h^{-1}) during 12:00 - 13:00 CET as a function of both modelled and emission inventory-derived NO_x and TNMVOC concentrations (in ppb) (see Sect. 2.3) after bivariate linear interpolation. The X-axis and Y-axis in both summer (A) and winter (B) are interpolated using a very fine-resolved 1000 × 1000 grid. $NetPO_3$ is shown using a rainbow color scale.

The revised paragraphs on how to estimate TNMVOC in lines 539 - 557 of the manuscript are as follows:

In the next step, the measured values of NO_x and observed O_3 change rate (dO_3/dt) for the years 2000, 2005, 2010, 2015, and 2019 were used to indicate for each year the location in the isopleth diagram. A key challenge, however, was the lack of measured TNMVOC concentrations in all years. To work around this, three simulation batches (see Sect.2.3 for details) were performed to achieve a sensible range of resulting TNMVOC and NO_x concentrations in the total of 800 and 1200 model runs for summer and winter, respectively. Points distribution of resulting TNMVOC and NO_x concentrations is shown in Fig. S6. Besides, the averaged $NetPO_3$ during noon time of 12:00 - 13:00 CET for each simulated scenario in both summer and winter conditions was obtained for each run. The resulting $NetPO_3$ was interpolated onto a regular 1000 × 1000 grid in the TNMVOC vs. NO_x space to generate Fig. S7. The O_3 isopleths (Fig. 10) were then fitted to this high-resolution grid from Fig. S7. At the same time, two tables present summer and winter data obtained after interpolation. From those, one can identify similar NO_x values along with their corresponding $NetPO_3$ and TNMVOC concentrations.

As depicted in Fig. S8, measured dO_3/dt and modelled $NetPO_3$ agreed reasonably well, particularly from 06:00 to 12:00 in summer and 08:00 to 12:00 in winter. This indicates that the measured dO_3/dt during these periods serves as a good proxy to the value of modelled $NetPO_3$, which is why it is considered valid to interchange them in the present application. By picking the known NO_x and dO_3/dt (finding the close values of $NetPO_3$), the TNMVOC concentration is then identified. For further clarification, in Tables S9 and S10, these TNMVOC estimates are shown together with a comparison of measured and modelled NO_x and dO_3/dt for the station types.

As a result of adding two new figures (as Fig. S6 and S7), the numbering of other figures (from Fig. S6) in the Supporting Information has changed accordingly.

R2-C10: *Line 499-501: "… because of lowered emissions through environmental mitigations." Are you talking about reductions in anthropogenic VOCs through emission controls/regulations or reductions in biogenic VOCs? Would be useful to clarify what is meant by "environmental mitigations" here / which VOC category is assumed to be affected.*

A10: We were referring to anthropogenic VOCs reductions, yes. The revised sentence in the lines 558-560 of the manuscript now reads:

Notably, simulated TNMVOC concentrations, primarily emitted or locally photochemical in origin, are expected to be similar to or lower than previous measurements (Knobloch et al., 1997) because of lowered anthropogenic emissions through existing European regulation and corresponding mitigation measures.

R2-C11: *Table S4. I appreciate the authors giving the smiles strings of the VOC compounds, but it would be useful for reproducibility if they also supplied the MCM species names of each in another column. As an MCM box modeler, if I wanted to recreate this study, I would have to go through the MCM mechanism and identify each one of these compounds by hand in order to recreate their simulation. Thus, while giving the strings does mean that these concentrations could be used in other mechanisms (and why it is important to retain that information), for reproducibility, it would also be useful to have the MCM compound name (since there are indeed existing "mapping" tools to map those tracer names to the tracers of other mechanisms).*

A11: Some SMILES strings are available on the MCM website, while others are not. It is challenging to provide the name of each MCM species. Additionally, there is no publicly accessible webpage in the CAPRAM link for referencing these SMILES strings - they are only accessible to CAPRAM users.

Instead of listing the SMILSE strings twice, one column with the full SMILES strings, one column with their IUPAC names and one column with their chemical structures are presented in the updated table.

As a result of adding one new table of Table S3, the numbering of Table S4 now should be Table S5. The revised caption of Table S5 has been added in line 38 of the Supporting Information, now reads:

See below the revised Table S4.

Table S5. Dominant initial gas-phase concentrations applied in the final 24-hour simulations for summer and winter scenarios. Each species is given by its SMILES string, and its IUPAC name generated using the PubChemPy package in Python (https://pubchempy.readthedocs.io/en/latest/), along with the corresponding chemical structure. It is noted that for some compounds, IUPAC names could not be retrieved from the PubChem database by the tool and they are therefore left blank.

See attached Excel file named *Table S5_Dominant initial gas-phase concentrations.*

References

Brümmer, B., Lange, I., and Konow, H.: Atmospheric boundary layer measurements at the 280 m high Hamburg weather mast 1995-2011. Mean annual and diurnal cycles, Meteorologische Zeitschrift (Berlin), 21, 2012.

Jaidan, N., El Amraoui, L., Attié, J.-L., Ricaud, P., and Dulac, F.: Future changes in surface ozone over the Mediterranean Basin in the framework of the Chemistry-Aerosol Mediterranean Experiment (ChArMEx), Atmospheric Chemistry and Physics, 18, 9351-9373, 2018.

Knobloch, T., Asperger, A., and Engewald, W.: Volatile organic compounds in urban atmospheres: Long-term measurements of ambient air concentrations in differently loaded regions of Leipzig, Fresenius' journal of analytical chemistry, 359, 189-197, 1997.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmos. Meas. Tech., 16, 433-479, 10.5194/amt-16-433-2023, 2023.

Lyu, X., Wang, N., Guo, H., Xue, L., Jiang, F., Zeren, Y., Cheng, H., Cai, Z., Han, L., and Zhou, Y.: Causes of a continuous summertime O 3 pollution event in Jinan, a central city in the North China Plain, Atmospheric Chemistry and Physics, 19, 3025-3042, 2019.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmospheric Chemistry and Physics, 3, 161-180, 2003.

Thürkow, M., Schaap, M., Kranenburg, R., Pfäfflin, F., Neunhäuserer, L., Wolke, R., Heinold, B., Stoll, J., Lupaşcu, A., and Nordmann, S.: Dynamic evaluation of modeled ozone concentrations in Germany with four chemistry transport models, Science of the Total Environment, 906, 167665, 2024.

Wang, W., Li, X., Cheng, Y., Parrish, D. D., Ni, R., Tan, Z., Liu, Y., Lu, S., Wu, Y., Chen, S., Lu, K., Hu, M., Zeng, L., Shao, M., Huang, C., Tian, X., Leung, K. M., Chen, L., Fan, M., Zhang, Q., Rohrer, F., Wahner, A., Pöschl, U., Su, H., and Zhang, Y.: Ozone pollution mitigation strategy informed by long-term trends of atmospheric oxidation capacity, Nature Geoscience, 10.1038/s41561-023-01334-9, 2023.

Wiegner, M., Emeis, S., Freudenthaler, V., Heese, B., Junkermann, W., Münkel, C., Schäfer, K., Seefeldner, M., and Vogt, S.: Mixing layer height over Munich, Germany: Variability and comparisons of different methodologies, Journal of Geophysical Research: Atmospheres, 111, 2006.

Yu, S., Eder, B., Dennis, R., Chu, S. H., and Schwartz, S. E.: New unbiased symmetric metrics for evaluation of air quality models, Atmospheric Science Letters, 7, 26-34, 2006.