

Many thanks to both reviewers for this round of comments. We have implemented these, the details of which are described below.

Report #1

I would like to thank the authors for taking an extensive revision of the manuscript. I would recommend a few technical revisions before potential acceptance.

1) This paper focuses on urban ozone, but the discussion of "urban" part is completely lost after introduction. It is desirable to add their implications specifically for urban chemistry and air quality.

We have added the following to the conclusions to more clearly frame the discussion w.r.t urban conditions

“The goal of this study was to determine trends in O₃ metrics relevant to human exposure, in urban locations. Due to non-linear relationships with precursor concentrations, policy intervention or environmental changes may not have straightforward impacts on ambient O₃.”

“Recent models show positive trends in HCHO to NO₂ ratios over the USA and Europe demonstrate a tendency towards NO_x-limited regimes (Fadnavis et al., 2025), but high heterogeneity within urban areas and differences between urban areas means that this behaviour cannot be generalised to all monitoring sites.”

“This is particularly true for urban sites, for which O₃ precursor emissions are dependent on population density and growth, as well as climate and topography.”

2) Although I know the rationale, it seems to lack an explicit explanation on why 2004 and 2018 are chosen as benchmark years (Fig 4 and further discussion).

Added the following to the end of section 3.1

“Later comparisons use 2004 and 2018 as "snapshot" years informed by this, as to avoid these potential edge or COVID-19 effects and the major 2003 heatwave in Europe and still being near the beginning and end of the study period. In these comparisons, 2018 was used despite some European sites experiencing a less widespread heatwave, as it is shown to have a lesser effect on the O₃ metrics (section 3.3), and figure SI1 does not show a difference in change points from non-heatwave years.”

Minor comments:

L28, technically the TOAR activity is not the first global review of trends in surface ozone, e.g. Cooper, O. R., Parrish, D. D., Ziemke, J., Balashov, N. V., Cupeiro, M., Galbally, I. E., ... & Zbinden, R. M. (2014). Global distribution and trends of tropospheric ozone: An observation-based review. *Elementa*, 2, 000029.

“... was the first global review of trends in surface ozone ...” replaced with:
“... was an extensive review of trends in global surface ozone ...”

L32, these results are explicitly quantified in Chang et al. (2017), which is also a part of TOAR-I: Chang, K. L., Petropavlovskikh, I., Cooper, O. R., Schultz, M. G., & Wang, T. (2017). Regional trend analysis of surface ozone observations from monitoring networks in eastern North America, Europe and East Asia. *Elem Sci Anth*, 5, 50.

The following text has been added following L35:

“Chang et al. further quantify summertime ozone trends in eastern North America and Europe, over the same period, using several metrics including: the monthly mean of the daily maximum 8-hour average (DMA8) and NVGT070, a modified form of NDGT70 where only summertime days are counted. It is shown that trends in DMA8 across all sites are decreasing in areas in both regions, but in urban areas the decreasing trend is only seen in eastern North America (-0.25 ppb yr⁻¹). For NVGT070 both eastern North America and Europe showed decreasing trends (-1.03 and -0.26 days yr⁻¹ respectively).”

Section 2.1, some citations need to be fixed, e.g. (204 Europe, 149 USA).

These refer to the number of sites, the sentence has been adjusted for clarity:
“This resulted in 353 O₃ time series, 204 in Europe and 149 in USA.”

Report #2

The reviewers' reports on the original version of this paper acknowledged the effort expended in examining trends in the 22-year datasets of urban ozone but also raised substantive concerns regarding choice of ozone metric, the methodology of the trend analyses, and the overall motivation for the work.

The authors have taken these comments seriously and have essentially re-done all their trend analyses. The last paragraph of the Introduction is now much clearer about the aims of the work.

The authors now use the maximum daily 8-hour running mean (MDA8) rather than the daily median as their measure of daily ozone. They also investigate trends in other metrics that are commonly used to capture levels and human exposures to ozone. To compare how trends in ozone levels have changed (or not) through their 22-year time period they now more simply compare the magnitude of trend for a timepoint towards the end of their time period (2018) with that at a timepoint towards the start of their time period (2004). They also investigate trends separately for the warm season and for the cold season.

The revised paper also contains a new analysis, which is the grouping of time series with similar trends to visualise the extent to which trends at different sites do or do not group in distinct

geographic areas (within Europe and the USA separately). This is an interesting visualisation of the spatial homogeneity/heterogeneity in 22-year ozone trends across different measurement sites.

Trying to explain all the observed temporal and spatial patterns in ozone trends is of course a difficult matter, given the huge number of factors that influence ozone concentrations at any given time at any given location. The authors endeavour to provide some explanations, although discussion is still a bit lightweight in addressing the questions “so what have we learnt” and “what does it mean going forward”? The final sentence of the conclusion is one instance that does describe scientific conclusion. On the other hand, detailed interpretation would need very substantial investigations using full process-based atmospheric chemistry modelling, which is not what this paper is about, so perhaps the authors have gone as far as they can in providing ‘user-relevant’ information.

We appreciate the comments regarding the discussion, but agree that without a substantial increase to the scope of the work, we believe we have gone as far as we can. However, as part of our response to reviewer #1 comment 1), we have added briefly to the conclusions which may be relevant to the above.

The paper can be recommended for publication.

Some minor points:

The paper would have benefited from a careful proof read as there are quite a lot of missing or extra words.

Thank you for those you have already found below, in addition we have extensively checked the document and corrected spelling and grammatical errors we found.

The new Section 3.3 is potentially informative analysis but there is no cross-referencing to which of the many similar figures the reader needs to be looking at. Likewise, Section 3.4 contains a lot of description of different results of analyses but with no cross-referencing to tell the reader where they need to be looking for the visual support for these descriptions (just a blanket one-off statement that the observations are based on the twelve figures S10 to S21).

Thanks for bringing this to our attention, we have added more cross-referencing throughout sections 3.3 and 3.4 hopefully improving the readability of these sections.

L5: “the period”

Corrected

L30: Make clear that the 4MDA8 is calculated using only days in the warm season whilst NDGT70 is for all days in the year.

Updated to: "... 4th highest daily maximum 8-hour O₃ in the warm season (4MDA8), and the number of days in a year with MDA8O₃ > 70 ppb O₃ (NDGT70) ..."

L65: Insert "to" before "be"

Done

L84: This definition of an MDA8 comes after the acronym MDA8 has already been used in the intro and methods sections.

Definition moved to 2.1

L88: The last part of this sentence doesn't make grammatical sense.

Language simplified to: "...and here we subtract monthly mean climatologies from each time series."

L89: The acronym QR appears here but has not yet been defined or described.

Now defined at the beginning of the paragraph: "These were calculated via quantile regression (QR) following the methodology in..."

L92: Delete extraneous "and the"

Done

L119: The start of this sentence is non-grammatical.

Reworded to: "To aid analysis it was desirable to organise similar time series into groups, which was achieved by applying hierarchical clustering..."

L142: "corresponding"

Done

L160: Reword to "closest to the beginning"

"...avoiding years closest to the beginning or the end of the series"

L197: Choose "were" or "are"

Done

L200: Sort out grammar in "Similar to the Europe"

"Similarly to those in Europe..."

L235: Suggest replacing “by region” with the more informative “by continent” and inserting the word “geographical” before “structure”

As this is the only location ‘continent’ would appear over ‘region’ we opted to reword to the following: “So far, trends have only been separated spatially by region (USA or Europe), but it is clear when visualising these that there is more geographical structure.”

L239 onwards throughout the rest of Section 3.3: The reader’s attention needs to be drawn to Figs. 7 and 8 and other supplementary figures, as relevant, when introducing the results from the hierarchical clustering and when referring to specific features shown in these figures.

As above, this section has had improved cross-referencing added.

L255: Visible where? Cite to the figure where these enhancements are visible.

Reference to figure 7 added earlier in paragraph

L361: Correct text to read “sites in Europe were almost exclusively”

Done

Figure 2 caption: The description of what metrics are plotted does not correctly reflect the text at the top of each column of plots says is plotted.

Text corrected to match figure

Figure 4: The font sizes need to be larger for all the labels, axis titles, panel titles, legend, etc.

The entire figure has been made larger, and we will liaise with typesetting to ensure legibility is preserved.