

# Urban Ozone Trends in Europe and the USA (2000-2021)

Beth S. Nelson<sup>1,2,\*</sup> and Will S. Drysdale<sup>1,2,\*</sup>

<sup>1</sup>Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK

<sup>2</sup>National Centre for Atmospheric Science, University of York, York, UK

\*These authors contributed equally to this work.

**Correspondence:** Beth S. Nelson (beth.nelson@york.ac.uk) and Will S. Drysdale (will.drysdale@york.ac.uk)

## Abstract.

Trends in urban ~~OMDA8O<sub>3</sub> and NO<sub>2</sub>~~, alongside annual O<sub>3</sub> episodic and exposure metrics, across Europe and the United States of America were explored between 2000-2021. Using surface monitoring site data from the TOAR-II and European Environment Agency databases, piecewise quantile regression (PQR) analysis was performed on ~~228-0353 MDA8O<sub>3</sub> time series (144 European, 84 USA) and 322 NO<sub>2</sub> times series (245 European, 77-205 European, 149 USA)~~. The PQR analysis permitted 2 break points over ~~the 23-year~~ period to balance the intent to describe changes over a large time period, while still capturing the abrupt changes that can occur in urban atmospheres. ~~Regressions were performed over quantiles ranging from 0.05 to 0.95 and indications of a slowing in the increase of high European O<sub>3</sub> levels was observed. In Europe, more trends~~ We found that there were many sites across Europe with high certainty increasing trends in the 5<sup>th</sup> and 50<sup>th</sup> quantiles, whereas the majority of high certainty trends in the 95<sup>th</sup> quantile were found to ~~having an increasing O<sub>3</sub> trend between 2015-2021 compared with 2000-2004. The reverse was true in be decreasing. A similar pattern was observed across~~ the USA, with ~~a reduction in the 5<sup>th</sup> quantile trends increasing and 95<sup>th</sup> quantile trends decreasing, though a small but increasing number of sites with increasing O<sub>3</sub> trends when the same periods were compared. An analysis of the change points revealed a large proportion of sites in Europe~~ ~~were the second change point in NO<sub>2</sub> switched from a positive to negative trend, occurred in 2020 (41/43 second change points in this year).~~ This was attributed ~~a reduction in NO~~ showed a return to increasing trend at  $\tau = 95$ . To group trends, hierarchical clustering with dynamic time warping was employed and these groups used to guide analysis. Clustering was typically regional across Europe and the USA, and increasing trends were identified across Southern and the Central alpine regions of Europe, and in California and the Intermountain West of the USA. Recent high certainty increasing trends in MDA8O<sub>23</sub> ~~due to the COVID-19 pandemic, however, in some cases these increasing trends have sustained beyond the recovery from restrictions. in~~ the Intermountain West may be related to warmer summers and increased wildfire events in the region, highlighting the need to monitor changing trends in MDA8O<sub>3</sub> with climate change, to assess human exposure risk to elevated levels of O<sub>3</sub>.

## 1 Introduction

Tropospheric ozone (O<sub>3</sub>) is a greenhouse gas and air pollutant harmful to human health, and plant growth (???). It is a secondary air pollutant, formed from the photochemical reactions of primary pollutants NO<sub>x</sub> (NO + NO<sub>2</sub>) and volatile organic

25 compounds (VOCs). The chemistry of O<sub>3</sub> formation is non-linear and the effect of changing precursor concentrations can cause O<sub>3</sub> concentrations to increase or decrease depending on the photochemical regime found at a given location (??). Despite global successes in reducing primary pollutant emissions over the past few decades, global exposure to O<sub>3</sub> has been increasing throughout the 21st century. This is particularly observed in urban areas, where the vast majority of the global population live, projected to increase to 68 % in 2050 from 55 % in 2018 (?). In a study of 12946 cities located worldwide, the average mean  
30 weighted O<sub>3</sub> concentration increased by 11 % between 2000 and 2019, and the number of cities exceeding the 2021 WHO peak season O<sub>3</sub> standard (60 µg m<sup>-3</sup>) increased from 89 % to 96 % (?).

Due to the complexity of O<sub>3</sub> production, its trend direction, magnitude, and significance varies by location. ~~A previous study calculated~~ The Tropospheric Ozone Assessment Report Phase I (TOAR I) was the first global review of trends in surface ozone, covering 1970 - 2014 (??). As part of the TOAR-I review, trends in two O<sub>3</sub> metrics were calculated globally over the period  
35 of 2000 - 2014: 4th highest daily maximum 8-hour O<sub>3</sub> (4MDA8), and the number of days with ~~MDA8~~ MDA8O<sub>3</sub> > 70 ppb O<sub>3</sub> (NDGT70)(?). The study used data from 4801 global monitoring sites over this time period. For both of these metrics, downward trends were observed for most of the USA, and some sites in Europe. However, whilst trends were decreasing, over the period of 2010-2014 (2,600 sites utilised), sites some of these sites were located in regions with the highest O<sub>3</sub> precursor emissions across North America, Europe and East Asia that had the highest values in 4MDA8 and NDGT70. ~~In North America and Europe, this was particularly true for California and parts of southern Europe(?)~~ values, particularly California and Southern Europe. These corresponded with the regions with the highest O<sub>3</sub> precursor emissions illustrating that downward trends had not yet reduced these peak metrics.

Since the 1990s, a general downward trend in urban O<sub>3</sub> pollution has been observed in the United States (?). This reducing trend has been linked to stricter limiting regulations on the emissions of primary pollutants such as NO<sub>x</sub> and VOCs. Although  
45 NO<sub>x</sub> and VOC emissions in Europe have also been declining since the late 1980s, the trend in O<sub>3</sub> is less clear due to large inner-annual variation, driven by climate variability and the dispersion and transport of pollutants from other regions (??). Between 1995 and 2014, negative trends in the highest O<sub>3</sub> levels across urban sites in Europe were identified due to pollutant emission restrictions across Europe. However, increasing background levels, particularly in northern and eastern Europe, make it difficult to identify strong trends in urban O<sub>3</sub> when transboundary effects are considered (?). Despite this, a study of 93  
50 suburban and urban sites across Europe identified notable enhancements in O<sub>3</sub> seasonal and annual means between 1995 - 2012, even with the continuous downward trend in anthropogenic emissions across the continent (?).

~~Since these earlier studies, much of the literature focuses on the impact of the COVID-19 pandemic on air pollutant concentrations and trends (????). A study of eleven cities across the world, all of which implemented stringent lockdown measures in early 2020, captured sudden decreases in deweathered NO<sub>2</sub> concentrations, concurrent with sudden increases in~~  
55 ~~O<sub>3</sub>, in most locations (?). Another study employing machine learning models to predict business-as-usual levels of pollutants, estimated that NO<sub>2</sub> concentrations were 32% lower than expected across 102 European urban background locations, whereas O<sub>3</sub> was 21% higher (?). It is highly likely that this global event has resulted in perturbations of both long-term NO<sub>2</sub> and O<sub>3</sub> trends, particularly in locations where lockdowns were stringent or lengthy.~~

~~This study examines~~ In this study, we aim to investigate a longer time-series of 21st century trends (2000 - 2021). Here, we focus on the trends in urban O<sub>3</sub> and NO<sub>2</sub> using, using urban monitoring site data from 2020–2023. It employs Europe and the United States of America across the 22 year period. We employ quantile regression and change point detection to construct trends that capture the broad structure of a complex time series, while remaining explainable with concise statistics. Section 2 details the method used to create the trends, section ?? provides a high-level overview of the trends with section ?? focusing on the significance of the trends and highlighting features that are discussed in detail later on. Section ?? explores the information that can be gained from trends calculated over a range of quantiles and finally section ?? provides details on some of the features revealed earlier in the analysis. To gain a clearer understanding of the group behaviour of sites, we use hierarchical clustering (HC) with dynamic time warping (DTW), to group together 22 year time series with comparable time-series structures within Europe and the United States of America independently. This is then used to investigate whether trends in MDA8, and both episodic and exposure-relevant O<sub>3</sub> metrics, are varying regionally. Our analysis focuses comparisons between warm-season and cold-season MDA8O<sub>3</sub> values, as well as probing different changes across quantiles, to allow for an assessment of whether low or high O<sub>3</sub> distributions have different behaviours, across Europe and the USA, and between hierarchical clusters.

## 2 Methodology

### Hourly NO<sub>2</sub> and

#### 2.1 Data Preparation

Hourly O<sub>3</sub> data were obtained for urban sites in the USA and Europe using the TOAR-II (?) and European Environment Agency (EEA) (??) databases, ~~using the r-packages toarR (?) and saqgetr (?) respectively~~. Those obtained from the EEA are categorised as urban traffic, urban background and urban industrial, while those from the TOAR database are categorised as urban and suburban. A full list of sites can be found in the supplementary information (table S1S1). TOAR-II data was retrieved between 2000-01-01 and 2021-12-31 (the latest available) and EEA data between 2000-01-01 and 2023-12-31, the latter being extended ~~longer so the effects of allowing break points to be fit around the changes due to the COVID-19 pandemic could be observed more clearly. Time-~~ though we remain focused on 2001-2021 in this study. Time series were quality controlled via threshold, persistence and variance checks similar to those by ?: O<sub>3</sub> concentrations were required be between 0 and 500 ppbv, periods where an identical value was repeated 8 or more times in a row were removed as were days where the difference between the minimum and maximum concentration was  $\leq 2$  ppbv.

After quality control, time series were required to have ~~90 > 80 %~~ data coverage between 2000-01-01 and 2021-12-31 and ~~retained time series were averaged to daily medians~~ for the calculation of annual metrics each individual year was retained if it had more than 60 % coverage. Additionally, a small number of time series were removed following visual inspection inspection due to large changes in the mean of data before and after a missing period, indicating possible sensor issues - these are listed in table S2S2. This resulted in 228–353 O<sub>3</sub> time series (144 Europe, 84 USA), 322 NO<sub>2</sub> time series (245 Europe, 77 USA) and 126 O<sub>x</sub> (NO<sub>2</sub> + O<sub>3</sub>) time series (101 Europe, 25 USA) 204 Europe, 149 USA).

Firstly, the time series were de-seasoned by calculating a monthly median climatology and subtracting this from the time series, producing an ‘anomaly’ time series. The  $\text{NO}_2$  and  $\text{O}_3$  anomaly time series were summed to produce  $\text{O}_x$  anomaly time series for sites that have both  $\text{NO}_2$  and  $\text{O}_3$

### 2.1.1 Metrics

95 In addition to the series described above, we calculated six common health related  $\text{O}_3$  metrics which are explored in section 3.4 and used in conjunction with calculated trends in 3.5. More information on these metrics can be found in ? and ?.

- 4MDA8 - The 4th highest MDA8 value in the warm season (ppbv)
- NDGT70 - Number of days in a year where  $\text{MDA8} > 70$  ppbv (days)
- SOMO35 - Subtract 35 ppb from the MDA8 value and sum all positive values (ppbv days)
- 100 – 3MMDA1 - Annual maximum value of 3 month rolling mean of MDA1 (maximum daily 1 hour  $\text{O}_3$ ) (ppbv)
- 6MMDA1 - As 3MMDA1 with a rolling 6 month mean (ppbv)
- AVGMDA8 - Mean of the warm season MDA8 (ppbv)

## 2.2 Trend Analysis

Trends were calculated on the maximum daily 8-hour average  $\text{O}_3$  data available. Following this, several methods were defined to help describe trends. As the time series are non-linear Locally Estimated Scatterplot Smoothing (LOESS) was applied first. This method works by calculating many least squared fits over a moving weighted subset of the data — with this subset tuned based on the desired ‘smoothness’ of the result. In this case with a smoothing parameter ( $\alpha$ ) chosen as 0.5, which was determined by inspection to capture a balance of broad trends and medium-term non-linearity. LOESS works well for drawing the eye to features of the time series but does not allow for broad trends to be described numerically as the resulting slopes are changing continuously. concentrations ( $\text{MDA8O}_3$ ), which is calculated by taking a rolling 8-hour mean, then selecting the highest value per day between 0700 - 2300 L. The 8 hour rolling average is required to have 6 of the 8 hours present to be valid. Additionally, subgroups of warm (April - September) and cold (October - March) season were created. It is common to remove the seasonal component of an ozone time series to improve the accuracy of trends (?) and here we subtract monthly mean climatologies for each time series resulting ozone anomalies.

115 As such, quantile regression (QR) was used to define trends. These were calculated following the methodology in, and using code provided by, ?. QR calculates a linear model that seeks to minimise the residuals with a defined proportion ( $\tau$ ) of the points above and below the fit line. For example, the scenario  $\tau = 0.5$  splits the data 50:50 above and below the line. QR has the advantage of being insensitive to outliers and can be considered analogous to a “median” trend line. This QR has the advantage of being insensitive to outliers and the which is desirable for the longer-term trends being investigated here. The 1-sigma uncertainty for the QRs were calculated via a moving block bootstrapping method, where the data are subdivided into

120

overlapping blocks that are  $n^{1/4}$  points wide (or ~20 points for 20 years of hourly data). These blocks are shuffled to generate a new time series, and replicated 1000 times, and the standard deviation of these replicates calculated. This uncertainty was subsequently used in the determination of the p-value, providing a metric for the significance of the trends.

As described the QR will be a poor representation of a time series with Quantile regressions were calculated piecewise with 0 - 2 breakpoints to allow the trends to capture large non-linearities (hence first investigating the LOESS) as it will be using a single slope to define the whole trend across a complex time series. As urban concentration trends can change sharply (e.g. with policy intervention), this gives the model the model requires freedom to represent these, ~~therefore they were extended to use piecewise quantile regressions (PQR) with up to two break points (resulting in three trend lines). This allows~~. Limiting the model to ~~have some changes in the trends calculated,~~ a maximum of two breakpoints strikes a balance between capturing sufficiently large-scale changes while still being able to describe a ~20 year time series with a small number of coefficients.

To determine what break points to use on each time series, ~~109 scenarios per time series defined by 1~~ a range of candidate models were constructed, each with zero, one or two break points. These break points were restricted to be unable to occur within the first or last two years of the time series, 2 or 3 sequential QRs were created, separated by break points occurring on the 1st of January in up to two years excluding the first and last 2 years of a time series. In the cases where there were 2 break points, these were not allowed to nor could they occur within 5 years of each other. ~~The cases with 2 or 3 regressions were combined into a single PQR, (PQR1 and PQR2 respectively)~~ They were arbitrarily set to occur on the 1st January in a given year, which was deemed an appropriate degree of freedom given the 20 year span of the time series and locating the break points sub annually does not have a large effect on the resulting trends. For a time series that fully spanned the range of 2000-01-01 - 2021-12-31, 110 models would be created - one with zero break points, 18 with one break point, and 91 with two. All regressions were calculated at  $\tau = 0.05, 0.10, 0.25, 0.50, 0.75, 0.95$  and  $0.95$ .

To select ~~'change points'~~ 'change points' from the array of break points available, the model performance was evaluated via Akaike information criterion (AIC). ~~The piecewise scenario with the minimum AIC at  $\tau = 0.5$  were selected as the change points for a given time series. This is a decision that trades off between there being potential change points that only exist in cases where  $\tau \neq 0.5$ , and retaining the ability to relate trends to one another across all quantiles. As this study aimed to explore the trends across a large number of sites, defining the change points only using  $\tau = 0.5$  is more appropriate, but future studies that reduce the number of time series could consider investigating how change points differ with AIC was chosen as the evaluation criteria owing to its penalty term penalising models with more break points that do not appreciably improve the model fit over the simpler case. The model with the minimum AIC was selected from the candidate pool as the best fit for a time series at a given  $\tau$  in more detail.~~

Figure ?? demonstrates this process on a single time series. Change points could be assessed for their efficacy (i.e. is the effect of introducing a change point important, or has it only provided a trivial reduction in the AIC?) by comparing the QR and the best PQR1 and PQR2 cases, but as this does not substantively change the results presented here, this was not done and the PQR2 fits were determined to be the most appropriate in all cases.

Example trend determination for an  $O_3$  time series (London Bloomsbury - GB0566A). A -  $O_3$  concentration time series (black) and monthly median climatology (red). B - Anomaly time series (black, concentration minus climatology) and 4 trend

options, LOESS (red), QR (blue), PQR1 (green) and PQR2 (yellow). C—Anomaly time series (black) and PQR2 across  $\tau = 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95$ . (purple to green)

### 3 Results and Discussion

Median concentrations of  $O_3$  (top) and  $NO_2$  (bottom) across all sites in this study, separated into sites in Europe (left) and sites in the USA (right). Shaded regions represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles.

To provide context for the subsequently calculated trends, figure 1 shows the annual median, 25<sup>th</sup> and 75<sup>th</sup> percentile concentrations of  $O_3$  and  $NO_2$  across all the urban sites used in this study. Europe shows median  $O_3$  concentrations increasing from 18 ppb to 26 ppb between 2000 and 2019, and USA concentrations increasing, from 26 ppb in 2000 to 30 ppb in 2019. Both Europe and the USA show decreasing  $NO_2$  over the same period (16 ppb to 10 ppb in Europe, 15 ppb to 6 ppb in the USA). The USA does see a return to increasing  $NO_2$  concentrations after 2020, reaching 7.8 ppb in 2022. In section ?? evidence for this increase is discussed.

The dataset that results from the application of the method described in section 2 has several degrees of freedom to bear in mind during discussion. For a given time series (one species at one site) there are two break points that are common for all the trend lines, and as there are 7 quantiles, this results in 21 slopes, each with an associated p-value. When comparing between sites (or even species within a site) the change points can differ. This is necessary to capture the changing trends observed at urban sites but does complicate describing them. For this discussion we have grouped the sites into those from Europe and those from the United States, though details on individual European countries are available in table S1. To aid in summarising, in some cases the results have been grouped into time periods of a few years—in these cases if a trend changes due to a change point both trends have been counted. For example, if between 2000 and 2004 a site were to go from increasing  $O_3$  to decreasing  $O_3$ , this would be described as ‘between 2000 and 2004 there was one increasing  $O_3$  trend and one decreasing  $O_3$  trend’. In the case where a site had a change point in 2003, but the trend remained increasing, this will be counted as one increasing trend. This has the benefit that no one category can count more than the number of available time series. In visualisation, both trends would be shown.

Trends have been collated into significance categories:  $p \leq 0.05$  (high certainty),  $0.05 < p \leq 0.10$  (medium certainty),  $0.10 < p \leq 0.33$  (low certainty) and  $p > 0.33$  (very low certainty or no evidence) based off of the guidance of ?. For the most part slopes where the p-value is  $> 0.33$  are treated as ‘no trend’ regardless of their magnitude (generally we observe that as the magnitude of the trend decreases so does its significance), though sometimes are given with a direction when required by a visualisation.

#### 2.1 Overview of Trends Clustering

Summary of trends across Europe. Angle of arrow indicates magnitude and direction, with the following orientations: vertical up  $\rightarrow +2.5 \text{ ppb yr}^{-1}$ , horizontal  $\rightarrow 0 \text{ ppb yr}^{-1}$  and vertically down  $\rightarrow -2.5 \text{ ppb yr}^{-1}$ . The few sites with absolute trends  $> 2.5 \text{ ppb yr}^{-1}$  have been clamped to  $\pm 2.5 \text{ ppb yr}^{-1}$ . Colour indicates significance and direction, darkest colours being most significant

( $p < 0.05$ ) and green the least ( $p > 0.33$ ). Blues are decreasing trends and reds are increasing trends. The upper 4 plots show  $O_3$  trends, and the lower 4 show  $NO_2$  trends. Each group of 4 groups the trends into segments of the study period. If a change point occurs in a group an arrow for both trends is shown:

Summary of trends across the United States of America. Angle of arrow indicates magnitude and direction, with the following orientations: vertical up  $+2.5 \text{ ppb yr}^{-1}$ , horizontal  $0 \text{ ppb yr}^{-1}$  and vertically down  $-2.5 \text{ ppb yr}^{-1}$ . The few sites with absolute trends  $> 2.5 \text{ ppb yr}^{-1}$  have been clamped to  $\pm 2.5 \text{ ppb yr}^{-1}$ . Colour indicates significance and direction, darkest colours being most significant ( $p < 0.05$ ) and green the least ( $p > 0.33$ ). Blues are decreasing trends and reds are increasing trends. The upper 4 plots show  $O_3$  trends, and the lower 4 show  $NO_2$  trends. Each group of 4 groups the trends into segments of the study period. If a change point occurs in a group an arrow for both trends is shown:

~~The~~ To aid analysis it was desirable to group similar time series, clusters were calculated using hierarchical clustering (HC) with dynamic time warping (DTW) used as a distance measure (?). DTW provides a distance measure between two time series that allows for some deviation in the relation of features in time and one-to-many mapping of features, as opposed to use of e.g. the euclidean distance where mapping between series is one-to-one and cannot undergo deformation (??). We adopt a similar method to ? using the R package *dtwclust* (?) for both DTW and HC steps, and used to calculate cluster validity indices (CVIs). Clustering was performed by region and time series type (e.g sites in Europe for the MDA8- $O_3$  warm season). Time series were normalised and then, to provide clusters related to the range of  $\tau = 0.5$  trends of  $O_3$  and  $NO_2$  were examined and are presented on maps in figures ?? and ??. values used in the QRs, aggregated monthly by calculating the  $\tau^{th}$  quantile per month. Metrics were not aggregated as they are all annual values. Clustering was calculated with 5 - 75 clusters, and then evaluated with 6 internal CVIs (Silhouette, Dunn, Calinski-Harabasz, COP, Davies-Bouldin and Modified Davies-Bouldin) where the latter 3 were inverted such that all can be maximised. The final selection for the number of clusters was determined as the mean of the optimal number of clusters suggested by each of the CVIs.

In Europe, between 2000 and 2004 there were 108 trends where  $O_3$  was increasing, and 41 where it was decreasing, 32 showed no trend. By 2015-2021, 123 had an increasing trend, 32 decreasing and 41 with no trend, indicating generally increasing  $O_3$  concentrations across Europe. This is somewhat mirrored by the trends in  $NO_2$ , where in 2000-2004 69 trends were increasing, 181 decreasing and 59 with no trend, changing to 29 increasing, 241 decreasing and 39 with no trend by 2015-2021. This would suggest a general picture that urban locations in Europe are still on the VOC limited portion of

### 3 Results and Discussion

215 As an initial overview, figure 1 shows the annual mean concentrations across the range of all the urban sites in this study. Additional daily mean aggregations are shown to produce day (0800 - 1900 L) and night (2000 - 0700 L) subgroups. European  $O_3$  concentrations show a steady increase over the last 20 years, with the average across all data increasing from 21.2 ppbv in 2000 to 25.9 ppbv in 2019, the  $MDA8O_3$  production isopleth, and decreasing  $NO_2$  increasing from 30.8 to 35.3 ppbv and warm season  $MDA8O_3$  increases  $O_3$  production, however, it is not quite as ubiquitous as this, as shown in ?? which discusses how  $O_3$  responded to changing  $NO_2$  following restrictions in 2020.

220 This contrasts with the USA, where in 2000-2004, there were 10 sites with increasing  $NO_2$ , and 0 sites increasing between 2005-. In the USA average  $O_3$  has increased from 27.7 to 29.7 ppbv, but  $MDA8O_3$  and warm season  $MDA8O_3$  both decreased, from 41.5 to 40.9 and 2014, but in 2015-2021 17 sites had begun increasing in  $NO_2$  again — though notably as seen from figure ?? these are different sites to those that were increasing at the beginning of the century. This increase in 49.9 to 45.1 ppbv respectively. Annual average urban  $NO_2$  is also shown to provide some context to precursor concentrations - these sites were selected under the same criteria for site type and coverage as the  $O_3$  data. Both Europe and the USA have seen reductions in  $NO_2$  is not clearly reflected by the  $O_3$  trends with 62 increasing, 23 decreasing and 22 showing no trend in 2000—2004 and 44 increasing, 34 decreasing, and 25 showing no trend in 2015-2021. A decreasing number of sites with a positive trend in  $O_3$ , suggests that the majority of decreasing  $NO_2$  trends are slowing the increase in  $O_3$ . The counts of sites at all values 19.0 in 2000 to 13.0 ppbv in 2019 in Europe and 19.7 to 9.8 ppbv in the USA. The USA did see a  $NO_2$  minima in 2020 of  $\tau$  can be found in tables ?? and ??. 9.3 ppbv, but increased to 11.8 ppbv in 2022, whereas Europe continued its decrease to 10.6 ppbv in 2022.

235 Across the entire study period, for  $O_3$  in Europe The other features that do stand out appear in the sub groups containing the warm season, but are most clear in the warm  $MDA8O_3$  series correspond to years with significant heatwaves - 2003, 2006, at sites with increasing trends the median slope was  $0.36 \text{ ppb yr}^{-1}$ , slightly lower than in the US, where they were increasing by  $0.40 \text{ ppb yr}^{-1}$ . For  $NO_2$ , the median rate of increase was the same in Europe ( $0.36 \text{ ppb yr}^{-1}$ ), but higher at  $0.71 \text{ ppb yr}^{-1}$  in the US. For decreasing trends, 2015 and 2018 in Europe and 2012 in the USA. Heat waves are historically linked to enhanced  $O_3$  concentrations, and are expected to increase in frequency (????). The continued impact of heat waves on high  $O_3$  is however, dependant on the future levels of  $O_3$  and  $NO_2$  in Europe and the US were all similar,  $O_3$  precursors and reductions and indeed this can be observed at some sites (????), but urban sites may take longer to reach levels where  $O_3$  in Europe decreasing at  $-0.40 \text{ ppb yr}^{-1}$ , ( $-0.36 \text{ ppb yr}^{-1}$ ) in the US and for  $NO_2$   $-0.39 \text{ ppb yr}^{-1}$  in Europe and  $-0.47$  in the US. This information along with the 5 th and 95 th percentile (still at - temperature sensitivity is reduced, owing to a higher initial concentrations (?).

240 While QRs were calculated across a wider range of  $\tau$  values, this discussion mostly focuses on  $\tau = 0.5$  are shown in table ??.  
0.05, 0.5 and 0.95 as the omitted values did not sufficiently expand the results presented.

#### 3.1 Change Point Assignment

Regional differences were observed in the types (QR, PQR\_1, PQR\_2) of change point assigned. Figure 2 shows that MDA8O<sub>3</sub> series in Europe were rarely described by QR (1 - 3 % of series), whereas the United States of America was described by it 30 - 62 % of the time. As the European time series extend to 2023, some of these additional change points are likely due to the COVID pandemic, and indeed examining the location of change points by year (figures S1 and S2) shows an increased number of change points in 2019 - 2021 in Europe - this is also reflected in the (fewer) change points seen in the USA. This also highlights that there are more change points focused at the beginning and end of the series generally (for both O<sub>3</sub> and NO<sub>2</sub>) which may indicate some edge effects from the selection process. However, we don't expect this to have impacted our results as the method demonstrates the ability to choose regressions with zero change points, we do not observe large swings in trends overall at these times and that there is a good amount of other change points assigned within the time period. We use this information to inform the selection of "snapshot" years presented in this manuscript, avoiding years close the end beginning or end of the series

### 3.1.1 Significance of Trends

Time series of trends in O<sub>3</sub> (left) and NO<sub>2</sub> (right) for Europe (top) and the USA (bottom). Positive trends are contained in the bar above 0 and negative trends are contained in the bar below 0. Colour indicates significance and direction, darkest colours being most significant ( $p < 0.05$ ) and green the least ( $p > 0.33$ ). Blues are decreasing trends and reds are increasing trends. In this case trends with  $p > 0.33$  have been sorted by their direction.

### 3.2 Trends in MDA8O<sub>3</sub> Across Quantiles

Figure ?? counts the trends by significance and direction per year. 2000, 2001 and 2021 onward in Europe and 2000, 2001, 2020 The number of sites with increasing and decreasing trends for a selection of quantiles (5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup>), annually, as well as the degree of certainty of the trends, in the MDA8O<sub>3</sub>, MDA8O<sub>3</sub> warm season and MDA8O<sub>3</sub> cold season data is shown in Figure 3. We use p-values to describe the certainty of each trend, as defined by ?. Generally, there are a large proportion of high certainty positive trends in the 5<sup>th</sup> and 2021 onwards for the USA were excluded from this year on year analysis, as it is not possible to capture changing trends in these regions due to the limitations placed on where change points can occur as detailed in section 2.

This reveals the majority of trends are 50<sup>th</sup> quantile in the Europe data set (range of 34 - 72 % of trends across the 20-year period). In contrast there are fewer trends of high certainty . It also reinforces the observations from section ??, where both Europe and the US see O<sub>3</sub> increasing at number of sites, with there being proportionally more sites with a decreasing trend in the US. This additionally shows that the number of high certainty decreasing trends in O<sub>3</sub> has been present since ~2010, joined by a slow decline in increasing trends.

For NO<sub>2</sub> the feature of some sites beginning to show increasing trends in the US is clear, with this pattern beginning in 2015-2016. Between 2000 and 2019 nearly all European sites moved to a trend of decreasing NO<sub>2</sub>, however, this reversed for several sites in 2020. This trend reversal can be attributed to the reduction in NO<sub>2</sub> concentrations in Europe which were a side effect of public health measures (??), and is discussed in more detail in ??.

To investigate how the trends in urban  $O_3$  are generally changing across quantiles in Europe and the USA, the distribution of slopes in each region was determined for each  $\tau$ . This section necessarily needs to discuss two sets of quantiles referring to different calculations, as such  $\tau$  is used to refer to the quantile associated with the quantile regression, and other references to quantiles are associated with in the 95<sup>th</sup>, the analysis of the distribution of slopes within a given  $\tau$ .

285 The distributions at the start (year = 2000) and end (year = 2019) of the majority of which are decreasing (between 5 - 25 % across the 20-year period were compared (Figure ??). Generally, across all values of  $\tau$ , period). The warm-season data shows a similar pattern, but with fewer high certainty increasing trends in the 5<sup>th</sup> and 95<sup>th</sup> quantile. Many more high certainty increasing trends are observed in the cold season only data, particularly in the interquartile range (IQR = 75<sup>th</sup> and 50<sup>th</sup> quantile minus 25<sup>th</sup> quantile data (range of 29 - 73 % across the 20-year period). The proportion of high certainty increasing trends in the cold

290 season appears to be increasing with time.

In the US MDA8 $O_3$  trend data, the majority of 5<sup>th</sup> quantile trends are high certainty increasing trends (35 - 53 %) and the 95<sup>th</sup> quantile is dominated by high certainty negative trends (46 - 76 %), with a mixture found in the 50<sup>th</sup> quantile) of the slope distributions in  $O_3$  was greater in 2000 compared with 2019 ( $\tau = 0.5$ , IQR = 0.87 and 0.58 ppb yr<sup>-1</sup> for 2000 and 2019 respectively) quantile. This is consistent with the findings of ? who observed increasing  $O_3$  trends across urban sites in the US

295 (1998 - 2013) at the lower end of the  $O_3$  distribution, and decreasing trends at the upper end, leading to an observed compression of the  $O_3$  range. In the warm-season only data, there is a larger contribution of decreasing trends across all quantiles. This is particularly true for Europe, where a clear reduction in the higher  $\tau$   $O_3$  slopes is observed between 2000 and 2019. For  $\tau = 0.95$ , a much larger proportion of the slopes were  $> 1.25$  ppb yr<sup>-1</sup> in 2000 than 2019. For  $\tau = 0.95$  in the USA the 95<sup>th</sup> quantile data, where the vast majority of trends are high certainty decreasing trends (55 - 84 %). However, there is a reduction in the

300 number of slopes with very negative ( $< -1.00$  ppb yr<sup>-1</sup>) slopes between notable change year-on-year in the US 95<sup>th</sup> warm-season dataset that is worth some discussion. Although decreasing trends dominate, and the number of high certainty increasing trends are low (1-11 %), we observed a step-change. Between 2000 and 2019.

Slope density for NO<sub>2</sub> (left) and O<sub>3</sub> (right) for Europe (cols 1 and 3) and the USA (cols 2 and 4) in 2000 (top) and 2019 (bottom) for  $\tau = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9$  and 0.95

305 In Europe, the bulk of NO<sub>2</sub> slopes become more negative across all  $\tau$  (Figure ??). 2008, between 1-3 % of trends are high certainty increasing trends. However, in the USA, despite after 2008, the number of sites showing a high certainty increasing trend increases rapidly from 2 to 16 by 2016, with the majority of slopes being negative in both 2000 and 2019, there is still a proportion of slopes with positive slopes in 2019, not observed in the European distribution. This is particularly true for  $0.5 \geq \tau \leq 0.9$ . The IQR of the slope distributions was comparable between 2000 these sites located in Western and

310 Central US, and the top three largest magnitude slopes ( $\tau = 0.5$ , IQR = 0.50 ppb yr<sup>-1</sup> medium-large), located in the South. In the cold season data, there is very little year-on-year change in the distribution of increasing and decreasing trends in the 5<sup>th</sup>, 50<sup>th</sup> and 2019 ( $\tau = 0.5$ , IQR = 0.40 ppb yr<sup>-1</sup> and 95<sup>th</sup>) for Europe, comparable to the reduction in spread observed in the USA in 2019 ( $\tau = 0.5$ , IQR = 0.44 ppb yr<sup>-1</sup>) compared to 2000 ( $\tau = 0.5$ , IQR = 0.57 ppb yr<sup>-1</sup>) quantile.

Annual changes in  $\tau$  slopes were used to evaluate how absolute urban  $O_3$  mixing ratios and trends changed across Europe and the USA between 2000 and 2021. While certainty and magnitude tend to follow one another, to probe the trends more specifically we define a threshold of "medium magnitude" of  $0.5 \text{ ppbv yr}^{-1}$ , and 2021. For both Europe and the USA, the cumulative sum of the annual median of the trends across all sites was calculated for each value of  $\tau$ . This was followed by the subtraction of the 2000 value to get relative change since 2000. Using this analysis, Figure ?? shows the change in median  $O_3$  and  $NO_2$  mixing ratios since 2000. The median of these results for Europe and the United States of America was then derived for each year and for each  $\tau$ . From this, Figure ?? describes how the magnitude and direction of the trend line changes annually, explore the moderate or higher certainty trends with and without this threshold. Figure S5 shows full distribution of trends in the groups discussed. A summary of the proportion increasing and decreasing trends, highlighting the proportion of moderate or higher certainty and medium or higher magnitude trends in 2004 and 2018 is presented in Figure 4.

~~In both Europe and~~ Across the Europe data set at the start of the century (2004), 85 % and 78 % of slopes showed an increasing trend in the 5<sup>th</sup> and 50<sup>th</sup> quantiles respectively (73 % and 57 % of total at moderate certainty or higher respectively), compared to 74 % showing a decreasing trend in the 95 % quantile (31 % of total at moderate certainty or higher). By 2018, the USA, median trend-derived  $O_3$  mixing ratios have generally increased between 2000 and 2021 (figure ?? number of increasing slopes increased across the three quantiles, to 90 % and 86 % for the 5<sup>th</sup> and 50<sup>th</sup> quantile (74 % and 69 % of total at moderate certainty or higher), and from 26 % to 39 % in the 95<sup>th</sup> quantile (from 7 % to 15 % of total at moderate certainty or higher). This is most clearly observed in Europe, where all percentiles demonstrated a consistent increase in median  $O_3$  across the two decades. Larger increases in  $O_3$  were observed in the lower  $\tau$  cases ( $+6.33 \text{ ppb}$ ,  $\tau = 0.05$ ) supports an increase in the number of Europe sites with increasing  $O_3$  trends at the end of the two-decadal period compared to the higher  $\tau$  cases ( $+5.08 \text{ ppb}$ ,  $\tau = 0.95$ ), indicating that the lowest ambient  $O_3$  levels are continuing to increase while higher levels show some signs of reducing their rate of increase. The picture is more mixed in the USA beginning, for which the majority of sites showed increasing trends in the low-mid quantiles during both time periods. In the 95<sup>th</sup> quantile, the majority of sites were are showing decreasing trends in both time periods, but a larger proportion of increasing trends is apparent in 2018 compared to 2004.

In comparison, 82 % and 49 % of slopes showed an increasing trend in the 5<sup>th</sup> and 50<sup>th</sup> quantiles of the US data set at the start of the century (56 % and 36 % of total at moderate certainty or higher). Similarly to Europe, the largest increases in median  $O_3$  since 2000 were observed in the lower  $\tau$  cases ( $+4.95 \text{ ppb}$ ,  $\tau = 0.05$ ). the Europe, 95<sup>th</sup> quantile trends were decreasing the in US (87 % - 68 % of total at moderate certainty or higher). However, in contrast to Europe, by 2018 fewer trends were increasing, down to 68 % in the 5<sup>th</sup> quantile (40 % of total at moderate certainty or higher), from 82 % in 2004 (56 % of total at moderate certainty or higher). In the  $\tau = 0.50$  case, median  $O_3$  increased to ca.  $+2.5 \text{ ppb}$  by 2011 before plateauing between 2011 and 2021. In contrast, in 50<sup>th</sup> quantile, more trends are decreasing (51 % - 30 % of total at moderate certainty or higher) than increasing (49 % - 36 % of total at moderate certainty or higher), although more trends are increasing than decreasing when the data is filtered for moderate certainty or higher. In the highest  $\tau$  value cases, median  $O_3$  95<sup>th</sup> quantile, there was an increase in the number of sites with increasing MDA8 $O_3$  was lower in 2021 than in 2000 ( $-0.76$  and  $-0.60 \text{ ppb}$ ,  $\tau = 0.90$  in 2018 (26 %, 12.1 % of total at moderate or higher certainty) compared to 2004 (14 %, 3.9 % of total at moderate or higher certainty), and, whilst the majority of sites were still decreasing trends, a decrease in the number of these from 87 % (68 % of total at

moderate or higher certainty) to 74 % (60 % of total at moderate or higher certainty) between 2004 and 0.95 respectively). This suggests that higher ambient levels of  $O_3$  have reduced in magnitude since 2000, compared to lower ambient levels which are continuing to increase. In both Europe and the USA, median ambient  $NO_2$  mixing ratios have reduced, with the largest reductions observed in the  $\tau = 0.95$  case (-11.49 ppb and -15.94 ppb for Europe and the USA respectively). Smaller reductions were observed in the  $\tau = 0.05$  case (-4.48 ppb and -5.78 ppb for Europe and the USA respectively).

Absolute change in trend-line-derived median  $NO_2$  and  $O_3$  mixing ratios for Europe and the USA, coloured by trend line percentile. The shaded region represents the median absolute deviation. 2018 a reduction in the number of increasing trends was observed at the 5<sup>th</sup> and 50<sup>th</sup> quantiles, but the high extremes (95<sup>th</sup> quantile) saw an increase in the number of increasing trends, coupled with a decrease in the number of decreasing trends.

Taking a closer look at the annual magnitude and direction of the trends allows for a closer inspection of how  $O_3$  and  $NO_2$  separating the beginning (2004) and end (2018) of time series MDA8 $O_3$  trends have changed between 2000 and 2021 for each  $\tau$  value (Figure ??). For Europe, all median  $O_3$  trends were positive for all  $\tau$  values across all years. For higher  $\tau$  values ( $\tau = 0.75, 0.90$  and  $0.95$ ) into warm (April - September) and cold (October - March) seasons reveals much clearer trends, and reveals a regionality in the direction of MDA8 $O_3$  trends. Generally across Europe early century (2004) data, there more increasing trends in the cold season (72 %; 37 % of total at moderate or higher certainty) compared to the warm season (25 %; 7 % of total at moderate or higher certainty) in the 95<sup>th</sup> quantile. Although there are a high number of increasing trends in the warm and cold season for the 5<sup>th</sup> quantile (88 % and 78 % respectively), there are many more increasing trends with a moderate or higher certainty and medium or greater magnitude in the warm season (15 %) compared to the cold season (4 %). This pattern is retained in the 2018 slope data, with an increase to 19 % of 5<sup>th</sup> quantile trends increasing at moderate or higher certainty and medium or higher magnitude in contrast to Europe, the median slope in  $O_3$  steadily decreased between 2000 and 2014, before plateauing between 2014 to 2021. For the lower  $\tau$  values, the slope in  $O_3$  remained positive between 2000 and 2021, but reduced by ca. 0.34 ppb yr<sup>-1</sup> between 2000 and 2014 ( $\tau = 0.05$ ). Trends in the higher  $\tau$  cases were positive at the beginning warm season. A smaller increase in the cold-season 5<sup>th</sup> quantile was observed, and a larger increase in the 50<sup>th</sup> quantile (from 7 % to 17 % of total at moderate or higher certainty and medium or higher magnitude). In summary, although we are generally seeing increasing trends across the quantiles in both 2004 and 2018, this is particularly prevalent for the lower extremes in the warmer months, and the 50<sup>th</sup> quantile cold-season data, where the trends are not only increasing, but are of higher certainty and magnitude.

In the US cold season at the start of the century, with a transition to a negative trend observed for  $\tau = 0.75$  (in 2013), and  $\tau = 0.90$  and  $0.95$  (both in 2007). In the  $\tau = 0.50$  case, median trends in  $O_3$  plateau around 0 ppb-yr92 %, 83 % and 45 % of trends are increasing in the 5<sup>th</sup>, 50

385 Trends in  $NO_2$  in Europe and the USA between 2000 and 2021 are also contrasting. In both Europe and the USA, median trends remain negative between 2000<sup>th</sup> and 2021, indicating  $NO_2$  is decreasing in both cases, as observed in Figure ??.

However, the direction of the trend is different in each region. In Europe, the slope of the trend is becoming increasingly more negative with time, particularly for the higher  $\tau$  cases (reduction of ca. 0.72 ppb-yr<sup>-1</sup> for the  $\tau = 0.95$  case). However, 95<sup>th</sup> quantiles respectively, the majority of which are of moderate or high certainty in the 5<sup>th</sup> and 50<sup>th</sup> quantile. In comparison, there

390 are fewer increasing trends in the USA the negative trends are becoming increasingly more positive (or less negative) between 2000 and 2021 the warm season for the 5<sup>th</sup> (0.29 ppb-yr<sup>-1</sup> higher,  $\tau = 0.95$ ).

Median trend line slopes (ppb-yr<sup>-1</sup>) in  $NO_2$  and  $O_3$  for Europe and the USA, coloured by trend line  $\tau$  value. The shaded region represents the median absolute deviation.

### 3.3 Spatio-temporal Distribution of Change points

395 The direction and magnitude of the change points in each location were investigated to identify locations where a change in direction of 52 %, 50<sup>th</sup> (17 %) and 95<sup>th</sup> (9 %) quantiles. In the trend (i.e. a negative to positive trend) occurs. In Europe, 21 sites observed a switch from a negative  $O_3$  trend line to a positive one in the first change point. The biggest switches occurred in southern Europe, in Spain and Greece ( $\Delta 3.28$  ppb in *gr0022a* warm season, the majority of trends are decreasing in the 95<sup>th</sup> quantile (90 %), 2004). For a larger number of sites (41), mainly in central Europe, with 63 % of all trends in this quantile

400 having a moderate or higher degree of certainty, at a medium or larger magnitude. The majority of 50<sup>th</sup> quantile trends are also decreasing (82 %; 70 % at moderate or higher certainty), but with a smaller proportion with both moderate or higher certainty and medium or higher magnitude (14 %). Comparable trends are observed in the first change point in  $O_3$  switched from positive to negative, up to a maximum swing of 4.55 ppb (*fr31013*, 2003) 2018 slopes, but with a small increase in the number of increasing trends of moderate certainty or higher in the warm season 50<sup>th</sup> (from 12 % to 16 %) and 95<sup>th</sup> (from 3 % to 10 %)

405 quantile, and in the 95<sup>th</sup> quantile cold season (from 23 % to 32 %). In contrast, a larger number of sites observed a negative to positive switch (39) in the second change point, compared to a positive to negative switch (28). Eight of the top ten biggest switches from negative to positive in the second change point occurred in Spain and Italy ( $\Delta 3.99$  ppb in *es0124a*), whereas the biggest switch from positive to negative showed no distinct spatial pattern. This suggests that generally across Europe, more trends in  $O_3$  were switching from positive to negative in the first 5-10 years of the 21st-century, whereas more trends were

410 switching from negative to positive between 2010-2021. In both decades, southern European sites showed the biggest changes in  $O_3$  trend from negative to positive. Similarly to  $O_3$ , more sites showed a positive to negative trend (84) in  $NO_2$  compared to negative to positive (28) for the first change point, with the largest change points again observed in Spain ( $\Delta 3.25$  ppb, *es1453a*). A comparable number of sites showed changes in trend direction in the second change point, (43 negative to positive, and 33 positive to negative). The positive to negative trends were generally distributed around 2010-2014, whereas the vast majority of

415 negative to positive switches occurred in 2020 (41 out of 43 sites). The cause of this switch in 2020 across Europe is discussion

in further detail in section ?? In the cold season, the number of increasing trends with a moderate or high certainty and also medium or higher magnitudes in the 95<sup>th</sup> quantile has increased from 5 % to 14 % of the trends.

European sites with a change in direction of slope ( $O_3$  ← left,  $NO_2$  → right) of the first (rows 1, 3) second (rows 2, 4) change points. The size of the slope is relative to the magnitude of the change in slope (slopes  $> 2.5 \text{ ppb yr}^{-1}$  have been clamped), coloured by the year of the change point. Negative to positive change points (rows 1, 2) are presented separately to positive to negative change points (rows 3, 4).

US sites with a change in direction of slope ( $O_3$  ← left,  $NO_2$  → right) of the first (rows 1, 3) second (rows 2, 4) change points. The size of the slope is relative to the magnitude of the change in slope (slopes  $> 2.5 \text{ ppb yr}^{-1}$  have been clamped), coloured by the year of the change point. Negative to positive change points (rows 1, 2) are presented separately to positive to negative change points (rows 3, 4).

### 3.3 Hierarchical Clustering of MDA8O<sub>3</sub> time series

In the USA, a comparable number of directional switches in  $O_3$  trend in the first change point occurred in the positive to negative direction (21), compared to negative to positive (17). The three largest magnitude trend changes from negative to positive occurred in Florida ( $\Delta 2.43 \text{ ppb}$ , 15292), with all of the ten highest magnitude cases in either Florida or California. Trends switching from positive to negative in the first change point scenario were smaller in magnitude, the biggest switch being So far, trends have only be separated by region, but it is clear when displaying the trends on a map there is more structure. Figures 5 and 6 show this for the warm season MDA8O<sub>3</sub> trends (those for MDA8O<sub>3</sub> and cold season MDA8O<sub>3</sub> are shown in figures S6 - S9). From these there appear difference in the north and south of Europe, and the coasts versus central United States. However, it is not trivial to chose sub-regions through inspection alone, so dynamic time warping coupled with hierarchical clustering (DTW, HC, described in section 2.1) was implemented to identify time series with similar features. Here, we discuss the clustering observed in California ( $\Delta 1.53 \text{ ppb}$ , 8197), but with some noticeable contributions from the central states. The majority of second change points occurred between 2010-2019, though it is important to note that due to a lack of data availability, it is not possible for this analysis to pick out change points from 2020 onward in the USA dataset. In total, 14 sites switched from positive to negative, with a maximum change of  $\Delta 2.25 \text{ ppb}$  (8169, 2019). A comparable number of sites (12) showed trends switching from negative to positive, with a maximum change of  $\Delta 2.46 \text{ ppb}$  (1959, 2019). In the  $NO_2$  dataset, there is only one site showing a switch from negative to positive in the first change point ( $\Delta 1.30 \text{ ppb}$ , 1369, 2008), compared to 13 sites showing a positive to negative switch. Of these 13 sites, the largest magnitude is in Virginia ( $\Delta 2.99 \text{ ppb}$ , 15024, 2003). In contrast, 19 sites show a switch in  $NO_2$  trend from negative to positive after the second change point, with the biggest change in California ( $\Delta 2.16 \text{ ppb}$ , 1206, 2019), with only one site showing a switch from positive to negative ( $\Delta 0.48 \text{ ppb}$ , 1369, 2013). MDA8O<sub>3</sub> data at the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> quantile. HC was performed for each of the three types of MDA8O<sub>3</sub> data set, separately for the United States of America and Europe.

Generally, The HC of the European data set indicates that most sites are well described by one cluster only, in the largest changes from negative to positive O5<sup>th</sup> and 50<sup>th</sup> quantiles across all three types of MDA8O<sub>3</sub> trends in the USA were in California data aggregation. However, a few additional clusters are introduced in the 5<sup>th</sup> and Florida in the first half of the 20

450 year-period, with the reversal of this trend seen across California and Florida in the second change point. 8/14 50<sup>th</sup> quantile cold-season data, as well as the 50<sup>th</sup> quantile full data set and warm-season only data set. In the cold-season, these clusters don't appear to have any particular regionality in the 5<sup>th</sup> quantile. However, in the 50<sup>th</sup> quantile two new clusters are introduced, localised around South Central and East Europe (clusters 3 and 4/14 of the sites showing a positive to negative transition in ). In the second change point were located in California and Florida respectively. This is important, since a previous study highlighted that 4th highest daily maximum 8-hour O<sub>3</sub> (4MDA8) values between 2010-2014 were particularly high in California (??). 95<sup>th</sup> quantile (high extreme) data, more regional clustering is observed in the full MDA8O<sub>3</sub> data, and the warm-season only data. This regional clustering is more clearly observed in the MDA8O<sub>3</sub> data, with clusters location in North Eastern Europe (cluster 1); across the continent from South and Central Western France to East Europe (cluster 2); some selected sites from across this same area, but also including sites in Spain and the UK (cluster 3); Northern France (cluster 4), the Far East (cluster 5), and North East (cluster 6). In the warm season, the majority of clustered sites are located across Europe, with some exceptions in the North East (cluster 2), North West (cluster 3), and Spain (cluster 4).

To supplement this analysis, For all clusters in the 95<sup>th</sup> quantile MDA8O<sub>3</sub> European data, an enhancement in the mean AVGMDA8, 4MDA8 was calculated using the mixing ratio data for each site and slicing the fourth highest value on an annual basis. 4MDA8 is an important metric used by the US for determining compliance with the National Ambient Air Quality Standards for Ozone (??). To allow for comparison with previous studies, only the 6-month warm season (April-September) was used to calculate 4MDA8 values, which is a reasonable approach in the USA and Europe since the highest, NDGT70, SOMO35, 3MMDA1 and 6MMDA1 metrics is observed in 2003, when an extensive European heatwave occurred. Enhancement in other significant heatwave years are also visible in some data clusters in 2006 (clusters 1,2 and 5); 2015 (clusters 1 and 2); and 2018 (clusters 1,2, and 6). This relates to the features seen in the (warm season) MDA8O<sub>3</sub> averages in figure 1, the 2003 heat wave is notably larger in the average presented there. Later enhancements may be less clear due to those heat waves being less wide spread, or reductions in O<sub>3</sub> values are likely to occur in the warmest months. temperature sensitivities in those locations. This is indicated by the heat waves not being present in all clusters as the decades progress, but from this study it is not possible to say definitively whether one or both effects are causing this.

Figure ?? shows how the 4MDA8 value varies between 2000, 2007, 2014 and 2021 in The HC of the United States of America data set has only one large cluster for the USA. Across the US, 33 sites had a 4MDA8 >= 85 in 2000, compared to 19 in 2007 and 6 full MDA8O<sub>3</sub> data set, cold-season, and warm-season only data in the 5<sup>th</sup> quantile. However, some clear regional clusters appear in the 50<sup>th</sup> and 95<sup>th</sup> quantile data sets across all three data types. The main secondary cluster sites are located in Florida, also including some sites along the Gulf of Mexico and the South in the cold-season (cluster 2, or cluster 3 in the MDA8O<sub>3</sub> full 95<sup>th</sup> quantile data). Some clustering around the Intermountain West and the South West is observed in the full MDA8O<sub>3</sub> data set in the high extremes (cluster 2, 95<sup>th</sup> quantile). The East Coast generally falls into the largest cluster which doesn't appear to have a strong regionality, except for the in 2014. In 2016 there was an up-tick to 10 sites with 4MDA8 values at this level. Of these 10 sites, all are located full MDA8O<sub>3</sub> data 95<sup>th</sup> quantile, where there is an independent North East coast cluster (cluster 5). Also in this data type, and in the 95<sup>th</sup> quantile data, there is a cluster located in South Central USA (cluster 1).

485 For the US 95<sup>th</sup> MDA8O<sub>3</sub> data, clusters behave much less uniformly than they do in Europe. Whilst all MDA8O<sub>3</sub> metrics are generally decreasing across all clusters (clusters 1,3,4,5 and 6). However, many metrics are either clearly elevated, increasing across the 20-year period, or both, for cluster 7, located in and around Los Angeles, in Southern California. This is also observed to a lesser extent in cluster 2, located more generally in the West and in California, compared to 20 the Intermountain West region of the United States. This affect can be seen by looking at the trend data. For cluster 7, increasing trends are observed in the full MDA8O<sub>3</sub> data set. However, the MDA8O<sub>3</sub> trends in cluster 7 are generally of low certainty. In cluster 2, mean MDA8O<sub>3</sub> is elevated compared to other clusters, excluding cluster 7, and the slope in MDA8O<sub>3</sub> trends is generally quite small. However, these trends typically have a high degree of certainty ( $p < 0.05$ ), which are typically negative in the first half of the vicennial, and positive in the second.

### 3.4 Spatio-temporal distribution of O<sub>3</sub> metrics, relevant to health and exposure

495 As the health/33 Californian sites in 2000 and 18 exposure O<sub>3</sub> metrics are annually aggregated, the DTW/19 in 2007. All 11 sites in 2021 with HC method was not directly applied to these data sets, but observations from inspection of figures S10 - S21 is presented here.

The NDGT70 and 4MDA8 metrics are generally used to investigate variations in the highest values of O<sub>3</sub>. High values of these metrics are typically an indicator of episodes of high photochemical O<sub>3</sub> production. Here, we explore the frequency of high values of these metrics over the 20-year time period, separately in Europe and the United States of America.

Across Europe, patterns in high 4MDA8 ( $\geq 85$  ppb) were located in California. This highlights that although fewer Californian sites are exceeding 4MDA8 ppb) and NDGT70 ( $\geq 85$  ppb) with time over the 2-decade period, the Californian 4MDA8 "hotspot" identified in ? is still present. The more general picture across the USA is that 4MDA8 values 25 days) frequencies and site locations are similar. In general, the only sites that consistently exceed these metrics are coming down with time, particularly strongly in the central and Eastern portion of the USA but less successfully in the west.

4MDA8 values in ppb for sites in the United States in 2000, 2007, 2014 and 2021

Southern Europe, particularly Greece and Northern Italy, were also highlighted in ? as being 4MDA8 hotspots. Figure ?? shows how the 4MDA8 values are changing between 2000, 2007, 2014 and 2021 across Europe. In 2000, there were 6 sites in which located in Southern Europe. Exceptions to this can be seen in 2003, 2006 and 2018 for both metrics, known heatwave years, when a large number of high 4MDA8  $\geq 85$  ppb. This slightly increased to 7 sites in 2007 before decreasing to 4 sites in 2014. By 2021, no sites had 4MDA8 values and NDGT70 values occurs across the continent. However, the number of sites exceeding 85 ppb. Since the measurement sites in this dataset are not evenly distributed across the USA and Europe, it is difficult to attribute high 4MDA8 values to particular states or countries, However, it was noticeable that a large proportion of the sites were the prescribed limits during these heatwave events generally reduces with time, with 89, 45, 27, and 15 exceedances in 4MDA8  $\geq 85$  ppb were in Italy and Greece. In 2000 3/6 of high 4MDA8 sites were located in Italy, and 1/6 in Greece. In 2007, 3/7 were located in Italy and 1/7 in Greece; and in 2014 3/4 sites were located in Italy. Consistent with the ? study, the southern European 68, 19, 6, and 8 exceedances in NDGT70 for 2003, 2006, 2015, and 2018 respectively.

For the US metrics, a high number of sites across the country are found to have high 4MDA8 hotspots—are apparent throughout the and NDGT70 values in the early years of the 20-year period.

520 4MDA8 values in ppb for sites in Europe in 2000, 2007, 2014 and 2021

### 3.4.1 2020 In Europe

As noted in section ??, in 2020 in Europe several sites switched from a long term decrease in  $\text{NO}_2$  to increasing  $\text{NO}_2$  again. The public health measures enacted due to the COVID-19 pandemic had the side effect of decreasing  $\text{NO}_2$  concentrations due to reduced traffic emissions. The removal of these measures led to an increase in  $\text{NO}_2$  mixing ratios as normal activities resumed, resulting in a positive trend in  $\text{NO}_2$  concentrations up from their lowest during the restrictions. There are 42 sites that had a negative trend in 2019 and a non-negative trend in 2020. The 2019 trends ranged from  $-1.47$  to  $-0.04$   $\text{ppb yr}^{-1}$  and the 2020 trends ranged from  $0.00$  to  $3.74$   $\text{ppb yr}^{-1}$ . There were two sites that already had non-negative trends (*fr25039* decade, from 2000 up until 2007. From 2008, there are few sites with exceedances, generally located in the West, and occasionally around the Intermountain West region of the US. Due to a lack of data availability in 2012, a reported heatwave year, it is not possible for us to comment here on whether this event led to elevated 4MDA8 and *pl0048a*) by 2020, but neither were strong or significant ( $p > 0.33$ ). NDGT70 frequencies.

The effect of this on  $\text{O}_3$  concentrations at the sites is varied. 21 of the 42 sites also measured  $\text{O}_3$ , though only 3 (*es1038a*, *fr33120* and *gr0031a*) also had 2020 as one of the change points for the PQR. To compare the 2020 onwards trends, new PQRs for each site were selected by requiring the second change point to occur in 2020 (SOMO35, AVGMDA8, 3MMDA1, and then selecting where the first occurred on a time series by time series basis by minimising the AIC as before. Furthermore, only sites where the resulting slope for the 2020 onwards  $\text{O}_3$  trend had p values  $< 0.33$  were selected. One final site (*gr0031a*) was removed as although it meets the data coverage criteria for the main analysis, its missing data is focused around 2019 and 2020 which has a strong influence on the trends derived. This led to 15 sites remaining for this case study. The resulting alternative trends for these sites are shown in figure ??.

Additionally, as the focus is on sites with both  $\text{NO}_2$  and  $\text{O}_3$ ,  $\text{O}_x$  is displayed. If a site is under a VOC limited regime, the trend in  $\text{O}_x$  provides an indication whether any changes in  $\text{O}_3$  are only due to changes in  $\text{NO}_2$ , or other factors. Whether a change in  $\text{NO}_2$  concentrations results in an increase or decrease in 6MMDA1  $\text{O}_3$  concentrations depends on the if a site is sensitive to locally produced  $\text{O}_3$  metrics are used to describe the general  $\text{O}_3$  (which an urban site exposure, not just high  $\text{O}_3$  events, important for assessing  $\text{O}_3$  health burdens, rather than policy compliance. Across Europe, and across the 20-year period, values of SOMO35 are generally fairly low across the board (typically  $< 4000$  ppb day). SOMO35 values of  $> 4000$  ppb day are more widespread across Europe, excluding Northern Europe, in general (2003; and at a few sites across Southern Europe over the years, typically in Spain, Greece, Switzerland and Italy. This is reflected in the AVGMDA8 metric, where the lowest values are typically found in the UK, Scandinavia, and North of mainland Europe (typically  $< 40$  ppb). Slightly higher values of AVGMDA8 are observed in Central and Eastern Europe (41 - 45 ppb) in the early years (2000 - 2014). However, an uptick in AVGMDA8 is observed in Eastern Europe in multiple years from 2015 onward, with many sites observing levels of 46-50 ppb, would be expected to be), and what photochemical regime ( $\text{NO}_x$  or VOC limited) the site is under. For example, in a VOC limited regime, if  $\text{NO}_2$  increases,  $\text{O}_3$  decreases but  $\text{O}_x$  remains the same,

or even higher in 2018 (51 - 55 ppb). Few sites exceed 56 ppb, and those that do are located in Southern Europe. The exception to this is once more in 2003, when AVGMDA8 were typically > 56 ppb across Europe, reaching 72 ppb in the alpine region of Central Europe. Both the 3MMDA1 and 6MMDA1 metrics are low across Europe for the 20-year time period, with maximum values of 52 ppb and 44 ppb respectively observed in Northern Italy in 2003. Despite lower values across the continent, the change in  $O_3$  can be attributed to highest of these (31 - 40 ppb) is consistent observed in Southern Europe and the change in  $NO_2$ . If  $NO_2$  and  $O_3$  increase,  $O_x$  will also increase, indicating a  $NO_x$  limited regime. 14 of the sites saw increasing  $O_3$  central alpine region, consistent with previous studies (?).

For the US data, these general  $O_3$  with increasing  $NO_2$  between 2020 and 2023 exposure metrics are generally higher than the observed values for Europe. At the start of the 20-year period, the lowest values of SOMO35 were observed in the Central and Eastern US (typically 1000 - this is expected in  $NO_x$  limited regimes. Due to the need for both  $O_3$  4000 ppb day), with higher values observed in the South (typically 4000 - 6000 ppb day), and  $NO_2$  to be measured at both sites 13 of the sites were urban background and 2 were urban traffic, as  $O_3$  is less frequently measured at traffic sites. Urban background sites are more likely to be under  $NO_x$  limited regimes as they are, by design, situated further from major roads, which are still a significant source of  $NO_x$ . This does not necessarily mean that urban areas in Europe are completely  $NO_x$  limited and will depend on the specific location as to whether an urban traffic or urban background site is more representative of the highest values found on the West coast (up to 11981 ppb day). This is consistent with previous studies, which have observed that Southern California is a location. At one urban traffic site (es1529a) has seen decreasing  $O_3$  with increasing  $NO_2$  over the same period. This is more indicative of a VOC limited regime, and is more similar to the overall trends seen across Europe in the 21<sup>st</sup> century, with more increasing  $O_3$  trends in 2015-2021 than in 2000-2004.

The sites identified by this method do not appear to have resumed their pre-2020 trends, though this should be verified when sufficient data has been collected post-2020 where change points will have more freedom to occur. The pattern of sites switching to an increasing trend in  $NO_2$  will have varied effects on urban  $O_3$  in Europe if they continue, though as background sites are designed to be more representative of residential areas, which make up a large portion of urban areas then it could be expected to see urban  $O_3$  continue to increase. hotspot for  $O_3$  pollution (??). Over the 20-year period, SOMO35 levels in Central, Eastern, and Southern USA generally reduce. However, values in the West and Intermountain West are found to increase or stay high, with SOMO35 values of 10973 ppb day observed in one site in 2020. Across the vicennial, all sites exceeding 7000 ppb day are located in the West, with a general decrease in the number of sites observing this limit over the 20 years. Similarly to the European data, the AVGMDA8 metric shows a similar pattern to the SOMO35 data, and by 2014 there is a clear divide between Central, Eastern and Southern US, where the lowest values are observed (up to 50 ppb), and the West and Southwest, including the Intermountain West (51 - 80 ppb). This narrative is also reflected in the 3MMDA1 and 6MMDA1 metrics.

Sites in Europe that had decreasing  $NO_2$  in 2019 but increasing  $NO_2$  from 2020 onwards. Monthly averaged  $O_3$  (red),  $NO_2$  (orange) and  $O_x$  (green) anomaly are shown.  $O_3$  and  $O_x$  trends differ from those presented elsewhere in the manuscript as the piecewise quantile regression where the second change point has been restricted to be in 2020.

### 3.5 Coupling MDA8O<sub>3</sub> trends with 6MMDA1 O<sub>3</sub> exposure metrics

#### 3.5.1 Recent NO<sub>2</sub> in the USA

In the final years of the USA NO<sub>2</sub> dataset, an up-tick the number of positive NO<sub>2</sub> trends can be seen. From 2015 onwards there is an increasing number of sites with this trend reaching 17 sites trends (16 high, 1 medium certainty) in 2021, up from 0. To investigate how trends are developing in different clustered regions across Europe and the US variations in the annual exposure metric (6MMDA1) are compared to the slopes and significance of trends in MDA8O<sub>3</sub> in 2014. Anomaly time series and trends of can be viewed in figure ???. Trends have reversed from -0.98 2004 and 2018. Here, we only look at sites that have been assigned a cluster (where the number of sites > 3), using the HC clusters analysis described in section 2 to explore whether regional clusters were showing similar trends. Figures 9 - -0.03 ppb yr<sup>-1</sup> to 0.1 11 show the MDA8O<sub>3</sub> trend against 6MMDA1, coloured by the clusters shown in section 3.3. Figures S22 - 1.78 ppb yr<sup>-1</sup> 1000 1000 -1. Unfortunately, only 4 of the 17 sites have concurrent measurements of O<sub>3</sub> so it is difficult to investigate this changing trend in NO<sub>2</sub> from the time series analysed in this study. Only one site (8169) O S24 show these same scatters coloured by significance.

In the European MDA8O<sub>3</sub> trend has its original second change point occurring alongside its NO<sub>2</sub> increasing change point, so the same method has been applied as in ?? where the second change point has been forced to occur in the same year as the NO<sub>2</sub> change point. Doing this reveals that sites 8169, 14386 data and in the 5<sup>th</sup> quantile where sites are behaving more uniformly (one bulk cluster), we do not see a clear reduction in high certainty increasing slopes between 2004 and 15197 have reasonably significant ( $p = 0.01, 0.15, 0.002$  respectively) O<sub>3</sub> trends after change point 2, when this change point is the same as that in NO<sub>2</sub> 2018, but we do see fewer high certainty decreasing slopes in 2018. Across the cluster, which includes a broad range of sites across Europe, we observe similar maximum values of annual 6MMDA1 in the two years (c.a. 8169 and 14386 have decreasing O<sub>3</sub> trends 68 ppbv), but fewer 6MMDA1 values in the lower range in 2018 (-1.70 and -0.55 ppb yr<sup>-1</sup>) with increasing NO<sub>2</sub> and 15197 saw increasing O<sub>3</sub> at 1.15 ppb yr<sup>-1</sup>. These data suggest that the O<sub>3</sub> concentrations will be sensitive to this increase in NO<sub>2</sub>, but without the co-located O<sub>20</sub> - 40 ppbv). For the 50<sup>th</sup> quantile data, where there are more clusters, observe that a cluster located across Southern mainland Europe (cluster 2), includes several sites with the higher 6MMDA1 values (> 60 ppbv), in both 2004 and 2018. However, the certainty and direction of the trends in MDA8O<sub>3</sub> is mixed. A smaller cluster of sites located in Spain observed lower values of 6MMDA1 in 2004 (20 - 40 ppbv), which increases to 40 - 50 ppbv by 2018. These sites also have high certainty increasing MDA8O<sub>3</sub> trends in both 2004 and 2018. More broadly across all clusters, we again observe fewer high certainty decreasing slopes in 2018 than 2004. In the 95<sup>th</sup> quantile, where the clearest regional clustering can be observed, generally higher 6MMDA1 values are seen in 2018 than 2004. There is no notable grouping of the clusters their 6MMDA1 values in either 2004 and 2018, but there is generally a broader spread of values in 2004 compared to 2018. The location of sites with high certainty increasing or decreasing values is also mixed, showing no clear regionality. When we compare 6MMDA1 values to the 95 999 1000th quantile warm-season only MDA8O<sub>3</sub> measurement it does not provide definitive information on the direction on any changes. Indeed ? note that reductions in NO<sub>x</sub> suggested that the 99.9<sup>th</sup> quantile O trends, we again observed a compression of the range of 6MMDA1 values, at the higher mixing ratio end

(c.a. 25 - 70 ppbv in 2004, vs 40 - 70 ppbv in 2018). We also observe that clusters located in Northern Europe have the smallest 6MMDA1 values in both years, and trends are generally increasing but with low certainty.

Average 6MMDA1 values across the USA are generally higher than those reported across Europe. Some clearer regional clustering can be observed in the 50<sup>th</sup> quantile MDA8O<sub>3</sub> concentrations were also reduced between 1994 United States data, with a cluster located in Florida (cluster 2) generally showing lower 6MMDA1 values (40 - 60 ppbv) in both 2004 and 2010. While this study limits itself to 2018. Additionally, these sites are generally high certainty decreasing trends, and this is reflected in both the cold-season and warm-season only MDA8O<sub>3</sub> trends. In contrast, sites located in the West show the highest 6MMDA1 mixing ratios in both 2004 and 2018 (c.a. 70 - 90 ppbv) (cluster 7), and these are typically high certainty increasing trends in both years. These sites also have the highest 6MMDA1 values in the 95<sup>th</sup> quantile (cluster 7), but increasing trends in this quantile are generally of low certainty. More interestingly, the 95<sup>th</sup> quantile, there is some indication in figure ?? that there has been a small increase in high O<sub>3</sub> levels during the time increased NO<sub>10001000</sub> cluster located in the West including the Intermountain West region of the USA (cluster 2), typically observes 6MMDA1 values of 55 - 70 ppbv in both 2004 and 2018, but there is a upward shift in the slope magnitude of this cluster from approximately -1 - 0.5 ppb per year in 2004, to -0.5 - 1.25 ppbv per year. In addition, most of these trends shift from being of low certainty in 2004, to high certainty increasing trends in 2018. These trends are also reflected in the warm-season MDA8O<sub>3</sub> trends. Over the 21st century, the Intermountain West has experienced both increased wildfires, and hotter springs and summers, which can enhance photochemical activity forming O<sub>100010003</sub> is observed.

Sites in the USA that had increasing NO<sub>2</sub> since 2015. Monthly averaged O<sub>3</sub> (red), NO<sub>2</sub> (orange) and O<sub>x</sub> (green) anomaly are shown. O<sub>3</sub> and O<sub>x</sub> trends differ from those presented elsewhere in the manuscript as the piecewise quantile regression where the second change point has been restricted to match that of the NO<sub>2</sub> at each site. from regional anthropogenic emissions (????). The higher magnitude and high certainty increasing slopes in MDA8O<sub>3</sub> may be linked to increases in one or both of these factors.

## 4 Conclusions

Trends in urban O<sub>3</sub> and NO<sub>2</sub> MDA8O<sub>3</sub> were calculated from de-seasoned monitoring site data across both Europe and the USA. A Piecewise quantile regression analysis was utilised to assess long-term trends, with the method extended to a piecewise approach. This allowed for some freedom to capture changes in a complex dataset, whilst being able to describe the trends with a small number of coefficients. Break points. Piecewise trends were limited to two change points per time series to keep the focus on long-term trends and the largest changes. One limitation to this approach is that time series with 3 or more large changes will not be well described by this method—however, after inspection of the regressions and comparison with a LOESS, two points were found to capture the structure well. The trends constructed from the piecewise regressions were evaluated by their statistical significance, and those poorly described using this methodology ranked lower in significance and were subsequently investigated less in the analysis. Piecewise segments were required to have  $\geq 2$  years of data, which limited

the ability to define the most recent changes in a given time series, and such is why only Europe was analysed for the effects of the COVID-19 restrictions.

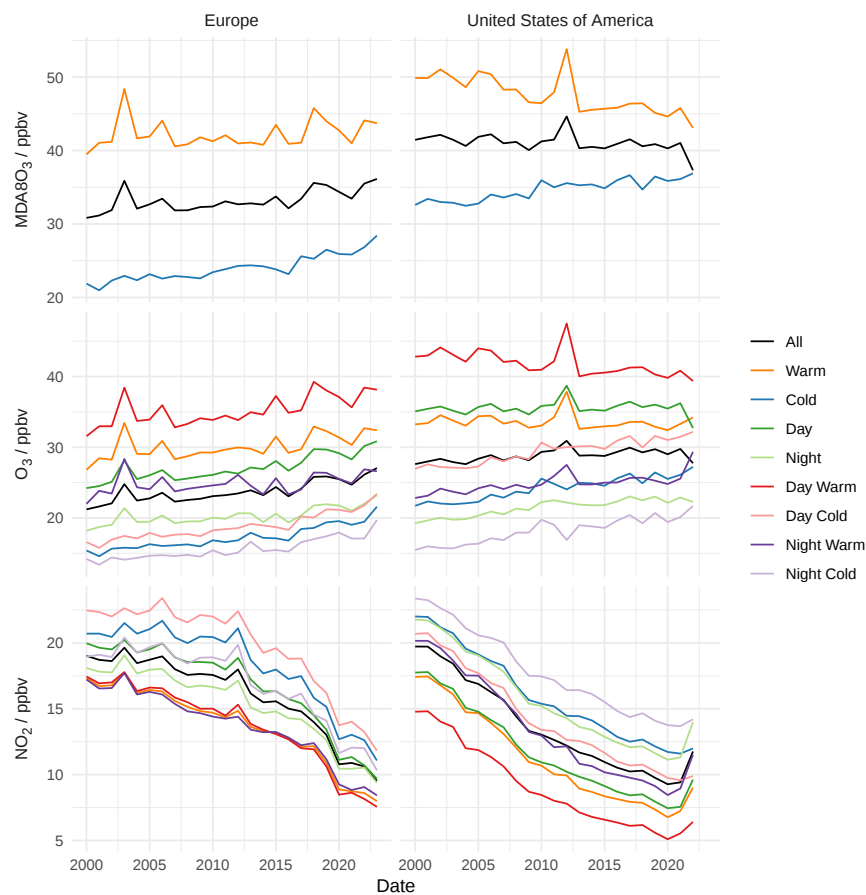
In Europe, more sites were found to have an increasing  $O_3$  trend between 2015-2021, compared to 2000-2004, with  $NO_2$  trends showing the reverse effect (fewer sites increasing, and more decreasing in 2015-2021 compared to 2000-2004). This broadly suggests that a VOC-limited relationship for  $O_3$  formation could be common across urban sites in Europe. In the USA, the reverse is true, with a reduction in the number of sites. Generally more sites in the United States of America were described by a trend line with no change points. In comparison, sites in Europe almost exclusively described by trends with increasing  $O_3$  trends in 2015-2021 compared with the 2000-2004 period, but with more sites with increasing  $NO_2$  in 2015-2021 compared with 2000-2004. The majority of  $O_3$  trends are positive and of high certainty ( $p \leq 0.05$ ) in Europe across the 20-year period. This is also broadly true for the USA, but the proportion of sites showing this trend reduces with time, alongside an increase in the number of sites with high-certainty negative trends in  $O_3$  between 2010-2021. Previous studies have attributed positive urban  $O_3$  trends in the USA to increasing winter  $O_3$  concentrations (?). However, differences in seasonality would not be observed in this analysis as the trends are derived from de-seasoned data. From the trends described here, the median ( $\pm$  median absolute deviation) change in  $O_3$  over the last two decades were calculated as  $5.0 \pm 4.6$  ppb in Europe and  $2.4 \pm 4.9$  ppb in the USA, with corresponding changes in  $NO_2$  as  $-6.8 \pm 3.6$  and  $-8.4$  ppb-1 or 2.

Given the nature of its long lifetime and the complexity of its chemistry, when describing trends in  $O_3$  By varying the selected quantile, quantile regression allows trends in different areas of the MDA8 $O_3$  it is important to look beyond the 50th percentile. A study of European ozone trends between 1995-2014 found that whilst the 50th percentile data did not show significant trends, significant trends were observed in distribution to be quantified. The analysis presented here focuses on the 0.05, 0.50 and 95<sup>th</sup> quantiles, relating to low, medium and high values of MDA8 $O_3$ . Additionally these trends were calculated for warm season and cold season subgroups. Across the 22 year period, there are many sites in Europe with high certainty increasing trends in the 5<sup>th</sup> and 1<sup>st</sup> and 50<sup>th</sup> quantiles, but the majority of medium-high certainty 95<sup>th</sup> percentiles (1<sup>st</sup> quantile trends in Europe were found to be decreasing. Separating the trends by season shows that the majority of these decreasing trends are found in the warm season in all quantiles and very few trends were increasing at  $\tau$  values) (?). An exploration of the  $\tau = 0.95$  with  $p < 0.33$ . In the USA the main time where increasing trends are observed is in the cold season, in all quantiles but to a lesser extent at  $\tau$  values in this study revealed smaller reductions in trend-derived change in  $O_3$  mixing ratios in 2021 since 2000 for the lower  $\tau = 0.95$ . Overall  $\tau$  values compared to higher values in Europe. This, coupled with the fact that generally  $O_3$  trends across Europe are positive, suggests that the lowest ambient  $O_3$  mixing ratios are increasing more rapidly than the highest values. This is consistent with the findings of ?, who found that although peak  $O_3$  concentrations were coming down across urban and suburban sites between 1995-2014 in Europe, an increase in the lower  $O_3$  concentrations was observed in more than 75 % of sites.

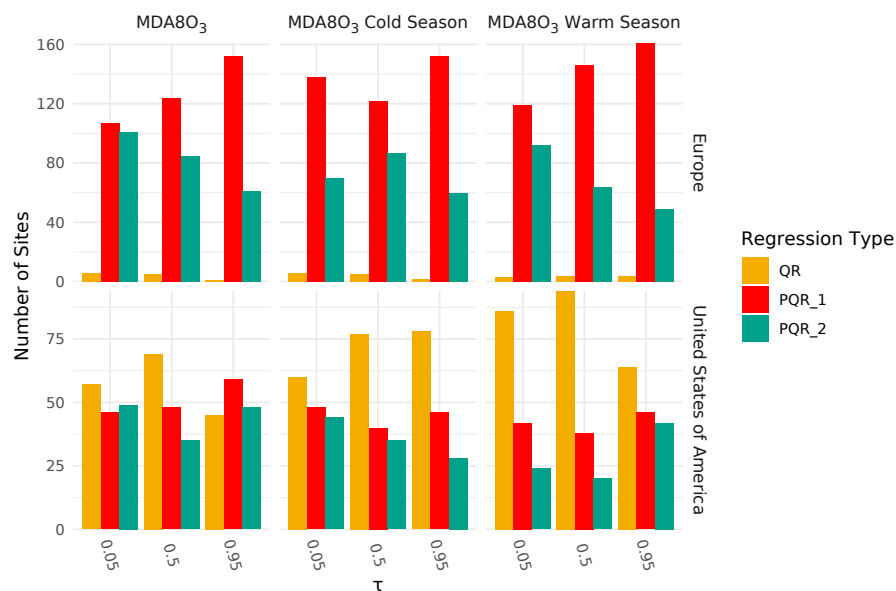
An assessment of the spatio-temporal distribution of the first and second change points revealed more sites with a directional switch in  $O_3$  trend across Europe were in  $\tau = 0.95$  trends are dominated by sites decreasing, driving by the warm season trends, though a small number of these trends at higher quantiles do appear to be increasing in the Intermountain region.

685 A primary challenge of interpreting monitoring station data is how to identify sites showing similar behaviour, without  
grouping based on pre-existing geographic regions. Hierarchical clustering with dynamic time warping was shown to be  
effective in grouping similar time series and returned differing clusters when time series were aggregated at different quantiles  
which agreed with those observed qualitatively when mapping calculated trends. Clusters tended to be more homogeneous at  
lower values of  $\tau$ , with differing subregional behaviour becoming more apparent at the  $\tau = 0.95$ . In the high extreme trends,  
690 a cluster located in California was showing increasing trends across the 22 year period, but these were typically of lower  
certainty. Higher certainty increasing trends were observed in a cluster located in the Intermountain West in the latter half of  
the century, which were of lower magnitude.  $O_3$  exposure metrics, SOMO35, AVGMDA8, 3MMDA1, and 6MMDA1 were  
typically higher in the positive to negative direction in the first change point, but with more sites with a negative to positive  
second change point United States of America than Europe, and the European sites that where  $AVGMDA8 > 56$  ppbv were  
695 located in Southern and the central alpine region of the continent. In the USA dataset, the biggest switches from negative to  
positive in the first change point occurred in California and Florida, but a reversal of this trend from positive to negative  
was also observed across these states in the second change point. To supplement this, a calculation of the 4th highest daily  
maximum 8-hour running mean for  $O_3$  (4MDA8) revealed high 4MDA8 hotspots in parts of southern Europe, but particularly  
in California, consistent with previous findings of ?. Furthermore, the USA has seen significant increasing trends in  $NO_2$  since  
700 2015 at several sites. However, there are limited sites that have been used in this study with co-located  $NO_2$  and  $O_3$ , and the  
 $O_3$  response is not consistent. This is not unexpected due to the wide range of conditions that an 'urban' site can be found  
in (e.g. variations in local VOC concentrations), but it should be expected that in some locations  $O_3$  sites located in the West  
observed  $SOMO35 > 7000$  ppbv day more consistently over the 22 year period, though the number of sites exceeding this limit  
decreased in frequency with time. More interestingly, although 6MMDA1 values in the Intermountain West were similar in  
705 2004 and 2018, the  $MDA8O_3$  trends shifted from low certainty in 2004 to high certainty increasing trends in 2018. This was  
also observed in the warm-season trends, and may be linked to an observed increase in wildfire events in the area over the 22  
year period, or warmer summers and springs as reported in previous studies (????). With the risk of increasing occurrences of  
warm summers, wildfires, and heatwaves, it is crucial that we continue to understand trends in  $MDA8O_3$ , will be sensitive to  
this increase in  $NO_2$ , especially if it persists.

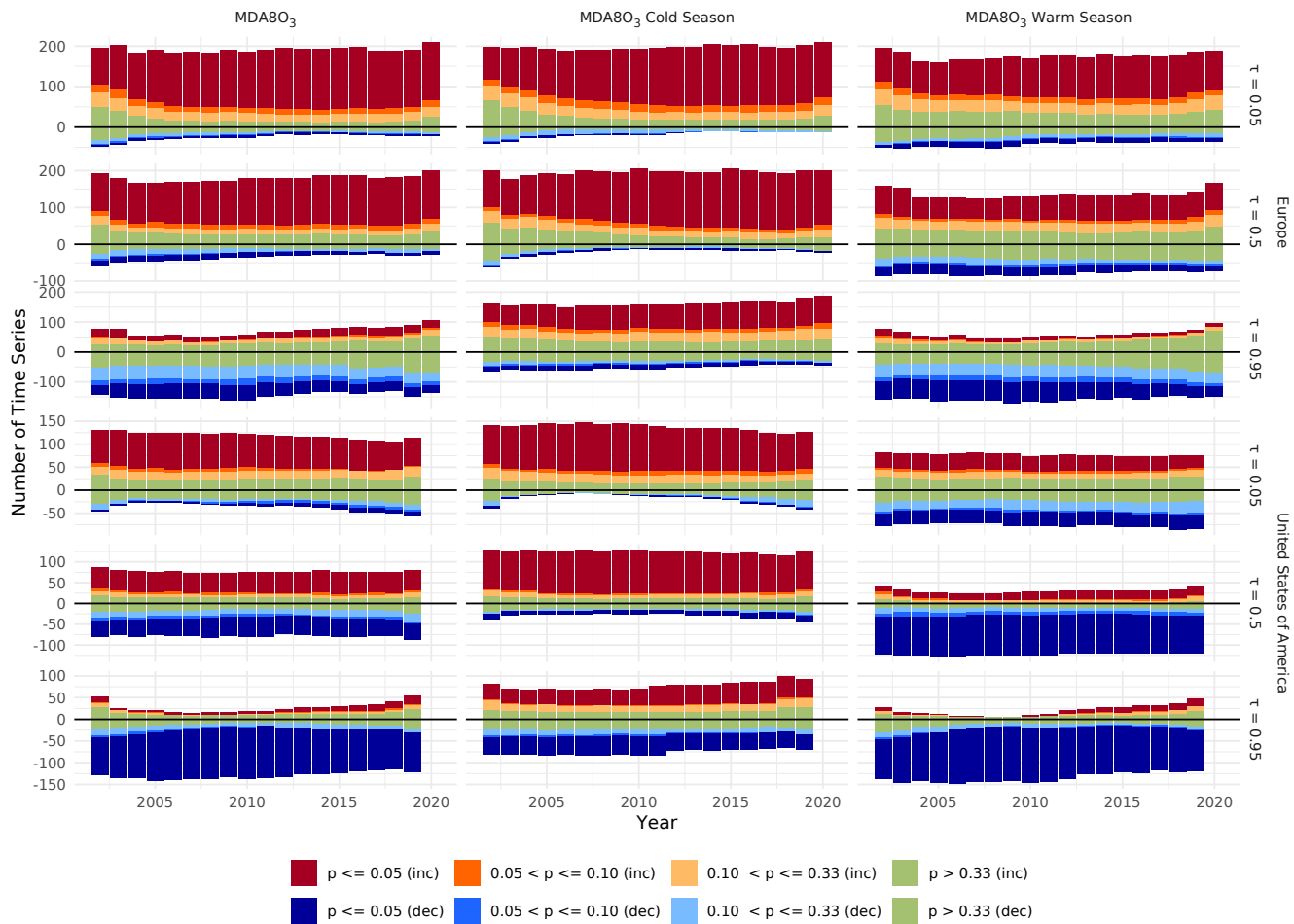
710 The impact of COVID-19 on ambient mixing ratios of  $NO_2$  is well studied, with one previous study estimating that  $NO_2$   
concentrations across urban background sites in Europe were 32 % lower than expected in 2020 (?). The impact of this  
COVID-19 effect was observed in the trend analysis of Europe described in this study, with the vast majority of directionally  
switching second change points in  $NO_2$  occurring in 2020 in the negative to positive direction. This can be attributed to lower  
ambient  $NO_2$  mixing ratios in 2020, leading to an increasing trend in the subsequent years as business-as-usual conditions resumed.  
715 These increasing  $NO_2$  concentrations have in some cases been accompanied by increasing  $O_3/O_x$  and should continue to be  
studied, since in many cases the prior downward trend in  $NO_2$  has not resumed, as this study has identified increasing trends  
in locations in the USA (the West and Intermountain West), and Europe (Southern and Central alpine region), that already  
experience hot and dry weather conditions, that may exacerbate human risk to  $O_3$  exposure.



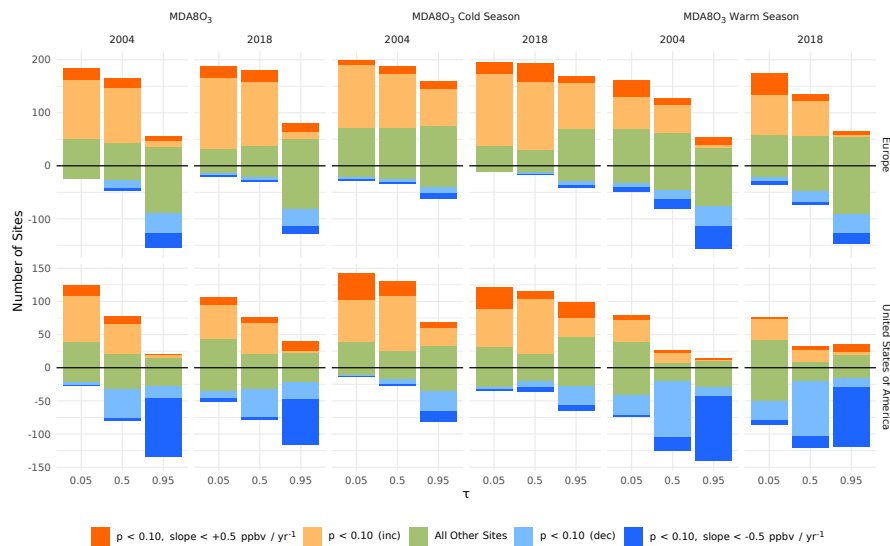
**Figure 1.** Time series of annual average MDA8O<sub>3</sub>, O<sub>3</sub> and NO<sub>2</sub> for urban sites in Europe and the USA. Day and night subgroups correspond to the hours of 0800 L - 1900 L and 2000 - 0700 L respectively, warm and cold refer to the months April - September and October - March respectively.



**Figure 2.** Number of sites with no change point (QR, gold), one change point (PQR\_1, red) and two change points (PQR\_2, green) in the Daily Mean, MDA8O<sub>3</sub> and MDA8O<sub>3</sub> warm season series, separated into Europe and United States of America regions.

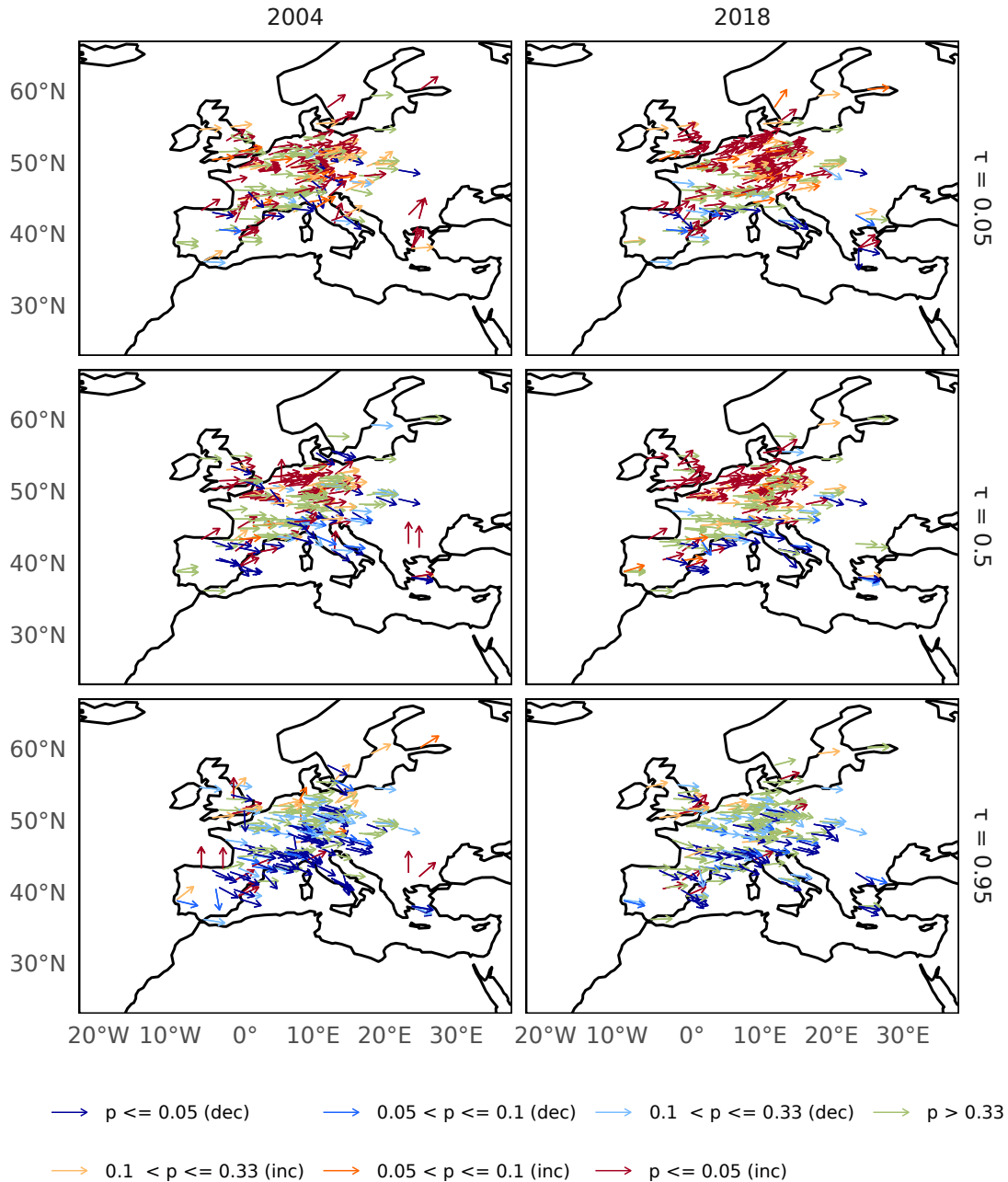


**Figure 3.** Time series of trends in MDA8O<sub>3</sub> for (left) annual, (middle) warm season, and (right) cold season, for Europe (top) and the USA (bottom). Within each region the panels are grouped by  $\tau = 0.05$  (top), 0.5 (middle), 0.95 (bottom)



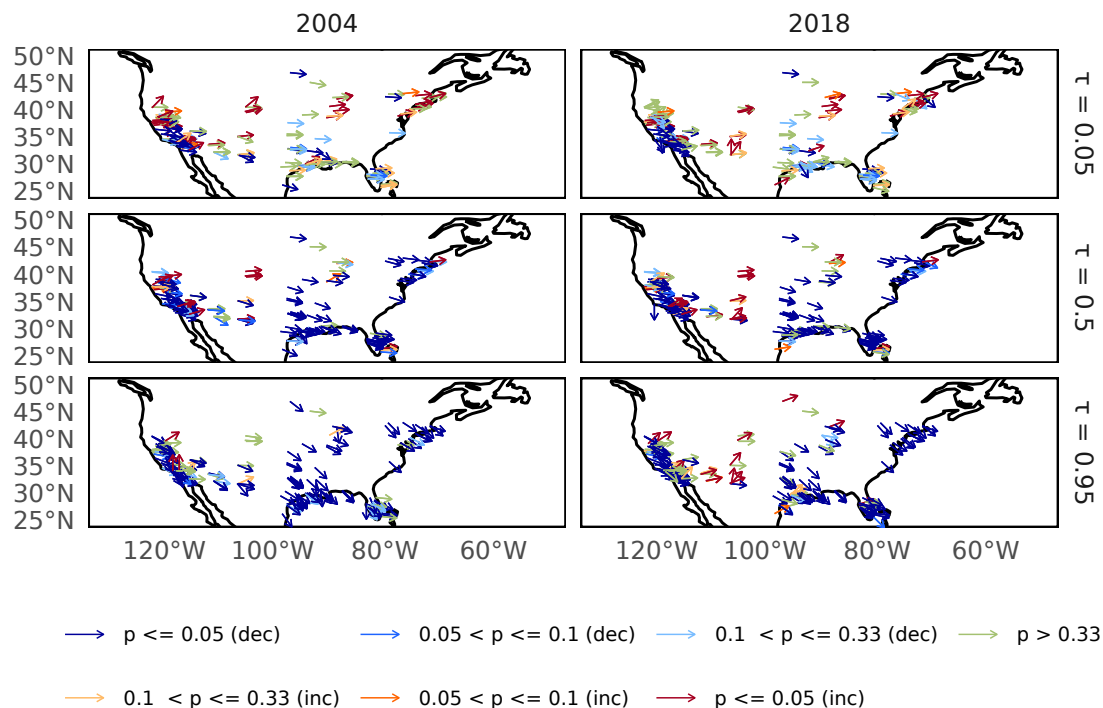
**Figure 4.** Number of  $O_3$  series across slope significance and slope magnitude categories. Those of medium significance ( $p > 0.10$ ) and medium significance and medium magnitude slope are highlighted.

# MDA8O<sub>3</sub> Warm Season

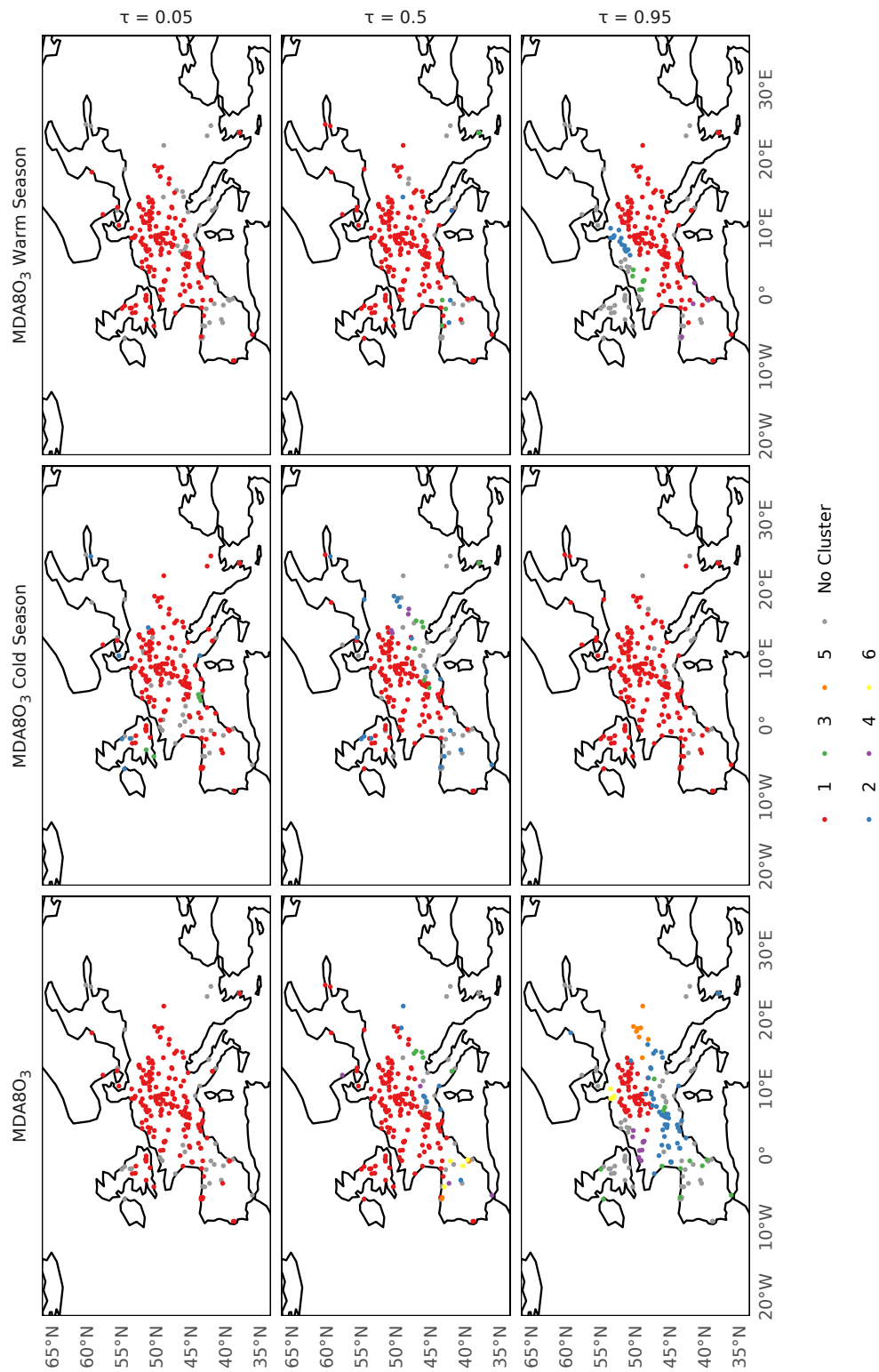


**Figure 5.** Trends in warm season MDA8O<sub>3</sub> in Europe in 2004 and 2018. Trends are shown by an arrow where the angle between vertically up and horizontal corresponds to  $2.5 - 0 \text{ ppbv / yr}^{-1}$ , and between horizontal and vertically down corresponds to  $0 - -2.5 \text{ ppbv / yr}^{-1}$ . The few trends that have a magnitude greater than  $2.5 \text{ ppbv}$  have been clamped. Colour corresponds to the direction and significant of the trend.

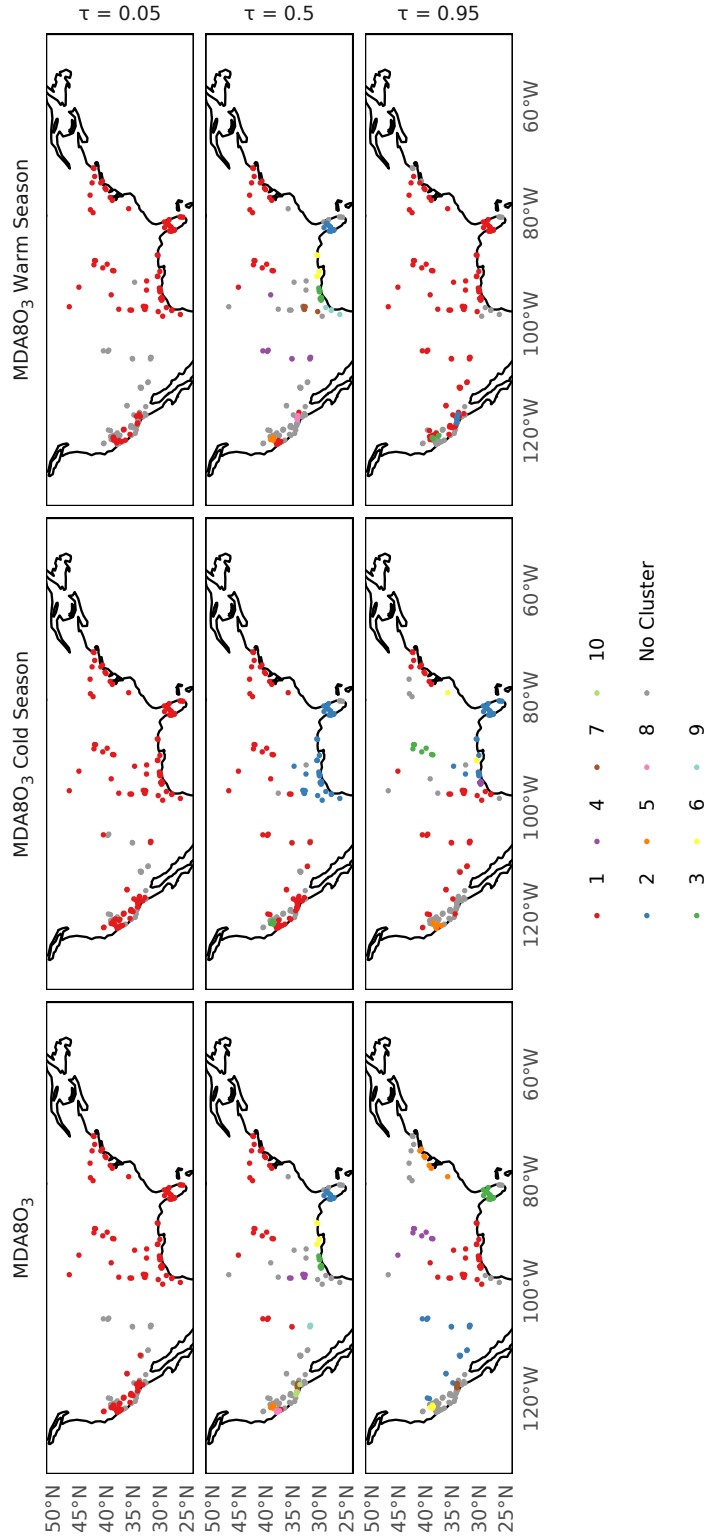
## MDA8O<sub>3</sub> Warm Season



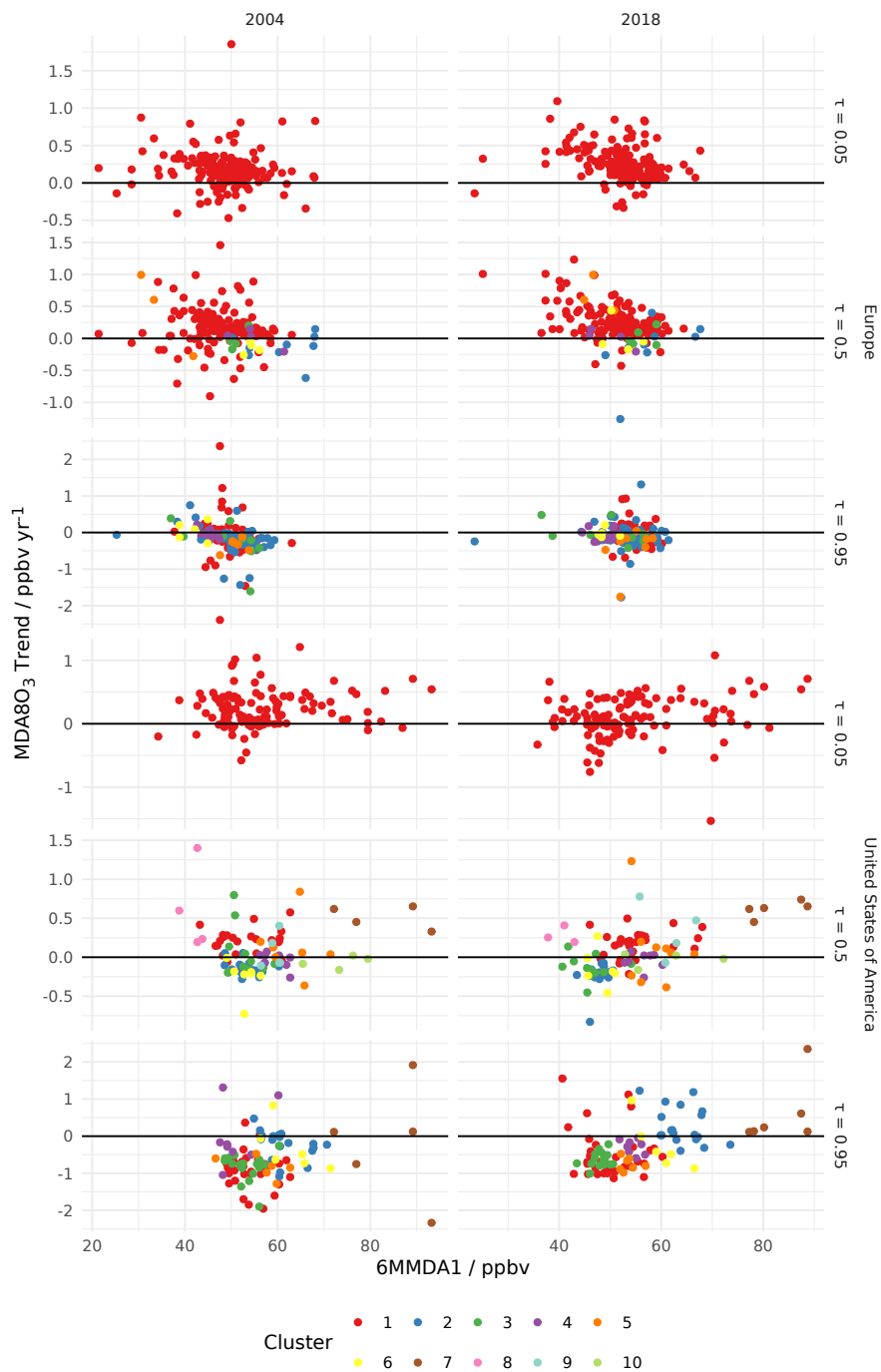
**Figure 6.** Trends in warm season MDA8O<sub>3</sub> in the United States of America in 2004 and 2018. Trends are shown by an arrow where the angle between vertically up and horizontal corresponds to 2.5 - 0 ppbv / yr<sup>-1</sup>, and between horizontal and vertically down corresponds to 0 - -2.5 ppbv / yr<sup>-1</sup>. The few trends that have a magnitude greater than 2.5 ppbv have been clamped. Colour corresponds to the direction and significant of the trend.



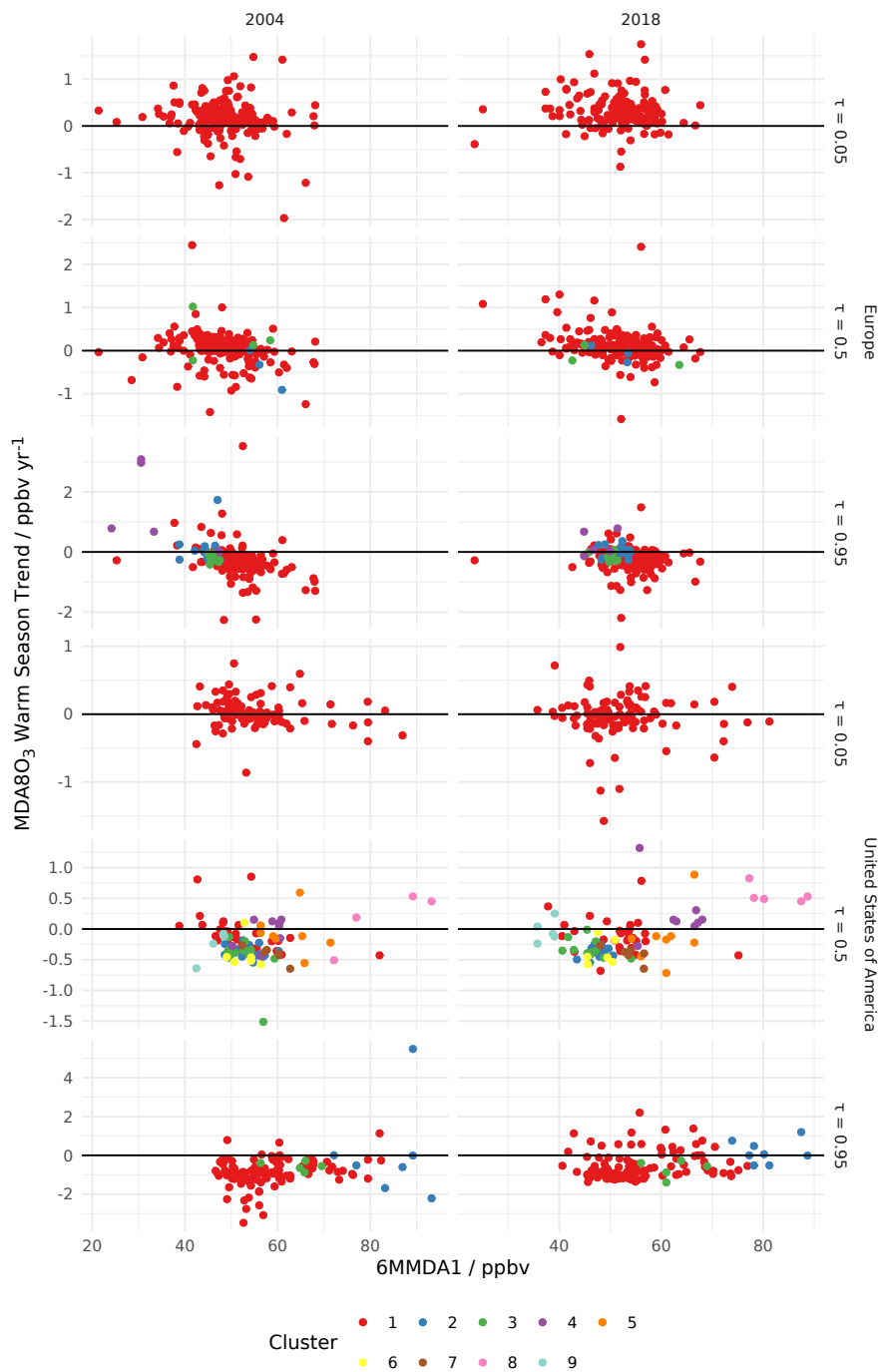
**Figure 7.** Clusters determined for MDA8O<sub>3</sub> warm season and MDA8O<sub>3</sub> cold season at  $\tau$  0.05, 0.50 and 0.95 in Europe. Clusters are coloured when there are  $> 3$  sites within a cluster, otherwise these are assigned 'No Cluster'. Clusters are numbered starting with the cluster containing the most sites as 1 within each pane.



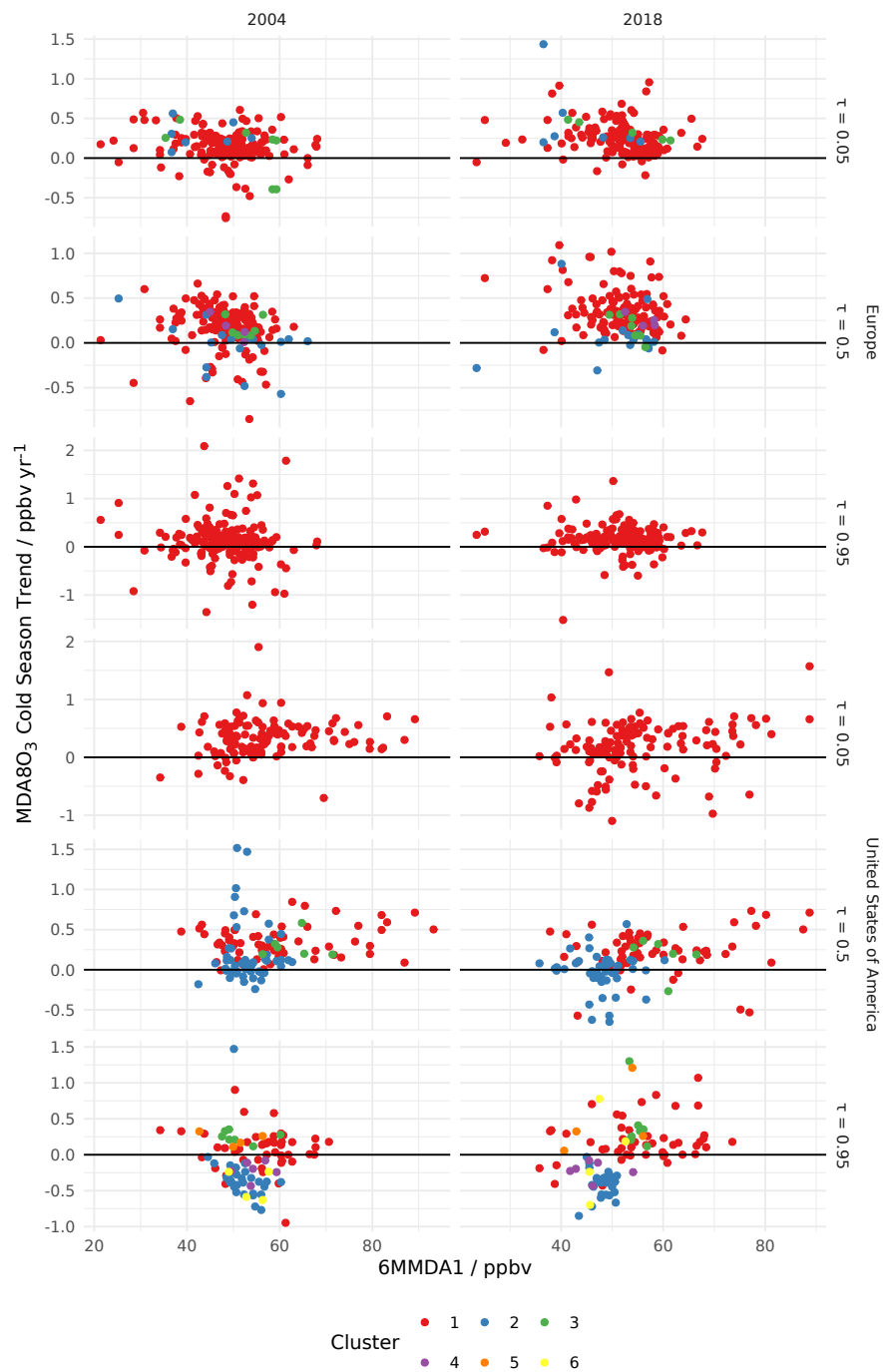
**Figure 8.** As figure 7 for the United States of America



**Figure 9.** For the years 2004 (left) and 2018 (right) scatter plots show that years' MDA8O<sub>3</sub> trend in (top to bottom) Europe  $\tau = 0.05, 0.5, 0.95$ , United States of America  $\tau = 0.05, 0.5, 0.95$  vs the 6MMDA1 for the same year. Colours correspond to clusters presented in figures 7 and 8.



**Figure 10.** For the years 2004 (left) and 2018 (right) scatter plots show that years' warm season MDA8O<sub>3</sub> trend in (top to bottom) Europe  $\tau = 0.05, 0.5, 0.95$ , United States of America  $\tau = 0.05, 0.5, 0.95$  vs the 6MMDA1 for the same year. Colours correspond to clusters presented in figures 7 and 8.



**Figure 11.** For the years 2004 (left) and 2018 (right) scatter plots show that years' cold season MDA8O<sub>3</sub> trend in (top to bottom) Europe  $\tau = 0.05, 0.5, 0.95$ , United States of America  $\tau = 0.05, 0.5, 0.95$  vs the 6MMDA1 for the same year. Colours correspond to clusters presented in figures 7 and 8.

. All data for this study can be downloaded from the relevant databases , this study used the r-packages *toarR* (?) and *saggetr* (?). Code for  
720 performing the download as well as the analysis can be found at 10.5281/zenodo.14538198 (?)

. BSN and WSD equally contributed to all aspects of this manuscript's production

. The authors declare that they have no conflict of interest.

. The Viking cluster was used during this project, which is a high performance compute facility provided by the University of York. We are grateful for computational support from the University of York, IT Services and the Research IT team.

725 The Authors acknowledge Prof. James Lee for their scientific advice and Prof. David Carslaw and Dr Stuart Lacy for their advice on the statistical analysis. We also thank Dr Stuart Lacy for their help with SQL and very useful suggestions on managing large datasets.