ANSWERS TO REFEREE #1

NB : For clarity, the suggestions/remarks of the referee are in italics, our direct answers in bold and our modifications in the manuscript in regular.

Referee 1 : Summary: This paper presents a very useful analysis of the extensive data set of vertical profiles of ozone and CO over the Frankfurt region collected by the MOZAIC-IAGOS program. The paper is much improved from the previous drafts. Nearly all of my concerns have been fully addressed, but one concern remains; when this concern is addressed, I recommend that this paper be accepted for publication. Final concern: In their trends analysis, the authors correctly point out the importance of autocorrelation of the data, and analyze the trends by methods that account for the correlation of annual and seasonal means between successive years. However, autocorrelation at other time scales (points 1 and 2 below) and with altitude (point 3) is also important, and should be accounted for in some of the analyses. Three specific issues are identified below.

ANSWER : Again, we are thankful to the referee for bringing new interesting scientific questions about our study.

Referee 1 : 1) The shaded region in Figure 5 is now defined, but I suspect that the ±2 standard error is underestimated here. I presume that it is calculated by dividing the standard deviation of the measurements by the square root of the number of measurements. However, not all measurements are independent, since there is likely a great deal of autocorrelation between successive measurements. That is, a given 4-second average of O_3 *is likely to be very similar to the immediately receding and immediately following 4-second averages. Was this autocorrelation accounted for in the calculation of the standard errors? If not, the standard errors should be recalculated based on the standard deviation of the measurements divided by the square root of the number of INDEPENDENT measurements. A short discussion of the method used should be given in the Supplement.*

ANSWER: In the seasonal variations shown in Fig. 5, both the average and the uncertainty on the average are calculated based on the monthly averages, and not based on the raw data as assumed by the referee. We fully agree with the referee that individual vertical profiles sampled within a few hours to a few days are likely not independent, and calculating the standard error on this population of points would indeed require to quantify the degree of freedom. However, in our case, the calculations are based on the monthly averages that we assume well determined due to a high frequency of sampling by MOZAIC-IAGOS aircraft. We thus do not take into account the uncertainty on individual monthly averages. For instance, in the LT in January, mean CO mixing ratios of 148, 179, 161, 148, 169, 154, 149, 140, 167, 143, 138 ppb have been calculated from 2001 to 2012. It gives a mean CO mixing ratio of 154 ppb and an uncertainty of ±8 ppb, as shown in Fig. 5. This assumes that the 11 monthly values are independent of each other (we see no evidence for multi-year correlations) and we report the 95% two-sided confidence interval = 1.96 times the standard error. We are not combining trends in the different layers (LT, MT, UT) to reduce any uncertainties and thus the vertical correlation of individual profiles (which can occur) is not relevant to our conclusions.

We modified the legend of Fig. 5 as follows : "Figure 5: Averaged O_3 (left panels) and CO (right panels) seasonal variations above Frankfurt/Munich in all three tropospheric layers, for the 95th percentile (top panels), the mean (middle panels) and the 5th percentile (bottom panels). The shaded areas show the ±2 standard error (i.e. the uncertainty in the average at a 95% confidence level) calculated based on the monthly averages assumed to be well determined (i.e. uncertainties on the individual monthly averages are not taken into account) and independent.**"**

Referee 1 : 2) I am also concerned that the confidence limits derived in Section 3.4.1 are overly optimistic. The shift in the seasonal cycle is now based on daily data, instead of the monthly average data considered in the previous version of the manuscript. However, ozone measurements from sequential days are fairly strongly autocorrelated, because ozone changes in the troposphere are largely driven by synoptic scale transport, which has something like a 5 to 7 day time scale. The authors must examine the degree of autocorrelation, and reduce the numbers of degrees of freedom in the confidence limit calculations accordingly. For example, if they find that the autocorrelation becomes insignificant only after a 5-day lag, then the numbers of degrees of freedom should be reduced by a factor of 5 from the number of daily averages. This would increase the confidence limits in Table 2 by about a factor of the square root of 5. In this regard, if the confidence limits are calculated with proper consideration of this autocorrelation issue, I would not expect the confidence limits to be much improved by using daily averages over those obtained from using monthly averages. The reason for this expectation is that the calculation of the monthly average minimizes the sum of the squares of the deviations of the daily averages from the resulting monthly averages. Then the regression fit of the monthly averages to the sine function minimizes the sum of the squares of the deviations of the monthly averages from the fitted sine function. On the other hand, the regression fit of the daily averages to the sine function minimizes the sum of the squares of the deviations of the daily averages from the fitted sine function. Formally, the process of fitting the sine function gives comparable results regardless of whether it is done in two steps (daily data averaged to months, and then a fit to monthly averages) or one step (a fit to daily averages). However, there may be subtleties in the data structure that actually do give more precise results from the one step process, so it is reasonable to maintain the fit to daily data, but the autocorrelation issue must be properly considered. A short discussion of these issues should be included in the Supplement.

ANSWER : As we are considering daily averages in this section, we agree with the referee that autocorrelation should have been taken into account. The correlograms obtained on the daily residuals (after detrending) indeed identifies some autocorrelation in the data, which leads to underestimated uncertainty estimates of the phase and amplitude. A simple autocorrelation shows that much of the variance is day-to-day independent, but a large fraction has synoptic-scale correlation.

To properly account for these autocorrelation in the estimation of the uncertainties, we assume that the data follows a first-order regressive process, which allows to estimate the effective sample size (n') based on the sample size (n) and the lag-1 autocorrelation coefficient (ρ_1) : $n'=n\,(1-\rho_1)/(1+\rho_1)$ (Wilks et al., 2006). The lag-1 autocorrelation coefficients over the two periods and the three layers range between 0.18 and 0.58. This leads to an increase of the initial

confidence intervals by a factor of 1.2 to 2.0. The seasonal shift remains significant in the LT and MT, but not in the UT.

We modified the text as follows :

- **From page 19 / lines 32-33 to page 20 / lines 1-3 :** “[…] The changes of amplitude and phase obtained with the sine fits are reported in Table 2. The uncertainties directly given by the standard linear least-square regression are underestimated since daily averages of O_3 show some synoptic-scale multi-day correlation (readily seen in the correlograms of the daily residuals of the sine fits, not shown). In order to take into account this autocorrelation in the estimation of the uncertainties, we assume that the data follows a first-order regressive process, which allows to estimate the effective independent sample size (n’) based on the sample size (n) and the lag-1 autocorrelation coefficient (ρ_1): $n' = n\ (1- \rho_1)/(1+ \rho_1)$ (Wilks, 2006). These lag-1 coefficients for the two periods and the three layers range between 0.22 and 0.60, which leads to an increase of the initial confidence intervals by a factor of 1.3 to 2.0. The uncertainties (95% confidence interval) reported in Table 2 reflect this calculation.

  Between the average 1995-2003 and the average 2004-2012, the amplitude of the O_3 seasonal cycle has decreased at levels that are statistically significant throughout the troposphere, with a difference of -2.5±1.7, -1.1±0.8 and -2.1±1.2 ppb decade^{-1} in the LT, MT and UT, respectively. Note that the difference between the two nine-year periods has been scaled to per decade. ”
- **The Table 2 is modified as follows :**

Layer	Amplitude			Phase		
	Amplitude 1995-2003 (ppb)	Amplitude 2004-2012 (ppb)	Amplitude trend (ppb decade^{-1})	Date of seasonal maximum 1995-2003	Date of seasonal maximum 2004-2012	Shift (day decade^{-1})
UT	18.0±0.8	16.1±0.7	-2.1±1.2	23th June ± 2.6 days	20th June ± 2.6 days	-3.3±4.1
MT	11.5±0.5	10.5±0.5	-1.1±0.8	23th June ± 2.4 days	16th June ± 2.9 days	-7.8±4.2
LT	9.9±1.0	7.6±1.1	-2.5±1.7	18th June ± 5.8 days	2nd June ± 8.6 days	-17.8±11.5

- **Page 20 / lines 16-19 :** “The shift of the O_3 maximum (typically in June) between average 1995-2003 and average 2004-2012 is statistically significant in the LT (-17.8±11.5 day decade^{-1}) and MT (-7.8±4.2 day decade^{-1}), but not in the UT (-3.3±4.1 day decade^{-1}). The difference of seasonal shift between the LT and the UT is also significant.”
- **Page 20 / lines 19-22, the sentence** “Note that applying the sine fit to the monthly O_3 mixing ratios gives similar shift estimates but much larger uncertainties, leading to insignificant differences among the tropospheric layers (-13.3±11.6 and -6.7±6.5 day decade^{-1} in the LT and MT, respectively).” **is removed**
- **Page 21 / lines 11-12 :** “The vertical profile observations provide unique data, allowing us to show that this seasonal change of the phase in the O_3 maximum to earlier days in the year extends above the surface at Frankfurt/Munich. The

magnitude of this shift in maximum is statistically significant through the middle troposphere. ”

- **Page 23 / lines 9-14 :** “Results highlight a statistically significant change of the phase in the LT, ozone maxima occurring earlier by -17.8±11.5 days decade^{-1} on average (at a 95% confidence level), in general agreement with what can be inferred from previous results from the literature (Parrish et al., 2013). A major contribution of this study is that it extends the analysis throughout the troposphere, and shows such shifts becoming smaller and less significant as one approaches the tropopause. In particular, the difference of seasonal shift between the LT and UT is statistically significant. The larger contribution from other regions (e.g. Asia) higher in altitude may explain the lower seasonal shift observed in the free troposphere and close to the tropopause, although further studies are obviously required to quantitatively assess this issue. ”
- **Page 2 / lines 1-5 : The text in the abstract is modified as follows :** “The O_3 maxima moves forward in time with a rate of -17.8±11.5 days decade^{-1} in the lower troposphere, in general agreement with previous studies. Interestingly, this seasonal shift is shown to persist in the middle troposphere (-7.8±4.2 days decade^{-1}) but turns insignificant in the upper troposphere.”

In the references :

Wilks, D. S. : Statistical Methods in the Atmospheric Sciences, 2nd edition, Academic Press, San Diego, 2006.

Referee 1 : 3) Autocorrelation is also important in the vertical. If I understand correctly, the daily averages were separately calculated for three atmospheric layers: the LT (a layer 1 to 2 km thick), the MT that is (several km thick), and the UT (~1.6 km thick). Thus, a day with a single altitude profile gives three averages, one for each layer. However, especially in the MT, I suspect that a single profile actually gives two or more independent measurements representing ozone or CO in the mid-troposphere. The authors should investigate the autocorrelation of the measured concentrations as a function of altitude offset. I expect that the autocorrelation is insignificant for vertical offsets of ~2 km, thus allowing several measurements of ozone or CO to be obtained from each profile, at least in the MT. If this issue is properly considered, then the numbers of degrees of freedom would be greater, implying that the confidence limits in Table 2 should be smaller, at least in the MT. A short discussion of these issues should be included in the Supplement.

ANSWER : Following the recommendations of the referee, we investigated the autocorrelation in the vertical in the O_3 profiles. First, we calculated the time series of daily averages of O_3 mixing ratios over the period 1994-2012, at 250 m-thick layers from 0 to 12,500 m (i.e. 50 layers). Second, we detrended and deseasonalised the 50 time series by applying the following linear, least-square regression :

$$\tilde{y} = a + bt + c.\cos\left(\frac{2\pi}{365}.t\right) + d.\sin\left(\frac{2\pi}{365}.t\right)$$

Third, we computed the correlation of the O_3 residuals between each pair of altitudes. The results are shown below for O_3.

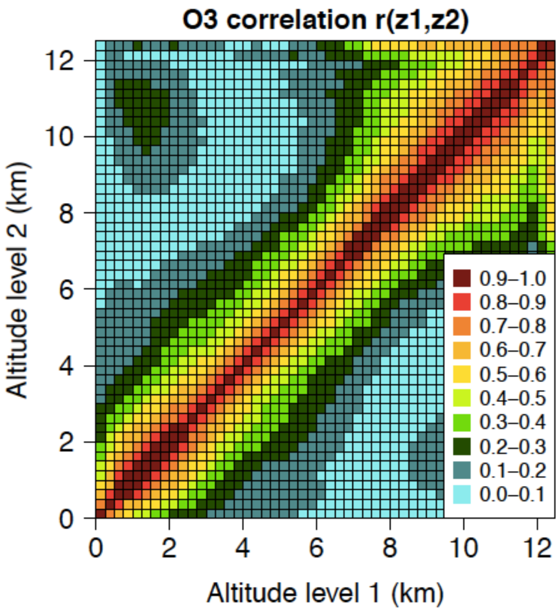

Figure : Correlation of O_3 daily time series (period 1994-2012 and 2002-2012, respectively) between the different levels of altitude (see text). Note that this correlation matrix is symmetric by construction.

For a given distance, the correlation between two layers depends on the location in the troposphere. For clarity, we define ρ_k as the average correlation between all pairs of layers distant from k*250 m (i.e. an average over 50 minus k layers). Results are reported in the Table below. A high average correlation is found between all pairs of adjacent layers (ρ_1=0.90). In order words, on average over the troposphere, 81% (=0.90^2) of the day-to-day variability of O_3 at a given altitude is described by the O_3 mixing ratios 250 m above or below. At a distance of 1000 m, this correlation is reduced to (ρ_4=) 0.64. At a distance of 2500 m, it falls to (ρ_8=) 0.33, i.e. only 11% of the O_3 variability in a layer is described by the O_3 2500 m away.

Table : Average correlations of O_3 residuals for several lag distances. For a given distance, ρ_k is calculated as the average of all correlations between all pairs of altitude distant by k*250 m (see text).

Lag-k distance (m)	Average correlation ρ_k
250	0.90
500	0.80
750	0.71
1000	0.64
1250	0.57
1500	0.51
1750	0.46
2000	0.41
2250	0.37
2500	0.33
2750	0.30

3000	0.27
3250	0.24
3500	0.21
3750	0.19
4000	0.17

Therefore, as expected by the referee, there is an autocorrelation in the vertical. And thus, taking into account the vertical autocorrelation would be required if we wanted to estimate the uncertainties of the daily averages in each layer. However, in our procedure to estimate the amplitude and seasonal changes and their corresponding uncertainties, we assume that the daily averages on each layer are well determined. The uncertainties derived for the amplitude and the phase thus only originate from the regression, now taking into account the autocorrelation (in time) of the daily time series (Cf. our answers to previous comments).

ANSWERS TO REFEREE #2

NB : For clarity, the suggestions/remarks of the referee are in italics, our direct answers in bold and our modifications in the manuscript in regular.

Referee 2 : Indeed, the manuscript strongly improved and the concerns raised by the reviewers were addressed in a very detailed way, very good! Also the English spelling is much better, although there are many (>40 or so) cases where e.g. the article was misplaced. I support the publications when the minor revisions below are considered.
ANSWER : We thank the referee for its positive appreciation of our last modifications. We took into account his new comments, especially the comment related the definition of the tropopause used here.

Referee 2 : Minor remarks:
p.1, l.15 delete "(due to dry deposition at ground and titration by NO)", as this is first only half of the truth and secondly not inferred from the data
ANSWER : Modification applied.

Referee 2 : p.1, l.17 delete "(due to stratosphere-to-troposphere in-mixing)", same explanation
ANSWER : Modification applied.

Referee 2 : p.7, l.13 In the mid- and high latitudes, where passenger aircraft can reach the tropopause, the ozone tropopause is far below 150 ppbv, see Bethan et al. (1996) and well measured by Zahn and Brenninkmeijer (Atmos. Environment, 2001) and thereafter verified by Thouret et al (ACP, 2006). Your argument "Therefore, the DT derived from PV values tends to be located below the 150 ppb O3 - isopleth, which may bias low the O3 mixing ratios in the UT." is thus not okay, especially because you may attribute data points with >100 ppb (which may be stratospheric) are attributed to the UT. The UT defined here using p_{2PVU} *+ 15 hPa is basically a* *conservative parameter, as 2 PVU is quite low as definition of the DT. Your DT study/statistics indicates (in my opinion) that the PV field from ECMWF is quite often not a suitable parameter to define the tropopause. You should discuss this shortly.*

ANSWER : We initially chose this criteria of 150 ppb in order to avoid the possible stratospheric intrusion and considering that O_3 vertical gradients are strong at the tropopause level (Thouret et al. (2006) reported O_3 mixing ratios ranging from 200 to 400 ppb in the lower stratosphere). However, the referee is right about the fact that this criteria probably includes too much stratospheric air masses. As the seasonal variations of O_3 mixing ratios at the tropopause are strong, we modified the sensitivity test by choosing a dynamic criteria based on the formula 97+26*sin((DayOfYear-30)*2π/365) ppb proposed by Zahn et al. (2002). This formula was based on the CARIBIC observations and is consistent with the O_3 mixiong ratios reported by Thouret et al. (2006) in the tropopause layer based on the MOZAIC observations.

Compared to our previous sensitivity test, quite large discrepancies can still be found on some profiles, but the mean bias of the DT pressure is reduced to +2 hPa, thus ensuring that there is no systematic bias with our definition of the tropopause. We modified the text as follows : "In order to assess the uncertainties introduced by an erroneous DT pressure, we compared it with the pressure at which O_3 reaches the typical tropopause O_3 mixing ratios and remains above at higher altitude (in order to avoid stratospheric intrusions in the troposphere). As O_3 at the tropopause varies seasonally, we consider a dynamic criteria given by the formula 97+26*sin((DayOfYear-30)*2π/365) (ppb) proposed by Zahn et al. (2002) based on CARIBIC observations (and consistent with Thouret et al., 2006). A maximum pressure is fixed to 600 hPa in order to avoid a wrong allocation if for instance a high O_3 plume is sampled in the BL with missing data above in the profile. This was done on all vertical profiles where it was possible, which represents 41% of the dataset. On average over the period 1994-2012, the mean bias of the DT pressure compared to the chemical tropopause derived from O_3 mixing ratios is +2 hPa, while the 5th, 10th, 50th, 90th and 95th percentiles of this bias are -138, -57, +10, +57 and +75 hPa, respectively. Therefore, although quite large discrepancies can exist on some profiles, the DT pressure derived from PV does not appear systematically biased. It is beyond the scope of this study to investigate in more details the influence of the method used to locate the tropopause."

And we removed the sentences page 13, lines 18-21 : "In addition, as shown in Sect. 2.2, the altitude of the UT (defined here based on the PV values) is biased low compared to the UT derived based on the chemical tropopause. Thus, the UT may be less influenced by the stratosphere and more by the free troposphere, which may increase the correlation between the MT and UT."

Referee 2 : p.9, l.10-12 delete "(due to dry deposition and enhanced titration by NO in the BL)" and "(due to STE)"

ANSWER : Modification applied.

Referee 2 : p.9, l.15 What are O3 episodes? high/low, short/long H episodes?

ANSWER : We replaced "O_3 episodes" by "short episodes of high O_3 mixing ratios"

Referee 2 : p.9, l.20 Citations for the « C » shaped profile

ANSWER : We modified the sentence as follows : "Thus, one might have expected higher mixing ratios in the BL than in the lower free troposphere (sometimes described as a « C » shaped profile), as sometimes observed in polluted cities (Ding et al., 2008; Tressol et al., 2008).”

Referee 2 : p.10, l.6 Not so simple to understand what you describe with “daily variability” on different scales. I guess “day-to-day variability” sounds more plausible here. Please use this synonym throughout the text.
ANSWER : Modification applied here and elsewhere.

Referee 2 : p.10, l.12 Do “transient exchanges” exist? _ “transient exchange processes”
ANSWER : “Transient exchanges” corresponds to the formulation given by Stohl et al. (2003b). Thus, to our opinion, no changes are needed here.

Referee 2 : p.13, l.3 I would first discuss the vertical profile (para starting at l.13) and thereafter the long-term time series (Fig. 6).
ANSWER : Modification applied.

Referee 2 : p.13, l.18 : No! The shown O3 values in the UT are significantly affected by unwanted attribution of stratospheric air (see my argument above), as indicated by the 95th percentile showing levels of up to 115 ppb. Modify this para.
ANSWER : Indeed, the high 95th percentile of O_3 found in the UT shows a contribution of stratospheric air masses. These high O_3 mixing ratio maybe be related to a wrong estimation of the tropopause height (i.e. an underestimated DT pressure). We saw in Sect. 2.2 (Cf. above our answer to the 3rd comment) that there is likely no systematic bias in the determination of the tropopause height, although there is a quite large error (with a few discrepancies as large as 140 hPa). However, it is worth keep in mind that these high O_3 mixing ratios may also be due to recent stratospheric intrusions in the troposphere since, as explained in Sect. 2.2, we do not consider purely tropospheric layer as recent stratospheric intrusions are not filtered. Thus, some high O_3 mixing ratios are expected in the UT.
We modified this paragraph as follows : “This may be explained by the fact that both the first kilometre and the tropopause layer are not taken into account in this study, which likely reduces the differences of interannual variation among the tropospheric layers as defined in this study. However, some high O_3 mixing ratios are still observed in the UT, with for instance a 95th percentile up to 115 ppb in summer, which suggests an influence of stratospheric air. They are likely partly due to a wrong estimation of the tropopause height since we saw in Sect. 2.2 that, despite a very low mean bias, quite large errors can be found compared to a chemical tropopause determined based on typical tropopause O_3 mixing ratios. As we do not consider a purely tropospheric UT (see Sect. 2.2), some of these high O_3 mixing ratios may also be due to stratospheric intrusions in the troposphere.”

Referee 2 : p.14, l.4 “... CO emissions at northern mid-latitudes when the photolysis is limited”. Upps & boah, I didn’t know that photolysis controls the decay of CO. Correct this somewhat embarrassing part!
ANSWER : Indeed, it was a mistake, we were meaning “photochemical destruction by OH”. We modified the sentence as follows : “The winter-time maximum results from the accumulation of the primary CO emissions at northern

mid-latitudes when the oxidation by OH is limited."

Referee 2 : p.16, l.12 : "the year 2000 is taken as a reference (i.e. the origin of the time series)". Why only as of 2000 and not 1994? Explain!
ANSWER : We simply chose the year 2000 as a reference (following the recommendation of a previous referee) in order to allow an direct comparison with the results of Parrish et al. (2014). We modified the text as follows : "[…] and is used for the normalization. The year 2000 is chosen in order to facilitate the comparison with the results obtained by Parrish et al. (2014). "

Referee 2 : p.18, l.12 : "… which is consistent with the trends found here over the period 1994-2012". Again, did you consider the entire period or only as of 2000?
ANSWER : Following the recommendation of a previous referee, we removed the analysis of trends over the subperiod 2000-2012. So this sentence correctly refers to the results obtained over the entire period (1994-2012).

Referee 2 : p.18, l.14 "The persistent positive trends found higher in altitude suggest that wintertime O3 has increased at a large scale". Refer also here to the supplement.
ANSWER : We modified the sentence as follows : "The persistent positive trends found higher in altitude (see Table S1 in the Supplement) suggest that wintertime O_3 has increased at a large scale (if not hemispheric) since air masses sampled by MOZAIC-IAGOS aircraft in both the MT and UT can be influenced by emissions from North America and Asia (as shown in Fig. 2)."

Referee 2 : p.18, l.19 The numbers given here differ from the ones given in the supplement.
ANSWER : It is normal : the O_3 trends given in the manuscript take into account the autocorrelation while the results reported (for information) in Table S2 in the Supplement do not (the trend best estimates are the same but their uncertainties are larger when the autocorrelation of the data is taken into account).

Referee 2 : p.18, l.29 "the reference year 2004" instead of "the 2004 reference year"
ANSWER : Modification applied.

Referee 2 : p.19, l.2 You often write "all the …". Skip there "the"
ANSWER : Modification applied here and elsewhere.

Referee 2 : p.19, l.5 … a decrease of the total column of CO over Europe
ANSWER : Modification applied.

Referee 2 : p.19, l.19 "The seasonal variation of O3 can be well approximated by a sine function". I don't see this in a figure.
ANSWER : To our opinion, it is a reasonable assumption, already done in other papers as Parrish et al. (2013), and thus it does not require an additional figure.

Referee 2 : p.20, l.5 "The differences of amplitude change between the different layers all remain statistically insignificant." I don't understand this sentence. The numbers in tab. 2 differ and more than the standard deviations indicate.

ANSWER : Actually, only the difference of amplitude change between the LT and the MT was significant (difference of 1.4±1.0 ppb decade^{-1}). However, following the recommendations of the other referee, we now take into account the autocorrelation of the daily-averaged data. This finally increases the uncertainties on the results of the sine fit (see our answers to the questions of referee 1), leading to statistically similar amplitude changes in the different layers (-2.1±1.2, -1.1±0.8 and -2.5±1.7 ppb decade^{-1} in the LT, MT and UT, respectively). Thus, no modification is required in the text.

Referee 2 : p.20, l.10 "…O3 on the 18th June in the LT and on the 23th June …" and later in the text the same.
ANSWER : Modification applied here and elsewhere.

Referee 2 : p.22, l.16 "day-to-day" instead of "daily"
ANSWER : Modification applied here and elsewhere.

Referee 2 : p.22, l.19 "Maximum day-to-day variability of CO" instead of "A maximum of variability"
ANSWER : Modification applied.

Referee 2 : p.22, l.25 "the entire troposphere" instead of "in all the troposphere"
ANSWER : Modification applied.

Referee 2 : A couple of times you write "variability and trends of …". Why once singular and the other time plural? "variability and trend of …" fits best.
ANSWER : Modification applied in the whole manuscript.

OTHER MODIFICATIONS

Page 9, line 22-25 : "Actually, such « C » shaped profile is only observed when considering the 95th percentile rather than the mean O_3 mixing ratio (Petetin et al., Diurnal cycle of ozone throughout the troposphere over Frankfurt as measured by MOZAIC-IAGOS commercial aircraft, under review in Elementa Science of the Anthropocene)." ➔ "Actually, such « C » shaped profile is only observed when considering the 95th percentile rather than the mean O_3 mixing ratio (Petetin et al., 2016).**" And we added in the corresponding reference :** Petetin, H., Thouret, V., Athier, G., Blot, R., Boulanger, D., Cousin, J.-M., Gaudel, A., Nédélec, P. and Cooper, O.: Diurnal cycle of ozone throughout the troposphere over Frankfurt as measured by MOZAIC-IAGOS commercial aircraft, Elem. Sci. Anthr., 4(000129), 1–11, doi:10.12952/journal.elementa.000129, 2016.

List of changes in the manuscript.
For clarity, the removed sentences are in red, the added/modified sentences in green. Note that the very minor changes (in response to the second referee) are not repeated here.

- **Page 2 / lines 1-5 : The text in the abstract is modified as follows : "**The O_3 maxima moves forward in time with a rate of -17.8±11.5 days decade^{-1} in the lower troposphere, in general agreement with previous studies. Interestingly, this seasonal shift is shown to persist in the middle troposphere (-7.8±4.2 days decade^{-1}) but turns insignificant in the upper troposphere.**"**
- **Page 7 / lines 13-23 : we modified the text as follows : "**In order to assess the uncertainties introduced by an erroneous DT pressure, we compared it with the pressure at which O_3 reaches the typical tropopause O_3 mixing ratios and remains above at higher altitude (in order to avoid stratospheric intrusions in the troposphere). As O_3 at the tropopause varies seasonally, we consider a dynamic criteria given by the formula $97+26*\sin((\mathrm{DayOfYear}-30)*2\pi/365)$ (ppb) proposed by Zahn et al. (2002) based on CARIBIC observations (and consistent with Thouret et al., 2006). A maximum pressure is fixed to 600 hPa in order to avoid a wrong allocation if for instance a high O_3 plume is sampled in the BL with missing data above in the profile. This was done on all vertical profiles where it was possible, which represents 41% of the dataset. On average over the period 1994-2012, the mean bias of the DT pressure compared to the chemical tropopause derived from O_3 mixing ratios is +2 hPa, while the 5th, 10th, 50th, 90th and 95th percentiles of this bias are -138, -57, +10, +57 and +75 hPa, respectively. Therefore, although quite large discrepancies can exist on some profiles, the DT pressure derived from PV does not appear systematically biased. It is beyond the scope of this study to investigate in more details the influence of the method used to locate the tropopause.**"**
- **Page 9 / line 20 : We modified the sentence as follows : "**Thus, one might have expected higher mixing ratios in the BL than in the lower free troposphere (sometimes described as a « C » shaped profile), as sometimes observed in polluted cities (Ding et al., 2008; Tressol et al., 2008).**"**

- **Page 9, line 22-25 : we modified the text "**Actually, such « C » shaped profile is only observed when considering the 95th percentile rather than the mean O_3 mixing ratio (Petetin et al., Diurnal cycle of ozone throughout the troposphere over Frankfurt as measured by MOZAIC-IAGOS commercial aircraft, under review in Elementa Science of the Anthropocene)." **as follows :** "Actually, such « C » shaped profile is only observed when considering the 95th percentile rather than the mean O_3 mixing ratio (Petetin et al., 2016).**"**

- **Page 12, legend of Fig. 5 modified as follows : "**Figure 5: Averaged O_3 (left panels) and CO (right panels) seasonal variations above Frankfurt/Munich in all three tropospheric layers, for the 95th percentile (top panels), the mean (middle panels) and the 5th percentile (bottom panels). The shaded areas show the ±2 standard error (i.e. the uncertainty in the average at a 95% confidence level) calculated based on the monthly averages assumed to be well determined (i.e. uncertainties on the individual monthly averages are not taken into account) and independent.**"**

- **Page 13, lines 18-21 : "**In addition, as shown in Sect. 2.2, the altitude of the UT (defined here based on the PV values) is biased low compared to the UT derived based on the chemical tropopause. Thus, the UT may be less influenced by the stratosphere and more by the free troposphere, which may increase the correlation between the MT and UT.**" is removed**
- **Page 13 / lines 16-21 : we modified this paragraph as follows : "**This may be explained by the fact that both the first kilometre and the tropopause layer are not taken into account in this study, which likely reduces the differences of interannual variation among the tropospheric layers as defined in this study. However, some high O_3 mixing ratios are still observed in the UT, with for instance a 95th percentile up to 115 ppb in summer, which suggests an influence of stratospheric air. They are likely partly due to a wrong estimation of the tropopause height since we saw in Sect. 2.2 that, despite a very low mean bias, quite large errors can be found compared to a chemical tropopause determined based on typical tropopause O_3 mixing ratios. As we do not consider a purely tropospheric UT (see Sect. 2.2), some of these high O_3 mixing ratios may also be due to stratospheric intrusions in the troposphere.**"**
- **Page 15 / lines 3-4 : we modified the sentence as follows : "**The winter-time maximum results from the accumulation of the primary CO emissions at northern mid-latitudes when the oxidation by OH is limited.**"**
- **Page 16 / line 14 : we added the following text : "[…]** and is used for the normalization. The year 2000 is chosen in order to facilitate the comparison with the results obtained by Parrish et al. (2014). **"**

- **Page 18 / lines 14-17 : we modified the sentence as follows : "**The persistent positive trends found higher in altitude (see Table S1 in the Supplement) suggest that wintertime O_3 has increased at a large scale (if not hemispheric) since air masses sampled by MOZAIC-IAGOS aircraft in both the MT and UT can be influenced by emissions from North America and Asia (as shown in Fig. 2).**"**

- **From page 19 / lines 32-33 to page 20 / lines 1-3 :** "[…] The changes of amplitude and phase obtained with the sine fits are reported in Table 2. The uncertainties directly given by the standard linear least-square regression are underestimated since daily averages of O_3 show some synoptic-scale multi-day correlation (readily seen in the correlograms of the daily residuals of the sine fits, not shown). In order to take into account this autocorrelation in the estimation of the uncertainties, we assume that the data follows a first-order regressive process, which allows to estimate the effective independent sample size (n') based on the sample size (n) and the lag-1 autocorrelation coefficient (ρ_1): $n' = n\,(1-\rho_1)/(1+\rho_1)$ (Wilks, 2006). These lag-1 coefficients for the two periods and the three layers range between 0.22 and 0.60, which leads to an increase of the initial confidence intervals by a factor of 1.3 to 2.0. The uncertainties (95% confidence interval) reported in Table 2 reflect this calculation.

  Between the average 1995-2003 and the average 2004-2012, the amplitude of the O_3 seasonal cycle has decreased at levels that are statistically significant throughout the troposphere, with a difference of -2.5±1.7, -1.1±0.8 and -2.1±1.2 ppb decade^{-1} in the LT, MT and UT, respectively. Note that the difference between the two nine-year periods has been scaled to per decade. "
- **Page 20, the Table 2 is modified as follows :**

Layer	Amplitude			Phase		
	Amplitude 1995-2003 (ppb)	Amplitude 2004-2012 (ppb)	Amplitude trend (ppb decade^{-1})	Date of seasonal maximum 1995-2003	Date of seasonal maximum 2004-2012	Shift (day decade^{-1})
UT	18.0±0.8	16.1±0.7	-2.1±1.2	23th June ± 2.6 days	20th June ± 2.6 days	-3.3±4.1
MT	11.5±0.5	10.5±0.5	-1.1±0.8	23th June ± 2.4 days	16th June ± 2.9 days	-7.8±4.2
LT	9.9±1.0	7.6±1.1	-2.5±1.7	18th June ± 5.8 days	2nd June ± 8.6 days	-17.8±11.5

- **Page 20 / lines 16-19 :** "The shift of the O_3 maximum (typically in June) between average 1995-2003 and average 2004-2012 is statistically significant in the LT (-17.8±11.5 day decade^{-1}) and MT (-7.8±4.2 day decade^{-1}), but not in the UT (-3.3±4.1 day decade^{-1}). The difference of seasonal shift between the LT and the UT is also significant."
- **Page 20 / lines 19-22 :** "Note that applying the sine fit to the monthly O_3 mixing ratios gives similar shift estimates but much larger uncertainties, leading to insignificant differences among the tropospheric layers (-13.3±11.6 and -6.7±6.5 day decade^{-1} in the LT and MT, respectively)." **is removed**
- **Page 21 / lines 11-12 :** "The vertical profile observations provide unique data, allowing us to show that this seasonal change of the phase in the O_3 maximum to earlier days in the year extends above the surface at Frankfurt/Munich. The magnitude of this shift in maximum is statistically significant through the middle troposphere. "
- **Page 23 / lines 9-14 :** "Results highlight a statistically significant change of the phase in the LT, ozone maxima occurring earlier by -17.8±11.5 days decade^{-1} on average (at a 95% confidence level), in general agreement with what can be inferred from previous results from the literature (Parrish et al., 2013). A major contribution of this study is that it extends the analysis throughout the troposphere, and shows such shifts becoming smaller and less significant as one approaches the tropopause. In particular, the difference of seasonal shift between the LT and UT is statistically significant. The larger contribution from other regions (e.g. Asia) higher in altitude may explain the lower seasonal shift observed in the free troposphere and close to the tropopause, although further studies are obviously required to quantitatively assess this issue. "
- **In the references, we added :**
    - Wilks, D. S. : Statistical Methods in the Atmospheric Sciences, 2nd edition, Academic Press, San Diego, 2006.
    - Petetin, H., Thouret, V., Athier, G., Blot, R., Boulanger, D., Cousin, J.-M., Gaudel, A., Nédélec, P. and Cooper, O.: Diurnal cycle of ozone throughout the troposphere over Frankfurt as measured by MOZAIC-IAGOS commercial aircraft, Elem. Sci. Anthr., 4(000129), 1–11, doi:10.12952/journal.elementa.000129, 2016.