ANSWERS TO REFEREE #1

We thank the referee for his detailed comments and suggestions which will greatly help improving this paper.

This paper presents an interesting analysis of the extensive data set of vertical profiles of ozone and CO over the Frankfurt region collected by the MOZAIC-IAGOS program. The analysis of the CO data is quite interesting and deserves publication when some problems are addressed. The analysis of the ozone data has serious problems. First, and very importantly, I think that the statistical significance of all the ozone analysis is not adequate, and even the analysis that is done is not carefully considered in the discussion and in reaching conclusions. One glaring example is the first sentence of Main Result #2 in Section 5 - Summary and conclusions: "On an annual basis over the 1994–2012 period, tropospheric ozone shows increasing upward trends with altitude, from +0.11±0.21 and +0.13±0.18 ppb yr-1 in the lower and mid- troposphere to +0.22±0.23 ppb yr-1 in the upper troposphere." None of these trends are statistically significantly different from zero, and there are certainly no statistically significant differences between them. This would be even clearer if the trends were expressed on a relative basis (as suggested in the first major issue below.

ANSWER: Detailed answers are provided below for all these comments raised by the referee (in particular, see our answers to the major issue 1).

Second, I think that there is an error in the sine function fitting to the ozone seasonal cycle. In order for this paper to be further considered for publication, a much more rigorous analysis of statistical uncertainties must be completed, and the discussion and conclusions extensively revised with that analysis fully and carefully considered. Fixing the error in the seasonal cycle analysis will be straight forward, but here again the statistical uncertainties must be carefully evaluated.

ANSWER: Except an obvious typographical error in the formula that does not impact the results discussed in the text, there is to our opinion no error in the sine function but we agree that a few information were missing in the text to help the reader understand the results. A detailed answer to this point is given below in the major issue 3.

REFEREE, major issue 1: In my opinion, the discussion of ozone trends (Section 3.3.1) is completely inadequate. The problem that the authors face is that their data record, although the densest record of vertical ozone profiles in existence, is short (1994-2012, or 19 years) in terms of that required to precisely quantify long-term changes. This problem is well indicated by the first sentence of the section, which can be shortened to read: "Considering mean ozone concentrations over the whole 1994–2012 period, annual trends are not statistically significant at the 95% confidence level." The same statement is close to correct for the 5th and 95th percentiles. Further detailed discussion of insignificant or nearly insignificant trends is not useful.

ANSWER: As recognized by the referee, we do not ignore that many of these O_3 trends are statistically insignificant, as it is mentioned in the initial manuscript. However, in hindsight, we fully agree with the referee that many of our formulations are awkward (in particular relatively to these insignificant trends) and that a part of our discussion is thus not relevant. A complete overhaul of Sect. 3.3 is proposed below, at the end of this first major issue.

Discussion of 10-years moving trends of is also not useful, since they are all of marginal significance, highly covariant since successive trends differ in only one year's of data, and

prone to biases. This is particularly true of this data set, since the year 2003, which had anomalously high ozone due to the European heat wave, occurs in the middle of the data record. During the early 10-year periods, this year comes at the end, which would tend to give positive trends. During the later 10-year periods, this year comes at the beginning, which would tend to give negative trends, or obscure what otherwise might have been a significant positive trend. I suggest the authors utilize the approach of Parrish et al. (2014). This reference shows that long-term changes are best analyzed by normalizing the data to a reference year, and expressing the changes as a percent of that reference. They suggest the year 2000 for this reference. This normalization then allows trends to be directly compared between studies at different locations or altitudes (as is the case in the present study). Importantly, this reference finds that all European sites have a common trend when expressed in this manner. If the authors follow the same process they can quickly and directly compare their derived long-term changes to those calculated from the parameters reported in Table 2 of Parrish et al. (2014). If their derived trends agree with those calculated within statistical uncertainty, then the discussion is ended, as the trends reported here support the conclusions in that reference, but provide no new information. I am confident that the authors will find such agreement. Quantifying changing trends is more quantitatively done on the basis of a statistically significant quadratic term (Logan et al., 2012; Parrish et al., 2012) rather than 10- year moving trends.

ANSWER: We agree that normalizing our trend results makes the comparisons between the tropospheric layers or with the literature easier. In the revised manuscript, we thus adopted the approach of Parrish et al. (2014), with all trends now expressed relatively to the reference year 2000. Note that we did the same for CO with 2004 as a reference year; this year was chosen because the mean CO concentrations in 2004 (101, 114 and 140 ppb in the LT, MT and UT, respectively) are very close to the mean CO over the 2002-2012 period (100, 114 and 141 ppb in the LT, MT and UT, respectively). We also agree with the referee that our trend analysis on 10-year moving periods is not so relevant due to the 2003 anomaly in the middle of the dataset. This part has thus been removed.

Concerning the trends calculation, we completely modified our approach by now using the non-parametric Mann-Kendall analysis (combined with Theil-Sen slope estimates) that does not make any distributional assumptions and is less sensitive to outliers. As explained in more details in our next answer, we also followed the recommendation of the referee by taking into account the autocorrelation of the data in the calculation of the trend uncertainties. Thus, our trend results have changed. Compared to the results given in the first version of the manuscript, some trends are now insignificant with our new approach, including the increase of the O_3 5th and 95th percentile in winter and the increase in autumn.

We added more details in the comparison of our results with the literature (see the new text at the end of this major issue). In particular, we mentioned the picture drawn by Parrish et al. (2012) (i.e. trends, when expressed relatively to the concentration in 2000, are of similar magnitude at many regional background sites whatever the season). However, it is to be mentioned that a direct comparison of our results with those of Parrish et al. (2014) at European alpine sites (as suggested by the referee) is complicated by the fact that the time periods are far too different. Indeed, there is a typo in the Table 2 of Parrish et al. (2014), the period over which concentrations are fitted being 1978-2011 rather than 1998-2011, as suggested by the text and the figures and confirmed directly by David Parrish (personal communication).

It is useful that the authors discuss trends in the 5th and 95th percentiles as well as means. However, these should be discussed in the context of whether there are any statistically

significant differences between those percentiles and the means. As far as I can tell, there are no such differences.

ANSWER: Yes, the referee is right, there are no statistically significant differences between the trends of mean O_3 and the trends of its 5th and 95th percentiles.

Finally, all of the error analysis for the trends implicitly assumes that the there is no covariance between data from successive years. Recent work (e.g., Lin et al., 2014) shows that such covariance can be important. Thus, the present data set must be examined for such covariance; if it is significant, then the confidence limits on derived trends must be widened accordingly. (I realize that most ozone trends analysis in the literature has not considered such covariance, but it is now apparently important.)

ANSWER: We agree with the referee that even if the problem of autocorrelation is often ignored in the literature, it has to be taken into account. As explained in our previous answer, we changed the statistical approach used for the trend analysis, and now use the Theil-Sen method. Calculations are performed with the OpenAir package (in the statistical programming language R) that allows to take into account (or not) the serial dependence of the data through a block bootstrapping method. As written above, we also briefly discuss the implications on the confidence intervals (for information, trend results without taking into account the autocorrelation are given in the Supplement). Indeed, results show that a few trends become insignificant when the autocorrelation is taken into account.

In response to the numerous points raised by the referee in its major issue 1, we propose the following overhaul of the part of Sect. 3.3 dedicated to ozone trends:

« 3.3 Trends

In order to easily compare trends between the different tropospheric layers, O_3 concentrations are normalized following the approach of Parrish et al. (2014) : (i) a quadratic least-squares regression is applied to mean annual concentrations in which the year 2000 is taken as a reference (i.e. the origin of the time series), (ii) the obtained intercept corresponds to the interpolated mean annual O_3 concentration in 2000 (hereafter designated as $O_{3,2000}$), and is used for the normalization. Trends are thus expressed in percentage of year 2000 intercept per year (hereafter referred as $\%O_{3,2000}$ yr^{-1}). The same approach is used for each season. Considering the relatively short time coverage of MOZAIC-IAGOS observations (in comparison with measurements at some historical surface sites traditionally used for long-term trend calculations), we limit the analysis to linear regressions. Besides mean O_3 concentrations, we also investigate trends of the 5th and 95th percentiles. Hereafter, all these quantities are referred as $M(O_3)$, $P_5(O_3)$ and $P_{95}(O_3)$ for clarity. The same approach is followed for CO, with the year 2004 as a reference (and the results expressed in $\%CO_{2004}$ yr^{-1}). This year is chosen because in the three tropospheric layers, the mean CO concentrations in 2004 (140, 114 and 101 ppb in the LT, MT and UT, respectively) are very close to the mean CO concentrations over the 2002-2012 period (141, 114 and 102 ppb). Ozone and CO normalized mean concentrations at the annual and seasonal scale are shown in Fig. 7 (similar plots of the 5th and 95th percentile are given in Fig. S5-S6 in the Supplement).

Trends are investigated using the non-parametric Mann-Kendall analysis combined with Theil-Sen slope estimate (Sen, 1968) which has several important advantages compared to the least-square regression, including the absence of distributional assumptions and the lower sensitivity to outliers. We use the OpenAir package (in the statistical programming language R) developed for applications in atmospheric sciences (Carslaw and Ropkins, 2012). This package provides an estimation of the uncertainties based on the bootstrap method, and allows to take into account the autocorrelation of the data. The autocorrelation of environmental

parameters is quite common (although often ignored in trend analysis), and tends to artificially decrease the uncertainties of trends, which can lead to the identification of trends that are actually insignificant (Weatherhead et al., 2002). Note that using our approach, the confidence intervals are not necessarily symmetric (around the mean slope estimate). The Theil-Sen slope estimates are reported in Table 1 (the corresponding absolute trends are reported in Table S1 in the Supplement). For information, the trend uncertainties obtained by ignoring the autocorrelation are reported in Table S2 in the Supplement.

Table 1: Annual and seasonal trends of mean O_3 and CO concentrations, 5th and 95th percentiles. Trends are estimated using the Theil-Sen slope estimate (see text). Uncertainties are given at the 95% confidence level (NS: non-significant trend). The significant O_3 trends over the 2000-2012 period are also reported in the footnotes below the Table.

		O_3 trend (%$O_{3,2000}$ yr^{-1}) (1994-2012)			CO trend (%CO_{2004} yr^{-1}) (2002-2012)		
Season	Layer	Mean	5th	95th	Mean	5th	95th
Year	UT	NS	+0.63 [+0.09;+0.99]	NS	-1.36 [-2.05;-0.74]	-1.22 [-2.27;-0.47]	-1.43 [-2.08;-0.89]
Year	MT	NS	+0.42 [+0.09;+0.68]	NS	-1.55 [-2.34;-0.72]	-1.57 [-2.52;-0.68]	-1.44 [-2.25;-0.59]
Year	LT	NS	+1.03 [+0.36;+1.62]	NS	-1.51 [-2.42;-0.44]	-1.59 [-2.58;-0.46]	-1.41 [-2.40;-0.12]
Winter	UT	+0.62 [+0.02;+1.22][1]	NS	NS[3]	-1.64 [-2.73;-0.80]	-1.39 [-2.76;-0.39]	-1.59 [-2.59;-1.06]
Winter	MT	+0.62 [+0.05;+1.22]	NS	NS	-1.50 [-2.53;-0.60]	-1.69 [-2.81;-0.31]	-1.22 [-2.24;-0.28]
Winter	LT	+0.83 [+0.13;+1.67]	NS	NS	NS	NS	NS
Spring	UT	NS	NS	NS	-1.67 [-2.83;-0.48]	NS	-1.97 [-3.13;-0.86]
Spring	MT	NS	NS	NS	-2.00 [-2.97;-0.69]	NS	-2.01 [-2.71;-1.07]
Spring	LT	NS	NS	NS	-1.91 [-2.72;-1.09]	NS	-2.22 [-4.04;-0.87]
Summer	UT	NS	NS	NS	NS	NS	-1.53 [-2.22;-0.59]
Summer	MT	NS	NS	NS	-1.83 [-3.25;-0.56]	NS	-2.29 [-3.77;-1.21]
Summer	LT	NS[2]	NS	NS	-2.31 [-3.61;-0.97]	-2.08 [-2.83;-0.76]	-2.63 [-4.54;-1.42]
Autumn	UT	NS	NS	NS	NS	NS	NS
Autumn	MT	NS	NS	NS	NS	NS	NS
Autumn	LT	NS	NS	NS	NS	NS	NS

[1]: +1.08 [+0.29;+2.06]%$O_{3,2000}$ yr^{-1} over the 2000-2012 period.
[2]: -1.00 [-3.17;-0.02] %$O_{3,2000}$ yr^{-1}
[3]: +1.22 [+0.63;+2.27]%$O_{3,2000}$ yr^{-1}

3.3.1 Ozone

All the annual and seasonal trends of the M(O_3) appear insignificant, except in winter for which a weakly significant increase is found in the three tropospheric layers (+0.83[+0.13;+1.67], +0.62[+0.05;+1.22] and +0.62[+0.02;+1.22]%$O_{3,2000}$ yr^{-1} in the LT, MT and UT, respectively). Concerning the $P_5(O_3)$, a significant increase is found at the annual scale in the three tropospheric layers (+1.03[+0.36;+1.62], +0.42[+0.09;+0.68] and

+0.63[+0.09;+0.99]%$O_{3,2000}$ yr^{-1} in the LT, MT and UT, respectively). Conversely, trends of the $P_{95}(O_3)$ are all insignificant. Note that ignoring the autocorrelation of the data leads to some additional significant positive trends, including the $M(O_3)$ at the annual scale, the $P_5(O_3)$ in winter and autumn, and the $P_{95}(O_3)$ in winter, although not in all tropospheric layers (see Table S2 in the Supplement). It is beyond the scope of this study to investigate why the autocorrelation has a stronger effect on these specific seasons or layers, but this illustrates the strong influence of the serial dependence on the trend analysis and the necessity to take it into account.

Most of the few positive trends found here over the whole period are due to an increase of O_3 in the 1990s. Over the 2000-2012 period, among the previous significant trends, the only persistent significant trends concern the $M(O_3)$ in the UT during winter (+1.08[+0.29;+2.06]%$O_{3,2000}$ yr^{-1}). However, interestingly, a few other trends become significant over that period, including the decrease of the $M(O_3)$ in the LT during the summer (-1.00[-3.17;-0.02] %$O_{3,2000}$ yr^{-1}), and the increase of the $P_{95}(O_3)$ in the UT during the winter (+1.22[+0.63;+2.27]%$O_{3,2000}$ yr^{-1}). Previous trend analysis at the alpine sites (Zugspitze since 1978, Jungfraujoch and Sonnblick since 1990) have highlighted (i) a strong increase of O_3 during all seasons in the 1980s (around 0.6-0.9 ppb yr^{-1}), (ii) a persistent but lower increase in the 1990s during all seasons except summer where O_3 has levelled off, (iii) the extension of that levelling off in the 2000s to the other seasons and a slight decrease in summer (Logan et al., 2012; Parrish et al., 2012). Qualitatively, this picture is in general agreement with our results in the lower part of the troposphere (e.g. significant increase in winter, negative trend in summer for 2000-2012). Interestingly, at regional background sites in Europe over the 2-3 last decades, Parrish et al. (2012) highlighted that O_3 trends, when they are expressed relatively to the concentration in 2000, are quite similar (around +1% $O_{3,2000}$ yr^{-1}) whatever the site and the season. Although not directly comparable due to a different (and shorter) time period, our study shows that the increase of wintertime O_3 over the 1994-2012 period is slightly lower but differences remain insignificant. At low altitudes, this increase of O_3 in winter has been observed at several sites in Europe and North America (Cooper et al., 2012; Derwent et al., 2013; Wilson et al., 2012) and is mainly attributed to a reduced O_3 titration by NO due to decreasing NO_x emissions (e.g. Ordóñez et al., 2005). The persistent positive trends found higher in altitude suggest that wintertime O_3 has increased at a large scale (if not hemispheric) (see Fig. 2). Based on the MOZAIC dataset at Frankfurt/Munich over the 1995-2008 period, at about 3 km, Logan et al. (2012) highlighted a significant increase of O_3 concentrations in winter (around +0.5±0.2 ppb yr^{-1}) and to a lesser extent in spring (around +0.25±0.2 ppb yr^{-1}), and insignificant trends during the other seasons. At the annual scale, the trend is around +0.2 ppb yr^{-1} up to 4 km and +0.4-0.6 ppb yr^{-1} between 4 and 8 km. Over the same period and using the same statistical approach (i.e. multiple linear regression of the annual cycle and the four seasonal trends from the monthly time series), we also found in the MT an increase in winter and spring, as well as at the annual scale in all tropospheric layers. However, our trends in the MT and UT (+0.19±0.10 and +0.17±0.13 ppb yr^{-1}, respectively) are lower than those reported by Logan et al. (2012) (which may be due to the fact that only the troposphere is considered here), although differences do not appear to be significant.»

REFEREE, major issue 2: I think that the treatment of the carbon monoxide trends also requires reconsideration. The authors correctly conclude in Section 3.2 that their data show statistically significant negative trends. However, the interpretation of those trends requires further thought. Specific problems include:

Careful thought must be given to whether absolute concentration trends (as the authors choose) or relative concentration trends (i.e. % change/year) should be discussed. As Figure

2 clearly demonstrates, CO is much lower in the MT and UT than in the LT. If the total emissions of CO to the atmosphere were changed by 10%, then one would simplistically expect a 10% change in the LT, MT and UT. In that case, analysis of relative concentration trends would give the same trend in the 3 layers, but analysis of the absolute concentration trends would give much larger trends in the LT, simply because the CO concentrations are larger there. This problem becomes exacerbated when one considers that CO has a background that is independent of anthropogenic emissions; this background arises from secondary CO produced from oxidation of naturally emitted methane and biogenic VOCs. Thus, changes in anthropogenic emissions do not give proportional changes in CO concentrations. I suggest that the authors reformulate the CO trends analysis as relative concentration trends, and carefully consider the influence of the background concentration of secondary CO in discussing the significance and causes of these trends. Specific issues in this regard are discussed in the following bullets.

ANSWER: We agree with the referee that it is more relevant to consider relative concentration trends rather than absolute concentration trends. In the revised version of the manuscript, this section has been largely modified. As mentioned in our previous answers (see the text added in the introduction of Sect. 3.3), we chose to keep the same approach as for O_3 with the year 2004 as a reference, and results are thus expressed in $\%CO_{2004}$ yr^{-1}. The mean CO concentrations in 2004 (101, 114 and 140 ppb in the LT, MT and UT, respectively) are very close to the mean CO over the 2002-2012 period (100, 114 and 141 ppb in the LT, MT and UT, respectively).

The discussion of CO trends was modified as follows: "As previously mentioned, CO trends are here investigated relatively to the 2004 reference year. Over the 2002-2012 period, the M(CO) at the annual scale significantly decreases in the whole troposphere, with trends of -1.51[-2.42;-0.44], -1.55[-2.34;-0.72] and -1.36[-2.05;-0.74]$\%CO_{2004}$ yr^{-1} in the LT, MT and UT, respectively. Similar negative trends are also obtained for the P_5(CO) and P_{95}(CO) in all the tropospheric layers. At the seasonal scale, the M(CO) and P_{95}(CO) show negative trends in winter, spring and summer, although not always in all the tropospheric layers, while the P_5(CO) is decreasing only in winter (in the MT and UT) and summer (only in the LT). Conversely, all trends in autumn are insignificant. Note that the results without taking into consideration the autocorrelation of the data show significant negative trends of the P_5(CO) in most layers and during all the seasons, except autumn (see Table S2 in the Supplement). These results are in general agreement with previous studies in Europe (e.g., Karlsdóttir et al., 2000; Novelli et al., 2003; Dils et al., 2009; Worden et al., 2013). Based on satellite observations, Worden et al. (2013) highlighted over Europe a decrease of the CO total columns, around -1.44+0.22% yr^{-1} with MOPITT over 2001-2011 and -1.00±0.33% yr^{-1} with AIRS over 2003-2011, thus in the range of our results over Frankfurt. Over the 1995-2007 period, Gilge et al. (2010) found trends of -3.36±1.08 and -1.51±0.64 ppb yr^{-1} (reduced to -2.65±0.04 ppb yr^{-1} by filtering the background values (Zellweger et al., 2009)) at two alpine sites from the WMO GAW network, in reasonable agreement with our absolute Theil-Sen slope estimates at Frankfurt/Munich in the LT and MT (-2.24[-3.59;-0.65] and -1.85[-2.79;-0.85] ppb yr^{-1}). »

p. 23855, beginning on line 18 - The authors state: "This decrease of mean CO concentrations is actually associated to a decrease in occurrence of high CO episodes". This statement is supported by a statistically significant difference between the trends of the mean and the 95th percentile only in the LT, not in the MT or UT. This should be clarified.

ANSWER: Indeed, the differences of absolute concentration trends are significant only in the LT. However, relatively to the CO concentration in 2004, the differences between

these trends are insignificant. The text was modified accordingly (see our answer to major issue 2).

p. 23855, beginning on line 18 - The authors state: "Interestingly, trends highly depend on altitude, with decreasing negative values while one moves away from the surface." At best, the difference is statistically significant for the absolute concentration trends only between the LT and UT. When relative concentration trends are considered, even that difference likely will become statistically insignificant.
ANSWER: Indeed, considering relative concentrations, none of these differences of trends are statistically significant. The text was modified accordingly (see our answer to major issue 2).

Discussion beginning on p. 23856, line 8 - This largely deals with statistically insignificant trends. Such discussion is not useful and should be eliminated, or put on a more solid statistical basis.
ANSWER: This part of the discussion was removed.

The comparison of the MOZAIC-IAGOS LT data with those from German WMO stations is useful. Perhaps it should come early in the discussion of the CO data.

ANSWER: The discussion about the correlations between the MOZAIC-IAGOS LT data and the GAW data is moved with more details in Sect. 3.2: "As previously mentioned, the local emissions both from the neighbouring agglomeration and from aircraft — on tarmac and/or during the take-off/landing phases (in case the MOZAIC-IAGOS aircraft closely follows other aircraft) — may add some variability depending on the local dispersion conditions and thus influence the CO measurements in the LT. To assess more precisely the spatial representativeness of these MOZAIC-IAGOS LT data, some surface measurements are available at four German stations from the World Meteorology Organisation (WMO) Global Atmosphere Watch (GAW) database: Hohenpeissenberg (47.8°N, 11.0°W; at 50 km South-West from Munich), Neuglobsow (53.1°N, 13.0°W), Schauinsland (47.9°N, 7.9°W), Ochsenkopf site (50.0°N, 11.8°W). We investigate the correlations of both the seasonal and annual mean CO concentrations between the MOZAIC-IAGOS measurements (in the LT) and these surface observations (see Fig. S4 in the Supplement). At the annual scale, a reasonable agreement is found, with correlations (R) of annual mean CO between 0.56 (at Neuglobsow) and 0.81 (at Schauinsland). Considering the monthly time series, correlations are improved due to the seasonal variations (from 0.61 to 0.90). At the seasonal scale, correlations remain satisfactory in winter, with values between 0.65 and 0.87. However, results appear more contrasted among the GAW stations during the other seasons. In spring, all correlations are above 0.69 except Hohenpeissenberg (0.52). In summer, both Hohenpeissenberg and Ochsenkopf have a low correlation with MOZAIC-IAGOS (0.34 and 0.52, respectively), while very high correlations (above 0.96) are found for the two other stations. In autumn, the correlation is very low for Neuglobsow (-0.21), moderate for Ochsenkopf (0.51) and satisfactory for Hohenpeissenberg and Schauinsland (0.73 and 0.85). »

The authors may wish to discuss the trends of the 5th percentile of CO.
ANSWER: A discussion of the 5th percentile of CO was added in Sect. 3.2.2.

REFEREE, major issue 3: Equation (2) and its discussion are incorrect. For example, if $\psi = 0$*, then* $\psi_{month} = 365/4 = 91.25$*. Thus the equation gives the phase in days, not months. If the*

365 is replaced by 12, then the units are in months. However, then if ψ = 0, ψ_{month} = 3, which corresponds to the maximum at the end of March. Thus, ψ_{month} = to 0 and 5.5 (not 1 and 6.5) correspond to an ozone peak on 1 January and the 15 June, respectively. This misinterpretation has led to an error of 1 month in the date of ozone ordinate in Figure 6. It is clear from the left panel of Figure 3 and the discussion in Section 3.2 that the seasonal peak of the sine function fit to the seasonal cycle should have a maximum in mid-June to early July, not one month earlier as Figure 6 indicates.

ANSWER: First point, indeed, it is 12 and not 365 in equation (2). The error only concerns the text (and is corrected in the revised version of the manuscript), but does not impact all the calculations and figures.

Second point, to our opinion, there is no other error in equation (2) because in this equation and in all our calculations, the time variable (t) is expressed in month ranging between 1-12, and not 0-11 as implicitly assumed by the referee. This information was missing and we propose to add it as follows (page 23858, line 5): « with t the time (in months, values ranging between 1 and 12) ». **We agree that it is somewhat counter-intuitive to see in Fig. 6, for instance in UT, a maximum around the 6 June, and a peak in July in the averaged monthly profile (Fig. 3). But as already (likely too) briefly discussed in the text (page 23859, lines 19-23), this is due to the fact that the monthly variations of ozone do not exactly follow a sine profile. This can be demonstrated if we perform a similar sine fit on the averaged monthly profiles shown in Fig. 3 (this figure will be included in the Supplement):**

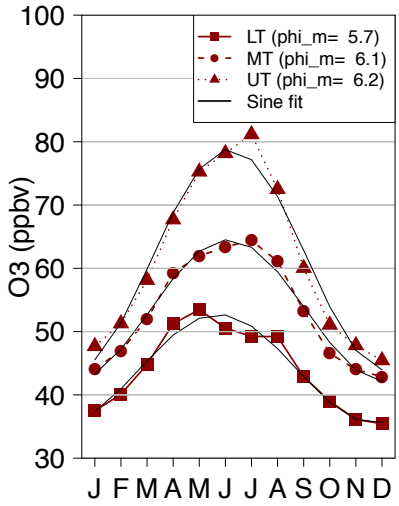

Figure S7: Averaged O_3 monthly profiles in the three tropospheric layers (as in Fig. 3), and their corresponding sine fits (performed similarly than explained in Sect. 4.1). The obtained ψ_{month} values are reported for each layer in the legend, and expressed in decimal month starting from 1 (i.e. for instance, a ψ_{month} of 6.2 corresponds to the 6th of June). They are consistent with the phase indicated in Fig. 6 (end of May in LT, early June in MT and UT).

Compared to the sine fit of the O_3 seasonal profile in the UT, the maximum of July is an "outlier", and its weight is thus low compared to all the other months (the concentration of which closely follows the sine function). As indicated in the legend, this fit gives in the UT a ψ_{month} of 6.2, which corresponds to the 6th of June, in agreement with the results shown in Fig. 6. As mentioned in the text (page 23859, lines 24-25), this is not a so important problem in our analysis since we are interested in the relative changes of the phase.

In order to clarify this point in the text, we propose the following modification: « It is worth noting that the dates of maximum O_3 obtained here (for instance, the 28 May in the LT and the 6-7 June in the MT and UT, over the 1995-2003 period) do not exactly correspond to

those given by the averaged seasonal variations, in particular in the MT and UT where O_3 is maximum in July (see Fig. 4). This is due to the fact that the O_3 seasonal pattern does not follow exactly a sinusoid. For instance, in the UT, the July maximum is driven by several high episodes during this month (with a high interannual variability). However, compared to the rest of the year, the weight of this July month is minor and thus it does not influence so much the fitted sinusoid that gives a maximum earlier in the year. This can be clearly demonstrated by fitting a sinusoid on the averaged seasonal profiles shown in Fig. 4 (see Fig. S7 in the Supplement). In addition, this is not so problematic since we are here more interested in the relative changes of the seasonal pattern than in the seasonal pattern itself. ».

REFEREE, major issue 4: It is critical to clearly explain the description of the error calculation (p. 23858, lines 13-15). In particular, how is the "number of points" calculated? Very importantly, it is not the number of points shown in Figure 6, since the points co-vary to a high degree as each of the points has 90% of the same data as the previous or following point. The "number of points" must be properly calculated as the number of independent points, i.e. the degrees of freedom. There are really only about 2 independent pieces of information in each curve in Figure 6. I suspect that the authors may be underestimating the errors, and that the apparent shift of phase and change of amplitude of the seasonal cycle in Figure 6 are not statistically significant. If this suspicion is correct, then all of the discussion in Section 4 will require major modification. Perhaps an approach similar to that given for the seasonal amplitude as discussed in "Significant Issue" #9 below will provide a useful approach. Also the error shown in the figure does not indicate the precision of the date of the seasonal maximum; this is a critical piece of information.

ANSWER: The text probably lacks of clarity on this point, as we are not sure to talk about the same kind of errors with the referee. The error shown on this figure (blue line) simply corresponds to the error of the sinusoidal fit. It is simply expressed as the square root of the square distance between points from the fit sine function and points from the observed monthly time series (i.e. a perfectly sine shape in the monthly observations would lead to a nil error) and normalized by the number of these points over the 10-year window (i.e. 10x12=120 points for a 10-year period without any gap):

$$\epsilon = \frac{\sqrt{\sum_{i=1}^{n}(o_i - m_i)^2}}{n}$$

With o_i an element of the observed monthly time series, m_i an element of the sine fit, and n the number of available monthly data over the 10-year time period. We noticed an artefact in our calculations, the square root being applied two times. This bug was corrected, which does not substantially affect the results since all points are affected the same way.

We agree that considering 10-year moving periods induces a strong co-variation between two consecutive data subset, but at the end, in this discussion, we are interested in the differences between the first (1995-2004) and the last (2003-2012) 10-year periods for which we only have 2 years of overlap. In order to cancel this overlap, we consider 9-year periods in the revised manuscript.

Note also that in the revised version of the manuscript, the seasonal shifts are now derived from a linear model that take into account the uncertainties on the date of maximum O_3 (in the initial version, the shifts were quantified by a simple linear regression, in which the uncertainties affecting each point were not taken into account). The procedure is derived from the numerical recipes of Press et al. (2007). The following text was added: "For each point (that corresponds to a 9-year time period), the 95% confidence level is derived from the linear regression for both the amplitude and the phase,

and represented by a coloured area on Fig. 7. These uncertainties need to be taken into account in order to fully quantify the trend and its corresponding uncertainty. Instead of using a simple linear regression, we thus follow the procedure precisely described in the numerical recipes of Press et al. (2007) (see equations 15.2.4 to 15.2.12) in which all uncertainties are considered in the calculation of the trend, its uncertainty and its goodness-of-fit (following Press et al. (2007), we consider the derived linear model as believable only when the goodness-of-fit exceeds 0.1). Note that the trends obtained with this procedure are always very close to those derived from a simple linear regression, but their uncertainties are substantially higher (up to a factor of 2-3 higher). We focus here on the changes of amplitude and phase. »

REFEREE, major issue 5: In Section 4.2, I suggest that the longest possible time periods, not just 1995 to 2012, be utilized for the sonde and surface measurements. This is because the longer the time period, the more precise the trend determinations. I realize that the authors wish to investigate the same time period in all of these data sets, but the primary obstacle they face is lack of precision in the determinations. If they were to ever find trends in the different data sets that were statistically significantly different, then they would return to the issue of differing time periods, but that is not likely to be a problem.

ANSWER: This paper focus on the MOZAIC-IAGOS data at Frankfurt, and this section on ozonesondes only serves as a comparison. This explains our choice to consider the same time period, even if considering a larger period may decrease the uncertainties. In the revised version of the manuscript, we added the shift values over the whole period of available data, both in the Table and in the discussion. Note that following the recommendation of the referee #2, the Sect. 4.2 is moved to the Supplement and briefly discussed in the revised version manuscript. The paragraph was modified as follows: **"**Vertical profiles obtained by ozonesonde are more interesting for a comparison with our results above Frankfurt/Munich. We investigate seasonal changes at four tropospheric levels, delimited by the following pressures: 900, 800, 600, 400 and 300 hPa. Seasonal shifts are calculated with the same approach as in Sect. 4.1 (over time periods of 9 years). Results are reported in Table S3.

Table S3: Seasonal shifts between 1995-2004 and 2003-2012 at the three ozonesonde sites and for different pressure levels, and uncertainties at a 95% confidence level. Negative values indicate a decreasing seasonality phase (i.e., toward an earlier O_3 maximum). For each station, the shifts over the whole period of available data (indicated with the station name) are reported in brackets.

Pressure level	Seasonal shift (days decade^{-1})		
	Hohenpeissenberg (1987-2012)	De Bilt (1993-2012)	Uccle (1994-2012)
300-400 hPa	NS (NS)	-9.4±8.0 (-10.1±6.0)	+7.2±6.5 (+6.7±5.7)
400-600 hPa	-4.2±2.7 (-2.9±1.2)	-6.9±4.0 (-5.0±3.0)	NS (NS)

600-800 hPa	-9.0±3.6 (-5.0±1.5)	-19.6±5.7 (-17.7±4.2)	-11.5±4.6 (-11.4±4.0)
800-900 hPa	-7.5±4.2 (-5.7±1.7)	-19.3±7.7 (-19.9±5.7)	-17.5±7.0 (-17.4±6.1)

Over the period of MOZAIC-IAGOS data (i.e. 1995-2012), the seasonal shifts are statistically significant at Hohenpeissenberg in the three first levels. The shift does no monotonically decrease with altitude as at Frankfurt/Munich, but reaches its minimum value at 400-600 hPa. Note that over the whole period of available data (1987-2012), both trends and uncertainties are reduced but seasonal shifts remain quite similar (i.e., differences are insignificant). Conversely, results at De Bilt show a high significant shift at the two first levels (around -20 days decade^{-1}) and a lower one above 600 hPa (again without significant differences with the results over the 1993-2012 period). The shift observed at Uccle is also strong in the two first levels, but turns insignificant at 400-600 hPa and goes in the opposite direction in the higher most level (i.e. toward later O_3 maximum). It is worth noting that due to a much lower number of vertical profiles compared to MOZAIC-IAGOS (on average: 11, 4 and 9 ozonesondes per month at Hohenpeissenberg, De Bilt and Uccle, respectively), these results may be less reliable than at Frankfurt/Munich (Saunois et al., 2012). » **Note also that an error in the mean number of ozonesondes at Uccle was corrected (from 12 to 9).**

Significant issues:
REFEREE, significant issue 1: Paragraph labeled 2) on p. 23848 - I do not fully understand the description of the criteria for defining the tropopause. In Figure 1, the tropopause is indicated to be at about 8.8 km, which appears to be the altitude at which the PV rises to 2 pvu. Yet, the 1.8 km thick layer with a minimum PV exceeding 2 pvu is above 8.8 km. I assume that the tropopause is set to the bottom of that 1.8 km layer. If this is correct, the description should be clarified.
ANSWER: Yes, the referee is right, in this case, the tropopause is set to the bottom of that 1.8 km layer. We give a precision on this point (page 23849, lines 3-5): « […] and the tropopause is considered to be reached only when the minimum PV over that window exceeds 2 pvu (this value of 60 being empirically chosen to handle most of these situations) (or when the previous criteria is fulfilled) and is set to the bottom of that window. » **Note that this situation is not illustrated in Fig. 1 since on this flight, there is no deep stratospheric intrusion into the troposphere (PV values reach 2 pvu at 8.8 km and remain greater than 2 above).**

REFEREE, significant issue 2: Since ozone concentration increases with altitude on average, it seems that truncating a vertical profile when the aircraft exceeded a 400 km distance from the airport would bias the average ozone low for the UT for that flight. Was such a potential bias considered and evaluated? A brief discussion of this issue should be included. One question is whether this bias varies with season, which could potentially distort the seasonal cycle derived for the UT.
ANSWER: As mentioned in the text, measurements along aircraft take-offs and landings are not sensu stricto vertical profiles as the aircraft ascents and descents are progressive. In our study, we make the assumption that the ozone variability is higher on the vertical dimension than on the horizontal one. As a guarantee, we thus decided to limit the horizontal range to 400 km. As mentioned by the referee, O_3 concentrations

increases with altitude on average, and one can think that increasing the distance threshold would necessarily increase the O_3 concentrations in the UT. However, a sensitivity case with a distance threshold of 800 km actually gives differences of mean O_3 concentration in the UT below 2% at the annual scale. At the seasonal scale, differences remain below 3% in winter and autumn, and below 1% in spring and summer (at the monthly scale, differences remain below 9% in 95% of cases, and below 18% in any case). We added in Sect. 2.2 the following sentence: "A sensitivity test with a distance threshold of 800 km leads to differences of mean O_3 concentration in the UT below 3% at the seasonal and annual scale. »

REFEREE, significant issue 3: p. 23852, beginning on line 22 - The authors note that "In comparison with ozone, the daily CO variability at the monthly scale is lower and similar in the three tropospheric sublayers (around 14–16 %)." It is worth mentioning that this is qualitatively expected given the differing lifetimes of these two species, as noted by Junge (1974). A brief discussion of this issue and a reference here would be useful.
ANSWER: We added the following text (page 23852, line 24): « […] (around 14-16%). Such a result is expected due to the longer lifetime of CO in comparison with O_3, which leads to a higher regional and hemispheric background (Junge, 1974). »

REFEREE, significant issue 4: p. 23852, beginning on line 28 - The authors note that "Such a seasonal pattern is rather consistent with maximum emissions and minimum photochemical activity usually occurring around winter combined with a maximum secondary formation in summer." Given the long (1 to 3 months) lifetime of CO, it is expected that its seasonal cycle is delayed with respect to that of the emissions and photochemical activity. This issue should be briefly discussed.
ANSWER: We added the following text (page 23853, line 2): « Such a seasonal pattern is rather consistent with satellite observations (Worden et al., 2013). It results from a maximum of primary CO emissions in winter (at northern mid-latitudes) associated to a limited photolysis which increases the CO lifetime and allows its accumulation in the atmosphere. This delays the maximum of CO concentrations to late winter/early spring, after which concentrations start to decrease due to a more effective photolytic destruction in summer, despite an enhanced secondary formation from biogenic compounds. »

REFEREE, significant issue 5: In Equation (1), the authors are defining the variable "a" as the peak-to-peak amplitude, while standard mathematical notation defines the magnitude of the sine function as the zero to peak amplitude. This issue should be mentioned.
ANSWER: We modified the equation, the values in Sect. 4.1 and the Fig. 6.

REFEREE, significant issue 6: p. 23858, lines 6-7 - I do not understand the significance of this note. It is true, but it is not clear to me that this information is ever used in the paper. Perhaps it should be removed.
ANSWER: This note was removed.

REFEREE, significant issue 7: The authors must clearly define their use of the term "baseline", and use it consistently through the paper. It is first used on pg. 23845 to indicate air "coming from the Atlantic Ocean". In the introduction to Section 3.3 it is equated to the 5th percentile of the data. In Section 4.1 it is defined as the average ozone over one or several years. During the HTAP program, "baseline" was carefully defined to be approximately the same as it is first used in this paper. I suggest that this use be maintained here, and that the

other two usages of the term be replaced by different terms. Importantly, the air "coming from the Atlantic Ocean" may well contain ozone concentrations much higher than the 5th percentile.

ANSWER: We followed the recommendation of the referee, and removed all inadequate occurrences of "baseline". In the discussion of the ozone seasonal cycle (Sect. 4.1), we replaced "baseline" by "offset concentration". In the introduction, we modified the text as follows: "baseline (originating from the Northern Hemispheric marine boundary layer) air masses**".**

REFEREE, significant issue 8: Section 4.1.1 gives no information beyond what was already discussed in Section 3.2.1. I suggest it be removed.

ANSWER: This section was removed.

REFEREE, significant issue 9: p. 23859, line 5 - Do the errors given correctly represent 95% confidence limits? If so, then the difference in the two numbers is 3.3 ± 3.0 ppb, which is significant at two sigma.

ANSWER: Yes, the referee is right, values were here given at a 95% confidence level, and the amplitude decrease was thus significant at 2 sigma (3.2±3.0 ppb). The sentence was corrected.

REFEREE, significant issue 10: Section 4.3 occupies 5 pages of text and does not seem particularly relevant to the main focus of the paper. I am not an expert in atmospheric transport, but I cannot find anything really new in this section. I suggest that this section be eliminated, or at least greatly shortened, limiting the discussion to only what is required to discuss the shift in the seasonal cycle and the long term trends, and perhaps to what is novel in this analysis.

ANSWER: This section was included in order to describe more quantitatively the origin of the air masses sampled along the vertical profiles above Germany, and the associated seasonality. To our opinion, such presentation of the geographical origin of the air masses sampled by MOZAIC-IAGOS aircraft is quite original. However, we agree with the referee that the discussion is too long and does not provide quantitative explanations of the seasonal shift. In the revised version of the manuscript, we removed this section but kept the Fig. 8 in Sect. 2.3 with the following discussion: « The FLEXPART Lagrangian particle dispersion model (Stohl et al., 2005) is used to investigate the origin of air masses sampled by the aircraft in the different tropospheric layers above Frankfurt/Munich. Input meteorological data are taken from the ECMWF operational analysis (00:00, 06:00, 12:00, 18:00 UTC) and forecasts (03:00, 09:00, 15:00, 21:00 UTC) and interpolated on a 1°x1° global longitude-latitude grid. The methodology used here basically consists in releasing along each vertical profile 1000 particles every 10 hPa and following them backward in time during 20 days. This duration corresponds approximately to the time during which a pollution plume is expected to remain significantly higher than the tropospheric background (Stohl et al., 2003a). The FLEXPART model computes the particles' residence time, sometimes referred as the potential emissions sensitivity (PES). Output are given on a 1°x1° global longitude-latitude grid, over 1-km width vertical layers up to 11 km plus a remaining layer ranging from 11 to 50 km (i.e., 12 vertical layers). The footprint PES between 0-1 km is presented in Fig. 2 for each tropospheric layer, averaged over the 1994-2012 period. As expected, air masses sampled in the LT spend most of their time in the European boundary layer (mostly in France, Germany, Benelux, England). In the MT, the influence of Europe persists but the influence of North America is greatly enhanced. In the UT, the PES is the

highest over North America but stronger winds at these altitudes (e.g. jet streams) also allow a fast transport of air masses from Asia.»

Minor issues:
REFEREE, minor issue 1: p. 23843, line 18 - The authors write "... several sinks are at stake in the troposphere...". The meaning is not clear; perhaps better is "... several sinks are active in the troposphere...". There are other similar issues throughout the paper. A native English speaker should edit the paper by for these small issues. However, overall I did find the writing generally clear and even elegant.
ANSWER: Yes, the referee is right about the meaning of the sentence. This paragraph was shorten following a comment of referee #2.

REFEREE, minor issue 2: The primary reference for instrumental details (Nédélec et al., 2015) is not included in the list of references.
ANSWER: The reference was added.

REFEREE, minor issue 3: p. 23851, line 26 - The authors are evidently discussing the "Highest monthly mean mixing ratios"; this should be made explicitly clear.
ANSWER: Yes, this sentence refers to mean mixing ratio, the suggested modification was applied.