ANSWERS REFEREE #2

We thank the referee for its numerous and constructive comments on our study.

REFEREE: The manuscript investigates the climatology of O3 and CO observed onboard passenger aircraft (projects MOZAIC and IAGOS) in the upper troposphere over Frankfurt (Germany), with emphasis on temporal trends. The used approach is also applied to O3 data recorded by 14 stations in Europe. The authors invested quite a lot of work, but in my opinion tried a bit too much. Instead of concentrating on one specific topic, the paper appears overloaded and does not (cannot) sell a clear message, yet. Major problem here is that it requires comprehensive and often tricky analyses to squeeze out statistically significant and for the reader plausible trends in noisy atmospheric (O3) data that is influenced by many processes at once. Thus, most of the inferred trends are not significant and also the trends derived for the European O3 GAW stations are different, partially contrasting and do not give a common picture. I do not claim that the different datasets have to and will indicate such a common picture, but it is somewhat venturous to believe that the application of one approach will tease out a common picture in all the different European monitoring sites analyzed. My second major concern is the often improper English verbalism; a problem that can be solved easily.

ANSWER: The first part of our study investigates the climatological vertical profiles, the seasonal variations and the trends of O_3 and CO in the troposphere, focussing only on the MOZAIC/IAGOS dataset. The second part focuses on the changes affecting the O_3 seasonal profile, and in particular its phase. The comparison between MOZAIC/IAGOS and some other datasets (ozonesondes and GAW surface observations) is only done in this latter analysis of the O_3 seasonal shift. That comparison was included because to our knowledge, only two other studies have tried to quantify this seasonal shift (Parrish et al., 2013 and Cooper et al., 2014). Indeed, contrasting results are obtained, but to our opinion, it is still worth providing them in the paper. However, we agree that it is better to focus on the MOZAIC/IAGOS dataset, so the Sect. 4.2 is moved to the Supplement and more briefly discussed in the paper. Concerning the English verbalism, we carefully checked and corrected many wordings.

Major concerns:

REFEREE, major issue 1: The manuscript focuses in my opinion too much on possible temporal trends in the data and loses this game due to the difficulty of the problem. Already the abstract limits itself on the description of the trend analysis and thus forgot to describe the measured distributions, vertical profiles, and seasonal variations (Figs. 1-4). Shift the focus more to the latter topics and describe them in more detail in the abstract and the conclusions.

ANSWER: In the revised version of the manuscript, the Sect. 3 was greatly modified. More details were added in the description of the climatological vertical profiles (Sect. 3.1) and seasonal variations (Sect. 3.2). The main conclusions are added in both the abstract and the conclusion. In order to take into account the numerous comments of both referees, the analysis of trends (Sect. 3.3) was completely modified. We invite the referee to look at our answers to the major issue 1 of referee #1 and to the major issue 2 of referee #2.

REFEREE, major issue 2: Trend analysis. As shown in a few papers, surface ozone maximized around the year 2000, with a very flat plateau or transition from increasing concentrations before 2000 and decreasing ones thereafter. Thus, it makes little sense to approximate the data between 1994 and 2012 by a linear regression. Thus, I question if table 1 makes sense? Moreover, you should explain much better why you have applied in addition a

10-year moving average. Were other techniques not successful? And why other datasets (listed in the text), usually taken at ground, but some also around the tropopause, indicate more significant trends? Is the reason the different type/nature of dataset (too bad statistics, too high atmospheric variability,...) or the different/unsuitable type of analysis (unsuitable choice of (too variable) regions, unsuitable technique,...)? That is, you have to give the reader the possibility to assess the results. Why for instance seems the trend derived from the ground-based stations to be less significant compared to the results in other relevant publications? It makes me especially sceptical, because J. Logan (2012) teased out a more significant temporal course in the same (MOZAIC) dataset. Why? Check also Lin et al (JGR, 2015, 2015GL065311).

ANSWER: To our opinion, as a first approach, it is interesting to investigate the trends over the whole period of available data (i.e. 1994-2012), as it was previously done for instance by Logan et al. (2012) (for 1994-2008). However, we agree with the referee that several studies have highlighted a levelling-off of O_3 concentrations in the 2000s. Indeed, considering the 2000-2012 period, almost all O_3 trends turn insignificant in the MOZAIC-IAGOS dataset. There are a few exceptions, including a persistent increase of O_3 in the UT during the winter (for both the mean and the 95th percentile) and a decrease of the mean O_3 in the LT during the summer. We added that information in the text.

As previously indicated, the trend section has been largely modified in the revised version of the manuscript. In particular, the analysis of trends over 10-year moving time periods was removed. This figure was initially included in order to progressively follow the changes of trend along the period. However, as noted by the referee #1, such approach is not so relevant in our case, in particular because of the high positive anomaly in 2003 that contaminates the trends in both the first and the last decade.

Concerning the agreement with Logan et al. (2012), it is worth noting that there are many different statistical approaches for the trend analysis that can give contrasting results. For instance, applying a linear regression on a deseasonalized monthly time series can give lower uncertainties than considering an annual time series. We made trend calculations following the method of Logan et al. (2012), and the agreement is better. In particular, our annual trends of mean O_3 concentrations are all significant. Some discrepancies persists in the MT and UT at the annual scale, our trends being lower than those reported by Logan et al. (2012). It is rather tricky to understand the origin of these discrepancies but to our opinion, they may be due to the fact that the stratospheric ozone is not considered in our data. Concerning the seasonal trends in the lower part of the troposphere, our results are also in general agreement with those given by Logan et al. (2012) (increase in winter and to a lesser extent in spring).

However, in the revised version of the manuscript, we decided to change the statistical method used for the trend analysis by considering the non-parametric Mann-Kendall approach with Theil-Sen slope estimates. This approach is less powerful than the least-square estimate, but relies on much less assumptions. In addition, following the recommendations of referee #1,we now take into account the autocorrelation of the data, which again increases our uncertainties. Thus our results are not directly comparable with these previous studies, but we still kept a discussion in our revised manuscript in which the text is modified as follows: « 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 3: In this respect, I do not understand why you have additionally included the trend analysis for the ground-based stations. In my opinion, this is counterproductive. You overload the paper and as also the result of this additional analysis appears to be less clear than in other studies, the reader comes somehow to the conclusion that the analysis applied is not the best. I suggest to skip this part and to limit yourself to a comprehensive comparison of your results gained from the MOZAIC/IAGOS data with other studies.

ANSWER: In Sect. 3, the annual and seasonal concentration trends are investigated only with the MOZAIC/IAGOS dataset. The GAW stations are used only for the comparison of the variability of CO with the MOZAIC/IAGOS observations in the LT, but no concentration trends have been calculated (and this comparison is judged useful by the referee #1). In Sect. 4, we investigated the changes of the ozone seasonal cycle in the MOZAIC/IAGOS dataset, leading to the following main conclusion: there is a strong seasonal shift in the lower troposphere that decreases in altitude. In this context, to our opinion, it is interesting to investigate in central/western Europe (i) how strong can be the seasonal shift at surface sites (for comparison with our result in the LT), and (ii) how does it depends on the altitude in the ozonesonde data (for comparison with our results in the three tropospheric layers). Many other studies have investigated the seasonal and annual trends of O_3 concentrations in Europe, but not in terms of seasonal shift as we did. Therefore, we do not have other published results to compare with except the two

papers already mentioned in the text (i.e. Parrish et al., 2013 and Cooper et al., 2014). However, as previously mentioned, we agree that this part can be moved to the Supplement and discussed more briefly in order to focus on the MOZAIC/IAGOS dataset in the paper.

REFEREE, major issue 4: Improper English. The wording is often lax and the grammar sometimes wrong. You often piece words together, e.g. "seasonal cycle phase", "vertical profile data selection" or "ozone seasonal changes". Often articles are missing. Sometimes you use wrong expressions, e.g. O3 "peaks", although the seasonal cycle does not show any peak, but just maximizes in certain months. In my remarks below, I sometimes just wrote "-> change" in such cases.
ANSWER: We carefully checked and corrected the wording and grammar.

Minor remarks:
REFEREE: Title: "... between 1994–2012" doesn't work
ANSWER: The title was changed in: "Characterizing tropospheric O_3 and CO around Frankfurt over the 1994-2012 period based on MOZAIC-IAGOS aircraft measurements"

REFEREE: Abstract. L.2-7. Far too long sentence.
ANSWER: The sentence was split in two sentences (and the number of flight corrected): « This study investigates the variability and trends of both species at several tropospheric layers above the Frankfurt and Munich airports. About 21,300 flights have been performed over the 1994-2012 period, which represents the densest dataset in the world (about 96 flights per month on average). »

REFEREE: Intro. The first introductory part (until p.23844, l.2) is far too long
ANSWER: This part was reduced to: « In the troposphere, O_3 is formed by photochemical reactions implying various compounds including volatile organic compounds (VOC), carbon monoxide (CO) and nitrogen oxides (NO_x), and it can be removed by photolysis, dry deposition and uptake on aerosols (Moise and Rudich, 2000, 2002). Despite the considerable scientific achievements made during the last decades, the O_3 budget remains difficult to quantify precisely (Wu et al., 2007). Major uncertainties are related to lightning NO_x production, isoprene biogenic emissions and degradation chemistry, biomass burning emissions, water vapour concentrations and stratosphere-troposphere exchanges (Stevenson et al., 2006). This leads to a large heterogeneity of the O_3 abundance and variability in the troposphere, making it difficult to draw a simple and global picture of the O_3 present-day concentrations and trends.»

REFEREE: p.23846. L.11. "trajectories" -> flight routes
ANSWER: The modification was applied.

REFEREE: p.23846. L.1ff. "In this paper, tropopause is considered in its dynamic sense". Very lax wording for referring to the dynamical tropopause. -> Change.
ANSWER: The sentence was modified as follows: « In this paper, we consider the dynamical tropopause, delimited by a potential vorticity (PV) of 2 pvu »

REFEREE: p.23848, L.20ff. DT is not a good tropopause, as PV is a model derived quantity. Also the DT threshold value is quite variable, see Kunz et al. (JGR, doi:10.1029/2010JD014343, 2011). Moreover, there is often a mismatch between model and *the real synopsis, also because the PV data is linearly interpolated between PV fields 6-hours apart. Best example is Fig.1. The real tropopause is not at 9 km as indicated, but around 11 km, namely there where O3 and CO show an abrupt step. Here, read the description by Sprung & Zahn (2010) where a O3-based height relative to the tropopause is suggested and also compare with Thouret et al. (2006) who found a seasonal variation of O3 at the TP.*

ANSWER: We agree with the referee that various uncertainties are associated to this choice, but it is still considered as one of the reference methods to define the tropopause, widely used in the community. We agree that on the example given in Fig. 1, the mismatch is quite important, but it is not always the case. Moreover, there are numerous cases where the aircraft does not reach the tropopause and where it is thus not possible to assess the tropospheric layer (MT or UT) to which observations belong. In such cases, one can only rely on model derived information (more precisely the pressure at which PV reaches 2 pvu). This situation occurs in about 1/3 of the cases (and not 1/5 as written in the text, there was an error that was corrected).

We added some precisions on these uncertainties in the new version of the manuscript: « It is worth noting that the determination of the tropopause altitude is associated to several uncertainties. Some uncertainties arise from the choice of the method used to locate the tropopause. For instance, the ozone criteria may give a lower dynamical tropopause (DT) compared to the thermal method (Bethan et al., 1996). In the determination of the DT altitude, other uncertainties can arise from the choice of a constant PV to locate the DT. Indeed, Kunz et al. (2011) showed that the PV values at the DT can vary between 1.5 and 5, with higher PV values in summer than in winter. In our case, there are also uncertainties related to the fact that the PV is here a modelled variable. In addition, it is linearly interpolated between PV fields 6-hours apart, which does not allow to entirely catch the variability of the DT. A good example is given in Fig. 1 where the abrupt O_3 increase (corresponding to the tropopause) occurs 2 km above the DT derived from PV values. However, our approach allows to assess in which layer (MT or UT) observations belong even when the tropopause is not reached by the aircraft (within the 400 km around the airport). It is beyond the scope of this study to investigate in more details the influence of the method used to locate the tropopause. Above the Frankfurt airport, a majority of vertical profiles (63%) reach the tropopause while most of the remaining profiles (36%) are selected according to the distance criteria. A similar proportion is found at Munich (63 and 35%, respectively). »

REFEREE: p.23848. "... with the Frankfurt–Boston flight of...". Please, avoid such a lax wording.

ANSWER: The sentence was modified as follows: « This is illustrated in Fig. 1 with the flight from Frankfurt to Boston on the 19 March 2002 during which the DT altitude is estimated at 8.8 km. »

REFEREE: Fig.2. Choose CO axis of 0-300ppb with 50ppb ticks

ANSWER: The modification was applied.

REFEREE: p.23849. L.15. Do you mix analyses and forecasts? Does this make sense? Explain

ANSWER: Yes, in the FLEXPART simulations, it is important to have meteorological data at a good time frequency. We thus use both analyses and forecasts in order to have data every 3 hours. This approach is widely used by the FLEXPART community (Stohl and James, 2004; Stohl et al., 2005).

REFEREE: p.23850. L.20. "... it is likely driven by intense shallow and transient exchanges." Do not understand what you mean
ANSWER: These two types of stratosphere-troposphere exchanges are defined in Stohl et al. (2003). Shallow exchanges are limited to the tropopause region. Transient exchanges correspond to stratospheric air masses that enter the tropopause during a short time before going back to the stratosphere. The sentence was simply modified as follows: « […] where it is likely driven by intense shallow and transient exchanges between the stratosphere and the troposphere (Stohl et al., 2003b). »

REFEREE: p.23851. L.1. "that last season". Change p.23851. L.9 and 13. "variability" -> "variation" and in the text p.23851. L10. "monthly profiles" -> "seasonal variation"
ANSWER: The sentence was modified as follows: « High seasonal variations are observed close to the surface, with concentrations in the first kilometre ranging from 156 ppb in summer to 233 ppb in winter on average. The increase during winter is likely due to a lower vertical mixing and higher emissions. » **The two other modifications were also applied.**

REFEREE: p.23851. L.17. "(including a secondary maximum in August)". Will not be significant, right?! -> Skip
ANSWER: This part was removed.

REFEREE: p.23851. L.23. On average
ANSWER: The modification was applied.

REFEREE: p.23852. L.13/14. What you mean? What is a significant O3 m.r.?
ANSWER: In this paragraph, we compare the highest monthly values between the three tropospheric layers, and show that they do not always coincide. In the sentence mentioned by the referee, "significant**" is not the appropriate term and was replaced by "**high**" in the revised version of the manuscript.**

REFEREE: p.23862. L.18. "Variation" and in text
ANSWER: The modification was applied.

REFEREE: p.23852. L.20ff. "The CO enhancement in the European lower troposphere represents about half of the CO concentrations observed higher in altitude, which illustrates the high contribution of the CO background at the hemispheric scale." I don't understand this sentence.
ANSWER: The sentence was modified as follows: « Concentrations in the UT are thus only 29% lower than in the LT (i.e. close to local emissions), which illustrates the high contribution of the CO background at the hemispheric scale. »

REFEREE: p.23852. L.23. "daily CO variability at the monthly scale". What you mean? The monthly mean of the daily variability?
ANSWER: Yes, the sentence was corrected.

*REFEREE: p.23857. L.13 A layer cannot have an impact***.** *p.23857. L.18/19. "In the light of this, ozone seasonal changes results ..." -> change. p.23857. L.21. " ... has highlighted significant differences of trend depending on the season" -> change. p.23857. L.22/23. "This section now investigates how these different trends affect the ozone seasonal cycle in the troposphere." Basically no. You would like to check if the trends come along a change of the seasonal variation, right?*

ANSWER: The paragraph was modified as follows: « In the previous section, we highlighted differences in the O_3 trends depending on the season and the tropospheric layer. Here, we investigate if these contrasted trends come along a change of the O_3 seasonal cycle above Frankfurt/Munich (Sect. 3.4.1). »

REFEREE: L.24. "Assuming ..." You can also assume a constant value. Better is "the seasonal variation can be well approximated using a sin function with ..."
ANSWER: The sentence was modified: « The seasonal variation of O_3 can be well approximated by a sine function fully characterized by three parameters […] »

REFEREE: L.12. "... considering windows of 10 years." -> change. p.23858. L.12. "The influence of that window width is discussed further below." -> change
ANSWER: The first sentence was modified as follows: "Results are presented in Fig. 7 for moving 9-year time periods. » **(and the second was removed).**

REFEREE: p.23858. L.22. "... previously obtained by linear regression over the 1994–2012 period." Give ref., e.g. see section ...
ANSWER: The discussion of the baseline was removed, following the proposition of referee #1.

REFEREE: p.23859. L.4. "trend is the most obvious" -> change
ANSWER: This paragraph was largely modified.

REFEREE: p.23859. L.12. "seasonal cycle phase" -> change
ANSWER: The sentence was modified as follows: "Concerning the phase of the O_3 seasonal cycle »

REFEREE: p.23860. Section 4.2. As written in my major concerns I find it counterproductive to include the analyses of ground based data and ozone soundings. There are many relevant and sophisticated papers. To refer to these papers and to compare the results makes in my opinion more sense. You may also write: "application of the same approach to ozone soundings at ... indicate ... (not shown)".
ANSWER: See our answer to the major issue (this section was moved to the Supplement).

REFEREE: p.23865. L.25. "In the lower troposphere, results indicate moderate residence times above Asia...". Never write "results indicate"! It's like "things do". It's just one of many, many further examples where the wording is far too lax. p.23865. L.26. "Higher in altitude, in both relative and absolute terms, ...". What? If 5% of the trajectories originate from boreal Asia, than it's 5%. What you then mean with "relative and absolute terms"? ... lax wording.
ANSWER: Following the recommendations of referee #1, this Sect. 4.3 was removed as it does not provide quantitative explanations of the O_3 seasonal shift. The Fig. 8 and a few sentences are moved to Sect. 2.3 in order to provide some information on the origin of the air masses sampled in our three tropospheric layers.

REFEREE: In my opinion, you exaggerate a bit. You often write in the discussion "much lower/higher", "strongly", ..., although all trends are quite small and partially not significant. To make the conclusions more clear, you could list the major results in bullet points.

ANSWER: In the new version of the manuscript, we carefully checked the terms used to describe the comparisons.

REFEREE: p.23868. Parrish et al... One conclusion you may draw that the downward transport of stratospheric ozone will not be the reason for changes near the ground, because you don't see a relevant change further up in the troposphere.

ANSWER: We agree with the referee. We modified the text as follows: "In terms of stratospheric contributions, stratosphere-to-troposphere (STT) ozone fluxes are known to peak in spring (Auvray and Bey, 2005; James et al., 2003; Tang et al., 2011) due to both enhanced downward transport and maximum concentrations in the lowermost stratosphere (e.g., Thouret et al., 2006). If the seasonal shift was induced by higher STT fluxes, one would expect stronger positive trends in spring and a similar (and even stronger) shift close to the tropopause compared to the LT, which is contradicted by our observations. Thus, the exchanges between the stratosphere and the troposphere are not likely the main reason explaining the shift of the O_3 seasonal pattern. »

REFEREE: p.23870. Summary and Table 1. Here again I don't understand why you give one linear trend of the entire period 1994 to 2012 although most data show a smooth maximum around 2000 (or a bit later).

ANSWER: See our answer to major issue 2. In the revised version of the manuscript, the conclusion was modified accordingly to the numerous modifications applied in the paper.