

## Response to Referee #1:

We are grateful to the referee for her/his careful reading of the manuscript and for her/his comments and suggestions. Responses to individual comments that have been quoted [...] are given here below.

### Major comments

*[The authors have performed a careful statistical analysis of a particular dataset. However, the manuscript does not discuss potential issues associated with the dataset that was used. The authors cite a number of papers that deal with characterization and validation of the IASI-FORLI dataset, but do not discuss any of the potential issues with the dataset that these works may have raised. Of particular relevance for this work on trends is the previous work by Boynard et al. [2016] (a paper on which all authors in this work were involved and cited in this work in the list of papers that deal with characterization of the IASI-FORLI dataset) that has shown that the IASI-FORLI dataset may have its own issues in terms of drifts with time. Figure 15 in Boynard et al. [2016] shows comparisons of IASI-FORLI with sondes over time. The figure appears to show a distinct negative drift in the IASI-FORLI surface-300 hPa ozone compared to sondes over the 2008-2015 time period. This is highly relevant to the results reported in this work, but was not discussed.]*

We thank the referee for pointing that important feature out.

At the time of the initial submission, no drift in the IASI dataset was reported in previous papers (neither in Boynard et al., 2016) or in IASI quality assessment reports ever. From the Boynard et al 2016 paper, it was not obvious that IASI-sondes comparison was showing a drift between the 2 datasets. It was obviously not our intent to leave that feature out of the discussion.

From an instrumental point of view there is no drift in the IASI radiance data. This can easily be assessed as there are currently 2 IASI flying which show similar radiance measurements. IASI is the reference instrument used in the Global Space-based Inter-Calibration System (GSICS). Its instrumental design (based on the Michelson interferometry which spreads and, hence, attenuates the effect of the degradation, if any, over the whole spectral range, as opposed to UV sounders) prevents any instrumental degradation/drift and assure a very good radiometric accuracy and stability. The good performance of IASI is indeed confirmed from the excellent stability in the recorded radiances that are monitored daily at the EUMETSAT ground segment, and from a series of successful validation studies which are mentioned in Section 2 of the manuscript.

However, it is true that two recent validation experiments lead by Arno Keppens/BIRA-IASB and Anne Boynard/LATMOS that were not available at the time of the submission but that are now submitted to this QOS special issue (and listed in the reference section) suggest a drift between IASI and the sonde data. Actually, the drift has been demonstrated in Boynard et al. (this issue) to result from a “jump” in the IASI O<sub>3</sub> time series between the period before and after September 2010. The reasons for this jump are still unclear. It translates to an “artificial” negative drift of around ~2.8 DU/dec in the N.H. (cfr Boynard et al., this issue) and, more particularly, of around ~2.7 DU/dec in the mid-latitudes of the N.H. (based on the stations characterized by the better temporal sampling). The amplitude of that drift is lower than the one of the averaged negative trend derived from the MLR in the N.H. (~5 DU/dec on average in summer; i.e. the drift cannot fully explain the trends reported in the present study). Furthermore, the drift strongly decreases (<|1| DU/dec on average) after the jump and becomes even non-significant for most of the stations (significantly positive drifts are also found for some stations) over the periods before or after the jump.

For overcoming the drift issue and avoiding any potential overestimation of the amplitude of the negative trends derived from the whole IASI dataset, the constant term used in the MLR model has been split into

two components: one covering the period before the jump and one after the jump. We show that the resulting trends are quite similar to the previous ones. In particular, the band-like pattern of negative trends in the N.H. in summer is still clearly observed (i.e. the impact of the jump was likely compensated by the adjustments of other covariates in the previous model regression). The only major difference between the regression results is that significant negative trends that were detected in the high latitudes of the S.H. are now turning non-significant (cfr Figure 1 here below which compares the distribution of O<sub>3</sub> trends derived from the two regression models). These new results are incorporated in the paper. The changes that have been made to address the reviewer's concern include the following:

- The drift reported in the two companion papers is now clearly mentioned in the revised manuscript:
- In Section 2, L.118-124: “Note, however, that a drift in the N.H. MLT O<sub>3</sub> over the whole IASI dataset is reported in Keppens et al. (this issue) and Boynard et al. (this issue) from comparison with O<sub>3</sub> sondes. This drift (~2.8 DU/dec in the N.H.) is shown in Boynard et al. (this issue) to result from a discontinuity (“jump” as called in Boynard et al., this issue) in September 2010 in the IASI O<sub>3</sub> time series, for reasons that are unclear at present. Furthermore, the drift strongly decreases (<1| DU/dec on average) after the jump and it becomes even non-significant for most of the stations (significant positive drift is also found for some stations) over the periods before or after the jump, separately.”
- In Section 2, L.137-140: “In order to take account of the observed “jump” properly, we modified the previously used MLR model so that the LT term is split into two components covering the periods before and after the September 2010 “jump, separately.”
- The figures 1 to 6 and 8 of the manuscript have therefore been reprocessed and they depict now the results derived from the improved regression model (including two constant terms to account for the “jump” in Sep 2010 instead of only one constant term over the whole IASI period).
- Finally, some words of caution have been added in the conclusion section about a possible impact of the reported drift on the trend estimates: “Nevertheless, it is worth noting that there could be a possible impact of the sampling (because of the cloud and quality filters applied) and of the “jump” in September 2010 that has been identified in the IASI dataset (see Section 2), in both MLR and SLR trends.”

*[Also, in considering trends from the IASI-FORLI ozone dataset, the influence of clouds on sampling ought to at least be mentioned somewhere. If I understand correctly, the IASI-FORLI retrievals are only performed for relatively clear-sky cases. We might expect there to be changes in cloudiness over time, and this could potentially impact trend estimates for thermal-IR ozone.]*

Actually, changes in cloudiness over time are not suspected to directly impact on the trend estimates. Only the FORLI retrievals with a cloud fraction in the field-of-view lower than 13%, i.e. only the clear or almost-clear scenes, are analyzed in this study. The maximum threshold of 13% for the cloud cover has been shown in previous studies to be good enough to consider the IASI pixel as clear for the O<sub>3</sub> retrievals (i.e. the atmosphere can be treated as a non-scattering medium in the radiative transfer code; cfr Clerbaux et al., 2009; Hurtmans et al., 2012). It is now clearly mentioned in the revised Section 2 (L.104-106) that the cloud contaminated IASI scenes are filtered out:

“... measurements (defined with a solar zenith angle to the sun < 80°) which are characterized by a good spectral fit (determined here by quality flags on biased or sloped residuals, suspect averaging kernels, maximum number of iteration exceeded,...) and which correspond to clear or almost-clear scenes (a filter based on a fractional cloud cover below 13% has been applied; cfr Clerbaux et al., 2009; Hurtmans et al., 2012)...”.

Note also that the use of quality flags (e.g. based on large residuals ...) that are specified in Section 2 further helps in filtering the cloud contaminated IASI scenes.

We agree, however, that the use of a cloud filter (and of other quality flags) might influence the sampling of the dataset and, hence, that it might impact on the trend estimates. The effect of the temporal and of the spatial samplings on the trend biases was already mentioned in Section 4.3. It is also now indicated in the conclusion Section (L.568-571):

“Nevertheless, it is worth noting that there could be a possible impact of the sampling (because of the cloud and quality filters applied) and of the “jump” in September 2010 that has been identified in the IASI dataset (see Section 2), in both MLR and SLR trends.”

*[There is no substantive discussion of how the trends from this analysis of IASI-FORLI data compare with those reported from radiosondes or from other satellite datasets. The authors do have some discussion in the introduction about difficulties and limitations associated with previous trend studies, and some rather vague, qualitative statements in Section 4.1 about how the trends determined from this work are consistent with findings in the literature. However, there is some implication here, from this paragraph in the introduction, and from the lack of specific discussion of results from other studies in the conclusions, that the trends reported here from this IASI-FORLI analysis provide definitive and absolute answers. I felt that there ought to be some more discussion of these results in the context of the recent Gaudel et al. paper associated with the Tropospheric Ozone Assessment Report. (This paper, for which the authors of this work were also involved as co-authors, had previously been available for public comment and is currently in review for Elementa.) I appreciate that a reconciliation of the differences in the trends from different satellite datasets reported in the Gaudel et al. TOAR paper is outside the scope of this manuscript, and I appreciate that the Gaudel et al. paper used a linear regression approach rather than the more rigorous multivariate approach advocated for in this work. Nonetheless, I feel strongly that the point that there are discrepancies between trends from different datasets included in Gaudel et al. ought to be raised more prominently in this manuscript.]*

We apologize if it is felt from reading the paper that we were so definitive in our conclusions. We are well aware that the accurate trend determination is a difficult task and we wanted to make the point that our results, in particular the comparison between MLS and SLR trends in the dedicated Section 4.3 which clearly highlights large differences in trend estimates, open perspectives for better determining accurate and realistic trends and for further resolving the trend biases between the existing datasets. Some clarifications have been brought in the last paragraph of the conclusion Section (L.580-582):

“This study supports overall the importance of using (1) high density and long term homogenized satellite records, such as those provided by IASI, and (2) complex models with predictor functions that describe the O<sub>3</sub>-regressors dependencies for a more accurate determination of trends in tropospheric O<sub>3</sub> - as required by the scientific community, e.g. in the Intergovernmental Panel on Climate Change (IPCC, 2013) - and for further resolving trend biases between independent datasets (Payne et al., 2017; the TOAR report) ...”

We understand the concern of the reviewer considering the different results/conclusions presented here in comparison to those from TOAR. However, as it is clearly stated in the manuscript (e.g. in the introduction, in the introductory paragraph of Section 4.1 and of Section 4.3), the lack of homogeneity in terms of time-varying instrumental biases, of measurement periods, of spatial and temporal samplings, of boundaries of the O<sub>3</sub> columns and of vertical sensitivity and resolution of the measurements (cfr the TOAR-climate assessment report), combined with differences in the methodology used (MLR vs SLR) makes impossible to “quantitatively” compare our results with those from previous/parallel studies. The best we can do is to “qualitatively” discuss them with respect to the recent published findings that, furthermore, mostly focus on changes in O<sub>3</sub> precursor emissions. It is what we have specifically done in Section 4.1.

It is true that direct comparisons between SLR trends obtained from a series of available independent measurements (among others, IASI) using the same period and the same tropopause definition to limit the

possible sources of discrepancies have been performed in the TOAR-climate assessment report. Nevertheless, large trend biases were reported between the different datasets and, more particularly, between the satellite datasets. The difficulty in comparing, because of the lack of homogeneity between the existing datasets and between the methodologies, the trends from our analysis with those reported in the TOAR-climate assessment report, as required by the referee, is now better underlined in the revised manuscript, especially in the introductory Section 4.3 (L.416-421):

“... Substantial effort in homogenizing independent tropospheric O<sub>3</sub> column (TOCs) datasets have been performed in the TOAR-climate assessment report (Gaudel et al., submitted to Elementa), but large SLR trend biases remain between the TOAR datasets, in particular, between the satellite datasets. The lack of homogeneity in terms of tropopause calculation (same tropopause definition but different temperature profiles are used), of instrument vertical sensitivities and of spatial sampling has been specifically pointed as possible causes for the trend divergence.

Reconciling trend biases between the datasets (e.g. by applying the vertical sensitivity of each measurement type to a common platform, as proposed in the TOAR-climate assessment report) is beyond the scope of this study, but the improvement in using a MLR instead of a SLR model for determining more accurate/realistic trends is explored here ...”

Understanding/reconciling the trend biases is still an open question which deserves further investigation. That huge piece of work could be attempted if there is a TOAR-2 project.

*[The authors raise some interesting speculative points about attribution of trends in tropospheric ozone, but since no rigorous attempts at attribution were made in this work, some care is needed with the language associated with these statements. Specific examples are provided in the minor comments below.]*

As required by the referee, we have now taken care to avoid making too strong statements in the sections specified in the minor comments below (see the minor comments below related to this comment for the changes made in the revised version).

Please, note that, by presenting our results in light of recent reported studies and by exploiting the simultaneous O<sub>3</sub> and CO measurements from IASI, we have investigated, as much as possible, the potential of IASI to derive trends and to help in understanding the origins of the air masses. The only way to more rigorously attempt to attribute trends would be to use a chemistry-transport model which would allow to trace back the sources of the transported air masses. The use of a CTM is beyond the scope of this paper and could be interestingly explored in a future study.

*[The discussion of attribution of trends (Section 4) is largely limited to changes in emissions. Why is this? What about long-term variations in stratosphere-troposphere exchange and the influence on tropospheric ozone? I see that Section 4.1 mentions interannual variability in stratosphere-troposphere exchange in the discussion of trends in IASI-FORLI tropospheric ozone in the SH tropical region, but I did not understand why this was not mentioned in the context of other regions. Presumably this could also be an important factor in mid-latitudes? (e.g. as per Verstraeten et al., 2015)?]*

The influence of the stratosphere-troposphere processes on the tropospheric O<sub>3</sub> trends is specifically discussed in Section 4.1 (for the S.H. tropical region and the mid-high latitudes of the S.H.) and in Section 4.4 where one of the objectives is specifically to help in discriminating the tropospheric from the stratospheric air masses at a global scale by using simultaneous O<sub>3</sub> and CO measurements from IASI. The contamination from the stratosphere was indeed shown to be largest in the mid-high latitudes of both hemispheres (see figure 9 of the manuscript). We agree that the air masses identified, with the O<sub>3</sub>/CO correlation analysis, as mainly originating from the troposphere may also reflect some minor stratospheric contributions and, hence, that the associated trends might be to some extent influenced by the variability in

the stratosphere-troposphere exchanges. This influence is now specifically mentioned in the paragraph related to the trends calculated in the N.H. in the revised Section 4.1 and some values quantifying the influence of the stratosphere into the IASI MLT columns (taken from the supplementary materials in Wespes et al., 2016, which estimates, with a global CTM, the stratospheric portion into the tropospheric O<sub>3</sub> columns from IASI) have been added in Sections 4.1 and 4.4:

- Section 4.1, L.316-318: “We should also note that, even if these latitudes are characterized by the lowest stratospheric contribution (~30-45%; see supplementary materials in Wespes et al., 2016), it might partly mask/attenuate the trends in the tropospheric O<sub>3</sub> levels.”

- Section 4.4, L.486-490 : ‘...the negative correlations for the high latitude regions might also reflect air masses originating from/characterizing the stratosphere due to natural intrusion or to artificial mixing with the troposphere introduced by the limited vertical sensitivity of IASI in the highest latitudes (stratospheric contribution varying between ~40% and 65%; see supplementary materials in Wespes et al., 2016).”

The study of Verstraeten et al. (2015) has also been added in the reference list and referred to in the revised Section 4.1 (L.271-275):

“... the tropospheric O<sub>3</sub> increases which have been shown to mainly result from a strong positive trend in the Asian emissions over the past decades (e.g. Zhao et al., 2013; Cooper et al., 2014; Zhang et al., 2016; Cohen et al., 2017; Tarasick et al., 2017; and references therein) but also from a substantial change in the stratospheric contribution (Verstraeten et al., 2015) ...”

## Minor comments

*[In general, the paper would benefit from editing by an English language service. There are small eccentricities in grammar throughout the paper. They are so numerous that I have not attempted to list issues of grammar in the minor corrections below. However, in most cases, these are not an impediment to understanding.]*

We have carefully proofread the paper in order to track down those grammatical eccentricities. We have found and corrected several of these errors in the revised version. In addition, an English language service will be provided by ACP during the proofreading phase before the final submission of the manuscripts to correct any grammar mistakes/incorrect word usages left in the revised manuscript.

*[The authors have chosen to describe the quantity of interest (tropospheric column from ground to 300 hPa) as tropospheric ozone columns (TOCs) in this work. In the previous Wespes et al. [2016] companion paper, the authors had referred to this ground-300hPa quantity as middle-low troposphere (MLT) ozone. The Gaudel et al. TOAR-Climate paper states that the IASI-FORLI TCO used in that study relies on the WMO definition of the daily tropopause height. It seems confusing to refer to the ground-300 hPa values as TOCs.]*

We thank the referee for highlighting that issue and we agree that it might be confusing. For consistency with those previous papers, we have substituted “TOC” for “MLT” through the revised manuscript.

*[Abstract, line 23-24: “This finding supports the reported decrease of O<sub>3</sub> precursor emissions in recent years”. It would be more appropriate to say that the finding “is consistent with”, rather than “supports”.] It has been changed in the revised version.*

*[Line 76-77: What is meant by “trend characteristics”? Consider an alternative choice of wording?]*

“Trend characteristics” meant both the sign and the amplitude of a trend. It has been replaced by “trend parameters” in the revised version.

[Lines 103-104: “These profiles are characterized by a good vertical sensitivity to the troposphere and the stratosphere”. I am not sure exactly what the authors mean here. Please consider an alternative choice of wording.]

This sentence has been corrected: “These profiles are characterized by a good vertical sensitivity IN the troposphere and the stratosphere”.

[Figure 2: What is the difference between gray areas and crosses in Figure 2? This is not clear from either the manuscript text or the figure caption. Also, the crosses in Figure 2 are almost impossible to see. The crosses are also tough to see in Figure 5, but are a bit more visible in that figure, possibly because of the lighter colour scale. Please find a way to make the crosses more visible.]

As in Wespes et al. (2016), the grey areas indicate that the covariate (here the linear trend term) is not retained by the stepwise backward elimination process in the grid cell, while the crosses indicate that the regression coefficient of the covariate (which is retained by the elimination process) is turning non-significant in the 95% confidence limits when accounting for the autocorrelation in the noise residual at the end of the elimination procedure (cfr Section 3, L.174-177 of the manuscript). The meaning of the grey areas is now given in the revised manuscript (Section 3, L.185-186):

“The grey areas in the LT panels refer to the LT terms rejected by the stepwise backward elimination process”.

The size of the black crosses in the Fig.2 and 5 is limited by the resolution of the grid cells (2.5° lat x 2.5° lon). The resolution of the figures has been improved in the revised version to make the crosses more visible.

[Lines 216-224: This is difficult to follow, possibly because the authors are trying to make a general statement that covers all eventualities. I was not sure what the main point of this paragraph should be.]

Lines 216-224 refer to the titles of Sections 4 and 4.1 and the description of Figure 5 (annual and seasonal distributions of the MLR trends). We think that the referee refers to the next lines (L.224-235). For clarifying the main point of that paragraph, the sentence has been rewritten (L.247-248 of the revised manuscript):

“... As a result, comparing/reconciling the adjusted trends with independent measurements, even on a qualitative basis, remains difficult. ...”

[Lines 246-251: I think the wording of this statement is too strong, given the scope of this study. I was also surprised that this section does not mention stratosphere-troposphere exchange.]

We are not sure how to interpret this comment considering that Lines 246-251 of the original manuscript refer to:

“The large O<sub>3</sub> enhancement of  $\sim 0.33 \pm 0.23$  DU/yr (i.e.  $3.1 \pm 2.2$  DU over the whole IASI period) stretching from southern Africa to Australia over the north-east of Madagascar during the austral winter-spring likely originates from large IAV in the subtropical jet-related stratosphere–troposphere exchanges which have been found to primarily contribute to the tropospheric O<sub>3</sub> trends over that region (Liu et al., 2016; 2017). Nevertheless, this finding should be mitigated by the fact that the trend value in the S.H. tropics is of the same magnitude as the RMSE of the regression residuals ( $\sim 2$ -4.5 DU; see Fig.1).”

The impact of the stratosphere-troposphere exchange is here clearly mentioned in line with what was found in previous studies, and our results have been also treated carefully by considering the amplitude of the RMSE of the regression residuals.

Note that, as required by the referee in his/her last major comment, the influence of the variability in the stratosphere-troposphere exchanges on the MLT trends has been specifically mentioned in the revised

paragraphs related to the trends derived in the N.H. and over the South-East Asia in Section 4.1. (see the related changes made in the revised version in the response to the last major comment above).

*[Lines 272-281: I found the idea that the annual and summer trends for 2008-2016 are “amplified” relative to the trends for 2008-2013 hard to reconcile with the language about “leveling off”. Can the authors please revise this paragraph for clarity?]*

What we meant is that the amplitude of the negative trend calculated from IASI is larger over 2008-2016 than over 2008-2013, which supports the recent assumption of a levelling off of tropospheric O<sub>3</sub> and, further, suggests a possible decrease in the tropospheric O<sub>3</sub> levels.

The sentence has been rewritten for clarity (L.307-313):

“This finding is in line with previous studies which point out a possible leveling off of tropospheric O<sub>3</sub> in summer due to the decline of anthropogenic O<sub>3</sub> precursor emissions observed since 2010-2011 in North America, in Western Europe and also in some regions of China (e.g. Cooper et al., 2010; 2012; Logan et al., 2012; Parrish et al., 2012; Oltmans et al., 2013; Simon et al., 2015; Archibald et al., 2017; Miyazaki et al., 2017). It even goes a step further by suggesting a possible decrease in the tropospheric O<sub>3</sub> levels”.

*[Line 433 (and also line 523): I do not think it makes sense to talk about “air masses” in the context of seasonal means. Consider changing “air masses” to “outflow regions”?]*

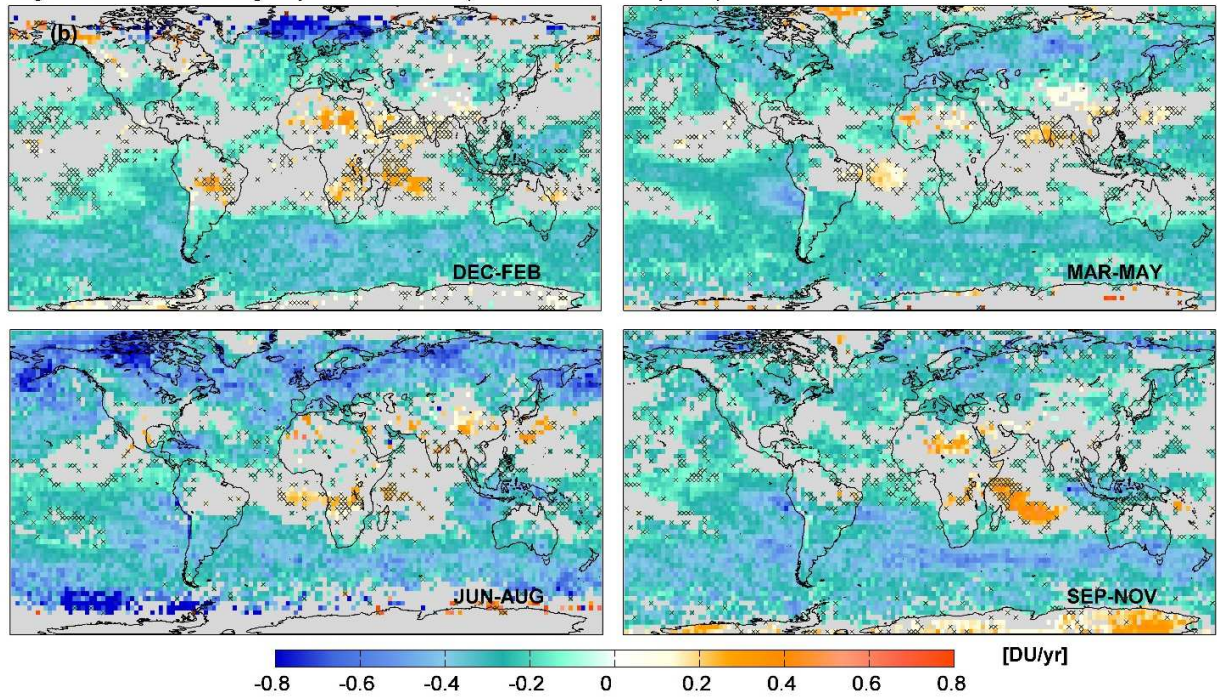
We thank the referee for that suggestion. “Air masses” have been changed to “outflow regions” and “patterns”.

*[Lines 471-482: I found this paragraph difficult to follow. China is not the only place where ozone precursor emissions have been decreasing in recent years. Perhaps it would be better just to say that the pollution outflow from Eastern Asia shows a stronger positive O<sub>3</sub>-CO relationship than the outflow from either the Eastern US or Europe and leave it at that? It does not seem that there is enough information here to make definitive statements about attribution.]*

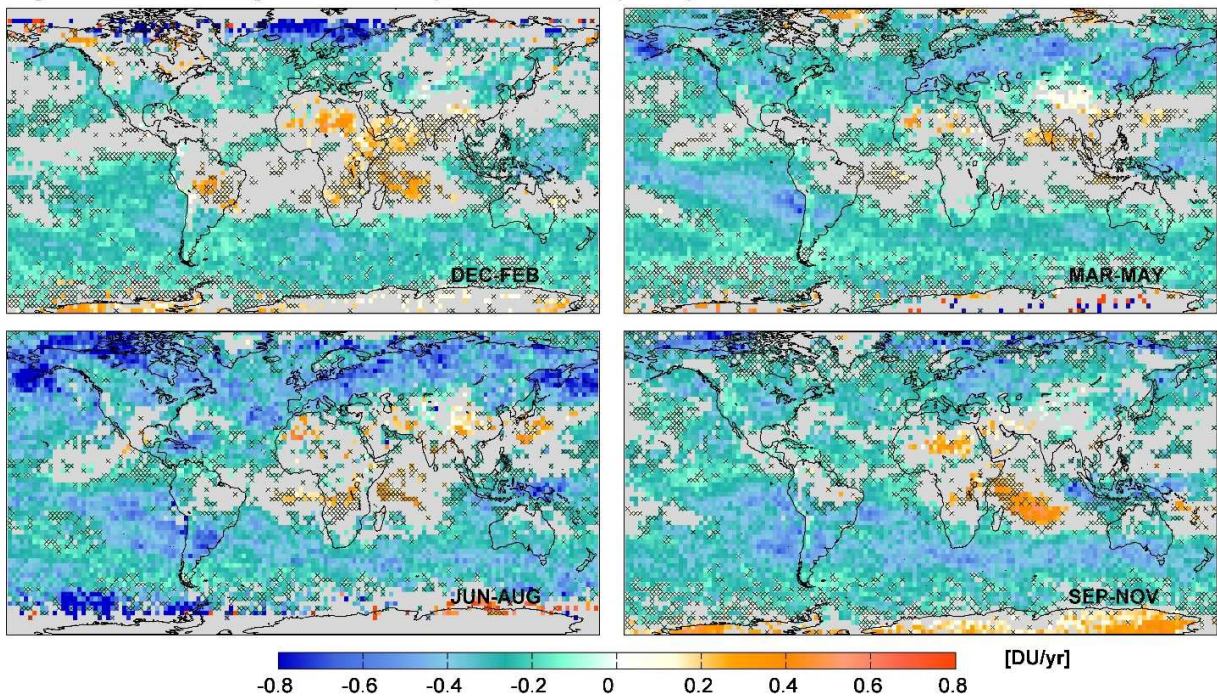
Actually the O<sub>3</sub>-CO covariance (COV<sub>O<sub>3</sub>-CO</sub>) that is discussed and analyzed in that paragraph provides additional information to R<sub>O<sub>3</sub>-CO</sub> and dO<sub>3</sub>/dCO (that are discussed in the paragraphs above in the manuscript). It describes the joint variability of O<sub>3</sub> and CO, and clearly allows to identify North-East of India and East of China in summer as the regions in the N.H. characterized by the largest O<sub>3</sub>-CO variability and, hence, by the most intense pollution episodes (in comparison with Eastern US and Europe). For avoiding possible misunderstandings, some clarifications have been made in the revised version (L.522-529):

“... To conclude, the particularly strong positive O<sub>3</sub>-CO relationship in terms of R<sub>O<sub>3</sub>-CO</sub>, dO<sub>3</sub>/dCO and COV<sub>O<sub>3</sub>-CO</sub> measured over and downwind North-East India/East China in summer in comparison with the ones measured downwind East US and over Europe indicate that South-East Asia experiences the most intense pollution episodes of the N.H. with the largest O<sub>3</sub>-CO variability (COV<sub>O<sub>3</sub>-CO</sub> > 40×10<sup>33</sup> mol<sup>2</sup>.cm<sup>-4</sup>) and the largest O<sub>3</sub> enhancement (dO<sub>3</sub>/dCO > 0.5) over the last decade. The strong O<sub>3</sub>-CO relationship in that region is associated with the significant increase that is detected in the IASI O<sub>3</sub> levels downwind East of Asia (see Section 4.1).”

Regression model including only 1 constant term (over the whole IASI period)



Regression model including 2 constant terms (before and after Sept 2010)



**Fig.1: Comparison between the seasonal distributions of the adjusted trends (in DU/yr) obtained from the MLR model including one constant term (over the whole IASI period) vs those obtained from the MLR model including two constant terms (one before and one after Sept 2010).**