**Response to Reviewer #3:**

**Review**

We thank the referee for his useful comments, which have helped us clarifying several points and improving the manuscript. Below are our point-by-point responses to his comments and suggestions. Before each response, the reviewer comments have been quoted between []. Corresponding information and corrections have been added to the revised version of the manuscript.

**General Comments:**

[1) The motivation for the choice of conducting a multivariate regression to understand the variability observed is not very clearly explained. I guess its main purpose is to be able to eventually extract trends from the full signal, is that right? Section 4.2, providing the detail of each fitting parameter, is missing conclusion comments to guide the reader through the interest of all this work on large datasets. For example: what amount of variability is explained by each process included/fitted. Can it be a way to efficiently analyze controlling processes without using model simulations?]

The use of a multivariate regression model is a commonly and widely used method for analyzing variations in ozone columns records. The goal of the method is to attribute inter-annual, seasonal and non-seasonal variations in ozone measurements to physical processes that are known to affect the ozone records, and, as a consequence, to allow a proper separation of a trend. This is mentioned in the first paragraph of Section 3.1.

Section 4.2 relies on the fact that the multivariate regressors patterns, uncertainties and residuals errors have first to be examined to evaluate the performance of the model before drawing any conclusions on processes controlling the variations in IASI ozone record, including a trend analysis. This is now clearly mentioned in the first paragraph of Section 4.:

"The performance of the multiple linear model is evaluated in subsection 4.2 in terms of residuals errors, regression coefficients and associated uncertainties determined from the regression procedure (Section 3) to properly characterize the physical processes that are expected to affect the IASI ozone records"

Numbers quantifying the contribution of the various physical processes on the measured IASI $O_3$ records have been added in the description of Fig. 8 (b) in Section 4.2. A conclusion on the model performance is also given at the end of this Section:

"Our results demonstrate the representativeness of the fitted models in each layers and latitude bands. These good performances of the model allow examining the adjusted linear term trends in Section 4.3 below."

[2) The authors have chosen to provide an analysis of several altitude levels, chosen according to the IASI sensitivity profiles (provided by the averaging kernels). In Section 2 the IASI retrieval, the AK and associated DOFs are discussed, as well as the correlations between the observed concentrations and the a priori information used as an initial constrain. Although this contribution is around 20-30Also, in Figure 1 the averaging kernels for the selected layers show significant overlap, so that all levels are not independent. This is mentioned throughout the paper but it would really be helpful to provide information on the vertical correlations between vertical levels in IASI. The authors discuss this problem in the supplemental material but I think that there are 2 effects combined in their discussion: the natural influence of the stratosphere on the tropospheric levels through STE, which has to be accounted for to understand observed variability of tropospheric O3 from any type of measurement; and the smoothing from the observational system used here. It is the later that is particularly critical because it may cause artefacts in the trends compared to what would be obtained using in situ observations for example. I guess it would be quite easy to evaluate this effect using the MOZART simulations and the stratospheric O3 tracer with and without applying the IASI AK. Another related question: could correlations in the observations be mistaken for dynamical processes (strat-trop exchanges) in the trend analysis? Because this part is confusing, the fact that tropospheric ozone from IASI can be used for trend analysis is still somewhat questionable after reading this paper... I think this could be easily improved.]

As mentioned by the referee, there are indeed two combined effects in the stratospheric contribution as seen by IASI in the troposphere: 1) the stratospheric influence calculated as the difference between the total simulated $O_3$ and the $O_3^{tagged\_NOx}$ and 2) an additional part due to the limited vertical sensitivity of IASI in the troposphere; the latter is accounted for by applying the IASI averaging kernels on the MOZART4 stratospheric contribution. This simulates the stratospheric part as observed by IASI in the troposphere. This is now better explained in the Supplement.

In addition, as suggested by referees #1 and #3, we now show in the revised Supplement (Figure S4a), the fit of MOZART-4 $O_3$ and of $O_3^{tagged\_NOx}$ time series, in addition to the stratospheric contribution (Figure S4 b), without accounting for the IASI sensitivity. This is done to evaluate the effect of the smoothing error from the observational system. Note also that the smoothing error $[(A-I)Sa(A-I)^T]$ can be evaluated from the a priori contribution [Xa- (AXa)] provided in Figure S5(b) since they are both correlated; i.e., if the IASI sensitivity is low in the MLT, the smoothing error will be large as well as the contributions from the a priori and from the upper layers. When comparing Fig. S4(b) and Fig.S5(b), the differences suggest that the limited vertical sensitivity of IASI contributes a smaller part (~10%-20%) to the IASI stratospheric contribution than the natural stratospheric influence (~20% to 45%). In addition, the contribution of the real natural variability (originating from both the troposphere and the stratosphere through STE processes) into the MLT $O_3$ columns is also now illustrated in an additional figure (Fig. S6(a) and Fig.2 here below) and is estimated to be larger than 50% everywhere. For example, we interestingly show that in the 30N-50N band where the DOFS is the largest (See Fig.2(b)), this contribution reaches ~85% from which only ~30% originate from the stratosphere (Fig. S4(b)) and ~55% from the troposphere (Fig. S6(b) and Fig.2(b) here below). This is now specifically mentioned in the last paragraph of Section 2 of the revised manuscript and in the Supplementary materials.

This further supports the findings that IASI is able to detect a large part of the real variability of $O_3$ in the N.H. troposphere, and that the increase in the observed concentrations and variability in the mid-latitudes N.H. during spring-summer likely indicate a photochemical production of $O_3$ associated with anthropogenic precursor emissions (Cfr. Section 4.1 of the manuscript).

Nevertheless, and as mentioned at the end of Section 2 of the manuscript and in conclusions of the Supplementary materials, we acknowledge that both the apriori and the stratospheric contributions resulting from the limited sensitivity of IASI partly mask the variability of tropospheric ozone measured by IASI and may bias the real trend in tropospheric $O_3$.

[3) Regarding the general performance of the fitting procedure: the figure (Fig. 4) is hard to read. Later in the study, the residuals are shown to actually be quite large (Figs 9 and 10): does this mean that significant processes are not considered? The performance should be better described and discussed at the beginning of section 4.2, with a dedicated figure.]
The figure has been provided at a high resolution and is in our opinion fully readable. The performance of the model is described throughout the manuscript, in Section 4.2 in terms of the covariates with the associated uncertainties, in Section 4.3.1 in terms of temporal sampling effect (daily vs monthly) and in Section 4.3.2 in terms of trend uncertainties.

As mentioned in Section 2, model residuals are less than 10% of the IASI $O_3$ measurements and good correlation coefficients between the IASI and the regression model time series (0.70-0.95) are obtained for all layers and all latitude bands. The residuals represent generally around 30-60% of the deseasonalised $O_3$ time series. This is now specifically mentioned in Section 4.3.1. This might seem quite large but this has to be mitigated by the fact that the non-seasonal variations are only minor contributors. Similar residuals have been reported in previous studies, suggesting that the regression model and the method applied here on IASI time series perform generally well.

We acknowledge however, that the model could be improved in several ways (e.g. including Eliassen-palm flux, EESC proxies) to further reduce the regression residuals. Further investigations on the regressors uncertainties and the total error on ozone measurements should be performed as well. This is now indicated in the Conclusion.

[4) Another point that needs to be improved: throughout the paper, the authors attribute enhanced lower tropospheric ozone to ozone production from anthropogenic emissions. But there are not only anthropogenic emissions that will contribute to lower tropospheric production, biogenic emissions and fires, for example, will also emit significant O3 precursors. This attribution to anthropogenic activities is not really demonstrated. The comparisons to MOZART-4 simulations in the supplemental material is not that convincing: constant anthropogenic emissions are used, but with daily fire emissions. For the ’realism of anthropogenic emissions’, the authors should provide some numbers on previous model evaluations against surface networks. Are there more uncertainties on anthropogenic emissions than on natural emissions? What about stratosphere-troposphere exchanges: is it well simulated by models? Has it been evaluated?]
Previous studies showed good evidence for an increase in tropospheric ozone downwind industrialized areas of the N.H. with a summer maximum and argue that this is due to high photochemical activity associated with anthropogenic emissions of NO, hydrocarbons and CO from combustion of fossil fuels (e.g. Logan et al., 1985; Fusco and Logan, 2003; Dufour et al., 2010; Cooper et al., 2010; Wilson, et al., 2012; Safieddine et al., 2013). Even if it is hard to reconcile the trends in tropospheric ozone with changes in emissions of ozone precursor, trends in emissions have already been able to qualitatively explain measured ozone trends over some

regions although the magnitude of the trend is not consistent with that from model simulations (e.g. Cooper et al., 2010; Logan et al., 2012; Wilson et al., 2012). This has been specifically added in the last paragraph of Section 4.3.2.

The anthropogenic emissions for $NO_x$ and CO are much larger than the fires and biogenic emissions, and they have the largest contributions in industrialized regions of the N.H. (Europe, US, China). In the POLMIP emissions used in this study, the anthropogenic $NO_x$ emissions represent 72% of the total $NO_x$ (see Emmons et al. ACP, 2015) while the fire and soil $NO_x$ emissions only represent 11% and 12% of the total, respectively. For CO, the anthropogenic emissions are also larger than the biomass and the biogenic emissions (~60% *vs* ~30% and <10%, respectively). On the contrary, the biogenic NMVOC (non-methane organic volatile compounds) emissions represent the largest contribution to the total NMVOC (~80%). This is added in Section 4.1.

The MOZART4-GEOS5 model has been evaluated against surface, ozonesondes and aircraft measurements in previous studies (e.g. Emmons et al., 2010; 2013; 2015; Pfister et al., 2008; Wespes et al., 2012). Result indicated that MOZART-4 is slightly biased low by around 5-15%, but that it reproduces generally well the variability of observations in space and time. This is now indicated in the Supplement.

The stratosphere-troposphere exchange is well known to be a common problem in global chemistry-transport models. However, we showed on Fig. S1 of the supplement that the concentrations, the amplitudes of the seasonal cycles and the timing of the maxima are well captured in MOZART4 in the UTLS region for all bands between 50S-50N. We also note that the stratospheric contributions are calculated to be the lowest in the bands south of 50° of the N.H. (~20% in summer in 30°N-50°N; See Fig. S4(b) of the Supplement) and that the stratosphere-troposphere exchanges are usually the weakest during the summer. As a consequence, this even reinforces the conclusion made about the ability of IASI to detect the variations in $O_3$ MLT columns, particularly in mid-latitudes of the N.H. This has been now specifically mentioned in Section 4.1.

[5) My final main comment is that the discussions generally lack quantitative comparisons to other studies. It is generally written that there is a 'good agreement' in trends but it would be interesting to provide orders of magnitude: maybe with a final table comparing results depending on the method chosen?]
As suggested by referee#1, when comparing to the previous publications of the trend analysis, "in agreement with previous studies" has been replaced by "comparable to the results published in the previous studies" throughout the paper. Quantitative comparisons cannot be made considering the different time period in our study with respect to others.

In addition, as suggested by the referee in his specific comments below, trend values and temporal resolutions of measurements from previous studies have been added in the discussion in Section 4.3.2 and we also now provide comparisons between trends inferred from a subsampled IASI dataset to match the temporal sampling of FTIR. The results are added in Table 4 of the revised manuscript. These new results show that fitted trends inferred from IASI and from FTIR are within the uncertainties of each other and that those associated with the subsampled IASI datasets are significantly larger than those obtained from the daily ones, leading to statistically non-significant trends. This gives further evidences for the ability of IASI, thanks to its high temporal sampling, to detect statistically apparent trends even on short periods.

**Specific comments:**

[1) Introduction:
a) P. 27577, l. 21: Why are there warnings? Only because too many unknowns or are their specific reasons?
b) P. 27578: 1st paragraph: It would be helpful to provide a few numbers for trends identifies in previous studies. Are the signs consistent? What were these studies based on? Observations at what resolution? etc.]
a) Warnings are mostly related to possible underestimation of the true uncertainties in the ozone trends that can be attributed to decreasing EESC. This has been added in the revised text.
b) It has been added. Numbers for trends from previous papers are now given later in the manuscript in Section 4.3.

[2) Section 2: Cf. main comment: a better evaluation of vertical correlations is the observations is required. Also provide information on the a priori used for the retrieval. The last sentence is disturbing and does not really help the reader. . . It would be good to confront each result to the identified sources of uncertainties in the discussion parts.]
See responses to general comment 2) above. The last sentence has been changed accordingly. The apriori and stratospheric contribution are now discussed as well in the discussion Section 4.1.

Information on the a priori has been added in Section 2:
"The a priori information (a priori profile and a priori covariance matrix) is built from the Logan/Labow/McPeters climatology (McPeters et al., 2007) and only one single $O_3$ a priori profile and variance-covariance matrix are used."

[3) Section 3:
a) Define ODS
b) l.12-13, P.27582: I am not an expert in this type of statistical analysis but I just don't get what this means. Maybe a little bit more explanations would be helpful.
c) Table 1 and corresponding text: the source of the data used should be detailed for all proxies: are they from model simulations? reanalyses? observations?
d) Sentence btw P. 27584-P. 27585: Would be good to detail briefly why harmonic and linear trends are appropriate for these effects, since these are among the main targeted features of the analysis.]
a) ODS is defined in Section 1.
b) We calculated the uncertainty of the fitted parameters by computing the standard error with an effective sample size (n*) of independent information based on the lag-1 autocorrelation coefficient correlation of the noise residual ($n^* = n \cdot \frac{1-\Phi}{1+\Phi}$) as in Santer al. (2000).This is now better explained in Section 3.1 of the revised version with the Eq. (3).
c) Details for the proxies have been added in the text.

d) The choice of using "statistical" harmonic and linear terms instead of "physical" proxies in the backward elimination approach is widely applied in previous ozone regression studies (e.g. Mäder et al., 2007; 2010). Physical proxies for the Equivalent Effective stratospheric chlorine (EESC) and for the Eliassen-Palm (EP) flux could be used for describing changes in ozone depleting bromine and chlorine substance and in the Brewer-Dobson circulation instead of the linear trend and harmonic terms. As explained in Section 3.2, the use of the EESC proxy could be an interesting alternative, in particular for the measurements starting before the turnaround point for the ozone hole recovery in 1996/1997. This would avoid adjusting two LT terms (one before and one after the turnaround).

[4) Section 4:
a) Also, the discussion should include reminders about expected uncertainties (that had to be kept in mind). For example P. 27586, last paragraph about the results in the tropics: what about the large contribution from the a priori? Does it matter in this discussion of the variability?
b) Several mention of the impact of anthropogenic emissions that has not been demonstrated: what about other sources? (Cf. comments above)
c) Section 4.2: Cf comments above about the residual. A lot of the processes have already been identified in the previous description of the O3 variability. This description could be a lot shorter, avoiding repetition. More importantly, it would be helpful to clearly explain why the regression is performed and what we learn with the coefficients plotted, what we learn in terms of O3 variability.
d) P.27589, last sentence: this statement seems really strong and is not well demonstrated. MOZART4-GEOS5 has probably already been compared to surface observations of O3 and NO2 in many regions, very probably providing a better evaluation of the quality of the anthropogenic emissions. Authors could provide a few numbers from the literature here.
e) P.27590, l. 13-14: Unless I missed it, this is the 1st mention of these numbers regarding the performance of IASI. This should be in section 2. And in this section, the authors should explain if the findings are still relevant considering this uncertainty (bias in this case).
f) It would be helpful to add a short conclusion at the end of section 4.2 to clearly explain how/if the multivariate regression approach provided original information, or if it is mainly used to get rid of other factors than the trends the authors are aiming at identifying.
g) Section 4.3.1: Results only discussed here for the US layer. What about other layers? Similar conclusions (I guess) or less critical to have daily obs? What about previous studies cited throughout the paper: what temporal resolution are they using?
h) Section 4.3.2: Provide numbers for trends obtained in the literature (cf general comment above).
i) Section 4.3.3: Are trends consistent if the full IASI dataset or IASI at the FTIR location is used? Hence: do we really need such high coverage to conclude? For regions where trends are insignificant if inferred from FTIR: same result if IASI used only when FTIR observation is available?]
a) We thank the referee for pointing that out. The discussion related to Fig.4 has been now extended to include the effect of the a priori contribution:
"A lower tropospheric column (e.g. ground-700 hPa) can generally not well be decorrelated because of the weak sensitivity of IASI in the lowermost layers (Section 2). However, the measurements in northern mid-latitudes in spring-summer are characterized by a larger

sensitivity. In the ground-700hPa columns, we find that the apriori contributions do not exceed 40% and they range between 10% and 20% over the continental regions….”
“This certainly helps in detecting the real variability of $O_3$ in the N.H. troposphere, and, the increase in the observed concentrations and the variability may likely indicate a photochemical production of $O_3$ associated with anthropogenic precursor emissions.”

b) Cfr responses to general comment 4) above. The small contributions from fires and biogenic emissions are now mentioned in the revised Section 4.1.
c) Please refer to responses to general comments 1) and 3).
d) Cfr responses to general comments 2) and 4).
e) Actually, this bias was already mentioned in Section 2. The bias is supposed to only affect the constant terms by ~10-15% in UTLS in the mid-latitudes and in the tropics, not the ozone variations. The sentence has been rewritten.
f) This has been added. Please refer to response to main comment 1) above.
g) We thank the referee for this comment. Section 4.3.1 and Figure 9 have been revised to clarify the discussion. The averaged relative residuals are now indicated in the middle panels in the revised Figure 9 which now also includes the differences between the both models (with and without the linear trend term) to highlight the offset which corresponds well to a trend over the IASI period. We also note that this offset is observed for most of layers and latitudinal bands particularly where either $O_3$ recovery or solar effect is important, consistently with the decrease in trend uncertainty. This is illustrated in Figure 1 below (same as Figure 9 of the manuscript) which shows one additional example (MLST in 30°S-50°S) of model adjustment, characterized by a large significant negative trend in both daily and monthly data (see Table 2 of the manuscript) and, as a result, by a large offset between the two regression models.

We add in the text: “The same conclusions can be drawn from the fits in other layers and latitude bands, especially those where the solar cycle variation of ozone is largest (MLST and UTLS) or where the ozone recovery occurs (UST). A larger trend uncertainty associated with the monthly data *vs* daily data is found in all situations (see Table 2, Section 4.3.2).”

Thanks to the high temporal sampling of the instrument, obtaining daytime daily measurements from IASI in the other layers is not critical at all, except for polar regions where daytime measurements are not possible during the polar night (data are only available during February-October and October-April over 70N-90N and 70S-90S, respectively; See Table 2).

To the best of our knowledge, the previous studies which focused on an analysis of inter-annual variability and long term trend in $O_3$ (total, stratospheric or tropospheric columns, either on regional or global scales) have all applied a regression model on monthly time series, not on daily measurements.

h) Cfr responses to General comment 5) and to Specific comment 1b). Trend values from previous studies have been added in the revised text for comparison.
i) We thank the referee for this comment.
As one can expect from the FTIR measurements illustrated on Figure 10 of the manuscript, IASI time series with the same sampling as the one of the FTIR leads to non-significant trends. This has been tested by applying the regression models on subsampled IASI dataset with the same

temporal resolution as that of the FTIR datasets. In all cases, we observe that the fitted trends inferred from IASI and FTIR are within the uncertainties of each other and that the uncertainties associated with the subsampled IASI datasets are significantly larger than the ones with the full daily dataset. These trend values have been added in Table 4 of the manuscript and this is mentioned in the revised Section 4.3.3.

Even if validating the IASI fitted trends with independent datasets is challenging due to the short-time period of available IASI measurements and the insufficient number of usable correlative measurements over such a short period, our results in Section 4.3.2 give convincing evidences for the ability of IASI to measure changes in $O_3$ records and sustain the need for continuous and high temporal frequency measurements.

[5) Conclusions:
a) P. 27598, l. 19: 'reasonably independent' is too vague considering that it is critical. It needs to be better evaluated.
b) Last sentence: I have not really seen this conclusion clearly appearing in the results.
c) P. 27599: I do not really agree with the conclusion on anthropogenic emissions as well since it is not really demonstrated here, and not referenced.]
a) Cfr general comment 2) and 4). The IASI vertical sensitivity in the MLT has been better evaluated throughout the manuscript.
b) As indicated in Section 1, only the MetOp-A measurements have been used. The extension of the $O_3$ records with the successive IASI instruments -B (2012) and –C (2018), and with IASI successor on EPS-SG after 2021, will allow assessing consistent trends in the different atmospheric layers.
c) See responses to General comment 4).

[6) Supplemental material
a) Information on the O3 tagging technique should be briefly provided in the S.1 section.
b) I do not agree with the conclusions provided. I do not see how the authors conclude that it's a problem due to anthropogenic emissions.
c) 2: The contribution from STE exchange does not really help with the problem of IASI's vertical resolution (Cf. comment above). A comparison of simulated contributions with and without smoothing would provide some indication on both aspects.]
a) Detailed information on the tagging procedure and all the photolysis and kinetic reactions used in MOZART4 for the tagged species can be found in Emmons et al. (2012). This has been mentioned in Section S1 of the Supplement.
b) Please refer to General comment 4) above.
c) Please refer to General comment 2) above.

**Technical comments:**

[1) Table 3: why 'Feb.-Oct' and 'Oct-Apr'? What do blank lines correspond to? (only '-')]
These time periods correspond to periods of available daytime IASI observations for the polar zonal bands. Because only daytime IASI observations (determined with a solar zenithal angle to the sun < 80°) are used in this study since they are characterized by a better vertical sensitivity to

the troposphere, we do not have measurement during the polar night (indicated by '-' in Table 3 of the manuscript). This is now mentioned in the revised text.

[2) Figure 4: No (a) and (b) in this Figure. . . The lines for the partial columns are hard to read, I suggest adding a specific figure. Small black titles in the figures are hard to read, real titles would be clearer.]
Titles have been enlarged and (a) and (b) have been added. A higher resolution of the figure has been provided.

[3) Figure 6: Color scale could be adjusted for the bottom plot since magnitudes change for each plot.]
It has been scaled

All the other technical corrections have been included in the revised version.

## Figure Caption

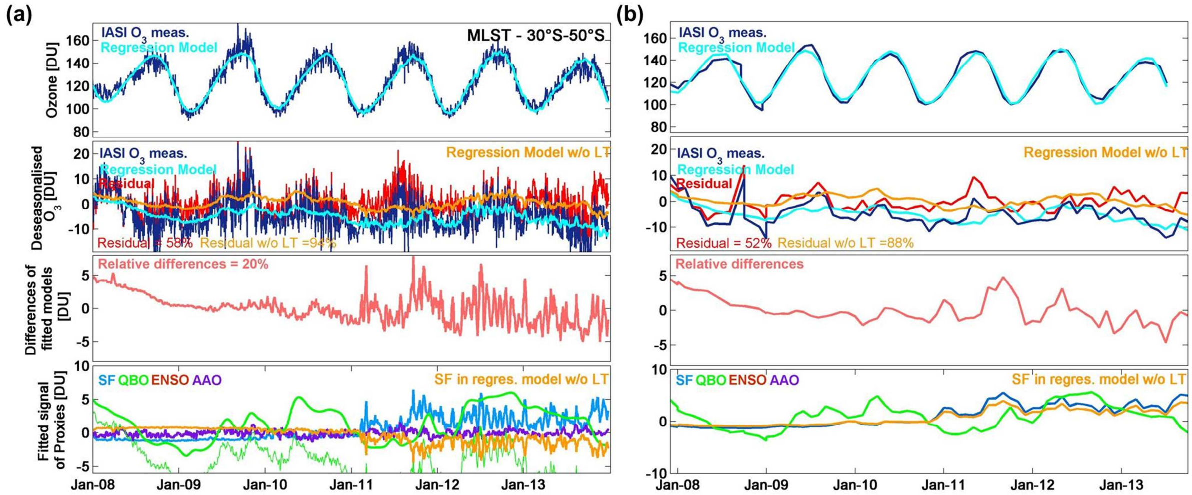


**Figure 1.** Daily (a) and monthly (b) time series of $O_3$ measurements and of the fitted regression model in the UST in the 30°S-50°S latitude band (top row), of the deseasonalised $O_3$ ($2^d$ row), of the difference of the fitted models with and without the linear term ($3^d$ row), and of the fitted signal of proxies ([regression coefficients*Proxy]): SF (blue), QBO ($QBO^{10}$ + $QBO^{30}$; green), ENSO (red) and AAO (purple) (bottom) (given in DU). The averaged relative residuals are also indicated (%).

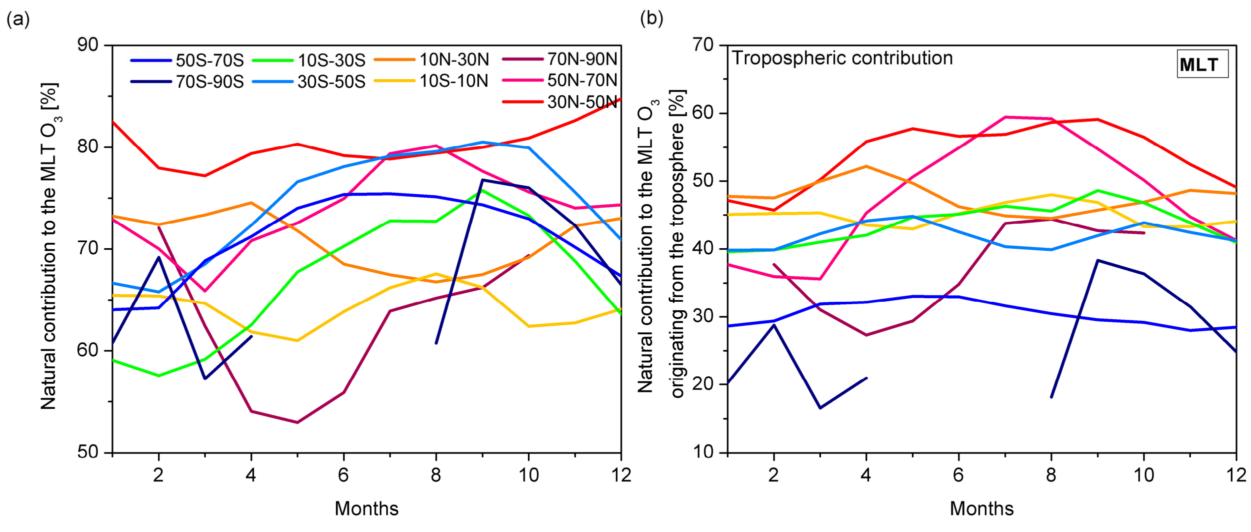


**Figure 2:** Contribution to the IASI MLT $O_3$ columns (%) (a) of the natural variability (troposphere and stratosphere) and (b) from the troposphere.