Referees#2 and #3 made comments on the review of referee#1. We therefore include all answers in this file (referee comments in blue). We would also like to apologize to Referees#2 and #3 for making frequent references to the answer to Referee#1, because we first answered the comments of the latter.

Answers to Referee#1

Interactive comment on "Particulate sulfur in the upper troposphere and lowermost stratosphere – sources and climate forcing" by Bengt G. Martinsson et al. Anonymous Referee #1 Received and published: 24 February 2017

This manuscript analyzes measurements of particulate sulfur taken from commercial aircraft altitudes in the IAGOS and CARIBIC programs. It attempts to derive the fraction of upper tropospheric sulfate that is from the stratosphere. This is an excellent data set but I find the analysis inadequate for several reasons. A previous paper by many of the same authors on a subset of these data (Friberg et al., 2014 but it only uses data through 2008) is a much better analysis. I would recommend resubmission with completely new analysis based on extending the Friberg et al. techniques to include the newer data and more emphasis than the Friberg paper on non-volcanic periods.

We appreciate the kind words about Friberg et al. (2014). The entire author group is well acquainted with the usefulness of O3 in interpreting particulate sulfur in the LMS. More than 10 years ago we developed an O3-based model to study particulate sulfur in the volcanically quiescent period at the turn of the millennium (Martinsson et al., JGR 2005). Following the Kasatochi eruption we introduced the ratio of particulate sulfur to O3 in order to express volcanic influence on the concentration of the former species (Martinsson et al., GRL 2009) as a method to make samples that were variably impacted by mixing across the tropopause more comparable. In the present study we seek a standalone methodology to be able to express concentration gradients and to integrate to obtain radiative impact. This is important because this data allow us to identify the volcanic component also close to the tropopause in contrast to lidar measurements that are biased by tropospheric sources (dust, condensed water and other species). This is altogether different in the goals compared to our previous studies, and to our knowledge the first such study. Therefore, a different methodology is required.

1) The manuscript (line 135) states that the analysis is based on concentrations per unit volume rather than mixing ratio. This is a mistake; straightforward analysis of fits of concentration versus a parameter like distance from the tropopause requires the use of mixing ratio. For an idealized example, suppose the aircraft ascends 1 km as it goes through a perfectly uniform air mass. That ascent would change both the x-coordinate (distance from the tropopause) and the y-coordinate (concentration not corrected to mixing ratio). This coupling of the independent and dependent variables makes it impossible to interpret the slopes in a simple fashion. The stated reason for using a volume concentration (integration over an altitude range) can always be done later in the analysis.

This problem was considered already before the analysis of the data, but was deemed insignificant due to the relatively narrow altitude range of the measurements, which, in principle, was a mistake. In response to this criticism we did make a change in the analysis to undertake the regression analyses based on STP-normalized concentrations ($C_{STP} = aZ + b$), which is a mixing ratio. This means that the regression results need to be converted to volume concentration in the final integration ($C_{STP}/C_V = Q_V/Q_{STP}$, where C and Q are concentrations and molar volumes). We used ECMWF data to estimate the slope of the molar volume

from pressure and temperature profiles in the LMS. Based on the average slope of all our samples we describe the molar volume with an exponential function:
$Q(Z) = Q_{TP}\, e^{wZ}$, where Z is the height above the TP and w = 0.0001535 m^{-1} (from ECMWF).
For each measurement we have a molar volume Q_m obtained at height Z_m above the TP. For this measurement the TP molar volume is obtained by $Q_{TP} = Q_m\, e^{WZm}$. We integrate to obtain the column concentration for the first 3000 m of the LMS:

$$C_{col} = \int_0^{3000} \frac{Q_{STP}}{Q_V(Z)} C_{STP}(Z)dZ = \int_0^{3000} \frac{Q_{STP}}{Q_{TP}} e^{-wZ}(aZ + b)dZ$$

2) The distance from a tropopause defined by potential vorticity (PV) is not a very good choice for the independent variable (Figures 1 and 2 and subsequent analysis). First, the PV values come from a meteorological analysis with substantial uncertainty. A colleague I spoke to estimated +/- 500 meters. There are also ambiguities with multiple tropopauses. Consider what Figure 2a would look like with uncertainties of +/- 500 m in the horizontal for most points, and more for a few points in the neighborhood of multiple tropopauses. Note that using an independent variable with significant uncertainty not only introduces noise into line fits but also biases the results to smaller slopes and, for positive data, larger intercepts. (This is worse than uncertainty in the dependent variable, which introduces noise but not bias.) Second, there is no reason why the gradient in PV has to be uniform with distance above the tropopause, so deeply stratospheric air could be close to or far above the tropopause. No tracer is perfect, but ozone, as used in Friberg et al., would be a far better choice than distance from a PV tropopause.
Locally there can be large uncertainties in the model tropopause. However, a sample was taken over typically 1500 km or more in flight range, meaning that the distance to the tropopause is based on typically 100 ECMWF positions along the flight track. Therefore, these samples become less sensitive to local errors of the model. The referee need to acknowledge the different purpose here compared to other studies of the ExTL. The objection "there is no reason why the gradient in PV has to be uniform with distance above the tropopause" is difficult to interpret. We do not use the PV as a tracer, and hence we do not need to make any assumptions concerning PV uniformity. Instead we use the dynamical TP at 1.5 PVU as the lower boundary of the stratosphere, which is generally recognized as the lower boundary of the stratosphere in view of chemical composition. Then we use the linear distance to this TP in the regression analyses, which tests the uniformity. The coefficient of determination (R^2) exceeds 0.6 in 72% of the 52 ordinary linear regressions undertaken. The deviations from the linear models consists of scatter without any trends. We do not force an erroneous shape of the gradient.

We considered using an O3-based distance according to Sprung and Zahn (JGR 2010) which is available in the IAGOS-CARIBIC dataset. However, we decided against that for two reasons:

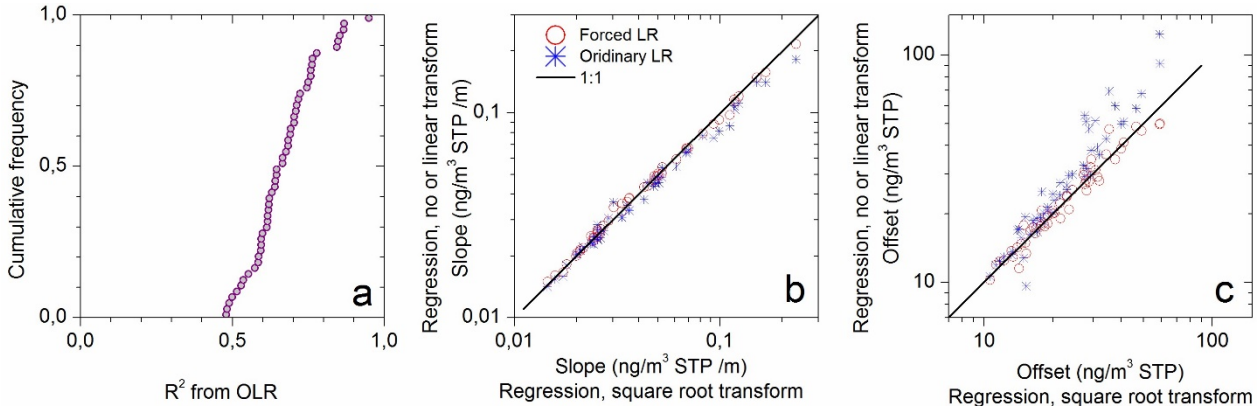
A) That parameter refers to a static tropopause implying that approximately 1000 m of the LMS falls below.
B) The use of a tracer would potentially introduce non-linearity in the vertical parameter that could introduce a bias in the results

3) The analysis of the stratospheric influence fraction is very convoluted with no propagation of uncertainty shown. There are three successive line fits to data, as shown in Figure 2a to 2c. After reading

through the manuscript several times, and having worked extensively with similar data, I still do not understand how the measurement uncertainties and atmospheric noise propagate into the results.

Due to the good detection capacity for particulate sulfur, the total measurement errors are 12% (Martinsson et al., AMT 2014), which is small compared with atmospheric variation of this species. In the analysis we use a method of forcing the linear regression to comply with the observations close to the TP, as a way to deal with the heteroscedastic nature of the data. In response to this comment we have simplified the analysis. We skipped the ordinary linear regression (now only used for illustration purpose) and used only forced linear regression, where the regression was forced to the average concentration C_0 and distance Z_0 of the measurements closest to the tropopause. With a simple linear variable transformation (forming $C - C_0$ and $Z - Z_0$) this becomes equivalent to forcing the regression through the origin. Thus, first a forced linear regression, and then a weighted regression between the obtained slopes and offsets of each month. The uncertainty in the slopes of the first step regression is propagated. The error estimates used in Fig 3 are based on the student t distribution (t70% and t95%, two-sided) because the number of measurements available to estimate C_0 in some cases is small. The error in b (at Z = 0) is based on the uncertainties in C_0 (at Z_0) and the slope, and is computed using the weakest slope at the upper error limit of C_0 and strongest slope to the lower limit.

To further advocate this method, we undertook a small study of the scedastic nature of our data using variable transformations. The most common transforms, square root and logarithm of the dependent variable, were investigated. Like most natural science data set our is heteroscedastic, showing increasing variance with distance from the TP. The log transformation changed the data in a way that the variance became smallest at large distance from the TP, whereas the square root transformation resulted in a rather constant variance in the dependent variable along the independent variable axis. This transform is thus more suitable for regression. After transformation the relation between the dependent ($y' = \sqrt{y}$) and the independent (x) variable is not linear (y and x has a linear relation). Thus the following expression should be minimized with respect to a and b: $(y' - \sqrt{ax + b})^2$. This resulted in rather tedious expressions that we had to solve numerically. In the Fig (planned new Fig. 2 in a revised manuscript) you can see a comparison of slope (Fig b) and offset (Fig c) between the sqrt-transformed regression and ordinary linear regression (OLR) and forced LR. As can be seen, the OLR deviates sometimes strongly, in particular the offset, from the sqrt-transformed due to the heteroscedastic data. The forced LR, on the other hand, shows only small differences from the analysis based on sqrt-transformation. Due to more direct determination of the offset as well as the simpler analysis we chose to use forced LR.



4) Putting aside the choice of independent and dependent variables and the propagation of uncertainty, there is a conceptual problem with defining the stratospheric influence from a corrected intercept derived from the line fits, as is done in this manuscript. Such an analysis of the slope and intercept of two variables in the lowermost midlatitude stratosphere generally assumes that both variables are conserved quantities controlled largely by transport and mixing (e.g. Plumb 1996 JGR tropical pipe paper). But sulfate mass in the lowermost stratosphere is mostly controlled by sedimentation (Wilson et al., Steady state aerosol distributions. . ., ACP, 2008). In the presence of sedimentation, it is not obvious what the slope and intercept of a correlation plot mean. Indeed, it is clear from Figure 5 in Wilson et al. that a line fit over an altitude range that goes deep into the stratosphere could easily produce an intercept unrelated to the tropopause value.

Yes, there is sedimentation going on in the stratosphere, which, through the Cunningham slip correction factor describing the viscous contact between a particle and the surrounding air, is strongly dependent on altitude. In the LMS, with typical size distributions (background or moderately volcanically influenced), the sedimentation velocity of the LMS is typically less than 0.15 km/month, implying that the particulate sulfur mainly is removed from the LMS by air transport (Martinsson et al., JGR 2005). Considering the recommendation above by the referee to use a tracer to improve the analysis, this comment is somewhat surprising. Whereas this effect in principle could affect the relation to a tracer, the observed gradient in our present study carries no such assumptions.

Lesser concerns are: (a) Distance from the tropopause is strongly correlated with latitude, since commercial flights generally only get well above the tropopause at high latitudes. This makes it difficult to separate latitude and altitude as causes of a correlation. (b) The introduction is too broad.

The referee is right in the assumption that observations deep into the LMS could be more frequent at higher latitudes. Ideally we would want to have enough observations to also study the latitude dependence. However, the salient features of the LMS with its ExTL are caught by the broad latitude band.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1142, 2017.

We thank you for your comments. In response we have adjusted the data handling by using mixing ratios which are converted to volume concentrations in a late stage of the evaluation. We also have simplified and clearer motivated the choice of regression methodology. We will re-write section 2 (mainly) accordingly. The changes did of course affect all Figs., but the changes do not affect the conclusions of the discussion paper.

Answers to Referee#2

Interactive comment on "Particulate sulfur in the upper troposphere and lowermost stratosphere – sources and climate forcing" by Bengt G. Martinsson et al. Anonymous Referee #2 Received and published: 22 March 2017

This manuscript reports on measurements of aerosol sulfur in aerosol samples collected from the IAGOS-CARIBIC platform over a 16 year period. Analysis focuses on a new regression technique that the authors suggest can be used to infer both the gradient of sulfur concentration and the integrated burden in the lowermost stratosphere (LMS), starting at the dynamical tropopause and extending to 3 km above it. It is also suggested that this analysis provides an estimate of the relative contribution of stratospheric sulfate mixed downward and tropospheric sources on the sulfur concentration in the upper troposphere (UT), and how these contributions vary seasonally. My biggest problem with the paper is that the authors do not show and discuss enough raw data to demonstrate that the regression approach is reasonable. Figure 1 does C1 give the reader a useful impression of both the range and seasonality of sulfur observed in the UT, and supports the authors impression that there is little correlation with distance below the tropopause. However, the LMS concentration data are never shown so the reader has no idea if fitting linear regressions is a remotely logical approach.
The reason we have not included figs of the first step linear regressions is that they are so many. In Fig 3 you in principle can see them all, but we agree that it is difficult to get the impression of the data from that Fig. However, a useful compromise could be to, instead of making 52 graphs, present the R^2 of the 52 ordinary linear regressions (OLR). Fig. a (in the answer to referee#1) shows the cumulative frequency of R^2. As can be seen, R^2 of the 52 OLRs span 0.48 to 0.95 and 50% of the OLRs have R^2 exceeding 0.66. We plan to replace the current Fig. 2 with this Fig.

This is compounded by the fact that the analysis apparently required multiple steps which are not well explained in section 2.3.
As explained in the answer to Referee#1 we have simplified the data evaluation, please see above for details.

I realize that this group has written a number of previous papers on this data set, and perhaps some of these have already presented spatial and seasonal distributions of sulfur in the LMS in ways that set the stage for this new analysis. However, I did not, and readers in general should not have to, read these earlier works to understand this one. I could provide a fairly long list of specific sentences and paragraphs that I found to be confusing or misleading.
That would have been helpful.

However, I just noticed that reviewer 1 has suggested major revision, starting with fundamentally changing the approach to analysis, which will clearly require rewriting most of the text.
Based on the comments from all the referees the revision will mainly concern the methods section.

Therefore it seems that specific editorial suggestions to improve clarity are premature I agree with the concerns reviewer 1 raised regarding the use of concentrations rather than mixing ratios, and relying on

distance from the tropopause as the independent indicator of degree of stratospheric character captured by a given sample. (I also note that simply defining this distance for sample intervals approaching 2 hours in length would often seem ambiguous, even without double tropopauses or crossings of the tropopause.). I cannot comment on the suggestion to extend the analysis used by Friberg et al., 2014, since, as noted above, I am not familiar with this paper.

We have made a new analysis based on mixing ratios, and extensively answered the comments by referee#1, please see above.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1142, 2017

Thank you for your comments. In response we have added Fig a above. In the planned revised manuscript this Fig will replace Fig 2 of the Discussion version. We will also undertake the changes according to the answers to the other referees.

Answers to Referee#3

Interactive comment on "Particulate sulfur in the upper troposphere and lowermost stratosphere – sources and climate forcing" by Bengt G. Martinsson et al. Anonymous Referee #3 Received and published: 28 March 2017

The paper describes the use of measurements of sulphur collected onto filters on board CARIBIC commercial aircraft over the past decade. The filters were analysed by PIXE. The data analysis develops a relationship between the concentration measurements and vertical distance to the tropopause derived from a PV analysis of re-analysis data. This is used to build a seasonal profile by combining data from multiple flights in each 3 monthly period. The analysis is used to show the amount of sulphur in the lower stratospheric column and derive an aerosol optical depth resulting from it.

I found the description of how the analysis was done to be less than clear. Some of the sentences were long and not transparent and on a number of occasions the sentence did to scan well or had typos in it e.g. "This was undertaken for 4 up to 7 groups of data for each season, and a total of 60 regression groups distributed over 12 overlapping seasons were used. This overlapping places each month in the center of a three-month season thus adjusting to smooth seasonal changes in the UTLS." I would recommend this whole section and how this relates to the further description of the approach in the results section need re-writing and clarifying.
We will look carefully on the formulations and try to improve. We have also simplified the analysis (see answer to referee#1), which hopefully will further help to improve legibility.

Further, it is not clear what is meant by "overlapping", does this mean that the some of the same data are used in multiple regressions? I assume this is a 3 month average centred on a particular month from the legend in figure 3. Please clarify.
Your interpretation is correct. We will clarify this in the planned revised manuscript.

A previous referee is critical of the use of concentration rather than mixing ratio and I can see why in principle. Equally I can understand the authors' use of concentration since the column abundance of sulphur can be retrieved from the regressions and hence the AOD which would not be the case if the mixing ratio was used. This is also true of the use of altitude deriving from a PV definition of the tropopause rather than ozone. However, as far as I can see this only works if the pressure changes over the altitude range of the samples are sufficiently small that the regression derived results from the relative position to the tropopause and not the absolute altitude, this needs to be clarified before the analysis can be verified.
We have changed the analysis in this respect to follow the recommendation of Referee#1 of using mixing ratios, and complemented with the use of a transformation to volume concentrations in the integration of the column concentrations.

The weighting of the regression isn't described in sufficient detail for a reader to follow and replicate. This needs clarification.
The weighting is undertaken based on the error estimated of each forced LR, where the inverted, squared linear error (t70%; see answer to Referee#1) of each first step regression is used as weights in the second step regression. We will clarify in the planned revision of the manuscript.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1142, 2017.

Thank you for your comments. We will re-write the methods section to describe the change in methodology, to clearer explain our choices of regression methodology (exchanged Fig. 2). We have already commented on your main concerns in the answer to referee#1.