Interactive Discussion on "High resolution assimilation of IASI ozone data with a global CTM": Final Response

We would like to thank the referees for their useful comments. Our answers to Referees #1 and #2 general and specific comments are detailed below. Note that the order of our answers does not always follow those of the referees' comments and are brought together according to the theme discussed.
Each answer is labelled A*.

```
**********************
* Referees introductions: *
**********************
```

* Referee #1:
The title of the paper might benefit from a revision because (i) it does not mention the assimilated MLS data, (ii) the term "global CTM" implies that a multi-species chemical mechanism is applied and (iii) the term "high resolution" is relative.

A1 Answer:
For the points (i) and (iii) the referee is right. Even if the term "high resolution" is relative compared to the resolutions now used by state of the art NWP models, it remains true for CTMs (c.f. notably the answer A54). We nevertheless think of revising the title: "Assimilation of IASI and MLS ozone data into a global CTM: assessment of the impact of observational and model spatial resolution".
For the point (ii) we keep the term "global" as in the context of a model it generally means that the model covers the whole Globe (as opposed to, e.g., just covering a regional area as the Europe). For example, this is the case in Flemming *et al.* (ACP 2011) for NWP models, or in Ziemke *et al.* (JGR 2006) for the Global Modeling Initiative's CTM. Here we associate the term "global" to the horizontal domain covered by the MOCAGE CTM that is therefore independent of the chosen chemical scheme. We also do not think that using a simplified parameterisation as a chemical scheme makes this version of MOCAGE no more a chemical transport model. Many models use simplified chemical schemes, and the work of Flemming *et al.* has shown that the linearized scheme used in our study can perform as well as the most sophisticated schemes in the assimilation framework.

- Flemming *et al.*, Forecasts and assimilation experiments of the Antarctic ozone hole 2008, ACP, 2011.
- Ziemke *et al.*, Tropospheric ozone determined from Aura OMI and MLS: Evaluation of measurements and comparison with the Global Modeling Initiative's chemical Transport Model, Journal of Geophysical Research, 2006.

* Referee #1:
The study is thoroughly carried out and includes some interesting aspects on the constructions of super-observations and the scale dependency of the background-error specification. However, it does not contain enough scientifically interesting results within the scope of ACP.
Referee #2:
Unfortunately, though the study presents a lot of detail it does not go into enough scientific depth in any area. It would benefit from a much greater focus on the interesting science, and a greater depth of investigation in those areas. As it stands there is not enough scientific content to recommend publication.

A2 Answer:

We still think this paper contain enough scientifically interesting results within the scope of ACP. In particular it contains useful information for those who would like to assimilate a combined product of columns and

profiles at a higher resolution than it is typically used. We hope that the following answers will convince the referees it is the case.

* Referee #2:
Separately, there is a need for much better presentation: the paper is confusing and vague, particularly in many important aspects like the different spectral and spatial representations, grid resolutions, and the superobbing. The results would be easier to follow and more interesting to read if condensed and restructured. The division of the results into three sections (first model (4.1), then analysis (4.2) and then high resolution analysis (5)) and then each of those into different subsections (e.g. ERAinterim, sondes, OMI) makes for much repetition. As mentioned before, some of these results are presented in too much detail given what they contribute to the conclusions of the paper.

A3 Answer:
In our revised manuscript, we clarify the important aspects which are mentioned. In particular, we will put forward a new designation for the grid resolutions. We agree that some results need to be condensed, but we believe the structure of this paper is logical as it allows distinguishing the impact of the model spatial resolution, without and with assimilation, and the spatial resolution of the observations.

****
* Referee #2 "main scientific points": *
****

A)
* Referee #2:
The effect of the background error correlation length scale is scientifically interesting and rightly given prominence in the paper's abstract. However, there is no justification for the choices made and no attempt at further scientific investigation in this area. For such a crucial part of the paper, this is completely insufficient. All that is said can be found on p. 29369: "horizontal length scales are set to 275km for the assimilation <of data superobbed at T42>."

A4 Answer:
The value of 275km has been chosen following arguments developed by Massart *et al*. 2009. Using ensemble methods they found that the length-scales is strongly inhomogeneous in time and space with values ranging from 100 km to 300 km in the meridional direction, and from 100 km to 600 km in the zonal one. They selected a trade-off value of 220 km that corresponds to the length of a grid cell at the equator for their model. We thus followed a similar approach with a horizontal length scale of 275km, that corresponds to a distance of 2.5° at the equator (the T42 low resolution corresponding approximately to 2.8° x 2.8°). We will explain our choices in the revised version of the manuscript. Note that Massart et *al*. (2011) analysed also the influence of the horizontal length scales on the analysis. They show that their effects are mostly of the second order compared to the specification of the background error standard deviation.

* Referee #2:
"As shown in Bouttier (1994), the horizontal length-scales decrease in areas with dense observations, so we choose to set the horizontal length scales to 77km for <assimilation of data superobbed at T170>". And in the conclusion "the horizontal correlation length scale of the background error has to be adapted to represent the small-scale structures present in the observations". The Bouttier reference is not justification: background errors are supposed to describe the characteristics of the first guess, not the resolution of the observations, so this is scientifically incoherent. Given that the authors already apply some of the Desroziers et al. (2005) diagnostics, they could perhaps have determined the appropriate background error length scales in a similar manner to Bormann et al. (2010).

A5 Answer:

Yes the background errors describe the characteristics of the first guess. But these characteristics evolve with the dynamics (Bouttier, 1993) and are constrained by analysis cycles (Bouttier 1994). See, for instance, the conclusion of Bouttier (1994) "Without a model error term, the estimated covariances exhibit very interesting features both in terms of variances and correlations, with a large variability in space and time. The largest variability is found where the observations are sparse, so that the error structures depend strongly on the model dynamics". Bouttier (1994) clearly discusses the effect of the spatial resolution of the observational network. This is also discussed in detail in sec. 4(b), p 2384, left col.: "During the prediction cycles, all correlations tend to spread and take a more complicated structure, as described in Bouttier (1993). The correlations between points that are far from the observations are not modified during the analysis". In a reverse way, correlation functions in the vicinity of observations are constrained along the analysis leading to shorter length scales. The effect of the global data heterogeneity has also been noticed by other author e.g. see Fig.10, p 2478 of Pereira and Berre (2006) where they show that length scales are broader in the North hemisphere than in the South one.

- F. Bouttier, 1993, The dynamics of error covariances in a barotropic model, Tellus A, 45: 408-423.
- M. Belo Pereira and L. Berre, 2006, The use of an Ensemble approach to study the Background Error Covariances in a Global NWP model. Mon. Wea. Rev. 134: 2466-2489.

* Referee #2:
Ultimately, the paper demonstrates that 77km is an inappropriate choice. But what would be a better choice? Would better quality ozone analyses be possible?

A6 Answer:
This conclusion is right but it needs to be specified: "Ultimately, the paper demonstrates that 77km is an inappropriate choice for the assimilation of this combination of MLS profiles and IASI total columns without averaging kernels." Obviously, the choice of horizontal length scales set to 275km "for <assimilation of data superobbed at T170>" cope with the problem of the deterioration of the vertical ozone distribution. But then the results are nearly identical to those obtained "for <assimilation of data superobbed at T42>", which is not satisfactory. Since no correlation is considered in our observation error covariance matrix R, our system does not allow to differentiate the shape of the increments brought by the data assimilation of MLS profiles from those of IASI total columns. Possibly we could set the horizontal length-scales to 165km that corresponds to a distance of 1.5° at the equator and represents the distance of two successive MLS scans. But it would only be a trade-off between the encountered problem and the objective of assimilating the maximum of information at the lowest horizontal scale. If it was possible, the best choice would be the use of estimated anisotropic and inhomogeneous horizontal length-scales.

B)
* Referee #2:
The combination of limb profiles and total columns in ozone data assimilation is an interesting subject that has seen much previous research (e.g. Struthers et al. 2002 and many following papers). The authors need to acknowledge this and to present their research in the context of all that previous work, particularly in section 3.2.

A7 Answer:
In our revised manuscript we will give more credit to papers that have used the combination of limb profiles and total columns in ozone data assimilation, e.g. Struthers *et al.* 2002, Massart *et al.* 2009 and other papers. But we will remind that in our work the combination of limb profiles and total columns is used to cope with the absence of vertical information of the total columns, in a specific situation where computational limitations do not allow the use of this vertical information.

C)
* Referee #2:

Given that so much of the paper concerns model resolutions and grids, the presentation is far too vague in the way these different concepts are described.

A8 Answer:
We will clarify the description of the different model resolutions in the revised manuscript.

* Referee #2:
There is an ongoing confusion between spectral and spatial representations, with for example "T170", which describes a spectral representation, being used to refer to a Gaussian spatial grid.

A9 Answer:
We agree that the choice made for the grids names can be misleading. In the paper the "T42" and "T170" referred to Gaussian spatial grids associated to the spectral representation. We will adopt an alternative notation. As the common notation to refer to a Gaussian grid is "N*" and as we want to distinguish the "Full" Gaussian grid from the "Reduced" Gaussian grid, we respectively call them "FN*" and "RN*". Thus in the revised manuscript "T42" becomes "FN32" and "T170" becomes "FN128".

* Referee #2:
This means that both the "T170" version of MOCAGE (on p. 29361) and the "T255" ERA interim analyses (on p. 29364) are described as having a 0.7 degree spatial resolution. This is certainly confusing and quite possibly wrong.

A10 Answer:
It is exact these both Gaussian grids have a 0.7 degree spatial resolution. As discussed above the grid of the horizontal resolution of the reanalysis products of ERA-Interim will be referred as "RN128" as it is a N128 reduced Gaussian grid.

* Referee #2:
The paper needs to clearly introduce these different kinds of grids and representations at the beginning of the paper, including references. There are dozens of confusing statements in the text, and indeed dozens of different grids and resolutions are mentioned: e.g. in the work of the other groups mentioned in the introduction; previous MOCAGE work on 2 by 2 degrees; ECMWF operations and ERA interim (which use both spectral and spatial grids); the two new MOCAGE grids described in the paper (whatever these are, spectral or spatial - I am still confused); the grids on which the Valentina analyses are made (spectral?); the grids used for the superobs; the original observation sampling patterns. This makes for a confusing read, and may reflect a real scientific confusion too.

A11 Answer:

We do not believe it is needed to introduce all the different kinds of grids at the beginning of the paper since most are only quoted one time. In our revised manuscript we will say that the Valentina analyses are made in the spectral space (see P. 29368.26-28) with the triangular truncations T42 and T170. The alternative notation (FN32/128) we will adopt elsewhere in the manuscript will clarify when physical fields are given in the physical space on Gaussian grids.

D)
* Referee #2:
ERA-interim ozone analyses are presented as a reference for validation. If that is so, is there any point in the current work?
* Referee #1 ("General comments"):
I am not so sure what can be learnt from the comparison with the ERA interim data set apart from a simple sanity check. Both ERA-interim and the MOCAGE system are similar in the horizontal resolution, the chemical scheme (Cariolle) and the assimilation technique (incremental 4d-VAR) and both assimilate MLS.

A12 Answer:

The ERA-Interim dataset and our ozone analyses are not independent (see P. 29365.3-5). Nonetheless, the ERA-Interim ozone reanalyses being a dataset, global in time and space, that has been largely validated by comparison with independent data (see P. 29364.16-18), they are well adapted to evaluate our results.

****
* Referee #1 "General comments": *
****

* Referee #1:
The motivation for the construction of the combined MLS and IASI data sets remains unclear. Why is a combination of the two data sets needed and why is beneficial?

A13 Answer:
Due to computational limitations, we had to discard the averaging kernels associated with the IASI retrievals (29363.1-4). As a consequence the distribution of IASI ozone data is not constrained on the vertical. In our experiment, a constraint on the vertical ozone distribution over the stratosphere and the upper troposphere is ensured by the joint assimilation of Aura/MLS observed profiles (29363.5-7) that is a low computational cost solution (29358.10-12). The ozone vertical profiles of the model could thus have a different shape after the assimilation of IASI and MLS data.

* Referee #1:
How does the assimilation of either MLS and IASI compare to assimilation the combined data set.

A14 Answer:
Even if the assimilations of either MLS or IASI have been done separately in a preliminary work, we choose not to present them in our article since it does not add much to the discussion. On the whole, the zonal mean of the monthly average of the ozone fields coming from the MLS data assimilation is very consistent with the one coming from the assimilation of the IASI T42 (FN32) & MLS dataset, with the largest differences at the polar vortex boundary around 75° S at 50 hPa. Total ozone columns coming from the assimilation of MLS data are higher (of the order of 20 Dobson) at the South pole than those coming from the assimilation of the IASI T42 (FN32) & MLS dataset that seems less realistic.

* Referee #1:
The Coriolle chemical scheme has proven to be very useful for the simulation of stratospheric ozone chemistry in assimilation systems. It should be noted that it is based on the relaxation to modelled 2D climatological temperature and ozone fields, which are of low resolution. This means that the study can not really explore the rather interesting aspect of the scale-dependency of the chemical conversion, in particular during the ozone hole formation. The "scale" of the applied linear scheme is the same in the T42 and the T170 experiments and it is genuine resolution is probably even coarser then T42.

A15 Answer:
As given by Cariolle and Teyssèdre (ACP 2007 page 2189) article, the parameters of the ozone linear scheme are determined using the MOBIDIC photochemical model working on a 2.8° resolution. In our study the coefficients of the parameterizaton are simply interpolated on the grid used. So it is true that by construction the specific "scales" involved in the formation of the ozone hole are not resolved. Although those scales corresponds to PSC formation and chlorine activation, and are associated to non-linear chemical interactions. So the problem is not only a "scale" problem but rather that in the situation the linearity of the system can be questioned. This problem has been addressed by Monge et al. 2011, where an attempt to derive coefficients from 3D calculations with a global CTM that include comprehensive chemistry is used. The conclusion is that the system can still be linearised and the simplified approach can still be retained.

- Monge-Sanz, B. M., M. P. Chipperfield, D. Cariolle, W. Feng. Results from a linear O_3 scheme with embedded heterogeneous chemistry compared with the parent full-chemistry 3D CTM. Atmos. Chem. Phys., 11, 1227-1242, 2011.

* Referee #1:
Since the study does not apply a full chemistry scheme, the model resolution can not be compared in a fair way to CTMs with a more less complex chemistry scheme. The model resolution should be compared to applications of simplified schemes for instance at operational NWP centres (e.g NCEP, ECMWF, UKMO). The horizontal resolution of these application is higher than T170 in most cases. In this respect, the term "high resolution" seems to be an exaggeration.

A16 Answer:
The aim of the paper is not to compare the CTMs or the chemical schemes they use themselves. We provided examples of the resolution of the most common CTMs just to illustrate the fact that they have a lower horizontal resolution than the pixels size of the measurement from new generation on-board instruments. Increasing the resolution is a natural way for most of the Geosciences models. The focus of the paper are (i) to show which improvements in the ozone fields we can expect when the horizontal resolution is increased and (ii) what is the caution that has to be adopted to assimilate data such as the IASI ones at these resolutions. And even if we would use a more complex chemistry scheme, the message would remain the same.
Note that, at least for ECMWF, ozone data are operationally assimilated to control the dynamics and not to control the ozone concentration. As far as we know, the assimilation of IASI ozone data is an on-going task. And this paper could help these centres for example to be convinced to use the IASI averaging kernels while assimilating IASI ozone data.

* Referee #1:
The methodology to combine the IASI and the MLS data needs to be better explained. Is the combined data set a partial column data set (as MLS) or a total column data set (as MLS).

A17 Answer:
The methodology to combine the IASI and the MLS data is described in Massart *et al.* (ACP 2009). We will nevertheless better explain the methodology in the revised version of the paper. But note that the combination of the two data sets does not mean the production of one data set including the two sources of information. This is not possible as we have both columns and profiles. This means that we assimilate together the two data sets with their own characteristics. This is the strength of the assimilation process.

* Referee #1:
If the IASI data are bias-corrected according to the MLS data, how big is this correction. How are IASI's total column observations corrected against the MLS partial columns.

A18 Answer:
The bias is monthly averaged for September and computed functions of the space. It could reach 15% of the total ozone column.

* Referee #1:
Is the aggregation to the T42 and T170 resolution done before or after the merge of the MLS and IASI data. Since IASI and MLS represent different horizontal resolutions, what is the effective resolution and the observation error of the combined data set.

A19 Answer:
The IASI super-observations are obtained from the T42/T170 (FN32/FN128) full Gaussian grids, and datasets are referred as T42/T170 (FN32/FN128) IASI data (29363.10-14), while the MLS super-observations are obtained from a regular 2° x 2° resolution grid (29363.27-28). Two datasets are assimilated:

the T42 (FN32) IASI & MLS dataset which combines/aggregates (but does not merge) the T42 (FN32) IASI data and the "2° x 2° MLS data", and the T170 (FN128) IASI & MLS dataset which combines/aggregates the T170 (FN128) IASI data and the "2° x 2° MLS data". Thus an effective horizontal resolution cannot be derived from the combined data set.

* Referee #1:
Differences in the modelled ozone total columns fields with different model resolution (Ml vs. Mh) are attributed only to differences in the vertical velocity. It should be better explained why the authors come to this conclusion. Convection or numerical aspects of the advection scheme could also play a role.

A20 Answer:
We agree numerical aspects of the advection scheme can also play a role in the differences seen between the Ml and Mh simulations (not convection since it has been ignored in our study). However previous studies (e.g. Monge-Sanz *et al.*, 2007) have shown that the meridional circulation (and hence the vertical velocities) given by the ECMWF analyses that we use in our study is sensitive to the resolution of the model. The highest resolutions tend to give a more intense equator to pole circulation in the lower stratosphere. This is confirmed in our study.

- Monge-Sanz, B., M.P. Chipperfield, A. Simmons, S. Uppala. Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophys. Res. Lett.*, 34, L04801, doi:10.1029/2006GL028515, 2007.

* Referee #1:
Further, it has to be better explained how the vertical velocities for the T42 experiment were derived from ERAI in a consistent way.

A21 Answer:
To obtain the vertical velocities for the T42 (FN32) experiment, we simply truncated the horizontal winds from the ERAI resolution to the T42 (FN32) one. And then we recomputed the vertical velocities from these T42 (FN32) horizontal winds in order to ensure the conservation of the mass using the divergence equation. We also recomputed the vertical velocities for the T170 (FN128) experiments.

* Referee #1:
The resolution dependency of the background error statistics and the observation error statistics is a vital point for the validity of the study. It seems that the background error standard deviation is considered to be the same for the T42 and T170 experiment, i.e. it is interpolated from values derived for a 2x2 grid. What is the motivation for this?

A22 Answer:
The background error standard deviation for the T42 (FN32) and T170 (FN128) experiments are interpolated from values derived for a 2° x 2° grid. Diagnosing the background error standard deviation is a difficult task. The way we did it is to combine the variational assimilation with ensemble techniques (see Massart *et al*. (2011) for more details). This methodology is very expensive and we can not afford such a methodology for the T170 (FN128) experiment. Nevertheless, as we used posterior diagnostics to adjust the errors, we can consider that the shape of the background error standard deviation is similar for the T170 (FN128) experiment.

* Referee #1:
The background errors statistics should be model-resolution dependent.

A23 Answer:

The referee is right. The background errors statistics should be model-resolution dependent and this aspect of our experiments may be questionable. We wanted to make experiments as much comparable as possible. Thus we made the choice to use same values in each experiment for the background and observational multiplicative correction factors sb and so. The factors sb and so have been chosen to be optimal for the Ahh experiment. This will be discussed in the revised article.

* Referee #1:
The construction of super-observations by averaging observation over a given area will change the observation error statistics. The random observation error of the super observations should be smaller than the one of the individual observations. The differences in the random and representativeness error will depend on the variability of the observation within the averaging area. Has this been taken into account? More detail on this is needed to better understand the impact of the technical averaging procedure to the T42 and T170 grid.

A24 Answer:
We agree that the information given on the IASI error statistics at the pages 29370 and 29371 does not answer this point. We have followed Massart *et al.* (2009) to compute the standard deviations of the observation error for IASI. We computed them by a comparison of the two sets of super-observation with previous MLS analyses respectively computed on the T42 (FN32) and the T170 (FN128) Gaussian grids. This allowed us to have different observation error statistics for the two sets of super-observation. We have thus replaced the sentence line 2 page 2937 "We also assumed that the standard deviation of the differences is the standard deviation of the IASI measurement error." by "We also used the gridded differences (both on the FN32 and FN128 Gaussian grid) between IASI measurement and the total ozone columns computed from the MLS analyses to compute the standard deviations of the observation error. Following Massart *et al.* (2009) we assume that the standard deviation of the differences in each grid cell gives the standard deviation of the IASI measurement error. We computed the standard deviations of the observation error for the two sets of super-observation by computing the standard deviation of the differences gridded respectively on FN32 and FN128 Gaussian grid."

* Referee #1:
An alternative approach to the averaging of the observation is to randomly pick one observation. The differences between the two approaches should be discussed.

A25 Answer:
Although it represents an important aspect of any assimilation system, the aim of the paper is not to compare the different possible approaches to assimilate a dataset when there are several of its observations within a model cell. We remind (see answer A16) the focus of the paper are (i) to show which improvements in the ozone fields we can expect when the horizontal resolution is increased and (ii) what are the precautions we have to take to assimilate data such as the IASI ones at these resolutions. Randomly picking one observation within a model cell would not have affected the impact of the background horizontal correlation length scales on the analyses. As a consequence, we do not think to discuss these two approaches in the revised version of the manuscript.

* Referee #1:
The validation of the model and assimilation result for only one month seems too short. What was the motivation to choose this period?

A26 Answer:
We agree this period of validation is short as it is recalled in the conclusions of the paper (29385.11-14). We chose this period as it sees the early formation of the Antarctic ozone hole with large ozone gradients at its boundary and filamentary structures extracted at the edge of the associated polar vortex, a situation that is particularly interesting to study the quality of a global ozone analysis (29361.11-16).

* Referee #1:

The comparison against ozone sondes and OMI is useful. It should be specified, which ozone sondes were used.

A27 Answer:
We will specify the 11 ozonesondes we used in a table.

Lat	Lon	Name
-37.69	144.95	Melbourne / Airport
37.93	-75.47	Wallops island
40.47	-3.58	Madrid / Barajas
47.80	11.02	Hohenpeissenberg
51.93	-10.25	Valentia observatory
52.10	5.18	De Bilt
-54.85	-68.31	Ushuaia
-5.49	-35.26	Maxaranguape (SHADOZ-Natal)
-69.01	39.58	Syowa
-70.65	-8.26	Neumayer
-7.98	-14.42	Ascension island

* Referee #1:
Using the IASI data as total columns without the application of averaging kernels seems to be a missed opportunity. In particular, the tropospheric signal of the IASI data could have been exploited in a better way. It is complementary to the MLS data which provide a more a stratospheric profile.

A28 Answer:
We agree with referee 1, the need to use IASI averaging kernels is one conclusion of this paper (29385.3). However the fact that we did not use the AK is not a missed opportunity but is a deliberate choice to limit the computational cost of the high resolution experiments.

****
* Referee #2 "detailed scientific points": *
****

1) 29358.8 (see point C) a "gaussian grid" cannot have a triangular truncation and cannot be described by the term T42 or T170 - those are spectral representations.

A29 Answer:
See the answer A9.

2) 29358.16-17 and 29384.9-10 "this modification results from a better representation of the vertical velocity". There is no scientific backing for this statement in the paper. Unless some can be added, these statements should be deleted.

A30 Answer:
See the answer A20. The reasoning leading to that conclusion will be described with more details in the revised paper.

3) 29358.18-19 "in a general way comparisons with independent data show large reductions in the ozone standard deviations when the resolution is increased" and 29384.11-17 "increase of the CTM ... resolution ... showed a mainly positive but small influence on the ozone analysis": there does not appear to be any basis for these statements in the paper, e.g. in Figs. 9 and 10. These statements should be removed.

A31 Answer:
The first sentence (29358.18-19) has to be mitigated. Indeed Figs. 9 and 10 show that when the resolution is increased, comparisons with independent data may show reductions at some locations in the ozone standard deviations as well as increases at other locations. However Figs. 7 and 8 show mostly ozone standard deviations reductions in the analyses when the model resolution is increased. As the comparisons with ERA-Interim dataset are relevant the statements should not be removed.

4) 29360.13-14 "... superobservations close to the datasets resolution...". This is confusing. What's the point of a superobservation that's almost the same resolution as the dataset itself? Here there are just one or two raw observations per superobservation. Why not just use the raw observation?

A32 Answer:
We agree that the sentence is confusing and is rephrased as: "However a model resolution increase is required to decrease the number of observations assimilated within a model cell and thus to better take into account the small scale information available in the dataset.".

5) 29361.2 (see point C again) "so-called T42 low resolution" - we need references and a precise definition of what this means.

A33 Answer:
See the answer A9.

6) -» 29364.1 "we had to construct super-observations". Is that really true? In the present study, MLS data could have been assimilated at its natural resolution. Wouldn't that be easier?
-» In the case of only one observation per superob (which would be quite common), the only difference between a superob and an ordinary observation would be that a superob would represent a model grid point whereas an ordinary observation would require an interpolation operator. The second approach is the normal one, as it can deal better with sharp gradients. As mentioned in technical point 2, aspects relating to superobservations need to be presented better.

A34 Answer:
-» The referee is right. In Massart *et al.* (2009) MLS measurements have been averaged in order to assimilate a dataset in a manner consistent with the use of the other assimilated data sets. MLS data could have been assimilated at its natural resolution but we preferred to assimilate superobservations that have been well characterized in Massart *et al.* (2009) and Massart *et al.* (2011).
-» The referee is wrong. In the case of only one observation per superobservation, there is no difference between the observation and the superobservation.

7) 29364.19-21 "N128 reduced Gaussian grid" (see point C) - be more consistent in describing the different grids.

A35 Answer:
See the answer A9.

8) 29365.6-8 (point C again) ERA interim data are interpolated from the N128 grid to the T170 grid. But it's already indicated that both these grids have the same 0.7 degree resolution with 512 by 256 points (on p. 29361 and 29364). Why interpolate if they are the same grid?

A36 Answer:

As said above (A9), they are not on the same grids. ERA interim data are given on a RN128 reduced Gaussian grid and our high model resolution data are given on a FN128 full Gaussian grid. Thus the data need to be interpolated before being compared.

9) 29366.14 Only 11 ozonesonde stations are used, with the result that it is very hard to assign statistical significance to the differences shown in Fig. 9 later. There are plenty more ozonesondes out there - why not use them too? The study referred to in this section, Geer et al. (2006), used data from 42 different ozonesonde stations.

A37 Answer:
At the time when the comparisons have been done, other ozonesondes data were not freely accessible.

10) 29368.2-9 "ECMWF operational analyses suffer from a lack of dissipation...too much energy remains in the small scales": What is meant by "dissipation"? Against which reference can these statements be justified? The production of "unrealistically noisy ozone fields" does not justify this. Perhaps there is a problem in the timeresolution, or in the way the interpolation or spectral truncation is done? Using analysed winds to drive CTMs is widely acknowledged to be tricky; many papers have been published on the subject.

A38 Answer:
By "lack of dissipation", we mean too much energy remains in the small scales of the dynamical forcing fields from ECMWF operational analyses as it is illustrated on the figure below. See the answer A20 for a reference. We do not believe there is a problem in the time resolution since we also tried to use these dynamical forcing fields with 3h forcing steps and that did not resolve our problem whereas it did with ERA-Interim.

Figure: Energy spectra of the meridian (left) and longitudinal (right) velocities for the 19 August 208 at 00h UTC at the 16th MOCAGE level (around 15 hPa). In black: spectra derived from the T799 ECMWF operational analysis. In blue: spectra derived from the T255 ECMWF ERA-Interim reanalysis.

11) 29369.10-13 "we apply diagnostics like those of Desroziers et al. (2005). A bit more detail is needed here. Also, scientifically, isn't it strange that Massart et al. (2011) determine s_o = 10 (s_b is not specified) but you get s_o = 2 and s_b = 10? Why are the correction factors so different?

A39 Answer:
We believe the referee made a mistake by reading Massart *et al.* (2011). It is written in Massart *et al.* (2011) at the 5th page: "The value of so is adjusted for each assimilation window. But for the assimilation experiments we ran [...], the average value is 1.69+-0.1. Figure 1 shows the average standard deviation of the observation error as the MLS instrumental error multiplied by the 1.69 factor." and at the 6th page: "But a posteriori diagnostics presented in Section 2.2.6 suggest to multiply the estimated standard deviations by a factor sb of about ten.". Thus they used s_o = 1.69+-0.1 and s_b = 10 in Massart *et al.* (2011) whereas we used s_o = 2 and s_b = 10. These results are consistent.

12) 29371.2 and 12. The standard deviation of departures is assigned as the IASI measurement error on line 2. In data assimilation, the departure covariance matrix is $R + HBH^T$, so this gives a maximum bound to the observation error R. So why then is an additional inflation factor, s_o = 2 applied on line 12?

A40 Answer:
The referee is partly right. $R + HBH^T$ is the background departure covariance matrix, i.e. $E[(y-Hx_b)(y-Hx_b)^T] = R + HBH^T$. But in that part of our study, we are using the departures between the IASI observations and the analyses from the MLS data assimilation. The methodology is described at the page 5083

of Massart *et al.* (ACP 2009). We will nevertheless better explain the methodology in a revised version of the paper.

13) 29373.5 "all values diverge". But why does one go up and the other two go down?

A41 Answer:
This is presumably due to the difference in the model resolution. Desroziers diagnostics are set up functions of a system, in terms of model and observations. During the absence of IASI data from 16 to 20 September, "all values diverge" and then the background function cost takes a larger part in the value of the total cost function. The values diagnosed for s_b shown on fig. 2 are greater than 1 for All and lower than 1 for Ahl and Ahh that confirms the influence of the model resolution.

14) 29373.7-14 s_o and s_b are described here. But how do these relate to the values mentioned in section 3.2? Are these the values applied in the assimilation, or are these new values? And why not put descriptions of the Desroziers (2005) method together in one place, rather than in both 3.2 and 3.4?

A42 Answer:
The s_o values from the fig. 2 could be applied to the combined IASI T42/170 (FN32/128) & MLS datasets. s_o and s_b are new values for the coefficients that may be applied but are just shown as consistency diagnostics. Since the results of the section 3.4 are only diagnostics whereas the a posteriori diagnostics of the section 3.2 are applied, we prefer not to put the descriptions in the same section.

15) Fig. 2. What is the reason for the spike in s_b in experiment All on 7th September?

A43 Answer:
We have no explanation to give to that spike. But it should definitely be an interesting point of investigation.

16) 29374.8 "T170 horizontal grid" (see point C again). Again a confusion between horizontal representations.

A44 Answer:
See the answer A9.

17) 29375.4-10 Why, physically, is there less ozone between 45S and 45N?

A45 Answer:
We believe it is due to a stronger meridian circulation (with the T170 (FN128) high resolution model) mainly due to the differences in the vertical velocities (see the answer A20).

18) 29375.24-29 This discussion on vertical velocity is vague and unsupported. Either remove, or provide supporting evidence.

A46 Answer:
See the answer A20

19) 29376.7-9 Low resolution results are being "projected" on the "T170" grid. "Thus the error of representativeness is the same for all the experiments". This can't be true. It would work if you averaged the high resolution results to the lower resolution, but not the other way round.

A47 Answer:
The referee is right and the sentence "Thus the error of representativeness is the same for all the experiments" will be deleted in the revised version of the manuscript. However, this does not influence the validity of the comparisons of our results with the OMI T170 (FN128) data.

20) 29377.23-25 "the Ahl experiment has globally a better agreement with the ERAinterim reanalysis than the All experiment". Is this any surprise given that MOCAGE is being driven by the ERA interim wind fields, which come on the N128 Gaussian grid?

A48 Answer:
We concede it is not surprising, but we think of it as a guarantee the data assimilation system drives well.

21) 29380.20 With 275km for the correlation length scale "we are not able to represent the structures that are smaller than that size". That is untrue. The model should be able to generate structures at its own length scales. It is only the analysis increments, not the model itself, that will be smoothed by the background errors.

A49 Answer:
The referee is right and we had replaced the sentence by "the analysis increments are not able to represent the structures that are smaller than that size".

22) 29383.22 "the solution is likely to involve using the IASI AKs". But also it needs a better choice of background error correlation length.

A50 Answer:
See the answer A6.

****
* Referee #1 "Specific comments": *
****

-> P 29358
l 11 The tropospheric columns of MLS have a big error and its use is not recommended.

A51 Answer:
Only stratospheric and upper tropospheric levels of Aura/MLS observed profiles (See P 29363 L 5-7) are assimilated.

L 23 If it so difficult to combine the data why combine them?

A52 Answer:
The sentence "This result is due to the difficulty to combine IASI data and MLS data." is ambiguous and it will be rephrased in the revised version of the manuscript by: "This result is due to the difficulty to extract IASI high resolution information without using its averaging kernels".

-> P 29359
L 13 MOCAGE of this study does not use a full chemistry scheme – please add references for applications with simplified schemes (Geer et al. 2007)

A53 Answer:

See the answers A1, A15 and 16.

L 27 The coupled chemistry assimilation system at ECMWF (using the full chemistry scheme of MOZART-3) has a resolution of about 1.1° and finer (TL159 and TL255) The system was applied for multi sensor ozone assimilation during the period August-December 2008 (Flemming et al. 2011, ACP)

A54 Answer:
The assertion of referee 2 is wrong. It is written in Flemming *et al.* (2011) in section 2.4 at the pages 1967 and 1968: "The IFS was run with a T159 spectral resolution and the grid point space was represented by a reduced Gaussian grid (Hortal and Simmons, 1991), which has a grid box size of about 125 km. [...] The CTMs MOZART-3 and TM5 were run in the coupled experiments on a regular latitude-longitude grid of 1.9°×1.9° and 2°×3° grid box length respectively; they used the same vertical discretization as the IFS." Thus only the linearization of the ozone chemistry has a resolution of about 1.1° (not finer), the full chemistry schemes being coarser.

-> P 29361
L 4 What is the vertical resolution and the model top? How is convective transport modelled?

A55 Answer:
The atmosphere is divided into 60 vertical layers from the surface up to 0.1 hPa. The vertical resolution is finer in the planetary boundary layer and coarser in the stratosphere and mesosphere (C.f. P. 29367 L 15-16 that refers to P.29364 L21-23).

-> P 29363
L 3 It would be very interesting to quantify the impact of the assimilation of the IASI data with and without AK for the low resolution to get a feeling of the importance of this simplification.

A56 Answer:
It is scientific interesting information that has seen much previous research. But as we recall in the answer A16, it is not the aim of this paper.

-> P 29364
L 20 The reduced grid has 512 longitude points only near the equator

A57 Answer:
The referee is right and it will be corrected in the revised version of the manuscript.

-> P 29366
L 14 What is the motivation of this choice of stations. Please provide better information on which stations were used.

A58 Answer:
See the answers A27 and A37.

L 23 This sort of stratification seems not justified if only 11 stations were used (mainly in the SH)

A59 Answer:

-> P 29367
L 9 Provide more information on the processing of the meteorological fields for the T42 and T170 runs.
Is it simply a truncation of the spectral fields? How are the vertical wind fields obtained?

A60 Answer:
See the answer A21.

-> P 29372
L 18 Why does this threshold does not depend on the resolution?

A61 Answer:
This threshold evaluates the consistence of the observation towards the running model, so it should not depend on the resolution.

-> P 29373
L5 Since the consistency measure deteriorates if no IASI data are available, does it mean that the error measures of MLS is not correctly specified ?

A62 Answer:
That could be an interpretation. But we rather believe that, as the multiplicative correction coefficients have been computed for whole data including IASI data and MLS data, they are worse adapted when no IASI data are assimilated.

-> P 29375
L 2 "significant" = is this based on a statistical test

A63 Answer:
The sentence is wrong (See A95) as the biases are only significant beyond 80° S. By "significant" we mean the bias is positive even considering the associated standard deviation.

L 26 Please clarify how the vertical winds are obtained and if they are consistent with the horizontal winds, or simply interpolated from the ERAI vertical winds?
How is convection treated in the model?

A64 Answer:
See the answer A21. Convection has been neglected in our study.

-> P 29376
L 1 The title of this is section is confusing – please mention that Ahl and All will be compared, i.e. the impact of the resolution of the model and not the observations.

A65 Answer:
It will be specified in the revised version of the manuscript.

-> P 29379
L 25 The following paragraph is not clear. OMI has been assimilated in ERAI.

A66 Answer:

In the revised version of the manuscript, the sentence "As for the comparison with the ERA-Interim reanalysis, the difference between the All and Ahl experiments in terms of standard deviation of the differences with OMI data does not exceed 1 %, with largest differences at the polar vortex boundary." Will be replaced by "As for the comparison with the ERA-Interim reanalysis, zonal averages of the standard deviations between All and OMI and between Ahl and OMI does not exceed 1 %. The largest differences between the two curves are at the polar vortex boundary.". ERA-Interim reanalyses assimilate many other observations than OMI data and are influenced by model laws. They do not necessarily share every OMI features, as it is namely shown by the peak in the standard deviations of the differences at the South Pole in Fig. 7 that is not on Fig. 10.

-> P 29380
L 1 Why is Ahh treated in a separate section ?

A67 Answer:
The focus of the paper are (i) to show which improvements in the ozone fields we can expect when the horizontal resolution is increased and (ii) what is the caution that has to be adopted to assimilate data such as the IASI ones at these resolutions. Thus Ahh is treated in a separate section in order to answer the point (ii).

L 18 The length scale is the correlation length of background (model) error. It should reflect the resolution of the model and not the resolution of the observations.

A68 Answer:
See the answer A5.

-> Figures:
Please avoid the comment "see text for details" in the figure captions. Try to say what the values mean (improvement etc.)

A69 Answer:
We think that the details are too long-winded for the figure captions.

Figure 9 Should be complemented with a figure of the biases

A70 Answer:
Unlike the standard deviations, the bias has no sufficient signification for the comparison with few observations.

****
* Referee #2 "Detailed technical points": *
****

1) 29362.24-26 "a lower error and better coverage Massart et al. (2009)": this is meaningless without a detailed knowledge of Massart et al. (2009). Be precise about what is meant and give some justification, or remove this statement.

A71 Answer:
In Massart *et al.* (2009), the observations recorded over icy and sandy surfaces (that represents about 12% of the overall data set) have been discarded. It will be specified in the revised version of our manuscript that they are assimilated.

2) -> 29363.11 "we build superobservations" .Information on the superobservations is too spread out through the text. Please gather together tables 1 and 3 and the discussions associated with them (e.g. the first part of section 5.1). Ideally this could go in the introduction, e.g. in section 2.1 next to where the assimilated observations are described.
-> Please also say how the observation error of the superobs is derived, given that the raw observations come with a specified error. This may be described in section 3.2 but I found that very hard to understand.
-> Also, what longitude and latitude is assigned to a superobservation? Or does a superob represent a model grid point?

A72 Answer:
-> We agree that the table 3 and the discussion associated with it (the first part of section 5.1) could go in section 2.1. We believe the tables 1 and 3 could be gathered together by doing this, but we do not want to merge them into one table. Indeed the table 1 deals with numbers of observation per day whereas the second table deals with numbers of observation per cell, and we want to separate these ideas.
-> See the answer A24.
-> The superobservation coordinates are determined as a geographical barycentre of the averaged observations.

3) 29364.12 Berrisford et al. (2009). There is now also a peer-reviewed reference for ERA-interim: Dee et al. (2011).

A73 Answer:
We thank the referee for this new reference that will be quoted in a revised version of the manuscript.

4) 29364.15 "The ERA interim analysis was made to improve the quality of the stratosphere..." - surely this was not the only reason?

A74 Answer:
The sentence will be moderated in a revised version of the manuscript: "The ERA interim analysis was notably made to improve the quality of the stratosphere..."

5) 29365.15 "OMI combines the advantages of .... with the characteristics of ...". This is meaningless unless you also say what those advantages or characteristics are.

A75 Answer:
OMI derives its heritage from NASA's TOMS instrument and ESA's GOME instrument (on the ERS-2 satellite). It can measure many more atmospheric constituents than TOMS and provides much better ground resolution than GOME (13 km x 25 km for OMI vs. 40 km x 320 km for GOME). That will be specified in a revised version of the manuscript.

6) 29365.24 "agreement better than 2%" in what sense, i.e. mean or standard deviation, accuracy or precision?

A76 Answer:
It is the average difference with the ground-based observations, the standard deviation being lower than 0.5%.

7) 29366.4 "using its nearest neighbours for the spatial interpolation". What's the point of a horizontal interpolation if you have already superobbed onto the spatial grid of the model?

A77 Answer:
The superobservation has a geographical position (see the answer A72).

8) 29366.7 "a rejection threshold of .. 40%". Why is this rejection done?

A78 Answer:
This rejection has been done because the OMI instrument has measurement problems, mainly at the North Pole, at the end of the month of September.

9) 29366/8-10 "results remain consistent for the whole month... ". What results, and in what way do they remain consistent? This statement is meaningless.

A79 Answer:
The results we are talking about here are the comparisons of our results with the OMI observations. They remain consistent in the way that, without this rejection threshold, a few (wrong) observations cause an explosion of the bias we get, this bias being greater than 40% whereas it is lower than 10% otherwise.

10) 29366.16 "random variability". What does this mean? Accuracy, precision, or just natural variability of the ozone field?

A80 Answer:
It is true as well for the accuracy and the precision and it will be specified in a revised version of the manuscript.

11) 29366.22 "the assimilation is output at 12:00 UTC". Why not interpolate in time like for OMI? Indeed, what time are the ozonesonde observations valid?

A81 Answer:
We did not have access to the exact time of the ozonesonde observations.

12) 29368.1 It is claimed that ECMWF analyses are computed on a "N800 reduced Gaussian grid". This appears incorrect. The analysis and forecast spectral resolution has been T1279 since 2009 and that appears to correspond to N640 on the Gaussian grid:
http://www.ecmwf.int/products/data/technical/gaussian/spatial_representations.html

A82 Answer:
The referee is right. ECMWF operational analyses are computed on a N400 reduced Gaussian grid at the time of our experiments (not N800 as we wrote, but neither N640 since our experiments are made before 2009).

13) 29368.10 "As a CTM, MOCAGE works with a linear time interpolation between two forcing steps." This is imprecise. What is being interpolated in time? And why?

A83 Answer:
The forcing fields (wind velocities and temperature) are interpolated in order to compute the chemical compounds transport. It is a common way of working for the CTMs.

14) 29369.4 "in the physical space" (see point C). This is confusing and would benefit from a more consistent terminology with respect to grids and representations.

A84 Answer:

See the answers A9 and A11.

15) 29369.9-11 "a global bias may appear" but "a multiplicative correction factor" is applied. "bias" tends to imply an offset, not a multiplicative correction, so this a bit confusing.

A85 Answer:
The referee is right. The term "bias" is misapplied here. In a revised version of the manuscript, we rather think of saying that the shape of the errors is correct but not their scale.

16) 29370.7-10 Very long winded. Why not just say "the observation errors are assumed uncorrelated?"

A86 Answer:
This part will be less long-winded in a revised version of the manuscript.

17) 29370.24-26 "for each IASI measurement we compute differences and grid these on the T42 and T170 grid". Aren't the differences already on the grid? Are IASI superobs or raw observations being used here?

A87 Answer:
As, at this time of the process, we used raw observations, we allocated the difference to model cell, and not to the position of the observation.

18) 29371.23 to 29372.6. This is summarised in table 2: a lot of text could be pruned here.

A88 Answer:
We thought detailing one time the nomenclature of our experiments would allow to better the comprehension of the article.

19) 29372.11-12. Line colours are specified in the figure legends. No need to repeat that here.

A89 Answer:
To lighten the manuscript, we agree to delete this description.

20) 29372.25 "divided by half the number of assimilated observations". This makes the simple formula J/2n sound quite confusing. Why not just give the formula?

A90 Answer:
We agree giving the formula $J/(p/2)$ would help the comprehension of the sentence but we believe that its understanding also needs to be formulated by words.

21) 29373.18-19 "observations constructed with MLS data only during that period". This is confusing. You're not merging the MLS and IASI observations together, except in the data assimilation, but that's what's implied here.

A91 Answer:
The sentence "in the observations constructed with MLS data only during that period" will be rephrased by "in the assimilated observations that are only made up of MLS data during that period" in the revised version of the manuscript.

22) 29373.25-27 This paragraph could be deleted.

A92 Answer:
To lighten the manuscript, we agree to delete this description.

23) 29374.9 "energy spectra". Please give a reference or equation for this.

A93 Answer:
A physical field Ψ develops on the unity sphere as:

where *P(n,m)* are the Legendre polynomials, *a(n,m)* and *b(n,m)* are the spectral coefficients, *θ* and *Õ* are the longitude and the latitude, *m* is the zonal wave number and *n* is the total wave number.
The energy spectrum *Sp(n)* is defined in functions of the wave number *n* as:
$Sp(n) = a(n,0)^2 + 2 * Sum(i=1,n) [a(n,i)^2 + a(n,i)^2]$

24) 29374.25 First this sentence says "differences are below 2% over the globe", which implies everywhere. But then it says they "reach 4% over the SP".

A94 Answer:
The sentence will be replaced in the revised version of the manuscript by "differences are below 2% over most the globe".

25) 29375.3 The biases in Fig.4 do not appear to be large or significant at 70S. 90S maybe?

A95 Answer:
The referee is right. The biases are only significant beyond 80° S. That will be changed in the revised version of the manuscript.

26) 29375.14 "Maximum values are located" - note that a figure in ppmv, not mPa, would likely give a very different location for the maximum.

A96 Answer:
This result remains the same in ppmv. The change is that tropospheric values are insignificant.

27) 29377.6-15 This text seems to be just describing the figure without adding any new information. Is that necessary?

A97 Answer:
The text points out the interesting features of the figure. Nevertheless it will be shorten in a revised version of the manuscript.

28) 29378.6,8 "above 68hPa" - this is ambiguous.

A98 Answer:
Looking at the figure helps to dispel the ambiguity. Anyway it will be rephrased: "at pressions lower than 68 hPa"

29) 29378.12-27. Being based on at maximum 47 observations, it's hard to believe that differences between Ahl and All in Fig.9 are significant. Given that, this discussion could be much shorter. Indeed the whole of section 4.2.3 could be reduced to just a sentence or two and joined with the slightly more interesting results given in section 5.2.

A99 Answer:
Even if 47 observations from ozonesondes have been used in the comparison, they represent a valuable source of information on the ozone vertical repartition. We believe the comparisons represent a tendency of the errors that deserves to be described.

30) 29379.15-28. This is quite a long winded paragraph. Could you pull out the key points and forget the rest?

A100 Answer:
We will be as synthetic as possible in the revised version of the manuscript.

31) 29380.24 This text duplicates what's on the caption. Could it be deleted or compressed?

A101 Answer:
It will be compressed in the revised version.

32) 29381.16-21 This text has an appropriate level of detail, and one that would be recommended for the results in section 4 as well.

33) 29384.16 "better agreement .. in terms of variability". Variability is a vague word.
Be more precise.

A102 Answer:
It will be specified: "better agreement .. in terms of variability since it reduces the standard deviation" in the revised version of the manuscript.

34) Table 2 column headings: "hor. res."; "assim" and "hor. len." should be expanded.

A103 Answer:
We believe the abbreviations are explicit.

35) Figure 9: x scale in RH panel should be expanded to include the spike in the Ahh (green) curve.

A104 Answer:
It will be done in the revised version of the manuscript.

36) Figure 11 panel (a). This illustrates one confusing thing about the superobs. They do not seem to be on a regular 0.7 degree Gaussian grid as might be expected. Also, "cumulated" -> accumulated.

A105 Answer:
Between 00:00 and 12:00 UTC, two traces of IASI observations have been made in the observed zone. Also, as it is recalled at the answer A72, longitude and latitude are assigned to a superobservation.

```
***************************
```
* Referee #2 "Typos / grammar": *
```
***************************
```

We thank the referee 2 for all the typos / grammar corrections that will obviously be included in the revised version of our manuscript.