We took into consideration the comments of the two Anonymous Referees and we feel that their suggestions have contributed to improve the quality of our manuscript. The revised manuscript has been significantly modified with the inclusion of one figure, the modification of most of the other figures, and important modifications of parts of the text.

Here we provide a point-to-point reply to the referees comments and suggestions.

Anonymous Referee #1

This paper present performance analyses for a concept geostationary observing system. In general this topic fits well into AMT. The paper is well organized and shows the capabilities of future air-quality geostationary mission over Europe. The instrument configuration and the pseudo-observation simulator are not sufficiently presented in details. The authors should refer more to their previous study Sellitto et al. 2013a and should provide a more descriptive overview of setups and methods. The vertical resolution and lowermost tropospheric ozone sensitivity and the general statistical analysis are interesting enough to make the paper scientific content sound. However, there are a lot of presentation issues. The quality of the figure is low and some parts of the text remains not very easy to understand. The English wording should be improved all through the text.

We thank the referee for the kind words and for the general comments, that helped improving the manuscript. In the revised manuscript, more details on the configuration of our simulator are given (cfr SC4, 6-10), most of the figures are improved (cfr SC19-21, Reviewer #2 SC2), the text has been clarified in several aspects (cfr SC1-3, 11, 15, 17).

Specific comments (SC):

We accept most of the SCs of the Reviewer #1 and we have modified the text and figures accordingly. We report here on some points that need some detailed explanation.

1) P6447, L9: Independent to what?

We have changed the sentence to “We have studied the expected sensitivity of MAGEAQ-TIR and we have found that it is able to provide a full single piece of information for the ozone column from surface to 6 km (about 1.0 DOF (degrees of freedom) and maximum sensitivity at about 3.0 km, on average), as well as a partially independent surface-3 km ozone column (about 0.6 DOF and maximum sensitivity at about 2.5 km, on average).

2) P6449, L16: Again independent to what? I would use a different wording here, for example “a full single piece of information”.

We agree and we accept the wording proposed by the reviewer

3) P6451, L8: I do not think the word “punctual” is appropriate here. Please rephrase this sentence.

We have changed to “...we have done a more detailed simulation...”

4) P6452, L15: “a horizontal resolution of 15 km x 15 km at sub-satellite point, a sub-satellite point at latitude= 0 and longitude= 0, and a field of regard of (15_W–35_ E, 35_ N–65_N).” Please re-consider the comment 1.a of the access review. A

horizontal resolution of 15 km at sub-satellite point is not necessarily 15 km at the field of regard. Please precise if the horizontal resolution at the field of regard is the same as the sub-satellite point or different (and then specify what resolution you really use over Europe).

We added the sentence: “In the present work, we don't simulate the variation of the pixel dimension and geometry for different lines of sight, but we use the same resolution of 15 km x 15 km over the whole field of regard.”

5) P6452, L20: Were you not able to reprocess the algorithm?

We couldn't reprocess the missing data (see also Reviewer #2, SC11)

6) P6453: L13: The authors should have made an evaluation of the model against independent measurements in order to show that the MOCAGE run is realistic. The authors can alternatively do an overview of the various studies which evaluate the MOCAGE model against independent measurements.

We have added the sentence “ MOCAGE trace gas outputs have been validated in the past with four-dimensional observations, e.g., [Dufour et al, 2004] and during measurement campaigns, e.g., [Bousserez et al., 2007].”

7) P6453, L14: What is the resampling in this case: interpolation, means, weighted mean (e.g. partial column calculation)?

The profiles are interpolated at the RTM input grid. We explicitly indicate this in the revised manuscript.

8) P6454, L14-17: This point needs to be more discussed. What is the impact of clouds on the evaluation the authors are doing in the next sections of the paper? For example, what is the level of significance of the tests made in table 4 5 6 and 7 if you flag the time series with clouds, as it should be done in reality?

We tried to cloud-screen a single image (19 August, 10:00 am) with a realistic cloud mask derived from the MOCAGE run, and we have found that all parameters, e.g., of the mentioned tables, have a maximum 10% difference with respect to non-screened data.

9) P6454, L18: Before beginning this section it should be useful if the authors specify the vertical resolution and the number of levels of the simulated data. A figure might be appropriated. As mentioned in the access review a plot of the full averaging kernel matrix/function (one typical example) would improve the quality of the paper.

We rather mention the KOPRAfit output resolution above (P6453, L20), see also the Reviewer #2's MP2. We included an additional figure of the AKs and the inherent discussion at the beginning of Section 3.

10) P6455, L1-2: What do you call surface- 3km TOC and surface-6km TOC. Supposing you have one level by kilometer of altitude is that the 3 or 6 first levels partial column? This point can be clarified by answering the one above.

As it should now be clear from the SC9, what we call surface-3km and -6km TOCs are the

columns calculated up to 3 and 6 km altitude, starting from the concentration profiles at 1 km resolution.

11) From P6456 L27 to P6457 L6: Please clarify this part of the text. Especially explain the last sentence: "The AKs for strongly positive thermal contrast (> 5 K) are only a bit more separated." Is that shown in any figures or has it been demonstrated before?

No, it is not shown in any figure in the manuscript (in the revised manuscript, we mention this in the text) but is a consideration we did after reviewing a similar figure as Figure 2, but for pixels with thermal contrast >5K.

12) P6458, L11-12: Please provide percentages.

Done

13) P6458,L21: We focus on the local scale

Change done

14) P6459, L11: Please specify/clarify which kind of comparison you are doing. Is it the MOCAGE pseudo reality smoothed by the averaging kernel or the MOCAGE raw pseudo reality?

It is the MOCAGE raw pseudo-reality. We specify this in the revised manuscript.

15) From P6459 L29 to P6460 L1: Please explain why?

In the revised manuscript, we have added one sentence to better explain. This part is now: "While the maximum of sensitivity is at about 12:00 UTC, the maximum of the pseudo-reality ozone concentrations at the lowest altitudes is at about 15:00 UTC. Consequently, the highest values at the lowest altitudes after 12:00 UTC are generally underestimated. This generates a quite strong artifact cycle at Milan, with a maximum which is shifted back of some hours."

16) P6460, L6: "mean of the mean biases" correct by "the mean biases"

Done

17) P6460, L26: A conclusive paragraph here is missing; please do a little summary about the capability of the MAGEAQ-TIR pseudo observations to follow the ozone partial columns temporal evolutions.

We added a short summary: "To summarize, we have found that our MAGEAQ-TIR pseudo-observation are capable to follow the pseudo-reality lower and lowermost tropospheric ozone columns temporal evolutions at local scale (at selected locations), even if artifacts in the columns time series, especially for the surface-3 km TOC, at Southern European urban locations are observed."

18) P6461, L12: Like in comment "P6459 L11" what kind of comparison are you doing?: Is it the MOCAGE pseudo reality smoothed by the averaging kernel or the MOCAGE raw pseudo reality?

We do comparisons with the raw pseudo-reality, and we integrated this information in the revised manuscript

19) Figure 3 and 4: the maps are a way too small. There is a lot of white wasted space between the maps. Please reduce this space in order to make the maps larger, or split the figures. Adapt the color scale of the differences to make it more accurate.

The spaces between the maps have been reduced and the label fonts have been augmented.

20) Figure 5 and 6: as for figure 3 and 4 reduce the space between the plots or split the figures.

The space between the plots is now smaller and the labels are bigger.

21) Figure 7: use more different colors for Surf-3km and Surf-6km DOF and reduce the wasted space between the plots.

Done

22)Figure 8: use error bars or a box plot instead plotting all the data points by levels. Reduce the wasted space between the plots.

We now use error bars; the wasted space is reduced

Anonymous Referee #2

The paper presents a sensitivity study for the observations of the MAGEAQ-TIR observing system. It is a Thermal InfraRed (TIR) spectrometer proposed as part of a project for Monitoring the Atmosphere from Geostationary orbit for European Air Quality (MAGEAQ). The subject of the paper is very interesting and appropriate for the AMT journal.

We thank the referee for the kind words.

Major points (MP)

The presented results, however, would be much more convincing and useful if complemented and modified to address the following major points:

1) Equations. No equations are reported in the paper. It is difficult to establish whether the authors have used the correct formalism to characterize the information contained in the MAGEAQ-TIR measurements. In particular, the description of the formulas used for the inversion and for the diagnostics is demanded to the papers describing the KOPRAfit algorithm. This algorithm usually retrieves vertical Volume Mixing Ratio (VMR) profiles on a 1km-fine vertical grid. It seems that the authors use the sum of the diagonal elements of the ozone VMR averaging kernels from 0 to 6

km (or from 0 to 3 km) as a proxy for the information contained in their synthetic measurements regarding ozone columns. Columns, however, are also connected with pressure and temperature distributions. It would be appropriate to show how this dependence is accounted for in the presented results.

For this study, we use the algorithm described by Eremenko et al (2008) and used for IASI, and we explicitly refer to that paper for a complete description of the inversion scheme, including formulae. In any case, we have added a few formulae to facilitate the reader, in case referring to the original paper is unpractical.

2) DOFs. On average the number of DOFs in the retrieved profiles from 0 to 6 km is around 1. This number seems very small to justify the retrieval of a "profile" with (as far as I understand) 1km vertical grid. It would be much more physically sound to retrieve directly the ozone column in the 0-3 or 0-6 km range, this approach would allow to limit the contribution of the a-priori information. Presently, with such a large contribution of the a-priori in the results, the discussion on biases seems quite speculative: the bias depends on the accuracy of the model used for the estimation of the a-priori profile, not on the actual measurements.

We are aware that the 1 km resolution of the KOPRAfit outputs is much finer than the actual vertical sensitivity of the instrument. In any case, as said for MP1, for this study we use the algorithm described by Eremenko et al (2008), and the design of a dedicated new algorithm is outside the scopes of the present paper. We added the following clarification on this aspect: "Our inversion scheme is based on an existing altitude-dependent Tikhonov-Phillips regularization method, which, as well, uses the KOPRA RTM. This algorithm, which has been developed to invert IASI radiance spectra measurements, is thoroughly described by Eremenko et al. (2007)". For this reason, at the present stage we start from the assumption that the retrieval is complemented by the AKs, which provide the information on the vertical sensitivity, and we added the following sentence in the discussion of the pseudo-observations simulator: "It must be kept in mind that the 1 km vertical resolution of the output is finer than the actual vertical resolution of the simulated instrument, as it will be shown in section 3. The AKs are then a necessary tool to interpret the retrieved profiles, in terms of their vertical sensitivity."

3) Contribution of a-priori. The (I expect "very large", but how much?) contribution of the a-priori information on the presented results could be adequately quantified with the so called "information gain" introduced in Rodgers (2000).

We find a bit unpractical to use of the "information gain" in our case. Indeed, it is mostly used in cases, like, e.g., wavelength selection during the design of inversion algorithms, where variation of information gain are analyzed as a function of some other quantities (in that case, number of bands or individual wavelengths). Instead, we decided to report on an error budget estimation, to allow the identification of the relative importance of the smoothing error, which is more linked to the a-priori contribution, and the measurement noise error. For a typical simulation (observation over land, + 1 K thermal contrast, 10:00 UTC, AKs for the lowest 12 km shown in Fig. 1 of the revised manuscript), we have found total errors of 1.21 DU (5.03%) and 1.42 DU (10.15%), for the surface-6 km and surface-3 km ozone columns. The smoothing error and measurement noise error contributions are, respectively, about 4.5% - 2.0%, for the surface-6 km ozone column, and about 9.5% - 4.0%, for the surface-3 km ozone column. In addition, please note that we use only two different a-priori profiles (depending on the tropopause altitude, see Sect. 2), so the structures observed in the pseudo-observations (cfr SC4) cannot derive from the a-priori.

Specific comments (SC):

We accept most of the SCs of the Reviewer #2 and we changed the text and figures accordingly. We report here on some points that need some detailed explanation.

1) Page 6453, lines 10, 11: stopping the model at 35 km may seem a quite rough approximation. Can you include a statement supporting this choice ?

We have used an existing run of MOCAGE. A typical MOCAGE output has altitude range between surface and about 35 km. We have completed the MOCAGE profiles with fixed representative trace gases profiles in the interval between about 35 km and 65 km, and we reported this aspect in the revised manuscript.

2) Page 6454, first paragraph of Sect. 3: as mentioned above, including some formulas here would help to understand if the used diagnostics is appropriate.

We don't feel that some formulas are required at the beginning of Sect. 3, as the (simple) derivation of the diagnostics is already indicated in the text: “The DOF for a partial column are calculated as the trace of the AK matrix, up to the top height of the column. The altitude of the maximum sensitivity of a partial column observation can be estimated by calculating the altitude of the maximum of the integrated AK for that partial column.” We think that explicit formulae would not add any information in this context.

3) Pages 6457-6458, Sect.4: see the general comment above. Can you demonstrate that the accuracy you are showing originates from the features of the measurements and does not depend on the a-priori information used to constrain the inversion ? Did-you try a test run with different a-priori and/or initial guess for the retrieval ?

Please refer to MP3.

4) Page 6459, line 22 and ff: note that the daily cycle in the DOFs could be avoided using a self-adapting constraint in the retrieval. See e.g. Steck (2002) who shows how to set up a Tikhonov constraint with the requirement of keeping constant the trace of the averaging kernel.

We plan to test self-adapting constraint schemes in the next future, see preliminary work like: Eremenko et al., Tropospheric Ozone Measurements with IASI/MetOp-A: Improvement of the Retrieval for the Lower Troposphere and Validation, Atmospheric Composition and Validation Evolution workshop, ESA-ESRIN, Frascati (Rome, Italy), 13-15 March 2013

5) Page 6460: see again the general comment above, regarding the use of a-priori information. What happens if a different a-priori profile is used for the Tikhonov constraint ? Will the observed biases change ?

We have tried different a-priori profiles and, indeed, the magnitude of the biases are similar as well as the station-to-station proportions, i.e., higher negative biases at Milan (of the order of -4 to -5%) and, to a lesser extent, at the marine station over the North Sea.

6) Page 6467, lines 21, 26: a synergistic use of both TIR and VIS observations is a good idea. For the future developments I would suggest also to adapt the inversion

code to retrieve only a few (1 or 2, max 3) ozone columns in pre-defined layers and avoid, as much as possible, the use of a-priori constraints. If these cannot be avoided, I would suggest to use at least Tikhonov with adaptive strength.

Thanks for the suggestion, and please refer to SC5 for more details on our ongoing work on Tikhonov algorithms with adaptive strength.

7) Page 6452, lines 20, 21: data missing from a synthetic set ? Why the calculations lost in the failure of the data processing system were not repeated ? Please explain or don't even mention the problem...

We didn't have the possibility to reprocess the data.

8) Page 6455, line 25: 0.61 or 0.71 as in Table 2 ?

It is indeed 0.61. 0.61 is the mean DOF for daytime surface-3 km TOC, i.e., theaverage of all pixels: "land"+"sea" (not reported in Tab. 2), while 0.71 is only for "land" pixels.

9) Labels and legendae are really tiny in Fig.s 3, 4, 5, 6, 7 and 9.

The fonts of the labels and legendae have been enlarged in most of the figures