The authors would like to thank the reviewers for all of their careful, constructive and insightful comments in relation to this work. We outline below our responses to these comments, and where appropriate indicate what has been changed in the manuscript as a result. For ease of viewing, the reviewers' comments are shown in bold. We believe that these comments help greatly in making our manuscript clearer and more useful to the reader.

Reviewer 1

We thank the reviewer for their constructive comments and suggestions and we discuss these in sequence below.

Abstract: It should be mentioned in the abstract that the paper is about nadir-measurements. Aircraft measurements are performed in four possible geometries/modes (nadir, limb emission, upward emission, upward absorption) and it is confusing for the reader not to know from the beginning which geometry/mode the paper is about.

We agree that this is an important issue that needs to be made clearer in the text. P10834, l8 has been changed to read "…representative of the ARIES system operating in the nadir viewing geometry."

Intro: The same applies to the body of the text: It is mentioned that Ts and εs are a prerequisite for the retrieval (p10837 l18), but this is only true for nadir retrievals. It has, however, not been stated yet that the paper is about nadir retrievals.

Thank you for pointing this out. In order to make it clearer from the outset that the paper is about nadir retrievals p10837, l14 has been changed to read, "…integrated ARIES measurement system, operating in the nadir viewing geometry."

p10837 l10: The abbreviation OEM is used here but it is defined only on page 10839.

Thank you for pointing this out. The abbreviation has now been defined on p10837, l10 (the definition has now been removed from p10839, l19 and replaced by 'OEM')

p10837 l24: I think (but I might have missed it) that the ARIES acronym is defined only in the abstract but not in the body of the text. This is not

sufficient; both the abstract and the body of the text must be able to stand alone. Thus the ARIES acronym needs to be defined when first used.

As well as being defined in the abstract (p10834, l1), the ARIES acronym was also defined on p10837, l11 in the main body of the text.

p10838 l7-9 (a very minor issue): when reading the abstract I was confused about the name ARIES because I wondered what the term 'Evaluation System' means. This becomes implicitly clear on p10838 l7-9 which I understand shall tell me that ARIES is an evaluation system for IASI. This rationale of the naming could be made a little more explicit.

That is correct. The ARIES instrument was originally designed as an IASI simulator to evaluate models and retrievals for the IASI project. In order to make this more explicit in the manuscript, the text on p10838, l7-9 has been changed to, "ARIES was originally designed as an airborne simulator for the IASI satellite instrument (hence the use of the phrase 'evaluation system') and, as such, has been used for calibration-validation activities for IASI…"

p10838 l18: The term 'scan' is ambiguous. It can be an interferometer sweep, it can be a geometrical scan over the swath, etc. Please be more specific here.

Thank you for pointing this out, by 'scan' we meant 'interferometer sweep'. The text on p10838, l18 has now been changed to read "…two complete interferometer sweeps per second…"

p10838 l22: Here the information is given that only nadir spectra are considered in this study, but this information is needed earlier.

Agreed. As in the response to above comments, this information now appears in the abstract (P10834, l8), and also earlier than p10838, l22 in the main body of the text (p10837, l14)

p10839 l11: 'for the first time' is unnecessary and should be deleted.

Thank you pointing this out, we have removed "from the first time" from p10839, l11.

p10840 l22: The threshold value of 85% does not tell me much if I do not know how the metric is constructed, thus it is pseudo-quantitative. Either remove the value or make it better traceable how this value is inferred.

This is a good point. We agree that this metric is pseudo-quantitative and potentially confusing to the reader. As such it has now been removed, and p10840, l20 now reads, "Key metrics such as hot and cold black body calibration statistics, phase correction, and other housekeeping parameters are used to ensure well-calibrated radiance spectra for retrievals."

p10842 l7: I am not sure if the term 'modulated' is ideal here. Modulation, I understand, modifies something which already exists, while the emitted radiance directly depends on the surface temperature and emissivity. I suggest 'which depends on Ts and εs'.

Thank you for pointing this out, we agree that the term modulation is used in the wrong context, and p10842, l7 has now been changed to, "…infrared radiation, which depends on T_s and ε_s...."

p10842 l8: The radiance is not attenuated at discrete wavelengths. Due to Doppler and pressure broadening, there is no attenuation at discrete wavelengths. Do the authors mean, that the numerical simulation is performed at discrete wavelengths?

We agree that the phrasing of this statement was potentially misleading, and that the atmosphere does not behave in this manner, rather it is the numerical simulations that are performed at discrete wavelengths. The text on p10842, l8 has now been changed to read, '…and is attenuated throughout the atmosphere by a large variety of atmospheric constituents.'

p10842 l8: There is a third term missing: The emission of the atmosphere. Since the temperature of the atmosphere and surface are similar, both contributions are important (contrary to, e.g., solar absorption), particular in the case of a temperature inversion. I know that the RFM is designed to consider atmospheric emission; please make sure that this feature has not been inappropriately deactivated, and change the text accordingly. I suspect that the calculations with the RFM are ok and only the text is incomplete but this should be checked.

This is correct. As the reviewer notes, this is an omission in the text and not in the use of the RFM, which in this study has been used with the inclusion of the emission by the atmosphere. We also note that this section does not also mention the back-scattered solar component of the radiation, which must also be

(and has been) included. The text on line p1043, l7 has been changed to reflect this, and now reads, “The radiation measured by ARIES is the sum of the upward infrared radiation from the Earth (which depends on T_s and ε_s, and is attenuated throughout the atmosphere by a large variety of atmospheric constituents), the emission by the atmosphere, and also a back-scattered solar component.”

p10842 l26: I think the term ‘theoretical’ would be more appropriate than ‘technical’. The latter, I understand, includes implementation issues etc.

We agree, and have changed the text on p10842, l26 from ‘technical’ to ‘theoretical’.

p10842 l26: It is funny to see Rodgers 2000 as a reference for OEM. While this method is thoroughly discussed in that book, it is called ‘maximum a posteriori’ there. The term ‘optimal estimation’ was used in Rodgers’ older publications only. Personally I have no problems with this but a student who is new in this business might be confused by the inconsistent naming. Perhaps it helps to add somewhere ‘OEM, later renamed maximum a posteriori’.

Thank you for pointing this out, it is a very important point that could lead to confusion for the reader. So as to correct for this we have now replaced all mention of the term ‘OEM’ with ‘maximum a posteriori’, because as the reviewer points out this is more in keeping with the correct terminology.

p10842 l28: ‘statistical knowledge’ is too vague. Please add ‘on the variability of the true state around xa’.

We agree that this statement was too vague. The text on p10842, l28 has now been changed to, “statistical knowledge on the variability of the true state around $\mathbf{x}_a$.”

p10843 l7: It is not generally true that the chi square is at a minimum equal to the number of measurements m. This is true only in a statistical sense, i.e. the expectation of chi square over a large number of retrievals equals m.

Thank you for pointing out this error. The text has now been changed to, “The expectation of χ^2 over a large number of retrievals is equal to the number of measurements m, the total number of degrees of freedom.”

p10844 appr l17-22: I could not fully understand this. This should be reworded for clarity.

We agree that this was perhaps poorly worded. There was also a typo with the text '2' on p10844, l19, which was supposed to have read 'NESR'. This typo has been corrected, and the text on p10844, l17-24 has now been reworded to the following, "The instrument noise spectral covariance matrix $\mathbf{S}_y$ was constructed as the sum of the square of two other covariance matrices: a diagonal matrix constructed from the NESR, where the errors in any singular channel were assumed to be independent of other channels; and a non-diagonal matrix computed from consecutive calibration differences. The NESR diagonal matrix was first divided by the square root of the number of co-added measured spectra (five in this case, as described in Section 3.1)."

p10850 l6-7: The first sentence of this paragraph does not help. There is no need to be so defensive. Sensitivity studies are a kind of research in their own right. Just delete this sentence and concentrate on what you have done; don't start the paragraph with a statement on what you have not done.

We agree, and we thank the reviewer for their support on this matter. This sentence has now been removed, and p10850, l6 begins with, "In the remainder of this study, we assess…"

p10854 l1: The formulation 1 − (Sa/S) is sloppy because it contains undefined matrix operations. I can see what you mean but a more careful formulation is needed, e.g. 'the diagonal terms of I − SaS−1 where I is unity'.

This is correct and it is a very important point. P10854, l1 has now been changed to, "Note that error reduction is defined here as the mean value of the diagonal terms of $\boldsymbol{I} - (\boldsymbol{S}_a/\boldsymbol{S})$, where $\boldsymbol{I}$ is an identity matrix."

Reviewer 2

We thank the reviewer for their constructive comments and suggestions and we discuss these in sequence below.

1. The first critical point is related to the organization of the whole study. I think that it would have been better if the work was not split in two parts or if the present paper showed at least a sample of preliminary results with

real observations. In fact, neither the instrumental set up here described nor the methodology are novel or original.

The decision to split the overall study into two papers, one that deals with the description of the retrieval scheme and theoretically optimal performance with simulated data, and one that presents retrieval validation with real observations, was made for the following reasons. For this manuscript, it was necessary to describe the retrieval scheme in technical detail, and to illustrate its potential performance in a range of simulated scenarios, before then going on to validate it against real data sets. This manuscript is a highly technical paper, which sets the scene for the later interpretation of operational retrievals with full traceability to the principles of retrieval theory and data handling practices, while the other manuscript presents first results and data from several field campaigns, each requiring unique description. We believe that to combine the two would (and did) result in a very unwieldy manuscript with two very contrasting themes. We maintain that two manuscripts are preferable and appropriate for the reader and end-user of ARIES data products.

(a) Regarding the instrument, for sure the authors know that the American Scanning HIS and NAST-I FTS instruments have similar spectral characteristic and both these instruments have shown (operated on board of aircrafts) their capability to retrieve geophysical parameters with an accuracy comparable with the one here stated (the authors can for sure add here relevant references).

We agree. Thank you for pointing this out. The end of Section 2 now includes the following text:

“The retrievals from the ARIES spectra that are discussed in this study follow on from the work that has been done by both the Scanning High-resolution Interferometer Sounder (S-HIS) and NPOESS / NASA Airborne Sounder Testbed - Interferometer (NAST-I) instruments on board the NASA ER-2 aircraft (Tobin et al., 2006; Larar et al., 2011). A potential advantage that the ARIES instrument has over the S-HIS and the NAST-I is these instruments, when flown on board the high altitude ER-2 aircraft, usually operate at an altitude of 20 km on board a NASA ER-2 aircraft, whereas the ARIES instrument typically operates between 1 and 8 km, resulting in a much smaller GIFOV and potentially improved sensitivity to the lower atmosphere (see Section 5)…”

(b) Moreover, the MARS methodology, as the authors state in the section 4.1, is not original and some parts seem to me to be less than state-of-art (e.g. see point 2 below) and it is fully based on Rodger Optimal Estimation Methodology.

We apologise for what appears to be some confusion abut what is novel about the MARS methodology. It is not the maximum a posteriori methodology that is original, but rather the extensive pre-processing, a priori construction, choice of auxiliary data etc.; the Maximum a posteriori technique described by Rodgers is just one small aspect of the overall retrieval algorithm. We believe that this has now been made clearer in the revised manuscript.

(c) The need of real observations in this paper is increased by the basic error in figure 1, the unique "Measurement" shown in this paper. This figure is misleading since it is not representative for the observations used for the retrieval. Contrary to what is stated, observed spectrum in Figure 1 is evidently a down-welling spectrum (Atmospheric windows radiance is lower than the Carbon Dioxide, Ozone and Water Vapour bands radiance). Please change this Figure with an aircraft up-welling radiance ARIES measurement.

Figure 1 shows a nadir-measured infrared spectrum whilst the aircraft was flying at a constant altitude of 8 km. The spectrum is most definitely measured in a nadir view and represents the sum of the attenuated spectrum and emission that would be expected at this altitude. Partial atmospheric window regions in Figure 1 are (for example) between 1400-1700 cm^{-1}. We suspect that the reviewer has confused the emission lines of carbon dioxide that are naturally present outside of the atmospheric window regions of Figure 1 (e.g. at 1000 cm^{-1}) at this altitude, with the idea that to see such emission, we must be viewing in the zenith. This is not the case. We can confirm that this is a nadir viewed spectrum, which is consistent with that expected (and simulated).

2. The second point is related to the references incompleteness in the introduction section, where the authors neglects some capability of IASI instruments and they show a French bias in the description of the state-of-art. Into details (page 10835, L26 to page 10836, L5) describing IASI retrieval capability the authors mentioned only CH4 and O3, but no mention is provided for CO2, CO and N2O. In 2004, Lubrano et al Tellus B, showed the possibility to retrieve N2O from IASI radiance using IMG radiance as a proxi (this article is also cited in the IASI ATBD). The IASI capability to retrieve CO2 has been shown in Grieco et al. Appl Opt, 2011. In this paper the authors exploit the partial scansion of the interferograms in order to insulate the CO2 signature enabling the retrieval of CO2 columnar amount with accuracy of about 7 parts per million by volume at the level of each single field of view. Similar methodology has been applied to IASI data to retrieve CO,CO2, CH4 and N2O (Grieco et al. Optic Express 2013, - both the Grieco et al. articles are mentioned in the CNES web page devoted to IASI publications). In the Grieco et al. work the columnar amount of these gases

are retrieved with better accuracy than the one in the current paper. And this point should be discussed in the paper.

We should note that this paper is not about the IASI instrument, and we make no mention here of either CO_2 or N_2O retrievals from ARIES; therefore some of these suggested references are not within the scope of the paper. However, we now include the suggested Grieco et al. (2013) reference to highlight other IASI retrieval studies.

3. Third Point. The cloud detection methodology is based on Brightness Temperature difference (8 – 11 μm). If this difference is larger than 1 K, the spectrum is classified as cloudy. This test is largely correct for sea surface, however for other surface features it could fails (e.g. think about desert sand, Masiello et al. Appl Opt 2004). It is not clear here if the authors propose their method for sea surface alone. Again this point should be clarified and discussed.

A full discussion and validation of the cloud-filtering algorithm will be presented in the second part of this study. The reason why such discussion is given in that paper is because of recommendations by the editor and referees at an earlier stage of review. This brightness difference method and threshold has been validated over both land, sea and ice scenes against lidar cloud detection by the aircraft, which will be discussed in Part 2.

4. Minor Point. Table 1-6. Please add absolute units for retrieved geophysical parameters.

The geophysical parameters in this table are given as percentage values for total columns and weighted mean profiles. The inclusion of absolute values/units for these biases would lead to a table that could be unclear and difficult to follow and we feel that such information is best presented by the figures and in the body of the text - absolute biases and differences for discrete vertical levels are readily apparent from Fig. 4, 7, 10, 13, 16.

Reviewer 3

We thank the reviewer for their constructive comments and suggestions and we discuss these in sequence below.

Some change in the structure needs to be made. It would be easier to read if putting all the description of Sa, Sy, xa etc. in the section of retrieval method.

All of the variables that are listed by the reviewer are indeed already defined in one section, Section 4.1: Solving the inverse problem. We believe that this section does describe each of these terms in one section, appropriately placed within the current manuscript.

The references in the introduction Section is incomplete and up to current. CO2, CH4, CO and O3 are standard products in Atmospheric Infrared Sounder, and similar products from IASI are generated in Eumetsat and NOAA. These works should be discussed and cited.

We have now modified the manuscript to include a more recent reference for IASI retrievals; in line also with one of the comments from Reviewer 2 (see above), the following reference has now been included Grieco et al. (2013). We have also included mention of the NOAA Atmospheric Infrared Sounder (AIRS) instrument, and more recent references for the GOSAT and TES satellites. Page 10835, from line 26 onwards now reads:

"Since the launch of MIPAS, a series of nadir-viewing Fourier Transform InfraRed (FTIR) satellite spectrometers measuring atmospheric composition now exists. These include greenhouse and trace gas retrievals by the Tropospheric Emission Spectrometer (TES) instrument aboard the EOS-Aura satellite (see e.g. Kulawik et al., 2006; Verstraeten et al., 2013), the Atmospheric Infrared Sounder (AIRS) instrument on board the EOS-Aqua satellite (see e.g. McMillan et al., 2005; Masiello and Serio, 2013), and the Infrared Atmospheric Sounding Interferometer (IASI) instrument on board the Metop-A and Metop-B satellites (see e.g. Turquety et al., 2004; Razavi et al., 2009; Grieco et al., 2013). Recently the Greenhouse Gases Observing Satellite (GOSAT) has also been used to obtain column averaged dry air mole fractions of CO_2 and CH_4 (see e.g. Yokota et al., 2009; Ross et al., 2013). Such measurement capability has dramatically increased our ability to assess the impact of human activities on the changing composition of our atmosphere and resulting climate change (see e.g. Eremenko et al., 2008; Crevoisier et al., 2009; Worden et al., 2013)."

P10843, L7-8: something is wrong and need to rewrite this sentence.

Thank you for pointing this out. The text has now been changed to: "'The expectation of χ^2 over a large number of retrievals is equal to the number of measurements m, the total number of degrees of freedom."

P10844. L8-10: The retrieval method is a straightforward application of Rodgers' Optimal. How to choose the damping Lambda is a very important part, so it would be helpful to give more detail.

Agreed. Thank you for pointing this out. The text has now been changed to "The MARS employs a Levenberg-Marquardt iterative technique, which makes use of a damping factor λ, chosen to minimise the cost function at each step of the iteration. An initial damping factor of 0.1 is selected, then after every iteration the cost function is calculated, and compared to the cost function of the previous iteration. If there has been an increase in the cost function then the damping factor is increased by a factor of 8, and if there has been a reduction in the cost function, then the damping factor is reduced by a factor of 4."

P10846, L1-8: It is not quite clear to me the steps for different gases retrievals, including the sequence of the retrieval, and how the variable of temperature, H2O, Ts are retrieved together with different gases (which are in different steps). This paragraph needs to revise.

We apologise for the potential confusion here. The retrieval does not take place in different steps; rather the temperature, water vapour, aerosol, and surface temperature are all part of the state vector, along with the trace gas retrieval product (e.g. CO), i.e. these are all retrieved simultaneously. We have rewritten p10845, l20 onwards, using the following text, to make this clearer:

"The auxiliary gases are defined as those gases that are not retrieved, but that exhibit significant spectral features in the chosen spectral window. Due to differing retrieval micro-windows this list of auxiliary gases was necessarily different for each of the retrieval parameters However, in the construction of the state vector, $\boldsymbol{x}$, it was decided that as well as the target gas, the retrieval of H_2O, temperature, and aerosol extinction profiles, as well as T_s, were also always operationally required. However, it should be clarified that for each of the target gases (and temperature) to be retrieved, a separate state vector is constructed, e.g. CO is not retrieved simultaneously with O_3."

P10851, L12: To test a retrieval system, the simulations should be made in assemble of profiles, then we can evaluate whether the algorithm works in different atmospheric conditions. Here two profiles were chosen. Maybe I missed something, and I do not quite clear whether this algorithm was

tested in enough simulated cases.

The algorithm was tested for two contrasting simulated environments/compositions that are representative of some of the wide-ranging atmospheric conditions under which the ARIES instrument will be operational. By considering European background and biomass burning scenarios for two different altitudes in each environment, and then considering the retrieval sensitivity to situations where the H_2O and temperature may not necessarily be well known, we believe that we have tested this algorithm sufficiently to assess the simulated capability of the ARIES airborne FTS system. The second part of this paper will present further validation of the retrieval scheme, using real atmospheric conditions and measurements.

P10852, L14-22: it is better to move it to Section 4.

The reason that the discussion of the construction of the a priori using the MACC-modelled data is given here is because it is particular to the simulations that are performed in this study. These values are not fixed, and are determined on a case-by-case basis; therefore it would be inappropriate to include this in Section 4, which is for the general retrieval methodology.

References cited in this response

Crevoisier, C., Nobileau, D., Fiore, A., Armante, R., Chédin, A., and Scott, N.: Tropospheric methane in the tropics–first year from IASI hyperspectral infrared observations, Atmospheric Chemistry and Physics, 9, 6337-6350, 2009.
Eremenko, M., Dufour, G., Foret, G., Keim, C., Orphal, J., Beekmann, M., Bergametti, G., and Flaud, J. M.: Tropospheric ozone distributions over Europe during the heat wave in July 2007 observed from infrared nadir spectra recorded by IASI, Geophysical Research Letters, 35, 2008.
Grieco, G., Masiello, G., Matricardi, M., and Serio, C.: Partially scanned interferogram methodology applied to IASI for the retrieval of CO, CO2, CH4 and N2O, Optics Express, 21, 24753-24769, 2013.
Kulawik, S. S., Worden, J., Eldering, A., Bowman, K., Gunson, M., Osterman, G. B., Zhang, L., Clough, S. A., Shephard, M. W., and Beer, R.: Implementation of cloud retrievals for Tropospheric Emission Spectrometer (TES) atmospheric retrievals: part 1. Description and characterization of errors on trace gas retrievals, Journal of Geophysical Research: Atmospheres (1984–2012), 111, 2006.
Larar, A. M., Smith, W. L., Zhou, D. K., Liu, X., Noe, A., Flood, D. O. M., Rochette, L., and Tian, J.: An update on the NAST-I airborne FTS, 2011,
Masiello, G., and Serio, C.: Simultaneous physical retrieval of surface emissivity spectrum and atmospheric parameters from infrared atmospheric sounder interferometer spectral radiances, Applied Optics, 52, 2428-2446, 2013.

McMillan, W. W., Barnet, C., Strow, L., Chahine, M. T., McCourt, M. L., Warner, J. X., Novelli, P. C., Korontzi, S., Maddy, E. S., and Datta, S.: Daily global maps of carbon monoxide from NASA's Atmospheric Infrared Sounder, Geophysical Research Letters, 32, 1-4, 2005.

Razavi, A., Clerbaux, C., Wespes, C., Clarisse, L., Hurtmans, D., Payan, S., Camy-Peyret, C., Coheur, P., and others: Characterization of methane retrievals from the IASI space-borne sounder, Atmos. Chem. Phys, 9, 7889-7899, 2009.

Ross, A. N., Wooster, M. J., Boesch, H., and Parker, R.: First satellite measurements of carbon dioxide and methane emission ratios in wildfire plumes, Geophysical Research Letters, 40, 4098-4102, 2013.

Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Best, F. A., Smith, W. L., Ciganovich, N. N., Dedecker, R. G., Dutcher, S., Ellington, S. D., and Garcia, R. K.: Radiometric and spectral validation of Atmospheric Infrared Sounder observations with the aircraft-based Scanning High-Resolution Interferometer Sounder, Journal of geophysical research, 111, D09S02, 2006.

Turquety, S., Hadji-Lazaro, J., Clerbaux, C., Hauglustaine, D., Clough, S., Cassé, V., Schlüssel, P., and Mégie, G.: Operational trace gas retrieval algorithm for the Infrared Atmospheric Sounding Interferometer, Journal of Geophysical Research: Atmospheres (1984–2012), 109, 2004.

Verstraeten, W. W., Boersma, K. F., Z\"orner, J., Allaart, M. A. F., Bowman, K. W., and Worden, J. R.: Validation of six years of TES tropospheric ozone retrievals with ozonesonde measurements: implications for spatial patterns and temporal stability in the TES bias, Atmospheric Measurement Techniques Discussions, 6, 1239-1267, 2013.

Worden, J., Wecht, K., Frankenberg, C., Alvarado, M., Bowman, K., Kort, E., Kulawik, S., Lee, M., Payne, V., and Worden, H.: CH4 and CO distributions over tropical fires during October 2006 as observed by the Aura TES satellite instrument and modeled by GEOS-Chem, Atmospheric Chemistry and Physics, 13, 3679-3692, 2013.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global concentrations of CO2 and CH4 retrieved from GOSAT: First preliminary results, Sola, 5, 160-163, 2009.