Response to Reviewer 2.

We thank Reviewer 2 for his/her comments which helped to improved, we hope, the quality of the manuscript. Reviewer 2's comments are in bold font, our answers are written with normal font.

General comments:

The English is generally very good, the paper is well structured and nicely concise. I am proposing major revisions because some of the plots should be remade, and the scope of the paper does not seem to fit the ambitions of the title. I found the choice of experiments a little strange, and was rather confused on the details. However, whilst I am suggesting some extensive changes, I think that these matters could be fixed quite easily by the authors.

To agree with the content of the paper, we suggest to change the title of the paper with « Homogeneity criteria from AVHRR information with IASI pixels in a Numerical Weather Prediction context ». In addition we have removed the data assimilation experiment with the only Obs-Hom criteria and only kept the experiment with the compromise criteria. Some parts of section 4 were rewritten as well as the conclusions.

I realised when I reached the final sections of this paper that I had totally misunderstood its intention. I had expected that the homogeneity criteria would be used to select additional observations, that are homogeneously cloudy, to assimilate in addition to the clear-sky channels accepted by the McNally and Watts check - i.e. to do something similar to McNally (2009). It took me almost the whole paper to realised that what is being proposed is additional quality control on radiances already accepted by McNally and Watts. This does not seem like "preparation for all-sky assimilation" and I think that the scope of the paper should be revised.

We recognised that the title of the paper may be confusing that is why we propose to change the title of the paper with « Homogeneity criteria from AVHRR information within IASI pixels in a Numerical Weather Prediction context » The objective of the paper in the last paragraph of the introduction was also modified into : « Our objective is to determine homogeneity criteria valid for both clear and cloudy conditions, suitable to an NWP context using collocated AVHRR and IASI information. »

At least, I assume that it is the case that this is just extra QC on clear sky calculations

I confess that I found the paper surprisingly confusing! There is no mention that RTTOV-CLD is being used in the assimilation experiments, therefore I assume the calculations are clear sky, and there is no mention of the use even of a single grey-cloud layer scheme in use as in McNally (2009).

We propose an ensemble of homogeneity criteria. In case of clear sky, the McNally and Watts is applied as an additional QC test afterwards. In case of cloudy scenes, these criteria could pave the way to an all-sky assimilation (methodology not decided yet). You are right in the assimilation experiments, only RTTOV was used (clear sky assimilation). RTTOVCLD was only used to compute the homogeneity criteria based on cloudy AHVRR simulations. In the operational version of ARPEGE , a single layer grey cloud scheme is used (Guidard et al 2011) but in the experiments carried out in this paper, this possibility has been switched off to focus on clear sky assimilation.

This has been specified in the text : « In these experiments, no cloudy observations detected with the CO2-slicing method and used with a single grey-cloud layer scheme was assimilated unlike in Guidard et al. (2011). RTTOVCLD was only used to compute the homogeneity criteria based on cloudy AHVRR simulations and RTTOV was used for the clear sky assimilation. »

Some curious choices are made throughout the paper: I believe that the E2014 method was adapted because Eresmaa's intention was to keep only clear scenes, whereas you wish to allow through homogeneously cloudy scenes also, but in fact because you are rejecting observations that had already been allowed through by McNally and Watts, I don't see why you don't just apply Eresmaa's method without modification. Why do you not include the scheme of McNally (2009)?

The objective of Eresmaa (2014) method is to keep clear pixels. Our objective is to keep homogeneous scenes both in clear and cloudy conditions. The Eresmaa method is not suitable in our case and this why we propose the modification based on this method. The single layer grey cloud scheme could have been used but the selection of cases to be assimilated this way woud need a dedicated study.

And finally, perhaps most surprisingly, you use the AVHRR clear/cloudy pixel fraction as a measure of whether the homogeneity criteria have "correctly" picked out homogeneous scenes, and on p15 you state that you are happy to accept a reasonable proportion of observations with >90% cloud cover. Why not just test the use of the AVHRR clear pixel fraction? And yet, you performed assimilation experiments with a scheme that you had seemed to reject based on the O-B statistics presented in Figure 4.
As suggested by reviewer 1 we added an evaluation with independent data of cloud type from SEVIRI on board MSG satellite in section 4 because the cloud cover from the AVHRR has some defects. We agree that observations with >90 % cloud cover are not necessarily homogeneous.
We have now removed the data assimilation experiments with obs_HOM criteria and we kept only the compromise method.

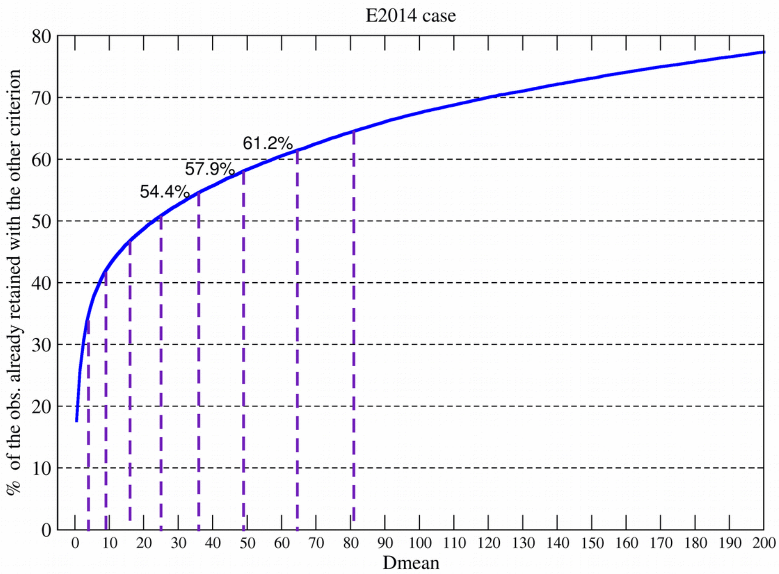
It is not really surprising that there is little impact, as very little seems to have changed in the experiments relative to the control. The work therefore seems rather immature for a publication. You still apply the CO2-slicing method (p16 line 17) - presumably this is designed to reject cloudy scenes? What effect does this have on the homogeneous scenes?
The sentence p16 l 17 was confusing and thus modified as said above : « In these experiments, no cloudy observations detected with the CO2-slicing method and used with a single grey-cloud layer scheme was assimilated unlike Guidard et al. (2011). RTTOVCLD was only used to compute the homogeneity criteria based on cloudy AHVRR simulations and RTTOV was used for the clear sky assimilation. »
3 % of observations are rejected between Exp and Ref, so you are correct that it is not very surprising that there is very little impact. Nevertheless, this impact is slightly positive. Thus it seems that this method is reliable enough to be used in a NWP assimilation. These criteria will reveal their full potential in an all-sky assimiulation. The design of a IR all-sky assimilation has still to be done.

Specifics:
P9 line 7: it is not clear whether this 7K check is an additional criterion over the 8% criterion in the previous paragraph. Why 7K? I also don't understand "interpolated using 12 points" - is that 12 points in 3D? This 7K check is an additional check verifying that both the observation and the model observe the same cloudy scene. In the original Martinet et al (2013) it was applied to the difference between the mean AVHRR brightness temperatures from the observed and simulated clusters. In the case or ARPEGE the horizontal mesh is coarser and we replaced this fine scale check with the difference between the brightness temperature from the guess profile and the observation. The guess profile results from a horizontal interpolation of 12 profiles surrounding the observation position coming from a 6-hour forecast.
We have added this point in the text : « In the original Martinet et al (2013) study, a third check verifying that both the observation and the model observe the same cloudy scene was done with the difference between the mean AVHRR brightness temperatures from the observed and simulated clusters less than 7 K. Here, the ARPEGE model has a coarser resolution and it is not possible to simulate the AVHRR clusters. This check was adapted with the difference between the AVHRR observation and the AVHRR simulation from the guess, which come from a horizontal interpolation of the 12 profiles surrounding the observation position coming from a 6-hour forecast. »

P10 line 14-17: I don't understand how this first bullet point is different from the original method. The difference lies in the fact that the simulation computation were done with RTTOV-CLD. This is mentioned now in the text. « All AVHRR simulations from background are made with RTTOV-CLD and the threshold of the background departure check was modified » The description of the threshold for inter-cluster homogeneity was thus removed.

P10 line 18-19: Why 49K2? Other than that it fits the 7K applied to M2013, it seems quite high relative to 1K.

Many trials were done for the value of D_{mean}. Below you will find the function of leaving observations as a fonction of D_{mean}. It appears that this value allows to keep more than 50 % of the observations. In addition it also fits the 7K threshold of M2013 but over the 2 IR AVHRR channels.



We used the D mean proposed by Eresmaa (2014) to perform here a kind of cloudiness consistency check between the observation and the model simulation if D mean is less than 49 K^2. This particular value of threshold allows to keep more than 50% of the observations compared to the initial threshold of $1K^2$ by Eresmaa (2014) which retains only 18% of the observations. In addition, this threshold compares well with the one applied by M2013, but it is applied over the 2 IR AVHRR channels. The text was modified accordingly.

P11 line 5: This is the same as the first test of Martinet but with 2 channels. It would be clearer if this was stated. Why change the L to R in the equation? What does the addition of the second channel bring in practical terms?

It was done : « It is the same test as in M2013 but in the brightness temperature space ». The change of L into R is practical as it is easier to work with brightness temperatures. We chose to have all value in brightness temperatures. If we only consider channel 1, we keep 68,2 % of the observations, with channel 2 69,6 % are remaining but with both 67,3 % of the observations are selected. This was mentioned in the text : « If this test is only applied over channel (10:5 m, 68,2% of the observations are selected, if applied over channel 11:5 m), 69,6% of the data are kept and if it is applied over both channels, 67,3% pass the test. »

P12: What is this dataset of 59 million observations? Is it just 24 hours' worth of observations? You state that 50% of the observations are 100% cloudy - that sounds potentially high for a normal dataset? There was a confusion between the number of observations and the number of channels. In fact the statistics were computed over 188090 IASI observations for 30 January 2017 and the number was corrected in the text and in the table. Half of the observations having a cloud cover of 100 % may seem a very high percentage, but with the independent validation we proposed with SEVIRI data, we find the same result.

P12/Figure 3: Are the numbers in the text for bias and SD an average over a number of channels, or the maximum value from the windows? It should probably be the latter. I cannot match the figures in the text with the plots - the numbers do not seem to match (e.g. 11.7K bias -> the bias looks over 12 K in the figure). It would also be better to just plot Band 1 so we can see the effect on the temperature channels. I think the numbers scattered over several paragraphs and two pages would be better in a table.

Figure 3 was changed and represents now only the spectral range of band 1. Section 4 was partly rewritten. This section was re-arranged and there are now less numbers in the text. Tables were also modified in order to show statistics on window channels and CO2 channels.

P14 line 15-20: I would disagree that the distribution asymmetry is small. I also disagree that the Obs_HOM approach reduces the range of the tropospheric water vapour channel distribution. Both sentences were removed and changed with. « The distribution asymmetry is reduced for mid and low tropospheric water vapour channels with M2013 and E2014 selection. » for the first one.

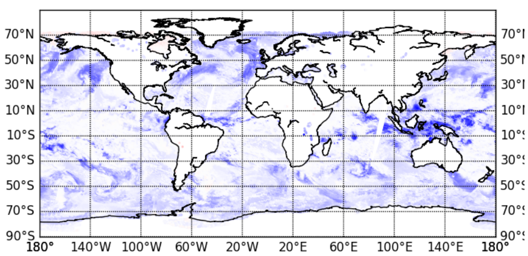
P15: The discussion focuses on letting through the most data - this isn't necessarily the best criterion, as you may be letting through inhomogeneous scenes. There is trade off between more data and better data. M2013 and Obs_HOM let through a lot of partially cloudy scenes (and even 100% cloudy scenes may have different cloud types in one pixel).
We agree that 100% cloudy scenes may have different cloud types in one pixel. Now the discussion is based on SEVIRI cloud which is an independent validation datum. We also agree that there is a trade-off between more data provided by M2013 and Obs_HOM and better data. That is why we propose a fourth method based on the Eresmaa (2014) test.

P16: It is not clear what the set-up for the assimilation experiment is - you do not mention RTTOVCLD - presumably this is still clear sky.
Now it is clearly stated : « RTTOVCLD was only used to compute the homogeneity criteria based on cloudy AHVRR simulations and RTTOV was used for the clear sky assimilation. »

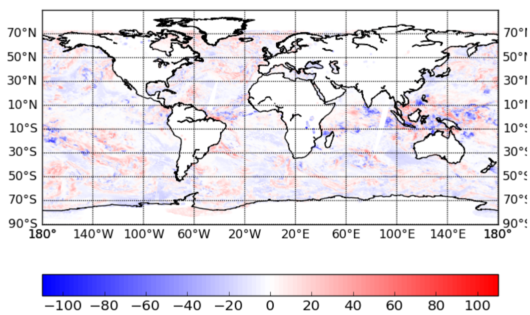
Figures:

- **Figure 1: would be better as two bigger O-B plots.** We have replaced the simulation plots with the suggested O-B plots. The comments on Figure 1 have been changed accordingly :

b) Observations minus RTTOV (clear-sky) simulation

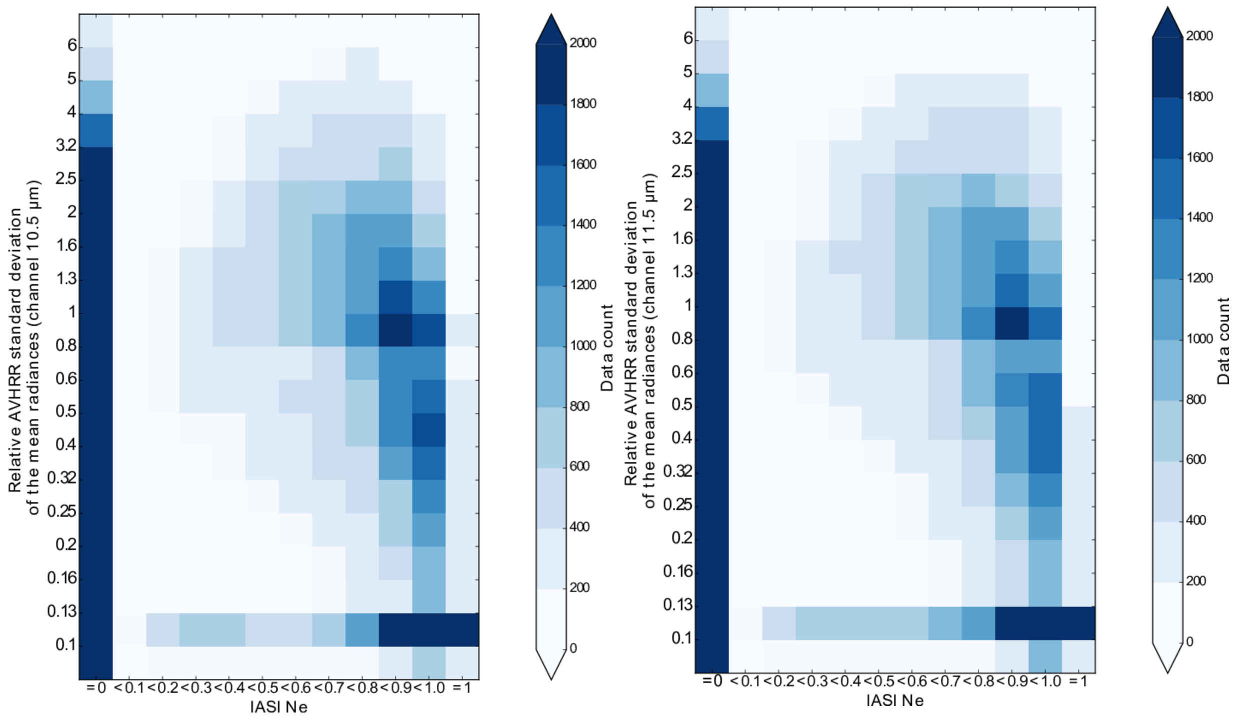


c) Observations minus RTTOV-CLD simulations



« To illustrate the benefit brought by RTTOV-CLD, Figure (1) shows IASI brightness temperature observations of a cloud-sensitive surface channel (1271, 962.5 cm −1) and differences between the observations and the simulations computed with RTTOV considering clear-sky and with RTTOV-CLD. Brightness temperatures less than 250 K are usually associated with higher elevation cloud structures. By using RTTOV in clear sky (figure 1.b) to simulate IASI observations, the observation departures are mainly below zero and may reach up to -60 K. This can be explained by the fact that the main cloud structures associated with low values of brightness temperature for the surface channel are missing in the simulation. On the contrary, differences obtained with the RTTOV-CLD simulations are in overall in better agreement with lower positive and negative values (figure 1.c). No major differences are found for example for the cloud structures located over the North Atlantic (30N-70N, 40W-0W) and above (30S-70S, 60W-0W) the Southern Atlantic Ocean. Large difference values are mainly obtained in the Tropics region. This may be explained by the fact that clouds are better simulated in the ARPEGE model for mid-latitudes than in the Tropics. »

- **Figure 2: very strange Y-axis. You can't see much on these plots. Is the Y-axis expressed as % as in the criterion on p11 line 11?** Yes the Y-axis is expressed as % as in the criterion line 10 page 11. The figure changed, with a log scale on the y axis, another color scale and as a function of data count.



- **Figure 4: Why not plot a temperature sounding channel? The x-axis has strange divisions. It would be better symmetrical.** Panels of Figure 4 were drawn again in order to have symetrical x axis as shown below. We do not consider temperature sounding channel as we are more interested in humidity and clouds.

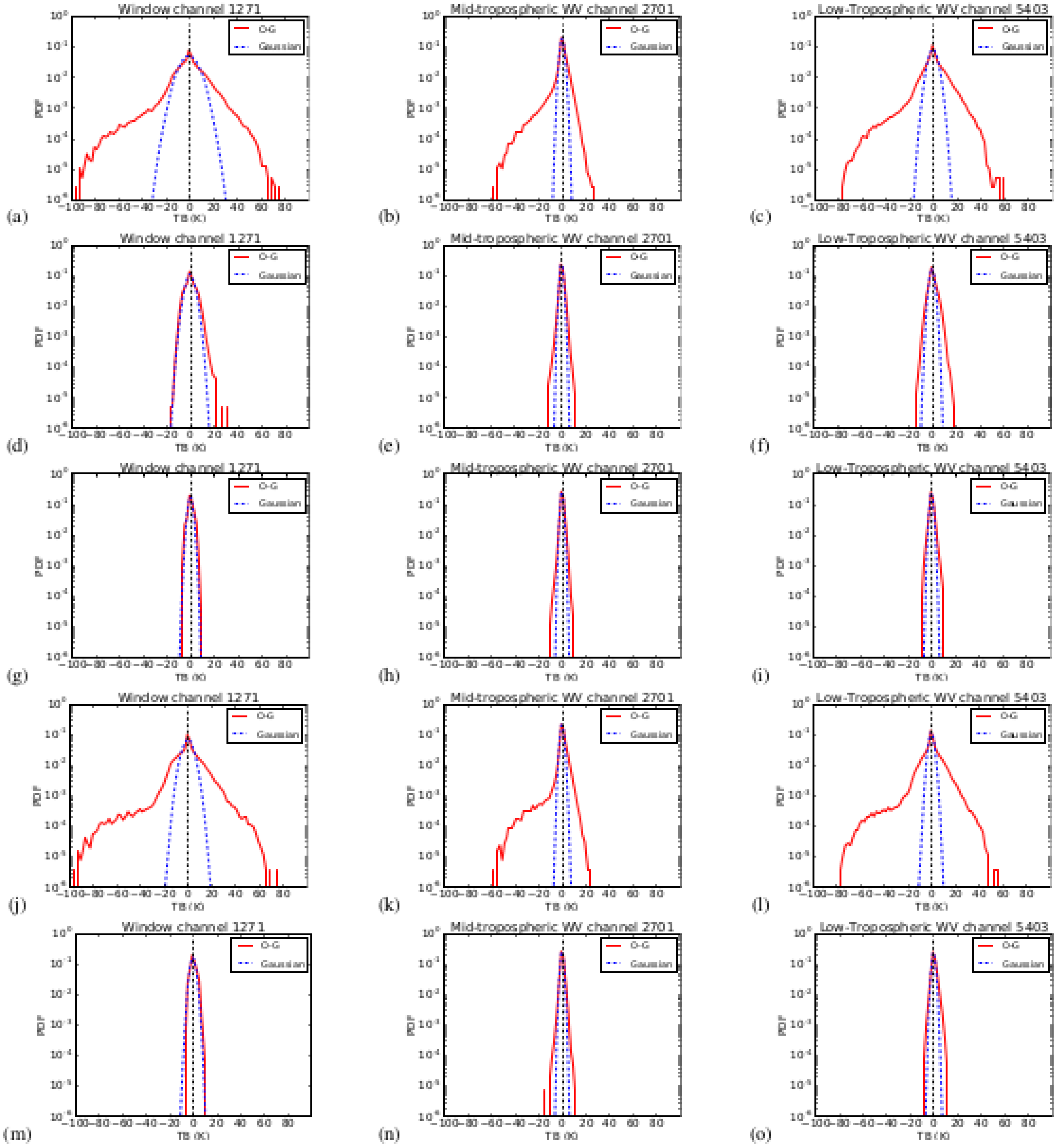
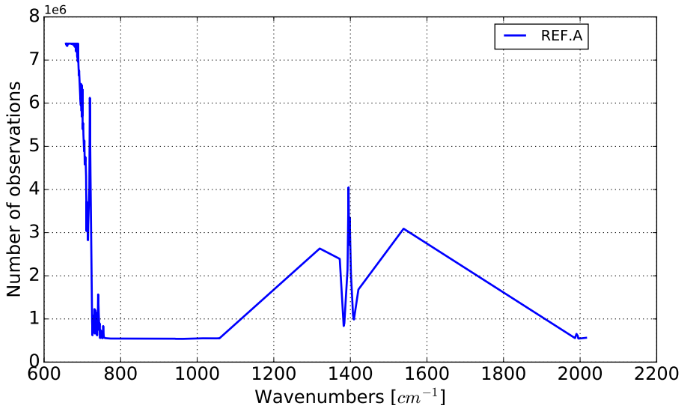


Figure 4. Frequency distribution of brightness temperature difference between observation and background (O-G) for all observations (a, b, c), after applying the homogeneity criteria derived from Martinet et al 2013 (d, e, f), the homogeneity criteria derived from Eresmaa 2014 (g, h, i,), the third method based on observation space method (j, k, l) and the compromised approach (m, n, o). The PDF are presented for three channels: window channel 1271, low-tropospheric water vapour channel 5403, and mid-tropospheric water vapour channel 2701). The Gaussian distributions with the same error characteristics (mean and standard deviation) are also shown in blue dashed lines.

- **Figure 6: I honestly cannot see any difference between these three plots. You need to revise the colour scale to highlight the differences.**
  As the figure is a bit redundant with figure 7 we propose to remove it.

- **Figure 7: I cannot see the REF line: is it under the green line or the red line? This is an important figure as it is the first time I realised this paper was about improved QC (more obs are assimilated with Experiment B than A).**



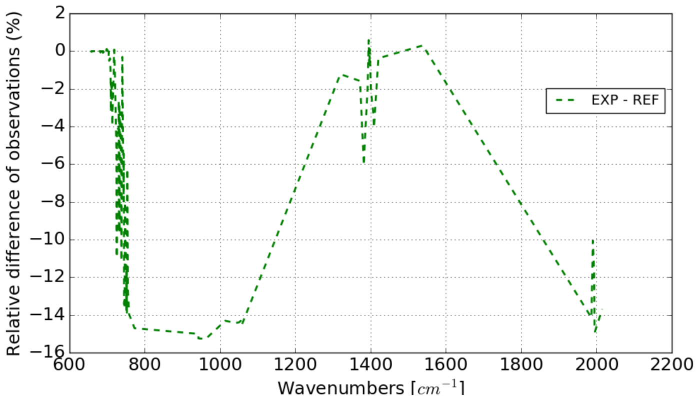


Figure 7 (a) has been changed with this upper-panel representing the number of assimilated data in REF and the second panel represent the the relative difference of number of observations for EXP (EXP.B) compared with REF (EXP.A being removed from this section). The caption was changed accordingly.

Tables:

- **Table 2: Should include the % partially cloudy**
  The column of partially cloudy observations was added in Table 2.
- **Tables A1 and A2 are unnecessary - this information is presumably included elsewhere. If not, a simple list of channel numbers would suffice.**
  Tables were removed.

Minor points:

- **P2 line 9: seems to be the first use of IR without the abbreviation being expanded.**
  Done
- **Section 2: this section is a light-touch description of the model and IASI, as it should be, but it is important to get the details correct in that case and make sure the writing is clear: P4 has a few poorly worded sentences, or poorly explained concepts.**
  - **P4 Line 7 - the background error statistics are not “derived from a climatological matrix”- it isn’t actually a matrix, and you do not explain how the ensemble information is incorporated.** We agree that this sentence was not well written and we replaced it with « The background errors are computed at each analysis time based on the 25-member assimilation ensemble (see Berre et al 2015 for further details). »

  - **P4 Line 22 - this area needs rewriting - Presumably you mean that the accuracy of the forward model calculation is limited by the accuracy of the NWP model, and that for some variables this is not sufficient to correctly model the observations?** This paragraph was rewritten : The observation operator allows to simulate observations from the model variables for comparison with the actual measurements. For satellites radiances, it includes a radiative transfer model.

The accuracy of the forward model calculation could be limited by the accuracy of the NWP model, for some variables this is not sufficient to correctly model the observations and these observations have to be discarded.

- **“Modelisation” -> “Modelling” in English!** Corrected.

- **P4 Line 29 - The McNally & Watts scheme is not clearly described. This scheme is now better described in the text :** The McNally and Watts (2003) scheme intends to detect clear channels and to assimilate channels unaffected by clouds even in a cloud-affected pixel. The channels are first re-ordered according to a ranking with respect to the altitude that reflects their relative sensitivity to the presence of cloud. After having applied a low-pass filter a search for the channel at which a monotonically growing departure can first be identified. Having found this channel all channels ranked more sensitive are flagged as cloudy and those ranked less sensitive are flagged clear.

- **P4 Line 31: In this section, there are numerous references to CTOP and Ne, but suddenly you switch to PTOP - maybe Pangaud (2009) used PTOP instead of CTOP but this switch is not necessary.** PTOP was modified into CTOP.

- **P5 line 1: “IASI is a key element of the payload of the Metop series of European: : :” P5 para 1: you may as well update this with Metop-C launch date.** The Metop-C launch data has been specified in then text: November 2018. « The third instrument was mounted on the Metop-C satellite, which was launched in November 2018. »

- **P5 line 22: It’s a bit far to say that failing to assimilate cloudy IR observations is a source of error.** This sentence was modified : “Assimilation of cloudy radiances is a crucial challenge for NWP centres as the cloudy observations discard represent an under-exploitation of hyperspectral sounders especially in sensitive meteorological areas (McNally, 2002; Fourrié and Rabier, 2004).”

- **P5 line 25: “allows to better describe” - not good english “allows a better desrcription of: : :”** Change made.

- **P5 line 29: You should reference the Baran parameterisation if you are going to mention it.** The reference to Baran, A. J., Cotton, R., Furtado, K., Havemann, S. , Labonnote, L.-C., Marenco, F., Smith, A. and Thelen, J.-C. 2014: A self-consistent scattering model for cirrus. II: The high and low frequencies. Q.J.Roy. Meteorol. Soc., 140: 1039–1057. doi:10.1002/qj.2193
- **P6 line 8: “an innovative challenge”? Remove the word innovative.** Done

- **P6 line 8: The sentence “In the context of...” doesn’t make sense.** The sentence was correted : « In the context of the preparation of all-sky assimilation, we plan to assimilate clear and cloudy observations that are completely covered in a homogeneous way, discarding the cases of fractional cloud observations. »

- **P6 line 21: They are not IASI L1c products - they are components of the L1c product.** This was corrected.

- **P6 line 24: this sentence is not clear either.** It has been reworded: For each class and each AVHRR channel, the cluster product provides the coverage of the class within the IASI pixel, the mean and the standard deviation of AVHRR brightness temperatures within the class.

- **P6: line 28: this sentence is not clear. I think it is a stretch to say something with one class can be less homogeneous than something with multiple classes - this is a bit subjective.** Our thought was that an important parameter is the standard deviation inside each class. The sentence was rewritten : A IASI FOV with several classes, each one having a small standard deviation and a mean radiance close to the ones of the other classes, can thus be more homogeneous than a FOV with a single class but with very large value of standard deviations.

- **P9 line 11: “aimed to propose” - that is a bit of a negative slant on this reference! “Proposed” would be better!** The change was made.

- **P17 line 15: No need to reference Table 3 here - it is a very basic table and you describe it all in the text,** The Table and the reference were removed.