Response to the Referees for *"Multi-sensor analyses of the skin temperature for the assimilation of satellite radiances in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS, cycle 47R1)"* by Massart et al.

We would like to thank each reviewer for their comments on the paper. taking them int account helped us improving the manuscript and the interpretation of the results. Please find below our responses.

1 Reviewer 1

1.1 Major Specific Comments

• ***Comment:*** *The introduction section lacks a description of the scientific context as it does not provide a description of related work that has already been performed at other NWP centres, internationally. For example, by saying "new approach" (on line 76), presumably this refers to being new at ECMWF and not within the international field of NWP. As an example, an early study by Garand et al (2004, Journal of Applied Meteorology and Climatology) described the Canadian system that already used a gridded skin temperature field for assimilating radiance observations (one of the "new" innovations described in this manuscript). They then evaluated the impact of including the background-error correlation between skin and air temperature, which in some respects goes beyond what is described in this manuscript.*

⋄ **Manuscript changes:** We added in the introduction the fact that skin temperature values were added to the control vector in other NWP centres in the early 90s.

We also added that our TOVSCV "approach differs from the one of some other NWP centres where the skin temperature is a gridded field optimised together with the atmospheric fields."

We changed the " A new approach (...) " sentence with "An enhanced approach to TOVSCV is presented here which follows the implementation of a gridded skin temperature field of some other NWP centres."

The work of Garand *et al.* (2004) is mentioned in the conclusion (see response to comment below).

• ***Comment:*** *The design of the experiments does not allow for a systematic evaluation of how the different modifications affect the results. As already mentioned, in some situations the specified standard deviation of skin temperature appears to have been changed in the new formulation in combination with the use of a common gridded skin temperature field for all instruments and* *introducing both spatial and temporal background-error correlations. This makes it difficult to draw any conclusions about the impact of the changes over sea ice, where it is noted that the standard deviation is significantly changed. This aspect should be more fully explored.*

⋄ **Response:** The study presented here extends a previous one published as an ECMWF technical memorandum (Massart *et al.* 2020). In that previous study, we presented an evaluation of the impact of both spatial and temporal background error correlations in the SKTACV approach.

We also showed that the skin temperature background error standard deviation derived from the EDA needs to be inflated for the TOVSCV approach to compensate for the lack of spatial and temporal correlations. In this study, we wanted the configuration of the control experiment (TOVSCV approach) to be as close as possible to the configuration of the operational analysis. In particular, we kept the value of 7.5 K for the background error standard deviation over sea-ice. This value is larger than over other surfaces because it accounts not only for random errors but also for the large systematic errors present in polar regions.

For the SKTACV experiment, using the same values of background error standard deviation as for the TOVSCV experiment was not relevant unless reducing the values to account for the spatial and temporal correlations. Instead of trying to tune the reduction factor, we made the choice of using directly the background error standard deviation derived from the EDA. We showed in Massart *et al.* (2020) that this produces similar analyses in both experiments except for over the polar regions as shown in the present work.

The difference found in polar regions has two origins. First, the EDA spread is underestimated and we addressed this by applying a scaling factor of 3. Second, the large systematic errors are not accounted for in the EDA. We could have applied another scaling factor to compensate for this and tuned this factor. Instead, we chose not to change this and to work in parallel on the reduction of the systematic errors (not part of this study). This means that it may be indeed difficult to draw any conclusions about the impact of the changes over sea-ice, but in the meantime we have a reference for future improvement of the SKTACV approach in polar regions.

⋄ **Manuscript changes:** We added that the 7.5 K value for the TOVSCV approach accounts not only for random errors but also for the large systematic errors present in polar regions.

We also added more details at the end of the section 3.1: "The final value of around 1.5 K is still significantly lower than for the TOVSCV experiment, even if we consider the effective standard deviation from taking into account the spatial correlations. We attribute the difference to the accounting of systematic errors in the TOVSCV experiment. We chose here not to account for these errors by inflating the standard deviation further, but to work in parallel on the reduction of the systematic errors in polar regions (not part of this study). This means that we may expect some differences between the two experiments in polar regions due to the difference in value of the background error standard deviation on top of the change of approach."

• ***Comment:*** *The reduced amplitude of the analysis increments with the new approach are assumed to be related to the additional constraints imposed on skin temperature increments in the new approach. These constraints result from the addition of spatial and temporal correlations and also the use of a single 2D skin temperature field for all microwave instruments and another for all infrared instruments. I would think that some additional analysis of the results could provide concrete evidence of this and possibly suggest if any of these differences are more important than the others. For example, the increments obtained with the two approaches could be compared (at observation locations) in terms of their spatial, temporal and between instrument variability. Showing a much higher variability when using the previous approach, either in* *space, time or between instruments, would support the assertion that one or more of the added constraints cause reduced increments. The authors could then also comment on if the higher variability is more or less physically realistic.*

⋄ **Response:** The Appendix A (page 12) of this reply illustrates the space, time and between instruments variability of the analysis increment at observation locations for the METOP-A AMSU-A and IASI instruments, and for METOP-A and METOP-B for the inter-instrument variability. We chose to present these results in order to have an infrared and a microwave instrument on board of the same payload and because the METOP-A and METOP-B trajectory are close to each other. We also present only the results for the month of July 2019. These selected results are representative of the results obtained for other instruments and for other months.

Overall the SKTACV experiment has lower variability in space/time and between instruments for its skin temperature analysis increments than for the TOVSCV experiment. Over land, the spatial variability of the SKTACV experiment is mainly lower than 0.4 K for the microwave increments and lower than 0.2 K for the infrared increments. These values are about 2 to 3 times lower than the ones of the TOVSCV experiment. We expect the model and therefore the background to have a good representation of the spatial variation of the skin temperature. We would then expect a low spatial variability in the skin temperature increment and in that sense the SKTACV experiment would be more realistic.

For the temporal variability of the skin temperature, the model is less accurate and we can expect higher values than for the spatial variability. It is nonetheless difficult to directly link the model temporal variability with the variability found in the analysis increments as METOP-A is sampling only the 9:30 and 21:30 local solar times and are also subject to the data screening. Therefore it is difficult to say if the values of the temporal variability found in the SKTACV experiment are more or less realistic than the higher values found in the TOVSCV experiment. We can nonetheless notice an increase of the variability for AMSU-A over the ocean in the SKTACV experiment.

For the between instruments variability, we have a lower global coverage than for the two others. Yet, the main behaviour is similar with lower values for the SKTACV experiment than for the TOVSCV experiment. With our criteria for computing the statistics, we have limited the influence of the temporal and spatial variability. The between instruments variability should then be of the order of the instrumental noise which makes the SKTACV experiment more realistic.

We have decided not to include these results in the revised version of the document, but we added a paragraph in the 4.2.2 section: "We found a reduction of the analysis increment variability in the SKTACV experiment compared to the TOVSCV experiment except for the microwave instruments over the ocean. The reduction is a combined effect of the addition of spatial and temporal correlations and of using one field per instrument type in the SKTACV experiment. The increase over the ocean in the microwave instruments can be attributed to an increase in the temporal variability in the SKTACV experiment."

1.2 Other Specific Comments

• ***Comment:*** *Equation 9: The cost function is not shown in the incremental form, which makes the meaning of H^T less clear, since there is no corresponding tangent-linear operator. Either the cost function and gradient should be presented in incremental form or at least the* *correspondence between the shown form and what is actually used in practice well explained in the text.*

◇ **Response:** We have introduced the incremental form (now Eq. (10)) and we have also changed the text. Similarly, we have changed Eq.(16) (now Eq. (18)). Note that the previous equations do not appear in the "track-change" version of the revised document to improve the readability.

• ***Comment:*** *Equation 18: Shouldn't this be sqrt of the B matrices? If not the variable transformation should be better explained and a reference given.*

◇ **Response:** This was a mistake, the change of variable uses $\mathbf{B}^{1/2}$. This was corrected.

• ***Comment:*** *Line 255: I don't think it's correct to simply refer to this approach for skin temperature as 4D-EnVar, since a major part of that approach is how the 4D ensemble covariances are implemented with spatial localisation. In the present study, the following section describes how covariance parameters in a wavelet-based representation are estimated from an ensemble. Also, the Liu et al reference is probably not the best for 4D-EnVar, since the approach in that paper does not include spatial localisation and uses ensemble covariances purely in observation space, making it quite different from what is currently considered the 4D-EnVar approach. It would be better to refer to Lorenc (2003) and Buehner (2005).*

◇ **Manuscript changes:** The new sentence is "our strategy is effectively a hybrid between a 4D-Var for most variables and, for the skin temperature fields, a hybrid 4D-Var where the background error is derived from an ensemble for the skin temperature fields (Lorenc (2003), Buehner(2005))."

• ***Comment:*** *Line 307: This choice of wording is unclear and possibly misleading. Consider rewording to avoid the use of "replace".*

◇ **Manuscript changes:** The new sentence is "We adapted the formulation such that for a given wavelet index we specifically added for the skin temperature a time covariance matrix on every point of a horizontal grid associated with this wavelet. "

• ***Comment:*** *Line 316: It is not clear if these diagnosed length scales are actually used as part of the B matrix specification. I thought that the (diagonal) wavelet approach directly computes the spatial correlations in wavelet space from the supplied ensemble. Please clarify if the diagnosed values referred to in the text (both spatial and temporal) are merely for diagnostic purposes or are somehow used in the B matrix specification.*

◇ **Response:** Using an ensemble is one option to build the wavelet **B**. Another option is to specify the spatial and vertical correlation length-scale and to build the wavelet **B** from those. We specifically developed this second option for our need. For the skin temperature, we do not have vertical correlation but time correlation instead. We changed the text to reflect the fact that the diagnosed values of spatial and temporal correlations are used to build the wavelet **B**.

◇ **Manuscript changes:** "We did not use directly the members of the EDA to build the wavelet covariance model for the skin temperature as it is the case for all the other variables of the control vector. Instead, to allow more flexibility, we developed the facility to build the wavelet covariance model from the estimation of the local spatial correlation length-scale and the estimation of the temporal correlation."

• ***Comment:*** *Line 341: What about MHS instruments on the platforms with AMSU-A? Are they not assimilated?*

◇ **Response:** The data from the MHS instruments are assimilated using the "all-sky" route (Geer *et al.*, 2018), and this presently does not have a skin-temperature estimation capability via SKTACV or TOVSCV. One reason for this is concerns over aliasing cloud information into skin-temperature increments, though this aspect could be revisited in the future.

• ***Comment:*** *Figure 3: Probably not appropriate to use a line graph when the x-axis represents distinct satellite sensors. A bar graph would be better.*

◇ **Manuscript changes:** We changed the plots with a bar graph.

• ***Comment:*** *Line 410: I suppose this is referring to the large values of stddev of the difference between the increments of the two experiments. This is not clear from the current text.*

◇ **Manuscript changes:** we changed the text with: "The large values of standard deviation of the skin temperature analysis difference over the sea-ice are due to large variability of the increments in the TOVSCV experiment. This experiment has indeed a more active skin temperature analysis due to the large background error standard deviation of 7.5 K as described previously."

• ***Comment:*** *Line 411: Would be helpful to provide average values for each of the 3 types of radiance instrument from each experiment to support this statement.*

◇ **Response:** We computed the standard deviation of the skin temperature analysis increment for each instrument and for each surface type for both experiments. We then computed the averaged value per instrument type (see Table 1 below). We also computed the normalised change in mean standard deviation of the SKTACV experiment with respect to the TOVSCV experiment (see Table 2 below).

Table 1: Mean of the skin temperature analysis increment standard deviation (in K) per instrument type, per season and per surface type, from the SKTACV experiment and from the TOVSCV experiment (in parenthesis).

		Land	Sea	Ice
Microwave	JAS 2019	0.606 (0.827)	0.236 (0.069)	1.281 (1.372)
	JFM 2020	0.692 (0.756)	0.214 (0.062)	1.280 (1.596)
Geostationary	JAS 2019	0.371 (0.082)	0.166 (0.002)	0.103 (0.182)
	JFM 2020	0.391 (0.148)	0.177 (0.003)	0.191 (0.145)
Hyper-spectral	JAS 2019	0.329 (0.706)	0.178 (0.239)	0.377 (0.787)
	JFM 2020	0.283 (0.482)	0.122 (0.129)	0.346 (0.685)

◇ **Manuscript changes:** We changed the text to introduce the fact we computed the above statistics: "We also computed the mean of the analysis increment standard deviation by instrument type for each of the two experiments."

We used the values of Table 2 in the text.

We also introduced the difference in the treatment of the geostationary infrared instruments in the two approaches: "In the TOVSCV experiment, the geostationary infrared instruments have little sensibility to the surface and therefore the analysis increment is close to zero on average and has little variability. Meanwhile, by construction, the average and variability in the

Table 2: Change in mean of the skin temperature analysis increment standard deviation (in %) per instrument type, per season and per surface type, for the SKTACV experiment compared to the TOVSCV experiment.

		Land	Sea	Ice
Microwave	JAS 2019	-26.75	242.57	-6.60
	JFM 2020	-8.55	247.37	-19.81
Geostationary	JAS 2019	353.79	-	-43.63
	JFM 2020	164.55	-	32.37
Hyper-spectral	JAS 2019	-53.42	-25.78	-52.18
	JFM 2020	-41.18	-5.42	-49.47

SKTACV experiment are similar for all infrared instruments and their values are constrained mainly by the hyper-spectral instruments."

We finally changed the last paragraph of the section to address the initial comment, with: "Due to the difference in background error standard deviation between the two experiments, the SKTACV experiment has a less active skin temperature analysis by about 7 % (JAS 2019) and 20 % (JFM 2020). The largest values of standard deviation of the skin temperature analysis difference over the sea-ice compared to the other surface types are thus partially due to a larger variability of the increments in the TOVSCV experiment and also to a larger variability over this surface type."

• ***Comment:*** *Line 417: Please briefly mention what these "other errors" could be. Is the accuracy of surface emissivity a concern?*

◇ **Manuscript changes:** The errors we are thinking of in this sentence are indeed the accuracy of surface emissivity, the cloud screening and the specular assumption for microwave frequencies over snow and sea-ice areas. We changed the sentence with: "The TOVSCV experiment in contrast allows for e.g. possible inaccuracy in the surface emissivity or in the cloud screening, or viewing-angle dependent biases from the specular assumption (when relevant) to be aliased into skin-temperature increments."

• ***Comment:*** *Line 430: What is the implication of this statement? This needs to be better explained.*

◇ **Response:** The larger increments in the SKTACV experiment in the Arctic region, allowing the analysis to better fit channel 6 of ATMS, do not have a significant impact on the analysis because it concerns only few points.

◇ **Manuscript changes:** We change the sentence with: "Looking at these outliers in more details, the larger increment in the SKTACV experiment allows the analysis to better fit channel 6 of ATMS on (not shown). Nevertheless, because it concerns only few points, this does not have an overall significant impact on the analysis. "

• ***Comment:*** *Line 436: How would inclusion of the background-error correlation between skin temperature and various atmospheric variables affect this? Is it technically possible to include such correlations with the hybrid 4d-var + "EnVar" approach used in the new approach?*

◇ **Response:** This is indeed a good point. We can imagine that if the skin temperatures are similar in both experiments, accounting for the cross-correlation would help bringing the weather parameters of both experiments closer.

Surely including the cross-correlation in the SKTACV approach is a promising perspective and we want to thank the reviewer for the suggestion. We do not forsee any technical issue to follow the approach of Garand *et al.* (2004) introducing a balance operator between the skin temperature fields and the atmospheric temperature.

◇ **Manuscript changes:** We added the following sentence in the conclusion: "Having the skin temperature as fields on the same grid as the atmospheric temperature will also allow to introduce background error correlation between them following the approach of Garand *et al.* (2004)."

• ***Comment:*** *Line 457: Please discuss the possible cause of the skin temperature being driven to fit the surface-sensitive channels.*

◇ **Response:** To fit the observed radiances, the assimilation is allowed to modify the atmospheric variables and the skin temperature. The atmospheric variables are in particular constrained by surrounding observations too due to the spatial correlation in the background error. On the other hand, the skin temperature has only its background as a constraint in the TOVSCV experiment. Without other constraints, it can be freely adjusted to improve the fit to the channels that are surface-sensitive.

◇ **Manuscript changes:** We changed the text with: "For the TOVSCV experiment, the skin temperature for each field-of-view is only constrained by its background in opposition to the atmospheric variables that are also constrained by surrounding observations due to the spatial correlation in the background error. Therefore, the skin temperature can be purposely adjusted to improve the fit to the channels that are surface-sensitive. This is illustrated in the two ATMS cases where the analysis can lead to large skin temperature increments."

• ***Comment:*** *Line 467-469: I do not understand this statement. Please clarify the reasoning behind it.*

◇ **Response:** We wanted to express the fact that in the TOVSCV experiment, the skin temperature could be unrealistically changed to try to gain a little improvement in the fit to the radiance profile.

◇ **Manuscript changes:** We changed the text with: "For the first scene, despite the large skin temperature increment, the analysis-fit is only marginally improved, which suggests that the TOVSCV approach can produce unrealistic skin temperature adjustments."

• ***Comment:*** *Line 481: This wording (i.e. "the fit is increased") is confusing. Please find a more straightforward way of expressing the change in the fit to the analysis.*

◇ **Manuscript changes:** We changed the text with: "For ATMS, the analysis-fit is larger by 10% for channel 6 and by up to 6 % for channels 18 to 22 (depending on the season) in the SKTACV experiment compared to the TOVSCV experiment."

• ***Comment:*** *First 6 paragraphs of "Conclusions": There is too much content here from the introduction and methods sections. This should be condensed significantly for the conclusions section.*

◇ **Response:** We have condensed the conclusions section.

1.3 Technical Corrections

• ***Comment:*** *Equation 10 and in equations elsewhere: The adjoint of H should be bold non-italic as in the main text.*

⋄ **Response:** Done.

• ***Comment:*** *Figure 1 and others: "micowave" should be "microwave".*

⋄ **Response:** Done.

2 Reviewer 2

2.1 General comments

• ***Comment:*** *However, radiance observation used for the skin temperature is the data set used for atmospheric analysis and surface sensitive channels are not sufficiently used. Therefore, it is unclear whether the consideration of spatial and temporal consistency in SKTACV is better than TOVSCV under the limited surface sensitive radiance data use.*

Response:

As a first step, we chose to have a conservative approach with a SKTACV configuration as close as possible to the TOVSCV configuration (apart over sea-ice, see reply to Reviewer 1). The purpose of this "model evaluation" paper was to assess the impact of the new approach before further improving it. We demonstrated that the new approach is mostly neutral in term of standard data assimilation and forecast skill measures, which gives us confidence in further developing it.

As mentioned in the paper (fourth paragraph of the introduction), the spectral channels are carefully selected to avoid those which are very sensitive to the surface, and at the same time to keep those which have a significant positive forecast impact in the ECMWF system (Bormann et al., 2017). We believe that with the SKTACV approach we will be able to add more surface sensitive channels in the near future and then we will be able to further assess in what measure the spatial, temporal and instrumental consistency of this approach is beneficial compared to the current one.

Concerning the spatial and temporal consistency in the SKTACV approach, see reply to Reviewer 1 and Appendix A (page 12).

2.2 Specific comments:

• ***Comment:*** *Line 52: "In the absence of snow, land emissivity varies very little, temporally or spatially".*

This is true for infrared case. However, for microwave range, land surface emissivity varies depending on surface vegetation and soil conditions. Add some words to specify this sentence is valid for infrared case only.

⋄ **Manuscript changes:** We changed the sentences with "Land emissivity varies very little, temporally or spatially for the infrared observations. An emissivity database is then used for those observations"

• ***Comment:*** *Line 74: "Because of this lack of constraint by other surrounding measurements, the skin temperature adjustment likely compensate for other errors in the background state,"*

Line 509: "other errors"

What kind of errors had been compensated in TOVSCV? The errors were IFSś systematic bias error? Please clarify what kind of error had been compensated in TOVSCV and justify the new approach SKTACV could use the observation information appropriately.

⋄ **Response:** As mentioned also in a response to a comment of Reviewer 1, the errors we are thinking of are mainly the inaccuracy of surface emissivity and issue in the cloud screening. For the "other errors" mentioned Line 509, they are probably linked to the inaccuracy of surface emissivity because the the cloud screening is less of an issue for the microwave data.

⋄ **Manuscript changes:** We added in the sentence of the Line 74: "(for e.g. possible inaccuracy in the surface emissivity or in the cloud screening)".

We modified Line 509 with: "This suggests that the TOVSCV approach may use the skin temperature to compensate for inaccuracy in the surface emissivity in the northern hemisphere for JFM 2020"

• ***Comment:*** *Table 1: Add channel numbers which are used in the data assimilation.*

⋄ **Response:** It would be possible to add the channel numbers for the microwave instruments, but not the wavenumber selection for the infrared ones (too many of them). Furthermore, the microwave channel selection is surface-dependent which would complicate the table entries. We decided to keep the Table 1 simple, especially because all the relevant information can be found in Table 1 of Bormann et al. (2017) which is mentioned in the text.

• ***Comment:*** *Figure 2: Over land, generally, microwave standard deviation is larger than infrared one. Is the reason insufficient microwave cloud-screening over land? Add some descriptions from the point of view in the text.*

⋄ **Response:** The difference in standard deviation values over the land between the two instrument types could be explained by several factors. First, the we assimilate more microwave channels over land with a higher surface-sensitivity than in the infrared. Secondly, more microwave instruments are available, providing more of the diurnal sampling, that could be large over land. We can not rule out that insufficient cloud screening may play a role, but that is probably less likely to be put into a skin-temperature increment in the SKTACV framework, due to the spatial and temporal background error correlations.

⋄ **Manuscript changes:** The number of microwave data assimilated over land and their higher surface-sensitivity was already discussed in section 4.1.1. We nevertheless changed the paragraph with: "The sensitivity to the surface and the type of information available from the assimilated microwave and infrared channels differs considerably, including the number of surface-sensitive channels assimilated. For the infrared sounders, many channels with relatively strong surface-sensitivity are assimilated over ocean from the hyper-spectral sensors, but many observations are screened out due to cloud contamination. The use of surface-sensitive channels is also more cautious over land. On the other hand, there is a wider range of microwave sounders available, with different overpass times, and these are less affected by cloud-contamination. These characteristics will affect the response of the skin-temperature estimation"

We also added a sentence to explain the difference in standard deviation values over the land between the two instrument types end of section 4.2.1: "The higher values over land for for the microwave field can be explained by both the higher surface-sensitivity and by the larger number of available instruments (providing more of the diurnal sampling)."

• ***Comment:*** *Line 515, and Figure 7 (c): Larger temperature difference at 850 hPa in Antarctica would be below the surface. Which depth are represented as red colored area?*

⋄ **Response:** That is right and we had checked before producing the figure that the value at 850 hPa corresponds to the value at the lowermost model level and does not result from some interpolation issues.

⋄ **Manuscript changes:** We appended the figure caption to reflect this, with: "Note that when the 850 hPa level is below the surface, the lowermost model level is used."

• ***Comment:*** *Line 582:*

The title of section 5 should be "Summary and Conclusions" instead of "Conclusions" because many sentences and their contents in section 5 are just a repeat of the previous section.

⋄ **Response:** We followed Reviewer 1 advice to reduced the length of the conclusions section. Therefore we will keep the current title.

2.3 Technical corrections:

• ***Comment:*** *Line 149:Ri-1 should be Ri*

Done.

• ***Comment:*** *Line 188:"," should be inserted.*

Done.

• ***Comment:*** *Line 224:"Tab 1" should be written as "Table 1" in the text.*

Done.

• ***Comment:*** *Line 518:"Figs. 7(a) and (b)" should be "Figs. 7(a) and (c)"*

Done.

• ***Comment:*** *Line 530:"Figs. 7(c) and (d)" should be "Figs. 7(b) and (d)"*

Done.

• ***Comment:*** *Line 677:Add URL of this reference.*

Done.

3 Executive editor

- ***Comment:*** *Even if it can not be made available to the public due to license issues state this and also make clear (by identifier) that you saved the exact version of the code, the results of the publication were produced with.*

Response: The version of the code is already specified in the text in the section 4.1 Experiments (IFS cycle CY47R1). This version is saved as an official ECMWF code release. Moreover, we added: "The source code of the ECMWF IFS model is not available for public use as it is intellectual property of the ECMWF and its member states." in the "Code and data availability" section.

A Variability of the analysis increment

We present hereafter some examples of the variability of the analysis increments of skin temperature at the observation locations, for the two approaches. We illustrate the spatial, temporal and inter-instrument variability.The examples are for the month of July 2019 and we focused on the METOP-A AMSUA and IASI instruments, and on METOP-A and METOP-B for the inter-instrument variability.

A.1 Spatial variability

To illustrate the spatial variability of the analysis increments, we first selected for each assimilation cycle all the increments from the same instrument, and in the same 1 hour time slot, and falling in the same a $4° \times 4°$ grid cell. We then computed the standard deviation of all the increments in each grid cell when the number of them was greater or equal to 5. This provided a value of the spatial standard deviation per cycle in the grid cells where there was more than 4 observations in a 1 hour time slot. We finally average all the cycles of the month (Fig. A.1).

A.2 Temporal variability

To illustrate the temporal variability of the analysis increments, we first selected for each assimilation cycle all the increments from the same instrument on a $1° \times 1°$ grid cell to reduced the spatial variability. The number of increment was very low and mostly less than 2 (ascent and descent trajectory of the satellite). To increase the sample size, we selected all the increments of the month in each grid cell and we then computed the standard deviation when the number of them was greater or equal to 5 (Fig. A.2).

A.3 Inter-instrument variability

To illustrate the inter-instrument variability of the analysis increments, we first selected for the whole month all the increments from two instruments, and in the same 1 hour time slot, and falling in the same a $4° \times 4°$ grid cell. We then computed the difference between the instruments of the mean increment of each time slot and each grid cell. We finally computed the standard deviation of the differences over the time slots for each grid cell when their number was greater or equal to 5 (Fig. A.3).

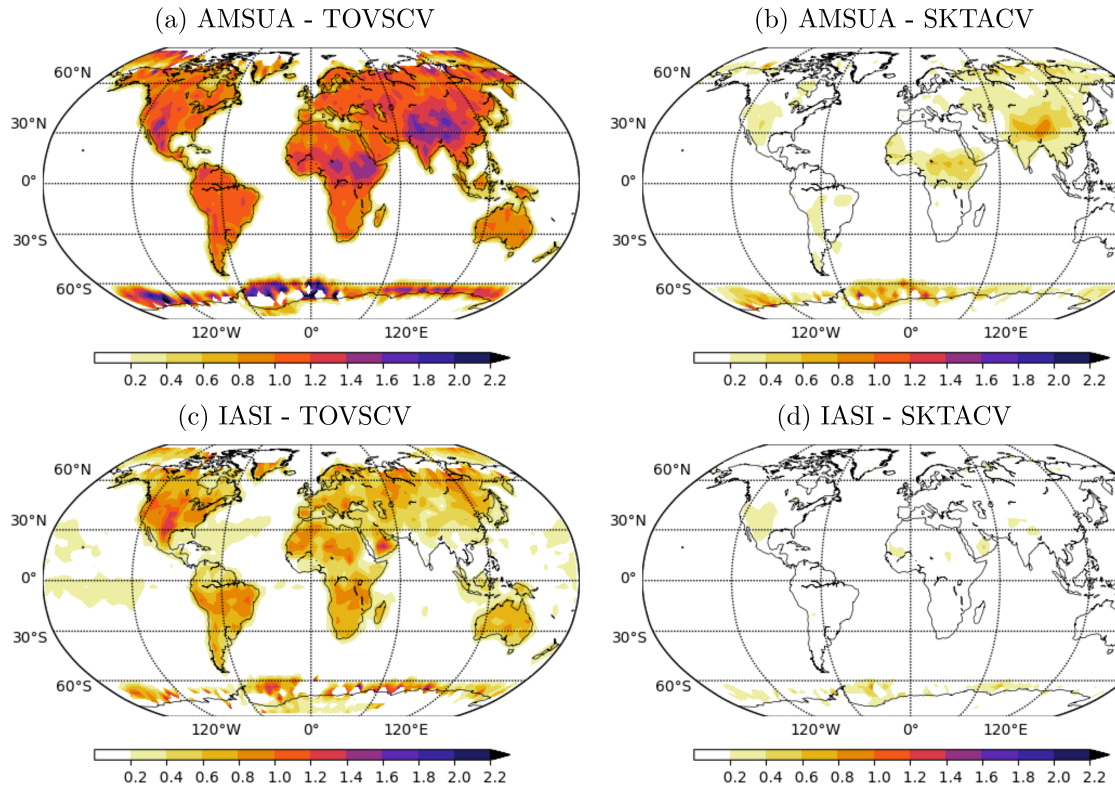

Figure A.1: Illustration of the spatial variability of the analysis increment (in K) at observation locations for METOP-A AMSUA (top) and IASI (bottom) in July 2019. Left: increments from the TOVSCV experiment. Right: increments from the SKTACV experiment.

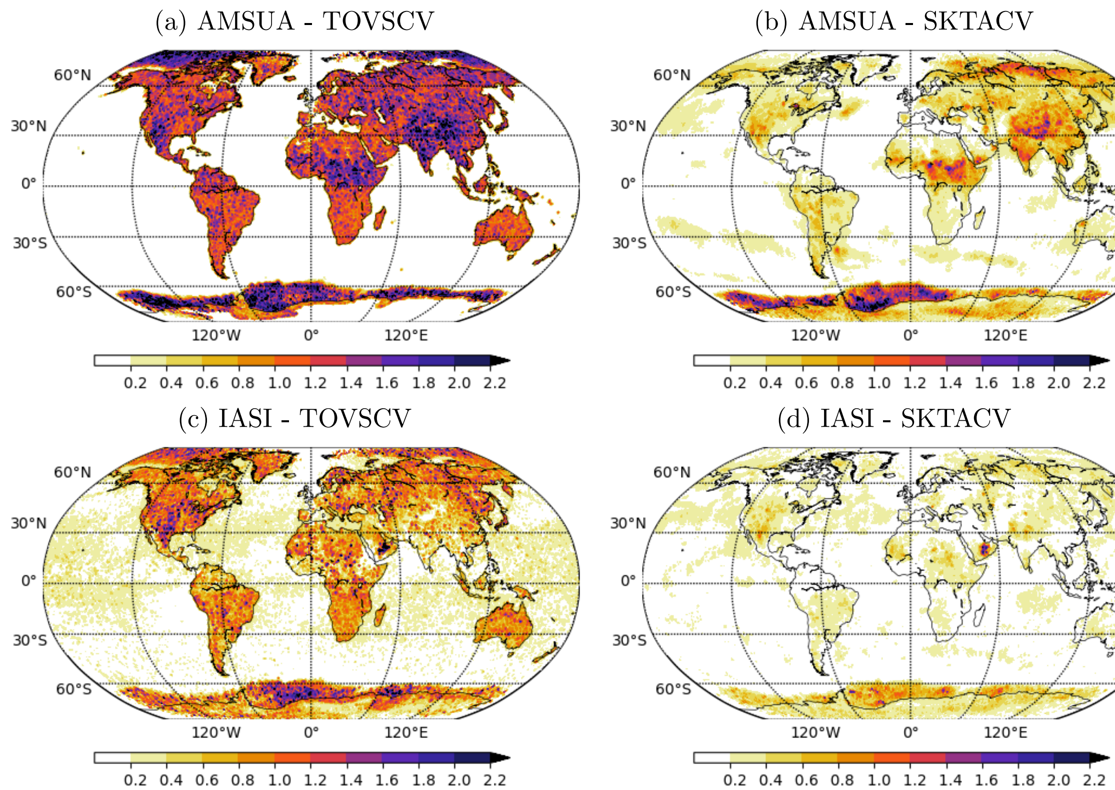

Figure A.2: Same as Fig. A.1 but for the temporal variability.

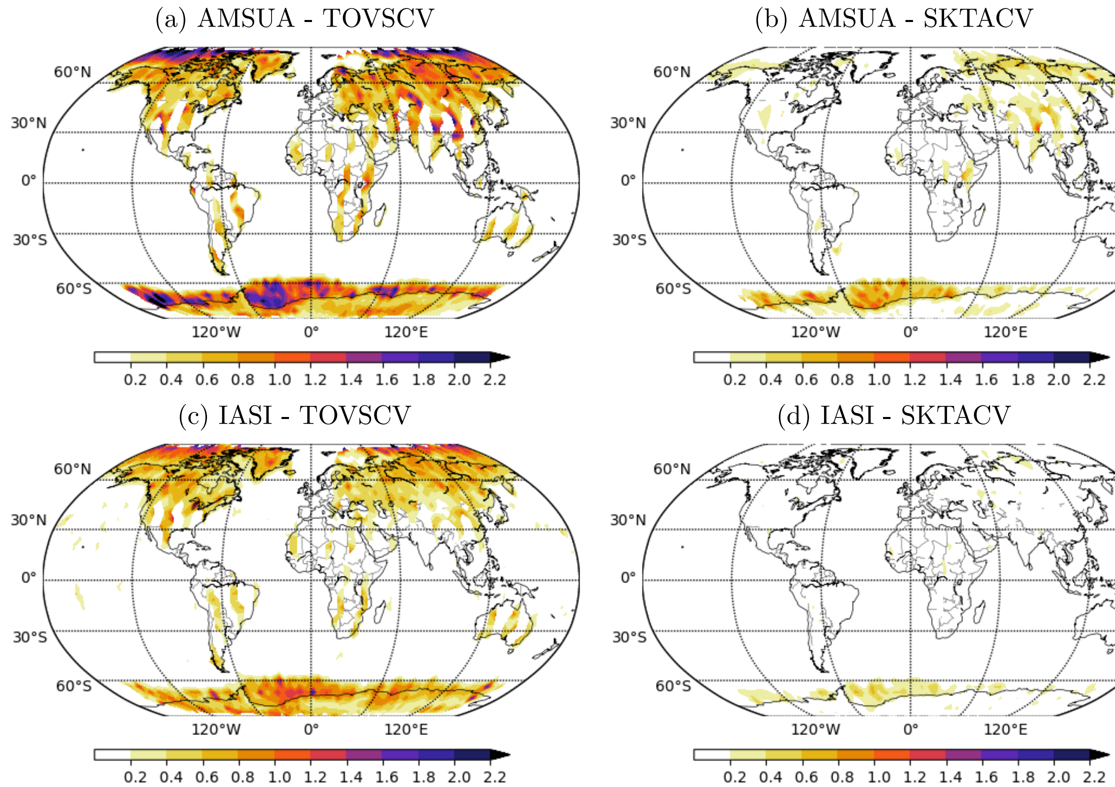

Figure A.3: Illustration of the METOP-A / METOP-B variability of the analysis increment (in K) at observation locations for AMSUA (top) and IASI (bottom) in July 2019. Left: increments from the TOVSCV experiment. Right: increments from the SKTACV experiment.