

Technical Memo

913

Extended control variable (XCV): skin temperature background correction for the assimilation of clear-sky microwave radiances

S. Massart

October 20, 2023

Series: ECMWF technical memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/en/publications/>

Contact: library@ecmwf.int

© Copyright 2023

European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

The microwave radiances are among the observations that have an important contribution when it comes to improving the Integrated Forecasting System (IFS) forecast skill. In clear-sky condition, these radiances have frequencies that make them sensitive to both the surface and atmosphere. Using them in the atmospheric analysis requires surface information, and therefore a specific treatment as the surface and atmospheric analyses are currently not strongly coupled.

For instance, the clear-sky microwave radiances observation operator expects a skin temperature value at the observation location and time, together with the profiles of the atmospheric variables along the viewing path. To get a skin temperature as accurate as possible, it was included in atmospheric analysis control vector (TOVS Control Variable or TOVSCV approach). This inclusion allows to adjust the skin temperature during the atmospheric analysis, but only at the observation time and location.

The SKTXCV approach was developed recently as an alternative of the TOVSCV approach with a significant difference that the skin temperature is optimised in model space. So far, the benefits of this new approach were neutral, but experimentation allowed us to highlight a potential biases in our analysis of the microwave-based skin temperature. Among the potential sources of biases that we discuss in this document, we believe that the largest contribution comes from the the skin temperature background. This bias is addressed, in model space, thanks to the SKTXCV approach, in a so-called persistence approach where the analysis of the previous days is used to predict it.

We ran an experiment over a northern hemisphere winter season with a new SKTXCV configuration based on the persistence approach. We compared to it to a reference SKTXCV experiment and a control experiment using the TOVSCV approach. The persistence configuration tends to use more clear-sky observations, and tends to reduce the first-guess departure to these observations.

The medium-range forecast from the persistence configuration is overall improved compared to the two other configurations. The largest improvement are located over the Arctic and to a lesser extend in the northern extra-tropics, and for the forecast up to day 5 to day 7 when compared to observations. For example, the improvement of the anomaly correlation of the geopotential at 500 hPa in the Arctic is between 3% to 4% in the first few days.

The correction of the background for the microwave skin temperature fields generally converges in 10 to 15 days, depending on the region and surface. Yet, there is no safeguard to prevent the correction to diverge in this approach. Considering a nudging constrain on the correction or adding a component relative to the correction in the weak constrain formulation are two paths that could be explored next.

Plain Language Summary

In the Integrated Forecasting System (IFS), a variety of observations are assimilated to provide the best estimate of atmospheric variables for forecasting. Microwave radiances are important observations that help improve the IFS forecast skill. These radiances are sensitive to both the surface and atmosphere in cloud-free conditions. To assimilate the microwave radiances, an observation operator is required that expects a skin temperature value at the observation location and time, along with the profiles of atmospheric variables along the viewing path. The SKTXCV approach is used here to adjust the skin temperature during atmospheric analysis to make it as accurate as possible. However, this approach relies on a prior estimate of skin temperature that is biased. In this document, an approach is presented to correct this bias, that results in improving the medium-range forecast up to day 5 to day 7.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 2 | Towards a bias-free α-skin temperature background | 4 |
| 3 | Results | 5 |
| 3.1 | Configuration | 5 |
| 3.2 | Mean analysis increment | 6 |
| 3.3 | Bias correction of the α_{mw} -fields | 6 |
| 3.4 | Analysis mean | 8 |
| 3.5 | Observation diagnostics | 9 |
| 3.6 | Forecast | 11 |
| 4 | Conclusion | 15 |
| 5 | Acknowledgements | 16 |
| | Appendix A Skin temperature background correction design | 18 |
| A.1 | Theory | 18 |
| A.2 | Bias model based on persistence | 18 |
| A.3 | Bias model from machine learning techniques | 20 |
| A.4 | Offline tests | 24 |
| A.5 | Online tests | 25 |
| A.6 | Choice of background bias model | 25 |
| A.7 | Conclusion | 26 |
| | Appendix B Mean correction per surface type and region | 28 |
| | Appendix C Mean analysis departure | 29 |
| | Appendix D Root mean square error maps of surface parameters | 31 |

1 Introduction

The European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric analysis relies on good quality observations and good usage of these observations by the observation operator. The clear-sky microwave radiances are among the observations that have most the important contribution when it comes to improving the forecast skill (McNally, 2015; Bormann *et al.*, 2019). These radiances have frequencies that make them sensitive to both the surface and atmosphere. The surface and atmospheric analyses are currently only weakly coupled in the analysis through the forecast model used in the cycling of the assimilation system. Therefore, a specific treatment of the surface in the atmospheric analysis is required. For instance, the clear-sky microwave radiances observation operator expects a skin temperature value at the observation location and time, together with the profiles of the atmospheric variables along the viewing path.

To get a skin temperature as accurate as possible, it was included in the ECMWF Integrated Forecasting System (IFS) atmospheric 4D-Var control vector. This approach is known as TOVS Control Variable (TOVSCV, ECMWF, 2019). The inclusion of the skin temperature allows to adjust it during the atmospheric analysis and circumvents the de-coupling between the atmospheric and surface analyses.

We recently proposed a new approach to this skin temperature analysis used for the assimilation of clear-sky radiance (SKTXCV approach, Massart *et al.*, 2020, previously named SKTACV). Practically, we have introduced additional hourly skin temperature fields in the atmospheric 4D-Var control vector. These fields are associated with the infrared (IR) instruments (referred to as α_{ir} -fields) and with the microwave (MW) instruments (referred to as α_{mw} -fields).

With this new approach, we have already simulated a winter and summer seasons with the IFS cycle 47R1 (Massart *et al.*, 2021). Compared to the current TOVSCV approach, we found that the SKTXCV approach is mainly neutral in terms of analysis and forecast skill. One main finding from the SKTXCV experiments was that the three months averages of the skin temperature analysis increments (analysis minus background) for the α_{mw} -fields could locally reach large values, sometimes three times larger than the standard deviation of the analysis increment. We then hypothesised that we have significant sources of bias in our analysis that needs to be addressed.

In an unbiased linear analysis system, the mathematical expectation of the analysis increment should be zero (Talagrand, 1999). In practice, the mathematical expectation is usually approximated with the time-average using the ergodic hypothesis. Therefore, a non null time-average analysis increment, as we found in our simulation, is a useful a posteriori diagnostic that points towards biases in the system. Such diagnosed biases can originate from the background, the observations or the observation operator, but this diagnostic alone can not disentangle the different sources of potential bias.

We may have these three potential sources of bias to explain the analysis increment bias of our simulation. First, the observation bias which is partially addressed by the variational bias correction (VarBC, Dee, 2004). As for the bias diagnosed by the mean analysis increment, the bias diagnosed by VarBC can not be attributed to either the observation or the observation operator. The expected residual observation bias after VarBC is much lower than the diagnosed analysis increment bias. We can assume that it is unlikely that the observations are the predominant source of bias.

The observation operator is the second source of potential bias. It is based on the Radiative Transfer for TOVS (RTTOV, Matricardi *et al.*, 2004; Saunders *et al.*, 2018) code. Errors in the cloud-screening can for example lead to biases that can be aliased into biases in the skin temperature. The description of the surface characteristics, including the description of the surface through an effective emissivity and the skin temperature may also be prone to biases. Additionally, the assumption of specular reflection has been found to be a considerable source of viewing-angle dependent bias over snow and sea-ice surfaces

for the forward calculations of microwave frequencies (Guedj *et al.*, 2010; Bormann *et al.*, 2017), and this is currently being addressed in parallel.

We do not know how much of the skin temperature analysis increment bias comes from the background and how much comes from the observation operator over snow and sea-ice surfaces. Yet, preliminary results using the Lambertian reflection in the assimilation of surface-sensitive microwave sounding radiances over these surfaces show that the biases coming from the observation operator are much lower than the ones observed in the mean analysis increment of the α_{mw} -fields (Bormann, 2021). In this document, we then assume as an initial step, that the contribution of the observation operator to the bias we want to address is negligible.

The last source of potential bias comes from the background. The background values of the α_{ir} and α_{mw} skin temperature are a copy of the the model skin temperature values. Over land, this skin temperature is computed by the surface scheme and based on a tiled approach (ECMWF, 2021). It represents the temperature of the vegetation layer, the top layer of the bare soil, or the top layer of the snow pack for each tile. In addition, it represents the temperature of the water in immediate contact with the atmosphere over lakes and open ocean, or the temperature of the top layer of the sea-ice or lake-ice tile.

Over land, each tile has its own properties which define separate heat and water fluxes used in an energy balance equation. The skin temperature is computed by a linearisation of this surface energy balance equation for each tile. The grid box skin temperature is then obtained by taking the weighted average of the skin temperature over the tiles. Errors on the weight of each tile can lead to biases, which can happen for example where the vegetation changes rapidly at the boundaries of desert regions. As the skin temperature has no heat capacity, it responds instantaneously to changes in its forcing (as the radiative forcing) and it is subject to their errors. Finally, misrepresentation of the surface processes, such as thermal conductivity in the soil, canopy, snow, water, ice, or built-up surfaces and their associated heat storage also contributes to the surface temperature error budget.

The sensitivity to the surface of the infrared instruments or more likely the microwave instruments may be deeper than the layer represented by the model skin temperature. This is also a potential source of bias if the temperature in the soil layers differ from the skin temperature or if there are vertical gradients of temperature in the ocean. We hypothesised that this is the main source of the bias found in the analysis increment and the aim of this document is to focus on it. We concentrate our attention on the background of the α_{mw} -fields as the mean analysis increment was large for them, and not so much for the α_{ir} -fields.

In the next part of this document we try to estimate the background bias and we assess the impact of a bias-corrected background on the analysis increment. Then, we present the results of a simulation with a bias-corrected background before concluding.

2 Towards a bias-free α -skin temperature background

We assume in this document that the model skin temperature does not represent the temperature layers to which the microwave instruments are sensitive to (Karbou *et al.*, 2006; Lv *et al.*, 2018). This mismatch is assumed to be the main cause of bias found in the analysis as the model skin temperature serves as background of the α -skin temperature fields. This assumption is strengthened by the fact that the mean analysis increment is larger for the α -skin temperature fields associated with the microwave instruments than for the ones associated with the infrared instruments. Indeed, the microwave channels sensitive to the surface are overall influenced by a deeper soil layer than the ones of the infrared channels.

The deeper the layer in the soil, the lower the temporal variability of the soil state. As we link the bias to the deeper soil layers, we can expect that it varies on slower time scales than the skin temperature. A

slow time variation means that we can expect to use information from the previous cycle(s) to estimate and remove the bias. We thus decided to use, as a correction, the analysis increment from the previous assimilation cycle valid at the same time in order to respect the diurnal variation of the bias (see appendix A page 18). With our assimilation window of 12 hour, we are then using the information from the last but one assimilation cycle $m - 2$, where m refers to the assimilation cycle number. The corrected background $\tilde{\mathbf{x}}_m^b$ of the α -skin temperature fields is then

$$\tilde{\mathbf{x}}^b = \mathbf{x}_m^b + \delta\mathbf{x}_{m-2}^{a,b}, \quad (1)$$

where \mathbf{x}_m^b is the background (model skin temperature) and $\delta\mathbf{x}^{a,b}$ is the analysis increment (difference between the analysis and the background) of two assimilation cycles prior.

The background bias η^b is added to the background \mathbf{x}^b before computing the model equivalent to the observations with the observation operator H_k at time t_k . This modifies the departure \mathbf{d}_k with the observations \mathbf{y}_k^o ,

$$\mathbf{d}_k = \mathbf{y}_k^o - H_k \left(\mathbf{x}^b + \eta^b \right) \quad (2)$$

This departure is used in various parts of the 4D-Var preparation, like the first-guess check or the variational quality control. For example, the first-guess blacklists the observations for which the associated departure is too large. Changing the background may result in a different observation check and usage, and may affect the number of used observations.

In our incremental 4D-Var formulation, the non-linear cost function is minimised through a series of minimisations of linear versions of the cost function (Courtier *et al.*, 1994). The linearisation of the cost function is performed around the first-guess which is the background for the first linearisation. Changing the background then also affects the tangent linear and adjoint versions of the observation operator.

3 Results

We present here the effect of having a bias correction for the microwave skin temperature fields when using the SKTXCV approach for both microwave and infrared instruments.

3.1 Configuration

All the experiments of this document are using the same configuration as the ones presented in Massart *et al.* (2021). They are based on the IFS cycle CY47R1 and using the incremental formulation of the 4D-Var. The outer loop has a TCo 399 horizontal grid (or spatial resolution of about 25 km) and a vertical grid with 137 vertical levels. The inner loop horizontal grids are successively TL 95 / TL 159 / TL 255 (or spatial resolution of about 200 km, 125 km and 80 km respectively), with the same vertical grid as the outer loop.

We ran the period going from December 2019 to March 2021 and we are using the control and SKTXCV experiments of Massart *et al.* (2021). The control which uses the TOVSCV approach is referred here as TOVSCV. The SKTXCV experiment is referred hereafter to as SKTXCV. The new SKTXCV experiment (referred hereafter to as SKTXCV-BPS for **B**ackground **P**er**S**itence) has the same configuration as SKTXCV except for

- the persistence approach is used for the α_{mw} -background fields only,
- the background error standard deviation is inflated by 2 over the sea-ice for the α_{ir} -fields (instead of 3),

- the background error standard deviation is inflated by 2 over the sea-ice (instead of 3), by 1.4 over the land (instead of 1) for the α_{mw} -fields.

In the following evaluation, we mostly remove the first month as a spin up period, unless we discuss time series. We simply refer to the background bias correction η_k^b of Eq. (7) for the α_{mw} -fields as correction.

3.2 Mean analysis increment

First, we computed the time average of the analysis increment of the α_{mw} -fields from the SKTXCV and the SKTXCV-BPS experiments. The mean values for the SKTXCV experiment are the highest over the land, and predominantly positive in the northern hemisphere and negative in the southern hemisphere (Fig. 1a). In absolute value, they are mostly larger than 0.5 K and locally over 3.5 K. The patterns of the mean analysis increment present little variation between the months (not shown). This emphasises that we are in presence of systematic bias and the persistence approach is relevant.

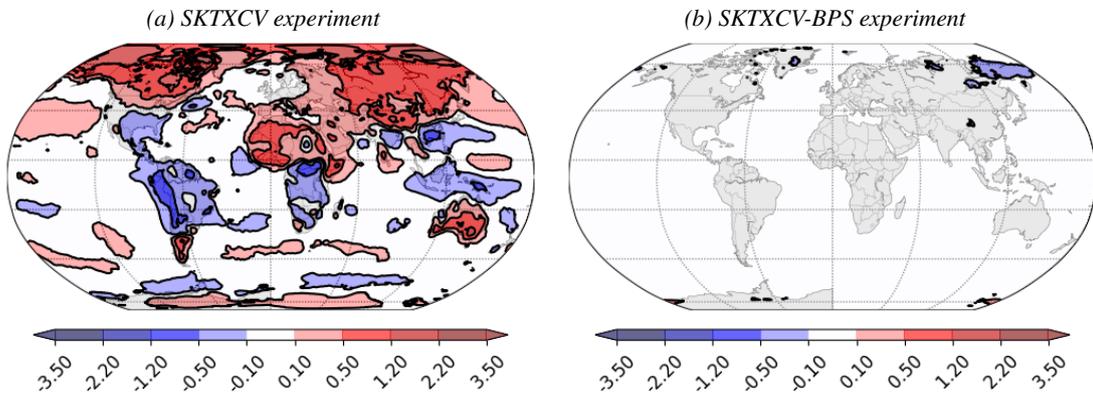


Figure 1: Mean analysis increment for the α_{mw} -fields for the period between 1 January and 31 March 2020: (left) SKTXCV experiment, (right) SKTXCV-BPS experiment.

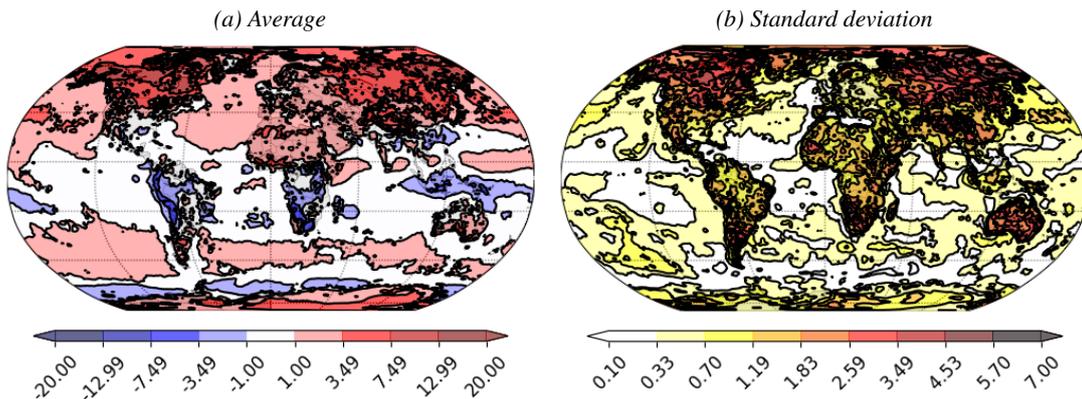


Figure 2: Correction for the α_{mw} -fields for the period between 1 January and 31 March 2020: (left) mean value, (right) standard deviation.

The persistence approach is effective at producing a small mean analysis increment (Fig. 1b). The mean values for the SKTXCV-BPS experiment are indeed in absolute lower than 0.1 K, except over part of Siberia where they are nonetheless lower than 0.5 K. There is some variability between the monthly mean, but it remains low (not shown).

3.3 Bias correction of the α_{mw} -fields

3.3.1 Time average

Figure 2 presents the time average and standard deviation of the correction for the α_{mw} -fields. On average, as expected, the patterns of the correction are similar to the mean analysis increment from the SKTXCV experiment (Fig. 1a). The values are however larger in the correction than in the bias. For example, over Siberia, the mean analysis increment is around 1 K while the mean correction is around 15 K. This can be explained by the fact that the amplitude of the analysis increment is constrained by the background error that are around 1 K. In the meantime, the correction does not have any constrain and can grow from assimilation cycle to assimilation cycle.

3.3.2 Time series

In order to assess if the absence of constraints on the correction could be an issue, we computed the time series of its mean and standard deviation (Fig. 3). Both stabilise during the first month over the land and sea-ice, but not over the ocean. To have a deeper understanding of the time evolution, the mean correction has been decomposed by region for each surface type (Fig. B.8, page 28). Over the land, the correction is positive in the northern extra-tropics and in the south polar region (after a month). It is close to zero in the tropics. The positive correction in the extra-tropics could reflect that the skin temperature is cooler in winter than the soil layers for which the microwave sounders are sensitive. Over the south polar region, the snow can act as an isolator. This makes the skin temperature cooler than the temperature within the snow pack (microwave-sensitive), and would also explain the positive correction.

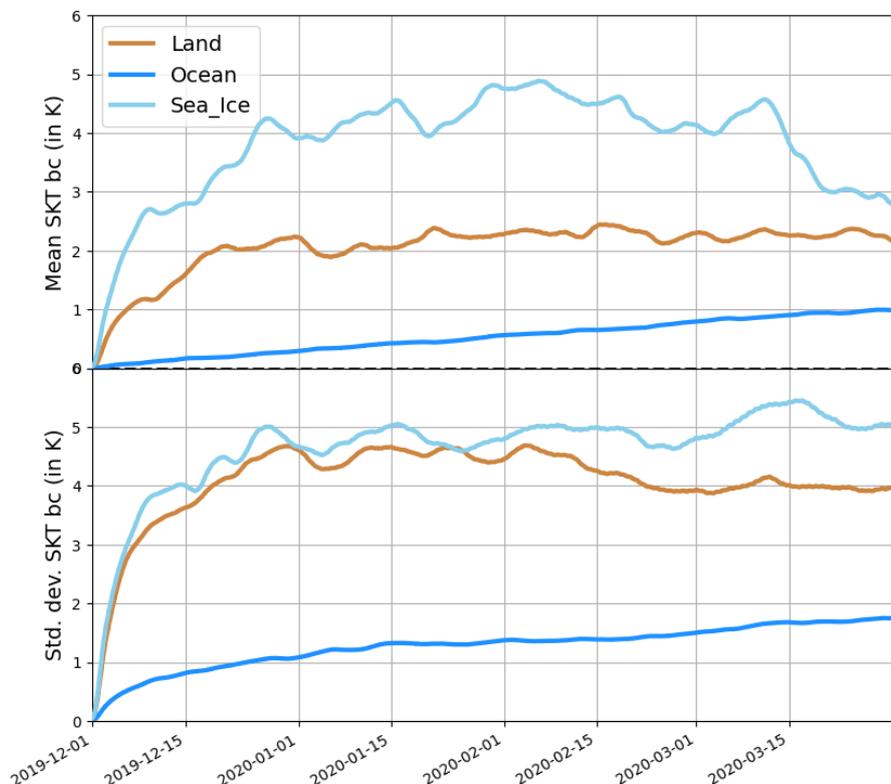


Figure 3: Time series of the mean (top) and standard deviation (bottom) of the α_{mw} correction averaged over land (brown), ocean (dark blue) and sea-ice (light blue).

Over the sea-ice, the correction has an opposite sign between the north and south polar regions. The temperature within the ice tends to increase with the depth. We therefore expect to have a positive correction as the microwave sounders are sensitive to deeper layers than the surface. This is probably what we observe in the north polar region. The profile can change shape during the melting period, making the surface temperature warmer than the layers below. This may explain the negative correction in the south polar region, the experiment spanning the austral summer season during which the ice can melt.

Over the ocean, the correction is small, under 1 K with a variability under 2 K. But, overall the correction does not seem to stabilise, even after four months. The correction is very small over the tropics, slightly larger over the south temperate region (around 1 K), and larger and increasing in time over the north temperate region. There, the correction reaches 3 K at the end of the simulation. The increase is dependent on the definition of the sea-ice mask and part of it comes from the correction over sea-ice which is larger.

3.4 Analysis mean

The modification of the α_{mw} -background fields affects the mean analysis of the SKTXCV-BPS experiment when compared to the SKTXCV experiment (Fig. 4).

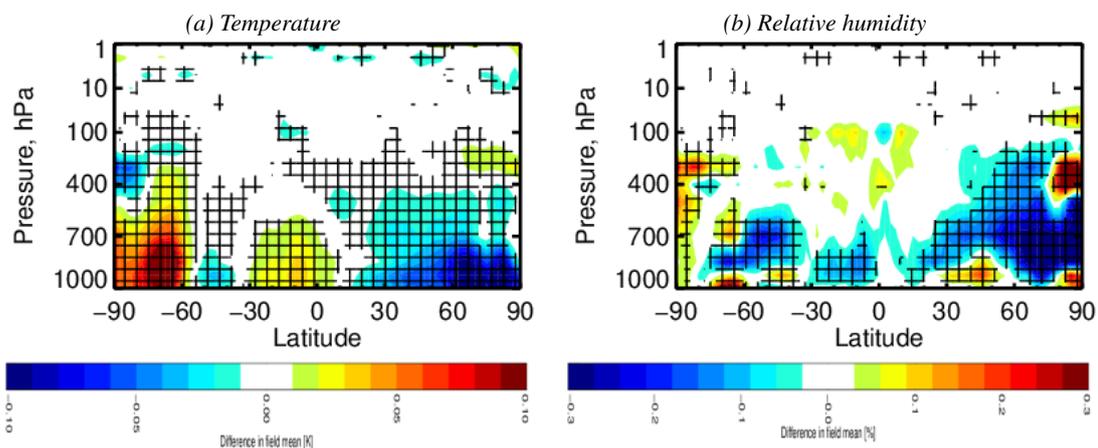


Figure 4: Mean change in the analysis field between the SKTXCV-BPS experiment and the SKTXCV experiment for the period between 1 Jan. to 31 Mar. 2020: (left) temperature, (right) relative humidity.

In terms of temperature analysis, the largest changes occur essentially in the troposphere with a warmer atmosphere over the Antarctic and a cooler atmosphere over the Arctic in the SKTXCV-BPS experiment. There is also a slight warming of the tropics. The changes of mean temperature analysis generate consistent signals in the mean geopotential analysis, but in the whole atmosphere. The geopotential decreases in the north polar region and increases in the south polar region (not shown).

The change in relative humidity analysis is located mainly in the troposphere. There is a reduction almost everywhere, with the largest change over the north polar region. There is an increase in the SKTXCV-BPS experiment over land in Central Asia and over sea-ice near the Antarctic coast close to the surface. Finally, there is a change in the circulation from the middle troposphere upwards in the south polar region. The U-component of wind analysis increases south of about 60°S and decreases north of it (not shown).

The changes in the analysis fields are mostly small compared to the underlying analysis biases in the troposphere and near surface as diagnosed by the mean analysis departure with the used observations

(Figs C.9 to C.12 pages 29 and 30). The change of temperature analysis in the SKTXCV-BPS experiment allows its analysis bias to be closer to zero in the polar regions, up to about 400 hPa, when compared to radiosondes data. It also has a positive impact on the northern hemisphere when compared to GPS-RO data. For the relative humidity, the impact of the change in the analysis is visible only when compared to radiosondes data and in polar regions. The bias is reduced in the SKTXCV-BPS experiment in the Antarctic but increased in the Arctic. For the wind, only the comparison with the radiosondes data shows an effect of the analysis change, and the difference is visible only in the Antarctic and above 400 hPa.

3.5 Observation diagnostics

The change of the α_{mw} -background fields in the SKTXCV-BPS experiment interacts in particular with the first-guess screening, the variational bias correction scheme and the emissivity retrieval. We present hereafter each of these changes focusing on AMSU-A channels 5 and 7, and ATMS channels 6 and 20, and on the period between 1 Jan. to 31 Mar. 2020, removing the first month as a spin-up period. While AMSU-A channel 5 and ATMS channel 6 are sensitive to the surface, the two other channels are less sensitive to the surface and more to the mid to upper troposphere.

3.5.1 Observation count

The change in the first-guess results overall in an increase of the number of used observations for the channels sensitive to the surface in the SKTXCV-BPS experiment when compared to the SKTXCV experiment. For example we have an increase of about 10% in the Northern hemisphere for the channels 5 to 7 of AMSU-A. For ATMS channels 18 to 20, the increase is around 5%, but for channels 6 to 8 there is some variability of the number both in time and space.

The additional data in the SKTXCV-BPS experiment from AMSU-A channels 5 and 7 are located primarily north of $60^{\circ}N$ and south of $60^{\circ}S$ over both land and sea-ice (Figs 5a and 5b). For channel 7 in particular, there is a significant increase of additional data in East Asia and, to a lesser extent, in the middle of central Africa, at the boundary between rainforest and grassland. These are areas where the skin temperature correction is positive and where the emissivity is decreased.

The change in data usage is more dependent on the channel for ATMS, expect south of $60^{\circ}S$ where less data are used in the SKTXCV-BPS experiment for both channels 6 and 20 (Figs 5c and 5d). The main other change are a decrease in data usage for channels 6 in central Asia and an increase for channel 20 in the western part of Europe.

3.5.2 Observation bias correction

It takes between 5 to 10 days for the variational bias correction scheme to adapt to the change of α_{mw} -background (not shown). The difference of the analysed bias correction is then stable between the SKTXCV-BPS experiment and the SKTXCV experiment for the period of interest. For the instruments and channel we focus on, the bias correction tends to be reduced (in absolute value) on average in the SKTXCV-BPS experiment (Fig. 6). The difference between the two experiments is more distinct over land and in the polar regions.

Over the polar regions, the difference between the SKTXCV-BPS and SKTXCV experiments is negative for all channels. Because the bias correction is positive in the experiments for AMSU-A channels 5, it becomes closer to zero in the SKTXCV-BPS experiment. For other channels, it is going further away from zero.

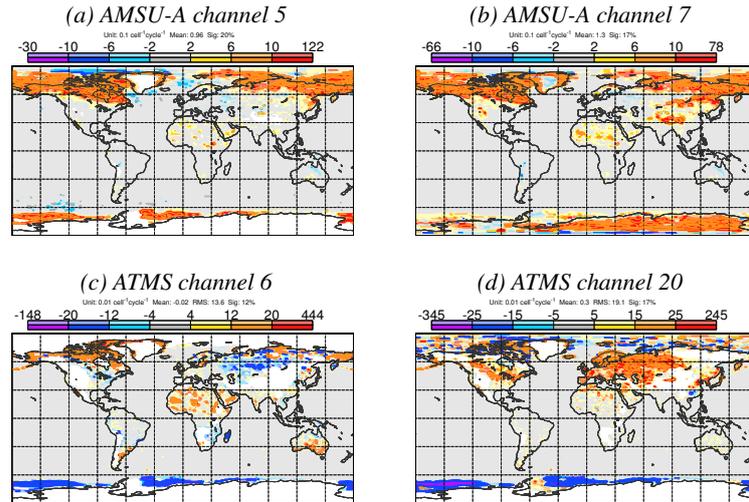


Figure 5: Difference in observation count between the SKTXCV-BPS experiment and the SKTXCV experiment for the period between 1 Jan. to 31 Mar. 2020: (a) AMSU-A channel 5, (b) AMSU-A channel 7, (c) ATMS channel 6 and (d) ATMS channel 20.

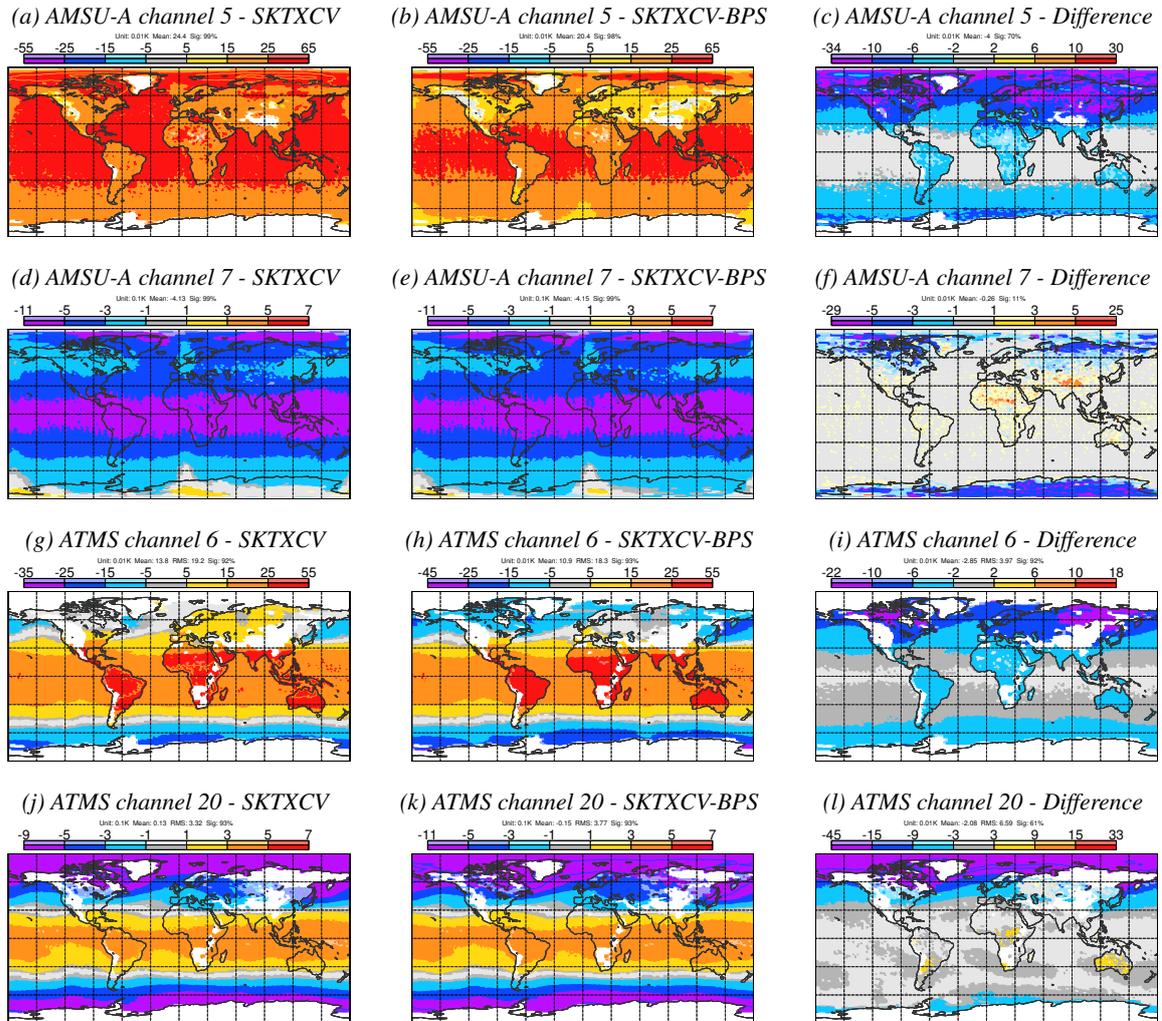


Figure 6: Bias correction for selected instrument and for the period between 1 Jan. to 31 Mar. 2020. From top to bottom: AMSU-A channel 5 and channel 7, ATMS channel 6 and channel 20. From left to right: SKTXCV experiment, SKTXCV-BPS experiment and difference between the SKTXCV-BPS and SKTXCV experiments.

A bias correction getting closer to zero is a good sign and may indicate that the skin temperature correction reduces a wrong surface forcing and moves the system closer to observed radiances. Nonetheless, only verification against independent observations and assessment of the forecast skill will tell whether the change of the variational bias correction is good or bad.

3.5.3 Emissivity

For the microwave observations, the surface emissivity is retrieved from the window channels using the model emissivity and the model skin temperature as background values (Karbou *et al.*, 2006). Changing the skin temperature background thus impacts the retrieved emissivity.

We present here the change in retrieved emissivity for two satellites only: NOAA-18 and NOAA-20. We present only channel 7 of NOAA-18 AMSU-A and channel 6 of NOAA-20 ATMS, as all the channels of a given instrument are using the same emissivity.

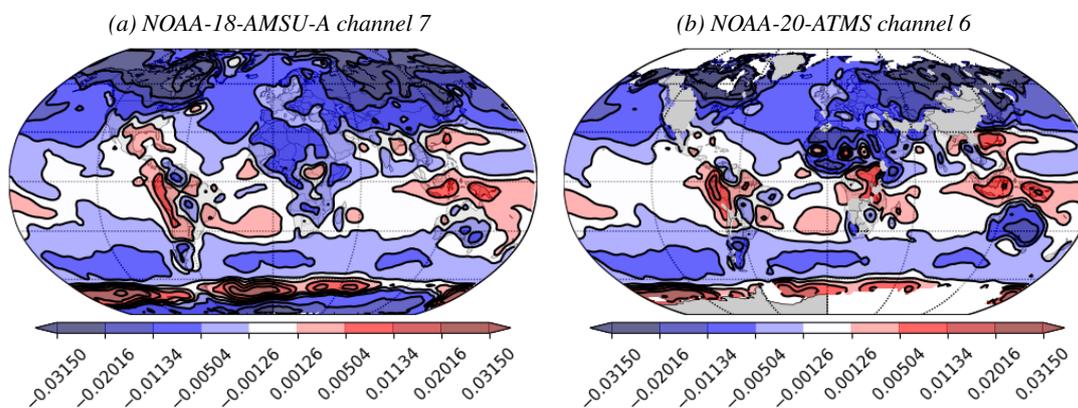


Figure 7: Mean change in emissivity between the SKTXCV-BPS experiment and the SKTXCV experiment for the period between 1 Jan. to 31 Mar. 2020. Left: NOAA-18-AMSU-A channel 7. Right: NOAA-20-ATMS channel 6.

Overall the changes in emissivity are similar for the two instruments (Fig. 7). The changes are strongly correlated with the skin temperature background correction, as expected. For example, in the north polar region, where the skin temperature background is increased in the SKTXCV-BPS experiment, the emissivity decreases. The link between changes in skin temperature background and changes in emissivity is also strong over the African and Australian continents.

3.6 Forecast

3.6.1 Mean first guess departure

We first compare the mean first-guess departure with bias corrected observations from the SKTXCV-BPS and SKTXCV experiments focusing again on the data from AMSU-A channels 5 and 7, and ATMS channels 6 and 20 (Fig. 8). Overall, there is a reduction of the first-guess departure for the SKTXCV-BPS experiment compared to the SKTXCV experiment. The reduction is a combination of a better first-guess and a change in the observation bias correction. The bias correction change is nevertheless small compared to the reduction of the first-guess departure except for ATMS-A channel 20 over the Antarctic. Elsewhere, the reduction suggests a better first-guess in the SKTXCV-BPS experiment than in the SKTXCV experiment.

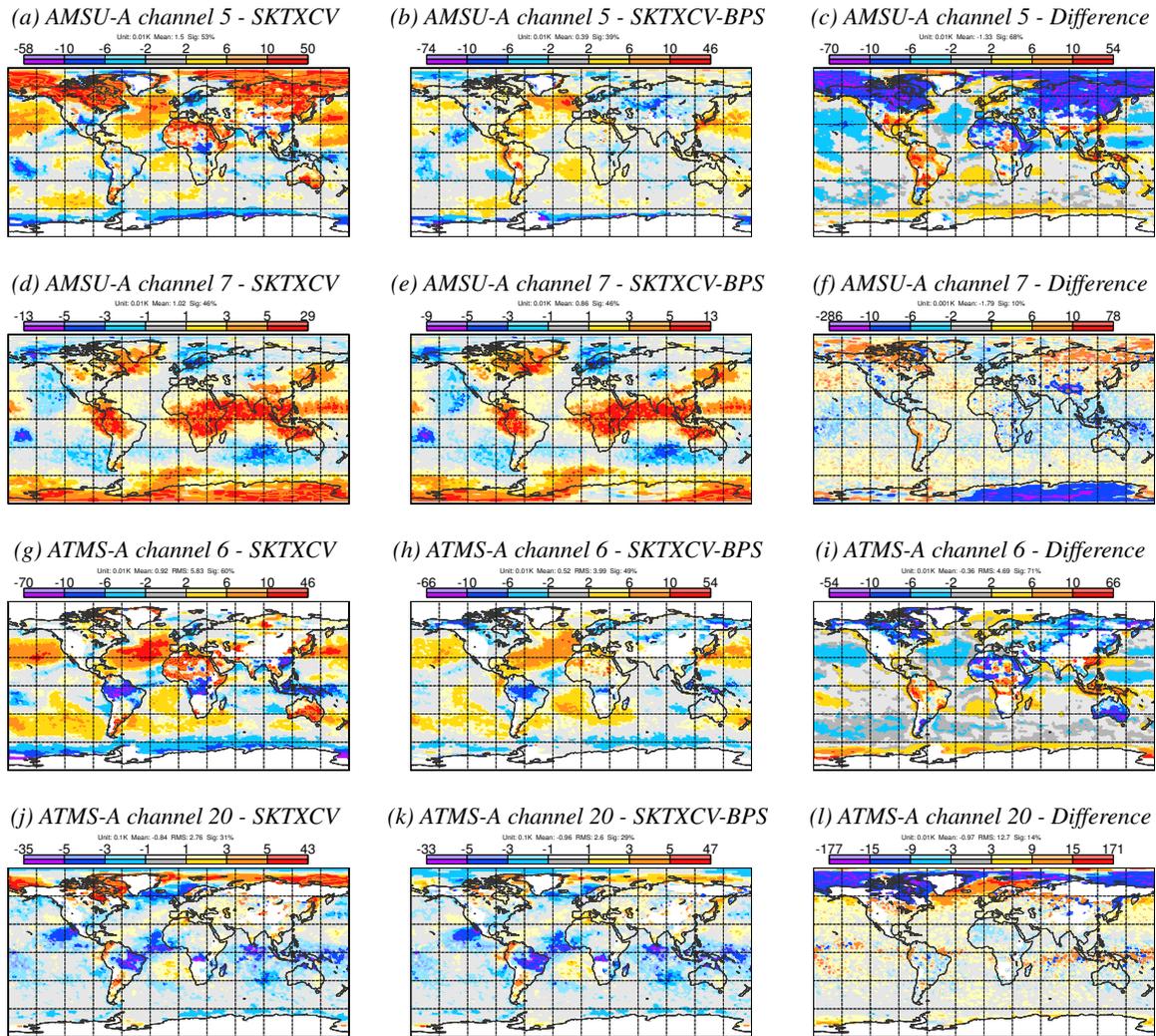


Figure 8: Mean first guess departure with bias corrected observations for selected instrument and for the period between 1 Jan. to 31 Mar. 2020. From top to bottom: AMSU-A channel 5 and channel 7, ATMS channel 6 and channel 20. From left to right: SKTXCV experiment, SKTXCV-BPS experiment and difference between the SKTXCV-BPS and SKTXCV experiments.

The reduction is present over all surfaces (land, ocean and sea-ice) and is more prominent in the Arctic regions. For example, the first-guess departure of the SKTXCV-BPS experiment with AMSU-A channel 5 is close to zero there, while it could reach up to 0.5 K in the SKTXCV experiment. The area of change of the first-guess departure is smaller for ATMS-A channel 20 than for AMSU-A channel 5 because this channel (as all channels between channel 18 and channel 22) is not assimilated over snow. Still, the first-guess departure of the SKTXCV-BPS experiment with ATMS-A channel 20 get closer to zero, compared to the SKTXCV experiment where it could reach up to 4 K.

Elsewhere, there is no much difference in the first-guess departure between the experiments for AMSU-A channel 7 and ATMS-A channel 20. The difference for these channels, mostly located in the Tropics, are most likely not linked with error in the skin temperature.

For AMSU-A channel 5 and ATMS-A channel 6, there are several regions over land with a significant reduction of the first-guess departure in the SKTXCV-BPS experiment: Asia, Africa and Australia. Over South America, the picture is more mixed with regions of increase and reduction.

Over the ocean, the largest difference is located in the Atlantic. The first-guess departure is reduced in the SKTXCV-BPS experiment, but the reduction is not large enough to get a first-guess departure close to zero. In this region, we saw previously that the correction of the α_{mw} -background was continuously increasing and with not sign of stabilisation after 4 months of simulation. As this behaviour allows to reduce the first-guess departure, we can be more confident in it.

When compared to the other data of the observing systems, the mean first guess departure is similar between the SKTXCV-BPS and SKTXCV experiments (not shown). When a significant difference is observed, this difference is always small compared to the mean first guess departure. This means that the SKTXCV-BPS experiment has a low impact on the forecast bias.

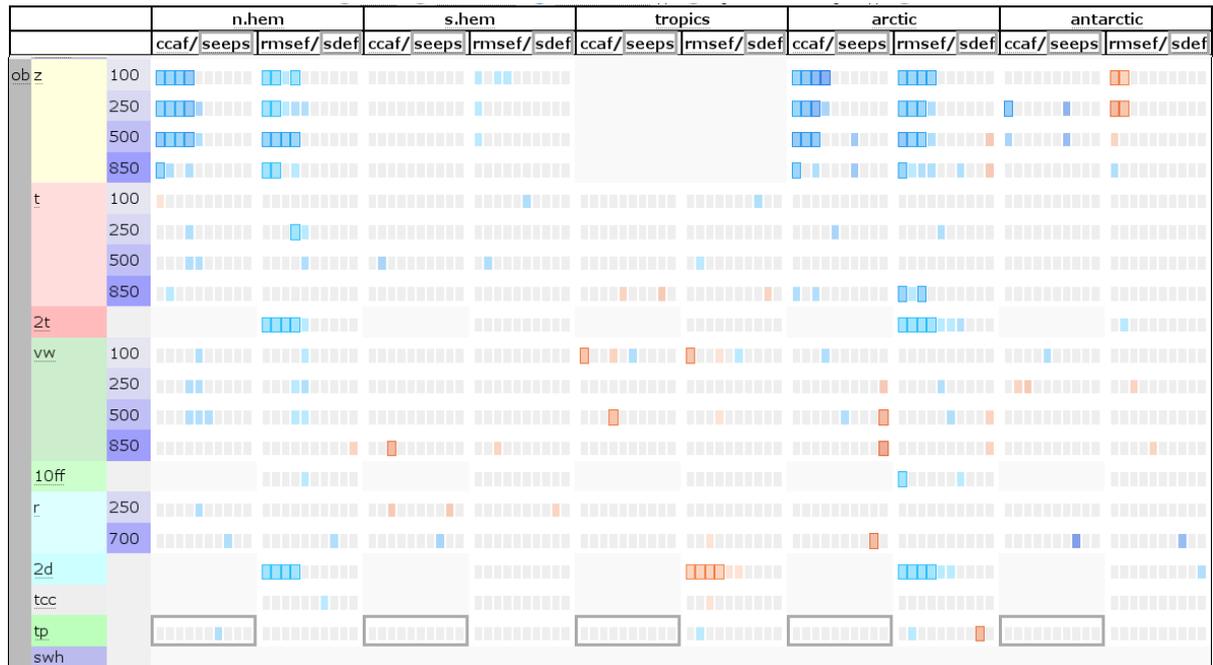


Figure 9: Scorecard of the SKTXCV-BPS experiment versus the SKTXCV experiment, for verification against observations and for the period between 1 Jan. to 31 Mar. 2020. Framed box where the difference is significant with 99.7% confidence, light colour box for 95% confidence, grey where the confidence is lower than 95%.

3.6.2 Forecast scores

We finally computed the scorecard of the experiments with observations: first comparing the SKTXCV-BPS and the SKTXCV experiments and then the SKTXCV-BPS and TOVSCV experiments.

Overall the SKTXCV-BPS experiment improves the quality of the forecast compared to the SKTXCV experiment, and mainly up to day 3 to day 5 with a confidence of three times the standard deviation (Fig.9). The largest improvements are for the geopotential in the Arctic region and to a lesser extent in the northern hemisphere. In these regions, the forecasts of the surface variables (2 m temperature and dew point) are also improved. The humidity and wind forecast are slightly improved.

The two main degradations can be found in the Antarctic region for the geopotential and for the 2 m dew point in the Tropics. For the geopotential, the degradation is on the root mean square error and not the anomaly correlation coefficient, and only for the first few days. This is likely linked to large change in the mean analysis there as discussed before.

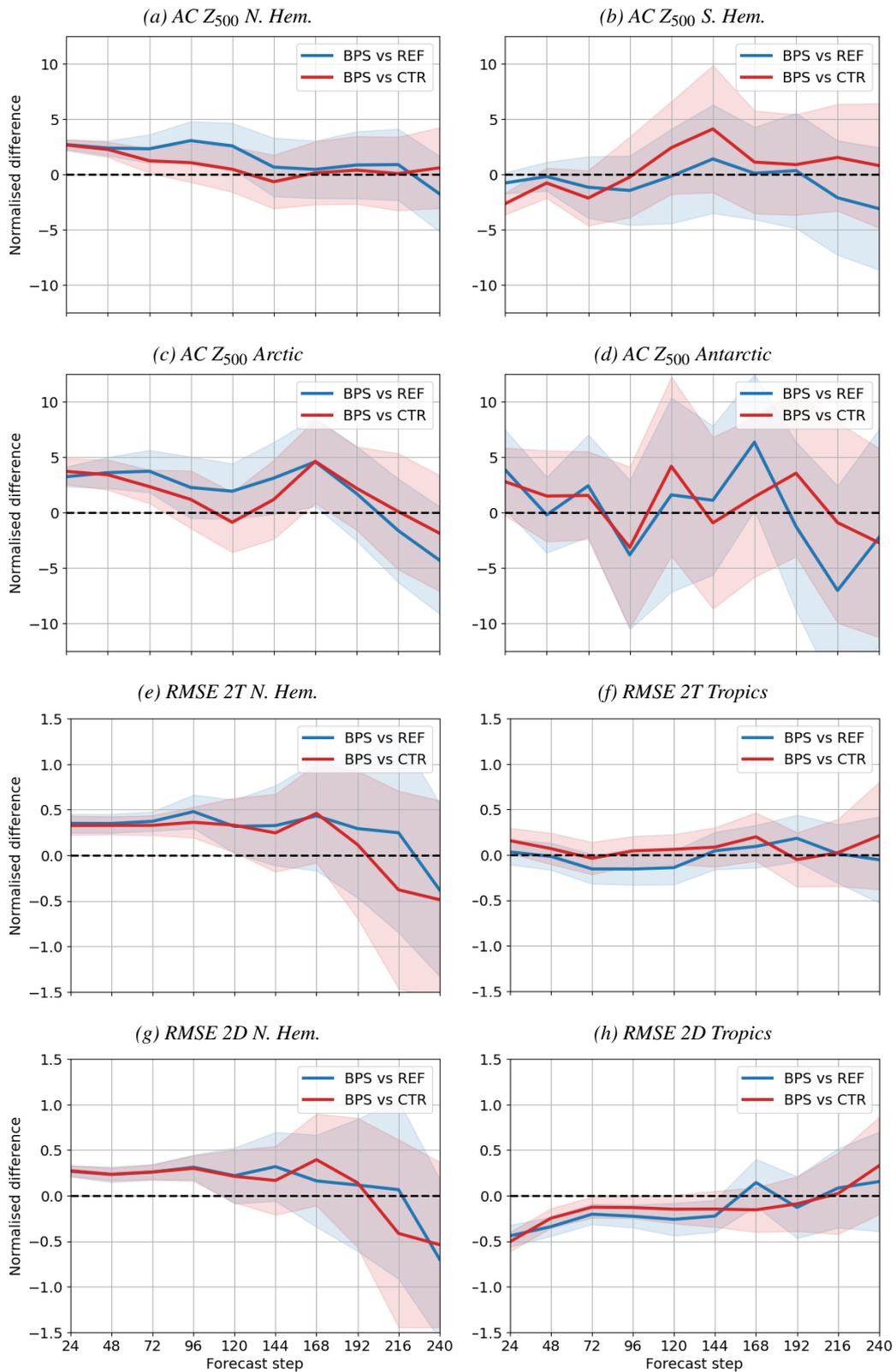


Figure 10: Normalised difference (in %) between the SKTXCV-BPS experiment and (blue curve) the SKTXCV experiment or (red curve) the TOVSCV experiment as a function of forecast lead time, for the anomaly correlation (AC) of geopotential height at 500 hPa (Z₅₀₀) and the root mean square error (RMSE) of 2 m temperature (2T) or 2 m dew point (2D) for various regions (see subfigure title). Note the difference in y-axis between the AC and RMSE statistics.

The scorecard of the SKTXCV-BPS experiment is similar when compared to the TOVSCV experiment instead of the SKTXCV experiment. The difference is that both improvement and degradation seem to last for one or two days less (not shown).

We have selected some of the largest differences in score for the SKTXCV-BPS experiment and plotted their amplitude as a function of the forecast lead time for the comparison with SKTXCV experiment and the TOVSCV experiment (Fig. 10).

The anomaly correlation of the geopotential height at 500 hPa is one of the headline scores as it integrates tropospheric information on the physics and dynamics. It is improved by 3 to 4% both over the Arctic (up to day 7) and the northern hemisphere (up to day 5) when compared to the SKTXCV experiment. When compared to the CRT experiment, the improvement is similar the first couple of days and then becomes not statistically significant after day 4. Over other regions, the changes are not statistically significant except a 2.5% degradation over the southern hemisphere for the 24 h forecast when compared to the CRT experiment (linked to the change in the mean analysis as discussed above).

For the surface parameters, first the root mean square error of the 2 m temperature is improved over the northern hemisphere by around 0.3% up to day 5 when compared to both experiments. The improvement is over the whole hemisphere (Figs. D.13a and D.13b page 31). After day 5, the improvement is not statistically significant. There is statistically no significant changes over the tropics.

The root mean square error of the 2 m dew point is improved in the northern hemisphere and degraded in the tropics when compared to both experiments. In both case, the change is statistically significant up to day 5. The degradation in the tropics comes mainly from west Africa, over the grassland and savanna (Figs. D.13c and D.13d page 31). There, the α_{mw} -background skin temperature was increase and the emissivity decreased. The first-guess departure for AMSU-A channel 5 and ATMS channel 6 was also largely improved.

4 Conclusion

Previous experiments with the SKTXCV approach highlighted a bias in our analysis that we attributed to the skin temperature background used for the microwave data. We proposed here an approach to address this bias and to provide an estimator of it based on the knowledge of the previous cycles (persistence).

One concern of this persistence approach is the lack of constraint on the estimator. There is no safeguard to prevent the correction to diverge. The experiment we ran over a northern hemisphere winter season shows that nonetheless the estimated values of the bias converge in few days to an apparent steady state, with slow variations during the experiment duration.

To prevent the correction to diverge, one could consider using a nudging approach to force the correction to have values close to some prescribed values (per surface type for example). Another promising way would be to use the weak-constrain framework to optimise the background bias in the 4D-Var. This would, by construction, provide a constrain of the amplitude of the correction.

The persistence approach could be further refined in the future. For example we should pay more attention at the surface type and should investigate, for example, if the temperature of uppermost soil layer may be more appropriated for the α_{mw} -fields over land. We could also investigate if the persistence for the α_{ir} -fields could be applied over some surface types only.

The experiment with the persistence approach has an increased number of used observations and changed values of bias correction when compared to a reference SKTXCV experiment. This results in a change of the mean analysis which reduces the analysis biases against observation except in the mid-troposphere

over the Antarctic and tropics, for both temperature and geopotential.

Overall the experiment with the persistence approach improves the first guess departure with the channels sensitive to the surface and also improves the forecast up to day 3 to day 5 (depending on the parameter), but mainly in the Arctic regions and to a lesser extent in the northern hemisphere. The analysis and forecast in the Antarctic region are slightly degraded.

The degradation in the Antarctic region is likely to be associated with a too large (positive) correction of the background over land and maybe to a too large (negative) correction over sea ice. The amplitude of the correction is controlled by the background errors for which the used scaling factor of 1.4 over land may not be relevant in this particular region. Ideally, we should run an EDA with the SKTXCV configuration in parallel and use it to build the flow-dependent background errors.

The assumption of specular reflection is also a potential source of bias over the Antarctic region. This can also artificially impact the amplitude of the correction. It is worth noticing that in the experiment using the background correction, the number of used data is increased for the surface sensitive channels of AMSU-A, but it is decreased for those of ATMS (only over sea-ice as data over snow are not assimilated). This highlights a potential conflict between the various microwave frequencies and in the future we may want to have separate α_{mw} -fields for the various microwave frequencies.

Starting from IFS cycle 48, the AMSU-A data are not longer considered as clear-sky only, but make use of the all-sky observation operator. This observation operator is based on a constant value of skin temperature and does not utilise the TOVSCV approach and therefore the SKTXCV approach is no longer useful for these data. There are some plans to include the skin temperature and the TOVSCV approach for the all-sky observation operator in the near future. When ready, we will integrate the SKTXCV approach into the new all-sky framework and re-evaluate its difference with TOVSCV.

5 Acknowledgements

The author would like to thank Massimo Bonavita, Niels Bormann and Tony McNally for their comments and recommendations during the production of this document.

References

- Bormann, N. (2021). Investigating the use of lambertian reflection in the assimilation of surface-sensitive microwave sounding radiances over snow and sea-ice. *ECMWF Technical Memoranda*, **Technical Memorandum No. 886**, doi:10.21957/hzcsrx8tt.
- Bormann, N., Lawrence, H. and Farnan, J. (2019). Global observing system experiments in the ECMWF assimilation system. *ECMWF Technical Memoranda*, **Technical Memorandum No. 839**, doi:10.21957/sr184iyz, URL <https://www.ecmwf.int/node/18859>.
- Bormann, N., Lupu, C., Geer, A. J., Lawrence, H., Weston, P. and English, S. (2017). Assessment of the forecast impact of surface-sensitive microwave radiances over land and sea-ice. *ECMWF Technical Memoranda*, **Technical Memorandum No. 804**, doi:10.21957/qyh34roht, URL <https://www.ecmwf.int/node/17674>.
- Courtier, P., Thépaut, J.-N. and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J.R. Meteorol. Soc.*, **120**, 1367–1388.

- Dee, D. (2004). Variational bias correction of radiance data in the ECMWF system. In *ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, ECMWF, Shinfield Park, Reading.
- ECMWF (2019). Part I : Observations. *IFS Documentation*, **IFS Documentation CY46R1**.
- ECMWF (2021). Part IV Physical processes. *IFS Documentation*, **IFS Documentation CY47R3**.
- Guedj, S., Karbou, F., Rabier, F. and Bouchard, A. (2010). Toward a Better Modeling of Surface Emissivity to Improve AMSU Data Assimilation Over Antarctica. *IEEE Transactions on Geoscience and Remote Sensing*, **48**(4), 1976–1985, doi:10.1109/TGRS.2009.2036254.
- Karbou, F., Gérard, E. and Rabier, F. (2006). Microwave land emissivity and skin temperature for AMSU-A and -B assimilation over land. *Q.J.R. Meteorol. Soc.*, **132**(620), 2333–2355, doi:10.1256/qj.05.216, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.216>.
- Lv, S., Zeng, Y., Wen, J., Zhao, H. and Su, Z. (2018). Estimation of Penetration Depth from Soil Effective Temperature in Microwave Radiometry. *Remote Sensing*, **10**(4), doi:10.3390/rs10040519, URL <https://www.mdpi.com/2072-4292/10/4/519>.
- Massart, S., Bormann, N., Bonavita, M. and Lupu, C. (2020). Skin temperature analysis for the assimilation of clear-sky satellite radiances. *ECMWF Technical Memoranda*, **Technical Memorandum No. 870**, doi:10.21957/goe0ads8z, URL <https://www.ecmwf.int/node/19723>.
- Massart, S., Bormann, N., Bonavita, M. and Lupu, C. (2021). 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). *Geosci. Model Dev. Discuss.*, **in review**, doi:10.5194/gmd-2021-37.
- Matricardi, M., Chevallier, F., Kelly, G. and Thépaut, J.-N. (2004). An improved general fast radiative transfer model for the assimilation of radiance observations. *Q.J.R. Meteorol. Soc.*, **130**(596), 153–173, doi:10.1256/qj.02.181, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.02.181>.
- McNally, A. (2015). The impact of satellite data on NWP. In *Seminar on Use of Satellite Observations in Numerical Weather Prediction, 8-12 September 2014*, ECMWF, Shinfield Park, Reading, URL <https://www.ecmwf.int/node/11061>.
- Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N. and Lupu, C. (2018). An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci. Model Dev.*, **11**(7), 2717–2737, doi:10.5194/gmd-11-2717-2018.
- Talagrand, O. (1999). A posteriori evaluation and verification of analysis and assimilation algorithms. In *Workshop on Diagnosis of Data Assimilation Systems, 2 - 4 November 1998*, pp. 17–28, ECMWF, Shinfield Park, Reading.

Appendix A Skin temperature background correction design

We present in this appendix two methods to estimate the background bias and we assess the impact of a bias-corrected background on the analysis increment. The first method consists on using the correction of the skin temperature performs the previous day by the analysis. The second method is based on machine learning techniques.

A.1 Theory

The analysis increment $\delta\mathbf{x}^{a,b}$ is computed as the difference between the analysis \mathbf{x}^a and the background \mathbf{x}^b ,

$$\delta\mathbf{x}^{a,b} = \mathbf{x}^a - \mathbf{x}^b. \quad (3)$$

We want an unbiased background $\tilde{\mathbf{x}}^b$ that leads to an unbiased analysis \mathbf{x}^a . This means that their mathematical expectation should be equal to the true state \mathbf{x}^t ,

$$\mathbb{E}[\tilde{\mathbf{x}}^b] = \mathbb{E}[\mathbf{x}^a] = \mathbf{x}^t. \quad (4)$$

Let us introduce the background bias η^b , as a correction to the background \mathbf{x}^b to get an unbiased background $\tilde{\mathbf{x}}^b$,

$$\tilde{\mathbf{x}}^b = \mathbf{x}^b + \eta^b. \quad (5)$$

The background \mathbf{x}^b contains the skin temperature at various times of the assimilation window which is specific to the SKTXCV implementation (Massart *et al.*, 2021). Therefore, the vector η^b contains the bias at the same times within the assimilation window. Using Eqs.(3) to (5), the mathematical expectation of the analysis increment is

$$\mathbb{E}[\delta\mathbf{x}^{a,b}] = \eta^b, \quad (6)$$

where we assume that the background bias is not a probabilistic variable.

Expression (6) assumes that the analysis is unbiased. The underlying assumption is that other components of the assimilation system, i.e. the observations and the observation operators, are unbiased or at least their biases are negligible compared to the bias of the background.

We would like bias-free backgrounds for the α_{ir} and α_{mw} skin temperature fields. Our purpose is to try to predict the background bias η^b .

A.2 Bias model based on persistence

We assume here that the background of the α -skin temperature fields is biased due to the difference in sensitivity to the soil layers under the surface between the model skin temperature and the IR and MW instruments. This assumption is suggested by the fact that the mean analysis increment is larger for the MW fields than for the IR fields. Indeed, the microwave channels sensitive to the surface are overall influenced by a deeper soil layer than the ones of the infrared channels.

The deeper the layer in the soil, the lower the temporal variability of the soil state. As we link the bias to the deeper soil layers, we can expect that it varies on slower time scales than the skin temperature. A slow time variation means that we can expect to use information from the previous cycle(s) to estimate and remove the bias.

If m refers to the assimilation cycle number and if we assume that the bias is constant over the two cycles $m - 1$ and m , $\eta_m^b = \eta_{m-1}^b$, then we can derive from Eq (6) the bias for the current cycle $\eta_m^b = E[\delta \mathbf{x}_{m-1}^{a,b}]$. This is not applicable due to the mathematical expectation. As a first approximation, we replaced it assuming a single realisation.

In order to take into account the diurnal variation of the bias, we decided to use the analysis increment from the previous cycle valid at the same assimilation time. With our assimilation window of 12 hour, we are then using the information from the last but one assimilation cycle $m - 2$,

$$\eta_m^b = \delta \mathbf{x}_{m-2}^{a,b}, \quad (7)$$

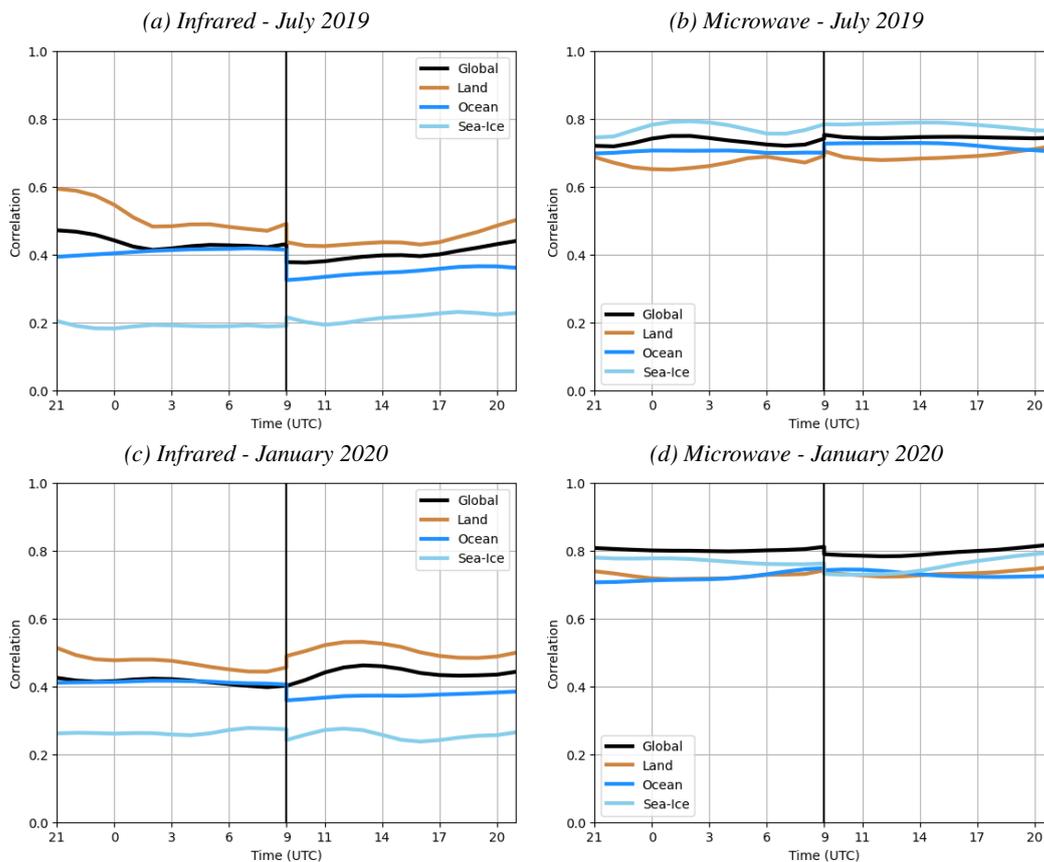


Figure A.1: Pearson correlation coefficient between the analysis increment of the α -skin temperature field from one assimilation cycle (m) and from the previous assimilation cycle valid at the same time ($m - 2$): (left) microwave, (right) infrared, (top): July 2019, (bottom): January 2020. The correlation is computed as a function of the analysis increment time and for all grid points (black), for the land grid points (orange) and for the seaice grid points (blue).

To assess the validity of Eq. (7) hypothesis, we computed the Pearson correlation coefficient between the analyses increments at time m and time $m - 2$ using the experiments of Massart *et al.* (2021) for a summer month and for a winter month (Fig. A.1). There are some differences between the two months, but the main conclusions are valid for both.

For the infrared fields, the correlation coefficient is around 0.4 (Figs. A.1a and A.1c). The values are slightly higher over land and up to 0.6. Still, these values are too low to assume that the analysis increment of the previous cycle (valid at the same time) is a skillful predictor of the the current analysis increment and thus the background bias.

For the microwave fields, the correlation coefficient is around 0.8 (Figs. A.1b and A.1d). The values vary depending on the surface type or season, but overall they remain mostly above 0.7. We can therefore assume that our hypothesis of using the previous analysis increment as a background value for the bias is valid for the microwave data. The hypothesis is also valid for all surface types. Instead of using the analysis increment from the previous cycle only, we additionally tried to use an average of the analysis increments over P cycles,

$$\eta_m^b = \frac{1}{P} \sum_{p=1}^P \delta \mathbf{x}_{m-2p}^{a,b}. \quad (8)$$

We did not see any significant improvement of the correlation coefficient while increasing P up to 3 days (not shown). A degradation appeared for $P = 4$ and more. Since the benefits are small, we use here only $P = 1$.

A.3 Bias model from machine learning techniques

Instead of using the analysis increment of the previous cycle as a proxy for the background bias, we want to assess here if we can predict the analysis increment $\delta \mathbf{x}_m^{a,b}$ of a given cycle m using machine learning techniques and then use this prediction as the background bias η_m^b .

A.3.1 Predictors

We discussed previously that the bias may be linked to the soil layers that are not accounted for in the skin temperature. If this assumption is valid, we may expect a dependency of the bias with the soil state in these layers. Therefore, we assume in this section that the background bias is dependent on the soil state, and we will use this to predict the background bias.

The first step is to select the model variables that we will use to predict the background bias. We selected

- the analysis increment of the previous cycle valid at the same time ($\delta \mathbf{x}_{m-2}^{a,b}$),
- the forecast model skin temperature (\mathbf{x}_m^b),
- the forecast model soil temperature for the first three soil layers (T_1^{soil} to T_3^{soil}),
- water content for the first three soil layers (W_1 to W_3) and
- ice temperature for the first three soil layers (T_1^{ice} , T_3^{ice}), and
- the model albedo.

For each model variable, we selected the 3-hourly fields of the current cycle m , leading to 4 fields per cycle.

The next step is to compute the cross-correlation between the above mentioned predictors. The cross-correlation matrices show some strong correlation between the predictors of the same family (temperature or water content) but for the various layers (Fig. A.2). These global values could be misleading as there are some strong spatial variations in the cross-correlation values (Fig. A.3). This means that we should be careful using all the soil layers of the same family as predictors. Instead, we chose here to use only one layer at a time.

The skin temperature is strongly correlated with the soil temperature at all levels as expected from the skin temperature model. We should then also avoid using the skin temperature and the soil layers as predictors together.

As already determined in the previous section, we also find here that the background bias for the infrared fields is weakly correlated with the analysis increment of the previous cycle (valid at the same time) with a value of 0.47 (Fig. A.2a). The values of the other correlation are low, mostly under 0.2. It is therefore unlikely that we will be able to get a good prediction of the background bias for the infrared fields.

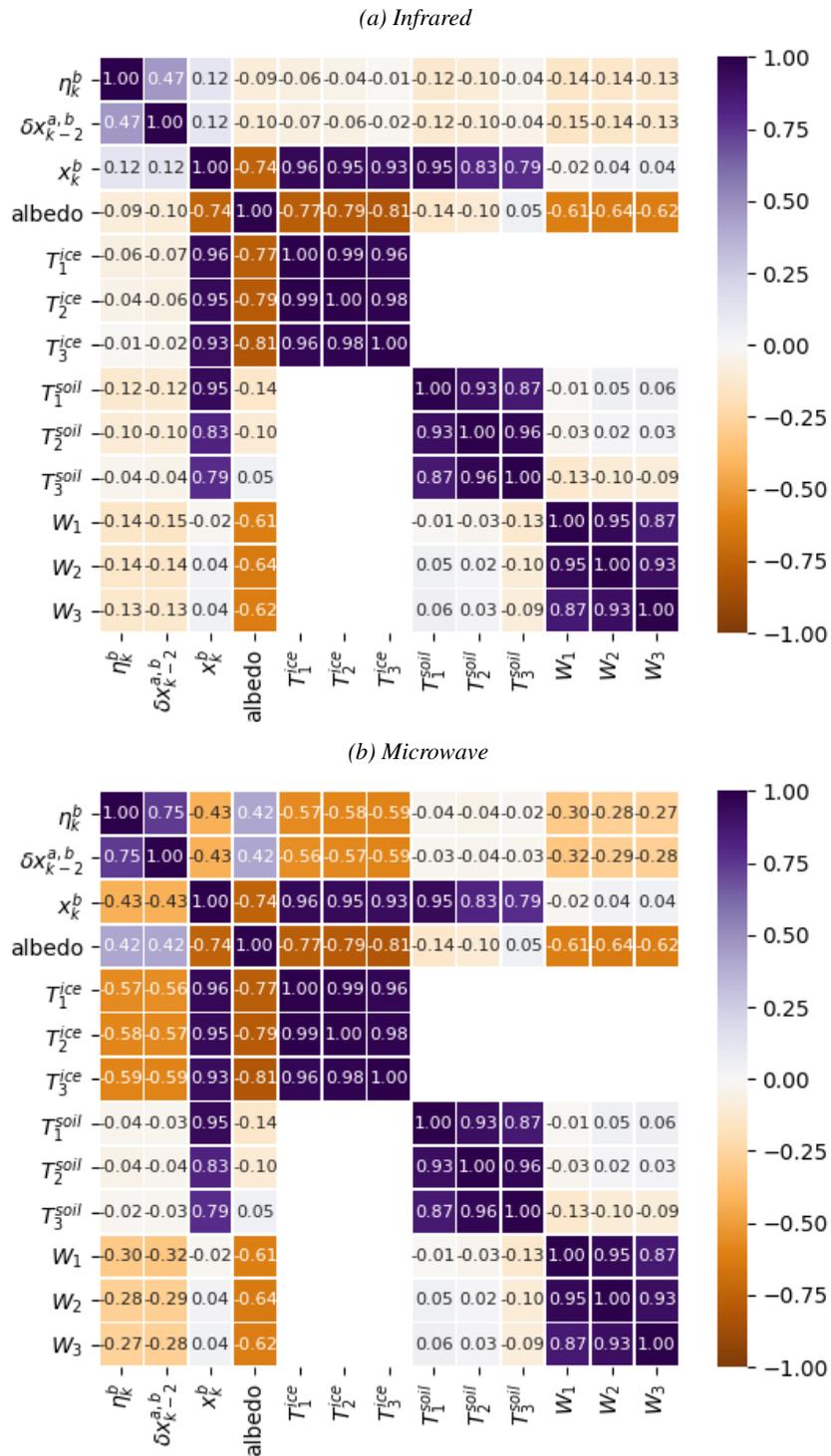


Figure A.2: Pearson correlation matrix between the analysis increment ($\delta x_m^{a,b}$) and the predictors, computed for January 2020. The predictors are the analysis increment of the previous analysis valid at the same time ($\delta x_{m-2}^{a,b}$), the background skin temperature from the model (skt), the model albedo, the ice temperature layers (T_1^{ice} to T_3^{ice}), the soil temperature layers (T_1^{soil} to T_3^{soil}) and the volumetric soil water layers (W_1 to W_3).

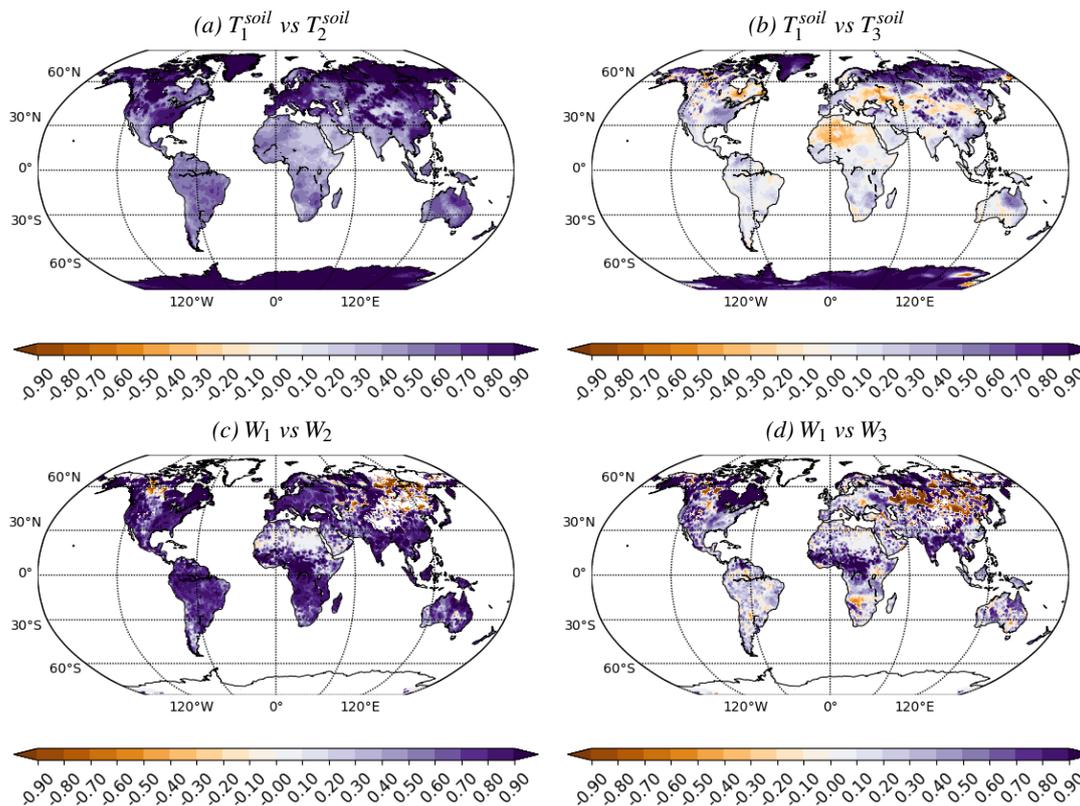


Figure A.3: Pearson correlation coefficient between the first and second (left) and the first and third (right) soil layers for the the soil temperature (top) and soil humidity (bottom) over land, computed for January 2020.

For the background bias of the microwave fields, we have overall larger correlation of around 0.4 in absolute value for most predictors. The exception is with the water content for which the correlation is close to zero. Note that the correlation with the water content was larger for the infrared fields. Contrary to the infrared case, it is likely that we will be able to get a good prediction of the microwave background bias through statistical regressions.

A.3.2 Linear regression

Given a set of predictors, we computed the best linear regression that fits a set of background biases. Note that we also tried other linear regression with up to 3 layers. We found that the additional complexity did not bring much improvement compared to the linear regression and therefore we did not use them in this document.

We selected the analysis increments of [Massart *et al.* \(2021\)](#) as the background biases. The learning period spans through July and August 2019. To be able to fit all the required fields in the used computer memory, we were restricted to use 3-hourly increments instead of hourly increments. We tried to use the latitude and longitude as predictors but the impact was neutral. This indicates that the bias is dependent on the soil properties and not on its location.

As we did not use the geo-location information, each grid point can be considered as an independent sample. The size of the sample is of about $350 \cdot 10^6$. We kept 10% of the sample out of the learning process to test the linear regression and we computed statistics on this test set.

To find the best set of predictors, we ran a set of linear regressions with various configurations. The basic configuration includes only the analysis increment from the previous cycle and the model skin temperature. The next configuration include the land-sea mask (lsm) and the sea-ice concentration (ci). This allows the algorithm to disentangle the various surface types as we saw previously that the background bias is dependent of them. Another configuration replaces the model skin temperature by the soil properties from the first layer. The albedo is also added to created some other configurations.

Table 1: Statistics on a test set from the linear regression to predict the background bias for several predictors. MAE is the mean absolute error. MSE is the mean squared error. R^2 is the coefficient of determination.

| Predictors | Infrared | | | Microwave | | |
|---|----------|--------|--------|-----------|--------|--------|
| | MAE | MSE | R^2 | MAE | MSE | R^2 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, skt | 0.1471 | 0.0527 | 0.2111 | 0.1964 | 0.1194 | 0.5505 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, skt, lsm, ci | 0.1472 | 0.0525 | 0.2139 | 0.1949 | 0.1173 | 0.5584 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, skt, lsm, ci, albedo | 0.1470 | 0.0524 | 0.2151 | 0.1947 | 0.1172 | 0.5589 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, T_1^{soil} , T_1^{ice} , W_1 , lsm, ci | 0.1472 | 0.0523 | 0.2178 | 0.1936 | 0.1174 | 0.5578 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, T_1^{soil} , T_1^{ice} , W_1 , lsm, ci, albedo | 0.1470 | 0.0522 | 0.2184 | 0.1934 | 0.1170 | 0.5596 |

Table 1 summarises the results from all the above configurations. First, independently of the configuration, the coefficient of determination R^2 is just above 0.2 for the infrared background bias. This means that, as expected, the ability to predict the infrared background bias is low, at least with the predictors we retained.

The ability to predict the microwave background bias is higher with a coefficient of determination R^2 above 0.55. Adding information on the surface type through the land-sea mask and the sea-ice concentration helps to improve the statistics. Adding the albedo also brings improvements. Lastly, the configurations with the soil layers instead of the model skin temperature performs better. This is likely linked with the memory of these layers where the time variability is slower than the one of the skin temperature.

Table 2: Same as Tab. 1 but for the microwave analysis increments only and split per surface type and with different predictors.

| Predictors | MAE | | | MSE | | | R^2 | | |
|--|-------|-------|---------|-------|-------|---------|-------|-------|---------|
| | Land | Sea | Sea-Ice | Land | Sea | Sea-Ice | Land | Sea | Sea-Ice |
| $\delta\mathbf{x}_{m-2}^{a,b}$, \mathbf{x}_m^b | 0.335 | 0.116 | 0.573 | 0.257 | 0.028 | 0.737 | 0.498 | 0.320 | 0.549 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, \mathbf{x}_m^b , albedo | 0.335 | - | 0.574 | 0.257 | - | 0.733 | 0.498 | - | 0.551 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, layer 1, albedo | 0.336 | 0.116 | 0.574 | 0.259 | 0.028 | 0.732 | 0.494 | 0.320 | 0.552 |
| $\delta\mathbf{x}_{m-2}^{a,b}$, layer 2, albedo | 0.336 | 0.116 | 0.574 | 0.259 | 0.028 | 0.732 | 0.494 | 0.320 | 0.552 |

For the microwave background bias, instead of using the land-sea mask and the sea-ice concentration as predictors, we separated the data points in three categories depending if the surface was land, sea or sea-ice. Then we reran the linear regression for each of these three categories separately. We also added a configuration using the layer 2 of the soil state. Table 2 summarises the results of the statistics on the test sample.

Over the sea, the mean absolute and mean squared errors of the test dataset are low compared to over the other surface. The choice of the predictors does not impact the prediction and with a value of 0.32, the coefficient of determination is lower than the one obtained with all the surface together (value around 0.55). The coefficient of determination is larger over the land (around 0.50) and over the sea-ice (around 0.55). Over the land, the second best predictor after the analysis increment of the previous cycle is the skin

temperature. Over the sea-ice, it is best to use the ice temperature from the first or second layers than the skin temperature.

These results show that it is difficult to find a set of predictors that suits all surfaces. In the rest of the document we will keep the configuration with the predictors from the first model layer and adding the albedo.

A.4 Offline tests

To assess the potential of the two types of approaches (persistence and linear regression), we tested them off-line. We are using here again the experiments from [Massart et al. \(2021\)](#) and the same two periods: June to September 2019 and January to March 2020.

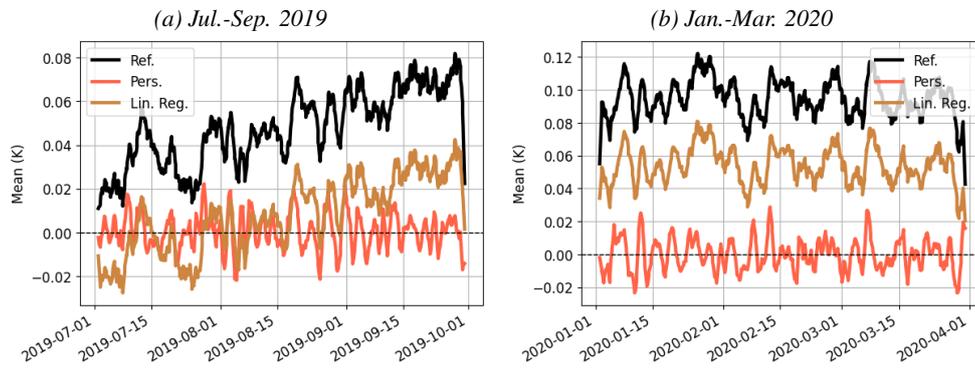


Figure A.4: Time series of the global mean of the α_{mw} -analysis increment (black) and the analysis minus the offline predicted background bias: persistence (red) and linear regression (orange). Left: summer period. Right: winter period. The training for the linear regression spans the period of June and July 2019.

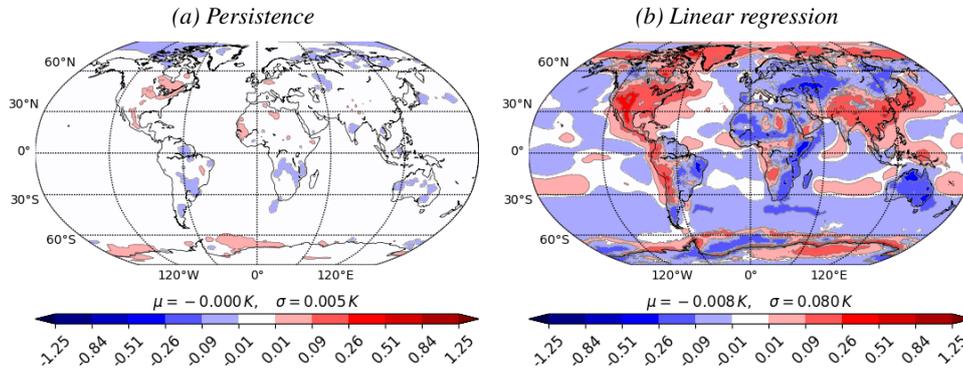


Figure A.5: Time average of the microwave α -skin temperature analysis increment minus the offline predicted background bias from (a) the persistence and (b) the linear regression for Jul.-Sep. 2019. The training for the linear regression spans the period of June and July 2019.

The learning period for the linear regression is the the first two months of the first period (June and July 2019). This means the regression was not trained on the last two months of the first period and not at all for the second season (January to March 2020).

The predicted analysis increment is assumed to be the background bias η_m^b . We compute its difference with the analysis increment obtained through the 4D-Var analysis $\delta x_m^{a,b}$. We expect the spatial and

temporal mean to be close to zero.

Figure A.4a presents the spatial mean for each cycle and compare it with the background bias used as a reference. For the learning period (June and July 2019), both approaches provide a background bias closer to zero than the reference. This result persists for the following month (August 2019), but the linear regression is slightly worse.

Figure A.5 presents the temporal mean of the difference between the predicted background error and the target (Fig. A.5). The persistence approach produces an almost bias-free background with little spatial variability. In comparison, the linear regression also produces a background bias with a global mean close to zero. but with large regional values, up to 0.5 K or higher.

When we change the season with the one for which the linear regression was not trained, it does not predict an accurate background bias. The mean difference with the analysis departure is large (Fig. A.4b). This means that the training period is important for the linear regression. In comparison, the persistence approach is as good for this season as for the previous one.

A.5 Online tests

The next step consisted of testing the two approaches while computing online the background bias and applying it to build the new background used in the next assimilation cycle. We ran two experiments based on the SKTXCV experiment of Massart *et al.* (2021) for the northern hemisphere summer. In the first one, we applied the persistence using the analysis increment of 24 hours prior the analysis time to bias-correct the background. In the second experiment we applied a bias correction based on a linear regression derived from the analysis increment of 24 hours prior the analysis time, the model skin temperature, the albedo, the land-sea mask and the sea-ice concentration.

The learning period for the linear regression is the same as the simulation period, i.e. July and August 2019. In practice, this would not be possible in operations as the learning would require a past period. On the other hand, this is the best possible configuration for the linear regression.

The time series of the global mean of the analysis increments shows, as expected, that on average it is lower for the bias-corrected experiments than for the reference from Massart *et al.* (2021) (Fig. A.6). There is no clear evidence which approach performs best between the persistence and the linear regression. For example the persistence experiment seems to perform better for the period between the 10th and 20th and to perform worse for the period between the 20th and 25th.

When we compute the time average of the analysis increments, it is globally almost everywhere under 0.05 K in absolute value for the persistence approach (Fig. A.7). It is slightly higher in the Antarctic region but under 0.35 K. For the linear regression approach, there are strong regional differences and the mean can reach more than 1.25 K over large patches of land.

A.6 Choice of background bias model

The remaining large values of mean α_{mw} skin temperature analysis increment locally in the experiment with the linear regression approach are concerning. This happens even if the linear regression was trained on the same period of the experiment. We expect a worse background for another period as suggested with the off-line experiment. Our measure of success is based on the mean error while, by construction, the linear regression aims at reducing the root mean square error (L2 norm). For all the above reasons, we conclude that the persistence approach outperforms the linear regression approach.

On the other hand, the computation of the skin temperature background bias using machine learning

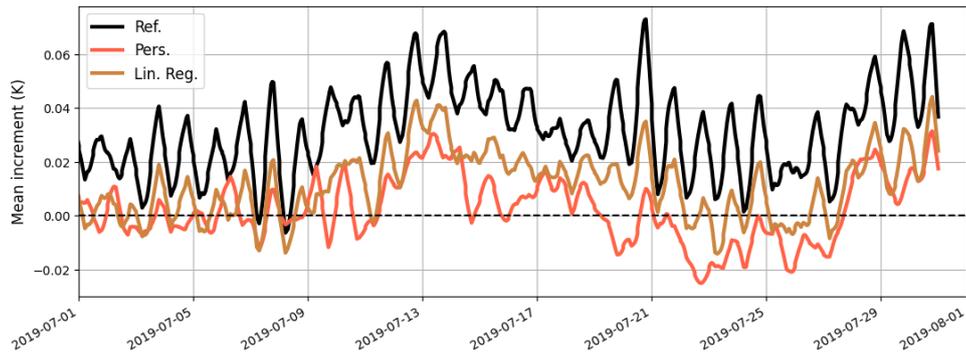


Figure A.6: Same as Fig. A.4 but for the online experiment and for July and August 2019 only. The training period for the linear regression is the same as the simulation.

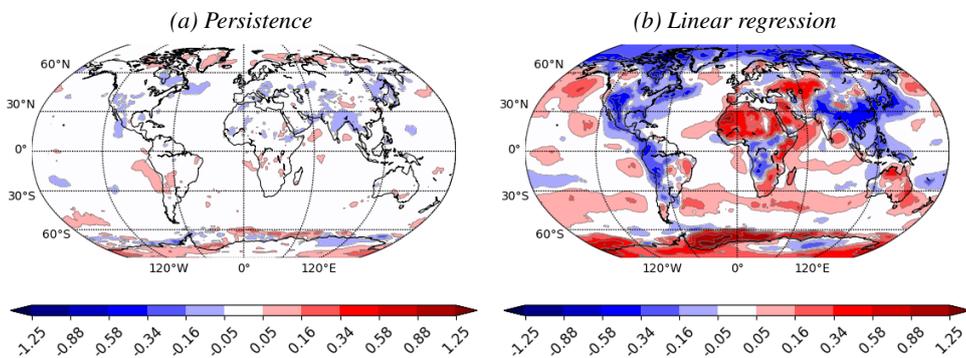


Figure A.7: Same as Fig. A.5 but for the online experiment and for July and August 2019.

techniques can be further improved. For example, using a technique aiming at reducing the LASSO (least absolute shrinkage and selection operator). More advanced techniques than a linear regression could also be used, and the training period could last several seasons.

Nonetheless, one should acknowledge that implementing a machine learning approach in operations would present some additional difficulties. For example, a change in the skin temperature model which would affect the skin temperature background, would require to retrain the algorithm which could be computationally expensive (depending on the length of the training period). The persistence approach would adjust to this change in few assimilation cycles at no additional cost.

A.7 Conclusion

Previous experiments with the SKTXCV approach highlight a potential bias in the background of the skin temperature extended control vector field associated with the microwave data. We proposed two approaches in this document to provide an estimator of the background bias: one based on the knowledge of the previous cycles (persistence) and one based on machine learning techniques (linear regression).

Despite trying various configurations of the machine learning techniques, we could not outperform the persistence approach. This may be due to our basic utilisation of the linear regression which provides an ordinary least squares linear regression, while we want an estimator that produces a background with the lowest possible bias. This may also be due to our choice of parameters for the training. In addition, we showed that the training period was important for the linear regression.

We concluded that the persistence approach is currently a better approach. One concern of this approach is the lack of constraint on the estimator of the background bias. There is no safeguard to prevent the correction to diverge. The persistence approach could be further refined in the future. For example we should pay more attention at the surface type and see if we can use the persistence for the α_{ij} -fields only over some surface types. Another promising way would be to use the weak-constrain framework to optimise the background bias in the 4DVar. This would, by construction, provide a constrain of the amplitude of the correction.

Appendix B Mean correction per surface type and region

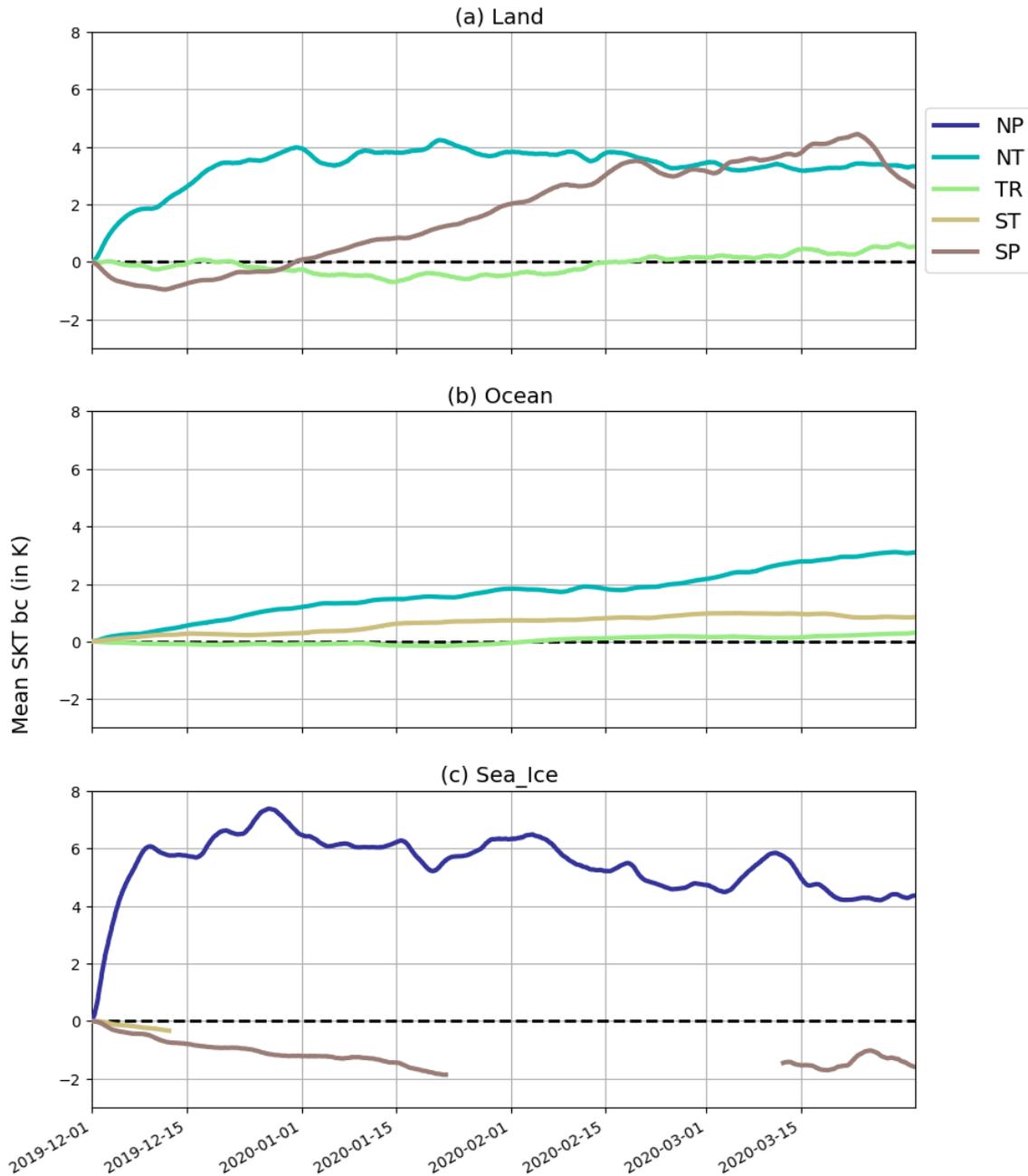


Figure B.8: Time series of the mean correction average over (top) land, (middle) ocean and (bottom) sea-ice. For each surface type, the mean is computed for the North Polar region (NP, $lat. > 66.5^{\circ}N$), North Temperate (NT, $23.5^{\circ}N < lat. < 66.5^{\circ}N$), Tropics (TR, $23.5^{\circ}S < lat. < 23.5^{\circ}N$), South Temperate (ST, $66.5^{\circ}S < lat. < 23.5^{\circ}S$), South Polar (SP, $lat. < 66.5^{\circ}S$). The mean is displayed only when the number of grid point for the selected region and surface is greater than 10% of the total grid point for that surface.

Appendix C Mean analysis departure

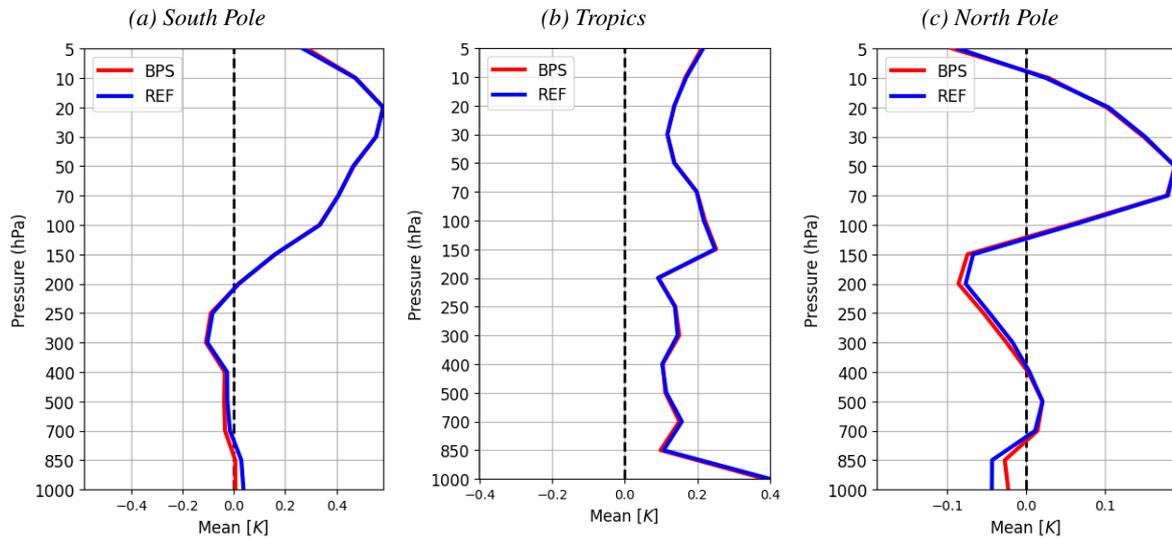


Figure C.9: Mean analysis departure (in K) with the temperature radiosondes for the SKTXCV-BPS (red) and SKTXCV (blue) experiments for the period between 1 Jan. to 31 Mar. 2020 and for 3 regions.

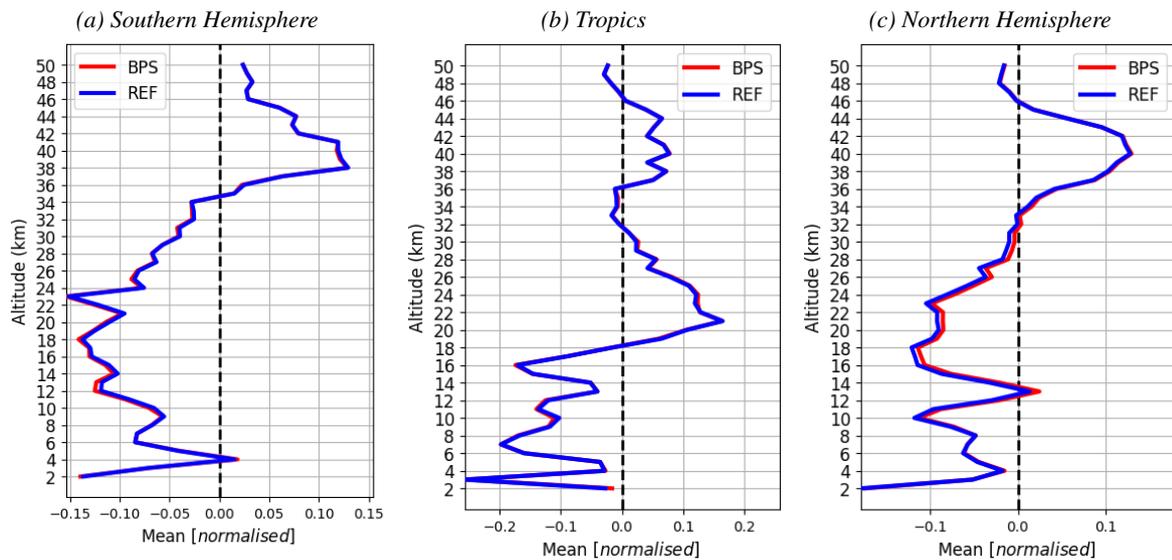


Figure C.10: Same as Fig. C.9 but for the departure with the GPS-RO data.

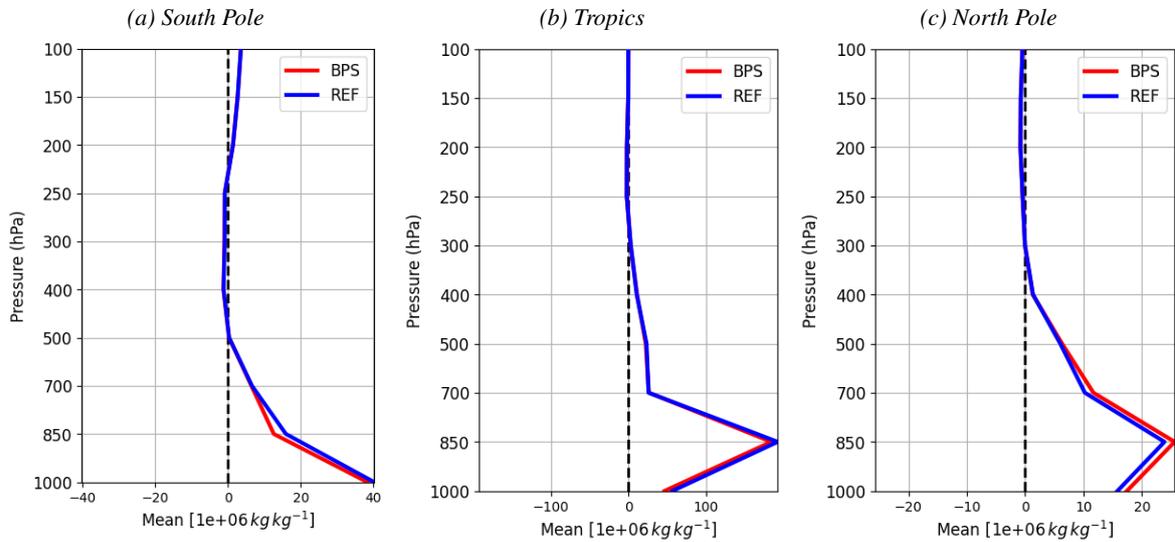


Figure C.11: Same as Fig. C.9 but for the departure with the relative humidity data (in $1.e^{-6} \text{kg kg}^{-1}$).

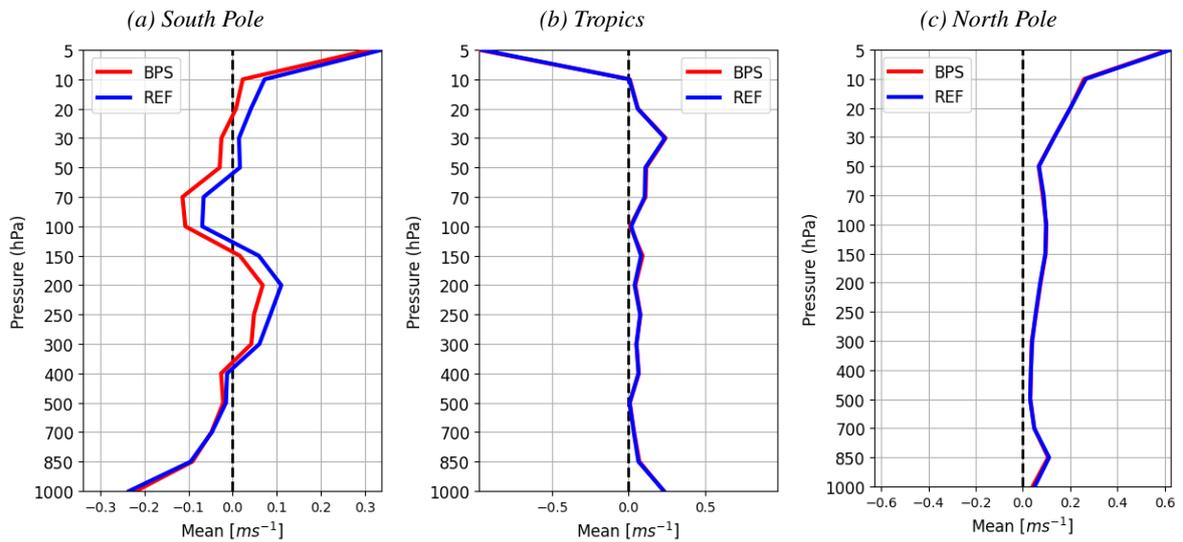


Figure C.12: Same as Fig. C.9 but for the departure with the U-component of wind data (in ms^{-1}).

Appendix D Root mean square error maps of surface parameters

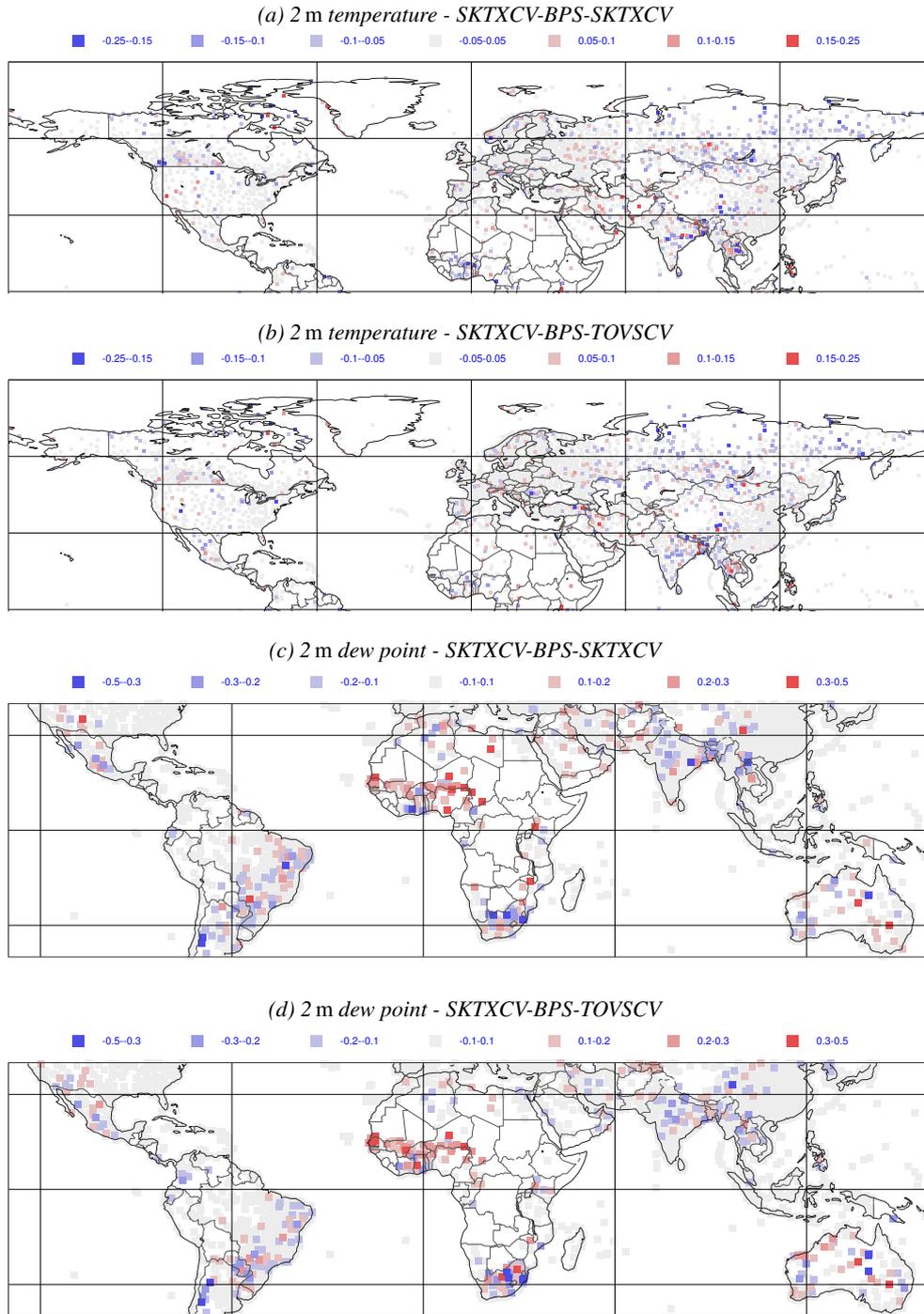


Figure D.13: Root mean square error (RMSE) of 2 m temperature and dew point at the forecast step 12 h, between the SKTXCV-BPS experiment and the SKTXCV or the TOVSCV experiment.