TECHNICAL MEMORANDUM

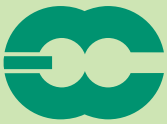
839

Global observing system experiments in the ECMWF assimilation system

Niels Bormann, Heather Lawrence, Jacky Farnan

Research Department

January 2019



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

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

Abstract

This study summarises results from observing system experiments with the ECMWF system, conducted over two seasons covering a total of 8 months. The experiments investigate the forecast impact of withholding selected observations from the assimilation system compared to using the full observing system. The observing systems considered are: conventional observations, microwave radiances, data from hyperspectral infrared instruments, bending angles from GPS radio occultation, as well as Atmospheric Motion Vectors.

Results show that conventional observations and microwave radiances are presently the main drivers of headline scores, with infrared sounders adding further robustness for a wide range of geophysical variables. GPS radio occultation measurements give significant impact in the upper troposphere/lower stratosphere, mainly on temperature, but also other variables, and the data have a clear influence on the mean state in these regions. Atmospheric Motion Vectors add benefits for tropospheric wind, particularly in the tropics and at the short range. The strong impact of the microwave satellite radiances is aided by the availability of an unprecedented number of instruments, providing good spatio-temporal coverage.

The observing systems considered have considerable effects on mean analyses, especially the conventional observations, resulting both from the direct assimilation of the observations as well as interactions with the variational bias correction.

1 Introduction

This memorandum gives an overview of the impact of some of the main observing systems in the current ECMWF assimilation system. The basis are assimilation experiments in which selected observing systems are denied from the assimilation and the results are compared to those from an experiment with the full observing system. The purpose for conducting these experiments is two-fold: Firstly, they have been performed in support of observing system experiments (OSEs) for the polar regions conducted under the EU-funded APPLICATE project, in order to put results for the polar regions into context of experiments with the global observing system. Secondly, the experiments highlight the value of microwave observations from satellites for the atmospheric analysis in Numerical Weather Prediction (NWP), in the context of the overall global observing system. This provides input to recent discussions regarding frequency protection and spectrum allocation (e.g., English et al. 2018).

Observing system experiments provide an evaluation of the complementarity and resilience of the present global observing system for Numerical Weather Prediction (e.g., Bouttier and Kelly 2001, Kelly and Thépaut 2007, Radnoti et al. 2012, McNally 2014). They are an important tool for analysing the interaction of different observations and can help in the planning of the future global observing system. When conducted periodically, they also help to document the evolution of the changing impact of different components of the observing system. Such changes are the result of either changing coverage or availability of observations, or improved use of the observations over time. For instance, a decade ago observations sensitive to humidity and clouds were providing comparatively little medium range forecast impact in assimilation systems (e.g., Kelly and Thépaut 2007), whereas their impact has grown in recent years, both through a larger number of available observations as well as developments of all-sky assimilation (e.g., Geer et al. 2017). It should be noted that OSEs therefore always only reflect the present usage of the observations in the context of the present observing system, and will neither give an indication of the potential of a given observing system, nor its performance in the context of a changed observing system or data usage. In contrast to cheaper adjoint-based estimates of short-range forecast impact (e.g., Langland and Baker 2004, Cardinali 2009), OSEs allow a detailed characterisation of the impact of observing systems both in terms of any geophysical variables as well as the medium-range forecast impact.

The last comprehensive OSEs in the ECMWF system have been conducted with cycle 40R1 by McNally (2014). He noted strong impact from conventional observations over the Northern Hemisphere, with microwave and infrared radiances providing leading impacts from satellites. The study demonstrated some resilience of the observing system, in the sense that the removal of one satellite observing system had a much smaller impact than the removal of all satellite data.

Major additions in the observation usage since then are the introduction of many more humidity-sensitive microwave instruments (MWHS, MWHS-2, SAPHIR, F-18 SSMIS, GMI, AMSR-2), the hyperspectral infrared sounder CrIS on S-NPP, as well as more high-resolution radiosonde data and aircraft observations. Observation usage has been enhanced through a number of observation error upgrades (including for GPS radio occultation, radiosondes, and IASI, with inter-channel error correlations taken into account for the latter), the use of a 2-dimensional observation operator for GPS radio occultation (GPSRO) and slant-path radiative transfer for clear-sky radiances, better usage of microwave and infrared data over land, improved aircraft bias correction (in 45r1), as well as a number of quality control refinements, among others. A key change of the data assimilation methodology has been the modification of the humidity background error formulation, making better use of flow-dependent characteristics as described by an Ensemble of Data Assimilations (EDA; Isaksen et al. 2010). A more complete summary of the changes in the ECMWF system can be found under: https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model

The structure of this memorandum is as follows. First we describe the observing systems considered in the experimentation, and the assimilation experiments conducted. This is followed by a discussion of the verification strategy employed in this study, and the verification results for the short- and medium-range forecast impact, respectively. We also provide an overview of the mean analysis changes introduced by the various observing systems, before summarising the main conclusions in the last section.

2 Observing systems and experiments

The present study characterises the forecast impact of five leading observing systems in the ECMWF system, using assimilation experiments in which each of these observing systems is withheld from an otherwise full system. Table 1 gives an overview of the observing systems considered. They comprise all conventional observations, all microwave radiances, infrared sounder radiances, bending angles from GPSRO, as well as Atmospheric Motion Vectors (AMVs). Observing systems not considered in separate denial experiments, but still included in the experimentation, are geostationary radiances, scatterometer observations, ozone retrievals, and precipitation estimates from ground-based radar observations. The choice of observing systems were primarily driven by the requirements of the APPLICATE study, which is considering the polar observing system, and hence some observations with no or very limited use over the polar areas in the ECMWF system were not considered.

The experiments were conducted with ECMWF's 12h 4-dimensional variational (4D-Var) assimilation system (e.g., Rabier et al. 2000, Bonavita et al. 2012), covering the two periods 1 June to 30 September 2016 and 1 December 2017 to 31 March 2018. They are based on cycles 43R3 and 45R1 of the operational system, respectively, but have been run at the lower spatial model resolution of T_{CO} 399 (approximately 25 km), with 137 levels in the vertical, and a multi-incremental analysis with three outer loops and a final incremental analysis resolution of T_L 255 (approx. 80 km). A Control experiment was run that includes all observations, and five denial experiments that exclude the five observing systems listed in Table 1, respectively. For computational reasons, the same background error specification was used in all experiments, based on situation-dependent estimates from an EDA that uses the full observing

Table 1: Observing systems considered in this study.

Observing system	**Types/instruments/satellites**
Conventional observations	Radiosondes Balloon and profiler wind observations Aircraft reports Synop stations Drifting buoys
Microwave radiances	*Sounders:* AMSU-A from 6 satellites (NOAA-15, -18, -19, Aqua, Metop-A, -B) ATMS on S-NPP MHS from 4 satellites (NOAA-18, -19, Metop-A, Metop-B) MWHS on FY-3B MWHS-2 on FY-3C SAPHIR on Meghatropiques *Combined imager/sounders:* SSMI/S from 2 satellites (F-17, -18[1]) GMI on GPM *Imagers:* AMSR-2 on GCOM-W1
Infrared sounder radiances	AIRS on Aqua IASI from 2 satellites (Metop-A, Metop-B) CrIS on S-NPP
GPSRO	Data from up to 9 satellites (Metop-A, Metop-B, TANDEM-X, TERRASAR-X, FY-3C, GRACE-A, and up to 3 COSMIC satellites)
AMVs	Data from up to 5 geostationary satellites (Meteosat-10/11[2], -7/8[2], Himawari-8, GOES-13[3], -15) Data from 7 polar satellites (NOAA-15, -18, -19, Aqua, S-NPP, Metop-A, -B) Dual-satellite AMVs from Metop-A/B

[1]Only 183 GHz humidity sounding channels used for F-18. [2]Depending on period. [3]Only up to 1 January 2018.

system. This means that the increase in the background error resulting from the denial of observations is neglected here, and the assumed background errors for the denial experiments are likely to be too small. Some of the loss in forecast skill in the denials could be recuperated through re-running the EDA without the denied observations, as highlighted, for instance, in McNally et al. (2014). The effect is, however, considered relatively small for experimentation with single observing systems.

Figure 1 provides an overview of the number of observations used in the Control experiment with the full observing system, separated by broad geographical regions. It shows the well know characteristics of these observing systems, with the majority of conventional observations located in the Northern Hemisphere extra-tropics, whereas the coverage of the four satellite observing systems is more equally distributed among these regions. Together, satellite data provide around 90 % of the observations, with the IR sounders the largest contributors (57 % of the total number of observations assimilated), followed by passive MW radiances (25 %). It is clear from this Figure that the number of observations denied in each experiment considered here is very different, and this will be one factor to bear in mind when interpreting the forecast impact obtained.

The spatio-temporal sampling characteristics for the two largest satellite observing systems, IR sounders and MW radiances, are very different. Despite the large data numbers, the spatio-temporal coverage for the IR sounders is actually much poorer than that of the MW sounders. This is illustrated in Fig. 2, which highlights the temporal coverage for a given longitude, in contrast to the commonly displayed spatial coverage over a given period. The 180E meridional was chosen here as it is mostly over ocean, therefore avoiding effects introduced through more conservative data use over land regions; qualitatively, the situation would be similar for other locations. As can be seen, the IR sounder coverage originates from only four satellites, some of which have very similar equator crossing times (e.g., Metop-A/B; Aqua/S-NPP). In the tropics and the mid-latitudes, large periods of the day remain unobserved by IR sounders. For every point observed by the IR sounders up to 191 channels are assimilated, subject to cloud screening and other quality control. In contrast, MW sounder radiances are presently provided by many more satellites, with different orbital characteristics, hence giving much more frequent coverage throughout the day, albeit with observations in much fewer channels (3-15, depending on instrument) and inherently much poorer vertical resolution. The coverage is currently particularly good for MW humidity-sounders, which are available from 11 satellites. They achieve better than $2\frac{1}{2}$-hourly sampling for most latitudes, with almost half-hourly coverage pole-ward of 70° latitude for around two-thirds of the day. However, for MHS for instance, only 3 sounding channels are assimilated, so the vertical resolution is very limited. This good coverage is of course not guaranteed to continue in the future, as many of the satellites present are well past their design life.

Another critical difference between the coverage for the IR and the MW data is the geophysical sampling. The MW radiances are assimilated over most surface types (ocean, land, sea-ice), including surface-sensitive radiances (e.g., Bormann et al. 2017), and the majority of the humidity-sensitive MW radiances are assimilated in all-sky conditions (Geer et al. 2017), with the exception of ATMS and MWHS radiances (Bormann et al. 2013, Chen et al. 2015). In contrast, the IR radiances are confined to channels not affected by clouds (and a limited sample of totally overcast scenes, McNally 2009), and over land only data not sensitive to the surface are used (for the Dec 2017 - March 2018 experiment). This further limits the sample of assimilated IR data in the troposphere.

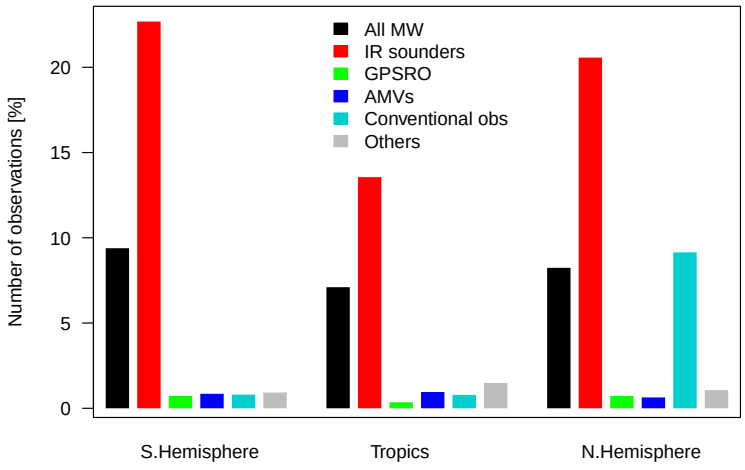


Figure 1: Percentage of observations over the Southern Hemisphere extra-tropics, Tropics, and Northern Hemisphere extra-tropics for the observing systems considered in this study. "Others" denotes all other observations (that is, geostationary radiances, scatterometer observations, ozone retrieval, and precipitation from ground-based radar observations). Data are based on the two experimentation periods combined.

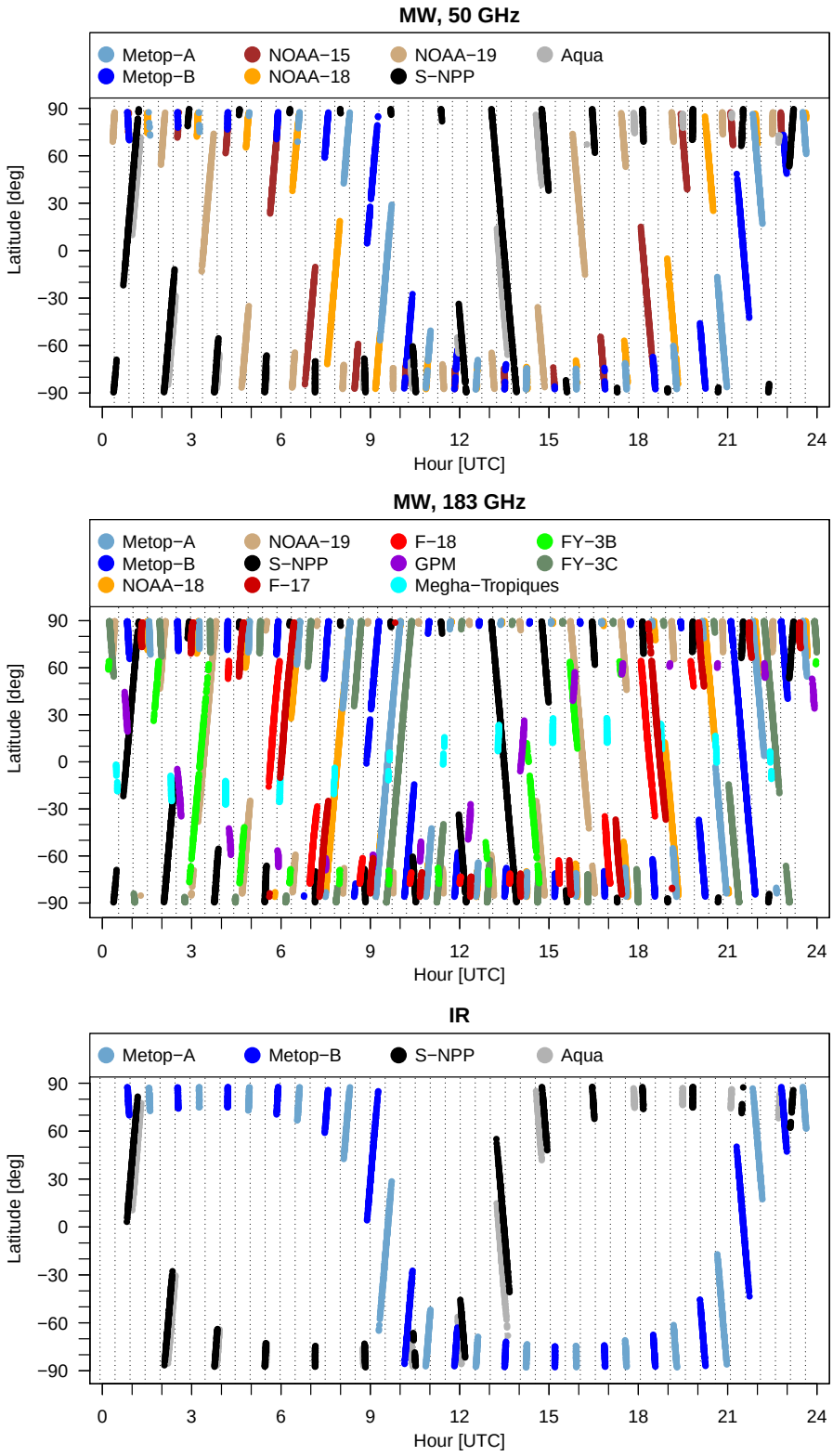


Figure 2: Temporal coverage for MW temperature sounders (top), MW humidity sounders (middle) and IR sounders (bottom). The plots show the coverage of assimilated observations in a meridional band of 500 km width centred on the 180E meridian, as a function of time of day (x-axis) and latitude (y-axis), for 15 February 2018.

3 Verification considerations

To evaluate the forecast impact obtained with the five observing systems requires comparing the forecasts against a reference to characterise the difference between the forecast errors in the denial experiments and the Control. The choice of this verifying reference can have a dramatic effect on the appearance of the forecast errors found, especially in the short range. This is highlighted in Fig. 3, which displays the change in the size of forecast error for wind at 200 hPa for the experiment with the denial of conventional observations as an example. Verification against radiosondes indicates an increase in forecast error of 8 % over the Northern Hemisphere after 12 h, so a considerable degradation when the conventional observations are withheld. In contrast, when the Control and the Denial are each verified against their own analyses, the scores suggest that forecast errors are smaller by 12 % when conventional observations are denied, ie a very significant improvement when the conventional observations are withheld (but with some degradation from day 3 onwards). Verification against the operational analysis instead suggests a massive increase in forecast errors exceeding 40 % at 12 h, ie, a massive degradation of short-range forecasts without conventional observations. It should be noted that the 200 hPa wind error is a particularly drastic example for verification differences for short-range forecasts, and other variables do not necessarily show such strong disagreements, but some significant differences are very common.

The differences in the short-range verifications are a result of the different error characteristics of the verifying references used, and in particular they reflect the degree of independence of the verifying analyses and the forecast errors. This aspect matters for short-range forecasts because the analysis and forecast errors are of comparable magnitude. Consider, for instance, the own-analysis verification shown in Fig. 3, suggesting strong degradations from the assimilation of conventional observations. As the assimilation system uses a short-range forecast to produce the analysis, errors in the analysis chosen for verification and errors in the short-range forecasts are necessarily correlated. This means that some forecast errors will be masked in own-analysis verification, and the size of the forecast error will be under-estimated. The apparent degradations seen in Fig. 3 likely reflect that the degree of this correlation is reduced when the conventional observations are added, and this aspect outweighs the accompanying reduction in the forecast error. Indeed, forecasts in a poorly observed area may apparently verify very well against an own analysis in the short range, simply because no observations have corrected the

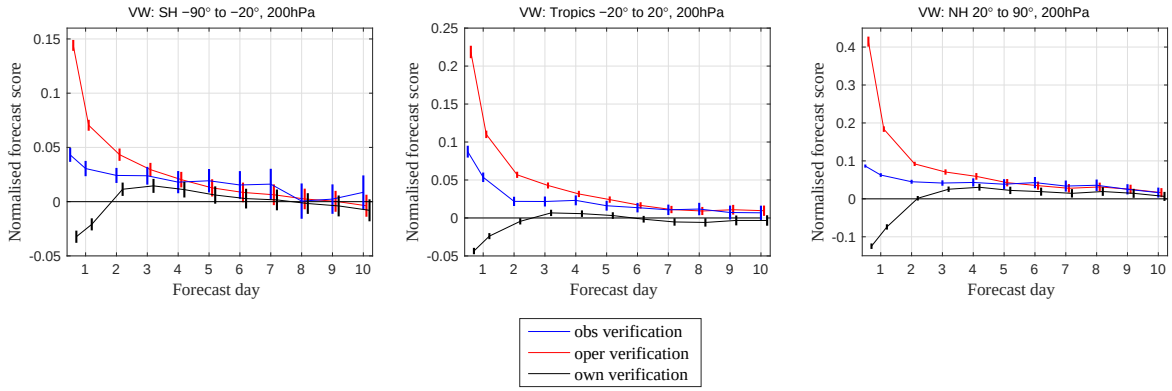


Figure 3: Normalised difference in the 200 hPa vector wind forecast error between the experiment without the conventional observations and the Control for the two experiment periods combined. Verification is performed against radiosonde observations (blue), operational ECMWF analyses (red) or each experiment against its own analysis (black). Error bars indicate 95 % statistical confidence intervals, following Geer (2016).

short-range forecast errors during the analysis (see also Geer et al. 2010). So own-analysis verification may incorrectly suggest that systems with fewer observations perform better for short-range forecasts, a particularly unfortunate situation for observing system experiments.

In contrast, verification against the operational analysis for short-range forecasts may appear somewhat optimistic in Fig. 3. This is because some aspects of the errors in the operational analysis are shared with the Control, as a result of using the same observing system, as well as the same forecast and analysis system. Verification against operational analyses may therefore favour the Control. However, the situation is not as bad as when all experiments were verified against the Control analysis, as the operational analysis is produced with a different assimilation set-up, at higher resolution and different computer configuration. The operational analysis hence provides a more independent sample of the random part of the analysis error, and the problems of correlations between random forecast errors and analysis errors are less severe than in own-analysis verification or verification against the Control analysis.

For short-range forecasts, the verification against observations may offer an attractive alternative, and the results shown in Fig. 3 for verification against radiosondes do not appear to show obvious flaws. However, verification against observations is also not without pitfalls. Ideally, independent observations that are not assimilated should be used for this, but of course in practice very few such observations of relevance are available. Forecasts can still be verified against observing systems that are assimilated (as in this example), and these observations are at least independent in the sense that each forecast has not yet seen the verifying observations. Nevertheless, similar quality control in the assimilation and the verification, or, for more complex observations such as radiances, the same observation operator or bias correction mean that there is still some dependence, and this dependence may mask some short-range forecast errors in a similar way as correlations between forecast and analysis errors do in analysis-based verification. Possibly the biggest disadvantage for observation-based verification is, however, the restriction to the sampling of the available observations, either spatio-temporally, or through the geophysical variables being measured. The geographical sampling is particularly an issue for radiosonde-based verification, which, for instance, cannot capture the short-range forecast error over large parts of the Southern Hemisphere ocean.

The above mentioned verification problems are less of an issue for medium-range forecasts from around day 3-4 onwards. This is because forecast errors dominate at this range, and analysis errors and their correlations with forecast errors become less relevant. As a result, there tends to be more agreement between different verification options. In the example shown in Fig. 3, there is qualitative agreement between the results from the verification against radiosonde observations and the verification against operational analyses from around day 3-4 onwards. Results from verification against own analyses are also more similar at this forecast range, indicating benefit from the observations, albeit to a smaller extent. While the ranking of the impact of the observing systems tends to agree for the verification against radiosonde observations and against the operational analysis, own-analysis verification can at times give a different impression.

Given the severe problems with analysis-based verification for short-range forecasts, we will in the following use observations to verify the short-range forecasts. The limitations in geographical sampling will need to be kept in mind, especially when looking at verification against radiosondes. We aim to counter-act these sampling problems by evaluating against a wide range of observations, including observations with more even geographical coverage. We will base our medium-range forecast verification on verification against the operational analysis. This tends to agree with verification against observations where good observational coverage is available, but is not limited to the observational coverage available, so allows a more representative evaluation in that respect.

4 Results

4.1 Short-range forecast impact

Figures 4 and 5 give an overview of the short-range forecast impact evaluated against a range of conventional and satellite observations, respectively. They provide a characterisation in terms of temperature, humidity, and wind over large hemispheric regions and at different levels in the atmosphere. The statistics highlight the complementary nature of the currently assimilated observing system, with almost all

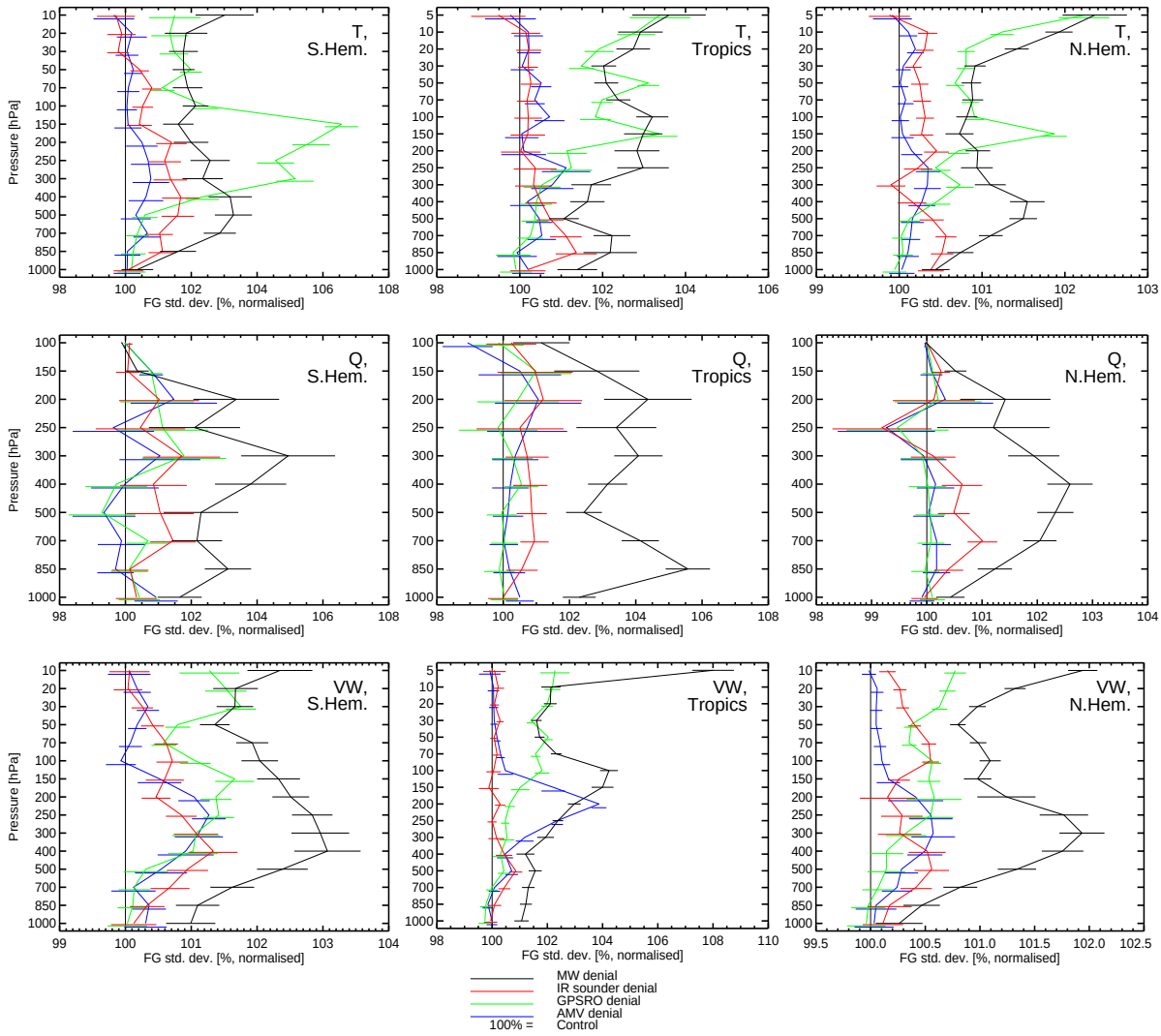


Figure 4: Standard deviation of background departures, normalised by the Control, for several conventional observing systems, for the Southern Hemisphere extra-tropics (left), the Tropics (middle), and the Northern Hemisphere extra-tropics (right). The observations are temperature from radiosondes (top), humidity from radiosondes (middle), and vector wind from radiosondes, profiler, pilot, and aircraft observations (bottom). Statistics cover the two seasons combined. Values for the four experiments with the satellite observing systems are shown. A value greater than 100 % indicates an increase in the error in the background due to the denial of the respective observing system. Horizontal lines indicate statistical significance at the 95 % level.

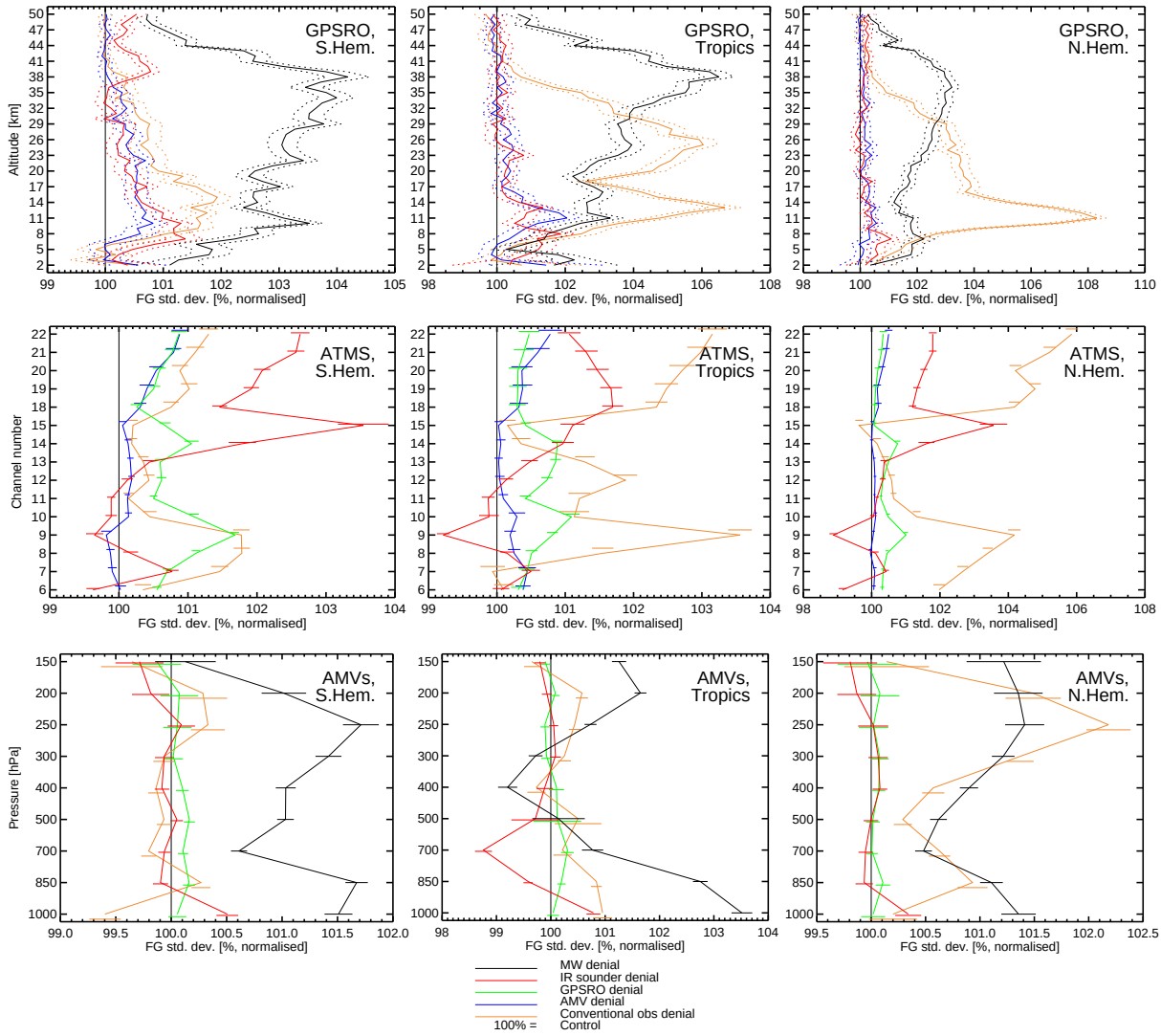


Figure 5: As Fig. 4, but for a number of satellite observing systems. The observations are bending angles from GPS radio occultation (top), MW sounding radiances from ATMS on-board S-NPP (middle), and Atmospheric Motion Vectors (AMVs, bottom). GPSRO is primarily sensitive to temperature, with some sensitivity to humidity in the troposphere. ATMS channels 6-15 are primarily sensitive to temperature, with sensitivity of channel 6 in the mid/lower troposphere and channel 15 peaking around 2 hPa. In contrast, channels 18-22 are tropospheric humidity-sounding channels, with the peak altitude increasing with channel number. AMVs provide estimates of wind at single altitudes derived by tracking cloud motions in sequences of satellite imagery.

observing systems considered here providing leading short-range forecast impact for at least some aspect of the atmosphere. The results also show the expected regional differences, with the conventional observations providing strong forecast impact over the Northern Hemisphere where the largest number of observations are available (Fig. 5, right column), whereas the satellite data dominate the forecast impact over the Southern Hemisphere. Somewhat surprising is the relatively large short-range forecast impact of the conventional data in the tropics (Fig. 5 middle column), given the relatively low data numbers shown in Fig. 1. There has been a marked increase in the number of available aircraft data in the tropics, such that more than half of the conventional data for the second experimentation period come from aircraft, and this likely contributes to the good impact found for conventional data.

The microwave or infrared radiances show significant impacts on temperature, humidity, and wind, in the troposphere as well as the stratosphere (Figures 4, 5). The very clear impact on wind, despite the data being only sensitive to temperature, humidity, and clouds, is likely a result of thermal adjustment processes combined with tracing effects in 4D-Var. For the MW data, the latter is aided by the high spatio-temporal sampling of the data presently available, as discussed earlier. The impact of the IR data is not quite as strong as that of the MW, probably partly a result of the poorer spatio-temporal sampling available with the current observing system, combined with the restriction to primarily cloud-free observations. An apparent degradation from using the IR sounder data is found with S-NPP ATMS channel 9, a temperature-sounding MW channel that peaks around the tropopause (Fig. 5, middle). The equivalent AMSU-A channel instead shows one of the strongest improvements for all AMSU-A instruments (not shown), so the signals here are not consistent between different instruments. A similar inconsistency, though in the opposite direction, is observed for GPSRO data, for which ATMS channel 9 suggests some of the largest benefits from introducing the data, whereas AMSU-A channel 8 suggests near neutral impact. The reasons for the inconsistencies for these equivalent channels are not fully understood. The channels show different geographical bias pattern, possibly the result of subtle differences in level-1 processing or pass-band characteristics. The possibly wider vertical sampling of the tropopause of ATMS channel 9 resulting from the wider swath compared to AMSU-A may also play a role. Given the good impact of the IR data seen in the radiosonde statistics, we do not think that the degradation seen here against ATMS channel 9 is a major concern.

Despite much lower data numbers, the beneficial impact of AMVs and GPSRO data is also very clear. The importance of GPSRO for temperature in the upper troposphere/lower stratosphere is very apparent, particularly over the Southern Hemisphere (Fig. 4, top). This also leads to benefits for wind at these levels, and there are indications of small benefits for tropospheric humidity, as highlighted through the ATMS humidity channels (channels 18-22 in Fig. 5, middle). AMVs, in contrast have their largest impact on upper tropospheric wind (Fig. 4, bottom), particularly in the tropics. Despite no sensitivity to humidity for AMVs, benefits for humidity are apparent from the ATMS humidity channels (Fig. 5, middle), most likely a result of improved humidity transport. Note that the AMV coverage had a gap over the Americas for part of the second experimentation period (after 1 January 2018), as GOES-13 retired and GOES-16 was still undergoing pre-operational testing (e.g., Lean and Bormann 2019). So the AMV impact may be slightly underestimated compared to full coverage from five geostationary satellites.

4.2 Medium-range forecast impact

We will now investigate the forecast impact for day 2 and beyond. This will primarily be evaluated through verification against the operational analysis, for the reasons discussed earlier.

Figure 6a, b shows the forecast impact of denying the various observing systems over the extra-tropics for the 500 hPa geopotential for the two seasons combined. For the Northern Hemisphere, the conventional observations show the largest impact, followed by the MW observations, with degradations of the forecast error at day 3 of 10 and 6 %, respectively, when the data are excluded. Statistically significant forecast impact is obtained from both observing systems out to day 7. Over the Southern Hemisphere, the MW observations show the dominant forecast impact, leading to an 11 % degradation at day 3 when the data are excluded, and some statistically significant impact can be detected out to day 9. Clear benefits are also obtained from conventional observations, IR sounders, and GPSRO data, and denial of these observations shows a statistically significant increase in the forecast error at day 3 of around 2-3 % each. The different behaviour over the Northern and Southern Hemisphere reflects the well known observation coverage, with relatively few conventional observations available over the Southern Hemisphere, such

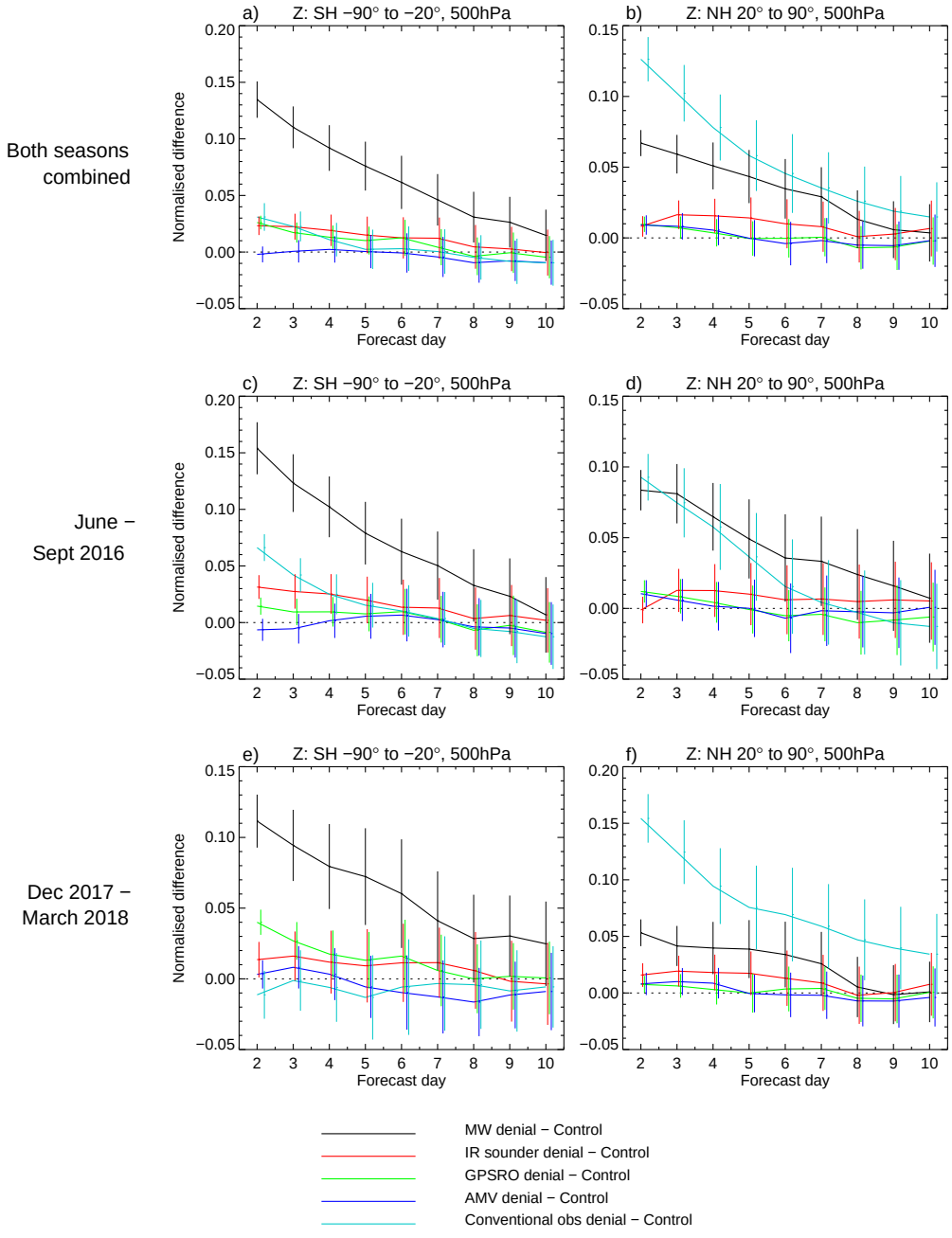


Figure 6: Normalised difference in the standard deviation of the forecast error in the 500 hPa geopotential versus the Control experiment, as a function of forecast range for the five observing system experiments as listed in the legend. The left column shows results for the Southern Hemisphere extra-tropics, whereas the right column shows results for the Northern Hemisphere extra-tropics. a) and b) show results for the two seasons combined, whereas c) and d) show the June – September 2016 period and e) and f) show December 2017 – March 2018 separately. Vertical bars indicate 95 % confidence intervals; corrections for temporal correlations in score differences and multiple comparisons (Sidak) have been applied, as described in Geer (2016).

that the analysis heavily relies on satellite data. The benefits from including the MW data in the Southern Hemisphere and those from including the conventional data over the Northern Hemisphere translate to a gain in forecast skill of around 6-7 hours at day 3.

There is some variation in these results over the two seasons considered (Figures 6c-f). The relative impact of the microwave data over the Northern Hemisphere appears to be larger over the summer period during which the MW impact is comparable to that of the conventional observations. Lower impact of MW data over winter has previously been found in earlier studies with surface-sensitive MW data over land and sea-ice (Bormann et al. 2017), and partially attributed to poorer data usage of these observations over snow and sea-ice during winter. However, it is likely that other factors such as different meteorological regimes also play a role for the present result. In contrast, the relative impact of the conventional data is larger during the respective winter seasons for both hemispheres. The impact of the conventional data for the Southern Hemisphere winter is remarkable given the low data numbers, and on a par with that of the IR sounder data. Similarly, GPSRO shows stronger impact on the 500 hPa geopotential during the Southern Hemisphere summer period, again despite relatively low data numbers (cf Fig. 1). Reasons for the larger impact of the conventional data over winter are not fully clear, but may be linked to benefits in terms of determining the thermal structure of the troposphere in the vertical. The stronger impact seen

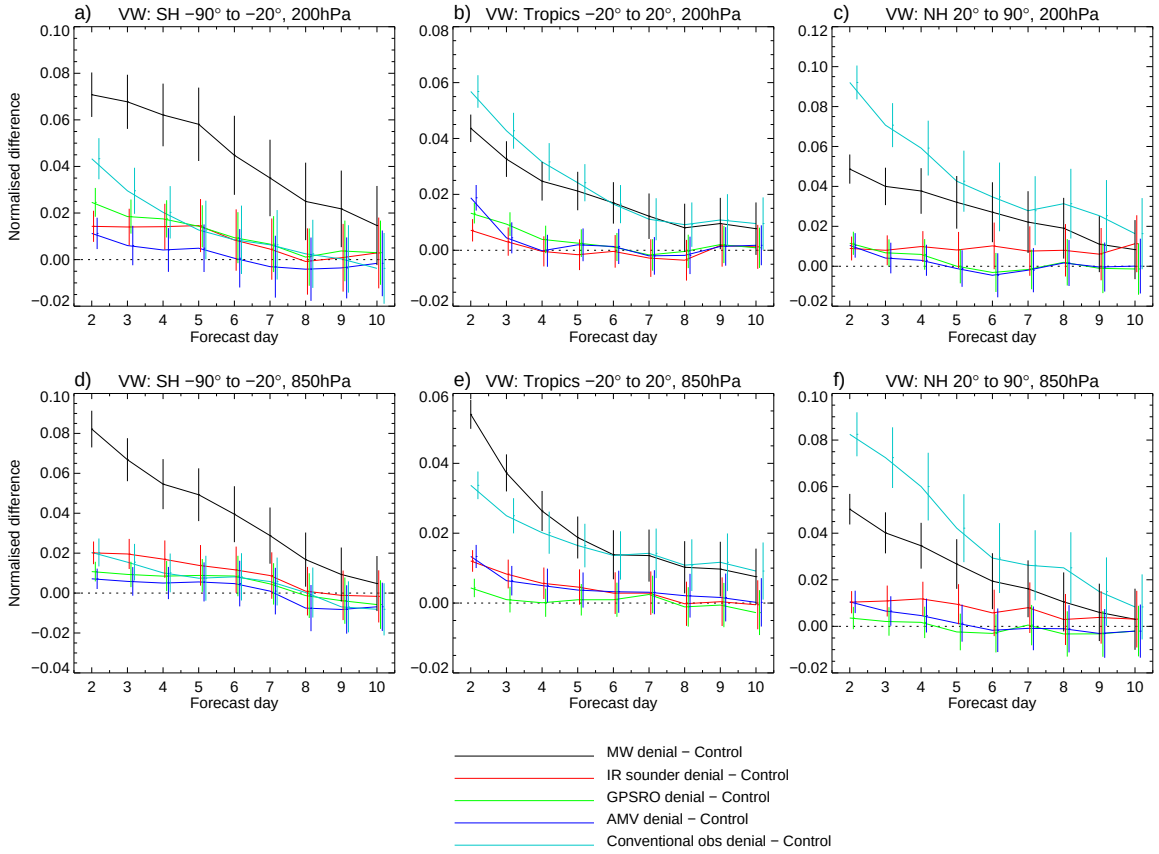


Figure 7: Normalised difference in the root mean square vector wind error versus the Control experiment, as a function of forecast range for the five observing system experiments as listed in the legend. The top row (a-c) shows results at 200 hPa, whereas the bottom row (d-f) shows results at 850 hPa. The left column covers the Southern Hemisphere extra-tropics, the middle one the Tropics, and the right one the Northern Hemisphere extra-tropics. Statistics are for both seasons combined. Vertical bars indicate 95 % confidence intervals, following Geer (2016).

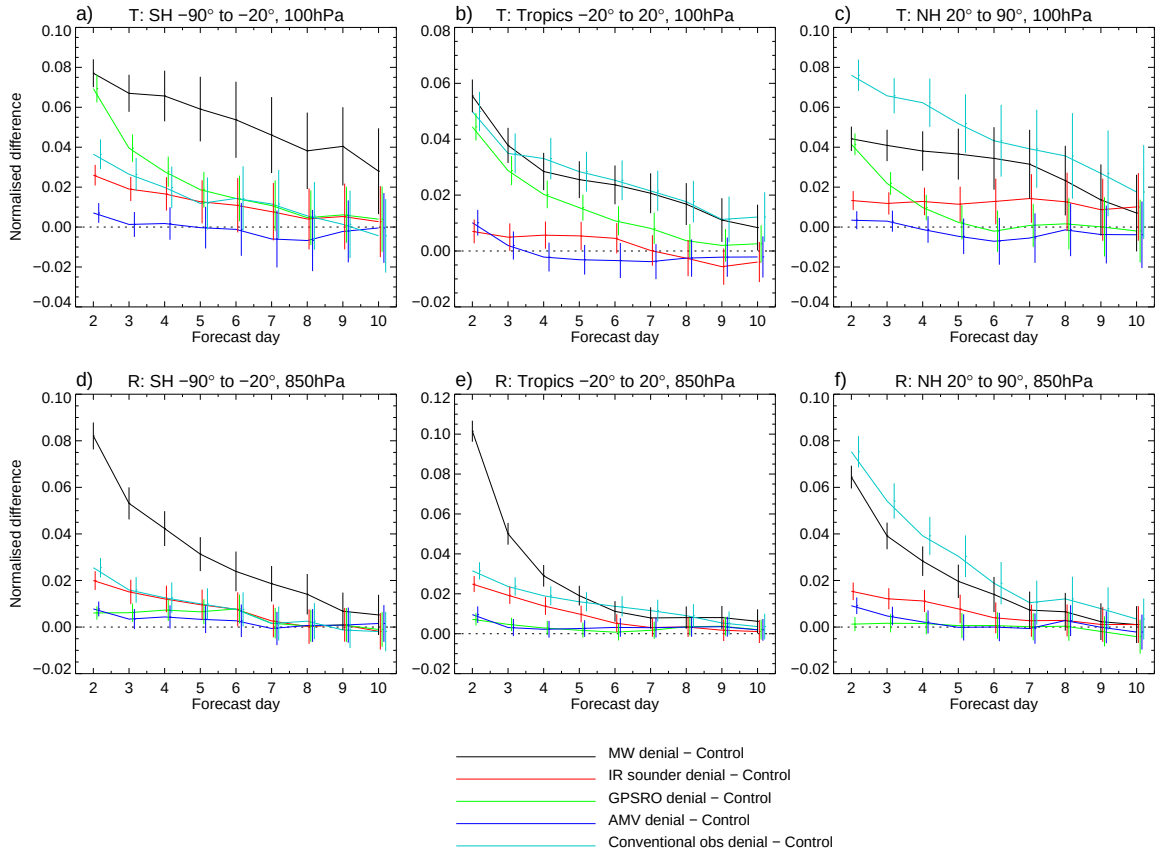


Figure 8: Normalised difference in the standard deviation of the forecast error versus the Control experiment, as a function of forecast range for the five observing system experiments as listed in the legend. The top row (a-c) shows results for temperature at 200 hPa, whereas the bottom row (d-f) displays statistics for the 850 hPa relative humidity. The left column covers the Southern Hemisphere extra-tropics, the middle one the Tropics, and the right one the Northern Hemisphere extra-tropics. Statistics are for both seasons combined. Vertical bars indicate 95 % confidence intervals, following Geer (2016).

for the Northern Hemisphere may also be a result of improved reporting practices for conventional data, with more high-resolution radiosonde and aircraft data available, and hence not only a seasonal effect.

The impact seen for the 500 hPa geopotential is also broadly representative of the impact on other tropospheric variables in the extra-tropics, at least in terms of the relative impact of the different observing systems (e.g., Figures 7a/d, c/f, 8a/d, c/f). The benefit of AMVs is somewhat clearer for wind forecasts, with statistically significant benefits from including the data out to day 3 for some levels.

For tropospheric wind forecasts in the tropics, the MW and conventional observations again show the largest impact, statistically significant out to day 7 and beyond, depending on level (Fig. 7b and e). The impact of AMVs on tropical wind is broadly comparable to that of the IR sounder data, and statistically significant out to day 3 or 4. Given the relatively strong impact of AMVs noted in the short-range forecast verification, the relatively swift loss of the AMV impact at 200 hPa may seem somewhat surprising. This may reflect a lack of overall thermo-dynamic adjustment during the assimilation, possibly because the observations themselves only provide single-level wind information, or because balance constraints are too weak in the tropics.

The significant impact of GPSRO data on upper tropospheric (and lower stratospheric) temperature noted in the short-range forecasts also continues into the medium-range. Adding the observations in an otherwise full observing system results in statistically significant benefits at 100 hPa out to day 6 over the Southern Hemisphere and Tropics (Fig. 8a-c), and smaller benefits in the mid- or lower troposphere (not shown). GPSRO play an important role in constraining biases and bias corrections in the assimilation system, and this aspect will be investigated further in the next section.

Finally, impact on low-level relative humidity is shown in Fig. 8d-f. Relative humidity is notoriously difficult to verify, as the problems discussed earlier regarding analysis-based verifications tend to be particularly severe. The results should hence be treated with some caution. However, the impacts of the observing systems considered on 850 hPa humidity is overall in line with that for other variables, and in particular consistent with that for wind at 850 hPa shown in Fig. 7d-f.

Given the strong impact of the MW data, Fig. 9 further separates this into impact from sounding and imaging channels, respectively. To do so, two further experiments have been run, a "MW sounder denial" in which all temperature- and humidity-sounding channels have been excluded (ie around 50-60, 118, and 183 GHz), and a "MW imager denial" in which all actively assimilated imager channels were denied (ie channels around 18.7, 23.8, 36.5, 89 GHz). Note that in these experiments we use data from only 3 instruments with imager channels, whereas data from 17 instruments provide temperature or humidity-sounding information. While the impact of the sounder data mostly dominates in the extra-tropics, the imager data nevertheless contributes significantly to the overall impact of the MW data. In particular, the imager channels add constraints on low-level humidity, which 4D-Var can use to infer wind information, as well as constraints on ocean surface wind through the ocean emissivity. These aspects appear particularly effective in the tropics (Fig. 9g, j). The impact on 200 hPa wind in the tropics is also notable, where the imager channels from 3 instruments appear to give around half of the impact of the sounding data.

The experiments discussed so far have only considered the denial of one observing system, and while the losses of one observing system are clearly significant, they are also not catastrophic. This suggests that there is good resilience of the present observing system and its use in the ECMWF assimilation system. However, this situation changes when we consider the loss of several observing systems. To highlight this, we conducted another assimilation experiment in which the microwave observations and the infrared sounder data are denied at the same time. In this case, the degradation in forecast skill is much stronger, and stronger than the sum of the individual losses of forecast skill (Fig. 10): in the absence of MW data, the IR sounders have a much stronger impact (compare the green line and the black line), and, equally, in the absence of the IR sounder data the MW observations have an even stronger impact than they have in the full system (compare the green and red lines). Note that for these experiments the background error characteristics are still those of the full control, and this will arguably over-estimate the loss of forecast skill even more in the case of the dual observing system. But it is apparent that the loss of one of these observing systems would result in a significant loss in resilience of the entire observing system, making forecasts more volatile to data outages and instrument failures and their quality therefore less reliable.

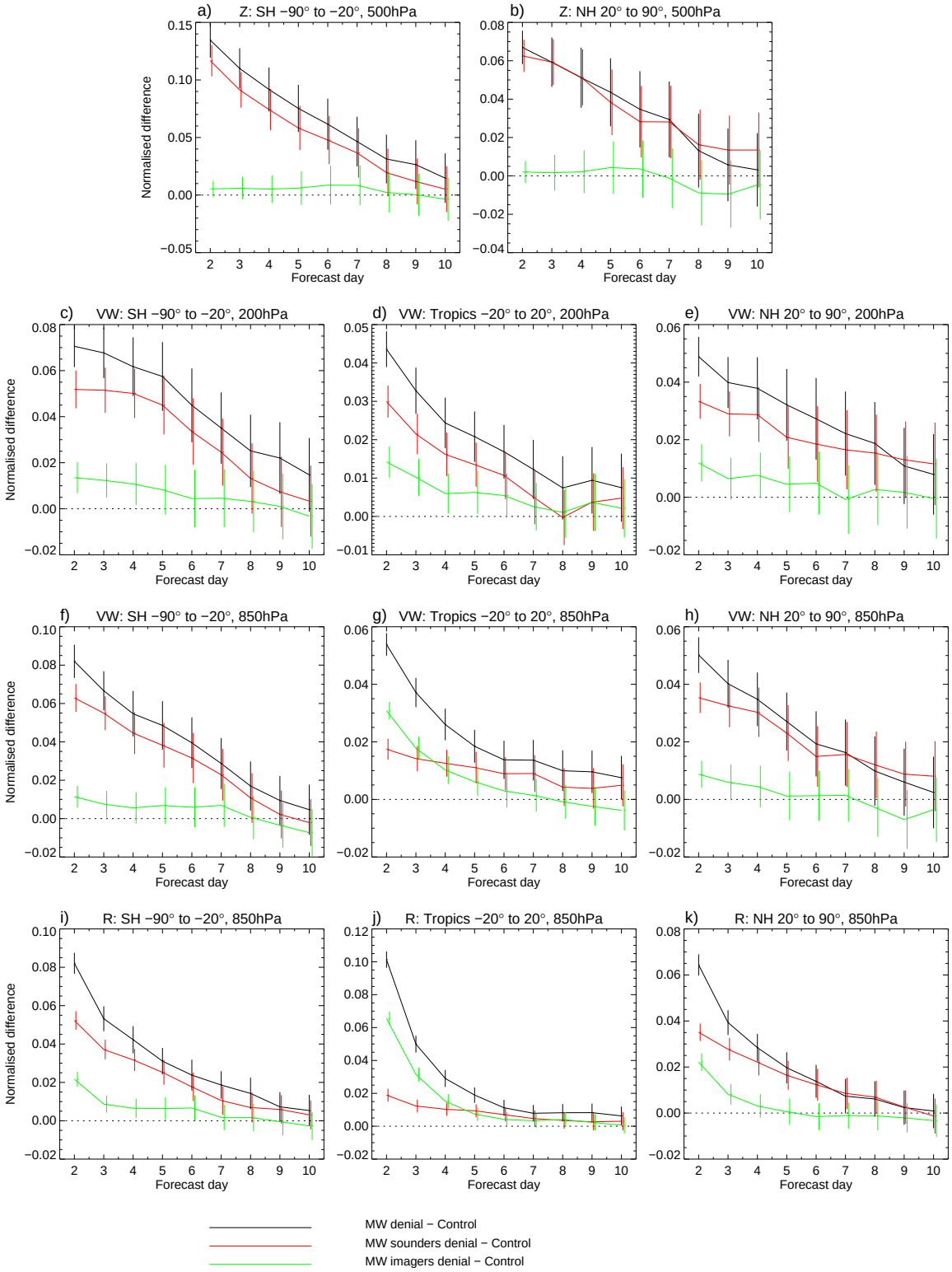


Figure 9: a, b) As Fig. 6a,b, but with the results of the denial experiment for all MW data (black) separated into impact obtained with MW sounder channels (red) and MW imager channels (green). c-h) As Fig. 7, but with the results of the denial experiment for all MW data (black) separated into impact obtained with MW sounder channels (red) and MW imager channels (green). i-k) As Fig. 8d-f, but with the results of the denial experiment for all MW data (black) separated into impact obtained with MW sounder channels (red) and MW imager channels (green).

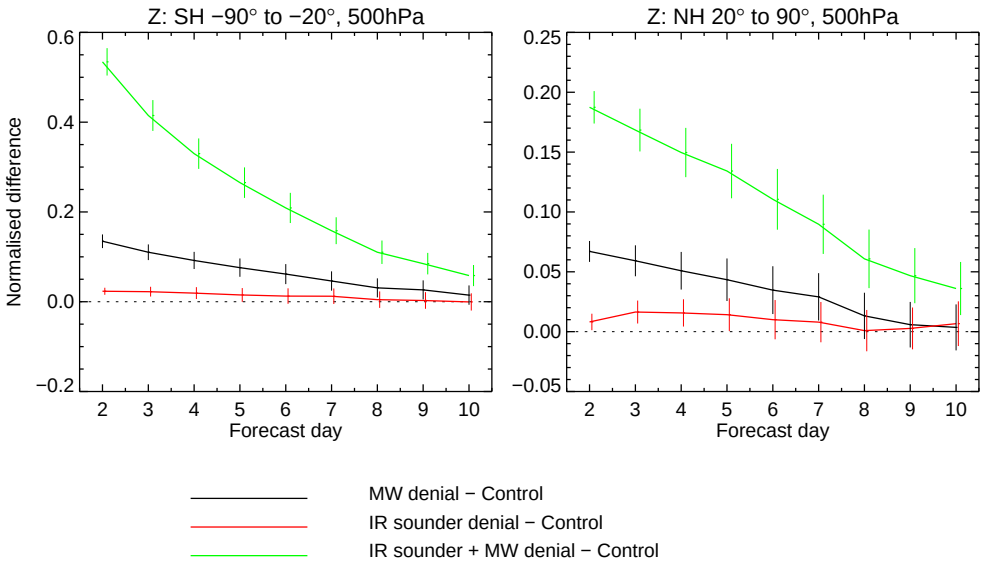


Figure 10: As Fig. 6a/b, but for observing systems in which the MW radiances (black) and the IR sounder data (red) are denied individually or together (green).

4.3 Impact on mean analyses

So far we have primarily looked at the reduction of random errors resulting from the addition of the observing systems considered here, and this is the most important aspect for NWP. However, there are also very considerable differences in the mean analyses, and in the following we will examine these more closely. Changes in the mean analysis from introducing observing systems tend to suggest that there are biases present, either in the forecast model, the observations, or the observation operators. Such biases are unavoidable, and the assimilation system includes some methods to address these. Biases in the observations or the observation operator are treated by variational bias correction (VarBC) during the assimilation (e.g., Dee 2004), but this is not without problems especially in the presence of forecast model bias (e.g., Eyre 2016). In the stratosphere, model biases are estimated through a weak-constraint formulation of 4D-Var (e.g., Goddard et al. 2017), with adjustment time-scales of the order of several months in the present configuration, but a bias-free forecast model is assumed elsewhere.

Validation of changes in the mean analysis is very difficult without assuming that certain observations are unbiased. Reference observations such as GRUAN[1] may offer their services here, but the changes observed in the following tend to be either too small or too localised to be convincingly captured by GRUAN. So in the following we will restrict ourselves to documenting the changes and pointing out some interactions which could be analysed further in the future. The changes reflect the uncertainty in the mean state, given unavoidable biases in the system. We focus on the winter period, but the summer period shows changes of similar magnitudes, albeit with different geographical structures.

Figure 11 shows considerable changes to the mean temperature analyses throughout the atmosphere from the introduction of conventional observations, GPSRO, MW instruments, or IR sounders. AMVs are not included here, as differences in terms of temperature are very small. Most of the differences are fairly large-scale and typically have primarily zonal features (with exceptions for the lowest tropospheric levels), so the zonal means displayed allow a convenient summary of the changes at a range of levels. The

[1]Global Climate Observing System (GCOS) Reference Upper-Air Network, www.gruan.org

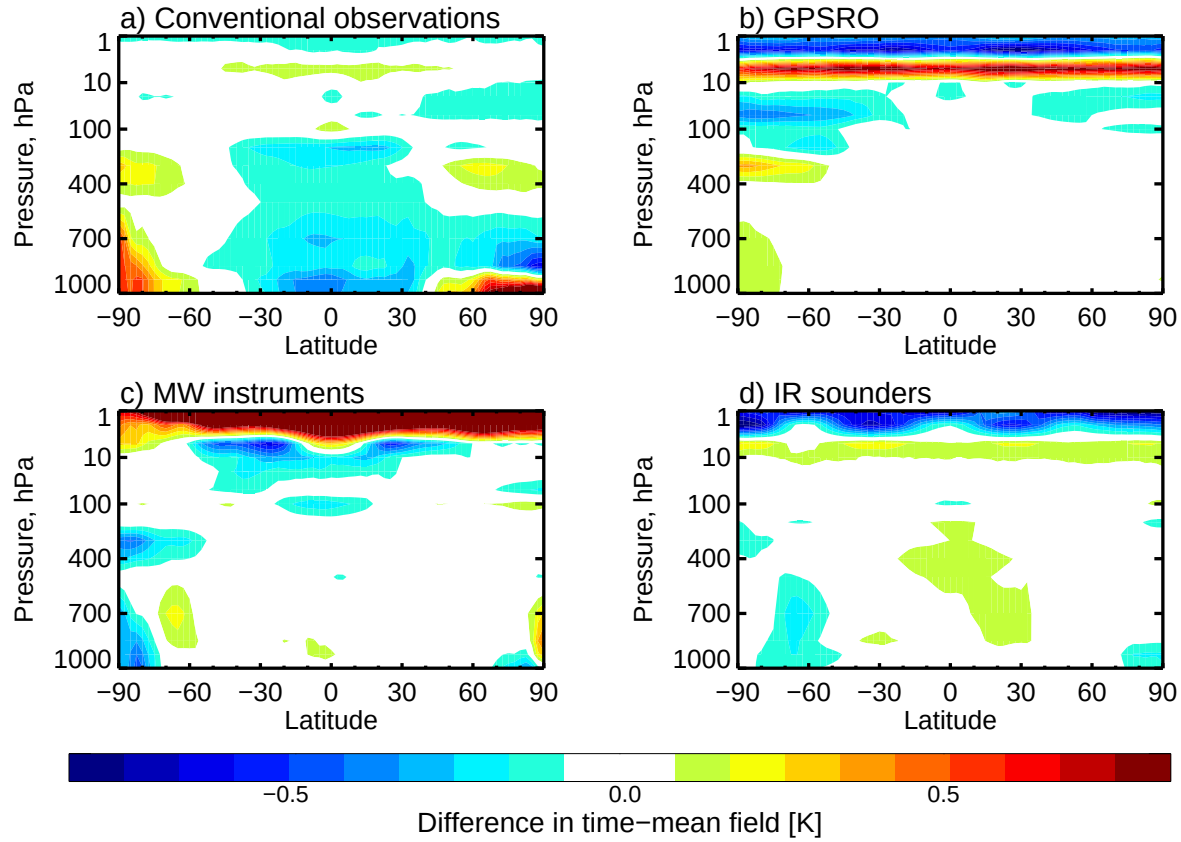


Figure 11: a) Zonal mean differences of the temperature analysis of the experiment with conventional observations denied and the Control for the December 2017 - March 2018 period. Red colours indicate a warming resulting from the denial of the observations. b) As a), but for the GPSRO denial. c) As a), but for the MW radiances denial, d) As a), but for the IR sounder denial.

largest changes in the troposphere result from the conventional observations, with vertically consistent changes of several tens of K in the tropics, and a more oscillating structure in the polar regions. Conventional observations and GPSRO appear to lead to consistent changes around 400-200 hPa over Antarctica, resulting in a cooling at these levels, whereas MW instruments and IR sounders lead to a warming at the same levels. The consistency of the signal from the conventional data and GPSRO may point to a forecast model bias here, as both observing systems are used as anchoring observations. In contrast, MW and IR radiances are prone to air-mass dependent biases and rely on the variational bias correction. As shown by Eyre (2016), VarBC in the presence of model bias is prone to at least partially re-enforce the model bias. So the finding that MW and IR radiances introduce changes of the opposite sign may still be consistent with the presence of model bias in these areas, but an uncorrected observational bias can also not be ruled out. In any case, the changes introduced by MW and IR radiances around 400-200 hPa over Antarctica act to increase the bias against radiosondes in the area.

The largest changes in the mean temperature analysis occur above 10 hPa. Conventional observations, GPSRO, and IR sounders introduce qualitatively similar changes, albeit of different magnitudes, whereas the MW radiances alter the analysis in the opposite direction. The strong influence of the MW instruments is a reflection of using the highest temperature-sounding channel as an anchor for the variational bias correction. For the period shown, this channel is treated by Constrained VarBC, penalising large bias corrections in order to counter-act strong model biases in the upper stratosphere (Han and Bormann

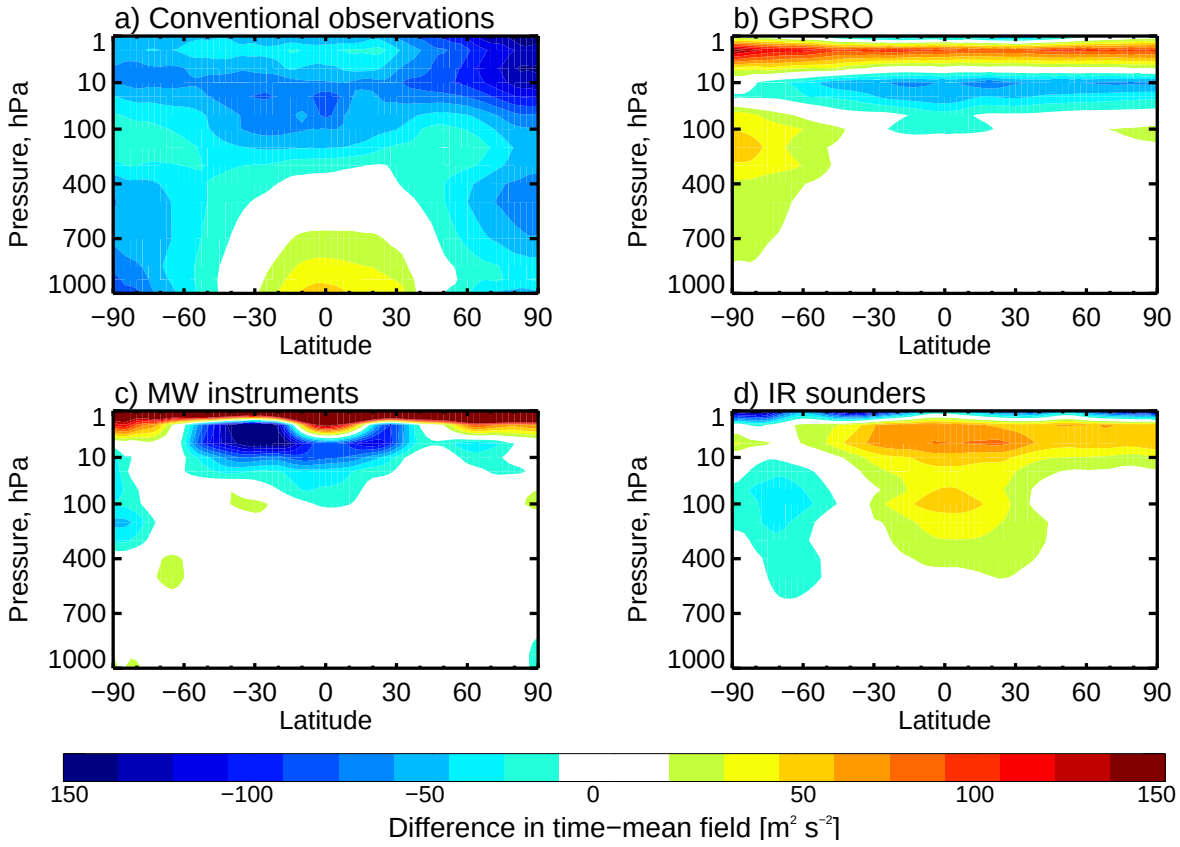


Figure 12: As Fig. 11, but for the geopotential.

2016). This provides a constraint for broad vertical structures only (as given by the weighting function of the selected channel), and little information on the vertical structure. The oscillating pattern in the zonal mean differences for the other observations are likely reflecting an interaction between the model bias and the further information on the vertical structure provided by these observations.

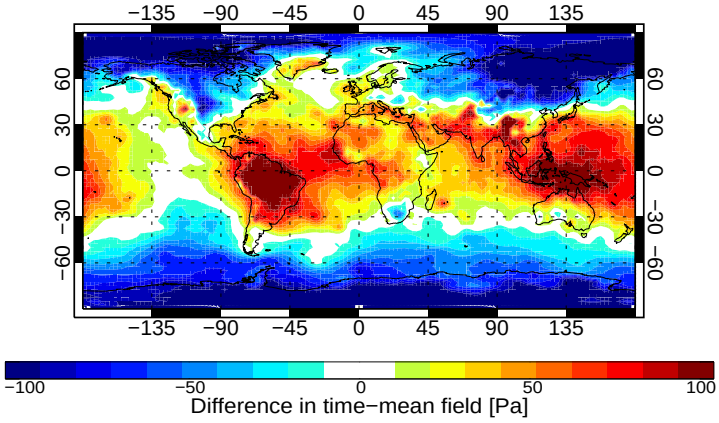


Figure 13: Map of the difference in the mean analysis of mean sea-level pressure between the experiment with conventional observations denied and the Control, for the period December 2017 - March 2018.

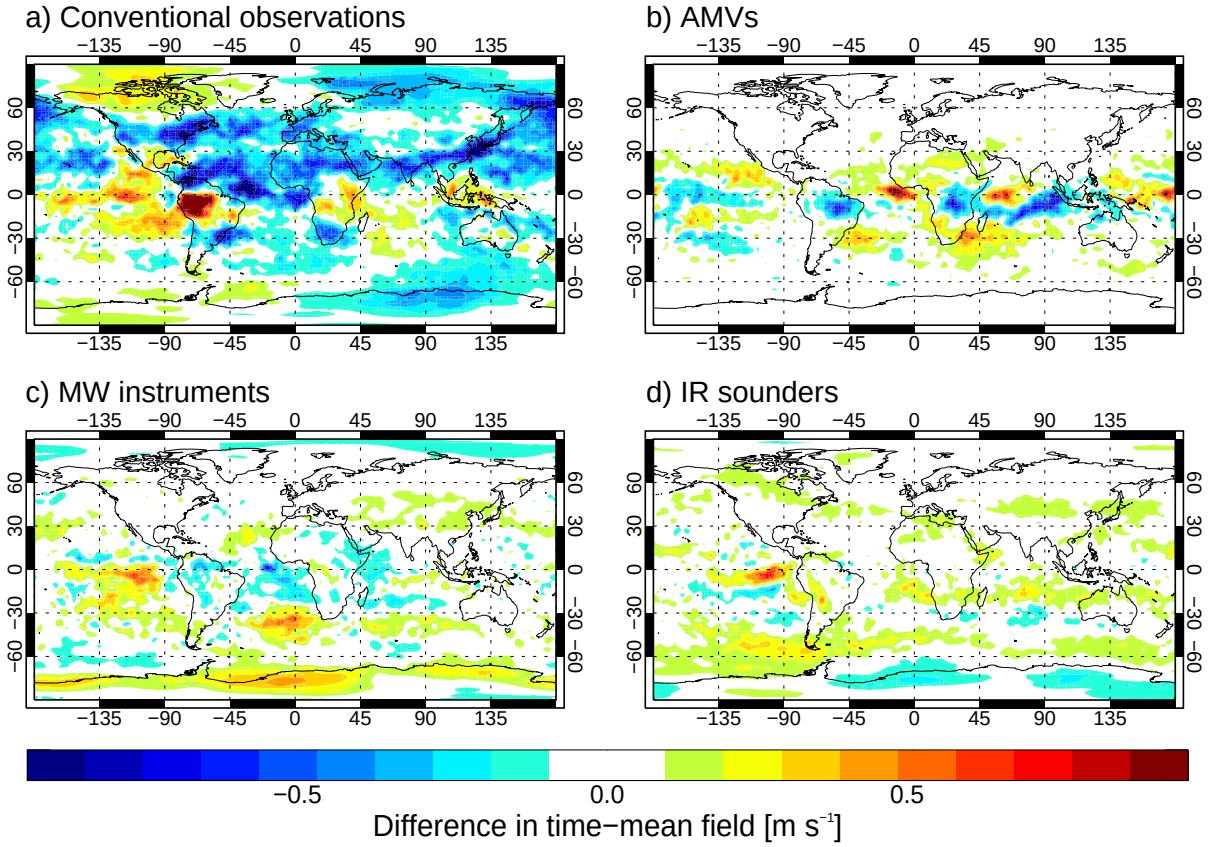


Figure 14: a) Map of the difference in the mean analysis of zonal wind at 200—hPa between the experiment with conventional observations denied versus the Control, for the period December 2017 - March 2018. b) As a), but for the AMV denial experiment. c) As a), but for the experiment without MW radiances. d) As a), but for the experiment without IR sounder data.

Mean analyses of geopotential also show considerable changes, especially for the conventional data, with values of several gpm for large parts of the troposphere. To put these into context, differences of these magnitudes are not uncommon when comparing mean analyses of geopotential from different NWP centres (e.g., McNally 2014), giving another estimate of the accuracy of NWP analyses in terms of bias. The differences are linked to changes in the mean analysis of mean sea-level pressure (MSLP) caused by the conventional observations which reach $\pm$ 1 hPa (Fig. 13) over the tropics and high-latitudes, respectively. For all other observing systems, changes in the MSLP are below 0.1 hPa for most of the globe, and the changes in the mean geopotential analyses reflect only the temperature changes discussed earlier.

Finally, very significant changes in the mean zonal wind analysis in the troposphere can also be seen for the various observing systems (with the exception of GPSRO), with conventional observations again causing some of the largest differences (e.g., Fig. 14). At 200 hPa, conventional observations lead to changes reaching almost 1 m/s in some places, particularly around the sub-tropical jet, but also in the mid-latitudes. The large changes in the mean wind analysis will obviously contribute to the apparent degradation from conventional data seen earlier in the own-analysis verification (Fig. 3). The four observing systems show relatively little consistency in the mean zonal wind changes, making it harder to attribute these to a forecast model or observation bias.

As indicated before, some changes to the mean analyses can be linked to the treatment of biases with VarBC, and further evidence of this can be seen in the variational bias corrections applied. An example of this are the bias corrections applied to AMSU-A data shown in Fig. 15. With the exception of AMVs, all other observing systems have some effect on the global mean bias correction solution VarBC applies, particularly for the stratospheric channels (channel 9 and above), but in the case of the conventional observations also for the tropospheric channels (5-8). GPSRO and radiosonde observations are not subject to variational bias correction (though offline bias correction is applied to day-time radiosonde measurements), and their anchoring effect on VarBC is well known (affecting the GPSRO and conventional experiments). The influence from the IR sounder data is interesting, given the IR sounder radiances are included in the variational bias correction, so are not anchoring observations in the classic sense. The changes introduced by the IR data are likely a result of adding vertical resolution, combined with interactions with model bias as mentioned earlier, and potentially observational biases. While these overall mean changes in the bias corrections applied are clearly notable, it should also be mentioned that they are all below 0.2 K and hence relatively small compared to the variability of bias corrections applied to AMSU-A data for different instruments (which can be ± 1 K).

Given the strong changes to mean analyses for the experiment without conventional observations, the role of the anchoring effect on the variational bias correction has been investigated further: to do so, another experiment has been conducted without conventional observations, but with bias corrections taken from the analysis of the Control experiment and kept fixed for each assimilation cycle. This latter experiment hence still uses the conventional observations to anchor the variational bias corrections, but it excludes the direct effect on the analyses of having the conventional observations present during the assimilation. Fig. 16 shows that when the anchoring effect is still present, the differences to the Control in terms of

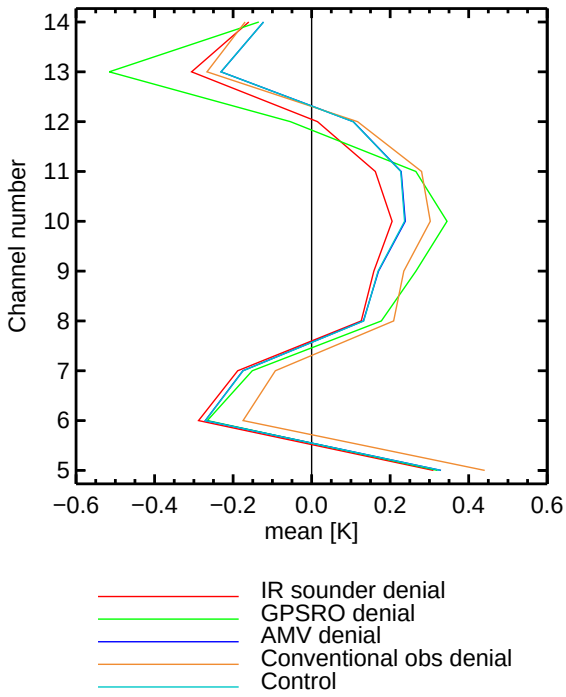


Figure 15: Global mean bias correction applied to all AMSU-A instruments assimilated for the Control and the denial experiments as indicated in the legend. Results are for the period 10-30 March 2018, ie the end of the winter experiment.

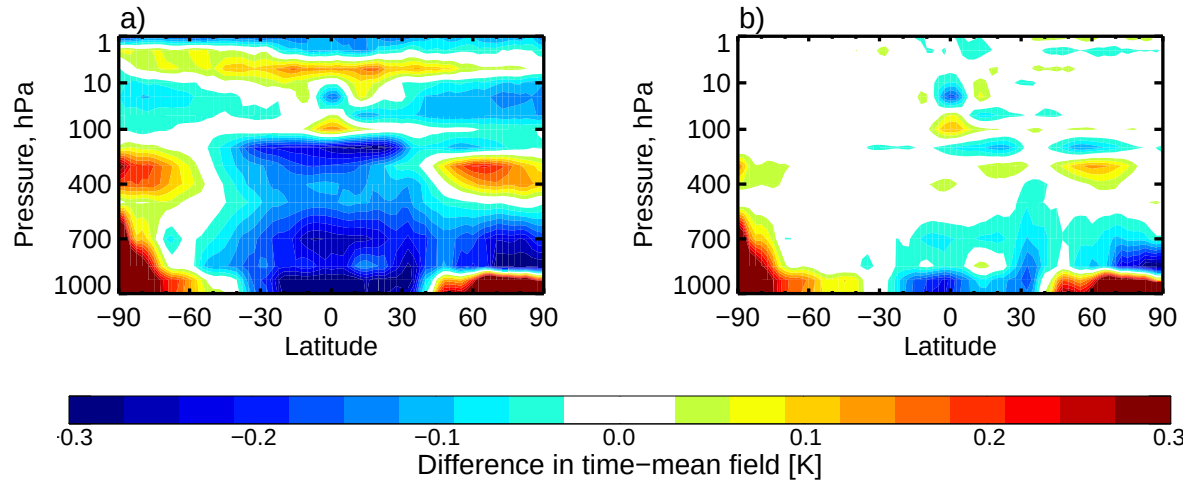


Figure 16: a) Zonal mean differences of the temperature analysis of the experiment with conventional observations denied and the Control for the period December 2017 - March 2018. Red colours indicate a warming resulting from the denial of the observations. b) As a), but for the conventional observation denial in which the variational bias corrections are inherited from the Control experiment.

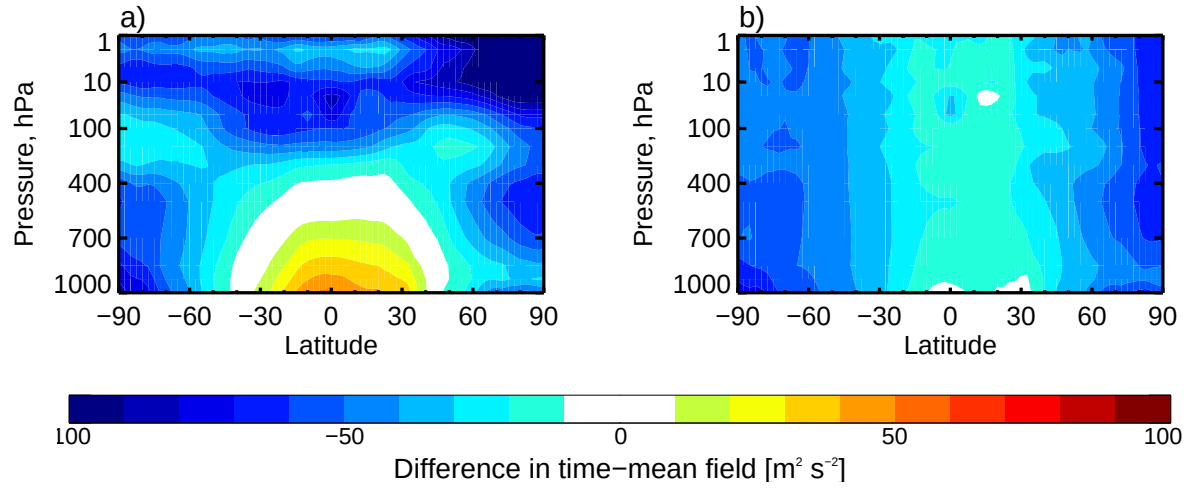


Figure 17: As Fig. 16, but for the zonal mean differences in the geopotential analyses.

the mean temperature analyses are considerably smaller in the mid troposphere to stratosphere, and the anchored bias corrections have a strong effect on the structure of the mean temperature analysis (aided also by the presence of GPSRO data). In contrast, at lower levels the anchoring effect plays a minor role, and most of the changes to the mean temperature analysis are still present even when the anchored bias correction is used. As a result, the changes in the mean geopotential analysis with the inherited bias corrections show a more zonal structure than for the standard conventional observation denial (Fig. 17). The results highlight that both the direct effect of having observations present in the analysis as well as the indirect effect of anchoring bias corrections play an important role in how conventional observations affect the mean temperature and geopotential analysis. Interestingly, the direct effect appears to be the dominant factor for wind analyses, and the differences in the mean wind analyses are fairly similar for the two conventional observation denials (not shown). While VarBC is not used for wind observations, the

indirect effect could still play a role through adjustments to temperature or humidity, but it appears that this effect is small. Standard deviations of forecast errors for the two conventional observation denials are relatively similar, with only minor benefits for the denial experiment with the inherited bias corrections, suggesting that the anchoring effect is only a minor aspect of the forecast influence of conventional observations (not shown).

It should be noted here that for most of the changes in the mean analyses discussed here, the differences introduced by each observing system appear to either persist throughout the forecast range, or to diminish with forecast range (as the model slowly reverts to its own climatology), and there is no indication that the biases grow during the forecast range. The role of the changes in the mean analyses in contributing to (random) forecast error growth is somewhat unclear.

5 Conclusions

The present memorandum has summarised results from recent observing system experiments with the ECMWF assimilation system. Experiments were conducted over a total of 8 months, covering a summer and a winter season, and observing systems considered include all conventional observations, all microwave observations, hyperspectral infrared radiances, bending angles from GPSRO, and AMVs. The main findings are:

- All observing systems considered here provide significant positive impact for at least some aspects of the NWP system. The results confirm the overall complementarity of the global observing system.
- Conventional observations and microwave radiances are the main drivers of headline scores. IR sounders add further robustness for a wide range of geophysical variables. GPSRO gives significant impact in the upper troposphere/lower stratosphere, mainly on temperature, but also other variables, and the data have a clear influence on the mean state in these regions. AMVs add benefits for tropospheric wind, particularly in the tropics and at the short range.
- As expected, conventional observations show the largest impact over the Northern Hemisphere, but strong impacts were also found in the tropics, which appears remarkable given the low data numbers. Of the observing systems considered, conventional observations have the largest impact on the mean state of the analysis.
- There are considerable seasonal variations in the impact of the considered observing systems. For instance, over the Northern Hemisphere, the impact of the MW data is strongest in summer, when it reaches the impact of all conventional observations.

The results presented here are broadly in line with results from similar observing system experiments conducted at ECMWF by McNally (2014), at least in terms of identifying the leading contributors to medium-range forecast impact over broad hemispheric regions. A strict comparison is not possible, given the different experimentation periods and some different choices in what observing systems to consider. Both studies demonstrate the complementarity of the observing system and the importance of the conventional as well as the satellite observing system.

The present study highlights the particularly strong impact of the MW observations, and it is clear that MW observations provide the largest impact of all the satellite data considered here. This large impact is

likely the result of several factors: firstly, an unprecedented number of instruments is being assimilated, with data presently available from 18 sensors, providing good spatial as well as temporal coverage. This appears to be a clear advantage over the IR sounder data, which is available from only 4 instruments for the study periods (albeit in much greater data numbers), with much more limited spatio-temporal sampling. Secondly, the MW data are used in a wide range of conditions, including the all-sky use of the majority of humidity-sensitive radiances, as well as the use of surface-sensitive data over land and sea-ice, as well as over oceans. Even the clear-sky use of MW temperature-sounding data allows sounding in some cloudy conditions, due to the lower sensitivity of MW data to clouds. Cloudy regions tend to be important for forecast error growth, so this wider geophysical sampling is a particular advantage for the MW. In contrast, IR observations are more strongly affected by clouds and do not provide information below clouds, and the use of cloud-affected IR observations is more challenging. Presently only data unaffected by clouds and a limited sample of overcast scenes are being assimilated. Clouds and other aspects also contribute to a more challenging observational error budget in the IR. Our results also show that both sounding as well as imaging channels contribute to the strong overall impact of the MW data. It is clear that a loss of MW sensing capabilities resulting, for instance, from the loss of protection for key frequencies would lead to significant degradations in the skill and reliability of numerical weather forecasts.

The seasonal dependence of some of the impacts found are very notable, with conventional observations showing stronger impact over the winter hemispheres and microwave data showing larger impact over the Northern Hemisphere summer. The reasons for this could be explored further, but contributing factors are likely different data usage over the two seasons as well as different meteorological regimes. The smaller impact of the microwave data for the Northern Hemisphere winter is likely to be linked to problems with using the data over snow and sea-ice (e.g., Bormann et al. 2017), but other factors may also play a role. It is clear that the seasonal dependence needs to be taken into account in the design of observing system experiments, in order to avoid un-representative results.

The observing system experiments also reveal how different observations affect the mean analysis, with particularly strong influence from the conventional observations. The withdrawal of the conventional observations leads to notable differences in the tropospheric mean analyses, including for temperature and wind. This appears to be the result of a complex interplay between the direct effect of assimilating observations that indicate a different mean state, as well as indirect effects through affecting the bias corrections of satellite radiances through VarBC. It is beyond the scope of the present memorandum to independently validate these mean changes to the analysis, but they may point to biases in the forecast model or uncorrected observational biases which could be investigated further.

References

Bonavita, M., L. Isaksen, and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. Technical memorandum 664, ECMWF, 31pp.

Bormann, N., A. Fouilloux, and W. Bell, 2013: Evaluation and assimilation of ATMS data in the ECMWF system. *J. Geophys. Res.*, **118**, 12,970–12,980, doi: 10.1002/2013JD020325.

Bormann, N., C. Lupu, A. Geer, H. Lawrence, P. Weston, and S. English, 2017: Assessment of the forecast impact of suface-sensitive microwave radiances over land and sea-ice. *ECMWF Technical Memorandum*, 804.

Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4-DVAR assimilation system. *Q. J. R. Meteorol. Soc.*, **127**, 1469–1488.

Cardinali, C., 2009: Monitoring of the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.*, **135**, 239–250.

Chen, K., S. English, N. Bormann, and J. Zhu, 2015: Assessment of FY-3A and FY-3B MWHS observations. *Weather and Forecasting*, **30**, 5, 1280–1290, 10.1175/WAF–D–15–0025.1.

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, Reading, UK, 97–112.

English, S., et al., 2018: Radio-frequency interference (RFI) workshop. ECMWF, https://www.ecmwf.int/en/learning/workshops/radio–frequency–interference–rfi–workshop.

Eyre, J., 2016: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties. *Quarterly Journal of the Royal Meteorological Society*, **142**, 699, 2284–2291, doi: 10.1002/qj.2819.

Geer, A., F. Baordo, N. Bormann, S. English, M. Kazumori, H. Lawrence, P. Lean, K. Lonitz, and C. Lupu, 2017: The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Q. J. R. Meteorol. Soc.*, **143**, 3189–3206. doi:10.1002/qj.3172.

Geer, A., P. Bauer, and P. Lopez, 2010: Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment. *Q. J. R. Meteorol. Soc.*, **136**, 1886–1905.

Geer, A. J., 2016: Significance of changes in medium-range forecast scores. *Tellus A: Dynamic Meteorology and Oceanography*, **68**, 1, 30229.

Goddard, J., P. Laloyaux, S. Lang, and M. Leutbecher, 2017: How to deal with model error in data assimilation. *ECMWF newsletter*, **153**, 14–15.

Han, W., and N. Bormann, 2016: Constrained adaptive bias correction for satellite radiance assimilation in the ecmwf 4d-var system. *ECMWF Technical Memorandum*, 783.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. Technical memorandum 636, ECMWF, 45pp.

Kelly, G., and J.-N. Thépaut, 2007: Evaluation of the impact of the space component of the global observing system through observing system experiments. *ECMWF Report*, 90pp.

Langland, R., and N. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, **56(3)**, 189–201.

Lean, K., and N. Bormann, 2019: Moving to GOES-16: a new generation of GOES AMVs. EUMETSAT/ECMWF Fellowship Programme Research Report 49, ECMWF, Reading, U.K., 21 pp.

McNally, A., 2009: The direct assimilation of cloud-affected satellite infrared radiances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.*, **135**, 12141229, doi: 10.1002/qj.426.

McNally, A., 2014: The impact of satellite data on NWP. In *Seminar on Use of Satellite Observations in Numerical Weather Prediction, 8-12 September 2014*, Shinfield Park, Reading.

McNally, T., M. Bonavita, and J.-N. Thépaut, 2014: The role of satellite data in the forecasting of hurricane sandy. *Monthly Weather Review*, **142**, 2, 634–646, 10.1175/MWR–D–13–00170.1.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143–1170.

Radnóti, G., P. Bauer, A. McNally, and A. Horányi, 2012: ECMWF study to quantify the interaction between terrestrial and space-based observing systems on numerical weather prediction skill. *ECMWF Technical Memorandum*, 679, 98pp.