ESA Contract Report

PEARL Cloud contract 4000128669/19/NL/CT

Contract Report to the European Space Agency

Optimising impact of radar reflectivity and lidar backscatter observations on forecasts

WP-3000 report

Preparations for EarthCARE Assimilation - Radar and Lidar Cloud Observations (PEARL Cloud)

Authors: Mark Fielding and Marta Janisková

Contract officer: Tobias Wehr

December 2021

ABSTRACT

To study and optimize the impact of cloud radar and lidar observations on numerical weather prediction (NWP) forecasts, Four-Dimensional variational (4D-Var) assimilation experiments have been performed using CloudSat cloud radar reflectivity and CALIPSO lidar backscatter observations. The work is an extension of WP-2000 within the PEARL Cloud project, which optimised the impact of assimilating cloud radar and lidar observations on NWP analyses. In addition to the three-month assimilation period considered in WP-2000, this work-package increases the number of assimilation cycles considered to up to 11 months, by combining four independent assimilation experiments. The extended period also allows the seasonality of the observations and model to be considered and the appropriateness of a bias correction scheme trained on one-months data to be tested.

Results indicate that assimilating CloudSat radar reflectivity and CALIPSO lidar backscatter has a positive impact on the globally-averaged forecast skill of large-scale variables such as temperature, wind and humidity at both short- and medium-range timescales. The impact is significant for the forecast of upper-troposphere temperature at short-range forecasts, and for other variables, such as vector wind, at forecast lead-times between 4-6 days. The positive impact is also verified against independent observations of top-of-atmosphere radiation fluxes, where significantly positive reductions in the fit to CERES daily average short-wave and long-wave radiation are found at forecast lead times between 1 and 5 days. Further positive impacts are possible when the sub-optimal use of lidar backscatter in one of the experiments is fixed and the reasons for regional degradations in forecast skill, such as in mid-level temperature and wind forecasts in the North Eastern Pacific stratocumulus to cumulus transition zone, are investigated.

Contents

1 Introduction

Recent advances in the performance of numerical weather prediction (NWP) data assimilation systems have been attributed to the greater use of observations related to clouds (Geer et al., 2017). At any one time, around 70 % of the world's surface is covered with cloud (Stubenrauch et al., 2013), therefore the ability to fully utilise the diverse information content on the current atmospheric state that clouds contain to improve weather forecasts is highly desirable. Currently, the only satellite observations related to clouds that are routinely assimilated are microwave radiances. While including these radiances has had a large beneficial impact on NWP forecasts Geer et al. (2018), they cannot resolve the vertical structure of clouds and precipitation. Active observations, such as those from radar and lidar, could fill the information-content gap and remove any ambiguities in the minimization. Further, as NWP systems resolve finer convective-scale features, higher resolution observations will be required to produce the best possible analyses.

This report describes results from the third work-package of the preparations for EarthCARE assimilation cloud (PEARL cloud) project. The main goal of the PEARL cloud project is to prepare the ECMWF assimilation system to monitor and assimilate observations from the Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE) satellite as soon as possible after launch. The previous assimilation project, (Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, Janisková and Fielding, 2018) focused on developments towards direct assimilation and monitoring to exploit cloud radar and lidar data for their assimilation in NWP models and demonstrated the feasibility to assimilate such observations using CloudSat (Stephens et al., 2002) and CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations Winker et al., 2009) data. The previous work-packages of the PEARL cloud project updated the assimilation system from the previous project (WP-1000) and optimised the impact of these radar and lidar observations on the analysis through an improved observation operator, screening and observation error characterisation (WP-2000).

In this work package, the evaluation in WP-2000 (Fielding and Janisková, 2020a) is extended to examining and optimising the impact of assimilating cloud radar and lidar observations on NWP forecasts within a Four-Dimensional Variational (4D-Var) data assimilation of system. Following the successful update of the developments from CY43R1 to CY46R1 (See WP-1000, Assimilation system adaptation and maintenance, Janisková and Fielding, 2020a, for more details), this work-package investigates the optimisation of the impact of assimilating cloud radar and lidar observations on forecasts using model cycle CY46R1 of the ECMWF Integrated Forecasting System (IFS). In addition to the three-month assimilation period considered in WP-2000, this work-package increases the number of assimilation cycles considered to up to 11 months, by combing four independent assimilation experiments. The extended period also allows the seasonality of the observations and model to be considered and the appropriateness of a bias correction scheme trained on one-months data to be tested.

The report is organised as follows. Section 2 provides the methodology for optimising the impact of assimilating radar and lidar observations on the subsequent forecast and also a description of the various observations used. Next, section 3 discusses the findings of the passive observation monitoring carried out using model cycle CY46R1 and provides an updated bias correction scheme for CALIPSO lidar backscatter. The main results are divided into two sections. Firstly, section 4 provides an evaluation of the impact of radar and lidar observations on forecasts by verification against own analysis, then section 5 evaluates the impact of the radar and lidar assimilation by comparing forecasts to independent observations. A summary and conclusions is provided in Section 6.

2 Methodology

This section describes the methodology taken for optimising the impact of assimilating radar and lidar observations on the subsequent forecast. It begins by describing the model and data assimilation system configuration used in the data assimilation experiments, before detailing the different observation types used. The section concludes by defining the experimental configuration used in this study.

2.1 Model and data assimilation system configuration

While the WP-2000 experiments (Fielding and Janisková, 2020b) were performed using both cycles 43r1 and 46r1, this work-package focuses solely on cycle 46r1 of ECMWF's assimilation and forecasting system. Cycle 46r1 was operational between June 2019 and July 2020 and is used in this work-package as it was the most up-to-date cycle when this work-package began that contained all the developments and optimisations to the observation operators and other components of the assimilation system that were made in WP-2000. However, the structure of the forecast system itself is similar in all recent cycles. Global analyses of large-scale variables such as temperature, humidity and wind are first produced by the data assimilation system. The data assimilation system combines short-term forecasts made with the ECMWF global atmospheric model with observations within a 12 h time window using an incremental formulation of 4D-Var (Courtier et al., 1994; Rabier et al., 2000). The atmospheric model, used for both the data assimilation and subsequent forecasts, has a semi-Lagrangian, spectral formulation with 137 levels from the surface to an altitude of ~ 75 km. In the configuration used for our experiments, the high resolution trajectory (used to calculate the departures between the model and the observations) has a horizontal resolution of TCo639 (corresponding to approximately 18 km on a cubic octahedral grid). To account for some non-linearity in the model and observation operators, the minimization consists of three inner loops all at a reduced resolution: the first at spectral resolution with truncation T95 (corresponding to ~ 200 km), the second at T159 (~ 125 km) and the third at T255 (~ 80 km). The minimizations are performed with the model containing simplified physical parameterizations of cloud and large-scale moist processes, convection, radiation, vertical diffusion, orographic and non-orographic gravity wave drag (Janisková and Lopez, 2013). More details can be found in WP-1000 (Janisková and Fielding, 2022) and WP-2000 (Fielding and Janisková, 2020b).

2.2 Radar and lidar assimilation configuration

The assimilation system for radar and lidar observations uses the same configuration as the 'updated' version as described in WP-2000 (Fielding and Janisková, 2020b), which employed the 'triple-column' approach to efficiently represent sub-grid inhomogeneity and cloud overlap. Only a brief summary of the assimilation system configuration will therefore be provided here. In terms of the observation operators themselves, which provide model equivalent to the observations, a combination of off-line look-up tables and on-line radiation integration through the model column is employed. The look-up tables for single-scattering, unattenuated lidar backscatter and radar reflectivity are generated as a function of model hydrometeor type, water content and temperature off-line given carefully chosen particle size distribution and particle types. Full details of the microphysical assumptions can be found in Fielding and Janisková (2020a). When the observation operator is run, this look-up table is searched, using a bilinear interpolation, for each model atmospheric profile and level where the model-equivalent radar reflectivity or lidar backscatter are required. To reduce computational cost, a parameterized version of the look-up table using a two-variable two-degree polynomial fit is used. Gaseous attenuation is then accounted

for by integrating from the top model level to the bottom level and summing the extinction coefficients provided by the Liebe gas optics model.

The radiative integration step, which computes the final model-equivalent observations, known as the 'driver', uses the 'triple-column approach'. The triple-column approach, inspired by the Tripleclouds radiative transfer algorithm (Shonk and Hogan, 2008), accounts for both sub-grid condensate variability and cloud overlap, and is an extension of the double-column approach (that only accounts for cloud overlap between model layers). The approach uses three separate uniformly filled columns: one clear and two cloudy. The cloud water content is increased in one-column (the 'thick' column) and decreased in the other column (the 'thin' column) so that the mean radiation across the column is consistent with an assumed probability distribution function (PDF) of sub-grid water content variability. While a broadband radiative transfer algorithm has to represent both direct and diffuse radiation, for our radar and lidar observation operator we make a single scattering approximation (radiation does not scatter more than once in the radar or lidar beam) so we only need consider the 'direct beam' radiation. The method was shown to perform well when verified against an explicit multi-column approach (Fielding and Janisková, 2020a). For radar, profiles of radar reflectivity where multiple scattering is likely to be significant are detected using the integrated reflectivity (Battaglia et al., 2011) and excluded. For lidar, multiple scattering is approximated using a tuned version of the Platt approximation.

Observation errors are specified using an 'error inventory' approach that accounts for different sources of error. Following the approach described in Fielding and Janisková (2020b), the observation errors are found by combining the contributions from instrument error, forward model error and representativity error. Under the hypothesis of uncorrelated errors, the individual components are added in quadrature to define the total observation error. The error components themselves are likely to be correlated, for example there are expected to be significant vertical correlations in the representativity error (Fielding and Stiller, 2019). To account for the vertical correlations, the observation errors are inflated by a factor of 2. The magnitude of the error inflation is somewhat arbitrary, but this scaling was found to perform best in Janisková and Fielding (2020b). Accounting for the vertical correlations explicitly was found to lead to degradations in forecast skill (Fielding and Janisková, 2020a), possibly due to the sensitivity of such approaches to inter-level biases.

Quality control is a vital component of the observation pre-processing to prevent 'bad' observations from degrading the analysis. The initial quality control for the observations is based on first-guess (FG) departures. Additional quality control measures include thresholds for indicators of signal strength, estimated total attenuation and, for the radar, integrated radar reflectivity as a measure of the expected multiple scattering. The bias correction scheme is based on a climatology of FG departures when the observations were passively monitored and is discussed in more detail in Section 3.3. To provide an implicit regime dependence for the correction scheme, indicators of temperature and height are used by the scheme. Variational quality control (VarQC; Tavolato and Isaksen, 2015), which reduces the weight of large departures in the minimization, is turned off. Full details of the observation operators and observation handling are described in Fielding and Janisková (2020a).

2.3 Observations

2.3.1 Cloud radar and lidar observations

In this study, measurements of cloud radar reflectivity (in dBZ), from the CloudSat 94 GHz radar and lidar backscatter (in $km^{-1}\ sr^{-1}$ using a logarithmic scale) due to clouds at 532 nm from CALIPSO are assimilated in the 4D-Var system. Cloudy-only data is obtained using the level-2 (L2) CloudSat geomet-

ric profile product (Marchand et al., 2008) and the Cloud-Aerosol Lidar with Orthogonal Polarisation (CALIOP) L2 lidar 5km cloud profile data versions 3 (Winker et al., 2009; Young et al., 2013). Version 4 CALIOP data is used for data between June 2008 and August 2008, due to the retiring of the older version data. The observation horizontal resolution is approximately 1.4 km for CloudSat radar and it varies between 300 m to 1.7 km for CALIPSO lidar depending on altitude. To reduce horizontal correlations in the data and increase the representativity, the observations are spatially averaged (often called 'superobbing') to a TCo159 grid, which corresponds to a 61 km grid spacing at the equator.

2.3.2 CERES

To help evaluate the impact of radar and lidar observations on forecast skill, fully independent observations of the net top-of-atmosphere (TOA) short-wave and long-wave radiation from the Clouds and the Earth's Radiant Energy System (CERES; Wielicki et al., 1996) SYN1deg-Day Ed4A product are used. The SYN1deg-Day product uses a variety of radiance measurements combining instruments aboard geostationary and polar orbiting satellites. The observation data is fused with a diurnal cycle model to compute daily average fluxes.

2.3.3 MODIS

Another useful independent dataset are observations or products derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. In particular, the A-train contained a MODIS instrument aboard the AQUA satellite, so the data is particularly useful for investigating the impact of assimilating CloudSat and CALIPSO observations on nearby cloud properties. In this work-package, we use the AQUA MODIS Level-2 5km 5-Min Wide Swath product (MAC06S1), which restricts data to within 200 km cross-track of the A-train orbit. To compare model cloud properties with the MODIS cloud optical depth and cloud water path, for each superobbed CloudSat or CALIPSO profiles, MODIS retrievals are collocated by averaging all observations within a 40 km radius.

2.3.4 GPCC

Precipitation from point measurements are challenging to use for verification due to their inherent high spatio-temporal variability. Nevertheless, we attempt a basic verification of precipitation forecast skill by comparing against the Global Precipitation Climate Center (GPCC; Schamm et al., 2014) daily global land-surface precipitation product. The product interpolates daily rain gauge totals to a regular grid with a spatial resolution of 1 by 1 degree. In our verification, we do not use any GPCC observations that are derived from less than two rain gauge observations per day.

2.4 Experimental framework

Several assimilation experiments have been run using the full system of regularly assimilated observations at ECMWF. Experiments have been performed for a selection of three month periods to enable the long-term and seasonal impacts of assimilating radar reflectivity and lidar backscatter to be investigated:

- ASO (August 1st 2007 - October 31st 2007), uses superobbing to TCo159 grid, original bias correction scheme.

- DJF (December 1st 2007 - February 29th 2008), uses superobbing to TCo159 grid, updated bias correction scheme for lidar tilt.
- FMA (February 1st 2008 - April 30th 2008), uses superobbing to TL255 grid, original bias correction scheme.
- JJA (June 1st 2008 - August 31st 2008), uses superobbing to TCo159 grid, updated bias correction scheme for lidar tilt.
- COMBI (Combined results from ASO, DJF, FMA, JJA, but not including the first month of FMA to avoid a repeated month).

Note that while the FMA experiment uses superobbing to a TL255 grid, the grid spacing is similar so we do not expect significant differences in the impact of the observations. For all of these periods, both a control and an experiment simulation have been performed with the following configuration:

- **REF**: reference run, i.e run with all regularly assimilated observations, but without new cloud radar and lidar observations included in the 4D-Var system;
- **RADLID**: experiment assimilating cloud radar and lidar observations on top of all other normally assimilated observations.

In addition to the standard 'RADLID' configuration, the sensitivity to

- **1.8err** - 1.8 times the single observation errors as defined in Fielding and Janisková (2017a, 2020a);
- **2err** - observation errors twice as large as defined (baseline 'RADLID' configuration).
- **2.2err** - observation errors ten percent greater than the baseline configuration.

3   Monitoring of observations and correction of biases

Before actively assimilating any observations in an operational setting, it is crucial to monitor the observations passively in the assimilation system to ensure that the observations are as expected, and so that the observation pre-processing can be validated. The same applies for experimentally assimilated observations. In this section we demonstrate the monitoring of CloudSat radar reflectivity and CALIPSO lidar backscatter over a 12-month period between August 2007 and August 2008. We also detail adjustments to the bias correction scheme that were found to be required as a result of the observation monitoring.

3.1   CloudSat quality monitoring

To begin with we monitor the 'used' first-guess departures (those that pass screening, so would have been assimilated) of CloudSat radar reflectivity from the control experiments. The mean of all first-guess departures in each 12 hour assimilation window (Fig. 3.1a) are remarkably stable over the one-year period, with 95 % of global means after bias correction of between 0-1 dB. For reference, the dynamic range of radar reflectivity measurements from CloudSat is between -30 and 20 dBZ. The bias correction scheme also performs well, at least in a global sense, with the mean bias reduced from around 2.5 dB to 0.5 dB averaged across the time period. There is also little seasonal variation in the global mean departures before bias correction. The slight change in global mean bias between the DJF and FMA experiments are due to the slightly different superobbing grid (the coarser the superobbing resolution,

the greater the grid-box mean radar reflectivity tends to be due to sub-grid variability (Fielding and Janisková, 2020a)). Also of note is that the reduction in the number of CloudSat observations entering the system for the last two weeks of July 2008 manifests in a slight increase in the variability of mean departures.
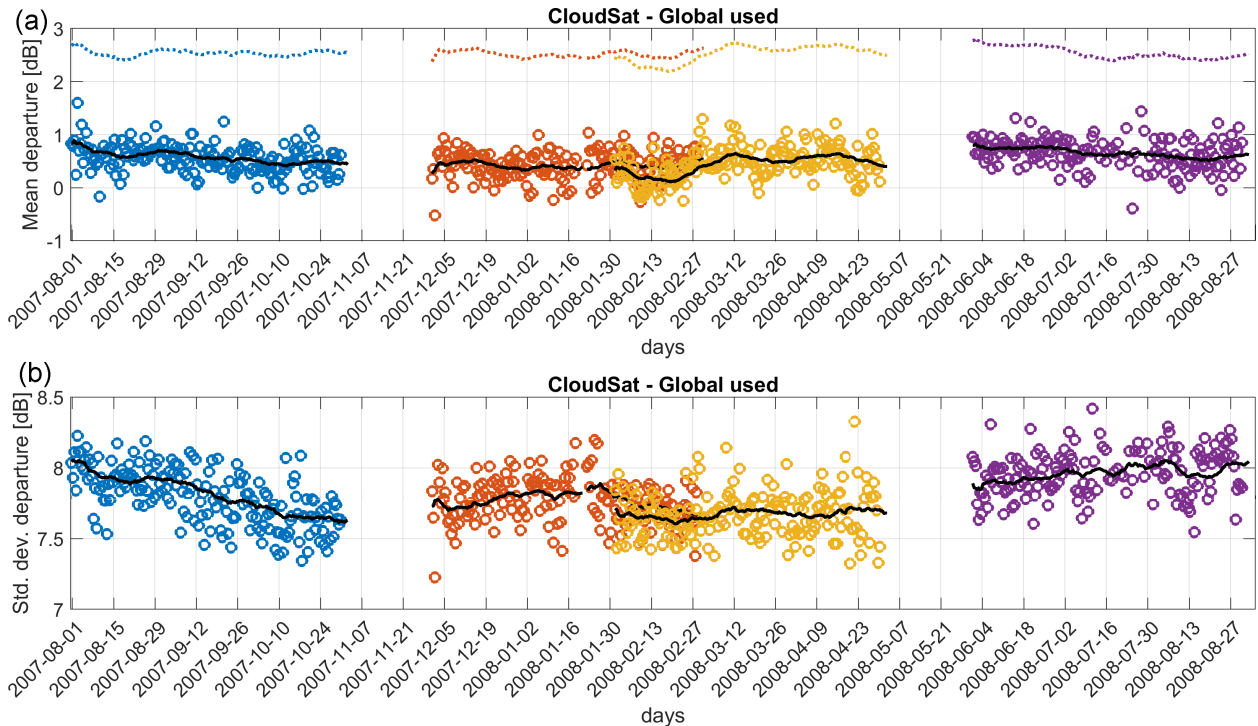

Figure 3.1: Passive monitoring of first-guess departures of CloudSat radar reflectivity. Plots show a) global mean departures for each 12-hour assimilation window, b) global standard deviation of departures for each 12-hour assimilation window. Solid black line shows a 7-day moving average, while dashed line shows a 7-day moving average before bias correction. The marker colours differentiate the different cycling assimilation experiments; see text for details.

The standard deviation of first guess departures gives a measure of the fidelity of the model to represent clouds and precipitation, and also the accuracy of the observation operators. Unlike the global mean bias, the global standard deviation of first guess departures (Fig. 3.1b) does show some seasonality. The greatest standard deviations are found at the start and end of the period, which coincides with Northern Hemisphere summer. The seasonal effects are more pronounced when looking at the data regionally (Fig. 3.3), where both the Arctic and North Hemisphere regions share the same signal. It is likely that the seasonality is due to the greater number of convective clouds and precipitation in summer months, which are more difficult for the model to represent. On the other-hand, the standard deviation of first-guess departures in the tropics are very stable throughout the year, with an average of around 9 dB. The Southern hemisphere statistics are also stable across the year, but the standard deviation is much smaller than the tropics, with an average of 7.5 dB. The smallest standard deviations of first-guess departures can be found in Antarctic winter, where the average standard deviation of first-guess departures is around 6 dB.

Figure 3.2 shows the bias in first-guess departures stratified by region. There is very little seasonal variation in the Southern Hemisphere and therefore the bias correction performs well for the whole period. As was seen for the standard deviation of first-guess departures, the northern hemisphere does

have some seasonal variation with biases slightly larger in summer months. This could be due to greater sub-grid variability in summer convective situations, which would tend to increase the superobs relative to the model. The polar regions show the greatest seasonal variability, with the superobs tending to be greatest in the polar region's summer months. The region with the least effective bias correction is in the Antarctic between December 2007 and April 2008, but it never exceeds more than 2 dB in the seven-day moving average.
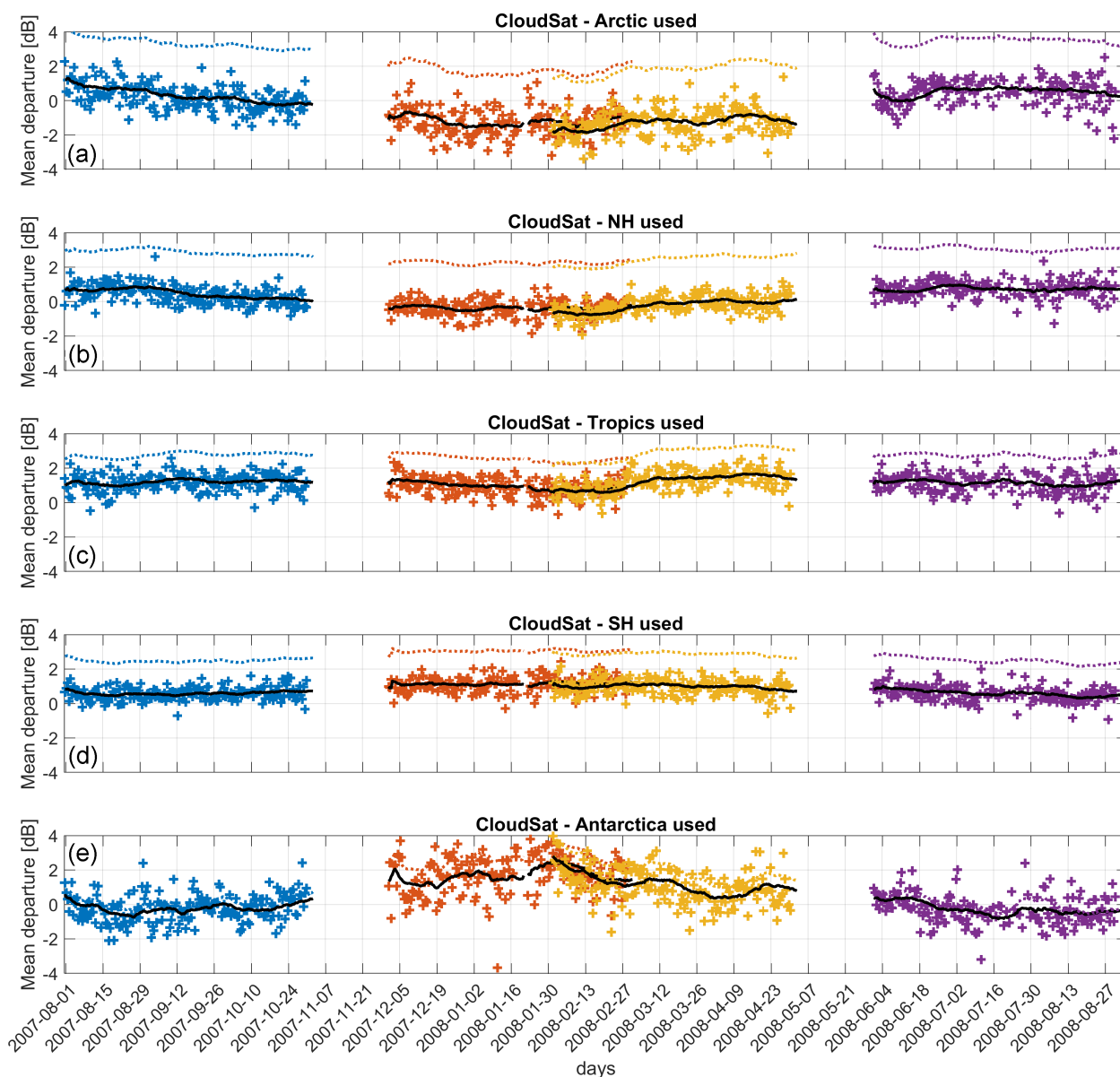

Figure 3.2: Same as Fig. 3.1, but stratified by regions. Plots show mean first-guess departures for each 12-hour assimilation window for the a) Arctic, b) Northern Hemisphere extra-tropics, c) Tropics, d) Southern Hemisphere extra-tropics, e) Antarctic.

It can also be informative to assess the first-guess departures as a function of height. Figure 3.4 shows the mean first guess departures as a function of atmospheric pressure. In a global sense, both before and after bias correction there is very little seasonal variation in mean first-guess departures when stratified by height. The bias correction also performs well, with mean first-guess departures after bias correction

less than 1 dB at all heights above 900 hPa. However, regionally, there are some variations. In general the bias correction works best for summer months (ASO and JJA experiments), but is also acceptable for the other experiments. The only occasions when the mean first-guess departures exceed 2 dB are for high clouds in the DJF and FMA experiments in the Arctic and for mid-level clouds in the DJF experiment. The standard deviation of first-guess departures as a function of height (Fig. 3.5) are also very similar for different seasons in a global sense. They tend to be greatest for low cloud and decrease with height. An exception to this rule is in the tropics, where the standard deviation of first-guess departures tends to be around 9 dB for all heights. Very high clouds in the extra-tropics show a slight increase in the standard deviation of first-guess departures, possibly because these clouds are likely to have a larger measurement error due to their likely lower signal to noise ratio.
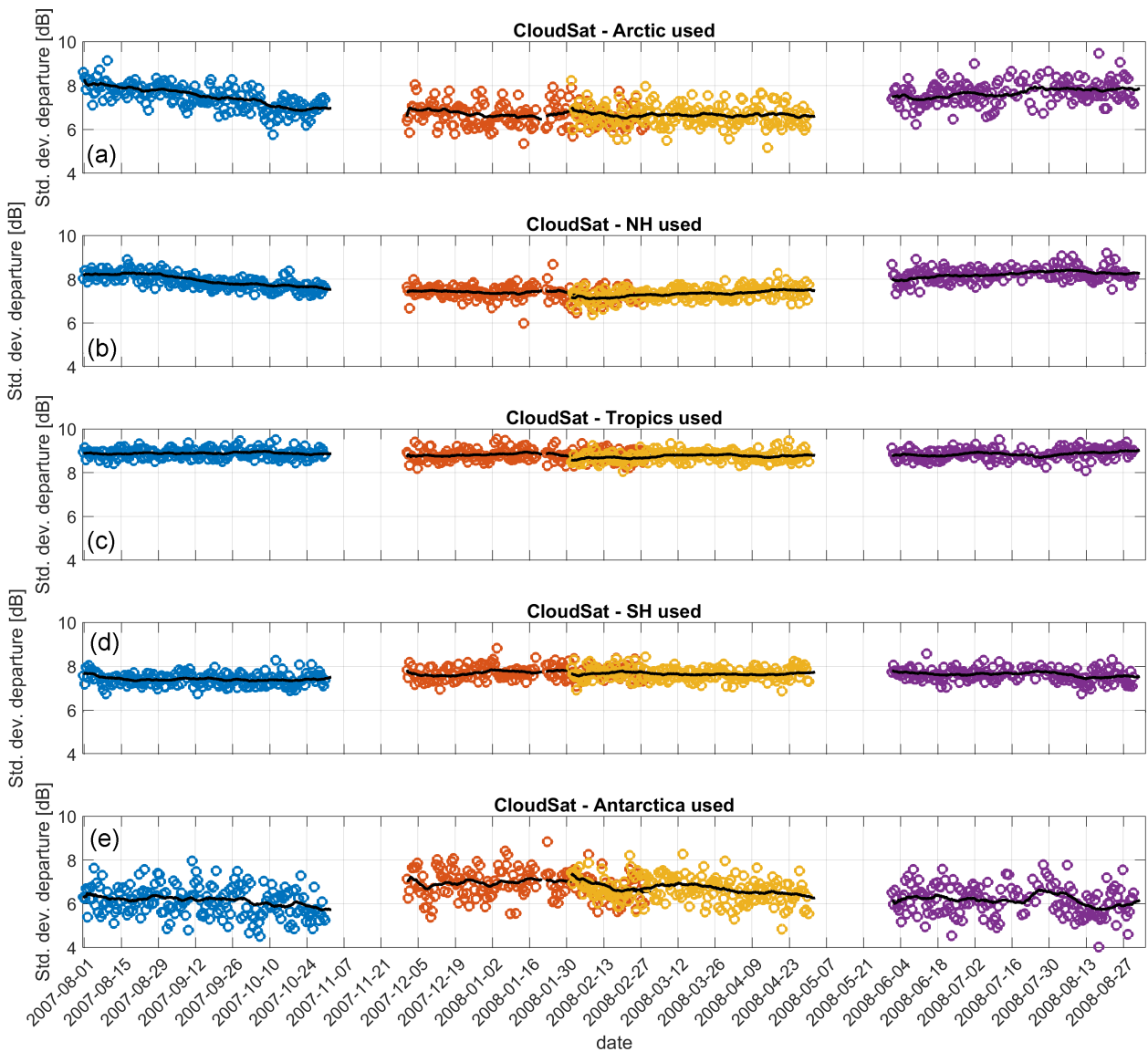

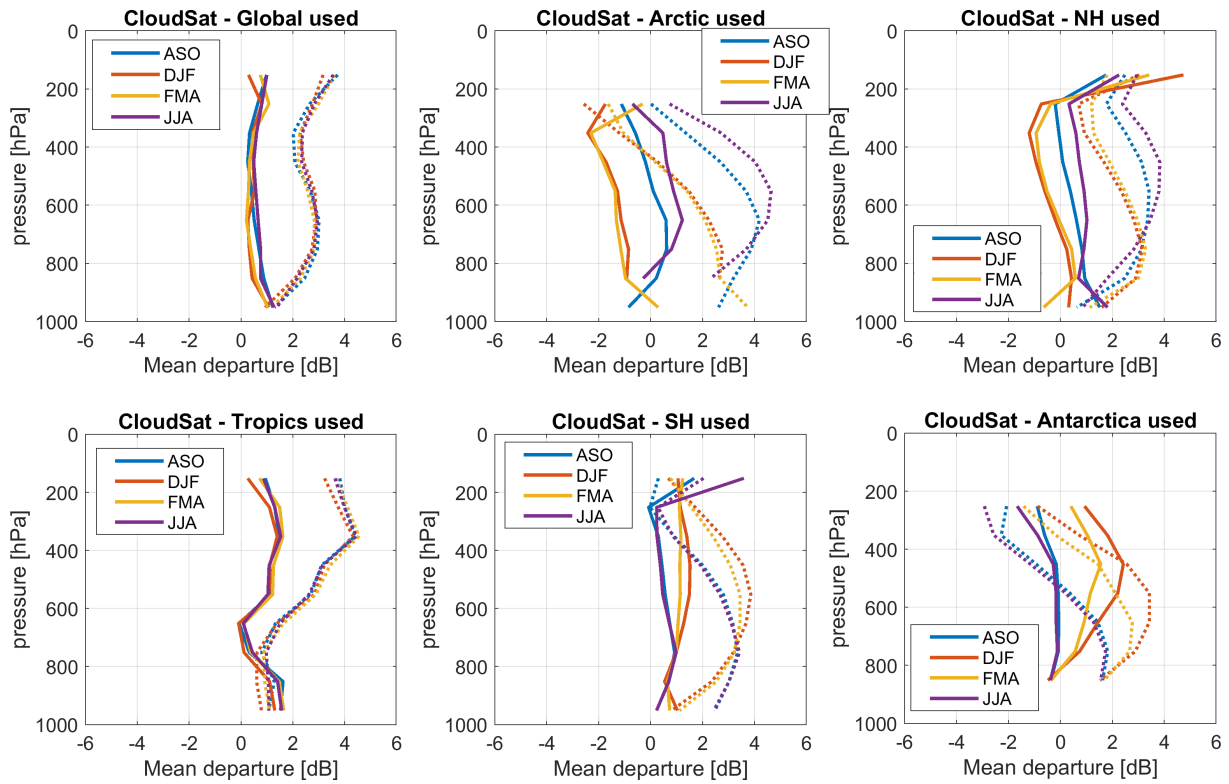Figure 3.3: *Same as Fig. 3.2, but for the standard deviation of first-guess departures.*

Figure 3.4: Mean first-guess departures for CloudSat radar reflectivity before (dashed line) and after (solid line) bias correction as a function of atmospheric pressure, for the four cycling assimilation experiments.
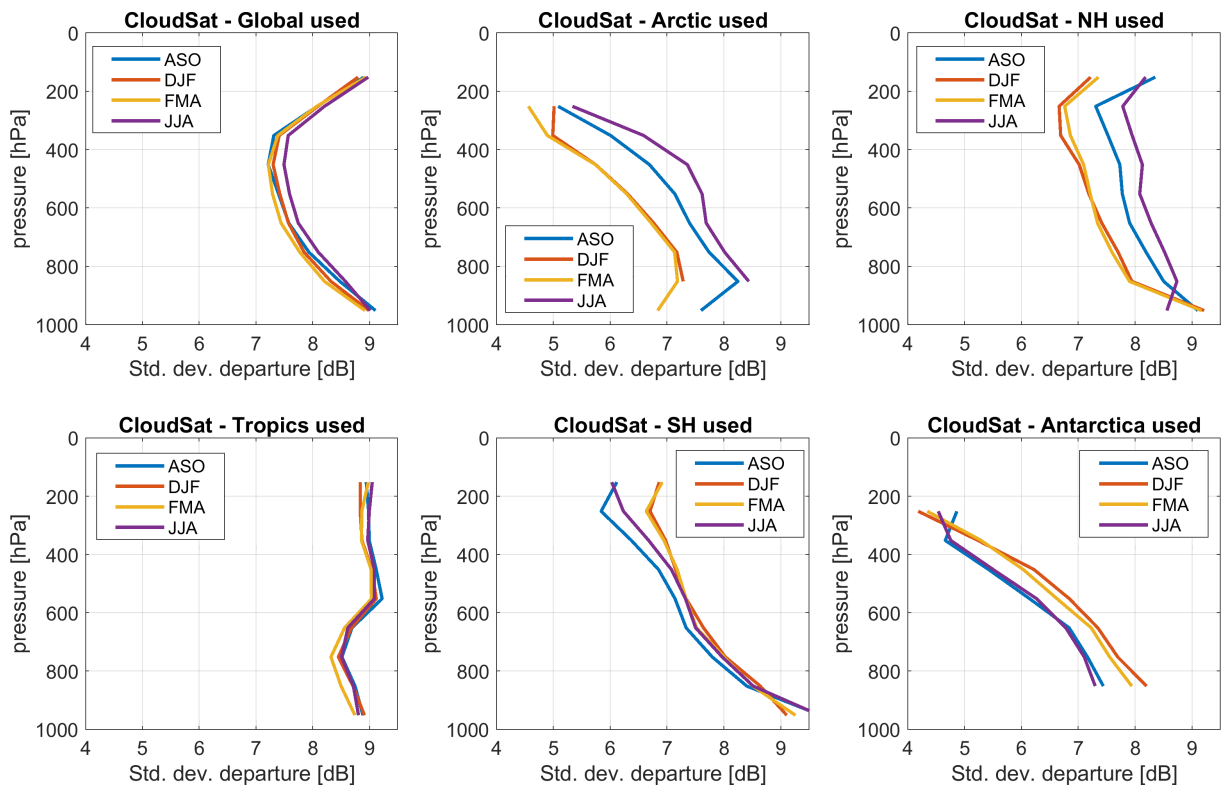

Figure 3.5: Same as Fig. 3.4, but the standard deviation of first-guess departures.

3.2 CALIPSO quality monitoring

The passive monitoring of CALIPSO lidar backscatter first-guess departures (Fig. 3.6) for the ASO experiment revealed a quality issue with the data that had not previously been noticed. The global mean first-guess departure can be seen to be around 1.3 dB for most of the experimental period, but drops to under 1 dB for the period between August 21st 2007 and September 7th 2007. The reason for the change in bias is due to the CALIOP lidar changing its tilt from 0.3 to 3 degrees during the period. Horizontally aligned ice crystals can cause enhanced backscatter due to specular reflection Yang et al. (2003), so, as expected, changing the pointing angle of the lidar to be off-nadir reduces the bias in the observations. A small reduction in the standard-deviation of first-guess departures can also be detected when the lidar was pointing off-nadir, suggesting that the specular reflection was also increasing CALIPSO's observation error. For the remaining experiments, the lidar was pointing at 3 degrees off-nadir, as CALIOP changed its tilt definitively from November 28th 2007 onwards.
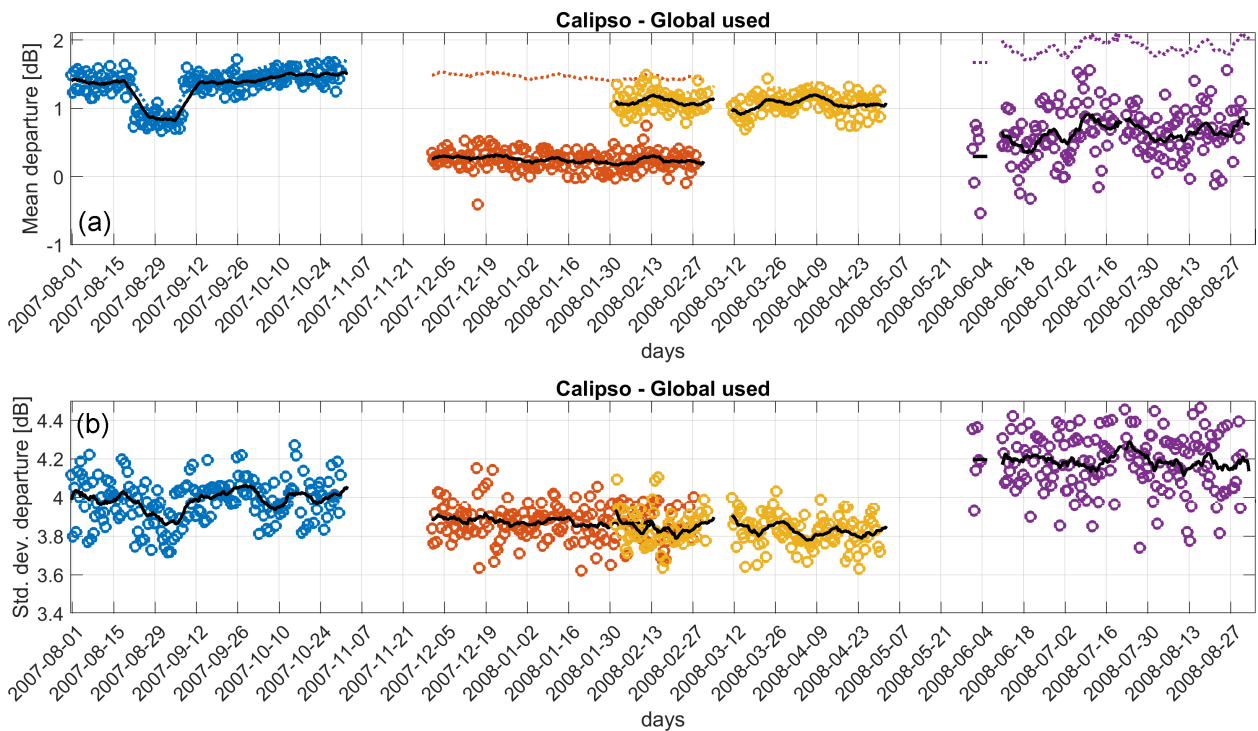

Figure 3.6: Passive monitoring of first-guess departures of CALIPSO lidar attenuated backscatter. Plots show a) global mean departures for each 12-hour assimilation window, b) global standard deviation of departures for each 12-hour assimilation window. Solid black line shows a 7-day moving average, while dashed line shows a 7-day moving average before bias correction. The marker colours differentiate the different cycling assimilation experiments; see text for details.

The FMA experiment was performed before the lidar tilt adjustment was noticed, so the global mean first-guess departures are also around 1 dB. A dropout in the lidar data is clearly visible for the first week of March 2008, which was due to the CALIPSO satellite experiencing a problem with the platform reaction wheel. According to NASA, the payload was down from March 03, 2008 at 23:18:30 UTC until March 11, 2008 at 12:36:22 UTC. A new bias correction scheme that was trained on CALIPSO data when the lidar was pointing off-nadir was implemented for the DJF and JJA experiments (see Sec. 3.3). For the DJF experiment, the new bias correction appears to work much better than for the ASO and FMA experiments, with the seven-day moving average in global mean FG departures hovering around 0.2 dB.

The standard deviation of first-guess departures is unaffected by the bias correction and is similar to the period when the lidar was tilted at 3 degrees in the ASO experiment.

Now considering the JJA experiment, there is clearly a problem with either the raw observational data, the observation processing, or the model clouds. The global mean first-guess departures are much more variable than for the other three experiments. Also, the standard deviation of first-guess departures is much larger and more variable than is expected. An investigation revealed that the observation processing was to blame; the experiment used a new version of the CALIPSO L2 5km cloud profile product, where the the number of vertical levels was increased from 345 to 399. As a result the vertical levels of the lidar data converted to BUFR and used in the assimilation system was offset. The radar data was unaffected, but it does mean results from assimilation experiments for the JJA period are contaminated by sub-optimal lidar data. The effect will be minimized by the quality control and screening, however improved results are expected when this period is re-run.

As was done for the monitoring of CloudSat observations, Figure 3.7 shows the mean first-guess departures for CALIPSO lidar backscatter that have been regionally subset. After bias correction, the DJF experiment has almost zero bias for all regions, indicating that the bias correction works very well. The change in bias due to the change in the CALIOP lidar pointing angle is most apparent in the Arctic, and to a lesser extent the northern hemisphere extra-tropics, where the summer atmospheric conditions are conducive to horizontally aligned ice crystals Noel and Chepfer (2010).A similar pattern in the seasonality of the standard deviation of first-guess departures (Fig. 3.8) to the monitoring of CloudSat can be seen, with the greatest variability in departures occurring in the summer months of each hemisphere. Both the extra-tropics and tropics regions have very similar magnitude of standard deviation of first-guess departures of around 4 dB.

Finally, we can consider the height dependence of the first-guess departures. In a global sense, before bias correction, the height dependence of the bias in first-guess departures is similar for all the four experiments (Fig. 3.9); the bias tends to be around 2 dB for mid- to upper-level clouds, and reduces to zero for low-level clouds. In agreement with Fig. 3.6, the bias correction of the DJF experiment performs the best. The bias correction for the DJF experiment also performs well regionally, with nearly all mean first-guess departures at all height levels being less than 1 dB. The effect of the change in lidar tilt on the mean first-guess departures can be seen by comparing the ASO and DJF experiments before bias correction. In the Arctic, the greatest difference in mean departures can be seen in mid-level clouds with pressure between 700 hPa and 500 hPa, where the ASO experiment has a bias of 3 dB, while the DJF experiment's bias is around 1 dB.

Excluding the JJA experiment, with its sub-optimal lidar data, the vertical structure of the standard deviation of first-guess departures is consistent through the different experiments (Fig. 3.10). The standard deviations are greatest for low-level clouds and smallest for high-clouds. When looking regionally, there is some seasonality to the standard deviations, but mainly in the Arctic and northern hemisphere extra-tropics. In these regions the standard deviations of first-guess departures are around 1 dB greater in the summer than the winter. Antarctic clouds between 400 hPa and 100 hPa have the smallest first-guess departures of around 3 dB, suggesting that the model performs well at predicting clouds in these conditions, or at least has the correct cloud amounts when CALIPSO observations detect clouds. This is in agreement with the standard deviation of first-guess departures for CloudSat observations for these types of cloud.

Figure 3.7: Same as Fig. 3.6, but stratified by regions. Plots show mean first-guess departures for each 12-hour assimilation window for the a) Arctic, b) Northern Hemisphere extra-tropics, c) Tropics, d) Southern Hemisphere extra-tropics, e) Antarctic
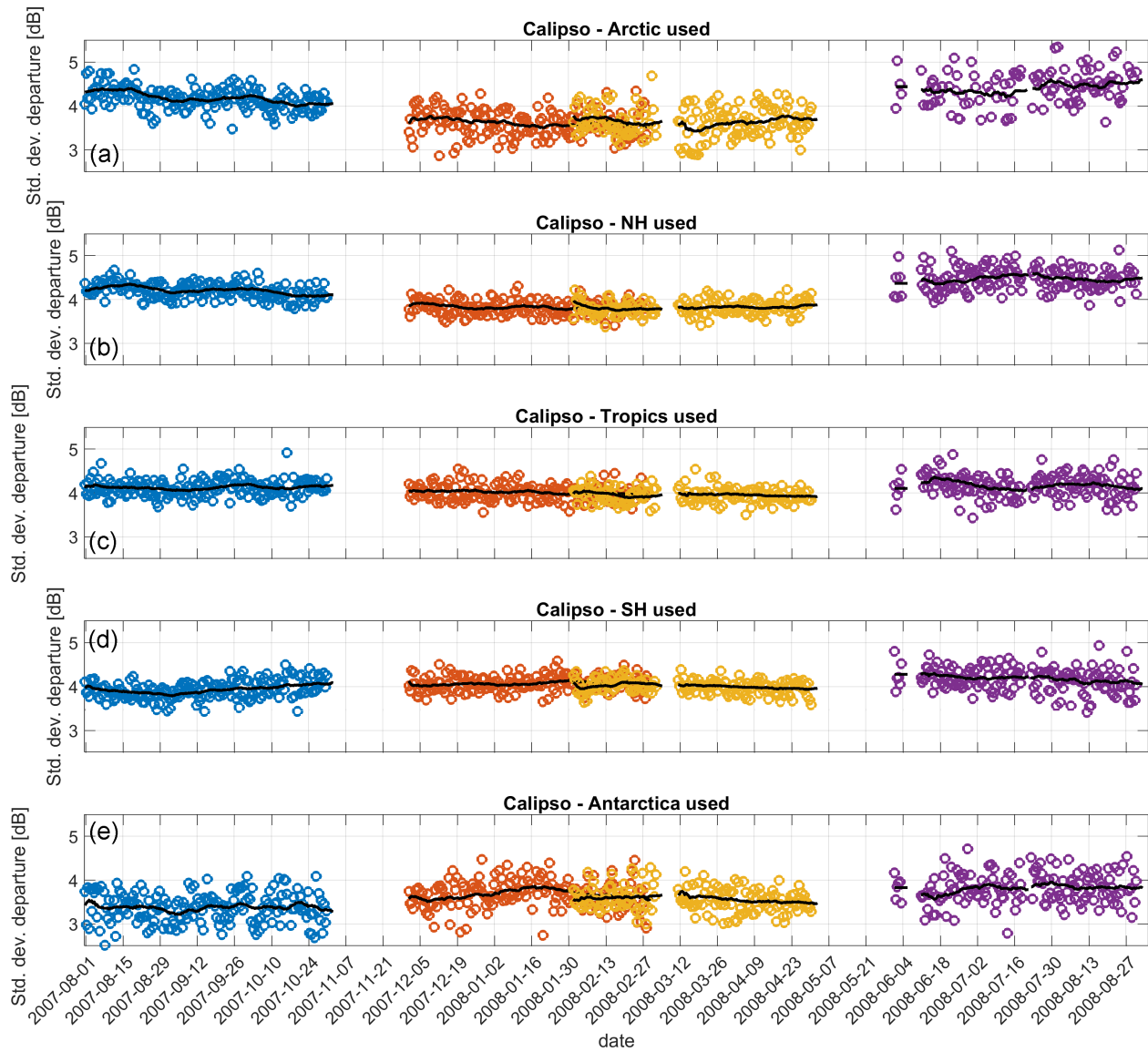
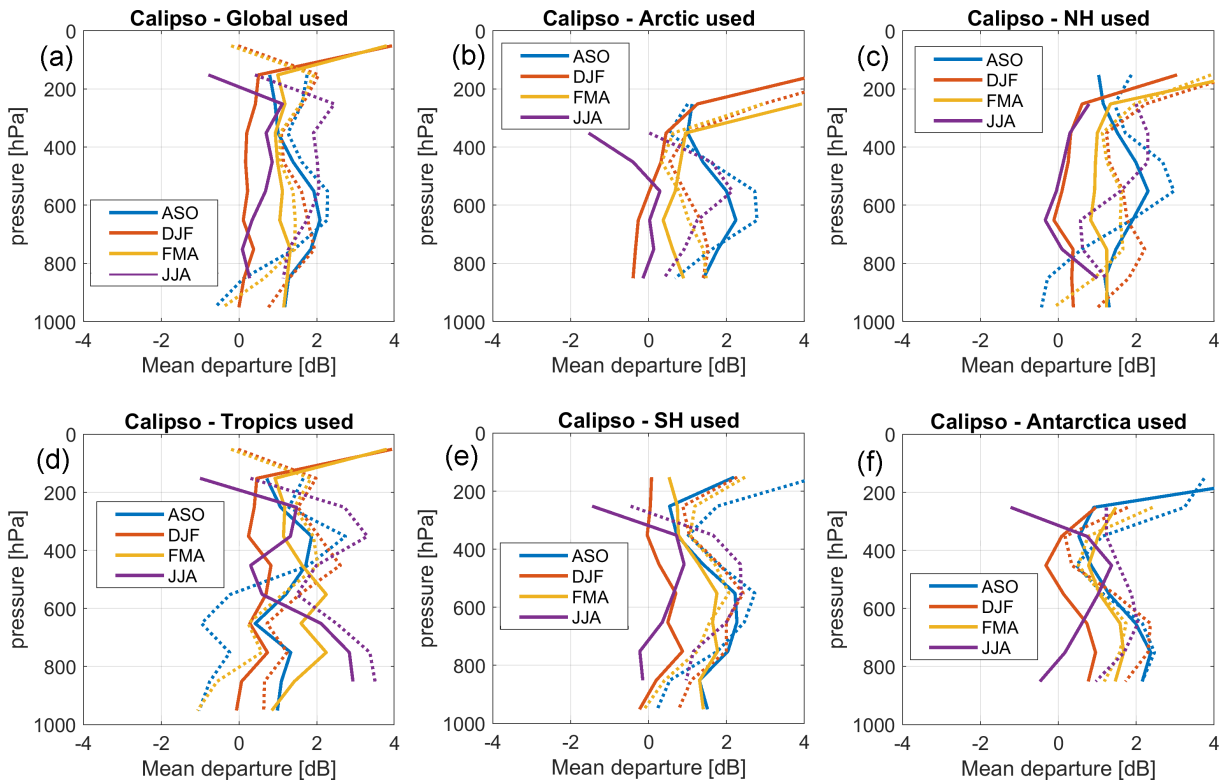*Figure 3.8: Same as Fig. 3.7, but for the standard deviation of first-guess departures.*

Figure 3.9: Mean first-guess departures for CALIPSO lidar attenuated backscatter before (dashed line) and after (solid line) bias correction as a function of atmospheric pressure, for the four cycling assimilation experiments.
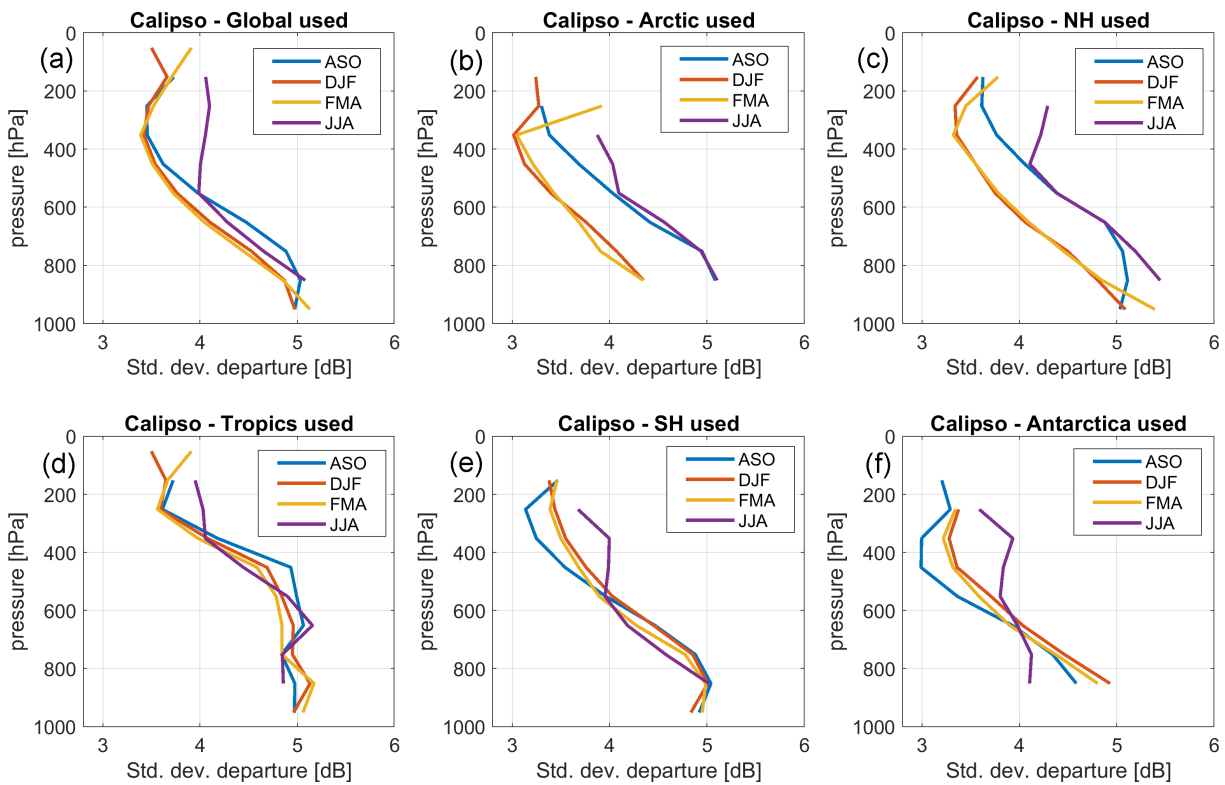

Figure 3.10: Same as Fig. 3.9, but the standard deviation of first-guess departures.

3.3   An updated bias correction scheme

Performing assimilation experiments for different months of the year allows the effectiveness of the bias correction scheme to correct for any seasonal biases to be evaluated. As a recap, a fully functioning climatological bias correction scheme for radar and lidar observations was developed in Fielding and Janisková (2017b). In this scheme, height and temperature were used as indicators to provide an implicit regime dependence. The scheme used look-up tables for the climatological bias based on first guess departures. However, when implementing this approach for the new triple-column approach in the observation operator, some regional biases remained.

The updated bias correction scheme developed in WP-2000 was chosen to be as simple as possible, whilst capturing the majority of the biases inherent in the model and observation operator. The bias correction scheme uses a fit to a 5th order polynomial as a function of model temperature, T:

$$\eta(T) = a_0 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4 + a_5 T^5, \tag{3.1}$$

where a_i is the set of fitted coefficients and η is the bias. Temperature is chosen as the indicator as the microphysical assumptions are sensitive to temperature, and the microphysical assumptions are one of the dominant sources of error. To account for regime dependent biases, five different regions are used (Antarctic, Southern Hemisphere, Tropics, North Hemisphere and Arctic) based on thresholds in Latitude. Also, because the biases are found to be sensitive to hydrometeor phase, two different parameter sets are used for each region (one for model temperatures greater than 273 K and one for model temperatures less than 273 K).

To fit the coefficients in eq. 3.1, a month-long monitor-only (radar and lidar observations are not assimilated) set of first-guess departures for August 2007 was used. The biases in first guess departures with respect to CloudSat and CALIPSO observations from the monitor-only experiment were then fitted to eq. 3.1 using non-linear regression to obtain a smooth bias correction.

Zonal plots of mean first-guess departures for the four different experiments before bias correction (Fig. 3.11 reveal to what extent the model biases vary seasonally. There are three key messages from the plots. The first is that by comparing the ASO and JJA experiments we can see that there is not much year-to-year variability in the biases in northern hemisphere summer months. The bias (observation minus model) is predominantly positive, with the greatest biases for high-clouds in the tropics and Arctic clouds. Secondly, there is significant shift in the zonal mean biases between the DJF and JJA experiments, which correspond to the shifts in the inter-tropical convergence zone (ITCZ). In northern hemisphere summer (JJA), the ITCZ sits in the northern hemisphere, while in winter it shift to just below the equator. This explains the shift in shape of the zonal mean biases. Finally there is a some seasonality to the biases in polar regions, most notably for high-clouds in the Arctic where the bias switches from positive in summer to negative in winter.

Overall, the bias correction scheme, despite only being trained on first-guess departures from August 2007, handles the seasonality of the biases well. Figure 3.12 shows that for all regions in all four experiments, the bias does not exceed 2 dB. Given that the observation errors are much greater than this (typically 15 dB after inflation for error correlations), there is no evidence that a more sophisticated bias correction scheme is needed. The bias correction scheme has the weakest performance in polar regions, for example the bias correction overshoots in the Arctic region of the DJF experiment. One possible approach to correct this would be to use an additional bias correction indicator for each region that slowly varied with time. This approach would be suitable for use with VarBC.
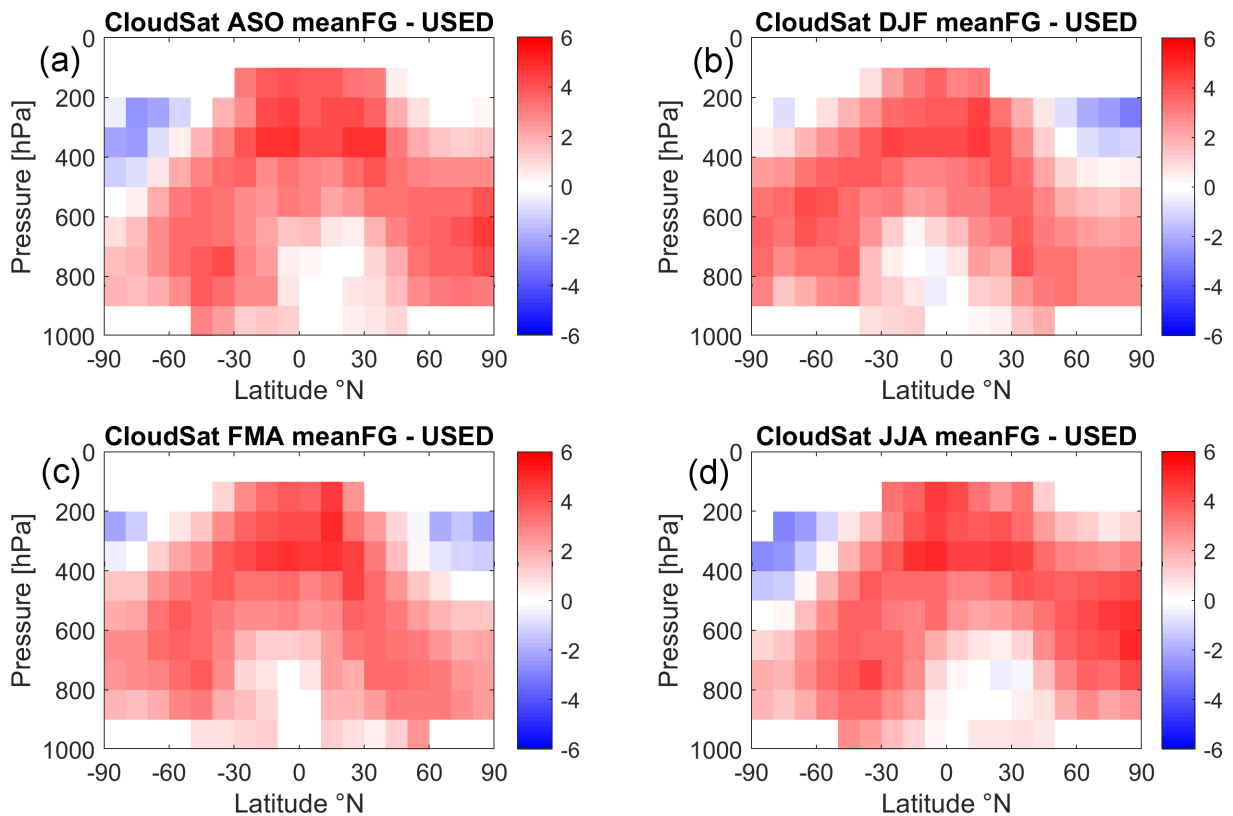
Figure 3.11: Zonal mean first-guess departures of CloudSat radar reflectivity before bias correction for the (a) ASO, (b) DJF, (c) FMA, (d) JJA, cycling assimilation experiments.

After the impact of the lidar pointing angle on the bias in first-guess departures was discovered through the monitoring system, it was clear that an adjustment to the climatological bias correction for lidar backscatter was required. To quantify the impact of the tilting angle, the first-guess departures for ASO experiment were stratified into 'nadir' pointing and ' off-nadir' pointing (Fig. 3.13). The majority of the differences in bias occur at temperatures between 240 K and 260 K, which agrees with microphysical studies (a review of studies can be found in Noel and Chepfer, 2010) that find horizontally aligned ice crystals occur most frequently at these temperatures. The greatest change in bias is found in arctic clouds at 250 K, and the least difference in clouds in the tropics, however the bias changes by over 1 dB for all regions. The bias correction coefficients (in eq 3.1), were updated so that all experiments performed after December 2007 use the 'off-nadir' bias correction.

Similar to the bias in CloudSat radar reflectivity, the observed lidar backscatter tends to be greater than model equivalent for all four experiments (Fig. 3.14). It is difficult to compare the seasonality of the bias given that the lidar was pointing at 'nadir' for much of the ASO experiment, but comparing the ASO and DJF experiments it appears that the biases are qualitatively similar to the CloudSat biases and the peak in bias in the tropics follows the position of the ITCZ. After the new bias correction is applied to the DJF experiment (Fig. 3.15b), the majority of the bias is removed and it can be concluded that the bias correction works well. For the ASO experiment, which used the original bias correction scheme trained on a mixture of nadir and off-nadir pointing data, the bias correction does not work as well, with a band of elevated bias around 400 hPa in the tropics and 600 hPa in the extra-tropics, corresponding to regions of the atmosphere with temperatures conducive to horizontally aligned ice crystals. The FMA experiment has less bias than the ASO experiment because the lidar was pointing off-nadir for the duration of the
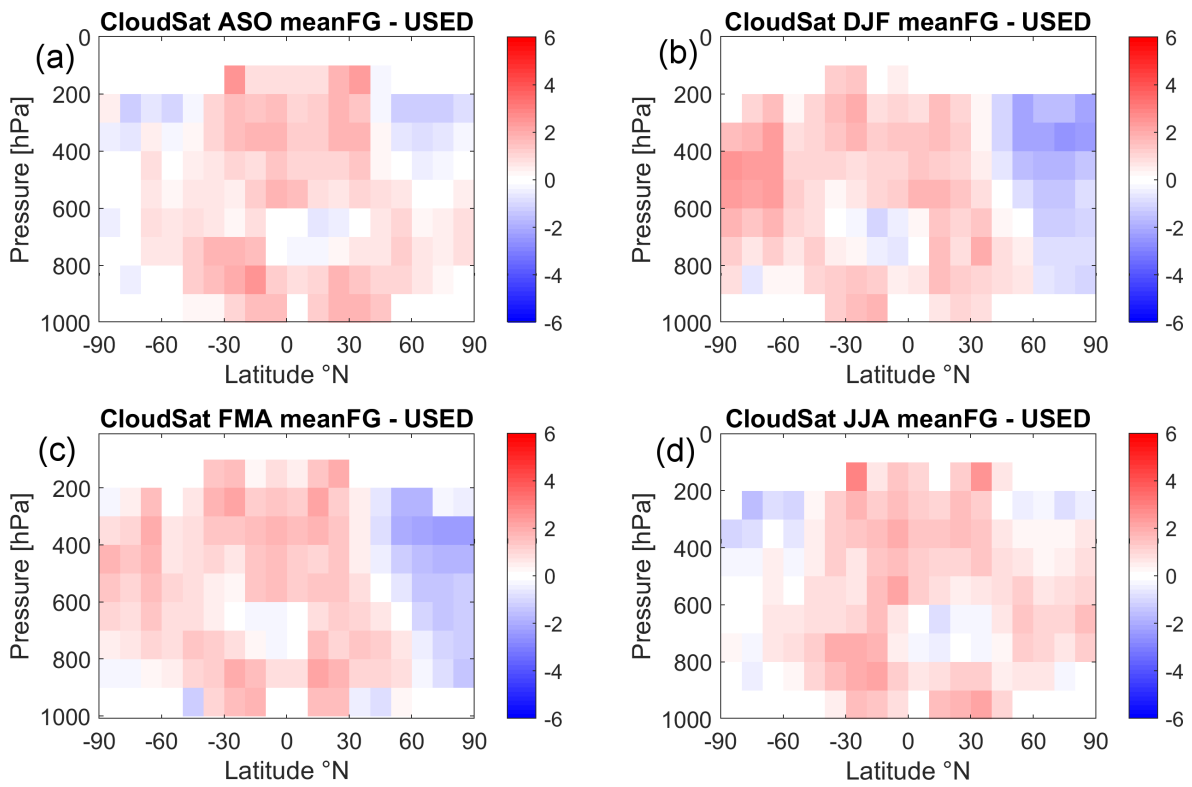
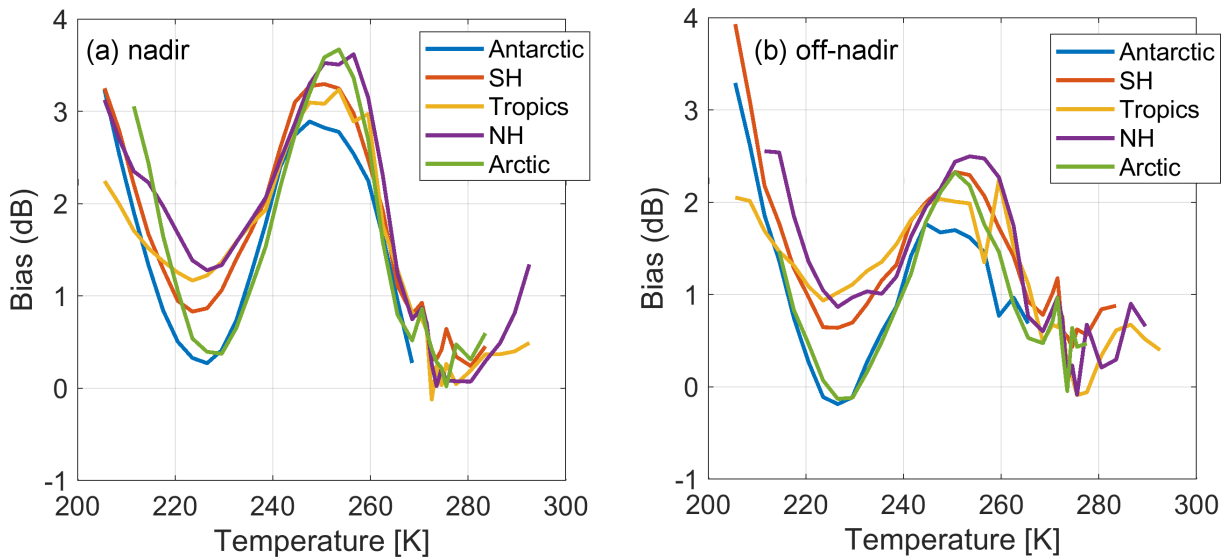*Figure 3.12: Same as Fig. 3.11, but after bias correction.*

Figure 3.13: Bias for CALIPSO lidar attenuated backscatter (before bias correction) when lidar is pointing at (a) 'nadir' with 0.3 degree tilt and (b) 'off-nadir' with 3 degree tilt, as a function of temperature.

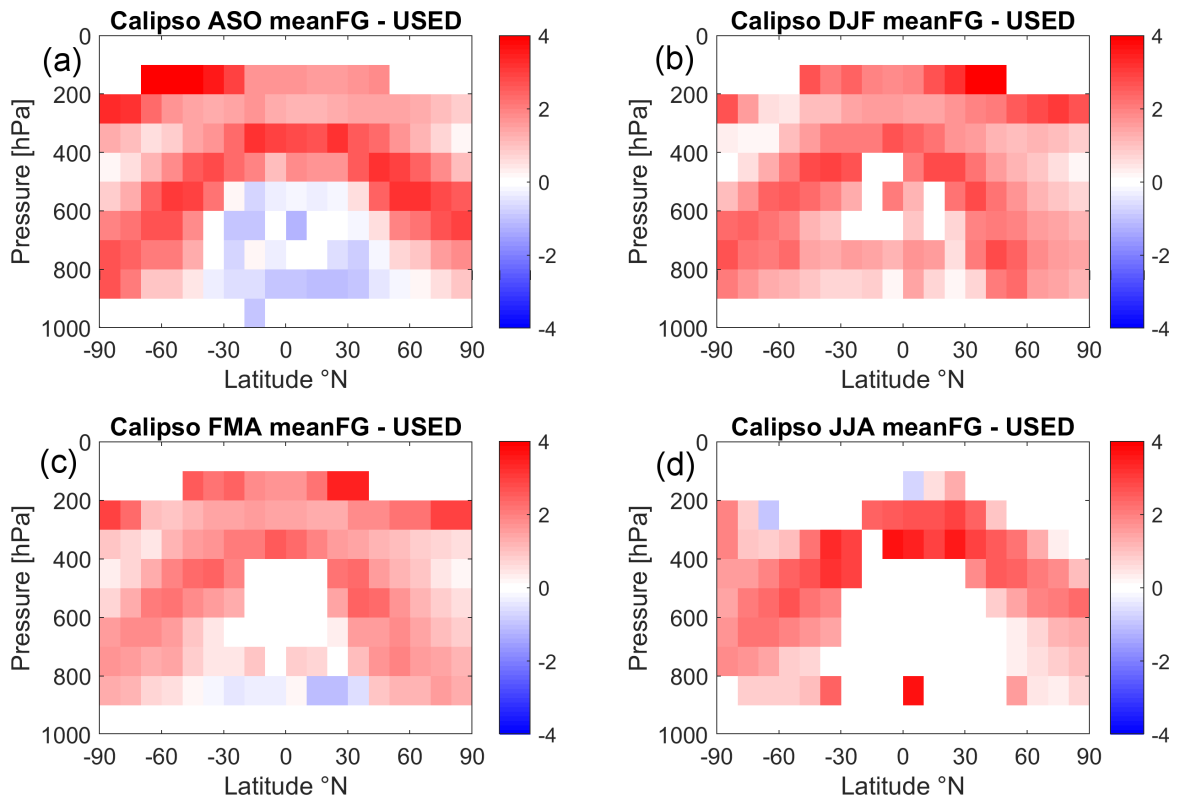experiment, however for optimal results should be re-run with the updated bias correction scheme.

Figure 3.14: Same as Fig. 3.11, but for CALIPSO lidar attenuated backscatter.
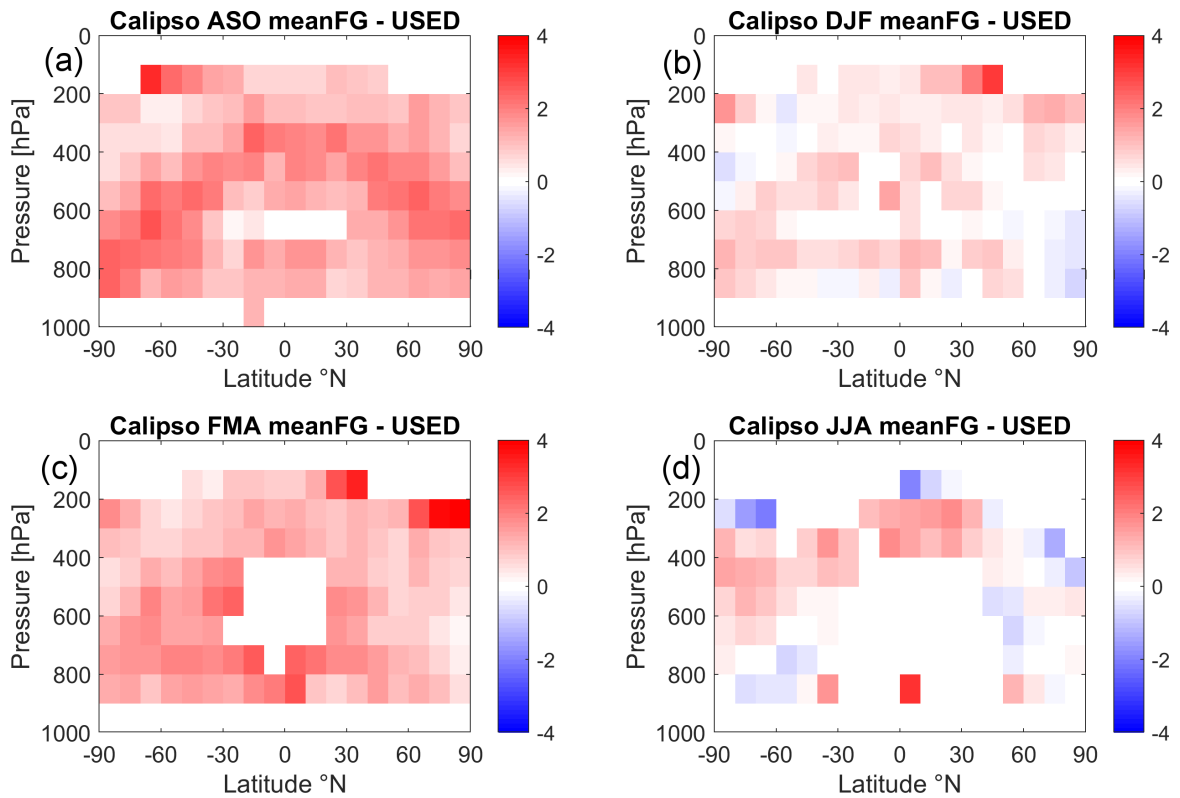

Figure 3.15: Same as Fig. 3.14, but after bias correction.

4   Evaluating the impact on forecasts by verification against own analysis

In this section, we present results from the assimilation experiments to show the impact of assimilating radar and lidar observations on NWP forecast skill. The evaluation in this section is achieved exclusively by comparing forecasts against each experiments own analysis. The evaluation of forecast skill against own analysis must be used in conjunction with other forecast verification techniques, as it can introduce mis-leading results, particularly for short-range forecasts, as it relies on the analysis to well-represent the truth. However, given that the coverage of radar and lidar observations in any given assimilation is relatively modest, we do not expect to see artefacts in the results. We first present the overall results from the combined, 'COMBI' experiment, before exploring seasonal variations in impact and the sensitivity of the results to observation errors.

4.1   Combined impact of experiments

Assimilating CloudSat radar reflectivity and CALIPSO lidar backscatter has a positive impact on the globally-averaged forecast skill of large-scale variables such as temperature, wind and humidity (Fig. 4.1) at both short- and medium-range timescales. The greatest short-term impacts are seen in the forecast skill score of temperature at 100 hPa, where there is a 0.3 % decrease in root-mean square error, which exceeds the 95 % confidence interval that the impact is positive. Positive impacts of similar magnitude are seen in the forecast of vector wind and relative humidity, both of which are significant with 95 % confidence at a forecast lead time of 24 hours. This short-term positive impact appears to persist and grow into the medium-range with forecast skill improvements reaching a maximum at day 5-6. The impact of radar and lidar observations on the globally averaged forecast skill of temperature and vector wind at upper and middle levels both exceed 0.5 % in the medium-range. The impact on the forecast skill of relative humidity appears to be shorter lived for upper levels, but persists into the medium-range at lower-levels, mainly due to the improvements in the forecast skill of temperature.

In general, the impacts on forecast skill appear to be greater in the southern hemisphere than the northern hemisphere, particularly at shorter lead times. This is evident when comparing the change in forecast skill for geopotential height at 500 hPa (Fig. 4.2) for the two hemispheres. Both hemispheres show a reduction in forecast error in the medium-range; at days 5-6 there is around a 1 % reduction in root-mean square forecast error when assimilating the radar and lidar observations in the COMBI experiment. Bearing in mind that the geopotential at 500 hPa is sensitive to the density of the underlying atmosphere, this could be because most of the short-range impacts are found in the atmosphere above 500 hPa and it takes time for the atmospheric model to propagate these increases in skill down through the atmosphere. However only the southern hemisphere has a significantly positive impact in the short-term range.

To understand the spatial distribution of forecast impacts at different lead times, Fig. 4.3 shows the zonal mean difference in RMS error for model temperature, wind and geopotential height relative to the RMS of the control experiment. For the first 48 hours, the majority of improvements in forecast skill are in the Southern hemisphere in two regions: one between -60 and -80 degrees latitude, the other in the tropics between -30 and 0 degrees latitude. This could be for a few reasons; there are less conventional observations in the southern hemisphere, so it could be that the radar and lidar observations provide greater unique information in these less observed regions. Secondly, all-sky radiances are currently screened out in convective outbreak situations, which are frequent in these regions. All-sky radiances are also not currently assimilated over land, so polar regions do not benefit from the information provided from passive microwave radiances. Thirdly, because CloudSat and CALIPSO are polar orbiting satellites, the observation density is much greater near the poles.
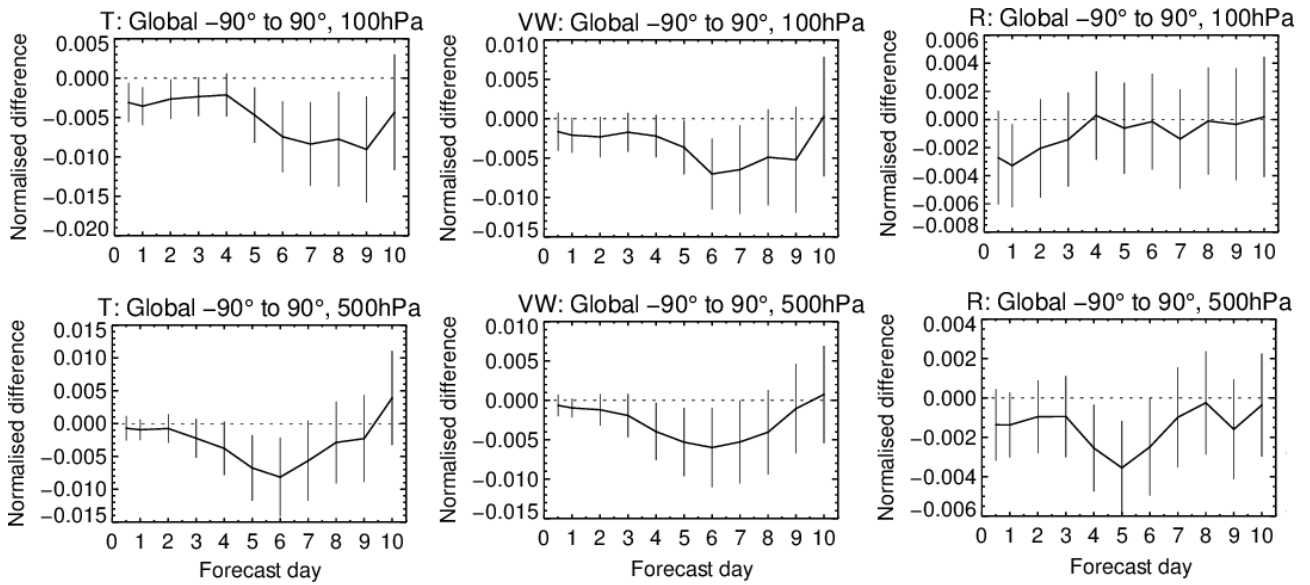
Figure 4.1: Global mean change in root-mean squared (RMS) error of temperature (left column), vector wind (middle column) and relative humidity (right column) compared to own analysis, relative to the control experiment. The statistics are generated from a combined 11-month period of assimilating CloudSat radar reflectivity and CALIPSO lidar backscatter in addition to the regularly assimilated observations between August 2007 and August 2008. The results are shown for 100 hPa (top row) and 500 hPa (bottom row). Error bars show 95 % confidence intervals accounting for autocorrelation and a Šidák correction for 4 independent tests.
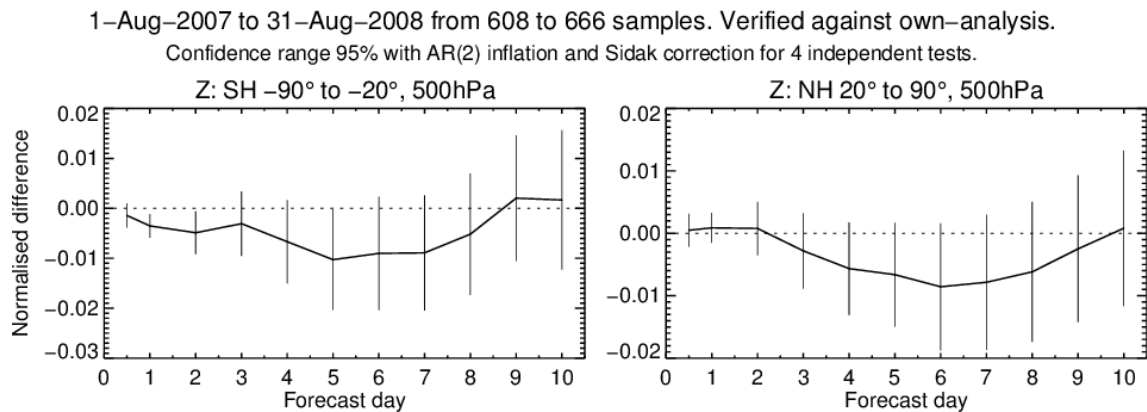

Figure 4.2: Same as Fig. 4.1, but for forecast skill in geopotential height at 500 hPa for (left panel) Southern Hemisphere and (right panel) Northern hemisphere.

The positive impact of the radar and lidar observations on the forecast appears to improve with forecast lead-time in most areas in all three large scale variables shown in Fig. 4.3. At around day 4, positive impacts begin to be seen in the extra-tropics in both hemispheres. This impact spreads to all almost all regions by day 6, apart from northern regions of the tropics. To try and understand this evolution of forecast errors, we can look at spatial maps of change in forecast skill for different variables at different heights, such as those shown in Fig. 4.4. Because there are less averaging of data in the lat-lon plots they the signals tend to be noisier. To mitigate this, spatial averaging of the statistics to 10 by 10 degrees is performed.
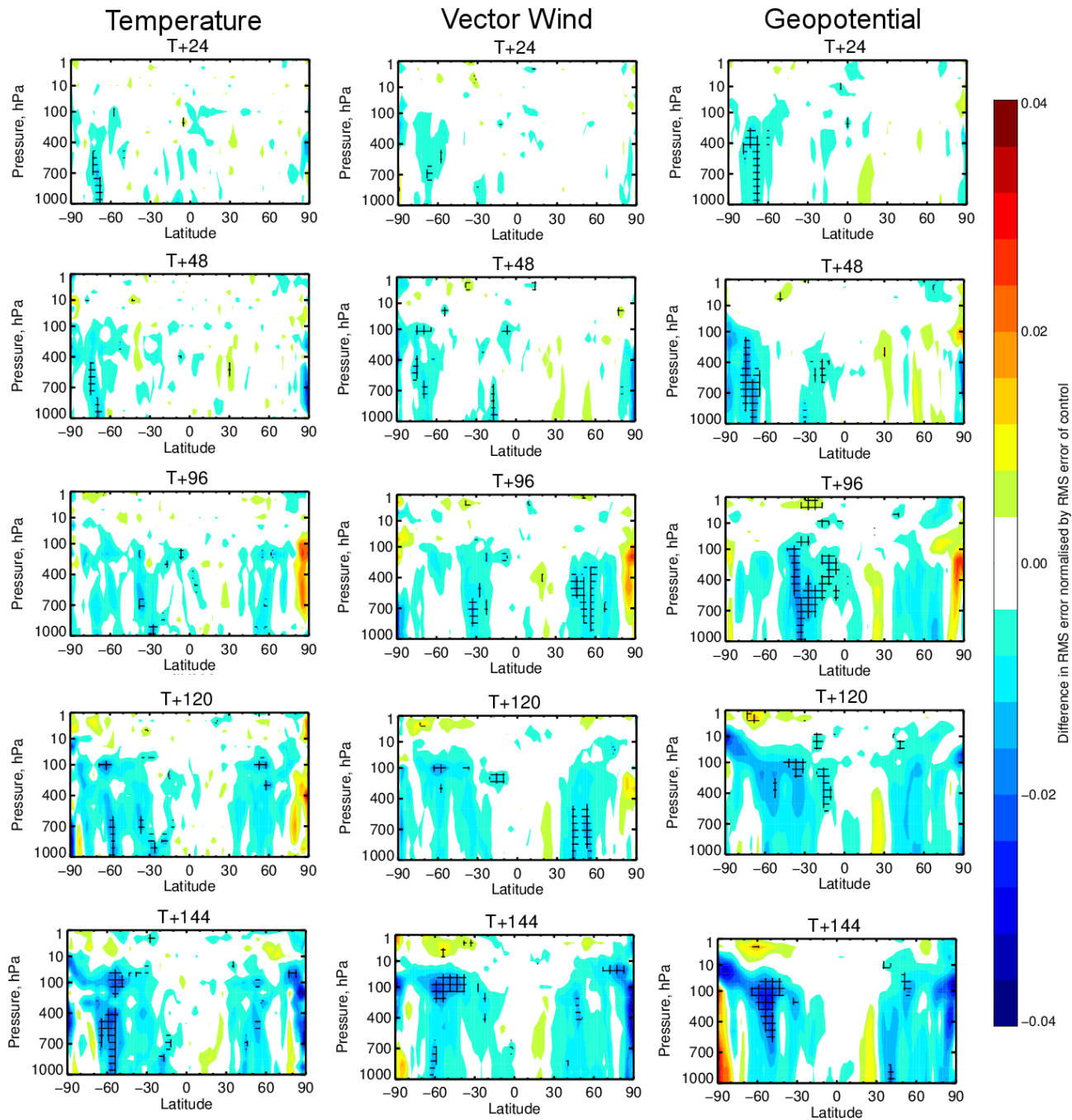
Figure 4.3: Impact of assimilating cloud radar reflectivity and lidar backscatter from CloudSat and CALIPSO respectively on forecast skill of temperature (left column), vector wind (middle column), and geopotential (right column) for a combined 11-month period between August 2007 and August 2008 (see text for details). Panels show the zonal mean difference in root-mean-squared (RMS) error of each large-scale variable compared to own analysis, normalised by the the RMS of the control experiment. The results are shown for forecasts at (from top row downwards) 24, 48, 96, 120 and 144 hours. Hatched areas indicate regions that have reached 95 % confidence in the sign of the change in forecast skill from assimilating the radar and lidar observations. Blue colours indicate an improvement in forecast skill.
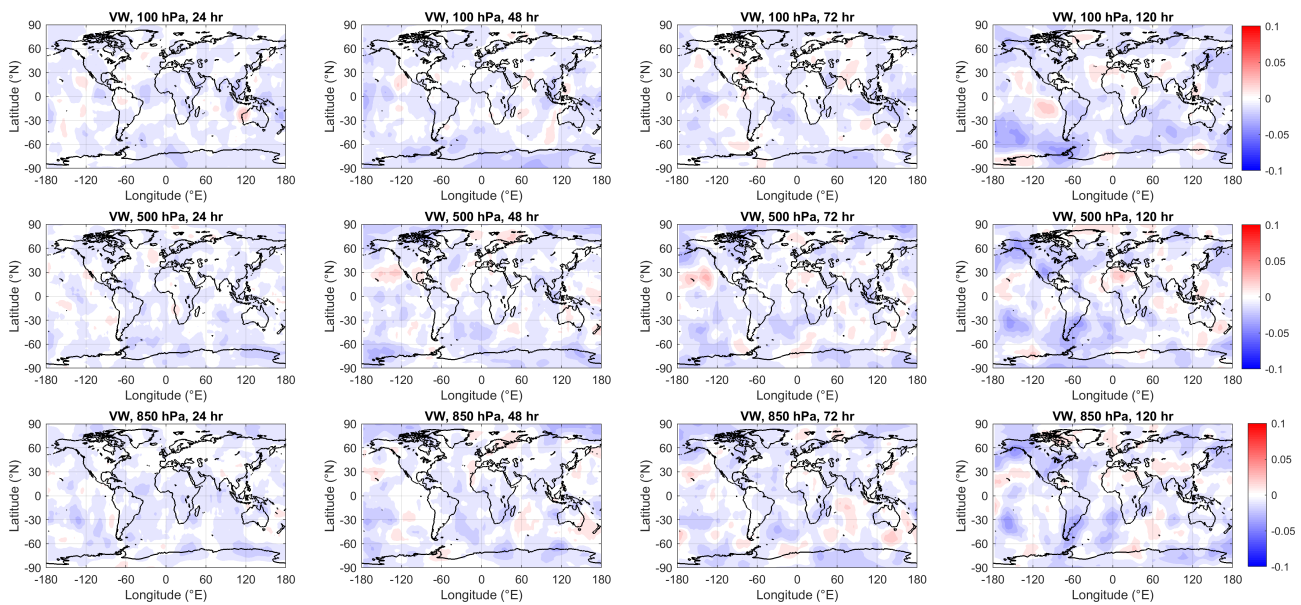
Figure 4.4: Change in forecast skill of vector wind at 100 hPa (top row), 500 hPa (middle row) and 850 hPa (bottom row), averaged to a 10 by 10 degree spatial resolution for forecast lead times of (left to right) 24, 48, 72 and 120 hours.

Impacts from cloud radar and lidar observations showing improvements in the short-term forecast of wind, are likely to be due to the 'tracer effect'. As explained by Geer et al. (2014), the tracer effect is the ability of the clouds fields in 4D-Var to be adjusted and fitted to observations through changes to the large-scale dynamics rather than just temperature and humidity increments. Later on in the forecast, the impacts to the forecast skill of wind are likely to be more complicated with the interactions between dynamics and physics playing a greater role. Nevertheless, with just small improvements in the initial wind fields could lead to large impacts later on in the forecast due to the chaotic nature of the atmosphere.

In Fig. 4.4, we can see that at a forecast lead-time of 24 hours, the positive 'tracer effect'-like improvements to the vector wind from assimilating cloud radar and lidar are evenly distributed across the globe. The greatest improvements appear to be for lower-level parts of the atmosphere, which agrees with the verification against first-guess departures for conventional wind observations (see WP-2000; Fielding and Janisková, 2020a). For upper-levels, such as at 100 hPa, the impact of radar and lidar observations is more mixed, but is generally positive in the southern hemisphere storm-tracks and over Antarctica where there tends to be fewer other observations assimilated. At longer lead-times of 3-5 days, the impact tends to increase relative to the 24 hour forecast time and in the same areas; for example, at lower levels, regions with larger impact include the Southern Atlantic and Southern Pacific oceans. At higher levels, both the Northern- and Southern-hemisphere storm tracks show the greatest impacts.

In contrast to impacts on vector wind, impacts on temperature in the short-term forecasts will tend to be due to the adjustments of clouds and precipitation directly in the 4D-Var by adjusting cloud water content through changes in temperature causing either additional condensation or evaporation. Improvements in temperature at short forecast lead-times are greatest at upper levels (Fig. 4.5) and in the tropics. This finding agrees with the reduction in FG departures of clear-sky AMSU-A radiances (see WP-2000) with weighting functions in the upper troposphere. The greatest impacts appear to be over Africa and in the north-Eastern Pacific near Hawaii. At lower levels, there is a slight degradation in the West Atlantic that could be related to biases in CloudSat first-guess departures (see discussion in Sec. 5.1). Later in the

forecasts the positive impacts tend to grow at all levels. The degradation over Greenland at 850 hPa can be ignored as it is likely to be an artefact due to vertical extrapolation over orography.
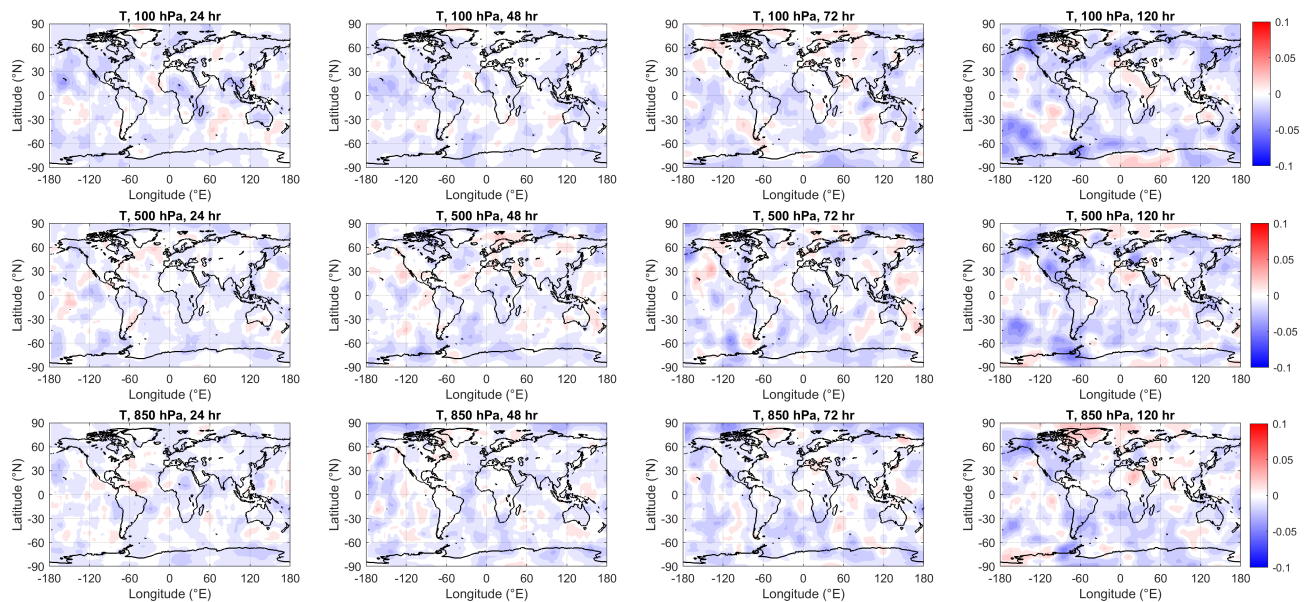

Figure 4.5: Same as Fig. 4.4, but for change in forecast skill of temperature.

Similar to the impact on temperature, the cause of impacts on relative humidity are likely to be directly related to the adjustments of clouds and precipitation in the assimilation system. The regional impacts on relative humidity (Fig. 4.6) at upper-levels can be noisy, but generally show improvements in the short-term forecast. At lower-levels, the impact of the new observations is similar to the impact on temperature. At day 1, there are almost no negative impacts over land, suggesting that the new observations have the greatest impact on humidity in these regions. At longer forecast lead-times at middle to lower levels, the southern hemisphere tends to have the most prominent positive impacts, particularly at day 5.

The spatial distribution of impacts on forecast skill of geopotential height at 500 hPa for the one-day forecasts is quite uniformly positive spread around the world (Fig. 4.7), with the exception of an area of slight degradation in the Northwest Pacific. There is no obvious reason for this degradation, but we will show later in Section 5.1 that this region conversely has an increase in skill of predicting TOA radiation fluxes (both shortwave and longwave), so it is possible that while the cloud radiative forcing has improved in this region, it has come at the expense of the accuracy of geopotential height. However, this area of degradation appears to be short-lived and by day 3 the impact turns positive in this region. Along the storm tracks in both Southern and Northern hemispheres, the skill in geopotential height increases with lead time, and the improvements are predominantly greater than 1 % in these regions. By day 5, there are some degradations in the Northern hemisphere, most notably in the stratocumulus to cumulus transition zone in the North East Pacific, and across much of North Africa. For the North East Pacific, this appears to be an evolution of errors that begins with small degradations in temperature and vector wind at shorter lead times in the same. For North Africa, it appears to be related to degradations in the forecast of temperature closer to the surface. However further investigations would be needed to confirm these hypotheses, bearing in mind that the negative impacts in these regions are not statistically significant. As was shown in Fig 4.2, the impact on forecast skill of geopotential at 500 hPa is positive overall.

Finally, we can also consider the change of forecast skill of surface and column-integrated model prop-
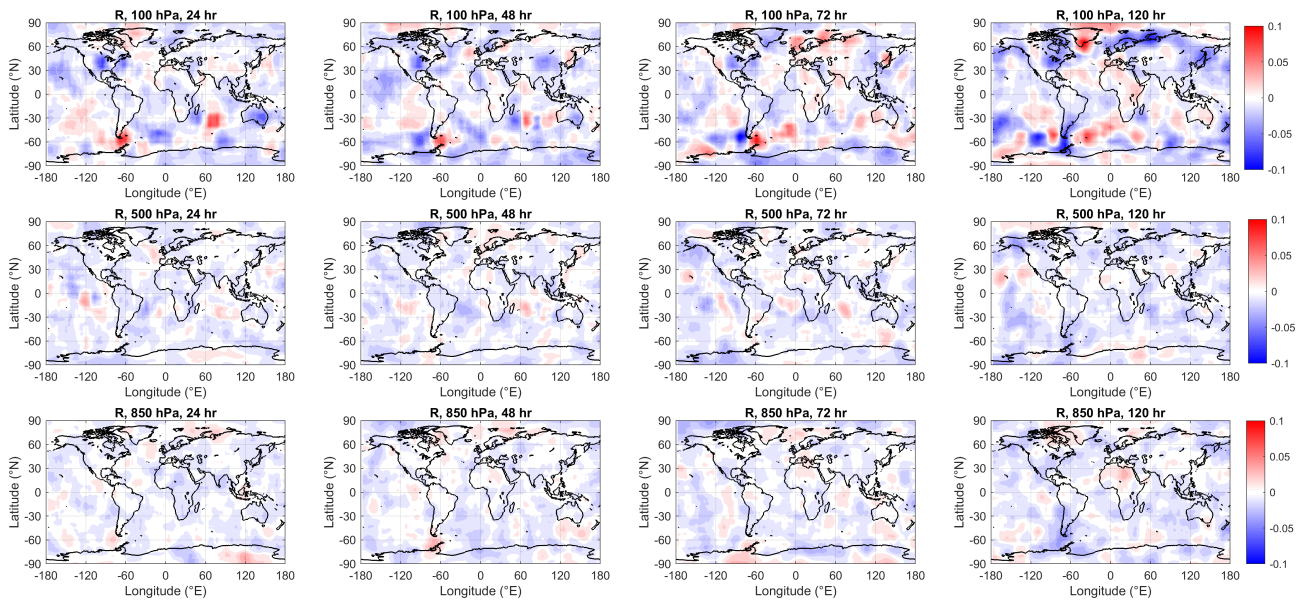
Figure 4.6: Same as Fig. 4.4, but for change in forecast skill of relative humidity.
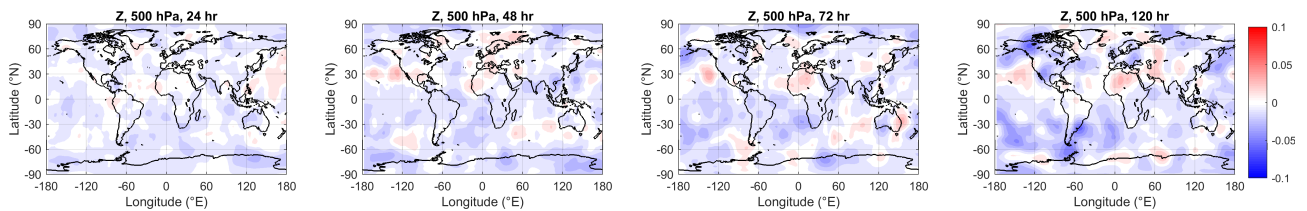

Figure 4.7: Same as Fig. 4.5, but for change in forecast skill of geopotential height at 500 hPa.

erties. Because these surface variables are not directly related to the control variables in 4D-Var, the analysis of these variables (which we are using for verification against the forecasts) is less constrained and therefore potentially noisier. Figure 4.8 shows the impact of assimilating radar and lidar on the forecast skill of precipitation, total column water vapour, total cloud cover and two-metre temperature. For nearly all lead-times, the sign of the results has not reached 95 % confidence, so the results should be treated with caution and corroborated with other verification techniques, such as verification against independent observations.

Figure 4.8 suggests that the greatest impact of assimilating radar and lidar observations on the forecast of precipitation at the surface is in the tropics at forecast day 2, although for all lead-times the sign of the impact has not reached 95 % confidence. The impact on very short-term forecasts is particularly difficult to verify against own analysis for polar orbiting satellites, such as CloudSat and CALIPSO, because they visit different locations in each assimilation cycle. Therefore the improved analysis could penalize a previous forecast because the precipitation was improved in a different location. Nevertheless, the precipitation appears to be marginally improved in the northern hemisphere at all lead-times up to day 5-6. The opposite is true for the Southern Hemisphere, where precipitation appears to be slightly degraded.

The change in skill of total-column water vapour in the Southern Hemisphere does not agree with the change in skill of precipitation which would expect to be correlated; the total-column water vapour fore-
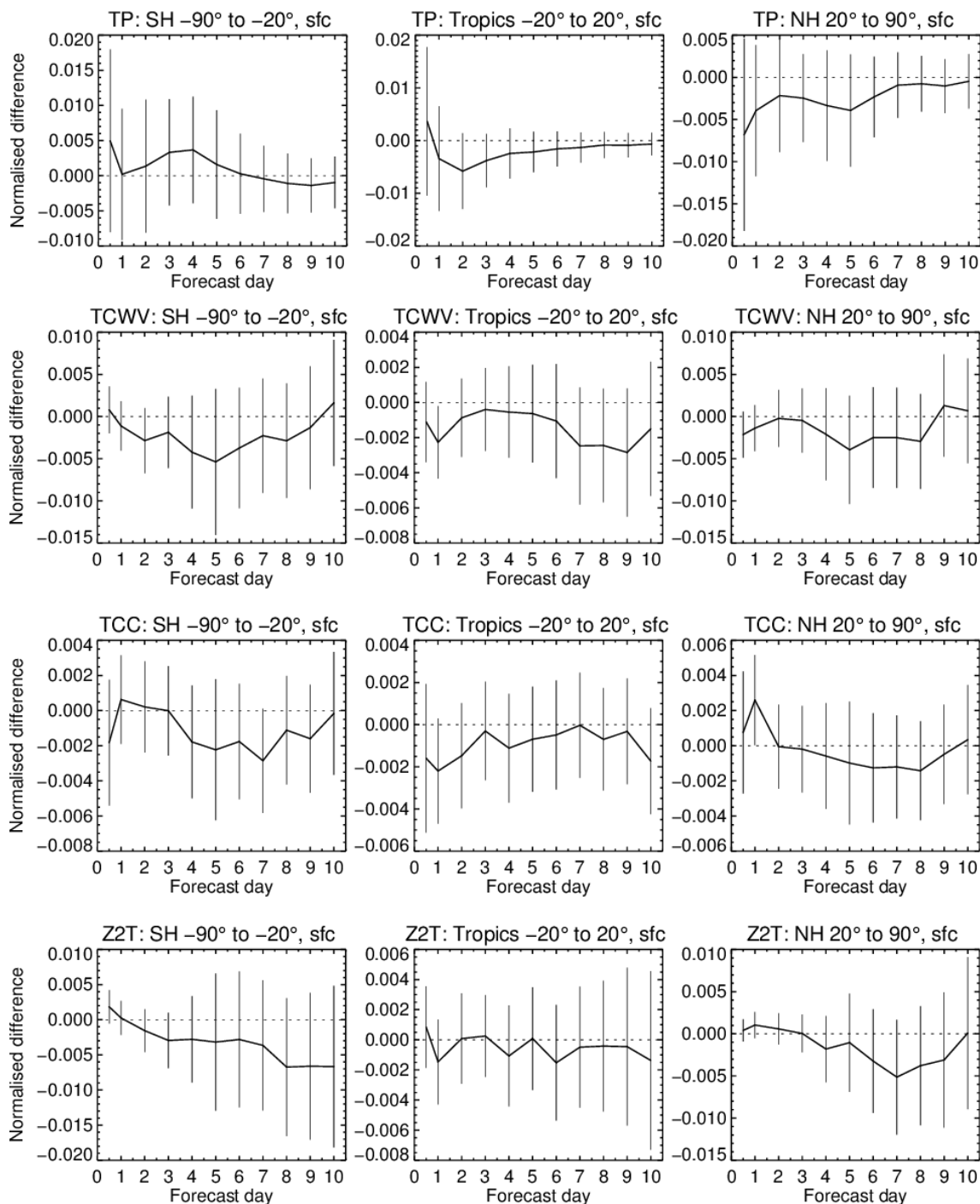
Figure 4.8: *Same as Fig.* 4.2*, but for forecasts of surface precipitation (top row), total-column water vapour (TCWV; second row), total cloud cover (third row) and 2m temperature (bottom row). Scores are stratified by regionally; left column shows southern hemisphere, middle column show tropics, right column shows northern hemisphere.*

cast skill actually improves up to a maximum at day 5, further justifying the need to treat the verification with caution. In the tropics at forecast day 1, there is a modest but significant increase in forecast skill of total-column water vapour, which corroborates the increase in skill of predicting precipitation at the surface. Similarly, in the northern hemisphere, the very short-term forecasts of total-column water vapour are improved, although not quite significantly. In terms of total cloud cover, the same caveats apply as to the verification of precipitation against own analysis. Most of the lead-times show a small positive, but not significant, impact, particularly for short-term forecasts in the tropics. Lastly, the impact on two-metre surface temperature is closely correlated with the impact of the total cloud cover; all the regions tend to have a slight apparent degradation at short lead times before transitioning to a positive impact from day 2 onwards.

In general, no obvious signs of seasonal variations in the impact of radar and lidar observations on the forecast were identified. This could either be because any differences are not detectable within the noise from comparing three-month impact studies, or that the differences are negligible. To summarise the seasonal impact of the observations, Fig. 4.9 shows the relative change in skill of forecasting geopotential height at 500 hPa as a function of forecast date and latitude. By their nature, the scores on a cycle-to-cycle basis are noisy, so a 30-cycle (15-day) smoothing is applied to enhance the signal to noise ratio. At lead-times of 24 hours (Fig 4.9a), the greatest positive impact on geopotential is in the tropics during March 2008. There is a consistently positive impact in the southern hemisphere extra-tropics. The largest negative impact is in the northern hemisphere extra-tropics around February 2008. On the whole, the day 5 impact (Fig 4.9b) is an amplified version of the 24 hour forecast impact, with the same forecast dates and regions showing positive impacts. The JJA experiment period appears to have less impact than the other periods, possibly due to the sub-optimal use of the lidar data.
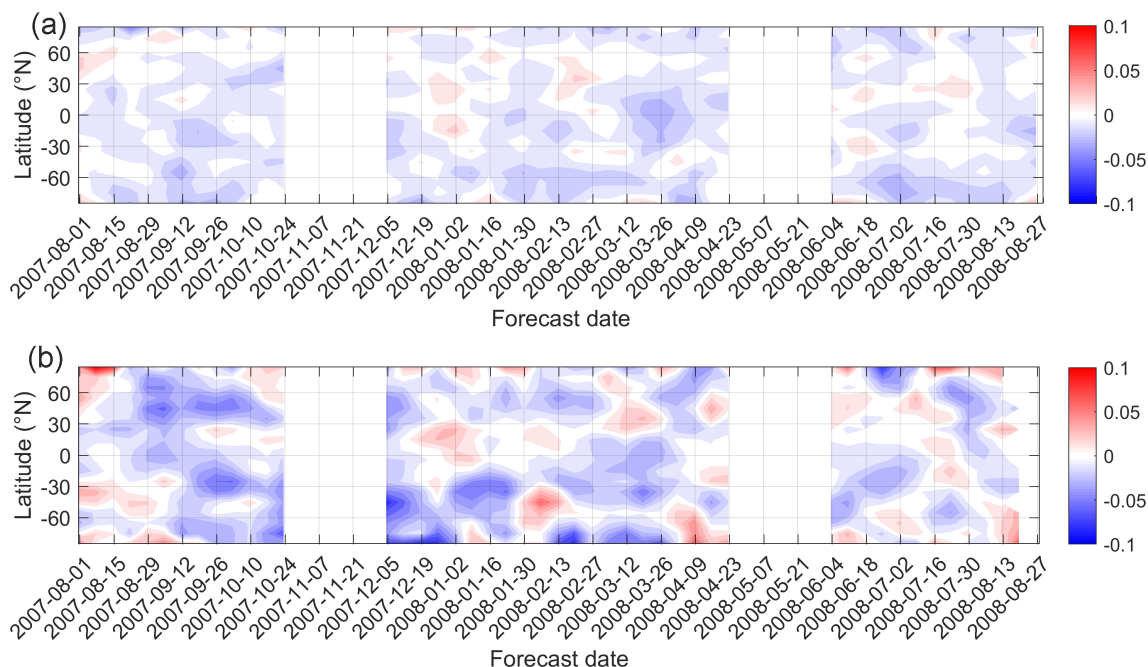

Figure 4.9: *Impact of assimilating cloud radar and lidar observations on zonal mean forecast skill of geopotential height at 500 hPa for (a) 24 hour forecast (b) 120 hour forecast, relative to the control experiment as a function of forecast date.*

4.2 Sensitivity to observation error

In WP-2000 (Fielding and Janisková, 2020b), the sensitivity of the observation errors assigned to the radar and lidar observations was investigated by running a set of assimilation experiments with scaled observation errors. In the study the observation errors were scaled by 1x, 1.5x and 2x the original observation error specification. By comparing the impact on first-guess departures of other assimilated observations, it was shown that the 2x error scaling gave optimum results. In this work-package we attempt to refine the observation error specification further by performing a set of assimilation experiments with 1.8x, 2x and 2.2x error scaling relative to the original observation error specification. The adjusted observation errors were applied to the DJF experiment time-period.

In general, using the 2x scaling of the radar and lidar observation errors produces the greatest positive impact on the forecast of temperature, vector wind and relative humidity (Fig. 4.10). For the very short-term forecasts of temperature at 100 hPa, the smaller the scaling the smaller the forecast error. However this rule does not extend to the medium-range, where the 2x scaling performs best. For global forecast skill of relative humidity, the 2x scaling experiment performs the best for the very short-term forecasts; this is confirmed by the first-guess departures for conventional observations of humidity (not shown). For vector wind observations, the 2x scaling experiment performs best at nearly all lead-times.
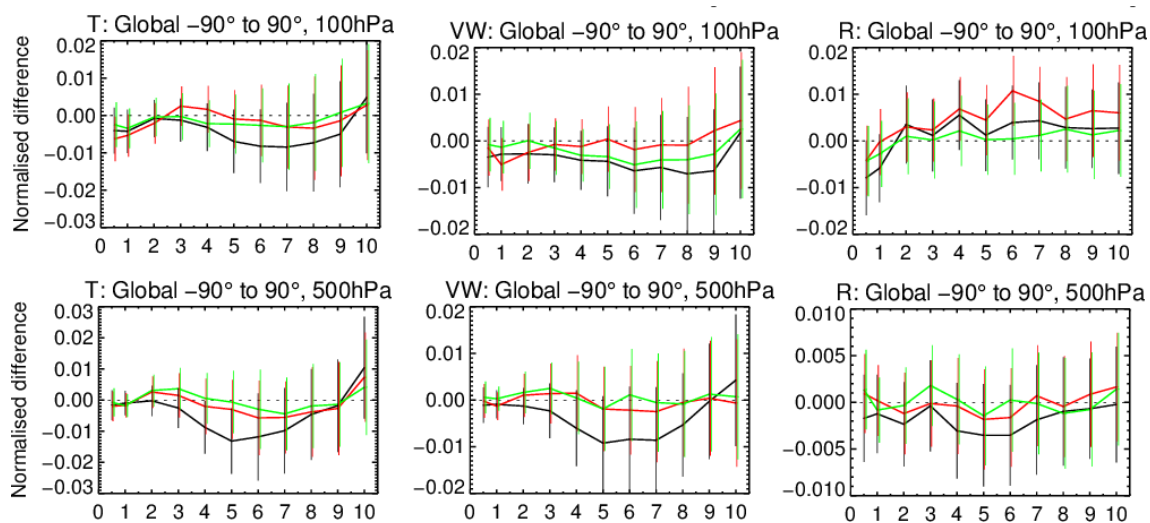

Figure 4.10: *Same as Fig. 4.1 but for the DJF experiments only and with observation error scaling of 1.8x (red), 2x (black), 2.2x (green).*

A similar story for the effect of the observation error scaling is seen for the impact on forecast skill of geopotential height at 500 hPa (Fig. 4.2). For the southern hemisphere, the 2x error scaling performs the best out to forecast day 7, with the 1.8x scaling performing second, and the 2.2x error scaling having the least impact. For the northern hemisphere, the 1.8x scaling has the best performance up to the first 24 hour forecast, but subsequently the 2x error scaling performs best. From these experiments we can conclude that the optimum value for the observation error scaling is likely to be between 1.8x and 2.2x error scaling, and that there is no reason to change from the 2x error scaling that was previously chosen.
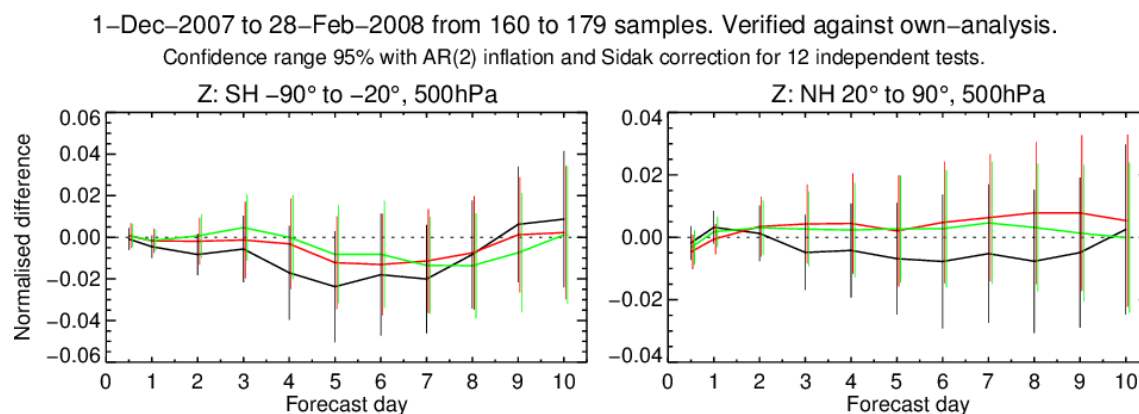
Figure 4.11: Same as Fig. 4.2 but for the DJF experiments only and with observation error scaling of 1.8x (red), 2x (black), 2.2x (green).

5 Evaluating the impact on forecasts by verification against independent observations

Verifying the impact of assimilation system changes by comparing forecasts against own analysis is a powerful tool and convenient as any forecast variable simulated can be verified, however it is important to use other tools to corroborate the findings from this type of analysis. One such method is to evaluate the impact on forecasts against independent observations (i.e. observations which were not assimilated by the system), which is the subject of this section.

5.1 Broadband radiative fluxes

CERES top-of-atmosphere (TOA) broadband radiative fluxes are very useful independent observations of the Earth's radiation budget for forecast verification as they provide a holistic view of many different model processes and, pertinent to the assimilation of radar reflectivity and lidar backscatter, are sensitive to clouds. The broadband shortwave radiation at top of atmosphere is sensitive to all clouds, but particularly boundary-layer clouds, which have small, highly reflective droplets. Because the source of the shortwave radiation is the Sun, it will only be sensitive to forecast changes in the day-time. It is less sensitive to clouds over highly reflective surfaces, such as sea-ice, so it is mostly useful for verification of cloud forecasts in the extra-tropics and tropics. The observed TOA net longwave radiation is predominantly caused by emission from the Earth's surface, and the absorption and emission of radiation from clouds. It is therefore sensitive to changes in the height and temperature of clouds, with the greatest sensitivity to high-level clouds. In contrast with the sensitivity of shortwave radiation, changes to boundary-layer clouds that are of similar temperature to the surface, will have little effect on the model broadband fluxes.

The assimilation of cloud radar reflectivity and lidar backscatter has a positive impact on the model's radiation budget when compared with CERES TOA radiative fluxes (Fig. 5.1). Both shortwave and longwave radiative fluxes show improvements of around 0.1 % for forecast lead times of up to 5 days. The greatest positive impact is in the first 48 hours of forecasts, with improvements in longwave radiation that are significant at a 95 % confidence level and significant in the shortwave for the 24-48 hour forecast period. The fact that the longwave radiation shows greatest improvements in forecast skill compared

with the shortwave, agrees with the improvements in the upper-level temperatures and winds that are seen early in the forecast period (e.g., Fig. 4.1).
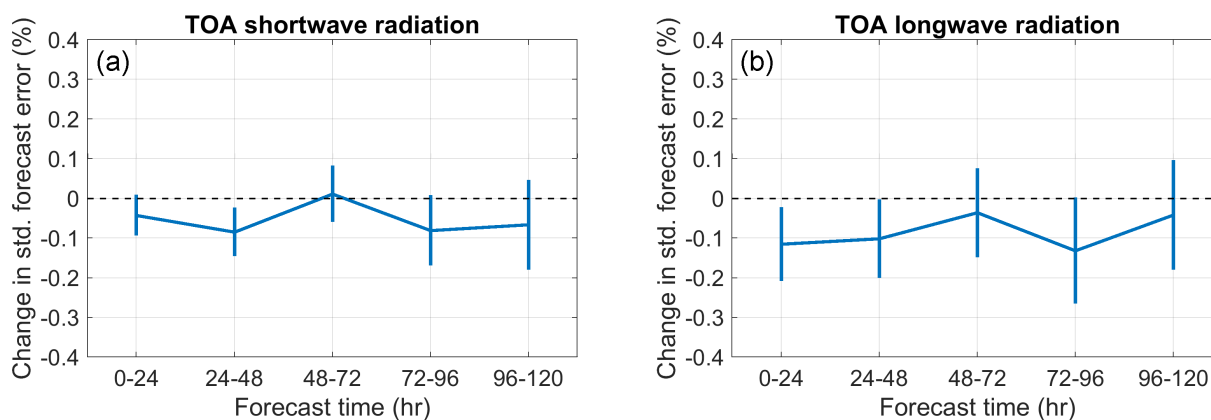

Figure 5.1: Change in the standard deviation of the forecast error of daily mean (a) TOA net shortwave radiation fluxes and (b) TOA net longwave radiation fluxes compared to the 1x1 degree CERES daily mean product for the COMBI experiment relative to the control. Error bars show 95 % confidence intervals.

To explore the regional impacts of the assimilation experiments on radiative fluxes, Fig. 5.2 shows the change in skill scores of shortwave and longwave fluxes for the first 24 hour of forecasts. For shortwave fluxes, the modest improvements to forecast skill are mainly spread evenly across the globe. The greatest impact is seen over western Africa, suggesting the assimilation improves the representation of deep convection in these areas as this is the dominant cloud type in this region. This is corroborated by the fact improvements are also seen in the TOA longwave radiation in this region (Fig. 5.2b). Whereas there were no obvious regions of forecast degradation in the shortwave, in the longwave, the north-eastern Pacific ocean shows a degradation in skill in the first 24 hour forecast of around 1 %. This also agrees with the slight degradation in upper-level temperatures seen in Fig. 4.5.
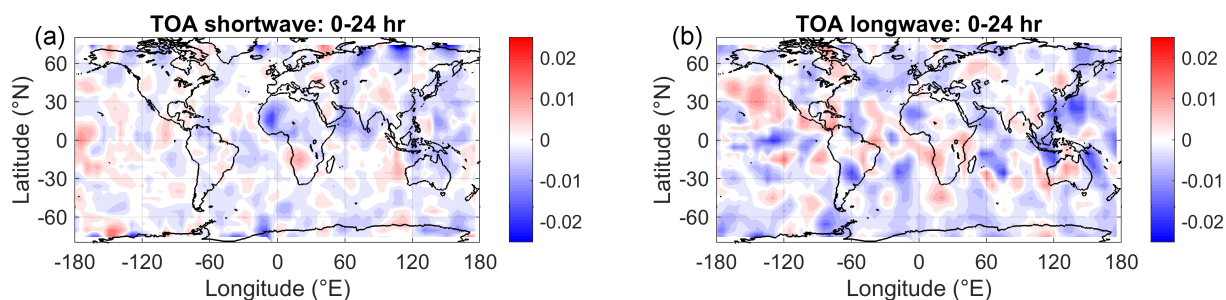

Figure 5.2: Same as Fig. 5.1, but for spatial maps, smoothed to a 10 x 10 degree grid. Scale is fractional change in the standard deviation of forecast error for the COMBI experiment relative to the control.

It is not clear what might cause the slight temperature degradation in the north-Eastern Pacific, so to explore if the errors could be related to the radar observations, Fig. 5.3 shows the spatial distribution of biases in the mean first-guess departures of the CloudSat observations. Interestingly, there are clear regionally dependent biases in the observations even after bias correction. In particular, there is a positive bias in CloudSat radar reflectivity in the north-eastern Pacific. However, the biases in this region are dominated by low-clouds, which should have a greater impact on shortwave radiation. It is therefore not clear what is causing the degradation in longwave fluxes, but it could be related to model biases in clouds in this region, which are well-known (Teixeira et al., 2011).
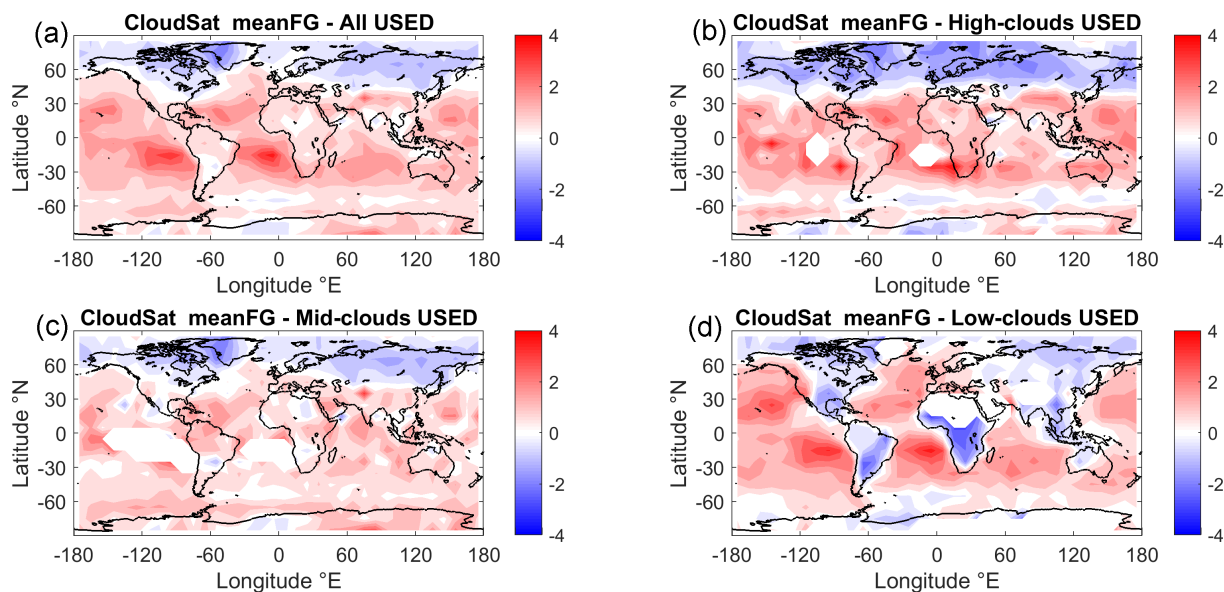
Figure 5.3: Mean first-guess departures after bias correction for CloudSat radar reflectivity in the COMBI experiment stratified by pressure (p) intervals. Panels show maps of (a) all FG departures, (b) FG departures for high-level clouds (p < 300 hPa), (b) FG departures for mid-level clouds (700 hPa > p > 300 hPa) and (c) FG departures for low-level clouds (p > 700 hPa).

In analysing the regional biases in CloudSat first-guess departures, it is clear that there is a strong land-sea contrast in biases for low-cloud, with observations having greater reflectivity over ocean and smaller reflectivity over land. Typically, clouds over ocean have smaller droplet number concentrations than over land, which means that the cloud droplets over ocean are larger and have greater radar reflectivities. These biases could be removed through an additional land-sea indicator in the bias correction, or perhaps a two-moment observation operator that used a climatological cloud droplet number concentration. The investigation of this is left for further research.

5.2   Precipitation

We conclude our analysis of the impact of the assimilation of radar and lidar observations on forecast skill by investigating changes in the forecast skill of precipitation. In Fig. 4.8, we saw that improvements in precipitation appeared to be greatest in the tropics. As an independent verification of the skill in forecasting precipitation, Fig. 5.4 shows the change in the standard deviation of differences between the model accumulated precipitation and gridded rain gauge observations provided by the GPCC dataset. Overall, the impact appears to be neutral for forecasts up to day 5, although four out of five forecast lead-times show an improvement in fit with the rain-gauge observations. Regionally (Fig. 5.5), the greatest positive impact appears to be over North America. This finding agrees with Fig. 5.2b, where improvements in the forecast of longwave radiation, suggest improvements to heavy rain producing deep-convection. There are also tentative signals for improvements in rainfall over Western Africa, which also saw improvements to the prediction of longwave radiation, but the sparse network of rain gauges in this region makes evaluation difficult.
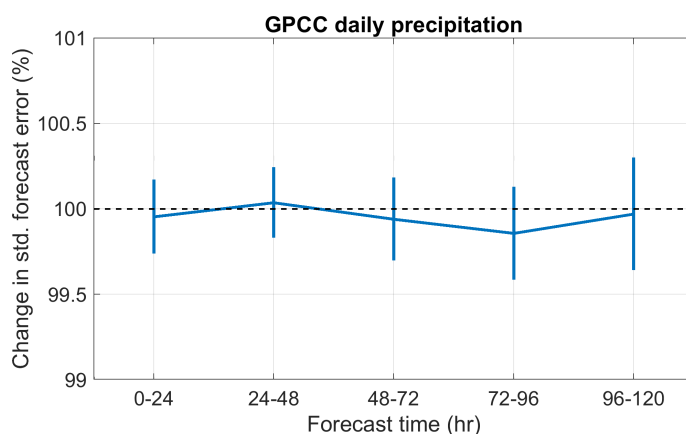
Figure 5.4: Global mean change in the standard deviation of difference between forecast accumulated precipitation at the surface and the Global Precipitation Climate Center daily global land-surface precipitation product, relative to the control experiment. Error bars show 95 % confidence intervals.
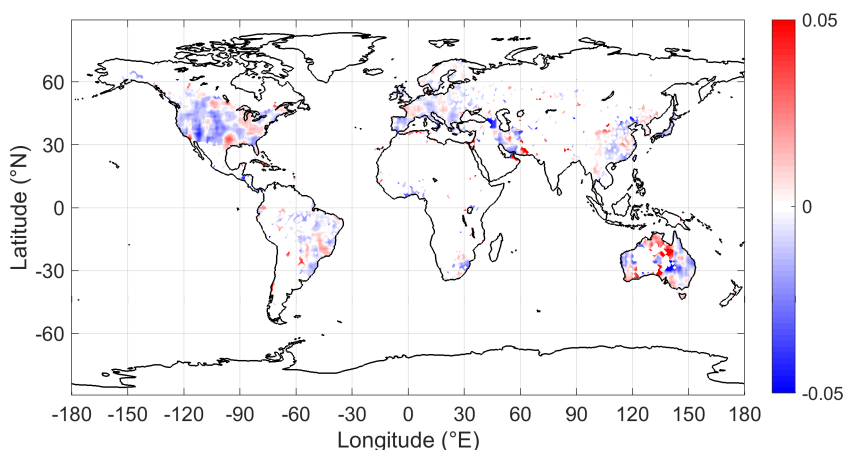

Figure 5.5: Same as Fig. 5.4, but for a spatial map of the relative change in the standard deviation of differences between model and GPCC daily precipitation for the first 24 hour forecast.

6 Summary and conclusions

Satellite observations of cloud radar reflectivity and lidar backscatter have the potential to improve numerical weather prediction in two distinct ways: indirectly through advances in our understanding and modelling of cloud processes and directly through their use in data assimilation to improve the analysis and subsequent forecast. This report focuses on the latter and describes the impact of optimisations on the assimilation of cloud radar and lidar to improve NWP forecasts. It leads on from its sister report, (WP-2000: optimisation of observation impact on NWP analyses, Fielding and Janisková, 2020a), by taking the previously optimised system and assessing the impact on longer-term assimilation experiments for different seasons. A thorough assessment of the impact on the forecast skill is achieved by assimilating CloudSat radar reflectivity and CALIPSO lidar backscatter on top of routinely assimilated observations and then comparing forecasts of a range of atmospheric variables against own analyses. Changes in forecast skill are also evaluated by comparing forecasts of top-of-atmosphere fluxes and surface precipitation with observations.

Overall, combining the results from four assimilation experiments to generate an 11-month period of assimilating CloudSat radar reflectivity and CALIPSO lidar backscatter has a positive impact on the globally-averaged forecast skill of large-scale variables such as temperature, wind and humidity at both short- and medium-range timescales. The greatest short-term impacts are seen in the forecast skill score of temperature at 100 hPa, where there is a statistically significant 0.3 % decrease in root-mean square error. Positive impacts of similar magnitude are seen in the forecast of vector wind and relative humidity, both of which are significant with 95 % confidence at a forecast lead time of 24 hours. This short-term positive impact appears to persist and grow into the medium-range with forecast skill improvements reaching a maximum at day 5-6. The impact of radar and lidar observations on the globally averaged forecast skill of temperature and vector wind at upper and middle levels both exceed 0.5 % in the medium-range. The impact on the forecast skill of relative humidity appears to be shorter lived for upper levels, but persists into the medium-range at lower-levels, mainly due to the improvements in the forecast skill of temperature.

In general, the impacts on forecast skill appear to be greater in the southern hemisphere than the northern hemisphere, particularly at shorter lead times. This is evident when comparing the change in forecast skill for geopotential height at 500 hPa for the two hemispheres. Both hemispheres show a reduction in forecast error in the medium-range; at days 5-6 there is around a 1 % reduction in root-mean square forecast error when assimilating the radar and lidar observations. However only the southern hemisphere has a significantly positive impact in the short-term. The experiment forecasts were also compared to independent observations. The assimilation of cloud radar reflectivity and lidar backscatter also has a positive impact on the model's radiation budget when compared with CERES TOA radiative fluxes. Both shortwave and longwave radiative fluxes show improvements of around 0.1 % for forecast lead times of up to 5 days. The greatest positive impact is in the first 48 hours of forecasts, with improvements in longwave radiation that are significant at a 95 % confidence level and significant in shortwave radiation for the 24-48 hour forecast period. The results of the comparison against GPCC rain-gauge data showed no significant increases in the skill of forecasting precipitation over land, but some areas, such as over North America, showed improvements in the first 24-hour forecasts.

Following observation quality monitoring of first-guess departures, a few issues in the observations and observation pre-processing were detected. Firstly, the pointing angle of the CALIOP lidar has a significant impact on the bias of FG departures due to specular reflection from horizontally aligned ice crystals. The original bias correction scheme was based on a period where the lidar had two different pointing angles, one close to 'nadir' and one at three degree angle 'off-nadir'. From December 2007 onwards the CALIOP pointing angle was definitively changed to 'off-nadir' so the bias correction scheme needed to be re-tuned for assimilation experiments after this date. Secondly, an issue with the observation pre-processing for lidar backscatter was detected in the 'JJA' experiment between June 2008 and August 2008. This sub-optimal use of the lidar data has now been corrected and results will be updated for the final project report. The detection and correction of these issues is an example of the power of observation quality monitoring against NWP data and is a demonstration of the capability to detect instrument problems faster than if only the observations themselves were monitored.

Acknowledgments

List of Acronyms

4D-Var	Four-Dimensional Variational Assimilation
AMSU-A	Advanced Microwave Sounding Unit - A
BUFR	Binary Universal Form for the Representation of meteorological data
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CERES	Clouds and the Earth's Radiant Energy System
CloudSat	NASA's cloud radar mission
CPR	Cloud Profiling Radar
EarthCARE	Earth, Clouds, Aerosols and Radiation Explorer
ECMWF	European Centre for Medium Range Weather Forecasts
ESA	European Space Agency
FG	First Guess
GPCC	Global Precipitation Climate Center
HIRS	High-resolution Infrared Radiation Sounder
IFS	Integrated Forecasting System of ECMWF
ITCZ	Inter-Tropical Convergence Zone
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NWP	Numerical Weather Prediction
OBS	OBServations
PEARL	Preparations for EarthCARE assimilation of Radar and Lidar
T95	Model grid with spectral truncation T95
T159	Model grid with spectral truncation T159
T255	Model grid with spectral truncation T255
TCo159	Model cubic octahedral grid with spectral truncation T159
UTC	Universal Time Coordinated
VarBC	Variational Bias Correction
VarQC	Variational Quality Control

References

Battaglia, A., T. Augustynek, S. Tanelli, and P. Kollias, 2011: Multiple scattering identification in space-borne w-band radar measurements of deep convective cores, *Journal of Geophysical Research: Atmospheres*, **116**(D19), n/a–n/a, D19201.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, **120**, 1367–1387.

Fielding, M. and M. Janisková, 2017a: Observation quality monitoring and pre-processing, WP-2000 report for the project Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, ESA ESTEC contract 4000116891/16/NL/LvH, 61 pp.

Fielding, M. and M. Janisková, 2020a: Optimisation of observation impact on forecasts, WP-2000 report for the project Preparations for EarthCARE Assimilation - Radar and Lidar Cloud Observations (PEARL Cloud), ESA ESTEC contract 4000128669/19/NL/CT.

Fielding, M. and M. Janisková, 2020a: Direct 4D-Var assimilation of space-borne cloud radar cloud radar reflectivity and lidar backscatter. Part I: Observation operator and implementation, *Q. J. R. Meteorol. Soc.*, **146(733),3877-3899**, doi:10.1002/qj.3878.

Fielding, M. and M. Janisková, 2020b: Optimisation of observation impact on analysis, WP-2000 report for the project Preparations for EarthCARE Assimilation - Radar and Lidar Cloud Observations (PEARL Cloud) to ESA ESTEC, contract 4000128669/19/NL/CT, 44 pp.

Fielding, M. and O. Stiller, 2019: Characterizing the representativity error of cloud profiling observations for data assimilation, *J. Geophys. Res. - Atmospheres*, **124**, 4086–4103.

Fielding, M. D. and M. Janisková, 2017b: Observation quality monitoring and pre-processing, WP-2000 report for the project Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, 4000116891/16/NL/LvH, pp.

Fielding, M. D. and M. Janisková, 2020b: Direct 4d-var assimilation of space-borne cloud radar reflectivity and lidar backscatter. part i: Observation operator and implementation, *Quarterly Journal of the Royal Meteorological Society*, **n/a**(n/a).

Geer, A. J., F. Baordo, N. Bormann, P. Chambon, S. J. 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, *Quarterly Journal of the Royal Meteorological Society*, **143**(709), 3189–3206.

Geer, A. J., F. Baordo, N. Bormann, and E. S. J., 2014: All-sky assimilation of microwave humidity sounders, ECMWF Technical Memorandum 741, 57 pp.

Geer, A. J., K. Lonitz, P. Weston, M. Kazumori, K. Okamoto, Y. Zhu, E. H. Liu, A. Collard, W. Bell, S. Migliorini, P. Chambon, N. Fourrié, M.-J. Kim, C. Köpken-Watts, and C. Schraff, 2018: All-sky satellite data assimilation at operational weather forecasting centres, *Quarterly Journal of the Royal Meteorological Society*, **144**(713), 1191–1217.

Janisková, M. and M. Fielding, 2018: Operational Assimilation of Space-borne radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, Contract report to the European Space Agency 4000116891/16/NL/LvH, 164 pp.

Janisková, M. and M. Fielding, 2020a: Assimilation system adaptation and maintenance, Vol. 1, WP-1000 report for the project Preparations for EarthCARE Assimilation - Radar and Lidar Cloud Observations (PEARL Cloud), ESA ESTEC contract 4000128669/19/NL/CT, 24 pp.

Janisková, M. and M. Fielding, 2020b: Direct 4D-Var assimilation of space-borne cloud radar cloud radar reflectivity and lidar backscatter. Part II: Impact on analysis and subsequent forecast, *Q. J. R. Meteorol. Soc.*, **146(733)**, doi:10.1002/qj.3879,3900–3916.

Janisková, M. and M. Fielding, 2022: Assimilation system adaptation and maintenance, WP-1000 report for the project Preparations for EarthCARE Assimilation - Radar and Lidar Cloud Observations (PEARL Cloud) to ESA ESTEC, 4000128669/19/NL/CT, pp.

Janisková, M. and P. Lopez, 2013: Linearized physics for data assimilation at ECMWF, in *S.K. Park and L. Xu (Eds), Data Assimilation for Atmospheric, Ocean and Hydrological Applications (Vol II), Springer-Verlag Berline Heidelberg*, pp. 251–286, doi:10.1007/978–3–642–35088–7–11.

Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens, 2008: Hydrometeor detection using cloudsat—an earth-orbiting 94-ghz cloud radar, *Journal of Atmospheric and Oceanic Technology*, **25**(4), 519–533.

Noel, V. and H. Chepfer, 2010: A global view of horizontally oriented crystals in ice clouds from cloud-aerosol lidar and infrared pathfinder satellite observation (calipso), *Journal of Geophysical Research: Atmospheres*, **115**(D4).

Rabier, H., F.and 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.

Schamm, K., M. Ziese, A. Becker, P. Finger, A. Meyer-Christoffer, U. Schneider, M. Schröder, and P. Stender, 2014: Global gridded precipitation over land: a description of the new gpcc first guess daily product, *Earth System Science Data*, **6**(1), 49–60.

Shonk, J. P. and R. J. Hogan, 2008: Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1d radiation schemes by using three regions at each height, *Journal of Climate*, **21**(11), 2352–2370.

Stephens, G., D. Vane, R. Boain, G. Mace, K. Sassen, Z. Wang, A. Illingwort, E. O'Connor, W. Rossow, and S. Durden, 2002: The CloudSat mission and the A-train, *Bull. Am. Meteorol. Soc.*, **83(12)**, 1771–1790.

Stubenrauch, C. J., W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. D. Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. C. Maddux, W. P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, and G. Zhao, 2013: Assessment of global cloud datasets from satellites: Project and database initiated by the gewex radiation panel, *Bulletin of the American Meteorological Society*, **94**(7), 1031 – 1049.

Tavolato, C. and L. Isaksen, 2015: On the use of a huber norm for observation quality control in the ecmwf 4d-var, *Quarterly Journal of the Royal Meteorological Society*, **141**(690), 1514–1527.

Teixeira, J., S. Cardoso, M. Bonazzola, J. Cole, A. DelGenio, C. DeMott, C. Franklin, C. Hannay, C. Jakob, Y. Jiao, J. Karlsson, H. Kitagawa, M. Köhler, A. Kuwano-Yoshida, C. LeDrian, J. Li,

A. Lock, M. J. Miller, P. Marquet, J. Martins, C. R. Mechoso, E. v. Meijgaard, I. Meinke, P. M. A. Miranda, D. Mironov, R. Neggers, H. L. Pan, D. A. Randall, P. J. Rasch, B. Rockel, W. B. Rossow, B. Ritter, A. P. Siebesma, P. M. M. Soares, F. J. Turk, P. A. Vaillancourt, A. V. Engeln, and M. Zhao, 2011: Tropical and subtropical cloud transitions in weather and climate prediction models: The gcss/wgne pacific cross-section intercomparison (gpci), *Journal of Climate*, **24**(20), 5223 – 5256.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. L. Smith, and J. E. Cooper, 1996: Clouds and the earth's radiant energy system (ceres): An earth observing system experiment, *Bulletin of the American Meteorological Society*, **77**(5), 853 – 868.

Winker, D., M. Vaughan, A. Omar, Y. Hu, K. Powell, Z. Liu, W. Hunt, and S. Young, 2009: Overview of the CALIPSO mission and CALIOP data processing algorithms, *J. Atmos. and Ocean. Tech.*, **26(7)**, 2310–2323.

Yang, P., Y. Hu, D. Winker, J. Zhao, C. Hostetler, L. Poole, B. Baum, and J. Reichardt, 2003: Enhanced lidar backscattering by quasi-horizontally oriented ice crystal plates in cirrus clouds, *Journal of Quantitative Spectroscopy and Radiative Transfer*, **79-80**, 1139–1157.

Young, S., M. Vaughan, R. Kuehn, and D. Winker, 2013: The retrieval of profiles of particulate extinction from cloud–aerosol lidar and infrared pathfinder satellite observations (calipso) data: Uncertainty and error sensitivity analyses, *J. Atmos. and Ocean. Tech.*, **30**(3), 395–428.