ESA CONTRACT REPORT

Contract Report to the European Space Agency

Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction

WP-4000 report: EarthCARE data handling and testing

M. Fielding, M. Janisková and R. Hogan

ESA ESTEC contract 4000116891/16/NL/LvH

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

ECMWF

Series: ECMWF ESA Project Report Series

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

Contract Report to the European Space Agency

Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction

WP-4000 report: EarthCARE data handling and testing

Authors: M. Fielding, M. Janisková and R. Hogan

ESA ESTEC contract 4000116891/16/NL/LvH

November 2017

ABSTRACT

This report details the specific modifications to the assimilation system at European Centre for Medium Range Weather Forecasts (ECMWF) that will be required for the inclusion of Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE) cloud radar and lidar data directly into the Four-Dimensional Variational (4D-Var) system. The work is divided into three sections. Firstly, the differences between EarthCARE observations compared to CloudSat (NASA's cloud radar mission) and CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) data are considered and updates to the forward models are made. Secondly, the off-line observation handling of raw EarthCARE L1b data into Binary Universal Format (BUFR) is outlined. In particular, the capability of EarthCARE's High Spectral Resolution Lidar (HSRL) to separate Mie and Rayleigh signals is accounted for. Finally, the system is tested using simulated EarthCARE nominal data of radar reflectivity and attenuated backscatter, where data flow through the assimilation system is tested.

Contents

1 Introduction

Preparations are underway to assimilate satellite cloud radar and lidar observations into the ECMWF Integrated Forecast System (IFS), with the aim for real-time assimilation of Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE; Illingworth et al., 2015) measurements upon its launch. Prior to launch, initial feasibility studies will make use of existing CloudSat (NASA's cloud radar mission; Stephens et al., 2002) and CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations; Winker et al., 2009) observations and, so far, developments have therefore been tailored towards these instruments. However, to ensure operational assimilation can commence in a timely manner, it is important to prepare the adaptions required of the system and the observation data handling before the EarthCARE satellite is launched and actual observations become available.

The necessary EarthCARE-specific developments can be roughly separated into two sections. The first involves adaptations to the 'off-line' observation data handling routines that convert the raw satellite data into the format recognised by the system. The second set of developments involve some adaptions to the 4D-Var data assimilation system itself related to the difference between EarthCARE and CloudSat and CALIPSO. These adaptions mainly relate to the forward operator, but also include changes to the observation pre-processing, such as the screening criteria. As was the case for previous developments, all the modifications will subsequently need to be tested thoroughly.

This document outlines the technical modifications to the ECMWF 4D-Var assimilation system to prepare for the operational assimilation of EarthCARE radar and lidar observations. In Section 2, the key differences between EarthCARE observations and CloudSat and CALIPSO observations are reviewed and the necessary updates to the forward models are described. The technical pre-processing and handling developments are given in Section 3. Section 4 gives examples of the testing of the system. A summary concludes the report in Section 5.

2 Differences between A-train instruments and EarthCARE

The CloudSat radar and CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) lidar share many characteristics of the EarthCARE CPR (Cloud Profiling Radar) and ATLID (ATmospheric LIDar) and can be considered synonymous for technical testing and feasibility studies. However, to assimilate the CPR and the ATLID with scientific meaning, several modifications are required. Tables 2.1 and 2.2 list the key differences between the CloudSat radar and the CPR, and CALIOP and ATLID respectively. For the radars, the main differences between the instruments relate to the larger antenna of the CPR, which leads to greater sensitivity and reduced multiple scattering. The CPR also detects the phase shift of signals, such that the Doppler velocity of targets can be measured. For the lidars, the differences are more significant; the wavelength of the instruments are also different, which leads to different cloud and molecular scattering properties. The smaller field of view of the ATLID also leads to reduced multiple scattering.

2.1 CPR-specific forward model developments

Radar sensitivity

Despite having a similar configuration to CloudSat, the EarthCARE CPR is expected to be 7 dB more sensitive, due to its lower orbit and larger antenna. This will allow a greater frequency of cloud detection, particularly for clouds with small particle sizes such as boundary layer clouds or cirrus. If all other parameters remain constant, this allows the detection of clouds with around 30 % smaller radii. To represent this change in the forward model, the screening for the lower bound on model-equivalent radar reflectivity will be lowered from -30 dBZ to -35 dBZ. Note the screening value used for CloudSat is lower than the reported minimum detectable reflectivity (see Table 2.1) as a greater sensitivity is obtained through noise reduction and averaging techniques. It is likely that the reported figure for the EarthCARE CPR is also conservative and therefore the screening threshold will be reviewed in the commissioning phase of the mission.

As a guide to the additional benefits that the increased sensitivity will bring, Fig. 2.1 shows the fraction of clouds missed by a radar for a range of sensitivities using the radar forward model applied to IFS model clouds. Overall the fraction of clouds missed by the CPR is halved (also reported in Stephens et al., 2002, but using high sensitivity ground based radar measurements) with similar absolute reductions across different hydrometeor types. The absolute amount of clouds missed is lowest for precipitation dominated scenes and greatest for ice cloud. Note that subgrid variability is not currently accounted for in the forward model so the probability distribution function of radar reflectivity 'observed' is likely to underestimate extremes leading to an averaging effect compared to actual observations. The increased sensitivity should benefit assimilation as it increases the dynamic range of the observations, which should increase the gaussianity of the observation errors by reducing the truncation of the observation probability distribution function and hence asymmetry of the first guess departures.

Multiple scattering

Multiple scattering occurs when transmitted pulses undergo more than one scattering even within the field of view of the instrument. To aid understanding, it is helpful to imagine two types of multiple scattering. The first, 'narrow angle' multiple scattering refers to photons that are forward scattered and remain within the field of view of the instrument as if they had not been scattered at all. The second, 'wide angle' multiple scattering, refers to all other scattering events, including photons that are scattered out of the field of view and then back in. Wide angle multiple scattering leads to so-called 'pulse stretching', where an apparent backscatter is detected beyond a cloud's boundaries due to the increased path length of these multiple scattered photons.

Characteristic	CloudSat	EarthCARE CPR
Specification	Nadir-pointing 94-GHz with 1.85 m diameter antenna	Nadir-pointing 94-GHz with 2.5 m diameter antenna
Satellite altitude	700 km	400 km
Resolution	480 m vertical, 1.1 km horizontal sampling, 1.7 km footprint	500 m vertical, 0.5 km horizontal sampling, 0.66 km footprint
Sensitivity	−28 dBZ	−35 dBZ
Misc.		Doppler capability

Table 2.1: Similarities and differences between the CloudSAT radar and the EarthCARE CPR relevant for data assimilation.

Characteristic	CALIOP	EarthCARE ATLID
Specification	3° off-Nadir dual wavelength (532 and 1064 nm) lidar with 1.00 m diameter telescope. Half-angle reciever field of view 65 μrad.	3° off-Nadir 335 nm lidar with 0.62 m diameter telescope. Half-angle receiver field of view 35 μrad.
Satellite altitude	700 km	400 km
Resolution	30 m vertical, 0.33 km horizontal sampling, 0.09 km footprint	103 m vertical, 0.285 km horizontal sampling, 0.03 km footprint
Misc.		High-spectral-resolution receiver with Rayleigh and Mie copolar and total cross-polar channels.

Table 2.2: Similarities and differences between the CALIOP lidar and the EarthCARE ATLID relevant for data assimilation.

For cloud radar wavelengths, where the wavelength is typically much longer than the hydrometeor size, wide-angle multiple scattering is the dominant source of multiple scattering. An option to account for this is included in the radar forward model by implementing the time dependent two-stream (TDTS; Hogan and Battaglia, 2008) scheme. The method involves tracking the propagation of radar pulses in time as well space. However, as this method is relatively costly to run and the explicit coding of the tangent linear and adjoint are not within the scope of the project, profiles that are suspected to be affected by multiple scattering are screened.

The full screening procedure for multiple scattering is described in Fielding and Janisková (2017), so we will only provide a short description here. Following Battaglia et al. (2011), the integrated reflectivity, dBZ_{int}, is used to blacklist profiles. Battaglia et al. (2011) showed that for CloudSat observations, integrating observations that exceed $8dBZ$ and using a threshold of 41.3 dBZ_{int} gave an optimum detection of cases where the magnitude of multiple scattering exceeds 3 dB. For EarthCARE's CPR, where multiple scattering is expected to be less due to a narrower field-of-view, they show the threshold for integrating observations increases to $12dBZ$. Applying the CloudSat thresholds to the CloudSat superobs results in the blacklisting of around 3% of observations (Fig. 2.2), in agreement with Battaglia et al. (2011). To estimate the fraction of observations that will be screened with EarthCARE, if we apply the EarthCARE thresholds to the CloudSat data the percentage of observation blacklisted is around 1%.

Doppler Velocity

In addition to radar reflectivity, the EarthCARE CPR will also measure Doppler velocity. While the scientific merits of assimilating Doppler velocity will not be considered in this project, the technical capability to include the measurement in BUFR format will be provided for future use. A basic forward model is also provided here

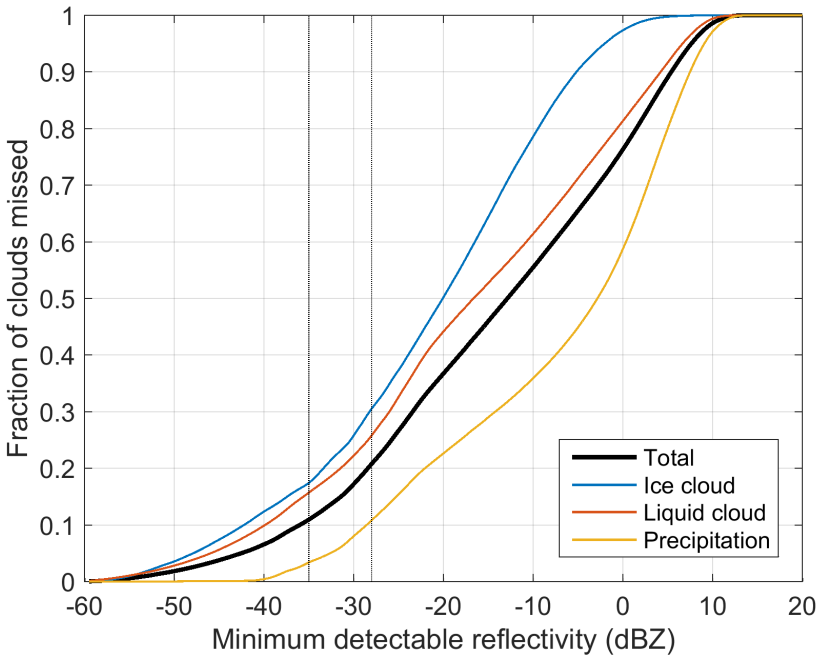

Figure 2.1: Fraction of the total number of cloud model gridboxes (black line) that have radar reflectivity below a given minimum detectable reflectivity along the CloudSat path during August 2007. Also shown is the fraction of ice cloud dominated (blue), liquid cloud dominated (red) and precipitation dominated (yellow) grid boxes.

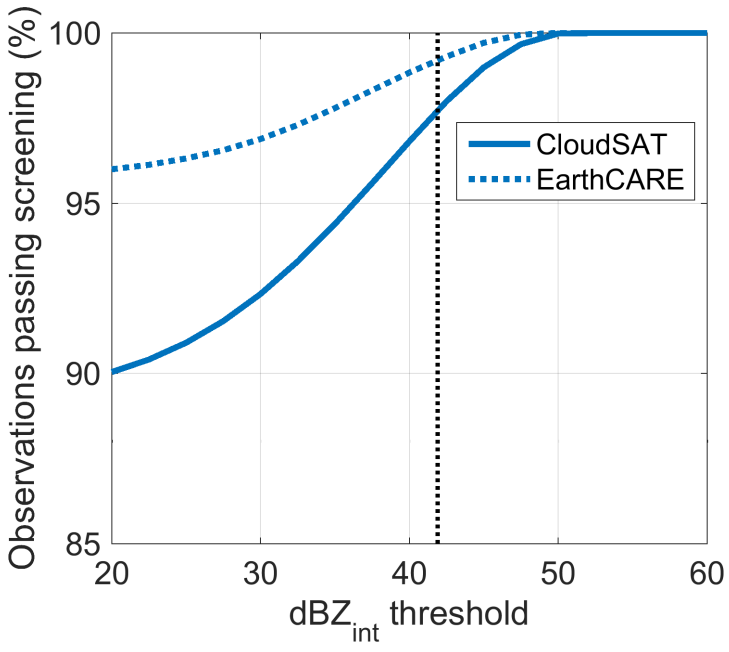

Figure 2.2: Percentage of observations passing screening for integrated radar reflectivity, as shown in WP-2000.

for testing purposes. The mean Doppler velocity, V_D, measured by a radar is the sum of the vertical air motion, w, and the mean radar reflectivity weighted droplet terminal fall velocity, v_d:

$$V_D = w + v_d, \tag{2.1}$$

Hydrometeor	c_x	d_x
Rain	386.8	0.67
Snow	16.8	0.527

Table 2.3: List of hydrometeor terminal velocity parameters used within the IFS.

where

$$v_d = -\frac{\int_0^\infty n(D)\eta(D)v(D)dD}{\int_0^\infty n(D)\eta(D)dD}, \tag{2.2}$$

$\eta(D)$ is the radar reflectivity for a hydrometeor of size D and $n(D)$ is the droplet number concentration (dependent on hydrometeor type and defined in WP-2000; Fielding and Janisková, 2017). The terminal velocity of a hydrometeor particle is defined by a power law:

$$v(D) = c_x D^{dx} \left(\frac{\rho_0}{\rho}\right)^{0.5} \tag{2.3}$$

where c_x and d_x are constants defined for each hydrometeor type in Table 2.3. The last term accounts for the decreased drag in less dense air where ρ_0 is a reference air density ($1\,\mathrm{kg\,m^{-3}}$).

Figure 2.3 gives an example of the revised forward model for the EarthCARE CPR and the corresponding Doppler velocity. Within ice cloud the Doppler velocity varies between $0\,\mathrm{m\,s^{-1}}$ to $1\,\mathrm{m\,s^{-1}}$, while for rain the absolute Doppler velocity is greater than $4\,\mathrm{m\,s^{-1}}$. Interestingly within ice cloud the Doppler velocity appears to be strongly correlated with radar reflectivity (in fact by definition from a single moment microphysics scheme) and therefore is unlikely to provide much additional benefit to assimilation. However, the benefit is likely to arise from a better constraint on the location of transitions between hydrometeor types, such as the melting layer. Also, Doppler is not affected by attenuation, so in rain it is likely to be a more direct measurement of rain rate. Finally, the Doppler velocity could be used to help optimize the forward model microphysical assumptions, particularly of the larger, faster falling, precipitation hydrometeors.

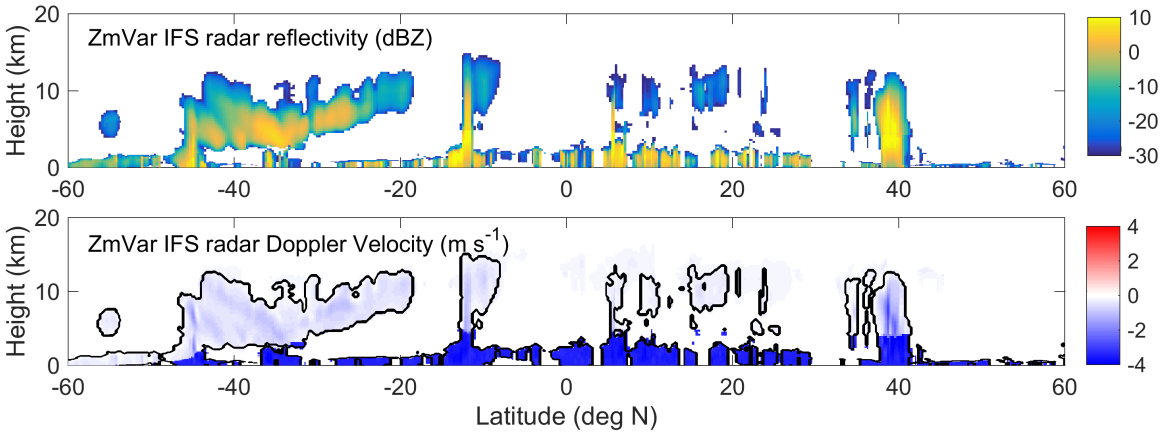

Figure 2.3: Panel (a) shows forward modelled radar reflectivity (dBZ) colocated with a CloudSat transect during August 2007, panel (b) shows the corresponding forward modelled Doppler velocity (ms^{-1}). To aid the reader, a black line denoting the cloud boundaries is also plotted.

2.2 ATLID-specific forward model developments

Hydrometeor scattering properties

The change of wavelength between the CALIPSO lidar and ATLID means the scattering properties for the forward model look up table need to be re-computed. Most differences in backscatter (Fig. 2.4) are superficial and much smaller than other sources of uncertainty. However, for larger rain drops (Fig. 2.4a and c) the backscatter is significantly greater at 355 nm. This is due to variations in the narrow backward peaks of the phase functions for large size parameter. Both the asymmetry parameter and single scattering albedo tend to be slightly less at 355 nm for all hydrometeor types (not shown).

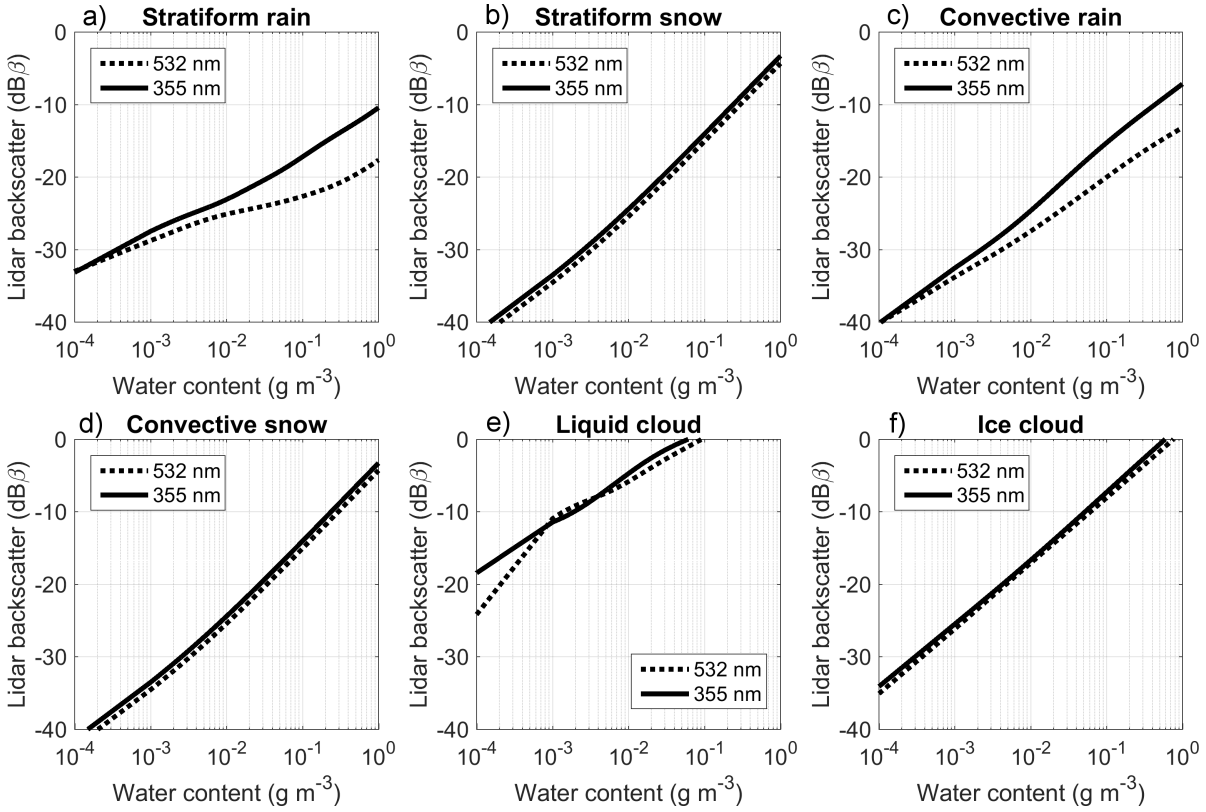

Figure 2.4: Comparison of hydrometeor lidar backscatter at 355 nm (black) and 532 nm (dashed) for six different hydrometeor types as a function of water content at 250K for solid hydrometeor types and 290 K for liquid hydrometeor types. dBβ is calculated from $10\log_{10}\beta'$*, where* β' *has units* $km^{-1}sr^{-1}$.

Molecular scattering and sensitivity issues

The molecular backscatter at 355 nm is roughly 5 times the value at 532 nm (Reverdy et al., 2015), which, while a simple change to the lidar forward model, has some implications for the detection of clouds. Unlike the CALIOP lidar, ATLID is an HSRL (High Spectral Resolution Lidar) with three separate channels: a Rayleigh channel, a co-polar Mie channel and a cross-polar Mie channel. Although the ability to separate the molecular backscatter (in the Rayleigh channel) from the cloud and aerosol backscatter (in the Mie channel) should theoretically allow the cloud extinction (a more direct measurement of cloud amount than attenuated backscatter) to be assimilated, initially only the total attenuated backscatter will be assimilated as its error characteristics are better known and it should behave similar to CALIPSO observations. However, the greater molecular backscat-

ter at 355 nm contained within the total attenuated backscatter will make cloud detection more difficult than at 532 nm (see Fig. 2.5).

To illustrate the additional difficulties in detecting cloud from attenuated backscatter at 355 nm, Fig. 2.6 shows the scattering ratio (SR), defined by Reverdy et al. (2015) as the ratio between total attenuated backscatter over the molecular-only attenuated backscatter. The ratio is equal to 1 in the absence of both particle scattering and any previous particle attenuation, and is generally greater than 1 in the presence of cloud/aerosol particles. The ratio can be below 1 anywhere where attenuation from cloud/aerosol has occurred.

Comparing Fig. 2.6a with Fig. 2.6b, we can see that the SR at 355 nm tends to be much smaller than that at 532 nm. At the shorter wavelength, there are also a significant amount of clouds where the SR is less than 1. The actual ability to detect clouds will partly depend on the noise in the molecular backscatter, which will have to be calibrated in the commissioning phase of the mission. Details of the cloud masking procedure and screening are provided in Sec. 3.3.

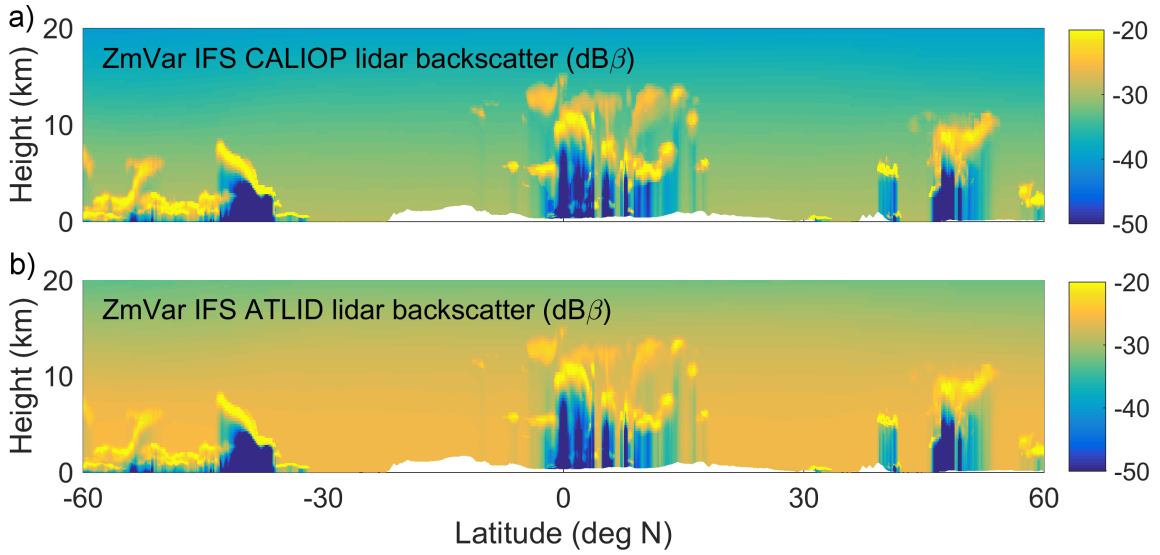

Figure 2.5: Forward modelled lidar backscatter at (a) 355 nm and (b) 532 nm using the full complex observation operator (including multiple scattering effects) for IFS model data collocated with a CALIPSO transect during August 2007.

Multiple scattering

A lidar signal entering a cloud is generally much more likely to be affected by multiple scattering than radar. At visible frequencies, the distance a photon travels before interacting with its medium (known as the mean-free path) is much shorter, hence the chances of two or more scattering events occurring are greater. In particular, in the optical limit, when the particles are much larger than the wavelength, the phase function becomes peaked at 0°, implying strong forward scattering. For cloud particles at lidar wavelengths, half of all scattered photons can be defined as scattering in a narrow forward-lobe.

Several methods exist to approximate the effect of multiple-scattering and, in particular, the Photon Variance-Covariance (PVC; Hogan, 2008) method is included in the complex observation operator and has been shown to be fast yet accurate compared with a slower explicit method. Indeed the PVC method is sufficiently fast for data assimilation (Fielding and Janisková, 2017), but, while the adjoint has been coded and validated, the explicit tangent linear would need to be coded for inclusion in the IFS, which is beyond the scope of the current project. Therefore, we will follow the approach of Di Michele et al. (2014b), who used the PVC method to train a simple multiple scattering approximation for CALIPSO, and train the method to EarthCARE's ATLID.

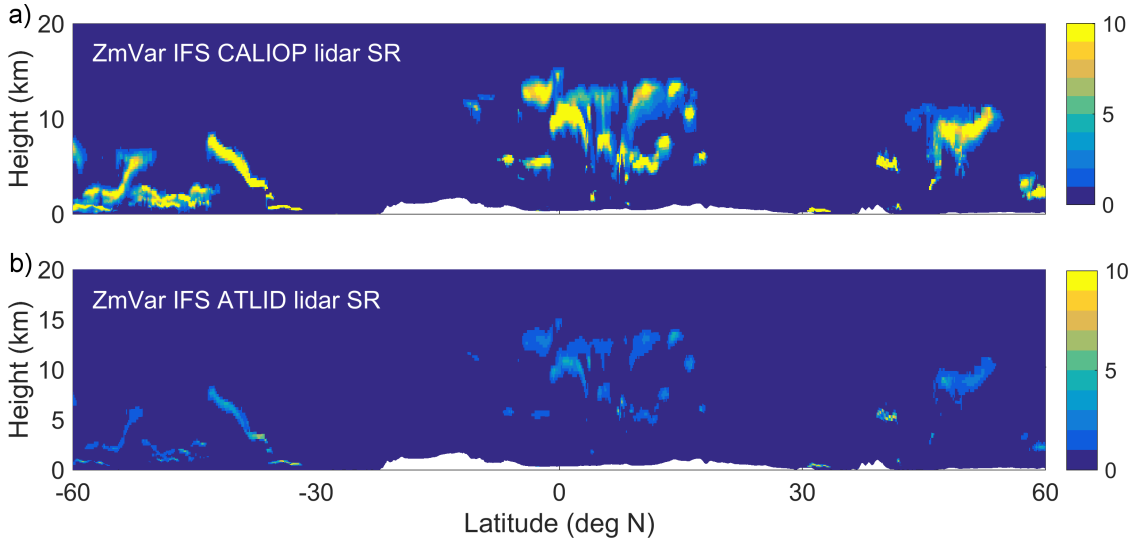

Figure 2.6: Same as Fig. 2.5, but showing the scattering ratio (SR; the ratio between total attenuated backscatter and molecular-only backscatter).

The simplest ways to approximate narrow-angle multiple scattering in the literature is using the so-called 'Platt coefficient', η, originally proposed by Platt (1973), where the apparent backscatter is defined as:

$$\beta(r)' = \beta(r)e^{-2(\eta\tau_{cloud}(r)+\tau_{gas}(r))} \tag{2.4}$$

and $\tau_{cloud}(r)$ and $\tau_{gas}(r)$ are the optical depth of hydrometeors and optical depth of gases between the instrument and a distance or range r respectively. Theoretically, the value of η is bounded between 1 (the single-scattering limit) and 1/2 (the wide field-of-view limit). The 'correct' value of η depends on the wavelength, scattering medium and the lidar geometry. For ATLID and typical hydrometeors, η lies somewhere between the two limits.

For any individual profile and lidar range gate, if the true apparent backscatter is known, it is possible to calculate an exact η by a straight-forward re-arrangement of Eq. 2.4:

$$\eta = -\frac{1}{2\tau_{cloud}}\log\left(\frac{\beta'}{\beta}\right) - \frac{\tau_{gas}}{\tau_{cloud}}. \tag{2.5}$$

While obtaining the true apparent backscatter and unattenuated backscatter from observations would be problematic, we can use the PVC to estimate the true apparent backscatter for a set of cloud profiles and then apply equation 2.5 to obtain an estimate for η. Note that this yields a range-dependent value for eta. In the following discussion, 24 hours of IFS model data collocated with real CALIPSO tracks during August 2007 have been used. Only model gridboxes where the lidar backscatter passes initial screening thresholds are included (see Fielding and Janisková, 2017).

Firstly, applying the approach to the CALIOP lidar configuration and stratifying by temperature (top row of Fig. 2.7) leads to two clear regimes related to hydrometeor type. Below 273 K, where hydrometeors are predominantly solid, η tends to be between 0.5-0.6, indicating that the effect of multiple scattering is close to the 'wide-angle' field-of-view regime. For small optical depths less than 1, an increase in η is observed for very low temperatures (around 200K), where particle sizes are smaller and the forward lobe is larger, so the lidar pulse is more likely to escape the field of view. At all temperatures, Fig. 2.7a contains larger values of η

than Fig. 2.7b because it includes cases where the optical depth is small enough for a non-negligible proportion of photons to undergo single-scattering only.

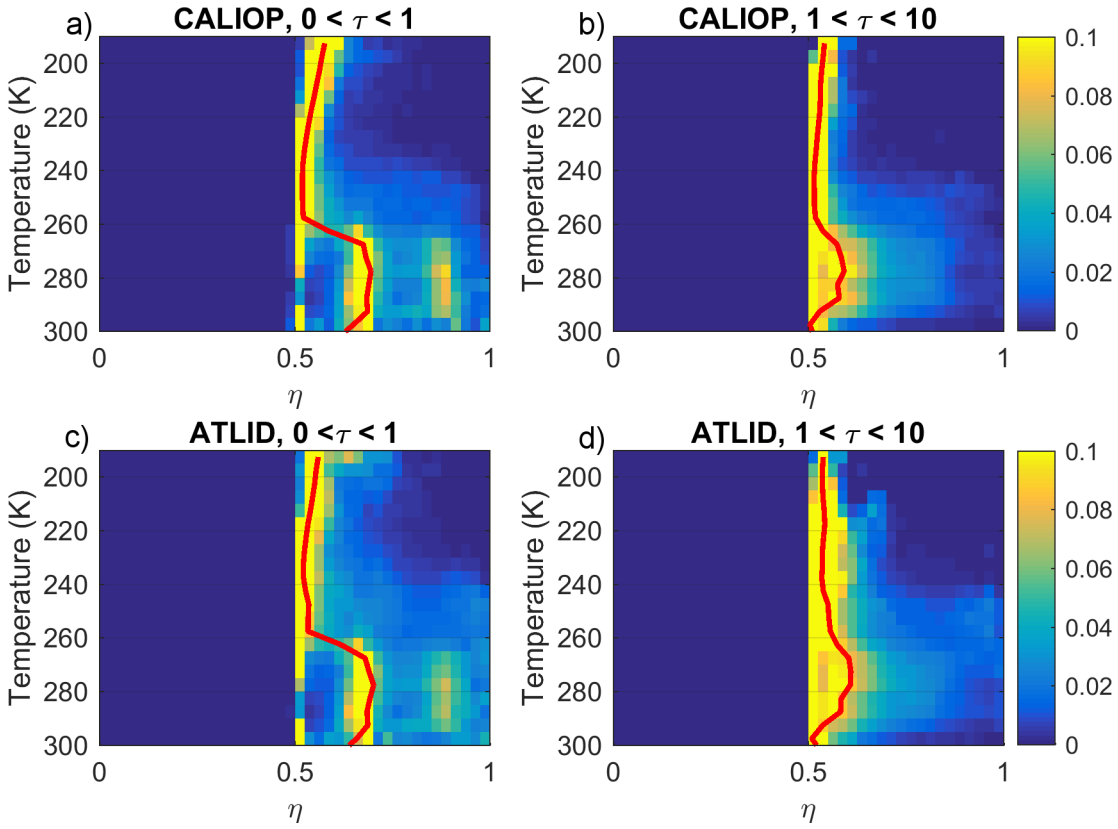

Figure 2.7: Frequency distributions of small-angle correction factor η derived using the Photon Variance-Covariance method to account for narrow-angle multiple scattering across a range of temperatures. Each panel only contains the backscatter from model levels where the optical depth integrated down from the top model level, τ, is within the range indicated in the plot title. Top row assumes the lidar configuration for CALIOP aboard CALIPSO, bottom row assumes the configuration for ATLID aboard EarthCARE (see Table 2.2). The red curve shows the median η for each temperature bin.

For optical depths below one and for temperatures above freezing, two modes are apparent. The greatest probability is for η of about 0.65, while another mode exists around 0.85. Broadly speaking, these relate to liquid precipitation and liquid cloud respectively. However, the optimum η at any given gate also depends on the hydrometeor properties encountered by the lidar beam above. Interestingly, for optical depths greater than 1 the bimodal shape is lost, with a broad range of possible η between 0.5-0.6. Again, this actually highlights the weakness in parameterizing η as a function of temperature; the most appropriate value of η for a particular profile depends on the cloud properties *above* rather than the layer itself.

Despite the differences in wavelength and footprint, the values of η for ATLID are surprisingly similar to those found for CALIOP. The median values (red lines in Fig. 2.7) are almost identical for optical depths less than 1, and slightly larger for ATLID for optical depths greater than 1. The similarity is a consequence of two competing effects. ATLID's shorter wavelength leads to a greater size parameter and hence a narrower forward lobe, which decreases η. At the same time ATLID's smaller footprint increases η as photons are more likely to leave its field of view.

Although ATLID's and CALIOP's field-of-view are much smaller than the CPR or CloudSat, wide-angle multiple scattering still has some effect on the apparent backscatter. To test this, the TDTS multiple scattering algorithm was also applied to the same cloud profiles analysed in Fig. 2.7. Because wide-angle multiple scat-

tering introduces interactions between lidar range gates, the physical justification for η to be bounded between 0.5 and 1 is no longer applicable. Further, there is no real physical basis for using η at all in cases where there is strong wide-angle multiple scattering, but the range of values η takes when applying Eq. 2.5 is a useful illustration of the range of uncertainty neglecting wide-angle multiple scattering can bring.

Figure 2.8 shows that when accounting for wide-angle multiple scattering, η has a greater range for CALIOP than ATLID due to its greater field-of-view. For optical depth less than 1, Fig. 2.8a and c are very similar to Fig. 2.7, indicating that wide-angle multiple scattering is only significant for greater optical depths. Wide-angle multiple scattering has the greatest effect for range gates where the temperature is warmer than freezing, because optical depths of liquid cloud tend to be greater than for ice cloud, and have a broader forward lobe. The greatest differences between CALIOP and ATLID are seen here. At 280K, the interquartile range for ATLID is confined to 0.48 to 0.61, compared to 0.35 to 0.58 for CALIOP.

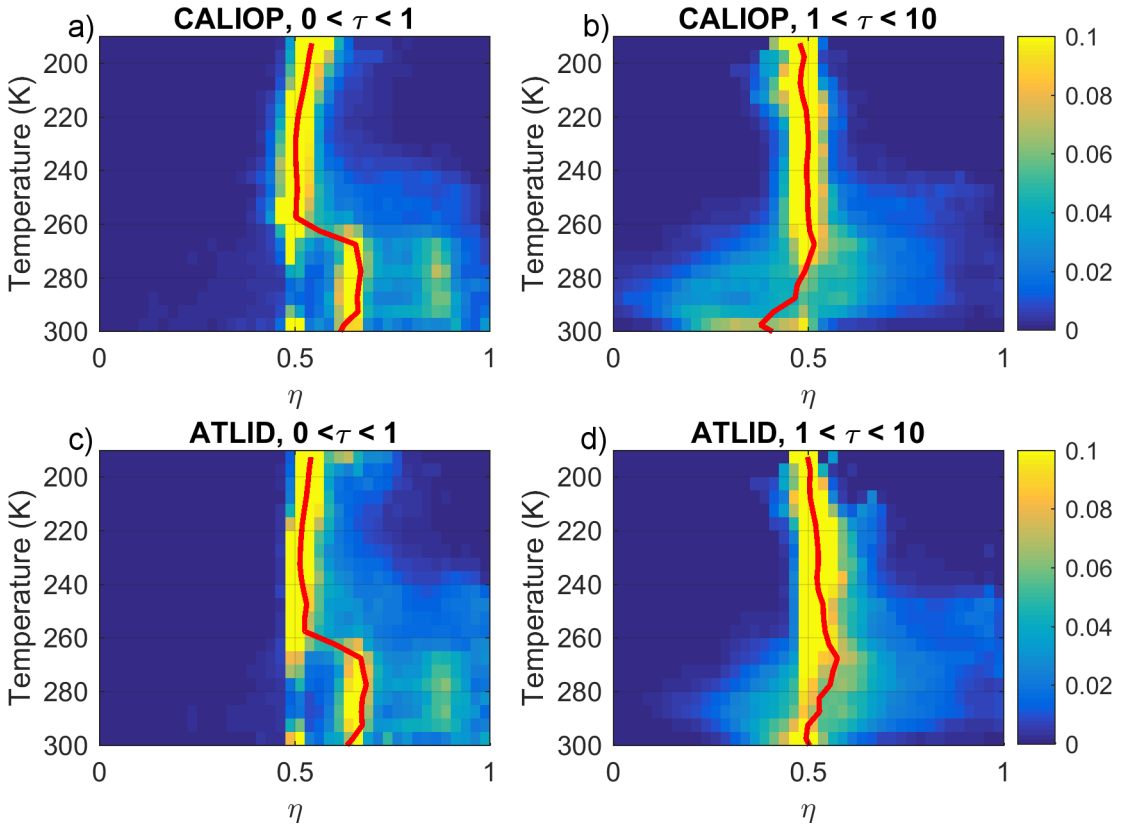

Figure 2.8: Same as Fig. 2.7, but also accounting for wide-angle multiple scattering using the TDTS method.

Alternatively, multiple scattering could be parameterized as a function of particle phase rather than temperature. Extending Eq. 2.4 to separate the optical depth due to solid hydrometeors, τ_{ice}, from the optical depth due to liquid hydrometeors, τ_{liquid}, yields the following equation:

$$\beta(r)' = \beta(r)e^{-2\left(\eta_{ice}\tau_{ice}(r)+\eta_{liq}\tau_{liq}(r)+\tau_{gas}(r)\right)} \tag{2.6}$$

where η_{ice} and η_{liq} are multiple scattering coefficients for solid and liquid hydrometeors respectively. Given that the multiple scattering is now parameterized by just two parameters, we can treat the problem as an optimisation exercise, where values for the parameters are found by minimizing differences between the backscatter from Eq. 2.6 and the 'true' backscatter given by the PVC or TDTS algorithms. Following a similar procedure used in

	η_{liq}	η_{ice}
CALIOP		
PVC	0.709	0.532
TDTS	0.350	0.525
ATLID		
PVC	0.709	0.543
TDTS	0.551	0.543

Table 2.4: Coefficients for the alternative phase-dependent multiple scattering parameterization (Eq. 2.6) for different lidars and multiple scattering algorithms.

WP-2000 (Fielding and Janisková, 2017) to optimize coefficients for the scattering look-up tables, we perform a non-linear regression, where the cost function is the root-mean square error of the backscatter in log space.

Using the same data as Figs. 2.7 and 2.8, Table 2.4 shows the optimum parameters for Eq. 2.6. The values broadly agree with the temperature dependent parameterization. Using just the PVC method, there is very little difference between the CALIOP and ATLID lidars; η_{liq} is identical to three significant figures, while η_{ice} is close to the wide-angle field-of-view limit for both lidars. The greatest difference is seen for η_{liq} found with the TDTS algorithm, with a lower value for CALIOP compared to ATLID, corroborating the results seen in Fig. 2.8.

In conclusion, there does not appear to be much justification for a complex temperature dependent parameterization for η. However, it is clear that η should be greater for liquid cloud and precipitation compared to ice cloud. If a temperature dependent parameterization is used (to be consistent with previous work), it would appear to be a reasonable approximation to use the Platt approximation with a value of $\eta = 0.55$ for temperatures colder than 273 K. Ideally, for liquid cloud and precipitation, the TDTS method should be used. As this is not possible within the scope of the project, using the Platt approximation with $\eta = 0.6$ for temperatures warmer than 273 K would seem to be reasonable compromise between the larger values obtained when the optical depth is low, with the smaller values seen for greater optical depth. As the lidar signal is strongly attenuated in liquid cloud, the observation error will automatically be greater, but it might need to be revised higher to take into account for the neglect of wide-angle multiple scattering. Alternatively, if a phase dependent parameterization is used, Table 2.4 provides coefficients for use in Eq. 2.6. This approach has the advantage over the temperature dependent approach of accounting for mixed-phase cloud.

3 EarthCARE data handling

In this section we will review the technical preparations for handling EarthCARE data. Firstly we will identify the data products to ingest with consideration of the likely latencies. Then we will discuss the conversion tools for shaping the data into a readable format by the data assimilation system. Finally we will adapt the selection and initial screening of the data using test data as an example.

3.1 Data identification and nominal test data

Choosing which radar and lidar products to assimilate is a trade-off between the product latency and the gain in information or value added by additional processing. Lower level products, which are typically closer to the physical measurement of the instrument, are likely to be available sooner compared to higher level products which are released with a longer delay. A further consideration, is that although higher level products may contain measurements of variables closer to model variables (such as cloud water content), their errors may be more complicated and/or contain biases that are difficult to characterize.

However, to be certain of reasonable availability of the data we will only assimilate data from the following L1B products:

- L1b C-NOM (CPR Nominal data). This product (generated by JAXA) will contain the calibrated radar reflectivity factor and basic quality control information.
- L1b A-NOM (ATLID Nominal data). This product (generated by ESA) will contain the calibrated co-polar and cross-polar Mie and Rayleigh channels, and basic quality control information.

Nominally, 60% of these data should be available within 5.5 hours of being taken (Illingworth et al., 2015), with the worst case delay typically 12 hours (depending on the choice of down-linking ground station location). As the data assimilation window at ECMWF is currently 12 hours long, we can expect to have around 80% of L1B data available in time, although as the worst case delay matches the data assimilation window, it is theoretically possible for no data to be available in time if there are successive 'blind orbits'.

3.2 Data ingestion (BUFR)

As explained in the assimilation system development for cloud radar and lidar observations work package (WP-3000; Janisková et al., 2017), all observations entering the ECMWF data assimilation system must first be converted to BUFR format. The BUFR sequences for the CPR (Table 3.1) and ATLID (Table 3.2) are similar to those for CloudSat and CALIPSO respectively. For the CPR, two additional descriptors has been added for Doppler Velocity and its corresponding uncertainty. For ATLID, four additional descriptors have been added to accommodate the additional Mie and Rayleigh channels and their respective uncertainties. In addition to updating the BUFR sequences, the BUFR conversion tool was modified to be compatible with the hdf5 data format used by the EarthCARE L1b products. The variable names of the L1b data mapped to the BUFR descriptors is also given in Tables 3.1 and 3.2.

3.3 Data selection, pre-processing and screening

The data selection, pre-processing and screening follows a similar method to that described for CloudSat and CALIPSO in WP-2000 (Fielding and Janisková, 2017) and WP-3000 (Janisková et al., 2017) . Once the raw

Code	Description	Scale	Ref.	Width	Units	Comment and/or L1b C-NOM variable name
0 01 007	Satellite identifier	0	0	10		satID=TBD
0 02 019	Satellite instruments	0	0	11		instrumentID=TBD
3 01 011	Year, month, day					'profileTime'
3 01 013	Hour, minute, second					'profileTime'
3 01 021	Latitude / Longitude (high accuracy)					'longitude' and 'latitude'
0 10 033	Altitude (Platform to Ellipsoid)	1	0	27	m	
0 25 182	L1 processing flag					
0 25 181	L2 processing flag					
0 21 194	Data classification type	0	0	4	CODE TABLE	0 Surface 1 Cloud likely 2 Cloud probable 3 Cloud possible 4 Unclassified 15 Missing value
0 33 003	Quality information				CODE TABLE	
0 08 049	Number of observations					
0 21 195	Cloud fraction	3	0	11		
0 31 000	Delayed replication factor					
0 02 153	Satellite Channel Centre Frequency	-8	0	26	Hz	94 GHz
0 21 197	Height	0	-1000	17	m	
0 21 192	Cloud radar reflectivity	2	-9000	15	dBZ	'radarReflectivityFactor'
0 21 193	Cloud radar reflectivity uncertainty	2	0	9	dB	see text
0 21 198	Doppler velocity	0	-20	6	m s^{-1}	
0 21 199	Doppler velocity uncertainty	0	-20	6	m s^{-1}	

Table 3.1: BUFR sequence for EarthCARE CPR observations.

Code	Description	Scale	Ref.	Width	Units	Comment
0 01 007	Satellite identifier	0	0	10		satID=TBD
0 02 019	Satellite instruments	0	0	11		instrumentID=TBD
3 01 011	Year, month, day					
3 01 013	Hour, minute, second					
3 01 021	Latitude / Longitude (high accuracy)					
0 10 033	Altitude (Platform to Ellipsoid)	1	0	27	m	
0 25 182	L1 processing flag					
0 25 181	L2 processing flag					
0 21 194	Data classification type	0	0	4	CODE TABLE	0 Surface 1 Cloud 2 Aerosol 3 Unclassified 15 Missing value
0 33 003	Quality information				CODE TABLE	
0 08 049	Number of observations					
0 21 195	Cloud fraction	3	0	11		
0 31 000	Delayed replication factor					
0 02 153	Satellite Channel wavelength	9	0	16	m	
0 21 197	Height	0	-1000	17	m	
0 21 202	Mie Copolar Attenuated Backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$
0 21 203	Uncertainty in Mie Copolar Attenuated Backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$
0 21 204	Rayleigh Attenuated Backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$
0 21 205	Uncertainty in Rayleigh attenuated backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$
0 21 206	Total attenuated backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$
0 21 207	Uncertainty in total attenuated backscatter	2	-9000	15	$m^{-1}\,sr^{-1}$	Range: 0 to $0.1\,m^{-1}\,sr^{-1}$

Table 3.2: BUFR sequence for ATLID observations.

input data has been converted to BUFR, the observations undergo a pre-processing task where the observations are averaged to a specified grid and height levels (typically the native model grid and height resolution). In this step, the 'cloud fraction' and standard deviation of the averaged observations are also computed. However, for the CloudSat and CALIPSO processing all these computations could make use of level 2 data. In particular, for the real-time assimilation of CPR and ATLID, a cloud mask will need to computed 'in house' as this is not contained within the L1b products. Further, it is expected that the radar reflectivity will need to be noise corrected to ensure the detection and correct values of clouds with small radar reflectivity such as cirrus. Some minor changes are also made to the screening thresholds.

CPR noise subtraction and cloud mask

The radar reflectivity contained within the L1b C-NOM product will not be corrected for noise, nor will it contain a cloud mask or an estimate of the measurement uncertainty. As these are all required for data assimilation, we will compute them 'in-house'. Firstly, the mean noise will be computed following the approach of Hildebrand and Sekhon (1974) as is implemented in the L2 C-PRO algorithm and documented in the algorithm theoretical basis document (ATBD; Kollias et al., 2016):

1. For each profile, the power, $P(i)$, is sorted in ascending order
2. Set n=N, where N is the total number of range gates
3. While the profile is determined to be coloured noise according to:

$$n\sum_{i=1}^{n} P(i)^2 - 2\left(\sum_{i=1}^{n} P(i)\right)^2 = \begin{cases} < 0, & \text{White noise} \\ > 0, & \text{Coloured noise} \end{cases} \tag{3.1}$$

   set n=n-1 and repeat.
4. When the profile is determined to be free from signal (white noise), the mean noise power is calculated by averaging the remaining gates.

The standard deviation of the noise, σ_N is also calculated from the remaining gates under the assumption of Gaussian white noise. It is possible that a given profile may not contain sufficient target-free range gates to provide a reasonable estimate of the mean noise. Therefore, after estimating the noise for each profile a threshold noise is computed as:

$$P_{NT} = \text{median}(\bar{P}) + 3\text{median}(\sigma_N). \tag{3.2}$$

Any profiles where $\bar{P}$ is greater than P_{NT} will have $\overline{P_N}$ set equal to P_{NT}.

The cloud mask (known as the 'significant detection mask', SDM, in the ATBD) can then be produced using the following rules:

1. For $i = 1...n$
2. $SDM(i) = 1$ if $P(i) \geq \overline{P_N} + 1\sigma_N$
3. $SDM(i) = 0$ if $P(i) < \overline{P_N} + 1\sigma_N$
   OR $\{\overline{P_N} + 1\sigma_N < \overline{P_N} + 1\sigma_N \leq \overline{P_N} + 3\sigma_N$
   AND $P(i-1) \geq \overline{P_N} + 3\sigma_N$
   AND $P(i+1) < \overline{P_N} + 1\sigma_N\}$

where $i+1$ refers to the range gate immediately above gate i. The third condition is in place to reduce the stretching of cloud top caused by the 500 m resolution of the CPR. An optional speckle filter is then applied to remove isolated false detections that are likely when using a low detection threshold of 1 s.d. above the mean noise (see e.g. Clothiaux et al., 1995). The speckle filter will be tuned in the commissioning phase of the mission.

The measurement uncertainty in radar reflectivity in logarithmic units, ΔZ_{dB}, can be estimated from the linear signal-to-noise ration (SNR) using the following equation (Hogan et al., 2005):

$$\Delta Z_{dB} = \frac{4.343}{\sqrt{N}}(1+\frac{1}{SNR}), \tag{3.3}$$

where N is the number of radar pulses averaged (around 500 at the native CPR resolution) and we have assumed each pulse is independent.

ATLID cloud mask

The L1b A-NOM will also not contain a cloud mask. Whereas significant detections away from the near-surface at cloud radar wavelengths are highly likely to be from hydrometeors, aerosols in significant quantities can be readily detected by the ATLID. Sophisticated classification algorithms exist for the separation of cloud and aerosol signals for CALIPSO based on the backscatter signals and supplementary information such as temperature, altitude and location (e.g., Liu et al., 2004).

For ATLID, there are several different L2 products, sequentially produced, to transform the L1b raw backscatter to the particle type classification product. Firstly, the feature mask product is produced, which does not distinguish between cloud or aerosol types. Next, Particle Optical Properties (POP) products are produced derived directly using the observed three channels of the HSRL. The planetary boundary layer height is then derived exploiting a signal in the gradient of attenuated backscatter. Finally the particle type classification product is generated using information from the various products.

Implementing in-house versions of all these processes is technically possible, but is not within the scope of this project. For data assimilation of cloud it will initially be sufficient to apply an empirically derived threshold in total attenuated backscatter (see Sec. 4) calibrated in the commissioning phase of the mission. However, one of the key advantages of ATLID versus CALIPSO is its capability to separate Mie from Rayleigh signals. It should be therefore possible for a greater sensitivity to be obtained by assimilating only the Mie co-polar channel. However, the exact noise characteristics are not known at this stage and will require further testing.

Taking advantage of the HSRL (if time allows), a more sophisticated method could be employed following the JAXA level 2 product ATBD, based on all three HSRL channels, using a diagnostic parameter P_r, equivalent to the total unattenuated particle backscatter:

$$P_r = \left[\frac{\beta_{\text{mie,co}} + \beta_{\text{mie+ray,cr}}}{\beta_{ray,co}} - \delta\right]\frac{\beta_m}{1+\delta}, \tag{3.4}$$

where δ is the depolarization ratio, β_m is the unattenuated molecular backscatter, $\beta_{\text{mie,co}}$ is the backscatter from the co-polar Mie channel, $\beta_{\text{ray,co}}$ is the backscatter from the co-polar Rayleigh channel, and $\beta_{\text{mie+ray,cr}}$ is the backscatter from the cross-polar channel. The value of P_r can then be used to distinguish between clouds, aerosol and molecular-only layers. The exact thresholds of P_r to determine the classification would be optimised after launch.

Indicator	Min	Max	Reason
Height (km)	1	20	Lower limit (relative to surface) to avoid surface return, upper limit (absolute) to discard spurious signals (although some stratospheric clouds may be removed)
CF_{IFS}	0.2	1.0	To avoid non-linearity and representativity issues
CF_{obs}	0.2	1.0	To avoid non-linearity and representativity
dBZ_{IFS}, dBZ_{obs}	−35	20.0	Plausible bounds for radar
$dB\beta_{IFS}$, $dB\beta_{obs}$	−36	0.0	Plausible bounds for lidar
FG departures	−20	20	Remove large departures
dBZ_{int}	0.0	41.3	Radar multiple scattering not modelled by observation operator
β_{int}	0.0	0.04	Avoid observations with excessive attenuation

Table 3.3: Screening thresholds for CPR and ATLID observations.

Screening

Not all the L1b radar reflectivity and lidar backscatter is suitable for assimilation; care must be taken to remove data that could be detrimental to the analysis. Unsuitable data include observations that the forward model does not represent (such as ground return or aerosols) and situations where the cost function may be strongly non-linear or ill-posed (such as when the model sees cloud and the observations do not, or lidar profiles with strong attenuation). The screening will use the same indicators as in WP-2000 (Fielding and Janisková, 2017) for CloudSat and CALIPSO, but with revised values (Table 3.3). The main differences are a reduction in the minimum radar reflectivity due to the increased sensitivity of the CPR and an increase in the minimum lidar backscatter due to the reduced sensitivity of the ATLID total backscatter relative to the background molecular signal.

4 Technical testing

To prepare and test the L2 EarthCARE products prior to EarthCARE's launch, L1b nominal test data have been prepared for both the CPR (thanks to Aleksandra Tatarevic) and ATLID (thanks to ESA). It is worth noting that the C-NOM is a best estimate prepared using information from the JAXA product description document. The test data itself is generated using simulations conducted with Environment Canada's high-resolution NWP model known as the Global Environment Multi-scale Model (GEM; Côté et al., 1998). GEM is a nested model that can run at a range of horizontal grid-spacings. For the test data, a high resolution of 0.25 km was applied for a 150 km wide swath along the A-train satellite orbit on 7th December 2014 (39316D). The so-called 'Halifax' transect begins at Greenland, crosses Eastern Canada and ends in the Caribbean. The GEM model output was then used as input for the EarthCARE simulator, ECSIM.

The Halifax transect contains three distinct regimes, which is ideal for testing the system. Between 20 °N and 30 °N there is a cluster of shallow cumulus convection over ocean. Further North, between 35°N and 45°N there is a region of deeper convection and a frontal system with mixed phase clouds occupying the whole free troposphere. Towards Greenland between 55°N and 65°N there are liquid water topped stratiform clouds with some convection embedded.

4.1 BUFR conversion and pre-processing

The first task of an assimilation experiment is to process the incoming data into BUFR format so that it can be read by the system. Figure 4.1 shows the direct output of the BUFR converter tool described in Sec. 3.2 when the Halifax test data is given as input (Fig. 4.1a). The radar reflectivity contained within the BUFR file is at native resolution, but is thinned to every 10th profile for plotting. Although not obvious in the plot of thinned BUFR data, once averaged to the model grid resolution (Fig. 4.1b) the cumulus regime is clearly visible with radar reflectivity maxima around -20 dBZ. The standard deviation of the observations (Fig. 4.1c) is reasonably large (between 3 dB to 5 dB), and coupled with low cloud fraction (Fig. 4.1d), suggests large representativity errors, as will be shown later.

Within the region of deeper convection, the radar reflectivity exhibits a greater standard deviation, particularly within the precipitating congestus clouds. As expected, there is less variability in the ice cloud of the frontal-like region. Note the number of observations (Fig. 4.1e) shows a striped pattern due to the way the satellite track traverses the model gridboxes. To account for this, superobs with very few observations (typically less than 5) will be blacklisted or have substantially greater observation error.

The simulated lidar backscatter for the Halifax transect has also been converted to BUFR and is shown in Fig. 4.2a. Using a simple threshold in lidar backscatter appears to be sufficient in masking most cloud and precipitation. The shallow cumulus regime can be seen at the native resolution, but again is clearer in the superobs. For the deep convective regime, regions of super cooled cloud liquid water are apparent where there are large backscatter values and strong attenuation of the signal. A band of ice-only cloud is located around 40°N. In conclusion, the adapted data handling and processing routines behave as expected and are ready for real data once EarthCARE is launched. Small changes to the incoming data format, such as variable name or type, will be straightforward to correct.

4.2 ODB integration and screening

Once the observations have been pre-processed, they are used to populate the Observation Data Base (ODB). Also included in the ODB is any additional information required for assimilation. Figure 4.3 shows a selection

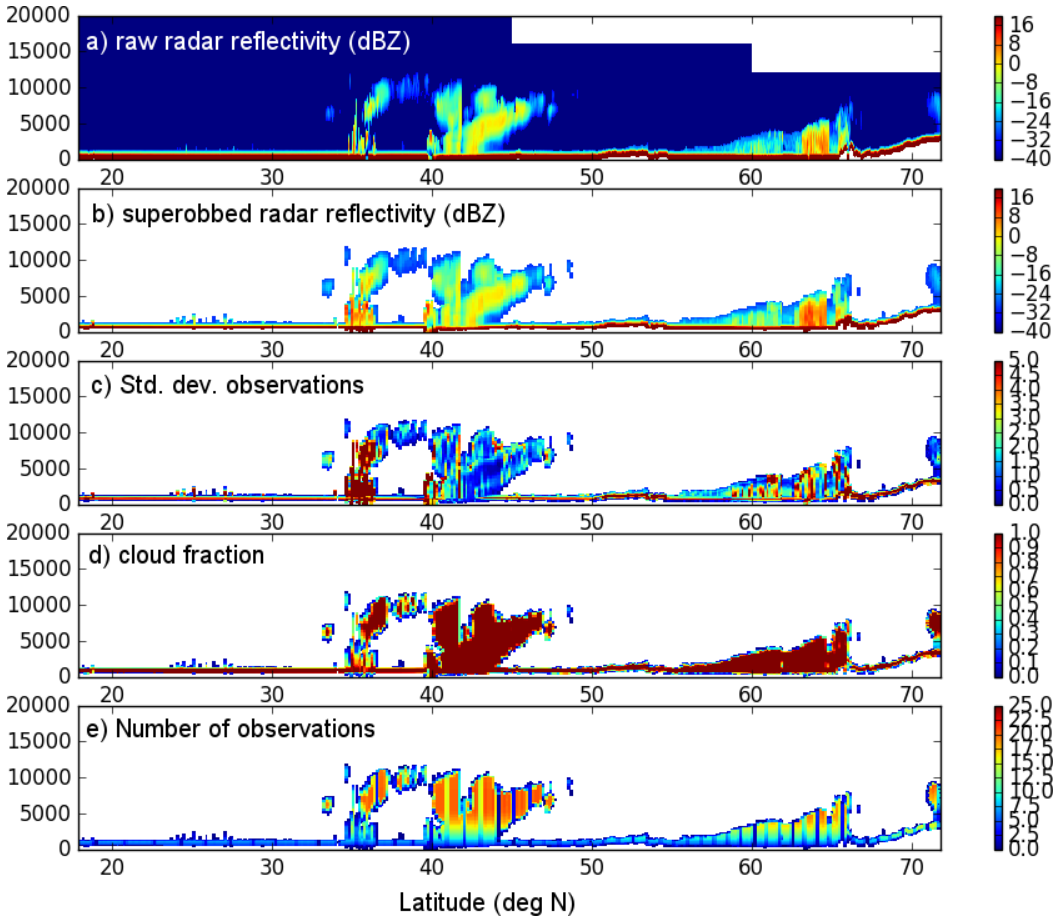

Figure 4.1: Direct output of the pre-processing of the EarthCARE L1b C-NOM test data. Panels show: (a) native resolution radar reflectivity from input BUFR test file, (b) Superobbed radar reflectivity to native model resolution (TCo639), (c) standard deviation of radar reflectivity within each superob, (d) cloud fraction of superob (number of cloudy observations divided by total number of observations), (e) total number of observations.

of fields contained within the ODB. The superobbed radar reflectivity is shown in Fig. 4.3a (identical to Fig. 4.1b), while the corresponding IFS model equivalent radar reflectivity is shown in Fig. 4.3b. Note no significant scientific insight should be drawn from comparing the two figures, as both observations are simulated. However, it is a useful testing exercise and confidence from the processing system is drawn from the fact the two models actually produce remarkably similar observed radar reflectivity.

The bias correction (Fig. 4.3c) uses the look-up tables trained from CloudSat data (see WP-2000), which combine biases in both the observation and the model. Unfortunately, there is no way to separate these two sources of bias, so until a climatology of EarthCARE observations is available we will use the CloudSat bias correction as an estimate of the CPR bias. The observation errors (Fig. 4.3d) are partially derived from CloudSat data and in a less direct way than the bias correction. For example, the representativity error is scaled using a climatology of correlation statistics derived from CloudSat observations. As both CloudSat and CPR have similar characteristics, the measurement correlation is expected to be sufficiently similar that the representativity error scaling is not significantly different.

A graphical representation of the screening of the C-NOM data is shown in Fig. 4.3f. The majority of data points actually pass screening (shown in red), with the main cause for rejection being surface return (all observations below 1 km above the surface are rejected). Also shown for illustrative purposes are data where the forward modelled radar reflectivity is below the sensitivity of the CPR (yellow); only a small amount of forward

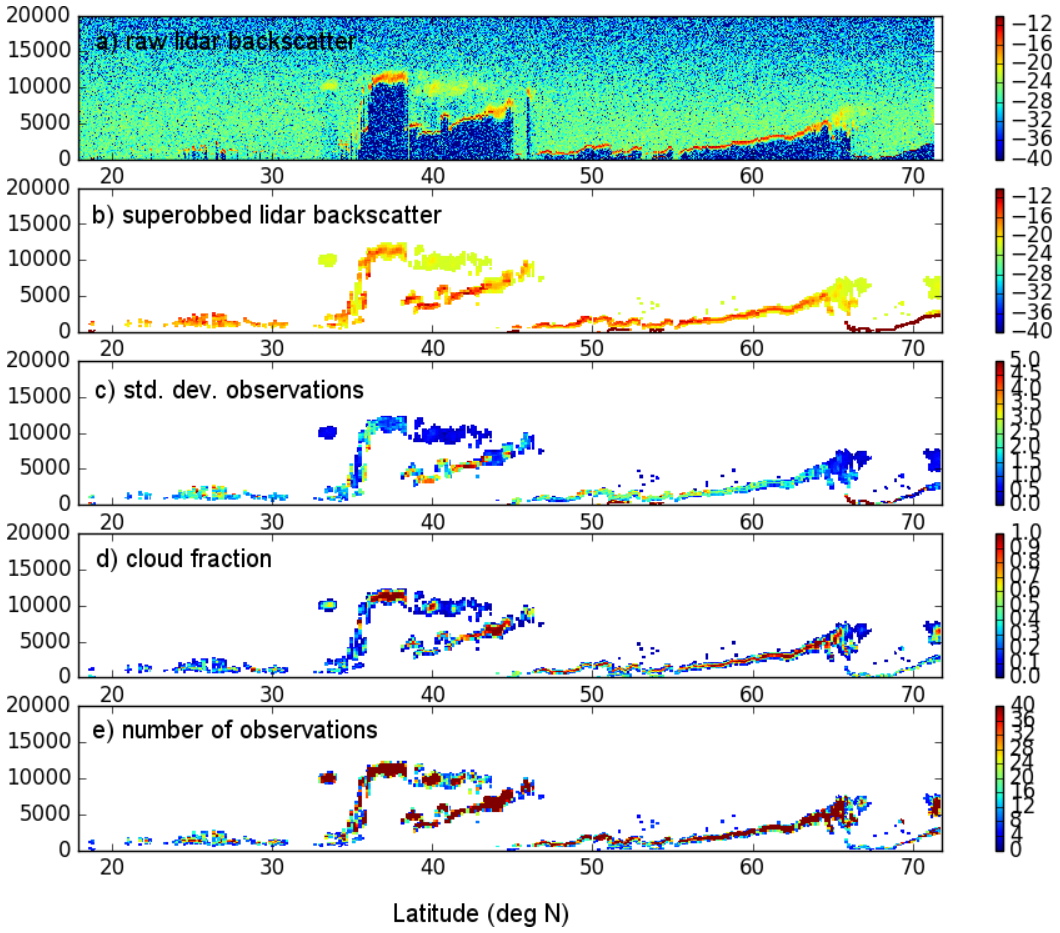

Figure 4.2: Direct output of the pre-processing of the EarthCARE L1b A-NOM test data. Panels show: (a) native resolution lidar backscatter from input BUFR test file, (b) Superobbed lidar backscatter to native model resolution (TCo639), (c) standard deviation of lidar backscatter within each superob, (d) cloud fraction of superob (number of cloudy observations divided by total number of observations), (e) total number of observations.

modelled radar reflectivity is below the sensitivity of the CPR.

Figure 4.4 shows the same fields stored in the ODB as Fig. 4.3, but for lidar backscatter. Due to the much greater attenuation of the lidar signal, less data is stored in the ODB than for the CPR. The first guess departures (Fig. 4.4e) tend to be quite large, particularly where super cooled liquid water exists in the A-NOM, but not in the model equivalent lidar backscatter. As for the CPR, the bias correction (Fig. 4.4c) look-up table uses a climatology built using data from a different instrument. However, the differences between CALIOP and ATLID are more significant than those between CloudSat and the CPR (see Sec. 2), so the bias correction must be used with caution and will be tested carefully in the commissioning phase of the mission. The screening for lidar backscatter (Fig. 4.4f) is generally either 'pass' (red) or 'excessive first guess departures' (green).

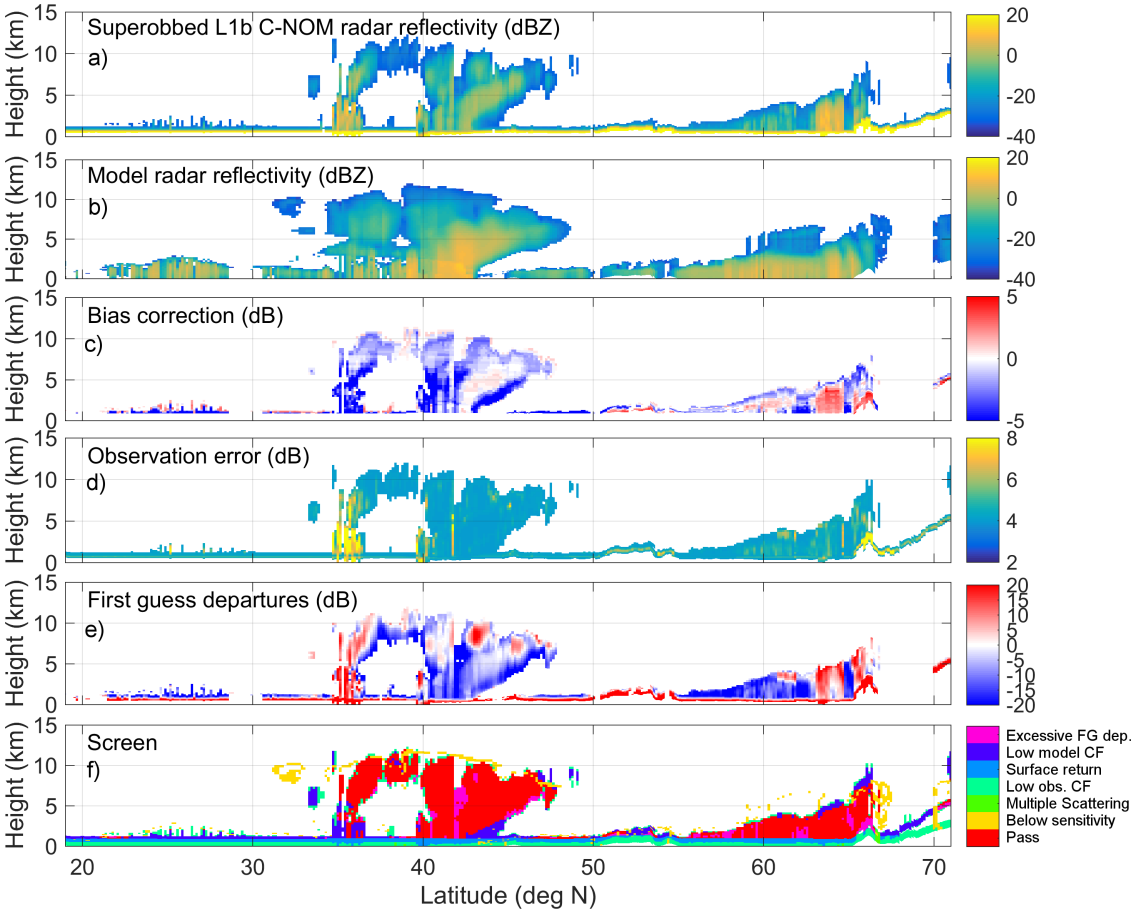

Figure 4.3: Example of observations stored in the ODB related to the EarthCARE L1b C-NOM test data. Panels show: (a) Superobbed radar reflectivity to native model resolution (TCo639; dBZ), (b) forward modelled radar reflectivity (dBZ), (c) bias correction applied to model equivalent (dB), (d) total observation error of each superob (dB), (e) observation minus background first guess departures (dB), (f) Screening value, blank areas have a cloud mask of zero in either observation or model equivalent.

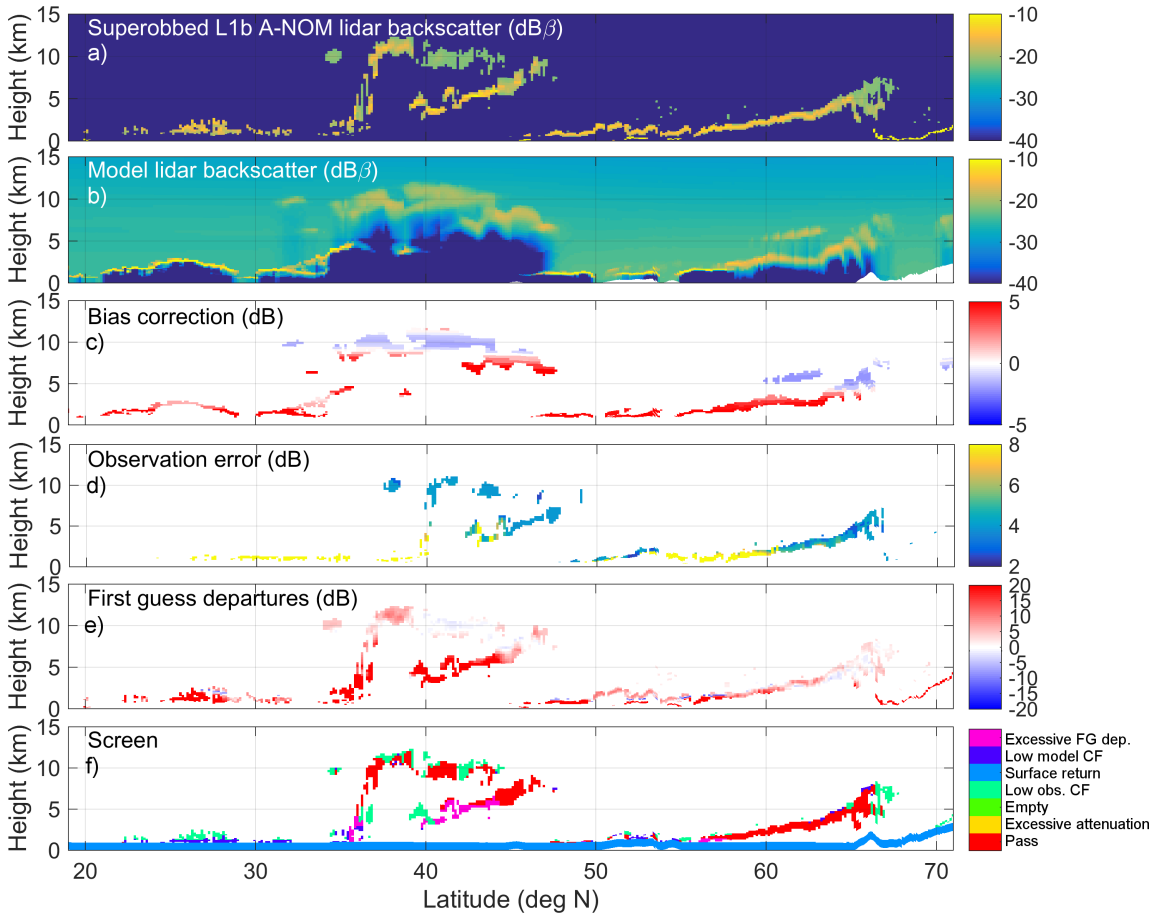

Figure 4.4: Example of observations stored in the ODB related to the EarthCARE L1b A-NOM test data. Panels show: (a) Superobbed lidar backscatter to native model resolution (TCo639; dBβ), (b) forward modelled lidar backscatter (dBβ), (c) bias correction applied to model equivalent (dB), (d) total observation error of each superob (dB), (e) observation minus background first guess departures (dB), (f) Screening value, blank areas have a cloud mask of zero in either observation or model equivalent.

5 Summary

In this report, the required modifications of the operational data assimilation system at ECMWF for EarthCARE cloud radar and lidar are documented. The work is separated to three sections. Firstly a summary of the required adaptations to account for the scientific differences between EarthCARE and CloudSat-CALIPSO data is given. While the CloudSat radar and CALIOP share many characteristics of the EarthCARE CPR and ATLID there are some differences that must be represented in the forward models. For the radars, the main difference is shown to be the sensitivity, with the CPR detecting significantly smaller hydrometeors (typically up to 30% smaller radii). For the lidar, the sensitivity in total attenuated backscatter is less for ATLID due to the increased molecular backscatter at 355nm. The effects of multiple scattering is shown to be similar in the two lidars due to compensating effects.

We then provide a summary of the technical changes needed to process the EarthCARE data into a format that can be ingested and used by the data assimilation system. Developments focus on the tools to convert the L1b data format into BUFR, which required the definition of new BUFR descriptors. Some changes to the data selection, pre-processing and screening were shown, particularly in relation to the cloud masks, which will not be provided in the L1b raw data. A short discussion on the likely latencies of the L1b is also made. In the final section a summary of tests demonstrating the technical capability of the system to assimilate EarthCARE data is shown and discussed. The tests make use of the A-NOM and C-NOM test data produced using a high-resolution model and the ECSIM.

Some fine-tuning of the developments will be necessary in the commissioning phase of the EarthCARE mission, but the framework for assimilating EarthCARE observations is now in place.

Acknowledgments

The authors would like to thank Aleksandra Tatarevic for her help in providing the L1b nominal test data. We would also like to thank Marijana Crepulja and Drasko Vasiljević for their help in writing the BUFR converter for EarthCARE data and adding new descriptors to the BUFR sequences.

List of Acronyms

4D-Var	Four-Dimensional Variational Assimilation
A-NOM	ATLID Nominal data
ATBD	Algorithm Theoretical Basis Document
ATLID	ATmospheric LIDar
BUFR	Binary Universal Form for the Representation of meteorological data
C-NOM	CPR Nominal data
C-PRO	Cloud profiling radar PROcessing
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CloudSat	NASA's cloud radar mission
CPR	Cloud Profiling Radar
EarthCARE	Earth, Clouds, Aerosols and Radiation Explorer
ECSIM	EarthCARE Instrument Simulator
ECMWF	European Centre for Medium Range Weather Forecasts
ESA	European Space Agency
ESTEC	European Space Research and Technology Centre
GEM	Global Environment Multi-scale Model
HSRL	High Spectral Resolution Lidar
IFS	Integrated Forecasting System of ECMWF
JAXA	Japan Aerospace eXploration Agency
NASA	National Aeronautics and Space Administration
NWP	Numerical Weather Prediction
ODB	Observation Data Base
POP	Particle Optical Properties
PVC	Photon Variance-Covariance
SDM	Significant Detection Mask
SNR	Signal-to-Noise Ratio
SR	Signal Ratio
TCo639	Model cubic octahedral grid with spectral truncation T639
TDTS	Time-Dependent Two-Stream
WP	Work Package

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.

Clothiaux, E. E., M. A. Miller, B. A. Albrecht, T. P. Ackerman, J. Verlinde, D. M. Babb, R. M. Peters, and W. J. Syrett, 1995: An evaluation of a 94-GHz radar for remote sensing of cloud properties, *J. Ocean. Atmos. Tech.*, **12**(2), 201–229.

Côté, J., S. Gravel, A. Mthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational cmcmrb global environmental multiscale (gem) model. part i: Design considerations and formulation, *Monthly Weather Review*, **126**(6), 1373–1395.

Di Michele, S., E. Martins, and M. Janisková, 2014b: Observation operator and observation processing for cloud lidar, WP-1200 report for the project Support-to-Science-Element STSE Study - EarthCARE Assimilation, AO/1-6441/10/NL/CB, 40 pp.

Fielding, M. and M. Janisková, 2017: 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.

Hildebrand, P. H. and R. S. Sekhon, 1974: Objective determination of the noise level in doppler spectra, *Journal of Applied Meteorology*, **13**(7), 808–811.

Hogan, R. J., 2008: Fast lidar and radar multiple-scattering models. part i: Small-angle scattering using the photon variancecovariance method, *Journal of the Atmospheric Sciences*, **65**(12), 3621–3635.

Hogan, R. J. and A. Battaglia, 2008: Fast lidar and radar multiple-scattering models. part ii: Wide-angle scattering using the time-dependent two-stream approximation, *Journal of the Atmospheric Sciences*, **65**(12), 3636–3651.

Hogan, R. J., N. Gaussiat, and A. J. Illingworth, 2005: Stratocumulus liquid water content from dual-wavelength radar, *Journal of Atmospheric and Oceanic Technology*, **22**(8), 1207–1218.

Illingworth, A. et al., 2015: The earthcare satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation, *Bulletin of the American Meteorological Society*, **96**(8), 1311–1332.

Janisková, M., M. Fielding, M. Crepulja, D. Vasiljević, T. Král, and P. Lean, 2017: Assimilation system development for cloud radar and lidar observations, WP-3000 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, 26 pp.

Kollias, P., A. Battaglia, A. Tatarevic, W. Szrymer, and D. Burns, 2016: Doppler radar and synergy products for EarthCARE Algorithm Theoretical Basis Document, Unpublished.

Liu, Z., M. A. Vaughan, D. M. Winker, C. A. Hostetler, L. R. Poole, D. Hlavka, W. Hart, and M. McGill, 2004: Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data, *Journal of Geophysical Research: Atmospheres*, **109**(D15), n/a–n/a, D15202.

Platt, C., 1973: Lidar and radioinetric observations of cirrus clouds, *Journal of the Atmospheric Sciences*, **30**(6), 1191–1204.

Reverdy, M., H. Chepfer, D. Donovan, V. Noel, G. Cesana, C. Hoareau, M. Chiriaco, and S. Bastin, 2015: An earthcare/atlid simulator to evaluate cloud description in climate models, *Journal of Geophysical Research: Atmospheres*, **120**(21), 11,090–11,113, 2015JD023919.

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.

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.