ESA Contract Report

PEARL Cloud contract 4000128669/19/NL/CT

Contract Report to the European Space Agency

Assimilation system development for additional EarthCARE observations: Doppler velocity, cloud extinction, Rayleigh backscatter

WP-4000 report

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

Authors: Mark Fielding and Marta Janisková

Contract officer: Tobias Wehr

July 2021

Series: ECMWF ESA Contract Report Series

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

Contact: library@ecmwf.int

ABSTRACT

This report details the developments to the assimilation system at ECMWF in preparation for the inclusion of EarthCARE observations other than cloud radar and lidar data, such as cloud extinction and Doppler velocity, directly into the Four-Dimensional Variational (4D-Var) system. The work is divided into three sections. Firstly, the scientific basis for the observation operators of cloud extinction, Rayleigh backscatter and radar Doppler velocity are outlined. Secondly, the extensive developments to the assimilation system are reported, including the definition of new variable names required for the conversion of the observations into BUFR and Observation Database (ODB) format are also described. Finally, the system is tested using pseudo-observations of EarthCARE cloud extinction, Rayleigh backscatter and radar Doppler velocity using retrievals and parameterizations of CloudSat and CALIPSO data.

Contents

1 Introduction

The Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE; Illingworth et al., 2015) multi-platform satellite, due for launch in 2024, will be the first of its kind, hosting a suite of instruments capable of simultaneously observing the vertical structure of clouds and aerosols. While observations of cloud radar reflectivity and cloud lidar backscatter have previously been made from space (from the CloudSat and CALIPSO satellites respectively), EarthCARE will also observe Doppler velocity from the radar and cloud extinction directly via a cloud and aerosol optimised high-spectral resolution lidar (HSRL). All of these novel observations will be based on the same platform.

In preparation for the operational monitoring and potential assimilation of EarthCARE observations, the ECMWF Four-Dimensional Variational (4D-Var) assimilation was recently adapted to ingest cloud radar and lidar observations. Significant technical developments were required to transform the raw observations received from the satellite into a format that is understood and useful. In particular, while the profiling nature of the observations will provide a wealth of information on the vertical structure of clouds to the numerical weather prediction (NWP) model used at ECMWF, such an observation type has not been assimilated previously. Having developed the capability to assimilate radar reflectivity and lidar backscatter observations has opened the possibility to consider monitoring or assimilating other observations from EarthCARE.

Expanding the assimilation system to include additional EarthCARE observations will have several benefits. Firstly, making the technical developments to the ECMWF 4D-Var system to include other observations such as Doppler velocity from the CPR, individual channels or appropriate L2 products from the HSRL and thermal-infrared radiances from the Multi-Spectral Imager (MSI) if successful will facilitate their monitoring in near real-time, which could provide invaluable warnings of instrument quality degradation to ESA. Secondly, additional EarthCARE observations can act as useful independent verification of the ECMWF analysis quality. Finally, retrieval experiments have shown the synergistic power of combining the information from the different EarthCARE sensors, so it is likely that further improvements in forecast skill could be obtained through assimilating additional observations, for example the profiling information provided by the active sensors can remove an ambiguity of cloud thickness that is present when assimilating thermal-infrared radiances.

The necessary technical developments for adding observations to the ECMWF system can be roughly separated into two sections. The first involves ‘off-line’ data handling where routines to convert the raw satellite data into the format recognised by the system need to be written. The second set of developments involve coding changes and additions to the 4D-Var data assimilation system itself. Updates and modifications will be required at all of these steps. Once both streams of developments are complete, all the modifications will subsequently need to be tested thoroughly.

As EarthCARE observations will not be available until after the satellite is launched, the initial data handling and testing will use CloudSat (NASA’s cloud radar mission; Stephens et al., 2002) and CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations; Winker et al., 2009). Although in a constellation rather than on-board the same satellite, the CloudSat radar and CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) lidar are sufficiently similar to the EarthCARE CPR (Cloud Profiling Radar) and ATLID (ATmospheric LIDar) to be used for testing.

This document outlines the technical developments required for the assimilation of EarthCARE-like observations of cloud radar and lidar. In Section 3, the technical pre-processing and handling developments are defined. Section 4 details the developments of the 4D-Var assimilation system. Section 5 gives examples of the testing of the system. A summary concludes the report in Section 6.

2 Observation operators for additional observations

To assimilate any kind of observation, an observation operator is required to convert the model variables into a so-called model-equivalent that can be compared to the observation. The observation operators for radar reflectivity and total attenuated lidar backscatter are described in Fielding and Janisková (2020a) and form the basis for the observation operators for the additional EarthCARE observations considered in this report.

2.1 Doppler velocity

In addition to vertical profiles of radar reflectivity, the EarthCARE CPR will also measure the Doppler velocity of clouds and precipitation within the radar beam. If we assume that the radar is perfectly nadir-pointing, then the mean Doppler velocity, V_D, measured by the radar is the sum of the vertical air motion, w, and the mean radar reflectivity weighted droplet terminal fall velocity, v_d:

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

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 WP2000; Fielding and Janisková, 2017).

Following the definition used within the IFS cloud physics scheme, the terminal velocity of a hydrometeor particle is defined by a power law:

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

where c_x and d_x are constants defined for each hydrometeor type in Table 2.1. The last term accounts for the decreased drag in less dense air where ρ_0 is a reference air density ($1\,\mathrm{kg\,m^{-3}}$).

Practically, for use in the observation operator, look-up tables of reflectivity-weighted fall speeds for different model hydrometeor types are computed off-line as a function of hydrometeor mass. Radar reflectivities as a function of particle size are themselves obtained from look-up tables (see Table 2.2) and combined with the particle size distributions given in Table 2.3. The resulting fall-speeds for a reference air density and temperature are shown in Fig. 2.1. Note that rain has fall-speeds that are an order of magnitude greater than snow; because of the inherently noisy characteristics of space-based Doppler radar measurements (e.g., Kollias et al., 2014) it is likely that most information in any observations will probably be gained from the Doppler velocity in rain. It is also worth pointing out that the expected reflectivity-weighted fall-speed is quite sensitive to the assumed rain drop sizes as can be seen by the differences in assumed fall-speeds between stratiform and convective rain.

Hydrometeor	c_x	d_x
Rain	386.8	0.67
Snow	16.8	0.527

Table 2.1: List of hydrometeor terminal velocity parameters used within the IFS.

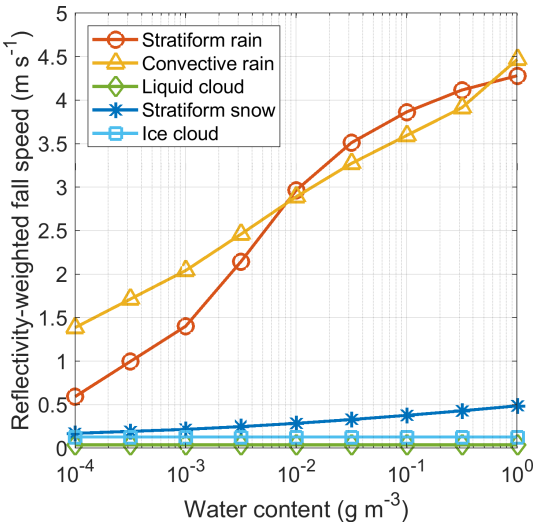

Figure 2.1: Reflectivity-weighted terminal velocity for different hydrometeor types as a function of water content using an air density of $1 kgm^{-3}$.

Within the observation operator, the contributions from different hydrometeor types to the total simulated Doppler velocity must be combined. This is straight-forward if we assume that all hydrometeors are distributed uniformly across the grid-box, which gives the grid-box mean Doppler velocity, $\overline{v_d}$, to be:

$$\overline{v_d} = \frac{\sum_{i=1}^{n} a_i Z_i v_{d,i}}{\sum_{i=1}^{n} a_i Z_i} \tag{2.4}$$

where n is the number of hydrometeor species, Z_i is the radar reflectivity for each hydrometeor species and a_i is the cloud fraction for each hydrometeor species.

The latest version of the observation operator for radar reflectivity includes options for accounting for sub-grid inhomogeneity and cloud overlap, which could be applied to the Doppler velocity observation operator, following the notation in Fielding and Janisková (2020a):

$$v^d_{triple} = (1-(C_1+C_2))v_{d,c} + C_1 v_{d,w1} + C_2 v_{d,w2}. \tag{2.5}$$

From this equation you can see that the Doppler velocity measured by a radar is not directly affected by attenuation (provided that the beam in each region is not fully attenuated; Kollias et al., 2018) Therefore, the benefit of using the triple-column approach for forward modelling Doppler velocity would be for representing the sub-grid inhomogeneity given the non-linear mapping between water content and fall-speed (see Fig. 2.1, noting the log x-axis). The representation of the effect of a fully attenuated beam within one of the columns is left for future work. Instead, the Doppler velocity is deemed unobservable if the total attenuated radar reflectivity falls below the radar sensitivity.

A demonstration of the radar Doppler velocity operator is provided in Fig. 2.2b for a satellite in a typical sun-synchronous orbit. The additional information Doppler velocity contains is apparent in the regions of heavier precipitation such as at 40 degrees North. If only radar reflectivity is considered, there is an ambiguity between attenuation and evaporation of the precipitation. By also considering the Doppler velocity, we can see that the reduction in radar reflectivity towards the surface is mainly caused by attenuation because the Doppler velocity remains constant towards the surface. In Fig. 2.2c,

Hydrometeor	Habit	Scattering model		Eff. density $[g/cm^3]$
		Radar	Lidar	$\rho_e(D) = aD^b$
Liquid cloud	Sphere	Mie	Mie	a = 1 b = 0
Ice cloud	6-bullet rosette	Liu (2008)	Yang et al. (2000)	a = 0.0094 b = -0.87
Stratiform Rain	Sphere	Mie	Mie	a = 1 b = 0
Convective Rain	Sphere	Mie	Mie	a = 1 b = 0
Snow	Aggregates	Hong et al. (2008)	Yang et al. (2000)	a = 0.0026 b = -1.42

Table 2.2: Definition of hydrometeor scattering assumptions used to create the bulk scattering properties for the radar and lidar observation operators. The effective density is the density that a particle would have if its volume was given by an encompassing sphere.

Hydrometeor	PSD	Parameters
Liquid cloud	$N(r) = \frac{N_t\sqrt{2\pi}(\ln\sigma_g)r}{e}^{-\frac{\ln^2(r/r_g)}{2(\ln\sigma_g)^2}}$ (Miles et al., 2000)	$N_t = 100\,\mathrm{cm}^{-3}$ $\sigma_g = 0.3$
Ice cloud	Field et al. (2007)) (fixed $T = -70\,^\circ$C)	a_d = 0.0094 b_d = -0.87
Stratiform Rain	$N(D) = N_0\exp(-\lambda D)$, $N_0 = x_1\lambda^{x_2}$ (Abel and Boutle, 2012)	$x_1 = 0.22$ $x_2 = 2.2$
Convective Rain	$N(D) = \frac{0.03N_L D_0^4\Lambda^{\mu+4}}{\Gamma(\mu+4)}D^\mu e^{-\Lambda D}$ (Illingworth and Blackman, 2002)	$N_L = 0.08\,\mathrm{cm}^{-4}$ $\mu = 5.0$
Snow	Field et al. (2007)	a_d = 0.0026 b_d = -1.42

Table 2.3: Definition of hydrometeor assumptions used to create the bulk scattering properties for the radar and lidar observation operators. For liquid cloud, N_t is the droplet number concentration, r_g is the median droplet radius and σ_g is the shape parameter of the assumed lognormal distribution. For stratiform rain, x_1 and x_2 are parameters in Abel and Boutle (2012), N_0 is the intercept parameter and λ is the shape parameter. For convective rain, N_L is the normalised drop concentration, D_0 is the median volume drop diameter and μ is the shape parameter in a normalised gamma size distribution.

an alternative version of the operator is shown with the ice fall speed fixed at $0.13\,\mathrm{m\,s}^{-1}$ and snow fall speed fixed at $1\,\mathrm{m\,s}^{-1}$ corresponding to the hydrometeor sedimentation rates in the IFS cloud physics in CY46R1. The relative performance of the two approaches will be tested when space-borne radar Doppler velocity measurements become available. Finally, Fig. 2.2d shows that the model vertical air motion is generally an order of magnitude less than the hydrometeor fall-speed, with the exception of near the top of some ice clouds where the hydrometeor fall speeds are low.

2.2 Lidar extinction

The observation operator for cloud extinction uses the same look-up table approach for obtaining cloud extinction from the model water content as the total attenuated backscatter forward model described in Fielding and Janisková (2020a). Figure 2.3 visualises the look-up table for cloud extinction as a function of water content, with the cloud optics assumptions listed in Table 2.3. Liquid cloud has the greatest extinction for a given water content due to the small droplet size and large droplet number concentration relative to the other hydrometeor habits. Conversely, convective precipitation has the smallest extinction for a given water content due to the relatively large drop size and small drop number concentration.

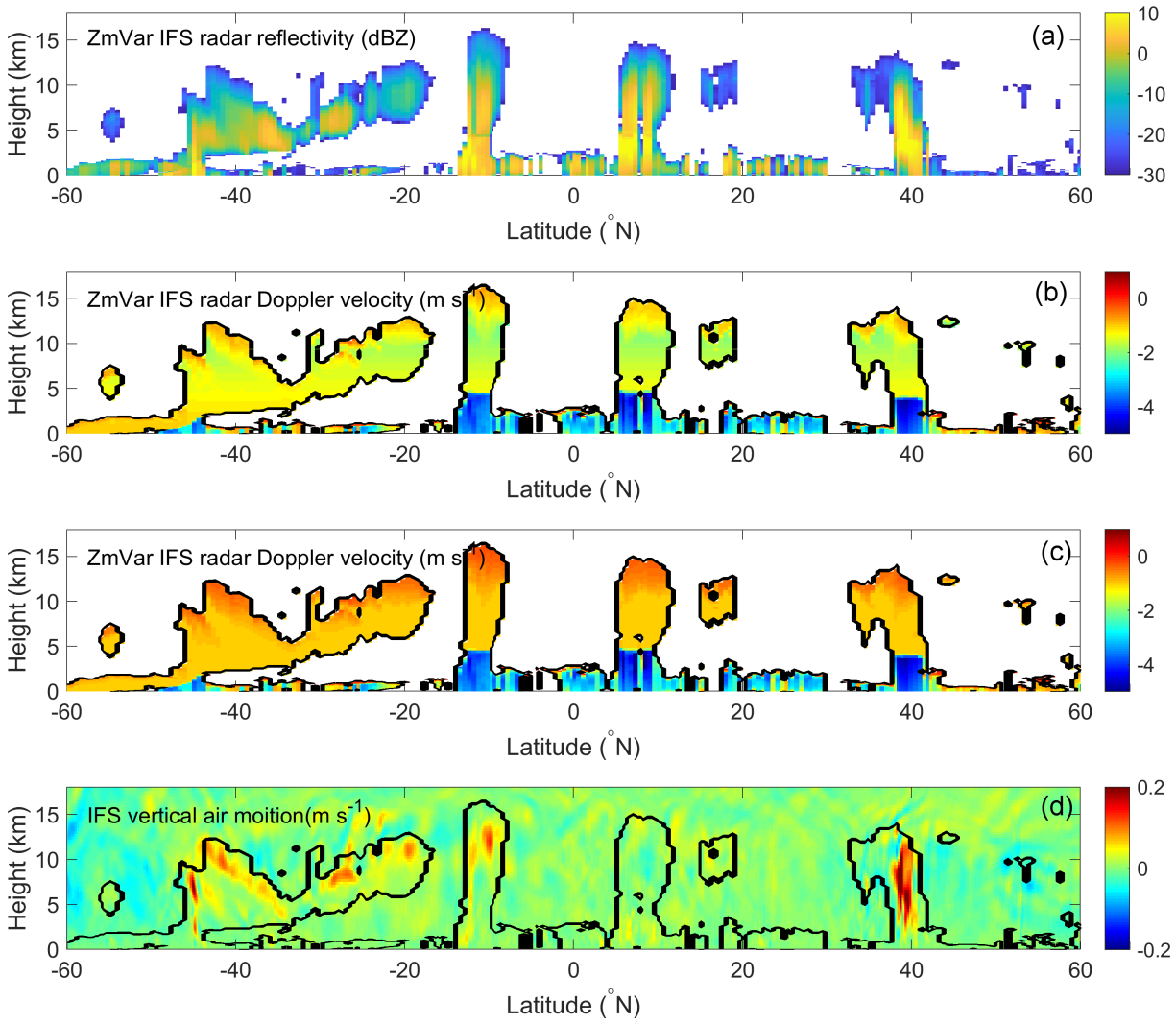

Figure 2.2: Transects of a) forward modelled radar reflectivity (dBZ), b) Doppler velocity using latest IFS cloud microphysics assumptions, c) Doppler velocity using CY46R1 cloud microphysics assumptions and d) simulated vertical air motion. The transects match a CloudSat orbit on 1 August 2007, where the input model fields come from a 12-hour model forecast with 16 km *horizontal grid spacing that has been horizontally averaged to a* 72 km *grid.*

To compute the grid-box mean cloud extinction, $\overline{\sigma_{\text{ext}}}$, the contribution to extinction from the different hydrometeor types is a simple sum:

$$\overline{\sigma_{\text{ext}}} = \sum_{i=1}^{n} a_i \sigma_{\text{ext}}. \tag{2.6}$$

Similar to Doppler velocity, the measurement of extinction is not directly affected by attenuation, however the lidar requires the total attenuated backscatter to at least be above the minimum detectable signal for the extinction of a given layer to be measured (see black lines in Fig. 2.4). In terms of sub-grid inhomogeneity, while the mapping of water content to extinction is quasi-linear (see Fig. 2.3) averaging observations of extinction where some profiles are fully attenuated could cause representativity issues.

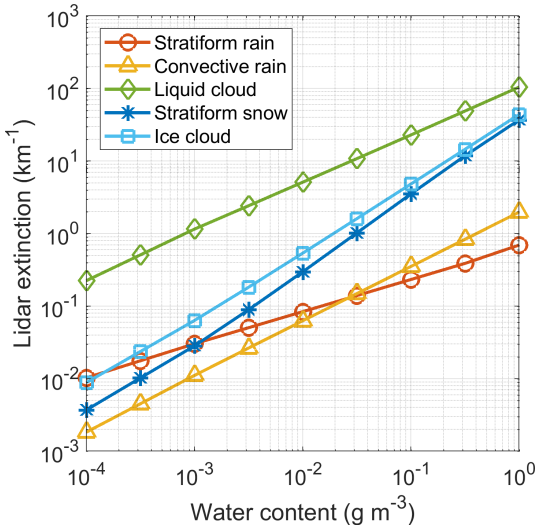

Figure 2.3: Same as Fig. 2.1, but for lidar extinction.

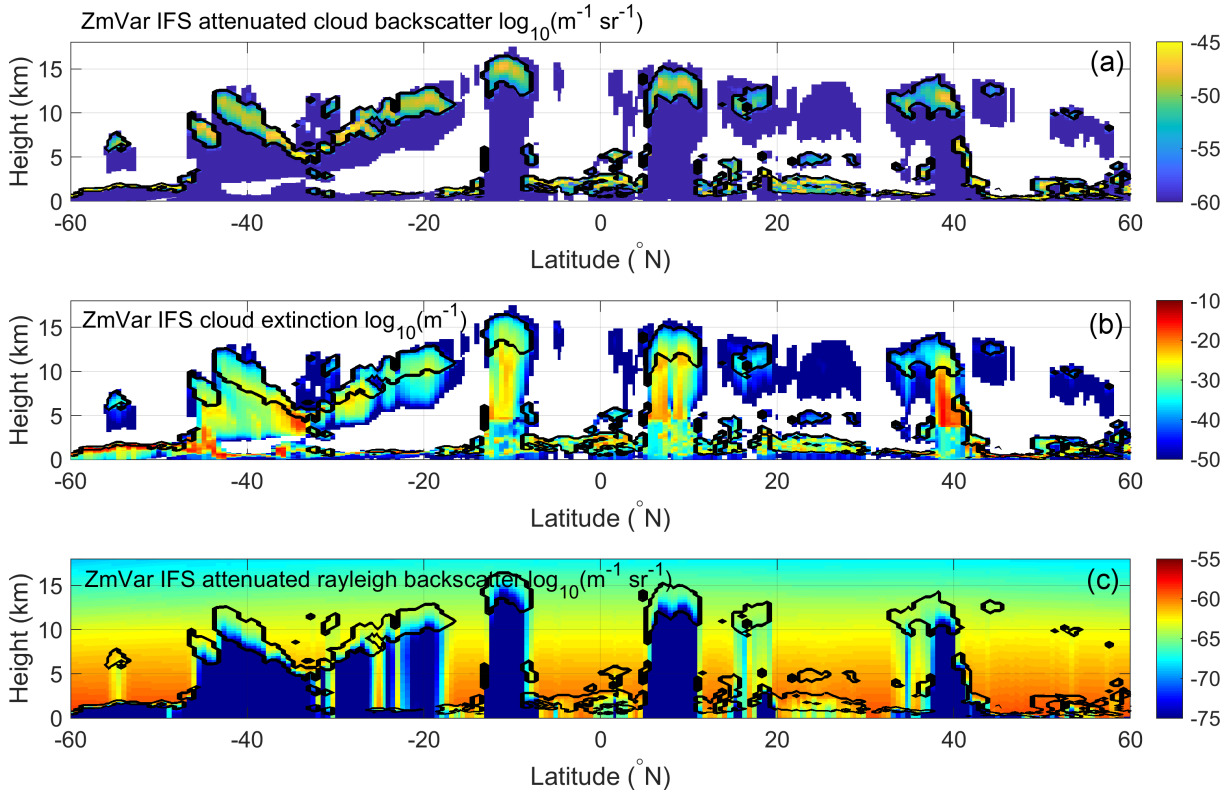

Figure 2.4: Same as Fig. 2.2, but for a) total attenuated backscatter at 532 nm, b) cloud extinction at 532 nm and c) Rayleigh attenuated backscatter at 532 nm.

As a demonstration of the observation operator, Fig. 2.4b shows the forward modelled cloud extinction corresponding to a CALIPSO orbit. The black lines show where the forward modelled attenuated backscatter is greater than the CALIPSO lidar sensitivity and hence show where we would expect extinction observations to be possible. The majority of observations are in ice cloud, where extinction is

lower and some observations of the tops of boundary-layer cloud.

2.3 Rayleigh backscatter

Observations of Rayleigh backscatter at visible wavelengths are sensitive to atmospheric pressure and temperature, while attenuated Rayleigh backscatter is also sensitive to the gaseous and particulate extinction integrated from the lidar to the level of the observation. The observation operator for Rayleigh backscatter is a modification of the observation operator for attenuated backscatter reported in Fielding and Janisková (2020a):

Molecular backscatter is straightforward to model using the following expression (Collis and Russell, 1976):

$$\beta_{\text{ray}} = 5.45 \cdot 10^{-32} \times \frac{p}{KT} \times \left(\frac{\lambda}{0.55}\right)^{-4.09} \tag{2.7}$$

where p is the atmospheric pressure, T is the temperature, K is the Boltzmann constant ($1.38 \cdot 10^{-23} J\,K^{-1}$) and λ is the lidar wavelength in microns.

$$\beta'_{\text{ray}}(r) = \beta_{\text{ray}}(r) \exp(-2\tau(r)) \frac{1 - \exp(-2\Delta h(r)\alpha(r))}{2\Delta h(r)\alpha(r)} \tag{2.8}$$

where

$$\tau(r) = \sum_{l=1}^{r-1} \exp(-2\Delta h(l)\alpha(l)) \tag{2.9}$$

is the optical depth of hydrometeors and gases between the instrument and a distance or range r and Δh is the depth of the model layer l. To approximate narrow-angle multiple scattering we apply the widely used 'Platt coefficient', η, originally proposed by Platt (1973), to the extinction due to clouds, α_{cloud}, while not adjusting the molecular extinction, α_{gas}:

$$\alpha = \eta\,\alpha_{\text{cloud}} + \alpha_{\text{gas}}. \tag{2.10}$$

Theoretically, the value of η is bounded between 1 (the single-scattering limit) and 1/2 (the wide field-of-view limit). The optimal value of η depends on the wavelength, scattering medium and the lidar geometry. For the CALIOP lidar at 532 nm we use $\eta = 0.55$ following tuning using the PVC method Hogan et al. (2006) as described in Fielding and Janisková (2020a).

A demonstration of the observation operator for attenuated Rayleigh backscatter is shown in Fig. 2.4c. The lidar signal undergoes the same extinction as the total cloud backscatter (Fig. 2.4a), so does not contain information in the middle and below thick clouds. The backscattered signal is generally much smoother than the cloud backscatter, so may be more conducive to 4D-Var assimilation, where strong non-linearities in an observation operator can lead to the minimization failing to converge.

3 Pre-processing and handling developments

For any observational data to be assimilated in the ECMWF operational system it must first be converted to Binary Universal Form (BUFR). A framework for converting radar and lidar data to BUFR was established in Fielding and Janisková (2017), but requires modifications to account for the additional observation types considered in this report. The conversion from the data's source format to BUFR is designed to be as close to a one-to-one mapping as possible and any differences in the context of the data should be limited to initial quality checks and reductions in precision to limit file sizes. In addition to developments related to BUFR, updates to the Observation Data Base (ODB) are also required. Whereas BUFR is optimised for the efficient storage of data, the role of the ODB is to provide fast I/O to the assimilation system on all observation related data. In this section, a brief overview of BUFR and ODB formats is provided before the new developments for additional observations are presented.

3.1 BUFR definitions

BUFR is a WMO standard for transmitting and storing observations of all kinds of meteorological data. Its flexibility lies with its use of 'data descriptors', which are used to access data values. Metadata is stored in external table files. Variables stored within BUFR must be selected from a finite pre-existing list of observation types. New variable types must be approved by WMO. The list of data descriptors for each observation type is known as a 'BUFR sequence'. Many of the descriptors are generic, such as the time and geolocation descriptors, however the more specific descriptors, such as 'lidar Rayleigh backscatter', are new variable types and will need to be approved by WMO for operational use as is needed for the radar and lidar descriptors defined in Fielding and Janisková (2017).

Each BUFR descriptor has a unique 6 digit code in the format FXXYYY, where F determines the type of descriptor, X determines the class of the descriptor and Y determines the name of the descriptor within its class. There are four types of descriptors: the most common are element descriptors (F=0), which are used to convey either meta-data or numerical data. The other types of descriptors are used to manipulate either data or the BUFR sequence itself. Replication descriptors (F=1) create loop-like structures in the BUFR template to allow descriptors to be repeated without the need for repeating the element descriptor explicitly. If Y=0, the replication descriptors are "delayed" and the number of replications needs to be provided by a subsequent elemental descriptor. Operator descriptors (F=2) perform actions on elemental descriptors, such as adding additional bits. Finally sequence descriptors (F=3) correspond to sequences of element descriptors that are often repeated. Sequence descriptors are not necessary, but can significantly reduce the size of the BUFR file if used wisely.

To allow for the additional observations to be stored in BUFR format, new descriptors have been created and added to the corresponding BUFR sequences. Table 3.1 shows the BUFR sequence created for spaceborne radar observations with new descriptors for Doppler velocity and Doppler velocity uncertainty included. In addition to their code, each descriptor has a *scale*, *reference* and *width* value that describes how the data is encoded to binary and vica versa. The *scale* determines the precision of the stored data, while the *width* and *reference* determine the range of the data. Specifically, the data is encoded by allocating a range of integers, **I**, from 1 to 2^{width} to data values, **R**, so that:

$$\mathbf{R} = (\mathbf{I} + reference)10^{-scale} \tag{3.1}$$

For Doppler velocity, the range of velocities stored is therefore between -20.00 and 20.47 with a precision of 0.01 $\mathrm{m\,s^{-1}}$.

F-X-Y	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					‘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 02 153	Satellite Channel Centre Frequency	-8	0	26	Hz	94 GHz
0 25 182	L1 processing flag					
0 25 181	L2 processing flag					
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	2	-20	11	m s^{-1}	
0 21 199	Doppler velocity uncertainty	2	-20	11	m s^{-1}	
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		

Table 3.1: Proposed BUFR sequence for cloud radar observations.

The corresponding BUFR sequence for space-borne HSRL observations is shown in Table 3.2. There are many similarities to the cloud radar BUFR sequence, with the main differences being the observation variables, such as lidar backscatter. For cloud extinction, we re-use the 'Extinction coefficient' descriptor originally included for aerosol observations. New descriptors include the Rayleigh backscatter and Mie co-polar backscatter.

3.2 ODB definitions

While BUFR is extremely efficient for storing data, a different format with fast I/O is required for operational data assimilation. At ECMWF, the ODB is 'in-house' data storage software to allow the 4D-Var system within IFS to store and access data. ODB is formulated on fast and efficient Structured Query Language (SQL) to define and retrieve observational data. There is no unique centralized ODB database: a new ODB is created each analysis time a 4D-Var analysis is made. Each ODB is stored locally in the ECFS (ECMWF's File Storage System) for post-processing and evaluation. Any new ODB variable name codes must be approved by the 'ODB Governance' to ensure consistency between ECMWF, member states and collaborators.

The ODB contains all the input data that is needed by the data assimilation system. Most data is stored in either a header (hdr) entry or a body entry (see left-hand side of Fig. 3.1). Header data typically consists of geolocation and instrument meta-data, while body data are observations or their meta-data. All body entries are linked to a header entry. For radar and lidar observations, the hdr contains the lat, lon and time the observations were taken. The observation value to be assimilated is stored within the 'obsvalue' variable within the body, alongside information needed for assimilation such as the observation error, screening status and bias correction. To identify the observation, each observation type has a unique variable number, 'varno', which tells the 4D-Var system which observation operator must be used to produce the model-equivalent.

Figure 3.1: Schematic of hdr and body storage within the ODB.

As each additional observation considered in this report will have a corresponding radar reflectivity or lidar backscatter value, it makes sense for the additional observations to share the same hdr information and simply extend the number of body rows to accommodate the additional observations (see right-hand side of Fig. 3.1). By way of an example, extending the number of body rows is the same procedure that would be taken for adding additional channels for a passive satellite-based remote sensing instrument. Because the observations are averaged to model levels (currently 137), the total number of body rows

Code	Description	Scale	Ref.	Width	Units	Comment
0 01 007	Satellite identifier	0	0	10		satID=787
0 02 019	Satellite instruments	0	0	11		instrumentID=303
0 02 153	Satellite Channel wavelength	9	0	16	m	
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 21 197	Height	0	-1000	17	m	
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 $m^{-1}\,sr^{-1}$
0 15 067	Extinction coefficient	9	0	30	m^{-1}	Range: 0 to 0.1 m^{-1}
0 15 068	Uncertainty in extinction coefficient	9	0	30	m^{-1}	Range: 0 to 0.1 m^{-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 21 202	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 21 194	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		

Table 3.2: Proposed BUFR sequence for space-borne HSRL observations.

Varname	Parent	Dimension	Description
obsvalue	body	model_levels (vertco_type 5)	Radar reflectivity averaged to model grid and level
biascorr	body	model_levels (vertco_type=5)	offline bias correction to be applied to radar reflectivity
datum_status	body	model_levels (vertco_type=5)	flag for storing screening/blacklisting/quality control information
clreflvalue	body	model_levels (vertco_type=5)	Model equivalent radar reflectivity
clreflvaluetl	body	model_levels (vertco_type=5)	Tangent linear variable for radar reflectivity
clreflvaluead	body	model_levels (vertco_type=5)	Adjoint variable for radar reflectivity
report_clreflflag	body	model_levels (vertco_type=5)	Flag for storing quality control information related to model equilavent
standard_deviation	superobs	model_levels (vertco_type=5)	Standard deviation of radar reflectivity within model grid and level
n_obs	superobs	model_levels (vertco_type=5)	number of samples used to compute obsvalue and standard_deviation
cloud_fraction	superobs	model_levels (vertco_type=5)	number of cloudy points defined by cloud mask divided by number of samples
repres_error	errstat	model_levels (vertco_type=5)	Flow dependent representativity error
obs_error	errstat	model_levels (vertco_type=5)	Measurement error
final_obs_error	errstat	model_levels (vertco_type=5)	Combination of meaurement, representativity and forward model error
surface_pressure	modsurf	scalar	background surface pressure from model
lat	hdr	scalar	Average latitude of observations
lon	hdr	scalar	Average longitude of observations
stalt	hdr	scalar	Height of satellite above sea level

Table 3.3: Selected ODB variable definitions for CloudSat observations.

per data instance is 137 multiplied by the total number of different observations to be monitored or assimilated.

The total observation error is stored in 'final_obs_error', but a breakdown of the error into representativity error, measurement error and forward model error is made possible by the inclusion of 'obs_error' and 'repres_error', which will be useful for the analysis of experiments. Also useful for analysis are the 'standard_deviation','n_obs' and 'cloud_fraction' variables. Note that much of the ODB variable definitions framework for CALIPSO observations (Table 3.4) is identical to that for CloudSat (Table 3.3), the interpretation of some variables differs, such as the 'datum_status' variable, which will contain different flags.

Varname	Parent	Dimension	Description
obsvalue	body	model_levels (vertco_type=5)	Lidar backscatter averaged to model grid and level
biascorr	body	model_levels (vertco_type=5)	offline bias correction to be applied to lidar backscatter
datum_status	body	model_levels (vertco_type=5)	flag for storing screening/blacklisting/quality control information
clbscvalue	body	model_levels (vertco_type=5)	Model equivalent lidar backscatter
clbscvaluetl	body	model_levels (vertco_type=5)	Tangent linear variable for lidar backscatter
clbscvaluead	body	model_levels (vertco_type=5)	Adjoint variable for lidar backscatter
report_clbscflag	body	model_levels (vertco_type=5)	Flag for storing quality control information related to model equilavent
standard_deviation	superobs	model_levels (vertco_type=5)	Standard deviation of lidar backscatter within model grid and level
n_obs	superobs	model_levels (vertco_type=5)	number of samples used to compute obsvalue and standard_deviation
cloud_fraction	superobs	model_levels (vertco_type=5)	number of cloudy points defined by cloud mask divided by number of samples
repres_error	errstat	model_levels (vertco_type=5)	Flow dependent representativity error
obs_error	errstat	model_levels (vertco_type=5)	Measurement error
final_obs_error	errstat	model_levels (vertco_type=5)	Combination of meaurement, representativity and forward model error
surface_pressure	modsurf	scalar	background surface pressure from model
lat	hdr	scalar	Average latitude of observations
lon	hdr	scalar	Average longitude of observations
stalt	hdr	scalar	Height of satellite above sea level

Table 3.4: Selected ODB variable definitions for CALIPSO observations.

3.3 Screening criteria

Preventing observations that may degrade the analysis from being assimilated is an important component of the observation pre-processing system. Known as 'screening', the selection of which observations to enter the minimization is achieved by checking each observation against a set of screening flags. The screening flags are represented as individual bits in an integer variable that is stored in the odb as the 'datum_status' and archived for diagnostic purposes.

Table 3.5 shows the various screening flags and associated bitfields for both radar reflectivity and Doppler velocity observations. The first 9 flags are generic for all observation types, for example if the observation is missing or if the observation is out of expected bounds. The last three are specific to cloud radar observations: the 'low_CF' flag is used when either the model or superob cloud fraction is below a threshold, the 'mscat' flag is used when multiple-scattering is suspected using the criteria in Fielding and Janisková (2017) and the 'FG_low' flag is used when the forward modelled value is below the sensitivity of the instrument. The Doppler velocity sensitivity is obtained by checking the corresponding radar reflectivity value minimum sensitivity.

Bitfield	Key	Description
	Radar reflectivity (CLREF)	
0	NRAD_CLREF_ACTIVE	Observation active if no other bits set
1	NRAD_CLREF_OBS	Observed clref out of bounds
2	NRAD_CLREF_NEGATIVE_Q	Negative Q in model profile
3	NRAD_CLREF_NO_OBS	No obs for this grid point (GP-space only)
4	NRAD_CLREF_NO_GRID	No grid point for this obs
5	NRAD_CLREF_PASSIVE	Passive observation
6	NRAD_CLREF_MISSING	Observation value is missing (=RMDI)
7	NRAD_CLREF_LAST_TIMESTEP	Last timestep,when TL/AD gp model doesn't run
8	NRAD_CLREF_FG_DEPARTURE	FG departure outside limit
9	NRAD_CLREF_LOW_CF	Low model or obs cloud fraction
10	NRAD_CLREF_MSCAT	Multiple-scattering
11	NRAD_CLREF_FG_LOW	Low FG value (below sensitivity)
	Radar Doppler velocity (CLDOP)	
0	NRAD_CLDOP_ACTIVE	Observation active if no other bits set
1	NRAD_CLDOP_OBS	Observed cldop out of bounds
2	NRAD_CLDOP_NEGATIVE_Q	Negative Q in model profile
3	NRAD_CLDOP_NO_OBS	No obs for this grid point (GP-space only)
4	NRAD_CLDOP_NO_GRID	No grid point for this obs
5	NRAD_CLDOP_PASSIVE	Passive observation
6	NRAD_CLDOP_MISSING	Observation value is missing (=RMDI)
7	NRAD_CLDOP_LAST_TIMESTEP	Last timestep,when TL/AD gp model doesn't run
8	NRAD_CLDOP_FG_DEPARTURE	FG departure outside limit
9	NRAD_CLDOP_LOW_CF	Low model or obs cloud fraction
10	NRAD_CLDOP_MSCAT	Multiple-scattering
11	NRAD_CLDOP_FG_LOW	Low FG value (below sensitivity)

Table 3.5: Screening flags for cloud radar observations.

Bitfield	Key	Description
	Lidar backscatter (CLBSC)	
0	NLID_CLBSC_ACTIVE	Observation active if no other bits set
1	NLID_CLBSC_OBS	Observed CLBSC out of bounds
2	NLID_CLBSC_NEGATIVE_Q	Negative Q in model profile
3	NLID_CLBSC_NO_OBS	No obs for this grid point (GP-space only)
4	NLID_CLBSC_NO_GRID	No grid point for this obs
5	NLID_CLBSC_PASSIVE	Passive observation
6	NLID_CLBSC_MISSING	Observation value is missing (=RMDI)
7	NLID_CLBSC_LAST_TIMESTEP	Last timestep,when TL/AD gp model doesn't run
8	NLID_CLBSC_FG_DEPARTURE	FG departure outside limit
9	NLID_CLBSC_LOW_CF	Low model or obs cloud fraction
10	NLID_CLBSC_FG_LOW	Low FG value (below sensitivity)
11	NLID_CLBSC_ATT_HIGH	Suspected excessive attenuation
	Lidar cloud extinction (CLEXT)	
0	NLID_CLEXT_ACTIVE	Observation active if no other bits set
1	NLID_CLEXT_OBS	Observed CLEXT out of bounds
2	NLID_CLEXT_NEGATIVE_Q	Negative Q in model profile
3	NLID_CLEXT_NO_OBS	No obs for this grid point (GP-space only)
4	NLID_CLEXT_NO_GRID	No grid point for this obs
5	NLID_CLEXT_PASSIVE	Passive observation
6	NLID_CLEXT_MISSING	Observation value is missing (=RMDI)
7	NLID_CLEXT_LAST_TIMESTEP	Last timestep,when TL/AD gp model doesn't run
8	NLID_CLEXT_FG_DEPARTURE	FG departure outside limit
9	NLID_CLEXT_LOW_CF	Low model or obs cloud fraction
10	NLID_CLEXT_BSC_LOW	Corresponding backscatter below sensitivity
11	NLID_CLEXT_FG_LOW	Low FG value (below sensitivity)
	Lidar rayleigh backscatter (CLRBSC)	
0	NLID_CLRBSC_ACTIVE	Observation active if no other bits set
1	NLID_CLRBSC_OBS	Observed CLRBSC out of bounds
2	NLID_CLRBSC_NEGATIVE_Q	Negative Q in model profile
3	NLID_CLRBSC_NO_OBS	No obs for this grid point (GP-space only)
4	NLID_CLRBSC_NO_GRID	No grid point for this obs
5	NLID_CLRBSC_PASSIVE	Passive observation
6	NLID_CLRBSC_MISSING	Observation value is missing (=RMDI)
7	NLID_CLRBSC_LAST_TIMESTEP	Last timestep,when TL/AD gp model doesn't run
8	NLID_CLRBSC_FG_DEPARTURE	FG departure outside limit
9	NLID_CLRBSC_LOW_CF	Low model or obs cloud fraction
10	NLID_CLRBSC_OBS_NOCLOUD	Observation cloud-free
11	NLID_CLRBSC_MOD_NOCLOUD	Model cloud-free
12	NLID_CLRBSC_FG_LOW	Low FG value (below sensitivity)

Table 3.6: Screening flags for cloud lidar observations.

For the lidar screening flags (Table 3.6), the first 10 flags are the same as for the radar observations (although the criteria for applying them might be different). As for the radar Doppler velocity observations, the sensitivity flag for the cloud extinction observation relies on a threshold in total attenuated backscatter. For the Rayleigh backscatter there are two additional flags for indicating if there are either no clouds in the observations or no clouds in the model. These flags could be useful for preventing spurious increments in temperature or pressure from assimilating Rayleigh backscatter when the model and observations disagree on the presence of clouds.

4 Modification and development of IFS code

This section details the modifications of the IFS and the 4D-Var assimilation system needed to assimilate the additional EarthCARE observations. A brief overview of the overall system is given before the modifications to the ODB tasks, observation operator tasks, 4D-Var system tasks and finally the observation processing and archiving tasks.

For reference of the different tasks performed in running a 4D-Var experiment and the dependency between them, Fig. 4.1 shows a snapshot from ecFlowView (the in-house task scheduler used at ECMWF). An analysis experiment '*an*' contains six main tasks that are all performed sequentially:

- *make*: This task is run once per experiment and prepares the environment for the experiment, including compiling the source code. This task also fetches the initial data required to run the experiment, including background model forecast data and initial model background error data.

- *obs*: In this task, the observations are prepared for each assimilation cycle. The raw observations are fetched from their long-term storage (usually ECFS), before pre-processing tasks are performed such as converting to BUFR, superobbing and basic screening.

- *main*: The main task is the most complex task and is also repeated for every assimilation cycle. Firstly the blacklist is compiled in task 'black', which tells the 4D-Var system which observations should not enter the system, even for monitoring. Next, the 'odb' suite of tasks is run which converts all observations from their BUFR format to ODB in an array of 'b2o' tasks. The next tasks, 'vardata', 'fetcherr' and 'lowres' gather data required for the 4D-Var assimilation. The '4dvar' task performs the main 4D-Var assimilation, which consists of three 'outer' loops of the 'ifstraj', 'ifstsave' and 'ifsmin' tasks. The ifstraj performs the high-resolution model trajectory run, while 'ifsmin' performs the minimization. See Janisková et al. (2017) for more details.

- *lag*: Archiving of the model analysis and subsequent forecast is performed in this task. Also archived are selected data from the ODB, which is used for diagnostic purposes.

- *wsjobs*: This task archives logfiles and cleans any data that is not required for the next cycle.

- *flush*: The flush task cleans all data on the supercomputer once the experiment has completed.

4.1 ODB tasks

To accommodate the new observation types, relatively minor modifications to the ODB system were required. For the system to recognise the new observations, new variable numbers (known as 'varnos') needed to be assigned (see Table 4.1) as well as a corresponding 'NVAR'. Additional changes were

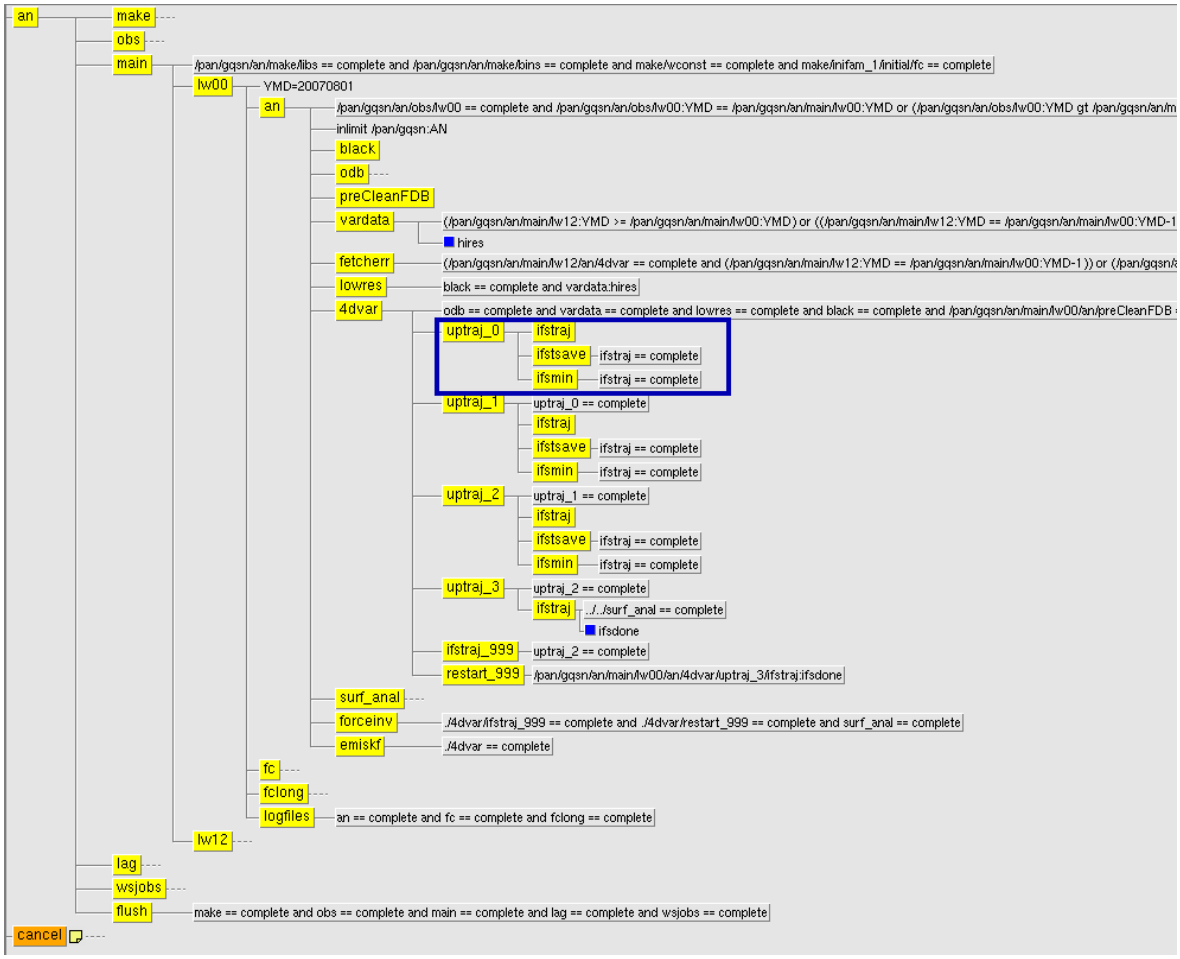

Figure 4.1: Visualising the main tasks in a 4D-Var experiment using ecFlowView.

required to the bufr2odb code and also to the sql requests used to communicate between the IFS and the ODB.

4.2 Modification of observation operator tasks

Nearly all routines in the observation operator needed to be modified in some way. Figure 4.2 shows a flowchart of the key routines within the observation operator. The 'hop' routine is a top-level routine that cycles through all observation data and calls the relevant observation operators depending on observation type. This routine was modified to call the correct observation operators when the new observations enter the system.

Once 'hop' has detected cloud radar or lidar observations, the 'obsop_radlid' routine is called. This routine prepares the interface for calling the wrapper routine for cloud radar and lidar observations. Modifications to this routine were made to include the additional observations in the sub-setting of observations according to varno. The wrapper routine itself – 'clradlid_wrapper' – provides the interface between hop and the radar and lidar observation operators. Physics variables, such as temperature and humidity at observation locations, are passed in as pointers. All other I/O for this operator (e.g., radar reflectivity/lidar backscatter, screening flags, observation error, bias correction) goes via the ODB, us-

Variable name	shortname	NVAR	varno
lidar_cloud_backscatter	CBSC	110	237
lidar_cloud_extinction	CEXT	112	238
‘ lidar_rayleigh_backscatter	RBSC	113	280
cloud_radar_reflectivity	CREF	111	239
cloud_doppler_velocity	CDOP	114	281

Table 4.1: List of variable names and associated numbers used within the ODB and IFS.

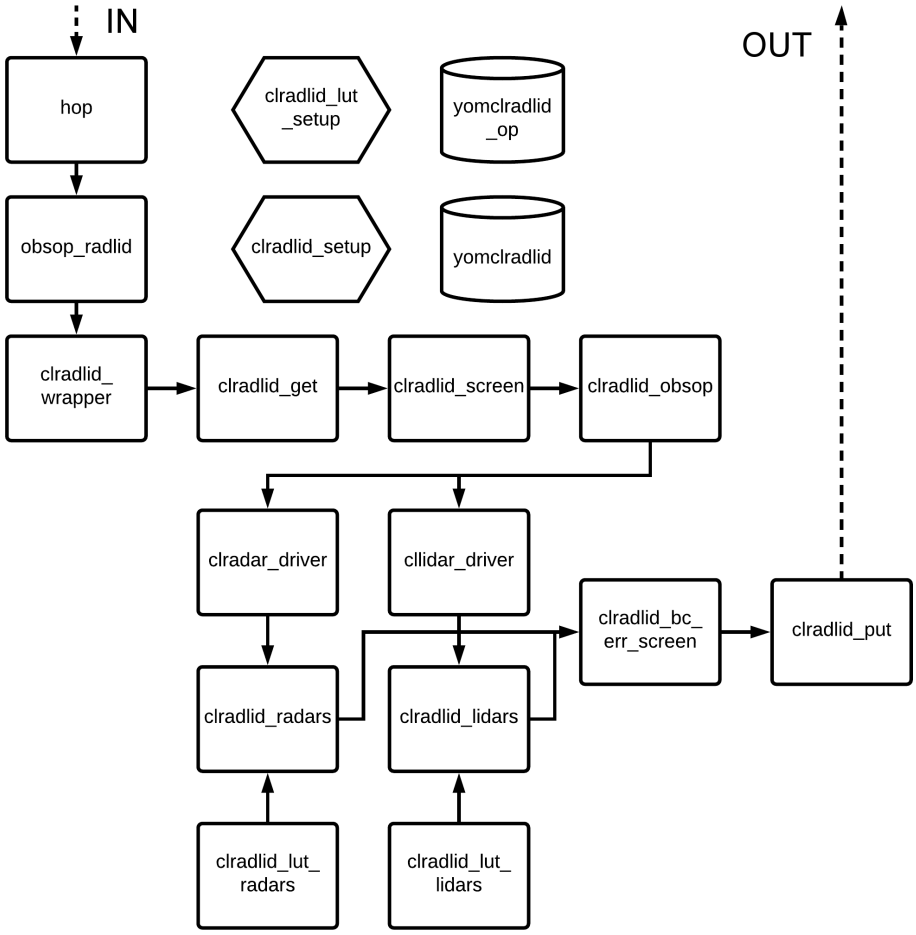

Figure 4.2: Flowchart for selected routines within the observation operator.

ing the ‘_get’ and ‘_put’ functions. The wrapper routine was modified to ensure the correct arrays sizes are allocated for observation profiles (the arrays must match the size of the ODB data structure, which depends on the number of additional observation types are to be considered).

The clradlid_wrapper routine first calls the ‘clradlid_get’ routine to fill local arrays of observation values and associated information from the arrays of data obtained directly from the ODB. The dimension of these local arrays was increased from 1 to 2 to allow for multiple observations at each level in each profile of observation. The local arrays are filled according to the observation varno. The clradlid_wrapper routine then calls ‘clradlid_screen’ that provides the first, model independent, screening of the observations. This routine was modified to include basic screening of the additional observations using the screening flags defined in Sec. 3.3.

After the initial screening is complete, the clradlid_wrapper routine then calls 'clradlid_obsop' which provides the final interface for computing radar and lidar observations. The routine checks the screening flags to see if there is a reason not to call the operators, then if all is OK, either the 'clradar_driver' or 'cllidar_driver' routine are called depending on the varno. Minor modifications to these routines were made to pass the additional observations as arguments through the routines.

The new observation operators defined in Sec. 2 were included in 'clradlid_radars' and 'clradlid_lidars'. In these routines, the model equivalents are computed by combining the optical and physical properties from different hydrometeor types according to the model cloud fraction and subgrid inhomogeneity whilst also accounting for attenuation. The radar look-up table routines were modified to include hydrometeor fall-speeds for the computation of radar Doppler velocity.

Before the model equivalents are stored in the ODB for use in the minimization, the clradlid_wrapper routine calls 'clradlid_bc_err_screen for a model-dependent screening of the observations. The bias correction and observation errors are also computed in this routine. Several modifications were made, including setting up place-holders for the specification of bias correction and observation errors for the additional observations. For testing, the bias correction is set to zero and observation errors set to 999.9. Finally, all the relevant data computed in the observation operator is stored in the 'clradlid_put' routine.

4.3 Modification of 4D-Var tasks outside observation operator

Some routines outside of the observation operator also needed to be modified for the additional observations to be recognised by the system. One such routine is 'defrun' which stores the parameters required for variational quality control (VarQC; Tavolato and Isaksen, 2015, which is applied to all observations assimilated in the ECMWF 4D-Var system). VarQC decreases the weight of spurious observations that have large departures from the model background. The VarQC weights are recalculated for each minimization, so an observation with a large departure could regain weight if the model state is brought closer to the observation in a previous outer-loop. For initial testing, VarQC is turned off for the additional observations.

5 Testing

In this section we give an overview of the testing for the technical developments outlined in the previous sections. Technical testing is required to ensure that the additional observations are recognized correctly and are fed through to the ODB properly. The observation equivalents generated by the observation operators passing to the ODB must also be verified. Once all the data stored in the ODB is known to be correct, a basic test of the 4D-Var minimization can be made; although the additional observations only enter into the system passively, it is important to check they do not have any unintended effect on the analysis. All testing is performed for a single cycle of the ECMWF 4D-Var system for the 00Z analysis of 1 August 2007.

5.1 Preparation and creation of observation test data

For testing the system, while physically plausible observations are required, they do not need to be scientifically accurate. When testing the radar reflectivity and lidar backscatter observation operators, CloudSat and CALIPSO are well-suited proxies for EarthCARE CPR radar reflectivity and ATLID total

backscatter respectively, however observations of particulate extinction and Rayleigh backscatter from an HSRL, and space-borne radar Doppler velocity are not readily available. Fortunately, at least for cloud extinction observations, we can use the CALIPSO retrieval products of cloud extinction that are contained within the CALIPSO L2 C-Pro data (Winker et al., 2009). To simulate observations of Rayleigh backscatter we use eq. 2.8, but instead of using the model cloud extinction, we use the same CALIPSO retrieved extinction product as for testing the assimilation of cloud extinction. This means the 'observed' Rayleigh backscatter above clouds should be identical to the FG model equivalent Rayleigh backscatter as they are generated using the same input.

Simulating radar Doppler velocity observations, V_{obs} (in units of $\text{m}\,\text{s}^{-1}$) , is achieved through a parameterization of CloudSat radar reflectivity, Z_{csat} (in units of dB):

$$V_{\text{obs}} = \begin{cases} 0.1 + 0.01(Z_{\text{csat}} + 30), & \text{if } T < 273K \\ 3 + 0.1(Z_{\text{csat}} + 10), & \text{otherwise} \end{cases} \tag{5.1}$$

where T is the model first guess temperature. The parameterization was conceived as a linear fit of the Doppler velocity (similar to Fig. 2.1) as a function of radar reflectivity using the look-up table for Doppler velocity.

5.2 Single cycle tests

As an initial test that the observations have successfully entered Figure 5.1 shows an example of the observation cost function that is printed to the output file documentating the model integration of the 'ifstraj' task. All five cloud radar and lidar related observations are represented by their 'varno' alongside the number of instances of the observations (DataCount) and the total contribution to the cost function (Jo_Costfunction). The magnitude of the cost function is given by the sum of the square of the first guess departures divided by the observation errors, so where the observation errors have been set to be very large (for example Rayleigh backscatter and Doppler velocity) the cost function contribution is small. The Rayleigh backscatter observations have the greatest number of observations because a first-guess departure is recorded in both clear and cloudy conditions.

All information relating to observations needed by the assimilation system is stored in the ODB, so its correct functioning is crucial and must also be tested. Figure 5.2 provides a visual check on the contents of the cloud extinction entries in the ODB. The observed cloud extinction stored in 'obsvalue' is plotted in Fig. 5.2a, while the model cloud extinction before and after the 4D-Var minimization is shown in Fig. 5.2b and Fig. 5.2c. Also shown is the analysis increments of cloud extinction, which shows where cloud has been added or removed by the assimilation. Recall that the cloud extinction observations entered the system passively, so the analysis increments in cloud extinction shown here are due to other observations. Despite the cloud extinction not being assimilated actively, the analysis does tend to get closer to the observations, for example the model analysis for extinction within the ice cloud in the tropics in Fig. 5.2 is closer to the CALIPSO cloud extinction observations.

Next, the processing of data through the ODB and 4D-Var system related to Rayleigh backscatter observations is demonstrated in Fig. 5.3. The 'observations' of Rayleigh backscatter generated in part by the cloud extinction shown in Fig. 5.2a, are shown in Fig. 5.3a. The corresponding model-equivalent Rayleigh backscatter before and after the analysis is shown in Fig. 5.3b and Fig. 5.3c respectively. As expected, the clear-sky regions that are above clouds are the very similar. Interestingly, the largest analysis increment (Fig. 5.3d) of Rayleigh backscatter in cloud-free conditions in the stratosphere is around 0.1 dB (not shown), which shows it would be very difficult to extract useful information from the lidar

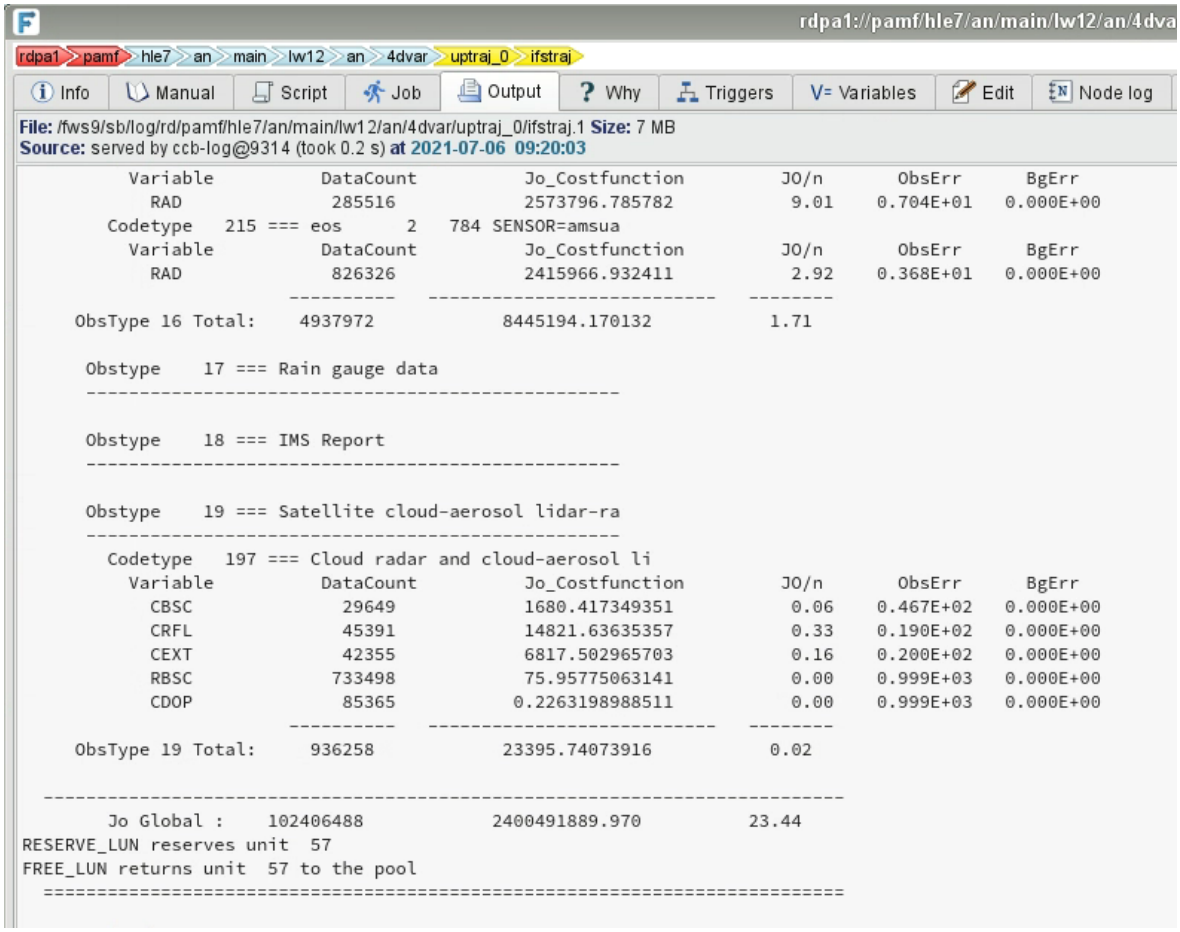

Figure 5.1: Screenshot of the output from ifstraj viewed with ecFlowview showing the cost function containing the additional radar and lidar observations.

backscatter in these regions unless the observation error was not significantly greater than this.

Finally, Figure 5.4 shows the equivalent testing of the ODB interface for the radar Doppler velocity test dataset. Again, Fig. 5.4a shows the 'observations' that were generated from the parameterization of CloudSat radar (eq. 5.1), while the next two panels show the model equivalent before and after the analysis. Note that the 'observations' of Doppler velocity in ice cloud are less than those simulated by the model because they do not take into account the reduced drag at lower pressures. However, it should be emphasised that the observations are not required to be scientifically precise for this testing. The size of the analysis increments (5.4d) provide a useful clue to how large the magnitude of the observation errors can be whilst still providing significant information to the analysis; only assimilating other observations does induce increments larger than the expected errors of the EarthCARE CPR Doppler velocity measurements. In rain, analysis increments sometimes exceed $1\,\mathrm{m\,s^{-1}}$, however for ice cloud they are rarely greater than $0.1\,\mathrm{m\,s^{-1}}$, which would generally be lower than the measurement noise before other errors are taken into account, such as representativity issues. Therefore it is likely to only be feasible to gain information via data assimilation of the Doppler velocity in rain.

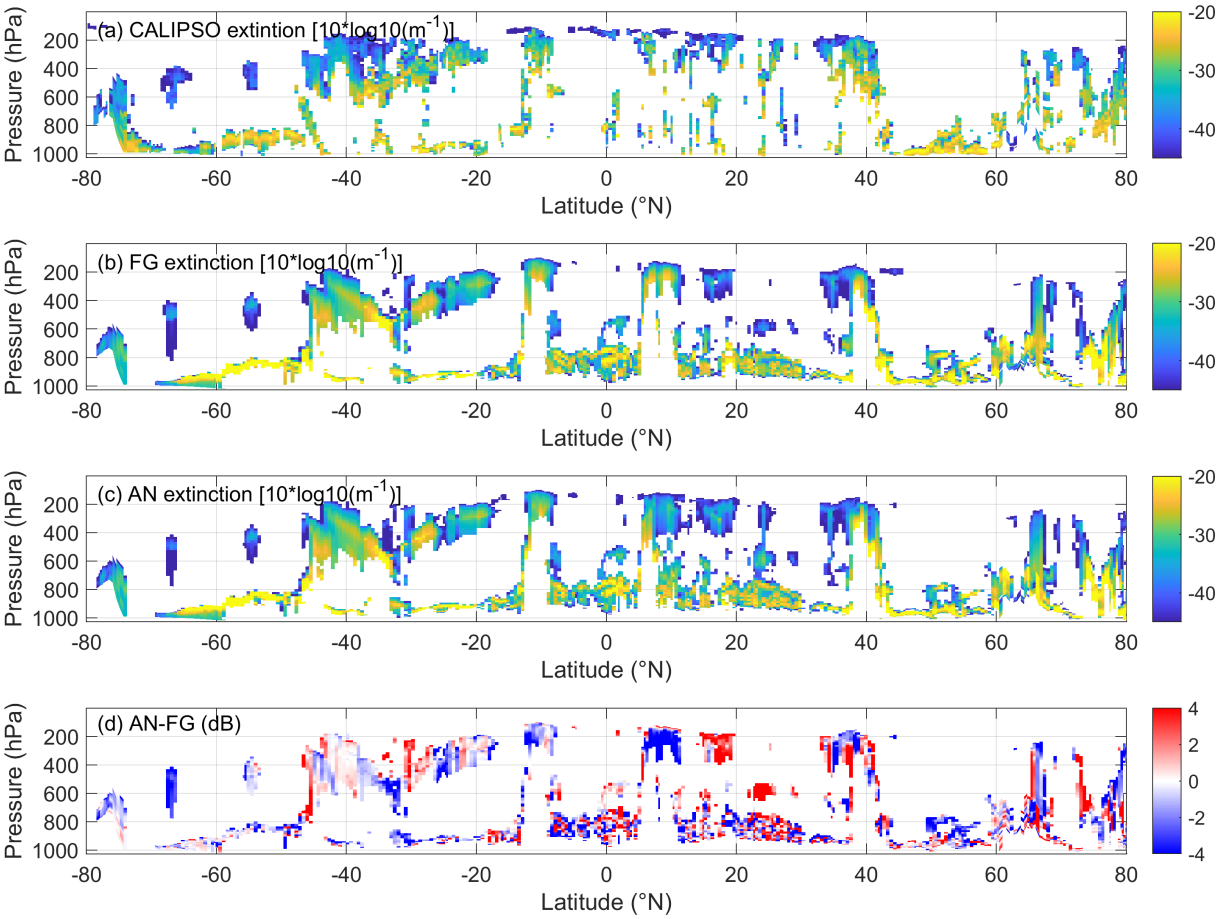

Figure 5.2: Example ODB output for (a) observations of cloud extinction, (b) first guess (FG) model equivalent cloud extinction, (c) analysis (AN) model equivalent cloud extinction and (d) the difference between analysis and first guess (AN-FG), for a CALIPSO transect during 1 August 2007.

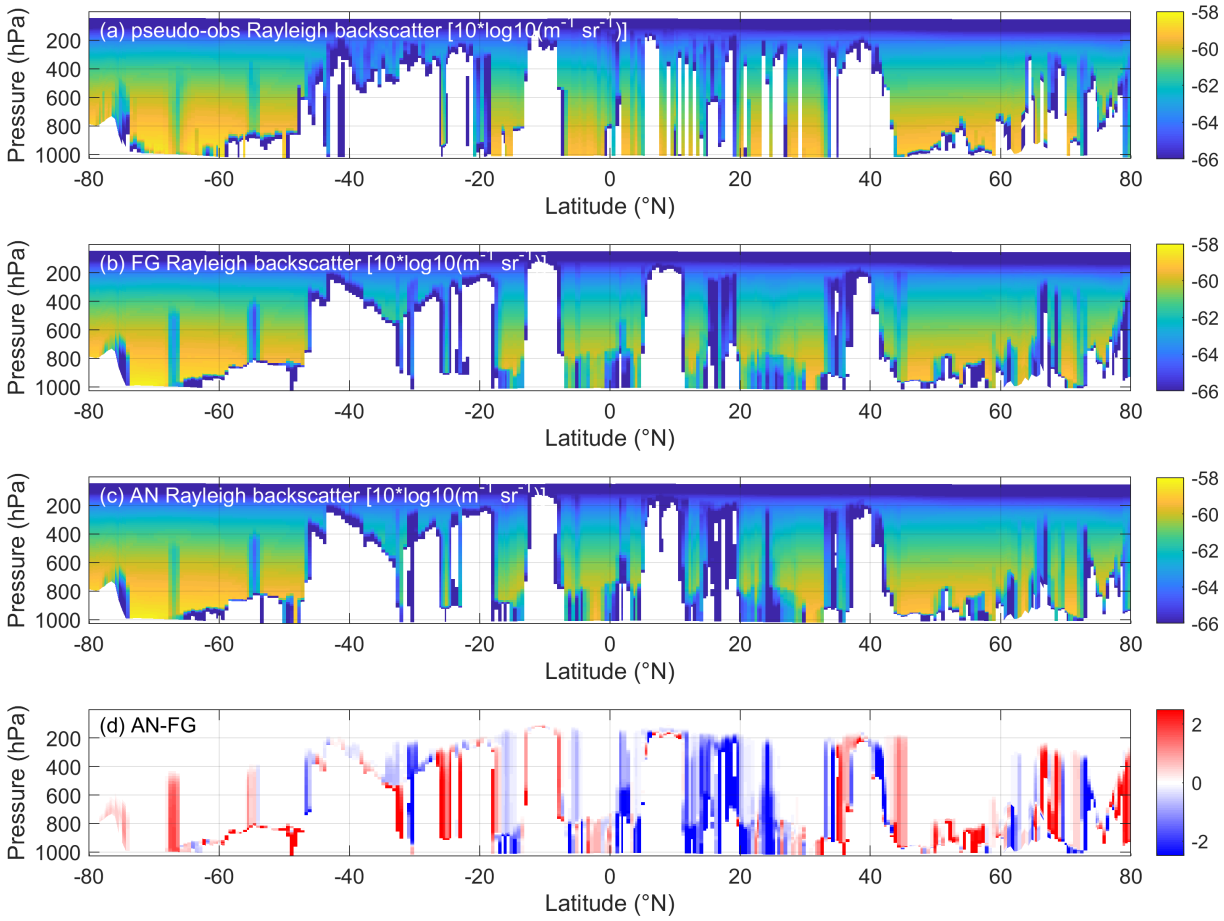

Figure 5.3: Same as Fig. 5.2, but for observations of Rayleigh backscatter at 532 nm.

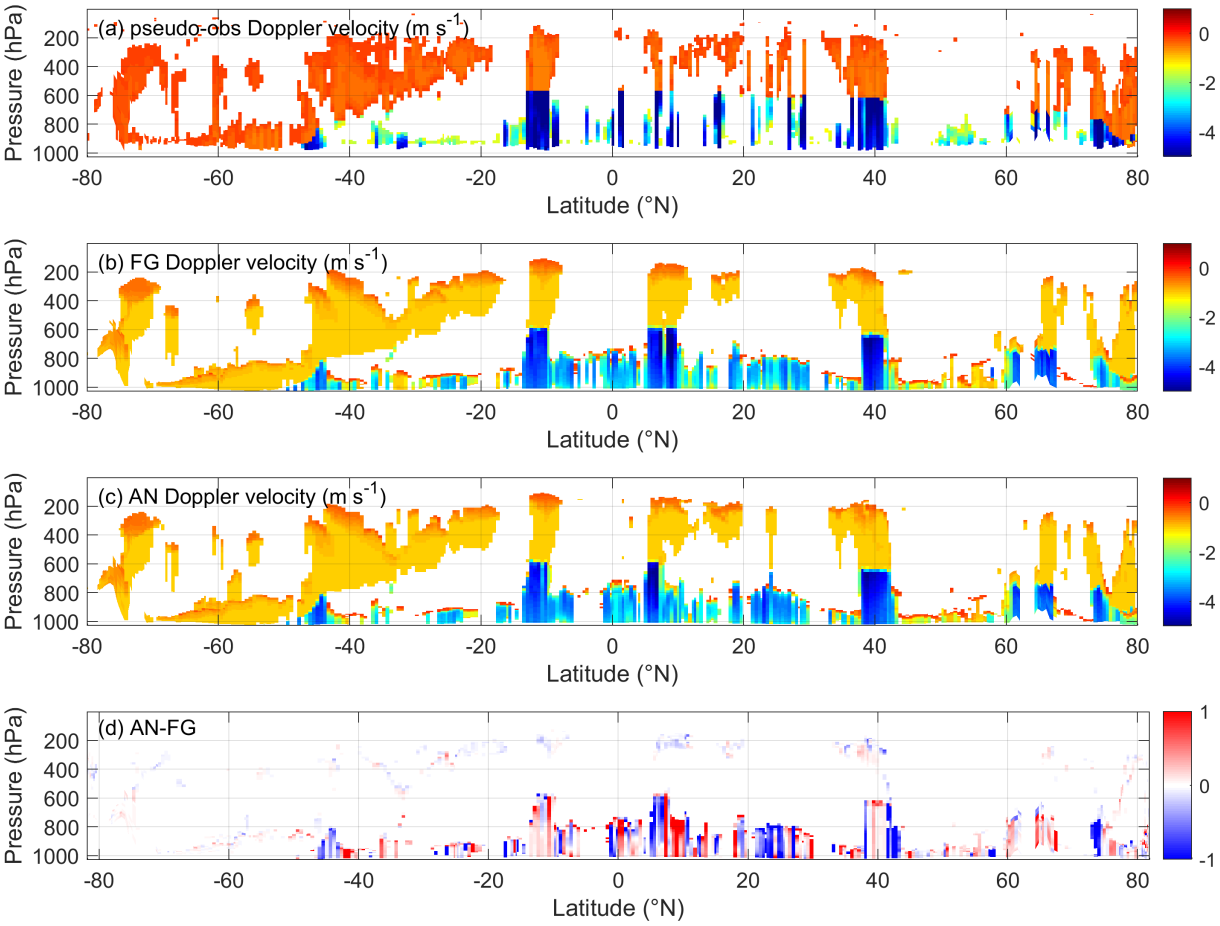

Figure 5.4: Same as Fig. 5.2, but for observations of radar Doppler velocity corresponding to a CloudSat transect.

6 Summary

In this report, the extensive developments to the operational data assimilation at ECMWF in preparation for the monitoring of EarthCARE-like observations other than cloud radar and lidar observations have been documented. Specifically, observation operators for radar Doppler velocity, cloud extinction and Rayleigh backscatter have been included into the IFS, so that observations of these types can be monitored against the ECMWF model. The report contains three sections. Firstly, the scientific basis for the observation operators was outlined. The second part of the report details all the coding developments that have been made to the assimilation system, the inclusion of the observations operators and the interface to the ODB. The off-line observation handling of the new observation types into BUFR format are explained, along with the necessary ODB developments. Finally the report concludes with an initial testing of the system using pseudo-observations derived from transects of CloudSat and CALIPSO data. All the above developments allow for the start of feasibility studies for the monitoring of cloud extinction, Rayleigh backscatter and radar Doppler velocity. For the new observation types to be assimilated in direct 4D-Var assimilation experiments, the tangent linear and adjoint codes for the observation operators need to be coded, which will be done in WP-5000 of this project. This will allow the potential benefit and synergy between the observations to be investigated, although the use of authentic measurements of Rayleigh backscatter and radar Doppler velocity from space will have to wait until the launch of EarthCARE.

Acknowledgments

The authors would like to thank Philippe Lopez and Alan Geer for their help in deciphering existing IFS code. Thanks also to Marijana Crepulja and Roberto Ribas for their help in preparing the preprocessing routines related to BUFR. Finally, the authors would like to thank Robin Hogan for invaluable discussions.

List of Acronyms

4D-Var	Four-Dimensional Variational Assimilation
AD	ADjoint
AN	Analysis
ATLID	ATmospheric LIDar
BUFR	Binary Universal Form for the Representation of meteorological 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
ECFS	ECMWF's File Storage system
ecFlow	ECMWF's work-flow manager enabling to run large number of programs
ecFlowview	graphical user interface to display the status of experiment tasks
ECMA	Extended CMA (used for all observations before screening)
ECMWF	European Centre for Medium Range Weather Forecasts
ESA	European Space Agency
FG	First Guess
HSRL	High-Spectral Resolution Lidar
IFS	Integrated Forecasting System of ECMWF
MSI	Multi-Spectral Imager
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NL	Non-Linear
NWP	Numerical Weather Prediction
OBS	OBServations
ODB	Observation Data Base
PVC	Photon Variance-Covariance method
SQL	Structured Query Language
TL	Tangent Linear
VarQC	Variational Quality Correction

References

Abel, S. J. and I. A. Boutle, 2012: An improved representation of the raindrop size distribution for single-moment microphysics schemes, *Quarterly Journal of the Royal Meteorological Society*, **138**(669), 2151–2162.

Collis, R. and P. Russell, 1976: Lidar measurement of particles and gases by elastic backscattering and differential absorption, *Laser monitoring of the atmosphere*, pp. 71–151.

Field, P., A. Heymsfield, and A. Bansemer, 2007: Snow size distribution parametrization for midlatituide and tropical ice cloud,, *J. Atmos. Sci.*, **64(12)**, 4346–4365.

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. D. 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, 4000116891/16/NL/LvH, pp.

Hogan, R. J., M. P. Mittermaier, and A. J. Illingworth, 2006: The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model, *Journal of Applied Meteorology and Climatology*, **45**(2), 301–317.

Hong, G., P. Yang, B. A. Baum, and A. J. Heymsfield, 2008: Relationship between ice water content and equivalent radar reflectivity for clouds consisting of nonspherical ice particles, *J. Geophys. Res.-Atmospheres*, **113**(D20), D20205.

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.

Illingworth, A. J. and T. M. Blackman, 2002: The need to represent raindrop size spectra as normalized gamma distributions for the interpretation of polarization radar observations, *Journal of Applied Meteorology*, **41**(3), 286–297.

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. Bataglia, A. Tatarevic, K. Lamar, F. Tridon, and L. Pfitzenmaier, 2018: The EarthCARE cloud profiling radar (CPR) doppler measurements in deep convection: challenges, post-processing and science applications, *Remote sensing of the atmosphere, clouds and precipitation*, **VII 10776**, 57–68.

Kollias, P., S. Tanelli, A. Battaglia, and A. Tatarevic, 2014: Evaluation of EarthCARE Cloud Profiling Radar Doppler Velocity Measurements in Particle Sedimentation Regimes, *Journal of Atmospheric and Oceanic Technology*, **31**, 366–386.

Liu, G., 2008: A database of microwave single-scattering properties for nonspherical ice particles, *Bulletin of the American Meteorological Society*, **89**(10), 1563–1570.

Miles, N. L., J. Verlinde, and E. E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds, *Journal of the Atmospheric Sciences*, **57**(2), 295–311.

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

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.

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.

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., K. N. Liou, K. Wyser, and D. Mitchell, 2000: Parameterization of the scattering and absorption properties of individual ice crystals, *Journal of Geophysical Research: Atmospheres*, **105**(D4), 4699–4718.