ESA CONTRACT REPORT

Contract Report to the European Space Agency

*Support-to-Science-Element (STSE) Study
EarthCARE Assimilation*
WP-1100 report: Observation operator and observation processing for cloud radar

August 2012, updated January 2014

*Authors: S. Di Michele, E. Martins,
          M. Janisková*

**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/publications/

Contact: library@ecmwf.int

Contract Report to the European Space Agency

Support-to-Science-Element (STSE) Study EarthCARE Assimilation

WP-1100 report: Observation operator and observation processing for cloud radar

Authors: S. Di Michele, E. Martins,
M. Janisková

ESA ESTEC contract 4000102816/11/NL/CT

ABSTRACT

This report summarizes the work done to prepare for the inclusion of the EarthCARE radar data in the ECMWF monitoring/assimilation system using CloudSat observations as testing data. The first part of the study describes the refinements brought to the radar forward operator. Improvements to the definition of the optical properties of frozen precipitation has led to a better agreement between simulated and CloudSat reflectvities. The impact of multiple scattering on the radar signal has been also investigated. Simulations show that this phenomenon affects all precipitation regimes, degrading the signal in conditions of heavy precipitation. The second part of the study focuses on the characterization of the observation errors, needed for data assimilation. Uncertainties due to the observation operator have been estimated perturbating some of the microphysical assumptions there made. In case of space-borne radar observations as the ones from CloudSat/EarthCARE, an important component of the observation error is the representativity one. In this work a flow-dependent estimate has been implemented following a previously-developed statistical approach. The final part of the study describes the quality control and bias correction schemes developed in order to have departures between CloudSat observations and model first guess suitable for assimilation.

Contents

1   Introduction

The Earth, Clouds, Aerosols and RadiationExplorer (EarthCARE) mission has the basic objective of improving the understanding of cloud-aerosol-radiation interactions by simultaneously measuring the vertical structure and the horizontal distribution of cloud and aerosol fields together with the outgoing radiation over all climate zones. The profiling of clouds will be provided by the combination of a lidar and a cloud radar (ESA, 2004). With the launch in 2006 of the CloudSat (Stephens et al., 2002) and CALIPSO (Winker et al., 2009) missions (as part of the A-Train constellation) observations of clouds from almost coincident spaceborne lidar and radar are already available. CloudSat carries the Cloud Profiling Radar (CPR), a 94-GHz nadir-looking radar that measures the power backscattered by clouds with a sensitivity threshold of -26 dBZ (Tanelli et al., 2008). Onboard CALIPSO, CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) is a nadir-viewing, three-channel lidar system (1064 nm and 532 nm parallel and perpendicular) with a 1 m receiving telescope (Hunt et al., 2009). A number of studies, including the ESA funded project Quantitative Assessment of the Operational Value of Space-Borne Radar and Lidar Measurements of Cloud and Aerosol Profiles (QuARL, Janisková et al., 2010), have shown that such observations are useful to evaluate the performance of current Numerical Weather Prediction (NWP) models and General Circulation Models (GCM) in representing clouds, precipitation and aerosols. The QuARL work has also shown that cloud radar and aerosol lidar observations can be correctly assimilated and that the information extracted from these data can improve the initial atmospheric state.

The objective of STSE Study EarthCARE assimilation is the development of an off-line system to monitor/assimilate Level-1 data from the CloudSat radar and the CALIPSO lidar in clouds within the ECMWF GCM model. The work will pave the way for the development of a pre-operational system for the exploitation of radar and lidar observations in data assimilation in the time frame of the EarthCARE mission.

This report summarizes the work performed within WP-1100 of the STSE-EarthCARE study. The aim of the work package was the development of the various components needed to handle CloudSat observations for their monitoring/assimilation, i.e.: a forward operator, the definition of observation errors as well as quality control and bias correction schemes. Section 2 describes the improvements to the CloudSat forward operator developed during the previous QuARL project. Emphasis has been put on the modelling of the radar multiple scattering. The definition of observation errors is described in Section 3, where forward modelling and representativity errors have been explicitly estimated. Finally, Section 4 describes the schemes developed for quality control and bias correction of CloudSat reflectivity.

2   Radar observation operator

The radar forward operator used in this study is ZmVar (reflectivity model for variational assimilation). It was originally developed to perform feasibility studies on the assimilation of ground-based (35 GHz) radar observations and the precipitation radar (13.5 GHz) on board the Tropical Rainfall Measuring Mission satellite (TRMM) (Benedetti and Janisková, 2004; Janisková, 2004; Benedetti et al., 2005; Lopez et al., 2006). In the context of the ESA funded project QuARL (Janisková et al., 2010), ZmVar was extended to simulate observations from the CloudSat 94 GHz nadir-looking radar (Di Michele et al., 2012). ZmVar is designed to meet the requirements of the assimilation system, i.e., to allow the coding of its adjoint counterpart and to be computationally efficient. It also has a certain degree of flexibility (e.g., for the definition of the optical properties).

2.1   ZmVar radar reflectivity forward operator

The way how ZmVar computes reflectivities is schematically shown in Fig. 2.1. Inputs to ZmVar are vertical profiles of the model cloud and thermodynamic variables coincident with the observations (Step 1). For computational efficiency, ZmVar uses a pre-calculated table of hydrometeor optical properties (extinction and

backscattering coefficients) (Step 2). This look-up table prescribes equivalent radar reflectivity factor (Z^h, called reflectivity hereafter) and volumetric extinction (α^h) at a given wavelength (λ) of predefined hydrometeor types (h). Values for Z^h and α^h for hydrometeor type h are generated off-line, integrating the backscattering cross section $\sigma_{bsc}^h(D)$ and the extinction cross section $\sigma_{ext}^h(D)$, over all particle sizes with diameter D, from an assumed particle size distribution (PSD), $N^h(D)$ (see Section 2.3 for PSD details):

$$Z^h = \frac{\lambda^4}{\pi^5 |K_w|^2} \int_{D_{min}^h}^{D_{max}^h} \sigma_{bsc}^h(D) N^h(D) dD \tag{2.1}$$

$$\alpha^h = \int_{D_{min}^h}^{D_{max}^h} \sigma_{ext}^h(D) N^h(D) dD \tag{2.2}$$

where D_{min}^h and D_{max}^h are pre-defined values of limits for the particle size and K_w is the dielectric factor of liquid water at $10°C$ (and 94 GHz). To ensure the same calibration format of CloudSat, ZmVar uses a value of 0.75 for $|K_w|^2$.

Z^h and α^h are saved in the look-up table for a range of temperatures and hydrometeor contents (w) using the following discretization:

- 70 values of temperatures, in the [234-303] Kelvin range for liquid particles and [204-273] K for frozen particles.

- 401 values of hydrometeor content, logarithmically scaled in the range [0.0001-1] g/m^3.

A value of reflectivity Z^h due to each single hydrometeor type h is evaluated by means of a bilinear interpolation on the values of temperature and hydrometeor content of the input profile (Step 3 in the schematic diagram of Fig. 2.1). Radar reflectivities of all hydrometeor types are then combined:

$$Z = \sum_{h=1,N_{HYD}} Z^h \tag{2.3}$$

where N_{HYD} is the number of hydrometeor types. Currently six hydrometeor types are modelled within ZmVar, consistent with the categories in the ECMWF forecast model: cloud liquid, cloud ice, stratiform rain, convective rain, stratiform snow, and convective snow.

The attenuation of the signal along the radar path due to both absorption by hydrometeors α^h and atmospheric gases α^{gas} must be taken into account as attenuation can be significant at 94 GHz (Step 4), therefore:

$$\alpha = \alpha^{gas} + \sum_{h=1,N_{HYD}} \alpha^h \tag{2.4}$$

In ZmVar, α^{gas} is evaluated using the model of Liebe (1985) and Liebe et al. (1992). For the $i-th$ layer, attenuated reflectivity Z_l^a are calculated (Step 5) as:

$$Z_l^a = Z_l \, e^{-2\bar{\tau}_{l-1}} \frac{1 - e^{-2\Delta h_l \alpha_l}}{2\Delta h_l \alpha_l} \tag{2.5}$$

where Δh_l is the depth of the $l-th$ layer. Note that levels are numbered from the top of the atmosphere, so $l=1$ indicates the top layer. The quantity $\bar{\tau}_l$ is the optical depth (for gases and clouds) between the layer
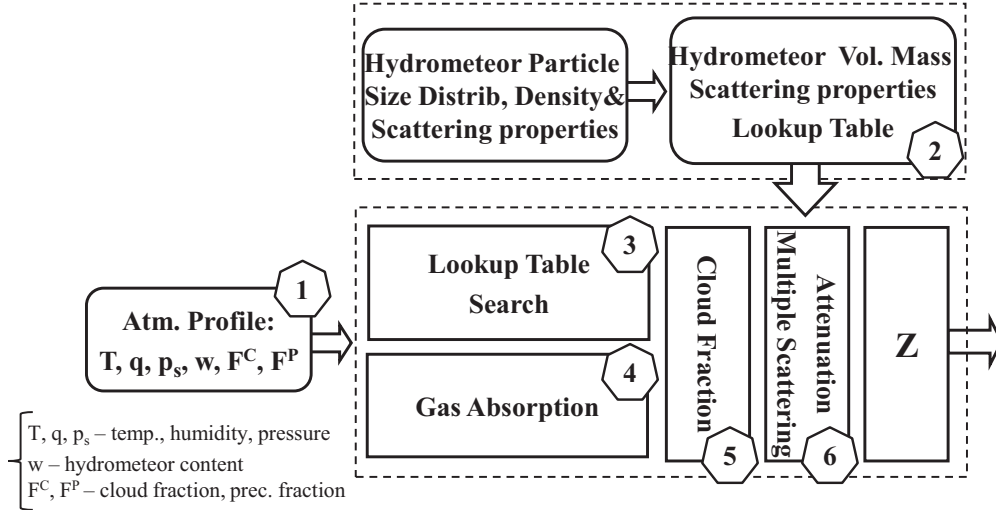
Figure 2.1: Schematic diagram of ZmVar operator. The numbers indicate different steps in operator computations.

l and the top of the atmosphere. The cloud contribution to $\bar{\tau}_l$ is obtained taking into account the profiles of cloud and precipitation fraction prescribed for the atmospheric state (Step 6, see Section 2.4.2 for details). It is worth noting that, the ZmVar forward model closely resembles the approach used elsewhere, including the QuickBeam radar forward operator (Haynes et al., 2007), which is part of the CFMIP (Cloud Feedback Model Intercomparison Project) Observation Simulator Package (COSP, Bodas-Salcedo et al., 2011). Indicated in Fig. 2.1 as part of Step 6, ZmVar has been updated to include the modelling of radar multiple scattering. The phenomenon can be important at the CloudSat frequency for the most intense clouds (e.g. Battaglia et al., 2010) and will be discussed in Section 2.5. In this work the ZmVar lookup table has been extended to include the volumetric single scattering albedo (ω^h) and the volumetric asymmetry factor (g^h) for each hydrometeor h which are needed for the evaluation of the multiple scattering. Similarly to Eqs. (2.1) and (2.2), they are evaluated integrating the corresponding single particle scattering cross section $\sigma_{sca}^h(D)$ and asymmetry parameter $g_1^h(D)$, over the PSD $N^h(D)$:

$$\omega^h = \frac{\displaystyle\int_{D_{min}^h}^{D_{max}^h} \sigma_{\text{sca}}^h(D)N^h(D)dD}{\displaystyle\int_{D_{min}^h}^{D_{max}^h} \sigma_{\text{ext}}^h(D)N^h(D)dD} \tag{2.6}$$

$$g^h = \frac{\displaystyle\int_{D_{min}^h}^{D_{max}^h} g_1^h(D)\sigma_{\text{sca}}^h(D)N^h(D)dD}{\displaystyle\int_{D_{min}^h}^{D_{max}^h} \sigma_{\text{sca}}^h(D)N^h(D)dD} \tag{2.7}$$

2.2 Treatment of cloud and precipitation fraction

ZmVar is designed to simulate radar reflectivity from an atmospheric model with horizontal resolutions which are so coarse that they to require some representation of sub-grid cloud and precipitation variability. Even when the model provides explicit information on sub-grid cloud and precipitation fractions, the overlap in the vertical

within a grid column must be taken into account. In this section, two alternative treatments are described as implemented in ZmVar to deal with the cloud/precipitation fraction of the Integrated Forecasting System (IFS) of ECMWF. The first one consists of a single-column treatment that was part of the (original) version of ZmVar used in previous studies (e.g. Benedetti et al., 2005). The second more sophisticated approach follows the multi-column concept recently implemented in ZmVar. Both are described here in order to highlight the benefit of the multi-column method used in this study compared to the previous single-column approach.

2.2.1 Single-column approach

In the single-column approach, sub-grid cloud and precipitation fraction are taken into account by the following steps:

1. For each layer l, the IFS specifies the cloud fraction F_l^C only for cloud liquid and cloud ice. ZmVar assigns a precipitation fraction F_l^P to rain and snow starting from F_l^C, making the assumption of random overlap for cloud and maximum overlap for precipitating hydrometeors.

2. The in-cloud value (denoted by ') of hydrometeor content w'^h_l in layer l for hydrometeor type h, is evaluated from the box-average values given by the model and scaled according to the corresponding sub-grid fraction:

$$w'^h_l = \frac{w^h_l}{F^h_l} \qquad (2.8)$$

   where F^h_l indicates F_l^C for cloud liquid and cloud ice and F_l^P for rain and snow.

3. The in-cloud reflectivity Z'^h_l and extinction α'^h_l are then evaluated using w'^h_l, as described in Section 2.1. Grid box averages Z_l and α_l are obtained scaling the in-cloud values according to the cloud and precipitation fractions and summing on the total number of hydrometeors N_{HYD}.

$$Z_l = \sum_{h=1,N_{\text{HYD}}} F^h_l \, Z^h_l \qquad (2.9)$$

$$\alpha_l = \sum_{h=1,N_{\text{HYD}}} F^h_l \, \alpha^h_l \qquad (2.10)$$

   Note that the values of α^h_l are very similar to the ones corresponding to the (unscaled) w^h_l since, for all hydrometeors, extinction coefficient has a (quasi-) linear dependence on the hydrometeor content. Note also that reflectivities are expressed in linear units (i.e. $\text{mm}^6 \, \text{m}^{-3}$).

4. Finally, the reflectivity measured by the radar Z^a is evaluated taking into account the attenuation along the path as in Eq. 2.5. In the single column approach, the optical depth between the $l - th$ layer and the top of the atmosphere $\bar{\tau}_l$ is evaluated as:

$$\bar{\tau}_l = \sum_{i=1,l} \sum_{h=1,N_{\text{HYD}}} F^h_i \, \alpha^h_i \, \Delta h_i \qquad (2.11)$$

   where Δh_l is the depth of the $l - th$ layer. With this methodology, no information is needed about the way clouds vertically overlap within the box. The main advantage of this approach is its computational efficiency which is an important factor for data assimilation. However, this is achieved at the expenses of accuracy since it neglects the (strong) non-linearity in the relationship between radar reflectivity and cloud/precipitation amount.

2.2.2 Multi-column approach

An alternative method that avoids the inaccuracy of the previous treatment of cloud/precipitation fraction is the multi-column approach as used in the Subgrid Cloud Overlap Profile Sampler (SCOPS), part of the International Satellite Cloud Climatology Project (ISCCP) simulator (Webb et al., 2001). This method uses a pseudo-random sampling process to generate an ensemble of subgrid cloud profiles representing the distribution within the model grid box. It takes the input vertical profiles of cloud fraction to generate a specified number of horizontally homogeneous cloud profiles. The SCOPS algorithm splits each grid column into a number of sub-columns, N_{COL}, in which each layer is either completely filled or completely free of cloud. The cloud cover in each vertical layer and sub-column is therefore either 0 or 1. The SCOPS version used here assumes maximum-random overlap for cloud (Geleyn and Hollingsworth, 1979; Tian and Curry, 1989) but maximum overlap for precipitation. The precipitation overlap scheme is discussed in more detail in Section 2.4.2. For the multi-column case, the attenuated reflectivity for layer l averaged across all sub-columns Z_l^a can be written as:

$$
\begin{aligned}
Z_l^a &= \frac{1}{N_{\text{COL}}} \sum_{k=1,N_{\text{COL}}} Z_{l,k}^a \\
&= \frac{1}{N_{\text{COL}}} \sum_{k=1,N_{\text{COL}}} \left(\sum_{h=1,N_{\text{HYD}}} F'^h_{l,k} Z'^h_{l,k} \right) e^{-2\tau'_{l-1,k}} \frac{1 - e^{-2\Delta h_l \alpha'_{l,k}}}{2\Delta h_l \alpha'_{l,k}}
\end{aligned}
\tag{2.12}
$$

where:

$$
\tau'_{l,k} = \sum_{i=1,l} \sum_{h=1,N_{\text{HYD}}} F'^h_{i,k} \alpha'^h_{i,k} \Delta h_i
$$

$$
\alpha'_{l,k} = \sum_{h=1,N_{\text{HYD}}} F'^h_{l,k} \alpha'^h_{l,k}
$$

and $F'^h_{l,k}$ (being $F'^C_{l,k}$ for cloud or $F'^P_{l,k}$ for precipitation) is either 0 or 1 in the l-th layer of the k-th sub-column. Note that, apart from numerical issues, the single-column and multi-column approximations produce the same grid box average un-attenuated reflectivity. Differences are only found for the attenuated reflectivity Z^a due to the treatment of optical depth, with the multi-column case providing a more realistic representation of overlap for the reflectivity profile. The ideal number of columns N_{COL} for the multi-column approach needs to be chosen to be sufficiently accurate but not too computationally expensive. The error of the IFS cloud fraction forecasts is certainly not expected to be smaller than 10%, hence for a reference case we assume that a hundred subcolumns (1%) are sufficient to reproduce a grid points cloud fraction profile with adequate precision, and to resolve the variability introduced to the simulated radar signal due to cloud overlap.

2.3 Microphysical assumptions in the forward operator

As mentioned in section 2.1, ZmVar generates radar reflectivities using a pre-calculated table containing volumetric extinction and equivalent reflectivity for the relevant frequency (94 GHz in this case) and for predefined hydrometeor types. For the generation of this lookup table, definitions of hydrometeor single particle scattering properties and PSDs are required. If possible, these assumptions should be consistent with the ones made in the forecast model used to generate the cloud profiles which are input to the simulator. However, there are two reasons why independent assumptions may be required. Firstly, a model may not prescribe consistent assumptions across the parametrizations in the model (clouds, convection, radiation) or even within a parametrization. Secondly, reflectivity is particularly sensitive to the large end of the particle size distribution (D^6 in Rayleigh regime), whereas microphysical processes depend more on the mass-weighted part of the PSD (D^3) and (scattering of) radiation on the particle cross sectional area (D^2). Therefore the emphasis on the PSD may be different

for the radar forward operator than for the model physics. In addition, there is significant spatial and temporal variability of hydrometeor PSDs and scattering properties in the atmosphere and significant uncertainty in representing this information appropriately in a global model. Hence the approach here is to make independent assumptions for PSDs and particle optical properties in the forward operator. A brief summary of the parameterizations for the PSDs and single particle scattering properties in the original configuration developed during the QuARL project is given below.

2.3.1 PSDs

The PSD for rain in ZmVar is modelled using an exponential function following Marshall and Palmer (1948). For cloud liquid, a lognormal PSD is used which, unlike the exponential, does not change shape with changes in the water content. The standard setup in ZmVar uses the values of median diameter and logarithm width for maritime clouds as reported in Table 3 of Miles et al. (2000). For cloud ice, the size distribution is modelled as a Gamma distribution, which has a spread wider than the lognormal function. Again, the shape of the Gamma PSD does not depend on the amount of ice content. For snow, a constant-slope exponential PSD is used, where the slope parameter Γ is fixed to 50 cm^{-1}. While for rain the same PSD has been used to model stratiform and convective precipitation, for snow it is different in the two cases. The Gamma shape parameter μ is set to 0 (reducing the PSD to an exponential) for stratiform snow and it is set to 5 for convective snow. This choice aims at reducing the number of precipitating ice particles with the smallest size (due to collection processes) in convective situations.

2.3.2 Single particle scattering properties

In the original ZmVar table, the backscattering cross section, $\sigma_{bsc}^h(D)$, and the extinction cross section, $\sigma_{ext}^h(D)$, for a single particle (of size D) are computed using the Mie solution (Mie, 1908) for a sphere (at the specified temperature and the given wavelength λ). To deduce the optical properties of water or ice particles one needs a model for the interaction with the radiative electro-magnetic field. The key quantity of such a description is the permittivity $\varepsilon = \varepsilon' + i\varepsilon''$ which describes the impact that the internal distribution of charges has on an external dielectric field. Generally, the response of the internal charges to an external field is time dependent and it is also a function of temperature. As a result, the permittivity depends on the frequency of the external field as well as the temperature of the responding medium. The real part ε' of ε describes the dispersion of the phase delay induced on an electromagnetic wave passing through a medium, while ε'' represents the loss of energy. The complex refractive index $m = m' - i\, m''$ is related to ε through $m = \sqrt{\varepsilon}$ ZmVar uses the water permittivity model of Liebe et al. (1991), while the model of Mätzler and Wegmüller (1988) is implemented for ice. Cloud ice is modelled as being solid ice, independent of size. Snow is modelled as particles composed of an air-ice mixture, i.e. with a density ρ lower than the one of pure ice. Under the assumption of random orientation and based on the particle projected area and volume, the actual density can be expressed as a power law function of the diameter:

$$\rho(D) = aD^b \tag{2.13}$$

where D is interpreted as the maximum dimension of the particle. Snow particles use the density relationship of Brown and Francis (1995) since Hogan et al. (2006) showed that this formulation leads to an excellent agreement between aircraft-calculated and radar-observed reflectivity. The corresponding effective permittivity is evaluated using the model of Maxwell Garnett (1904). It is also worth mentioning that ZmVar includes a model for treating melting ice particles below the freezing level (Bauer et al., 2000). However, the modelling of the melting layer is of secondary importance here, since the bright band phenomenon is weak at a frequency of 94 GHz due to attenuation (Kollias and Albrecht, 2005). In addition, the melting layer is resolved only very crudely by the coarse vertical resolution of the model. The PSD functional form, fixed parameters, particle shape and density for each hydrometeor type are summarized in Tab. 2.1 for the original configuration of ZmVar. Figure 2.2 shows the main scattering/extinction quantities assigned to cloud liquid, cloud ice, stratiform rain and stratiform snow as function of the size.

Hydrometeor	Particle shape	PSD and parameters	Density $[g/cm^3]$ $\rho(D) = aD^b$
Cloud Liquid	Sphere	$n(r) = \frac{N_t}{\sqrt{2\pi}(\ln\sigma_g)r} e^{-\frac{\ln^2(r/r_g)}{2(\ln\sigma_g)^2}}$ $\begin{cases} r = 0.5\,D \\ r_g = 30\,\mu m \\ \sigma_g = 7\cdot10^{-4} \end{cases}$	$a = 1$ $b = 0$
Cloud Ice	Sphere	$N(D) = N_t \frac{1}{\Gamma(\mu)D_n}\left(\frac{D}{D_n}\right)^{\mu-1} D^\mu e^{-\frac{D}{D_n}}$ $\begin{cases} D_n = 50\,\mu m \\ \mu = 2 \end{cases}$	$a = 0.916$ $b = 0$
Rain LS	Sphere	$N(D) = N_0 e^{-\gamma D}$ $N_0 = 0.08\,cm^{-4}$	$a = 1$ $b = 0$
Snow LS	Sphere	$N(D) = N_0 e^{-\gamma D}$ $\gamma = 50\,cm^{-1}$	$a = 5.615\,10^{-3}$ $b = -1.1$
Rain Conv	Sphere	$N(D) = N_0 e^{-\gamma D}$ $N_0 = 0.08\,cm^{-4}$	$a = 1$ $b = 0$
Snow Conv	Sphere	$N(D) = N_0 D^\mu e^{-\gamma D}$ $\begin{cases} \gamma = 50\,cm^{-1} \\ \mu = 7 \end{cases}$	$a = 5.615\,10^{-3}$ $b = -1.1$

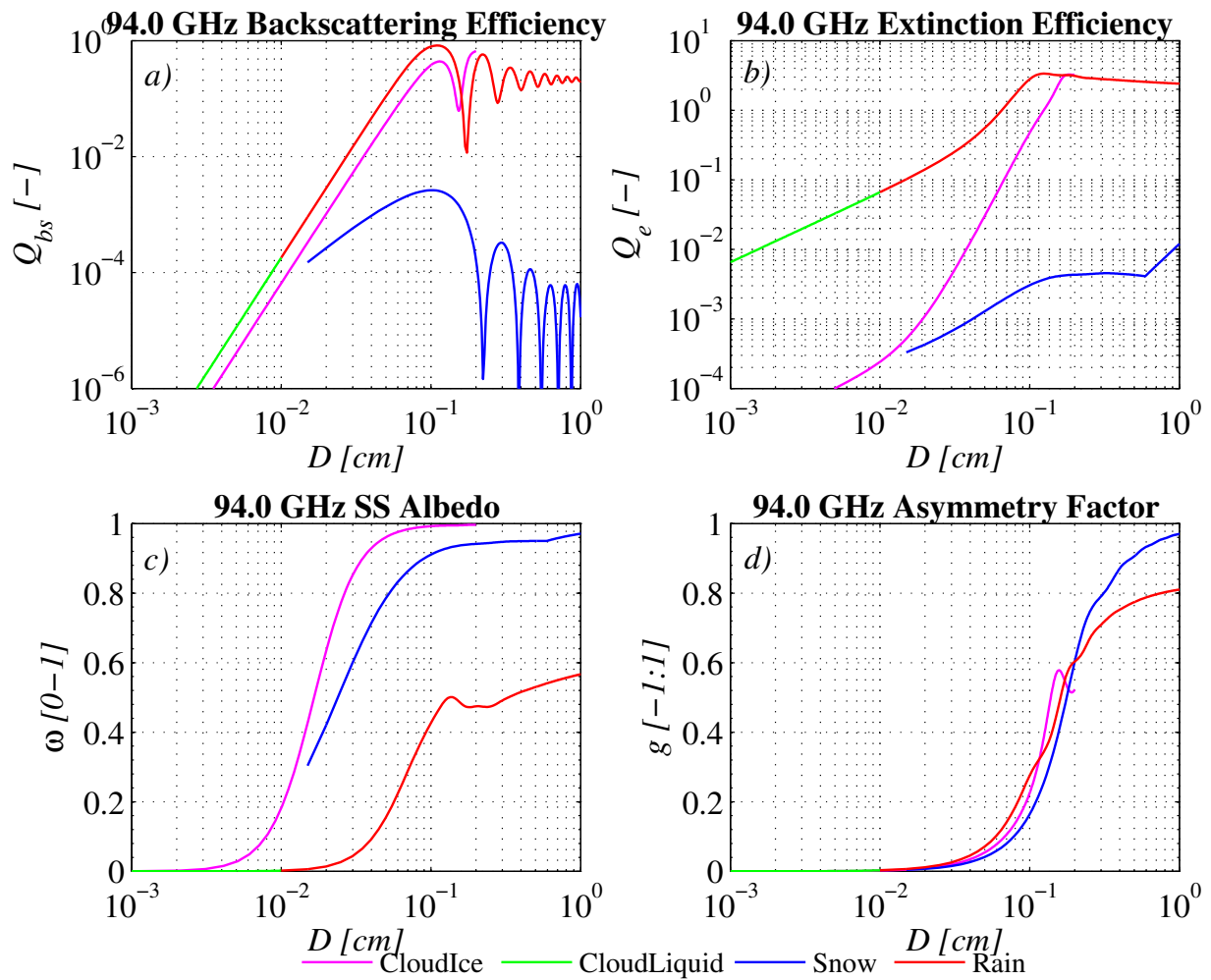*Table 2.1: Original parametrization of hydrometeor particle properties in ZmVar.*

Figure 2.2: Single particle backscattering efficiency (panel a), extinction efficiency (panel b), single scattering albedo (panel c) and asymmetry factor (panel d) as function of size used in the QuARL configuration of the ZmVar lookup table for each of the represented hydrometeor categories.

2.3.3 Reflectivity versus hydrometeor content relationships

Once the single particle scattering properties and a particle size distribution $N^h(D)$ are assumed, the equivalent radar reflectivity Z^h of each hydrometeor h can be evaluated by integrating over the size spectrum according Eq. 2.1. The resulting relationships between hydrometeor content and radar reflectivity specified in the lookup table are shown by the curves in Fig. 2.3. For cloud droplets, non precipitating ice, rain and solid precipitation (snow) hydrometeor types, two lines of the same colour are plotted, referring to the two extremes of the temperature-dependent relationships. The left one refers to the minimum temperature and the right one to the maximum temperature specified in the lookup table. The water permittivity model of Liebe et al. (1991) has been used for temperatures as low as -20°C, a value for which this model was not tested. However, this was a necessary choice since, at present, a model for the permittivity of supercooled cloud drops is not available.

As an example, Fig. 2.4 shows a cloud structure observed by CloudSat and represented by the IFS using ZmVar. CloudSat observations at the original resolution and averaged (horizontally and vertically) to the model grid resolution are shown in panels (a) and (b), respectively. The corresponding model data with simulated reflectivities from ZmVar are shown in panel (c). Simulated observations are evaluated at model resolution by averaging the reflectivities of the hydrometeor-filled subcolumns in the grid-box, considering only values above the minimum detectable threshold (-30 dBZ for CloudSat). Comparing panels (b) and (c), we note that the forecast model is able to realistically reproduce the structure of this event, and the general distribution and magnitude of radar reflectivity values. The contingency mask is given in Fig. 2.4(d) showing the correspondence between model and observations. Note that the difference in the lowest kilometre is due to the removal of CloudSat data contaminated by surface clutter. The grid-box hydrometeor fractions for CloudSat and the model (panels (e) and (f), respectively) show that the model broadly captures the differences between the higher cloud fraction region in the frontal system between 7000 and 8000 km (although with some underestimate) and the lower cloud fraction in the convective region between 4700 and 5500 km.
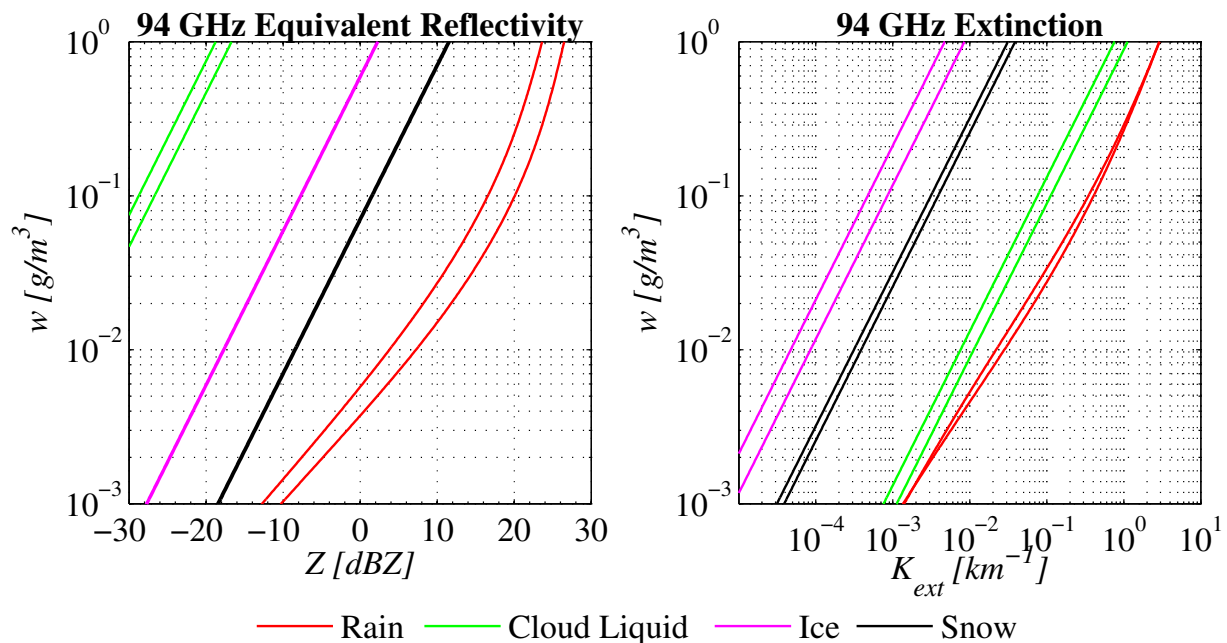

Figure 2.3: Relationships between mass content of each hydrometeor type and corresponding simulated reflectivity (panel a) and volumetric extinction (panel b) as prescribed in ZmVar lookup table. Red, black, green, magenta lines refer to rain, snow, cloud liquid and cloud ice, respectively. The two lines correspond to the minimum and maximum value of temperature range used.
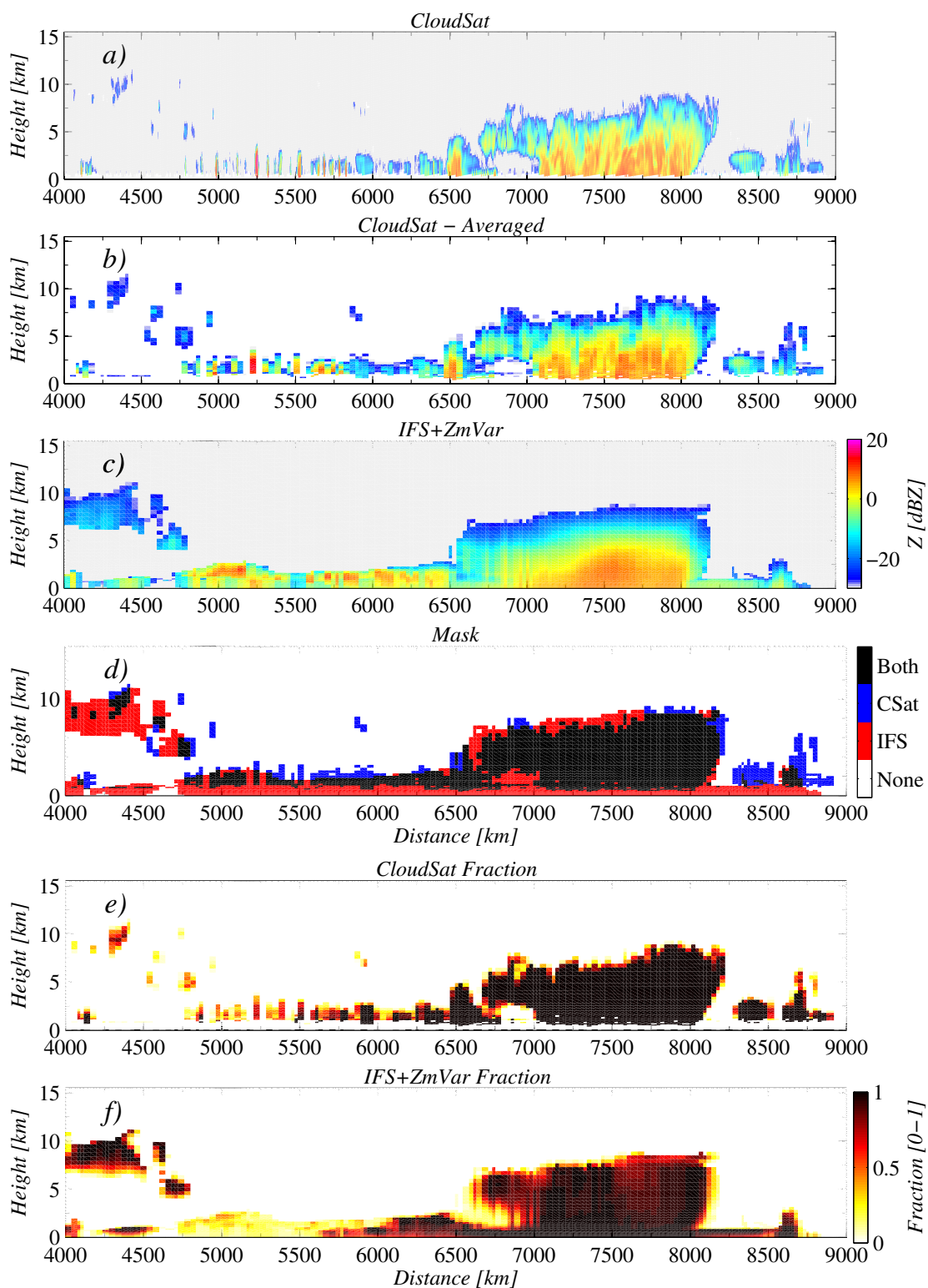
Figure 2.4: Reflectivity of a cloud system as observed by CloudSat on 2 January 2007 (panel a). Panel (b) shows the same CloudSat reflectivities averaged to the IFS model grid resolution. Panel (c) displays the corresponding simulated reflectivities using ZmVar on the ECMWF model short-term forecasts. Also shown: contingency mask between observed and simulated reflectivities (panel d) and the hydrometeor fraction in each grid box for observations and model respectively (panels e and f).

2.4 Updates to the ZmVar radar forward operator

The ZmVar reflectivity simulator has undergone a number of changes in order to improve the microphysical representation, to better model the measured reflectivity, and to be consistent with some changes in the IFS model. Each of them is described in the following subsections.

2.4.1 Updates to the microphysical assumptions

The lookup table of scattering properties has been updated to give a better representation of particles, especially of the frozen ones, by using more sophisticated assumptions.

Changes to PSDs

PSD is the quantity with the largest degree of variability and uncertainty. The rain PSD, has been made more general replacing the exponential distribution with the normalized gamma distribution proposed by Illingworth and Blackman (2002). The latter has the advantage of being able to describe the shape of the size distribution for a specific precipitation regime (from light stratiform to convective precipitation). This change does not affect the simulated Z since the parameterization used for this PSD in our operator corresponds to the Marshall-Palmer previously implemented. Different expressions may be used for convective and stratiform rain PSD, but in the current version of ZmVar same values are chosen. The snow PSD has been updated, using the parameterization developed by Field et al. (2007). Based on in situ observations, it prescribes an increase in the relative number of larger particles at warmer temperatures, representing the effect of particle aggregation on the shape of the distribution. In their study, two distinct expressions are proposed, one for mid-latitude and a second for tropical clouds. In our implementation, the mid-latitude PSD function is used for both stratiform and convective snow since it is representative of a wider range of meteorological conditions. PSD of cloud ice also has been changed employing a modified version of the Field et al. (2007) PSD which does not consider aggregation. This is obtained keeping the temperature dependence in the PSD constant, setting it to a value typical of regions where aggregation is not a dominant process (chosen in $-70°$C in our case). The main reason for this approach is to create continuity and consistency between the PSDs of precipitating and non-precipitating ice.

Changes to the single particle ice scattering properties

Sphericity, being a convenient approximation, was used during the QuARL project, but it is not a valid assumption in the majority of cases. In reality instead, ice particles can assume a variety of shapes depending on their growth environment and on past evolution. In this study, the sphericity hypothesis has been replaced by a more realistic ice particle habits (shapes). This implies a more complex (and computationally expensive) solution to the problem of finding the corresponding optical properties. The approach followed here makes use of the discrete dipole approximation, as done in many similar works (e.g. Evans and Stephens, 1995). For cloud ice, the values of single particle backscattering and extinction given by Liu (2008) for 5-bullet rosettes (Type 07) are used. This choice is justified by the fact that in-situ observations reveal that bullet rosettes are the most commonly observed shapes in the upper layers of ice clouds (Heymsfield et al., 2002). Following Korolev et al. (2000), who found that most particle size distributions are dominated by irregular shapes, an aggregate model is used for snow. The single particle scattering properties of both stratiform and convective snow have been modelled using the results for aggregates of columns evaluated in Hong (2007). To fully determine the PSD for a given total mass, the particle density of frozen particles also needs to be specified. However, this cannot be done arbitrarily as it must be consistent with the choice of particle shape. The assumed values are given in Tab. 2.2, which summarizes the updated setup of ZmVar. Figure 2.5 shows the single-particle extinction/scattering properties for the updated configuration. Comparing them with the previous ones in Fig. 2.5, we note that snow and cloud ice backscattering are now of the same order of magnitude. The new relationships of hydrometeor content versus reflectivity are in Fig. 2.6. The main change with respect to Fig. 2.3 is a much

wider temperature-dependence of the snow resulting from using the Field et al. (2007) PSD.

Hydrometeor	Particle shape	PSD and parameters		Density $[g/cm^3]$ $\rho(D) = aD^b$		
Cloud Liquid	Sphere	$n(r) = \frac{N_t}{\sqrt{2\pi}(\ln\sigma_g)r} e^{-\frac{\ln^2(r/r_g)}{2(\ln\sigma_g)^2}}$	$\begin{cases} r_g = 30\mu m \\ \sigma_g = 7\,10^{-4} \end{cases}$	a	=	1
				b	=	0
Cloud Ice	6-bullet rosette	*Field et al. (2007), no Temp dependence*		a	=	$9.4\,10^{-3}$
				b	=	-0.87
Rain LS	Sphere	$N(D) = \frac{0.03 N_L D_0^4 \Lambda^{\mu+4}}{\Gamma(\mu+4)} D^\mu e^{-\Lambda D}$	$\begin{cases} N_L = 0.08 cm^{-4} \\ \Lambda = \frac{3.6+\mu}{D_0} \end{cases}$	a	=	1
				b	=	0
Snow LS	Aggregate of columns	*Field et al. (2007)*		a	=	$5.615\,10^{-3}$
				b	=	-1.1
Rain Conv	Sphere	$N(D) = \frac{0.03 N_L D_0^4 \Lambda^{\mu+4}}{\Gamma(\mu+4)} D^\mu e^{-\Lambda D}$	$\begin{cases} N_L = 0.08 cm^{-4} \\ \Lambda = \frac{3.6+\mu}{D_0} \end{cases}$	a	=	1
				b	=	0
Snow Conv	Aggregate of columns	*Field et al. (2007)*		a	=	$5.615\,10^{-3}$
				b	=	-1.1

Table 2.2: Revised parametrization of hydrometeor particle properties in ZmVar.
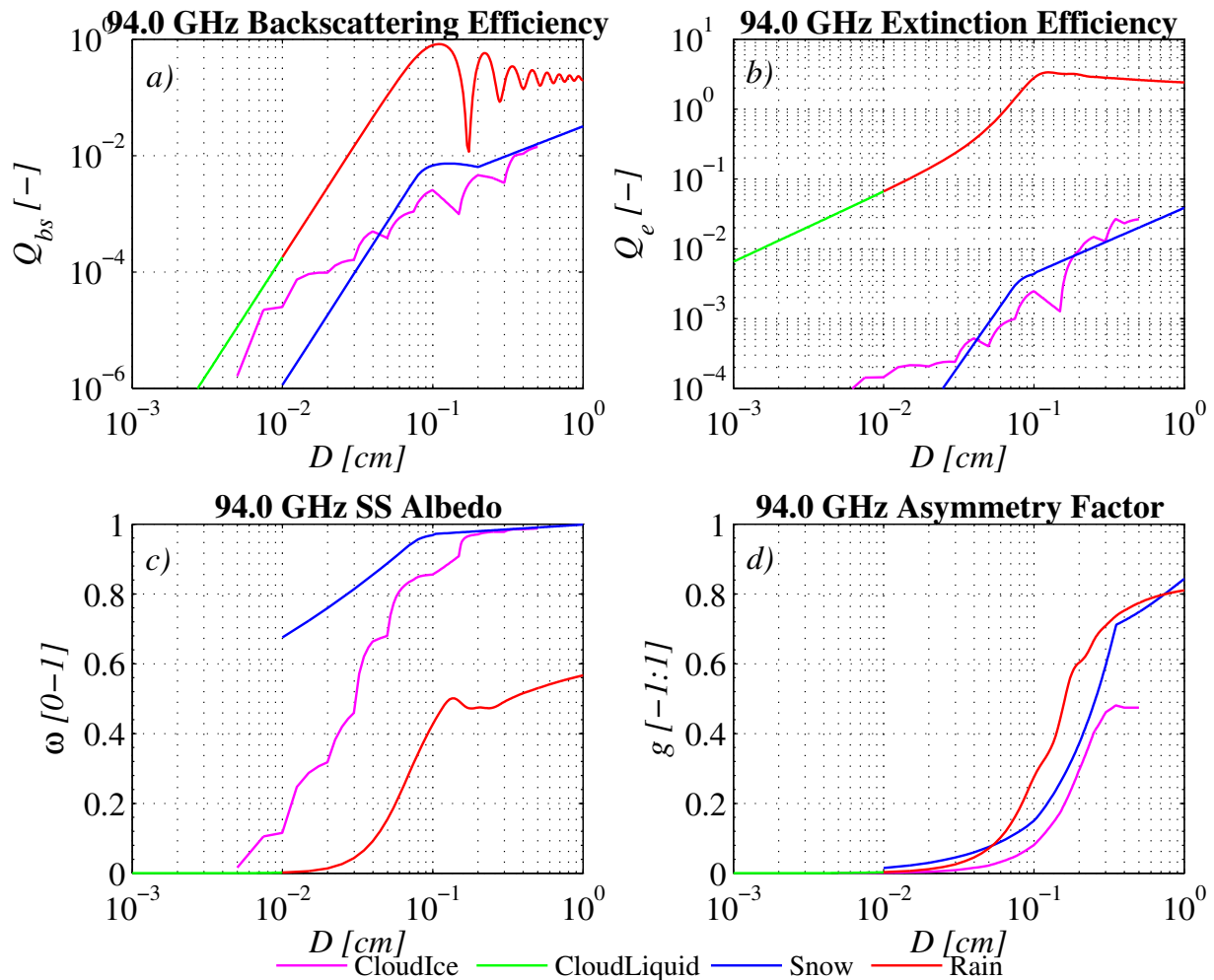

Figure 2.5: As Fig. 2.2, but for the new configuration of the ZmVar lookup table.
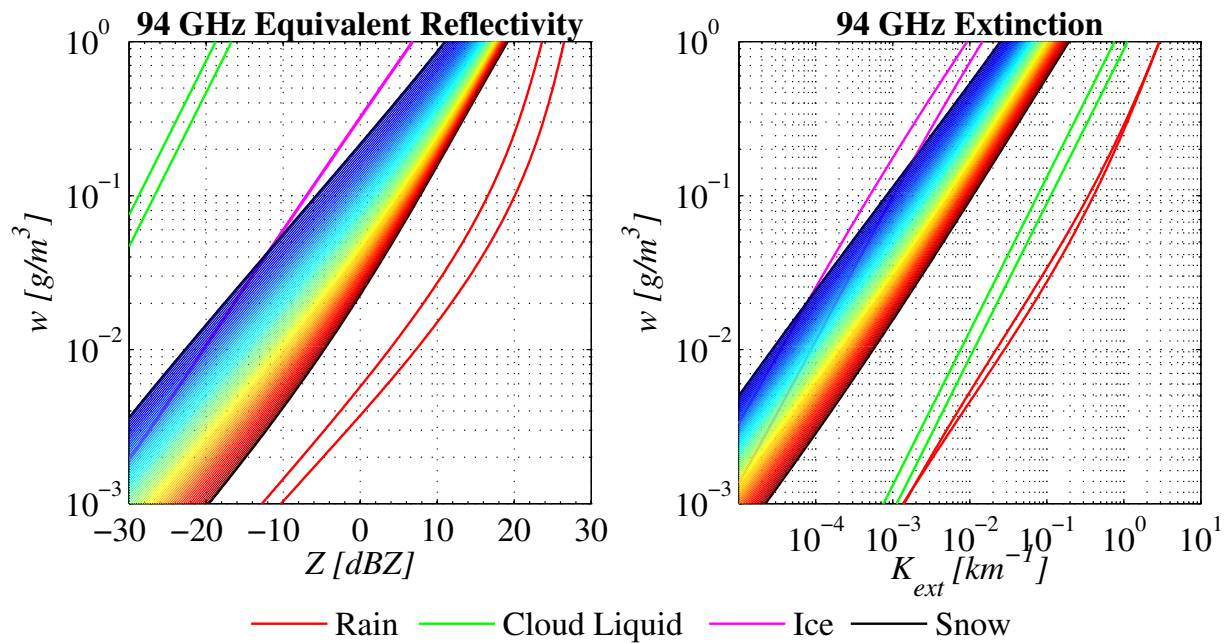
Figure 2.6: *As Fig. 2.3, but for the new configuration of the ZmVar lookup table. Envelopes of rainbowshaded-coloured lines show the temperature dependence of snow from $-70°C$ (deep bluemost left) to $0°C$ (brownmost right), with a $1°C$ stepping.*

2.4.2 *Updates to the sub-grid precipitation fraction*

Rain and snow precipitation make a significant contribution to, and often dominate, the radar reflectivity signal due to the larger size of the precipitating hydrometeors. While the vertical overlap of clouds is routinely considered for radiative transfer calculations in models, less attention is given to precipitation fraction and how precipitation and clouds are arranged in the vertical. In the IFS, a precipitation fraction is used to convert grid box average rain and snow contents into the equivalent of an in-cloud (or rather, in-precipitation, or local) value for the evaporation process. The strategy for determining precipitation overlap with ZmVar is illustrated in Fig. 2.7. For a given model cloud fraction (panel a), cloud overlap is calculated first (as described in section 2.3b) providing a set of subcolumns with binary cloud cover (panel b). The cloudy subcolumns are ranked according to the number of cloud-filled levels in each subcolumn (panel c). Panel (d) shows the stratiform precipitation fraction, as obtained from the IFS. Subcolumns are then flagged as precipitation-filled based on the existence of large-scale precipitation at a given model level, and the existence of either cloud at the current level, or precipitation in the level above (panel e). Thus precipitation fraction is directly related to cloud formation and is maximally overlapped with cloud (referred to here as the PMAX precipitation fraction). With the introduction of a prognostic scheme for precipitation, beside to the already available cloud/ice fraction, the IFS model also provides the fraction of stratiform precipitation. The ZmVar has been therefore extended with an optional step where the stratiform precipitation fraction on subcolumns given by ZmVar is adjusted to be consistent with the precipitation fraction parameterization of the IFS. As shown again in Fig. 2.7 (panel f), stratiform precipitation is set to zero first in the least-cloudy subcolumn when (at a given level) the fraction has to be reduced, and it is added to the most cloudy subcolumn if fraction needs to be increased. Overall, the adjustment leads to a reduction of the first-guess precipitation fraction in ZmVar. This may be interpreted as reintroducing a reduction in precipitation fraction due to re-evaporation, which is not explicitly accounted for in ZmVar, but is considered in the IFS (referred to here as the PEVAP precipitation fraction). Since IFS does not predict a fraction estimate for convective precipitation, a fixed fraction of 10% is assumed in ZmVar for the evaporation of convective

precipitation. If convective precipitation exists on a vertical level (marked by a black bar in panel (d)), 10% of subcolumns are flagged, again starting with the most cloudy subcolumn (panel f). If stratiform precipitation already exists, the total cloud fraction in the model column is not altered. Where no stratiform precipitation previously existed, the convective precipitation fraction is added.

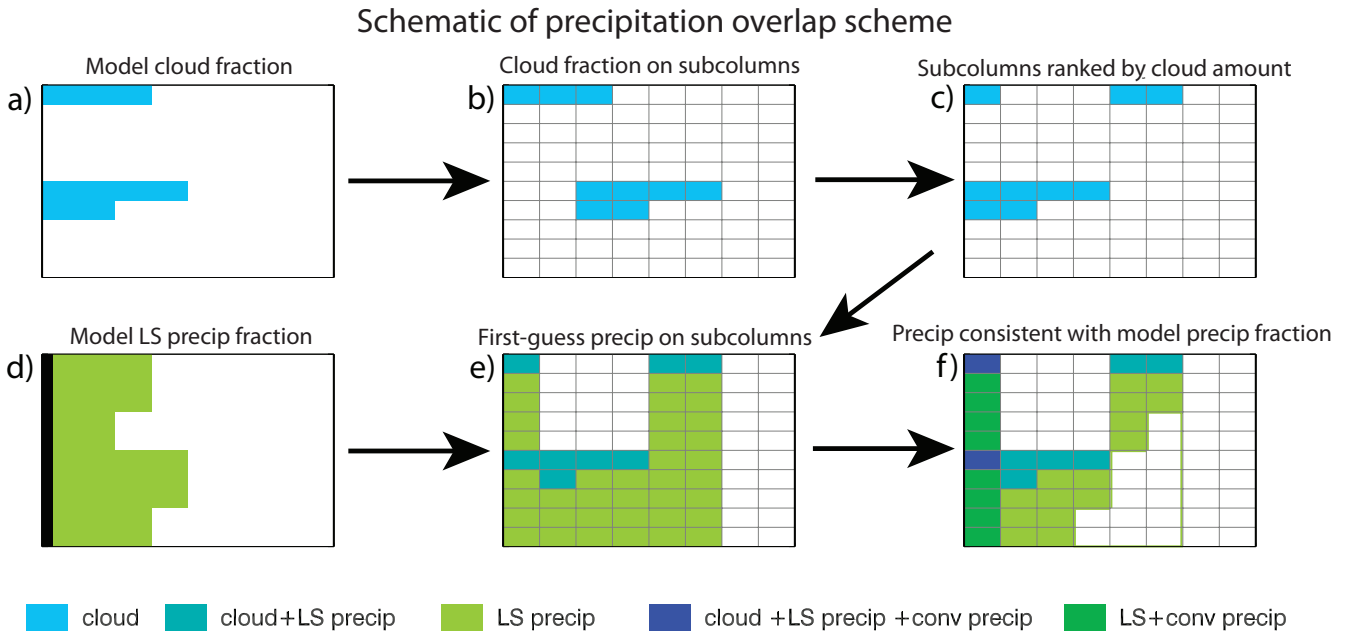## Schematic of precipitation overlap scheme

Figure 2.7: Schematic diagram describing the precipitation overlap scheme. (a) Cloud fraction as provided by the model. (b) Cloud overlap determined and expressed as a binary cloud fraction on subcolumns. (c) Subcolumns ranked by number of cloud-filled vertical levels. (d) Large-scale precipitation fraction provided by the model, together with a flag for the existence of convective precipitation (black bar). (e) Precipitation and clouds maximally overlapped as a first guess (PMAX). (f) As an optional step, adjusted first guess precipitation fraction to match the model precipitation fraction from (d) (PEVAP). If vertical levels are flagged as containing convective precipitation, a fixed fraction of subcolumns (one in this example) is flagged as containing convective, as well as large-scale precipitation.

2.4.3   Contoured Frequency by Altitude Diagrams

PDF of reflectivity along height (or temperature), commonly referred as Contoured Frequency by Altitude Diagrams (CFADs), can provide useful information on the quality of the simulated observations. Figure 2.8 shows these diagrams for the global CloudSat data for January 2007 at model grid resolution of 25 km as well as the simulated reflectivity from the original QuARL and new versions of ZmVar. The contour plots result from the PDFs of reflectivity at each temperature range (individually normalized). The comparison of each of the CFADs for ZmVar with CloudSat one shows that the new parameterization improves the transition around the freezing level and increases the spread of reflectivities in a better agreement with observations. However, some deficiencies associated with the reflectivity simulator, but also with imperfection in the input model cloud profiles for the simulator, still remain. In particular, the following can be noticed: (i) a wider range of reflectivities at very cold temperatures (< 220 K) associated with the general over-estimate in mid-latitudes and under-estimate in the tropics for high altitude ice cloud, (ii) the narrower distribution and underestimate of mean reflectivities closer to mid-levels (220 K to 270 K) and (iii) the over-occurrence of higher reflectivities (0 to 5 dBZ) at temperatures above freezing associated with small precipitation size particles, with a corresponding under-estimate of reflectivities associated with cloud size drops (-30 to -25 dBZ). These discrepancies (discussed in Di Michele et al., 2012) will be revisited in the future, when the cloud parameterization of the model will undergo modifications/improvements.
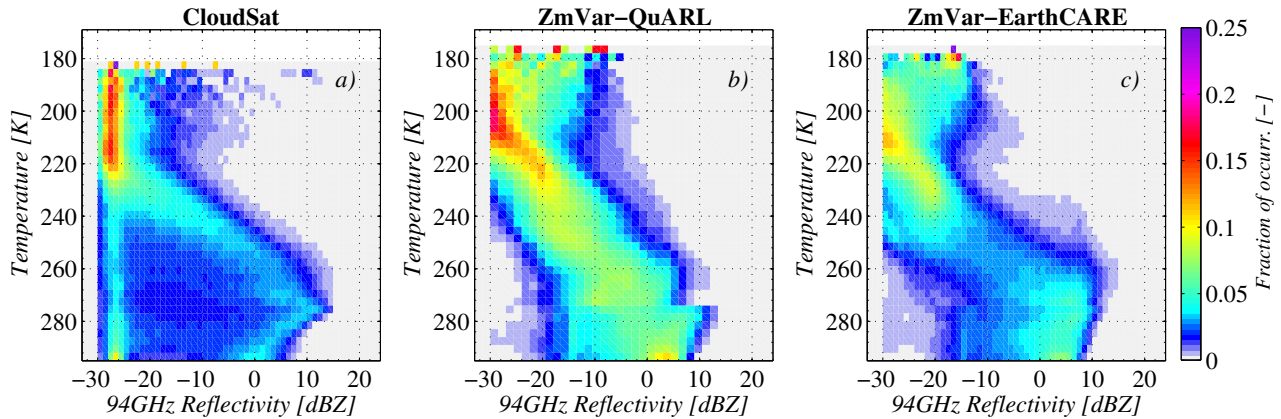
Figure 2.8: Frequency distribution of radar reflectivity with temperature. Panel (a) shows CloudSat observations for January 2007 after averaging at model resolution. Panels (b) and (c) show the corresponding simulated reflectivities (model grid box mean) using the original QuARL and updated versions of ZmVar, respectively.

2.5   Inclusion of a model for multiple scattering

Although well-known in the lidar community, multiple scattering (MS) is usually disregarded in the radar community. In the single scattering (SS) approximation, the reflected received power is assumed to come from radiation which has directly travelled from the transmitter to the scatterer and has been then back-scattered to the radar receiver (possibly being attenuated within the two-way path). The key assumption is that any scattered photon leaving the geometrical field of view (FOV) of the receiver is not detected, except those in the exact backscatter direction. Thus, all possible successive orders of scattering contributions are neglected. Such conditions are usually verified at the frequencies where most of the currently radar systems (e.g. ground-based weather radars) operate. However, this is not necessarily the case for spaceborne CloudSat and EarthCARE radars, both working at 94 GHz. As the frequency of operation and the distance from the observed medium increase, the following factors come into play which have the potential to enhance the significance of higher order of scattering in the radar echo and hence to break the validity of the SS approximation:

1. The increase in the optical thickness of the medium, i.e. the increase of the probability of interaction of radiation within the medium. As a rule of thumb, the critical condition to activate considerable MS contributions is the decrease of the mean free-radiation path (defined as the inverse of the extinction coefficient) towards values comparable or smaller than the radar footprint dimension (which can be of the order of few kilometres for typical space-borne systems).

2. The increase of scattering versus absorption cross-section, i.e. the SS albedo ω, allows the radiation to interact many times within the medium before being absorbed or leaving the cloud and eventually being redirected to the receiver.

3. The change of hydrometeors scattering phase functions, that become more peaked in the forward direction, tends to keep the radiation within the instrument FOV despite scattering, thus increasing the contribution of higher orders of scattering.

MS regimes
For the purpose of modelling the process, it is convenient to distinguish two MS regimes experienced by active remote sensors:

1. Small-angle multiple scattering: For lidar application, cloud particles are typically much larger than the wavelength; hence Babinets principle states that half of the extinguished energy is scattered into a narrow

forward lobe with a 1/e half-width of $\Theta = \lambda / \pi a$, where λ is the wavelength and a is the radius of the particle (details in Hulst and Van De Hulst, 1957). These small angle forward-scattered photons may remain within the FOV of the detector (lidar FOVs are within the range of 0.15 mrad) and contribute to the apparent backscatter (e.g. red path in the top panel of Fig. 2.9). Photons that experience wide-angle scattering will leave the FOV and they will not be detected (e.g. green path in Fig. 2.9). The condition for this to occur is:

$$\frac{\ell \Theta}{\tilde{\omega}} \leq X \ll \ell_t \tag{2.14}$$

where X is the width of the footprint of the receiver at the range of the scatterers (see Fig. 2.9), $\ell = 1/k_{ext}$ is the mean-free-path and $\ell_t = \frac{\ell}{\tilde{\omega}(1-\tilde{\omega}g)}$ is the transport mean-free-path (Liou, 2002), with g representing the asymmetry factor.

2. Wide-angle multiple scattering: If the footprint size X is on the same order or larger than ℓ_t (panel b) in Fig.2.9), i.e. $X \geq \ell_t$ wide-angle scattered photons may remain within the FOV and eventually they may be detected.

In radar systems, the first MS regime never materializes since hydrometeors have sizes smaller or at most comparable to the radar wavelength. Vice versa, given the hydrometeors more isotropic-like scattering pattern, the second regime could occur. Wide-angle MS (regime II) is the main mechanism for MS in radar systems because even the largest precipitation particles are too small to produce a significant forward lobe in the phase function.

Computing MS
Among exact solutions, Monte-Carlo techniques have been widely exploited to simulate MS effects on both lidar (e.g. Bissonnette et al., 1995) and radar (Marzano et al., 2003). The main disadvantage of the Monte-Carlo approach is the computational cost. This can be drastically reduced employing analytic approximation methods. Faster calculations allow simulations of larger number and more optically thick scenes, and can even accommodate iterative retrievals in which retrieved parameters are adjusted using on-the-fly radiative calculations. Such approaches open the possibility for retrievals based on inverse or adjoint radiative transfer calculations. While analytical approximations limited to narrow-angle scattering (scattering regime I) have been developed to simulate narrow field-of-view lidar signals (e.g. Eloranta, 1998), not so much effort has been devoted to wide field-of-view applications, to which spaceborne radars fall into. Recently, Hogan and Battaglia (2008) proposed the time-dependent two-stream (TDTS) approximation, in which the atmosphere varies only in the direction r at which the instrument is pointing (typically upward or downward), and assume homogeneities in the other two orthogonal dimensions. The diffuse radiation field is described by the outgoing and incoming "streams", I^- and I^+, whose evolution with time t is given by:

$$\frac{1}{c}\frac{\partial I^+}{\partial t} = -\mu_1 \frac{\partial I^+}{\partial r} - \mu_1 k_{ext}(\gamma_1 I^+ - \gamma_2 I^-) + S^+ \tag{2.15}$$

$$\frac{1}{c}\frac{\partial I^-}{\partial t} = +\mu_1 \frac{\partial I^-}{\partial r} - \mu_1 k_{ext}(\gamma_1 I^- - \gamma_2 I^+) + S^- \tag{2.16}$$

where $\pm\mu_1$ is the cosine of the angle between these streams and the outgoing direction and the coefficients γ_1 and γ_2 are given by:

$$\gamma_1 = \frac{1 - [\tilde{\omega}(1+g)]/2}{\mu_1} \tag{2.17}$$

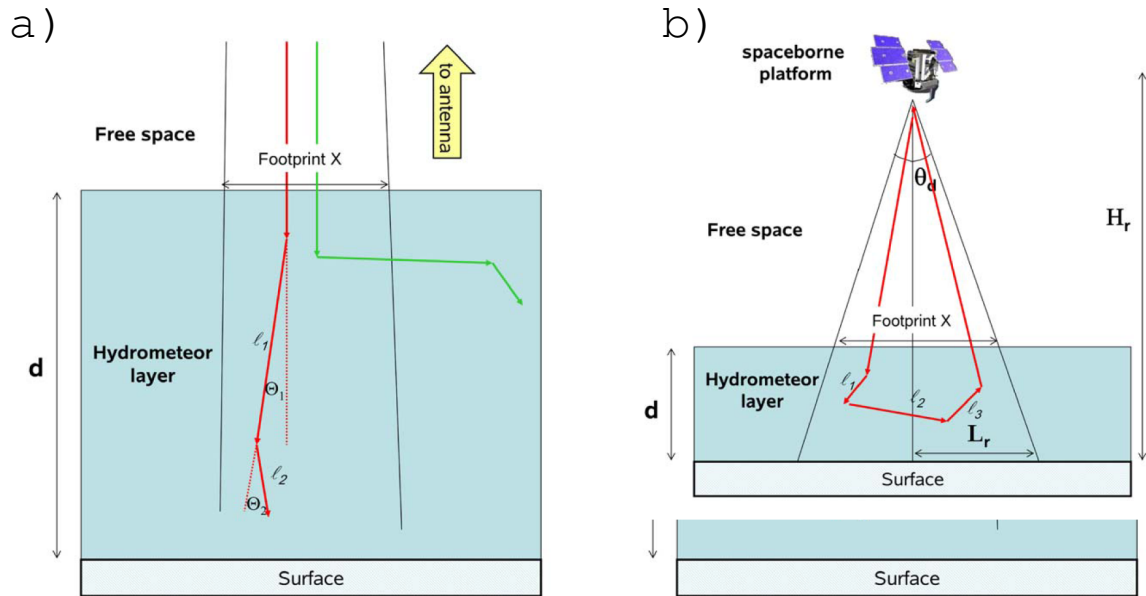$$\gamma_2 = \frac{\tilde{\omega}(1-g)}{2\mu_1} \tag{2.18}$$

Figure 2.9: Schematic representation of the two MS regimes. Radiation released from the active sensor at the top of the scene enters the scattering medium (blue shaded area) and undergoes small-angle (panel a) and wide-angle (panel b) scattering. Contributions to the received power are relevant only when coming from radiation interacting within the antenna footprint (marked by the black lines).

The terms $S^\pm$ are function of both t and r and represent the source of energy from the quasidirect beam. Eqs. 2.15 and 2.16 are simply the standard time-independent two-stream equations used in climate and weather forecast models (e.g. Meador and Weaver, 1980), but with the time derivative reintroduced from the original radiative transfer equation (Chandrasekhar, 1960). The three terms on the right-hand-side of these equations represent the transport of radiation from further upstream, exchange of radiation between streams due to scattering (and loss by absorption), and a source term due to scattering from the direct unscattered beam originating from the radar. Details on how these equations are solved can be found in Hogan and Battaglia (2008). The version of the TDTS as described in their paper, has been included into ZmVar as an optional module and has been used to perform simulated study of the MS in radar.

2.5.1 A case study

The version of ZmVar including the modelling of MS described above has been first tested on a particular case study. The selected situation is a cold front observed by CloudSat on the 1st of January 2007 in the North Atlantic. Figure 2.10 shows CloudSat observations (granule 3609) at the original resolution (top panel) and averaged along-track on the co-located IFS model grid boxes (at a resolution of about 25 km). Corresponding short-term forecast fields of cloud/precipitation from the IFS model, used as input to the ZmVar to simulate CloudSat Z are shown in Fig. 2.11 and Fig. 2.12. Interestingly, the height of the freezing level reaches the surface for most of the scene, increasing up to 2 km only on the most left 500 km of the scene. Figure 2.12 shows that in this portion of the event there is convection activity, generating some solid (convective snow) and liquid precipitation (convective rain). Synthetic reflectivities from ZmVar are given in Fig. 2.13, showing the values on the 20 single subcolums (top panel) and the result of averaging over the model grid boxes (bottom panel). The comparison of the simulated Z with the observed ones highlights a remarkable good agreement: the model is able to correctly represent the structure in the different sections of the event. There are of course discrepancies in the intensity that can be mainly attributed to the coarser model resolution and to the fact that the simulation does not necessarily exactly reproduce the position of the front structure at the time of CloudSat overpass. In order to understand how MS impacts simulated Z, the ZmVar has been run again switching off the
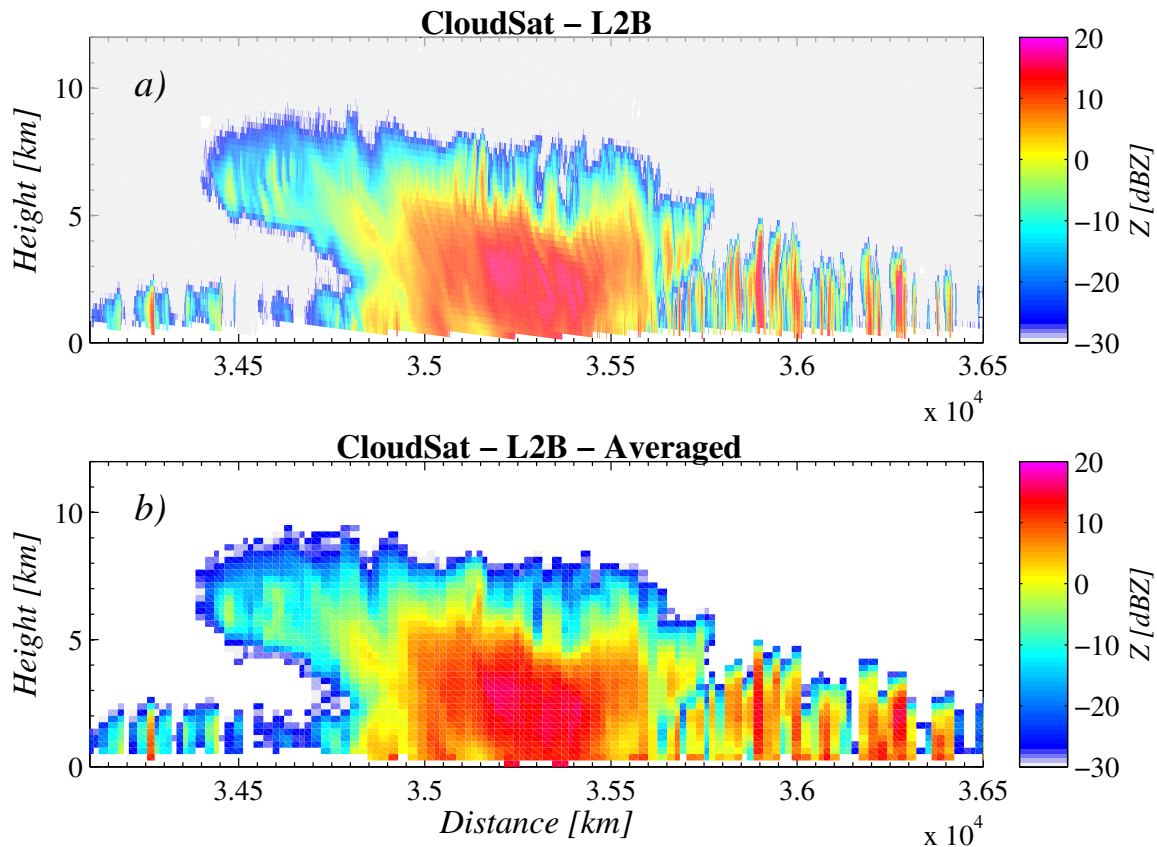
Figure 2.10: Reflectivity of a cloud system as observed by CloudSat on 1 January 2007 (panel a). Panel (b) shows the same observations after co-locating and averaging to the IFS model resolution.

MS module. The differences between Zs simulated with the inclusion of MS and the ones without are shown in Fig. 2.14. At level of the single subcolumns (top panel) we note that the increase in Z is negligible (up to 1 dBZ) everywhere except for those profiles affected by convective precipitation. In these cases the difference can be very large and likely to exceed 3 dBZ. This value in reflectivity difference is particularly meaningful since it corresponds to an increase of 100% over the SS case. The 3 dBZ value therefore can be considered a upper limit before the MS enhancement makes the return signal completely unrepresentative of the associated radar bin. When averaging at the model grid box resolution is done (mid panel), differences have the same qualitative pattern. However, there are few regions where difference is negative. These are situations where, at least in one of the subcolumns, the so-called pulse-stretching occurs, i.e. the MS pushes the Z from below to above the detectability threshold (chosen in -30 dBZ). The new simulated Z, when added to the grid box mean, lead to a lower value. This interpretation is confirmed considering the fraction of subcolumns with a Z larger that the detectability threshold. The difference between the fractions with and without MS, given in the bottom panel of Fig. 2.14, are indeed positive where differences in averaged Z are negative.
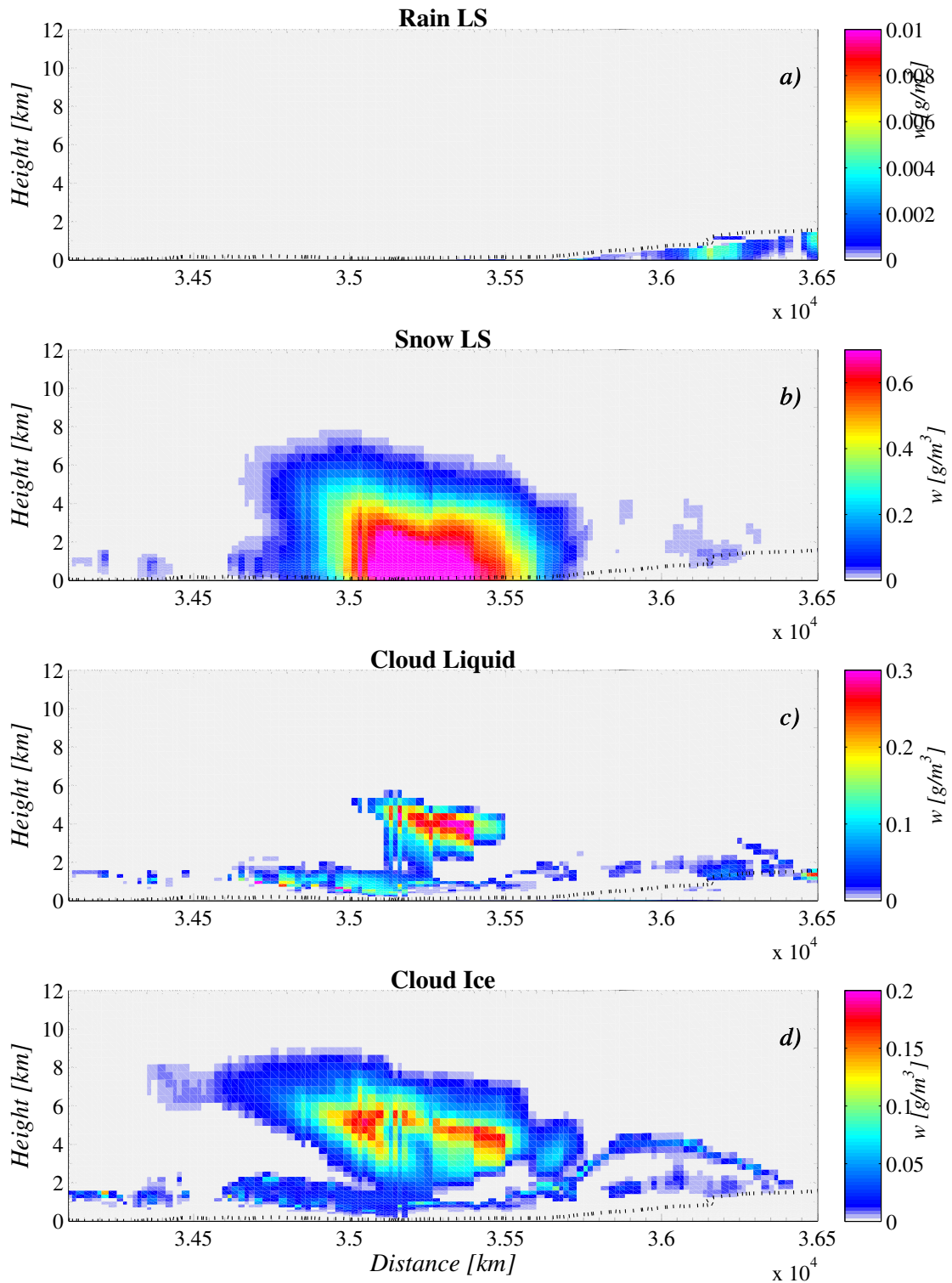
Figure 2.11: Cross section of cloud variables from the IFS model forecast corresponding to the event in Fig. 2.10. From top to bottom: large-scale rain, large-scale snow, cloud liquid, cloud ice. Dashed lines indicate the height of the freezing level.
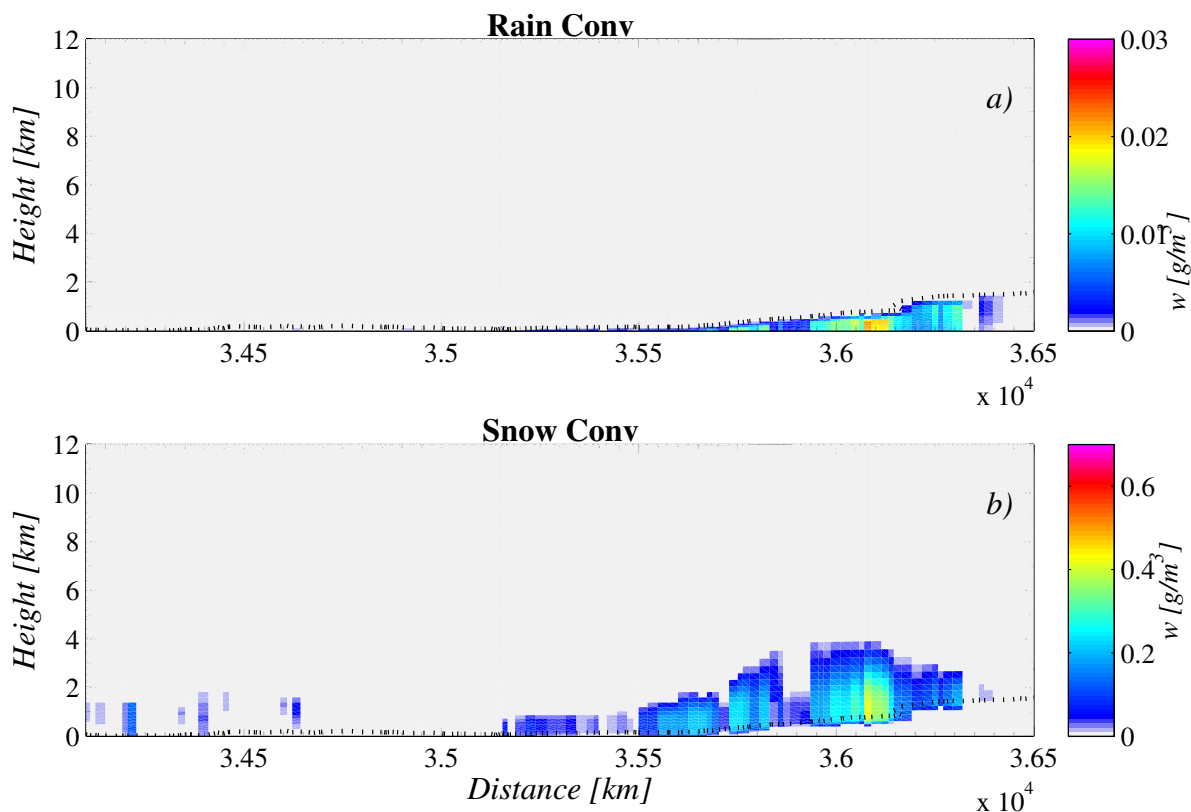
Rain Conv

Snow Conv

Figure 2.12: As Fig. 2.11, but showing convective rain and convective snow.

IFS+ZmVar − Subcols
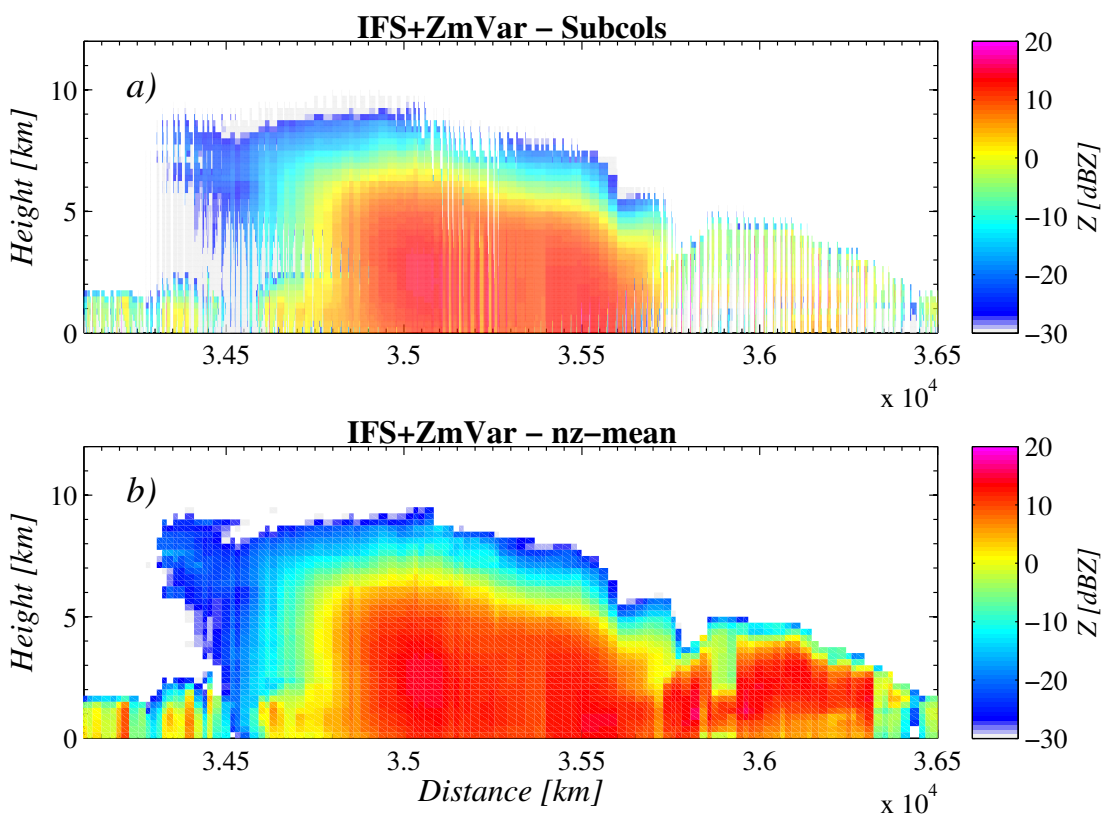
IFS+ZmVar − nz−mean

Figure 2.13: ZmVar-simulated 94 GHz radar reflectivity (including MS) relative to the case in Figs. 2.11 and 2.12. Top panel shows the output on each of 20 subcolumns, while bottom panel the grid-box average.
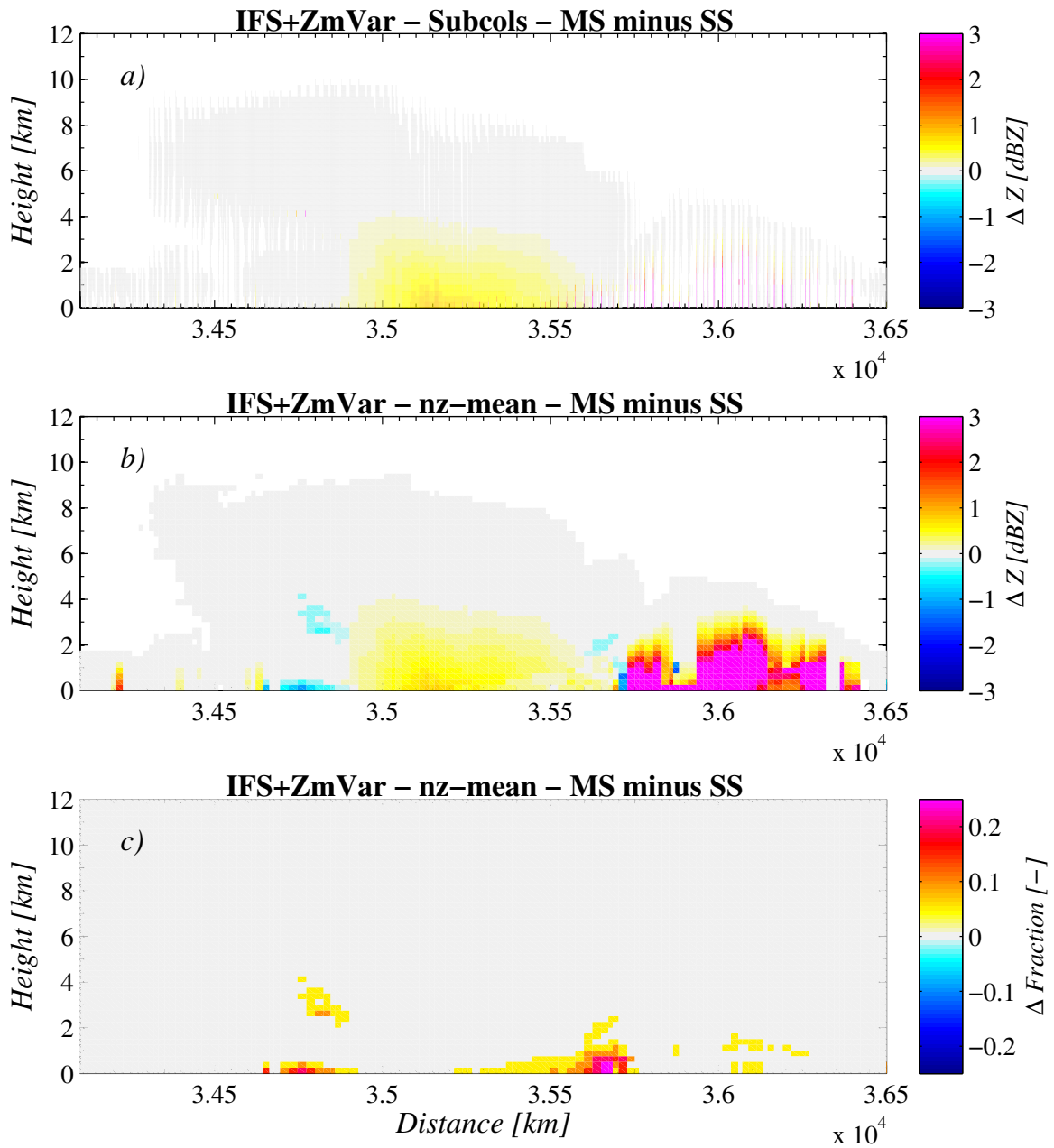
Figure 2.14: Differences using ZmVar with and without MS taken into account. Top panel and mid panels respectively show differences in reflectivity on each subcolumn and on the grid-box average. Bottom panel gives the difference of the grid-box fraction with reflectivity above the detectability threshold (chosen in -30 dBZ).

2.5.2   *Qualitative analysis*

A more comprehensive study of the effect of MS has been done by evaluating Z over a large dataset of atmospheric cases. In particular, IFS profiles collocated with CloudSat tracks over one day (about 7000 cases) were used. The analysis has been performed linking the MS enhancement to the cloud/precipitation. Over the single sub-column of ZmVar, we found that the "precipitation intensity", defined as the maximum values of precipitation along the vertical, is a convenient quantity because it correlates well to the maximum Z enhancement on the same profile. Figure 2.15 (left panel) shows the scatter between solid and liquid precipitation, while colour indicates the MS enhancement. To understand the relative role of large-scale and convective precipitation, they are also considered separately (mid and right panels). The plots clearly show that the cases with differences larger than 3dBZ (red colours) occur when precipitation intensity is very high. As expected, the right panel shows that it is convective precipitation which causes the largest enhancements. However, the MS is negligible when some small amount of convective precipitation is present. Figure 2.16 shows that the MS enhancement is small when only cases having the maximum amount of convective rain lower than 0.3 g/m^3 are considered ("moderate"). This condition, putting a limit on the amount of both liquid and solid convective precipitation, guarantees a small occurrence of MS. The few cases that escape this condition and still have large solid precipitation are removed setting a maximum threshold of 3 g/m^3 for convective snow. The complementary "intense" cases, i.e. the ones with larger convective rain, plotted in Fig. 2.17, are the ones where MS has a significant impact. We note however that, the *intense* class contains also a fair amount of cases where MS is still negligible. As shown in Fig. 2.18, they correspond to situations where the precipitation load is below 3 g/m^3 (left panel) and with low freezing level (right panel). These conditions produce a radar signal which is strong (well above 0 dBZ as in mid panel) and, because of the shallow precipitation structure, not very affected by attenuation. In these situations the signal enhancement given by MS is less evident. In fact, the MS can be seen as a reduction in attenuation and this reduction is more effective where the radar signal is not very strong. An example is given in Figs. 2.19 and 2.20 where two single precipitation profiles are considered. Below the freezing level (dashed red line), the second one differs only by the presence of some additional large-scale liquid precipitation. This extra rain content produces SS reflectivities 10 dBZ larger than in the first one (at all levels). However, counter-intuitively, the relative increase in Z due to MS is now smaller despite of having a larger amount of precipitation. Therefore, there is not a unique answer to the way how MS affect Z.

Table 2.3 shows how many IFS profiles fall into the the "moderate" and "intense" categories defined above when considering a dataset built matching CloudSat data with the IFS model forecast over 30 days. Cases over ocean, which are the ones we are considering in this study, constitute about 60% of the total. Of these, only about 1% are classified as intense, which means that on a global scale, the occurrence of significant MS effects into CloudSat observations is quite small. Battaglia et al. (2011) estimated that about 4% of CloudSat observations are affected by MS. This number is in line with our findings since our screening is based on model profiles, which have a coarser resolution (i.e. more that one MS-affected CloudSat profile can fall within a single model grid-box).

	Cases [-]	Fraction [%]	Fraction-ocean [%]
All	852233	100	-
Ocean	527689	61.9	100
Ocean-Moderate	521670	61.2	98.8
Ocean-Intense	6019	0.7	1.2

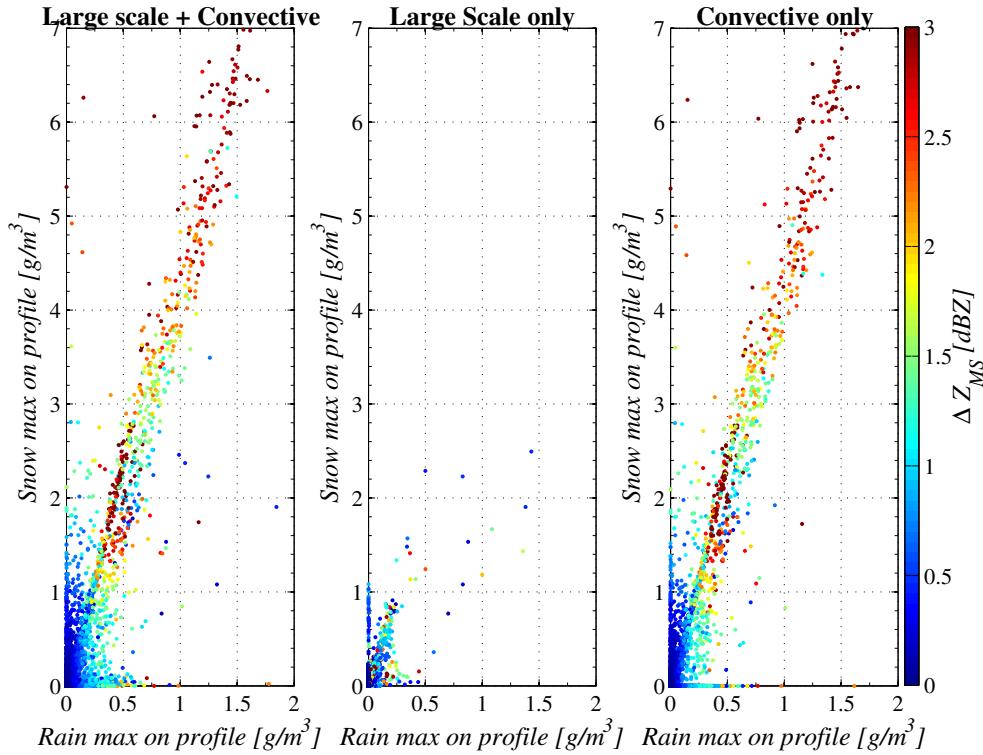*Table 2.3: Number of IFS profiles falling into the "moderate" and "intense" categories.*

Figure 2.15: *Maximum content of solid precipitation (snow) vs. maximum content of liquid precipitation (rain) based on a dataset consisting of approximately 7000 profiles. Colour of dots indicates the reflectivity enhancement due to MS as simulated by ZmVar. Left panel uses the total amount of precipitation (large-scale plus convective), while mid and right panels respectively are for the large-scale and convective components only.*
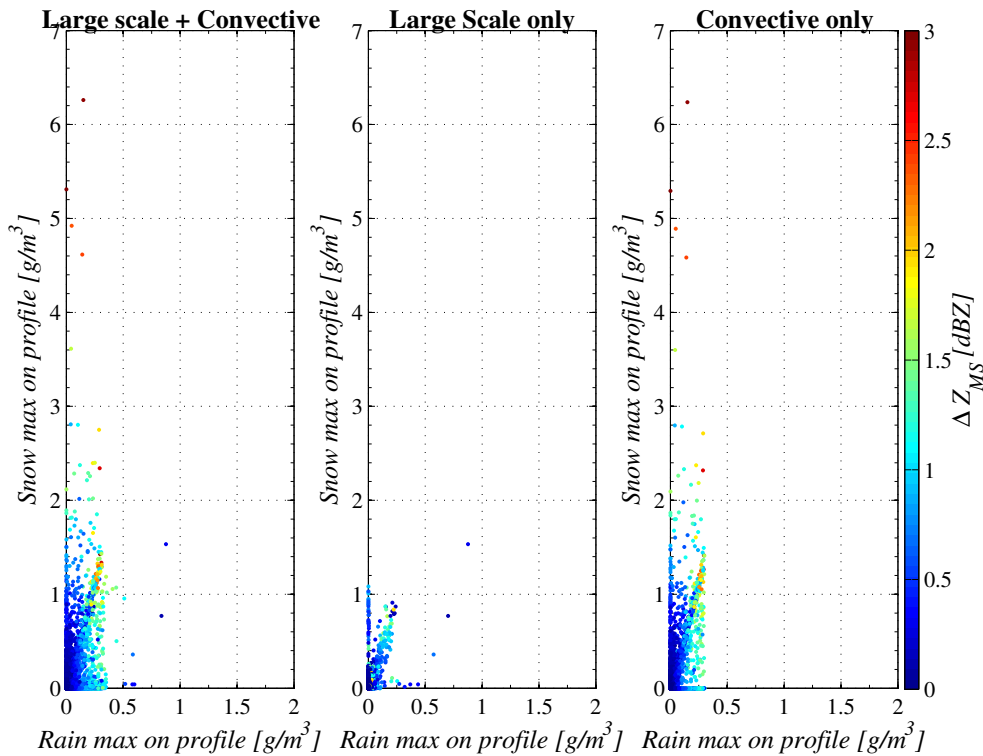

Figure 2.16: *As Fig. 2.15, but considering only model profiles with a maximum content of (in-cloud) convective rain lower than 0.3 g/m³.*
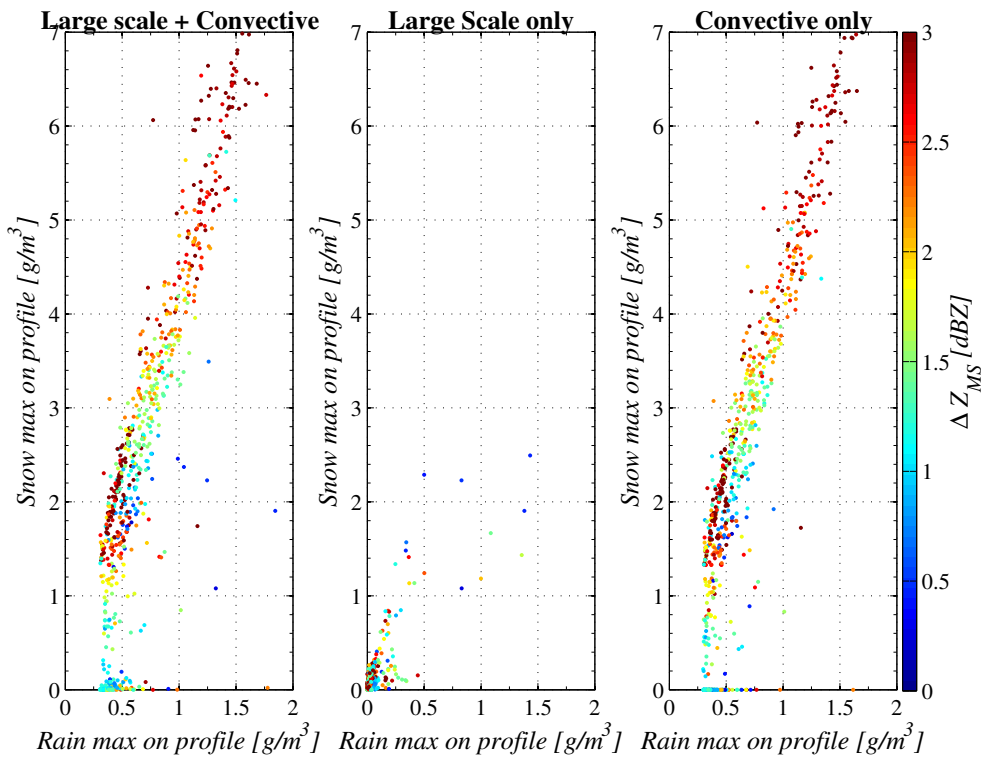
Figure 2.17: As Fig. 2.15, but considering only model profiles with a maximum content of (in-cloud) convective rain larger than 0.3 g/m^3.
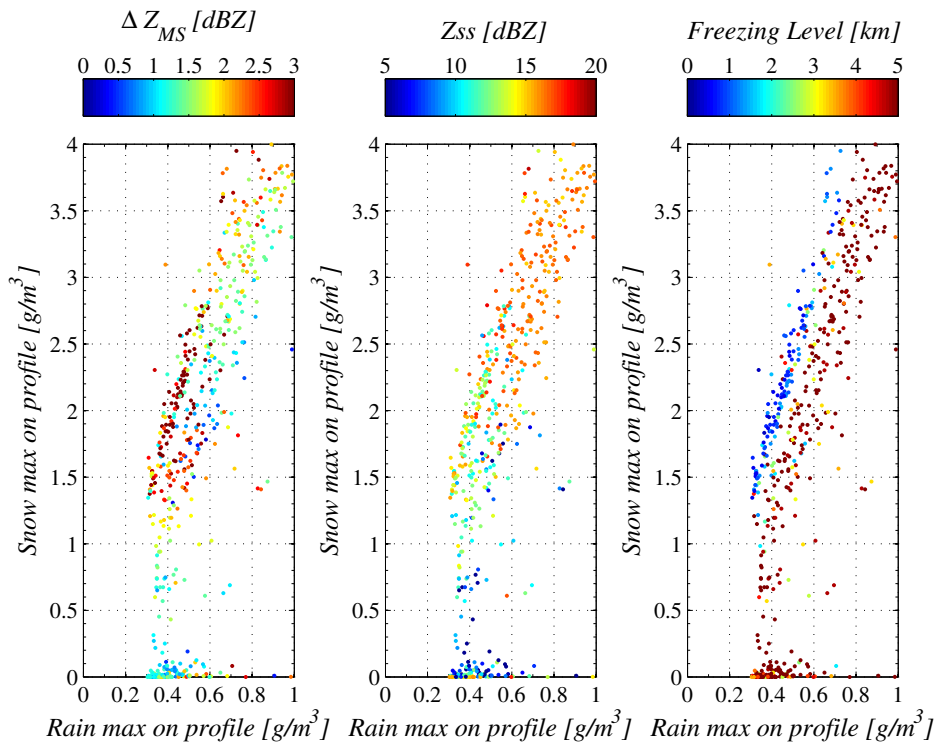

Figure 2.18: Left panel: as left panel of Fig. 2.17, but with the range of maximum content of convective rain up to 1 g/m^3. For the same cases, mid and right panels show (as colour code) the SS reflectivity and the height of freezing level, respectively.
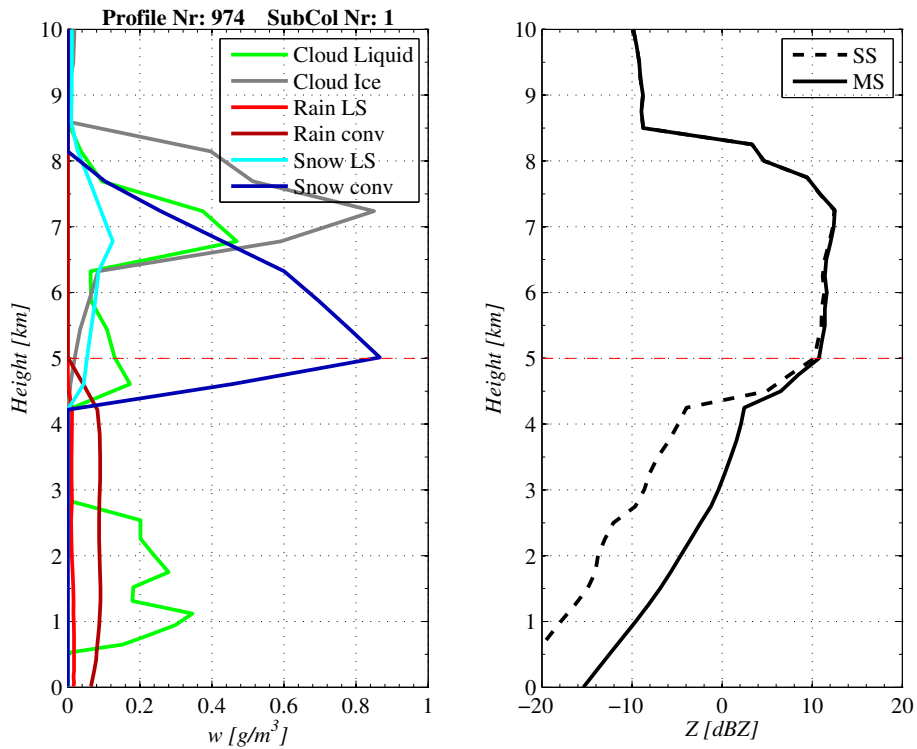
Figure 2.19: *Hydrometeor content of a single IFS model profile (left panel) and corresponding reflectivity simulated with ZmVar considering (solid line) and neglecting (dashed line) MS.*
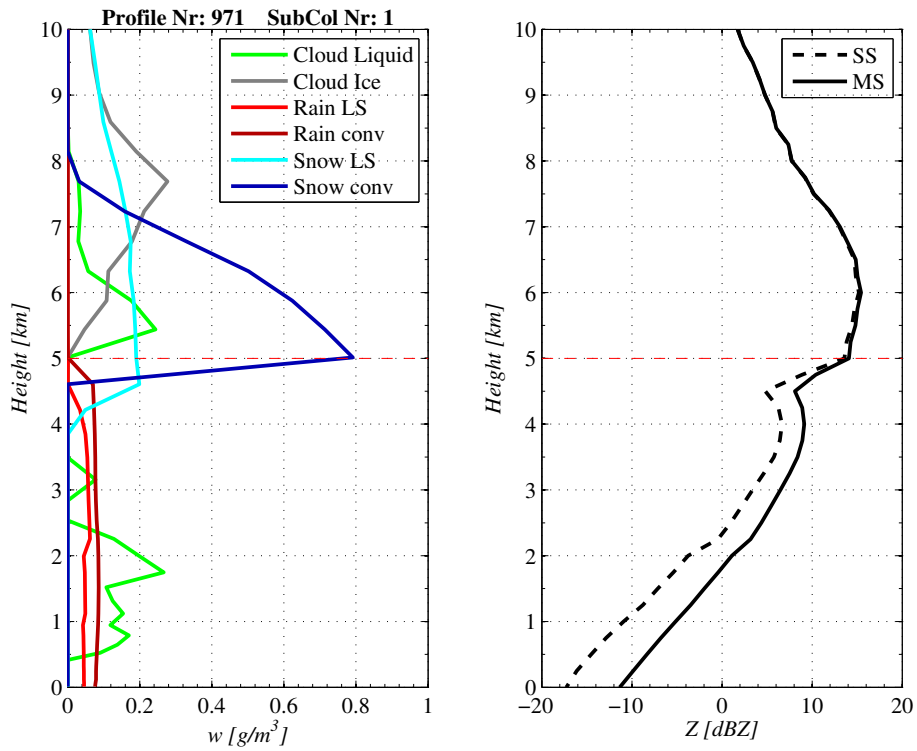

Figure 2.20: *As Fig. 2.19, but for a different profile.*

2.5.3 Comparison with CloudSat observations and computational cost

A legitimate question to answer is if the modelling of MS improves the agreement between Zs simulated by ZmVar and actual CloudSat observations. For this purpose, statistics of reflectivity differences between ZmVar and corresponding CloudSat measurements (averaged at model resolution) have been evaluated over a period of 10 days running ZmVar with and without the MS module. Results are summarized in Tabs. 2.4 and 2.5, separating between moderate and intense cases, according to the definitions given in Section 2.5.2. Each table contains standard deviation and bias of CloudSat-minus-ZmVar differences for the SS and MS runs, stratifying data over six intervals of temperatures. For the moderate situations, Tab. 2.4 shows values of error statistics for MS that are similar to the ones for SS, with only the mean difference (bias) slightly improved. Therefore, as expected, accounting for MS does not significantly affect the comparison with CloudSat.

Statistics using only intense situations are shown in Tab. 2.5. In this case, both bias and standard deviation improve at temperatures above freezing. Interestingly, the bias (CloudSat-minus-Simulated) becomes less negative in spite of MS prescribing larger simulated reflectivity on the levels of the each subcolumn. In fact, as shown in the previous section, the MS can produce lower gridbox mean reflectivity because of the pulse stretching effect. On the contrary, at temperatures below freezing, both bias and standard deviation are deteriorated by MS. This behaviour must not be attributed to a deficiency of the MS module in ZmVar itself to properly represent the phenomenon. Instead, it is most likely be interpreted as the IFS model not to be able to model the convective situations with the level of realism needed by ZmVar for a reliable evaluation of the MS enhancement.

Computational cost

The increase in computational cost of the MS is summarized in Tab. 2.6, where values correspond to a run consisting of ∼2000 profiles (and 20 subcolumns) on a machine with a Intel Duo 3GHz CPU. We note that the largest additional cost (+86 %) comes from the so called regridding process. This is a preliminary step required because the MS module works under the hypothesis of cloud/precipitation profiles uniformly spaced along the altitude. The IFS model prescribes instead profiles on irregularly distributed model levels and therefore a regridding of the model variables is necessary. In agreement with the CloudSat radar configuration, we set up a vertical grid over 250 meters (up to 20 km), resulting in 78 uniformly-spaced levels. Interestingly, the actual MS module adds only 11% to the cost of ZmVar.

In summary, the above results show that MS needs to be taken into account in situations of intense precipitation, however it is difficult to model it accurately. For cases of moderate precipitation (large scale or convective) it has a minor impact on the measurements and its modelling can be neglected, and instead its contribution can be taken into account with an additional term in the bias correction.

	Bias [$mm^6 m^{-3}$]		Standard dev. [$mm^6 m^{-3}$]	
"Moderate" class	SS	MS	SS	MS
T>169K - T<208K	0.050	0.050	0.412	0.412
T>208K - T<238K	0.045	0.045	0.595	0.600
T>238K - T<256K	0.195	0.189	2.472	2.515
T>256K - T<271K	0.280	0.271	3.849	3.875
T>271K - T<286K	0.060	0.122	3.736	3.670
T>286K - T<295K	-0.362	-0.322	1.792	1.782

Table 2.4: Statistics of difference between ZmVar simulated reflectivities (with and without considering MS) and corresponding CloudSat measurements. Only cases falling into the "moderate" class are considered.

	Bias [$mm^6 m^{-3}$]		Standard dev. [$mm^6 m^{-3}$]	
"Intense" class	SS	MS	SS	MS
T>169K - T<208K	0.162	0.161	1.441	1.441
T>208K - T<238K	0.020	-0.022	3.158	3.291
T>238K - T<256K	-0.131	-0.570	6.962	7.468
T>256K - T<271K	-0.783	-1.842	8.450	9.608
T>271K - T<286K	-1.140	-1.113	4.931	4.737
T>286K - T<295K	-1.968	-1.959	2.349	2.308

Table 2.5: As Tab. 2.4, but for cases falling into the "intense" class.

Test	Total (sec)	Incremental diff. (%)	Diff. (sec)	Incremental diff. (sec)
SS No regrid	140	-	-	-
SS Regrid 78 lev	261	+86.4	121	121
MS Regrid 78 lev	290	+11.0	150	29

Table 2.6: Computational cost of MS.

2.6 Summary from operator developments

The ZmVar forward operator that was developed during the QuARL project to simulate CloudSat data starting from the ECMWF model variables has been refined. A first improvement has resulted in a better representation of the PSD and single-particle radiative properties of frozen particles. Comparisons with CloudSat in terms of CFADs showed that, especially in the cold phase, the updated version of ZmVar provides a better agreement with observations. ZmVar has been also extended with a module for the simulation of MS effects. This new feature has been used to identify the conditions where MS is more detrimental for the radar return. It was shown that, consistently with studies based on observations (e.g. Battaglia et al., 2011) reflectivity enhancement due to MS occur in events with heavy convective precipitation. The intercomparison with CloudSat observations indicated that this phenomenon is difficult to replicate using the IFS model. For this reason, and in light of the non negligible additional computational cost, we therefore intend to neglect the modelling of MS in ZmVar for data assimilation. This choice will imply excluding any situations with heavy precipitation from the analysis, roughly corresponding to 15% of the cases.

3 Radar observation errors

In a variational data assimilation system, the goal is to find an optimal model state that simultaneously minimizes, in a least-square sense, the distance to the available observations $\mathbf{y}^o$, and to an *a-priori* model state $\mathbf{x}^b$ called background (usually provided by a short-range forecast valid at the time of assimilation). The misfit is measured by the following objective cost-function $J(\mathbf{x})$, to be minimized during the assimilation process:

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})) \qquad (3.1)$$

H indicates the observation operator which provides the model counterpart to the observations, called first guess (FG). $\mathbf{B}$ is the covariance matrix of background errors. $\mathbf{R}$ is the observation-error covariance matrix. The observation-error covariances are intended to be the sum of all errors relevant to the interpretation of the observations used in data assimilation. The magnitude of the observation error determines the weight to be given to the observation during the assimilation process. If the assumed error is large, this weight is small and so is the potential impact of the respective observation. An underestimation of the error on the other hand may have a detrimental impact on the assimilation process as the analysis would be drawn closer to the observation than the quality (reliability) of the measured values justifies.

3.1 Forward modelling and instrument errors

The assumed observation-error covariances, together with assumed background-error covariances, play an important role in determining the weight of a given observation in data-assimilation systems. Observation errors include instrument errors, errors in the observation operator, and errors of representation (from the mismatch of scales in the horizontal or vertical between the observations and the model). Under the hypothesis of uncorrelated errors, the observation-errors covariance $\mathbf{R}$ can be written as: $\mathbf{R} = \mathbf{E} + \mathbf{F} + \mathbf{O}$ where $\mathbf{E}$ the instrument-errors covariance, $\mathbf{F}$ the forward modelling-errors covariance, and $\mathbf{O}$ the representativity-errors covariance. In case of measurements from an active sensor, the off-diagonal elements of the observation-error matrices represent the covariance across measurements at different altitudes. In this work, this correlation is neglected, therefore we will use the term observation error to refer to the diagonal of the observation-error covariance matrix.

CloudSat instrument errors
In a radar measurement we can distinguish two types of errors, systematic and random. A systematic errors is typically a bias due to a calibration error. In case of CloudSat, it is corrected before data dissemination and therefore data can be considered unbiased. Radar random measurement errors are due to the combination of two factors: having a finite number of samples and the background instrument noise. Since due to the motion of the satellite it is valid to assume that each pulse is independent, the standard deviation of random errors (in dB) ΔZ_{dB} can be written as (Hogan et al., 2005):

$$\Delta Z_{dB} = \frac{4.343}{\sqrt{M}}\left(1 + \frac{1}{SNR}\right) \qquad (3.2)$$

where M and SNR are respectively the number of pulses averaged and the signal to noise ratio (in linear units). The SNR is not part of the CloudSat products, but it can be evaluated as difference between the Level-2 reflectivity Z_{dBZ} (where noise is subtracted) and the radar background noise N_{dBZ} as:

$$SNR = 10^{0.1(Z_{dBZ} - N_{dBZ})} \qquad (3.3)$$

In case of CloudSat, following Delanoë et al. (2010), the radar background noise N_{dBZ} is assumed as:

$$N_{dBZ} = -131.4 + 20\log_{10}(H_{sat}) \qquad (3.4)$$

where H_{sat} is the distance in meters of each radar bin from the satellite. Since M is known (from CloudSat Level-1 product), ΔZ_{dB} can be easily evaluated. Based on the equations above, Fig. 3.1 shows CloudSat random error as percentage fraction of the measured reflectivity for two different combinations of M and H_{sat}. It can be noted that the error rapidly decreases, becoming less than 20% for reflectivities larger than -20 dBZ. Before being assimilated, CloudSat observations are averaged to match the model resolution. The reduction in random noise that the averaging of observations brings is shown in the same figure (as dashed lines). For the evaluation, 30 samples are considered which correspond to the typical number of CloudSat bins falling into an IFS model grid-box.
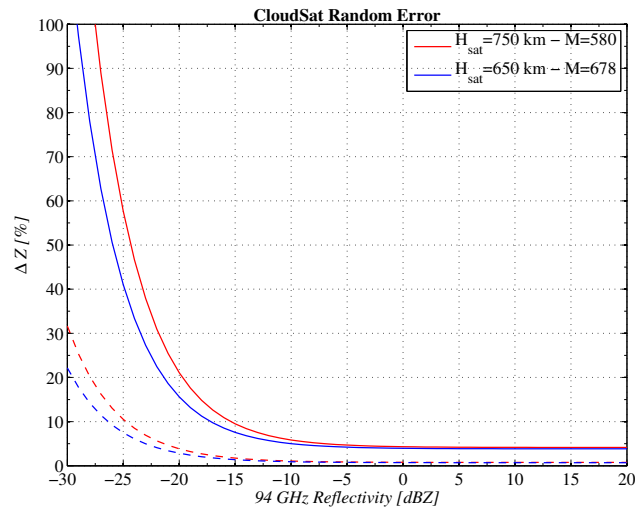

Figure 3.1: CloudSat instrument random error expressed as percentage fraction of the measured reflectivity (solid lines). Colour line indicates two possible combinations of satellite altitude and number of averaged samples. Dashed lines represent the error w 30 samples are averaged.

Forward modelling errors

As discussed in Section 2.3, a number of assumptions about particle properties are required in the process of modelling radar observations, e.g. their size distribution and shape/density (in case of frozen particles). The choices made may be generally valid, but they cannot be considered the most appropriate in every situation because of the large variability and uncertainty of cloud microphysics. For this reason, we can reasonably assume that the main source of uncertainties when simulating radar returns from clouds lay in the assumptions made to model the microphysical properties of hydrometeors. Our approach here is to perform, in a quantitative manner, an error analysis relating natural ambiguities in the above parameters to the uncertainties in the forward-modelled reflectivities. Considering the setup described in Subsection 2.4.1 as reference configuration, we have defined a number of perturbed states around it. The original properties (PSD or shape/density) of cloud liquid, cloud ice, stratiform rain and stratiform snow have been changed separately four times to possible alternatives (summarized in Table 3.1), resulting in a total of 16 perturbations. ZmVar has been then repetitively run on a dataset of several thousands of cloud/precipitation profiles obtained from the IFS, each time using one the configurations so obtained as new lookup-tables. Figure 3.2 shows (left and mid columns), the standard deviation of the differences between the perturbed and reference simulated reflectivity. Values are given as function of the reference reflectivity itself in terms of both linear (left panels) and dBZ value (mid panels). For better understanding, the statistics are built based on all 16 perturbations and separately on the 4 ones for each hydrometeor. The relative occurrence of reflectivity values (panels on the right column) helps to understand which ranges of standard deviation are more important. In addition, results have been stratified across six ranges of temperatures (panels from top to bottom). As expected, uncertainty in rain is the largest at temperatures above the freezing level, while below freezing snow and cloud ice dominate. At temperatures where both hydrometeors are present, the error due to cloud ice dominates over the one of snow when reflectivity values are below -10 dBZ. We note that the total standard deviation (black lines), being evaluated as squared sum of the

variances of each hydrometeor, is driven by the values of largest contributor. Interestingly, it assumes values that never exceed 3 dBZ, typically around 1 dBZ for Z larger than 0 dBZ.
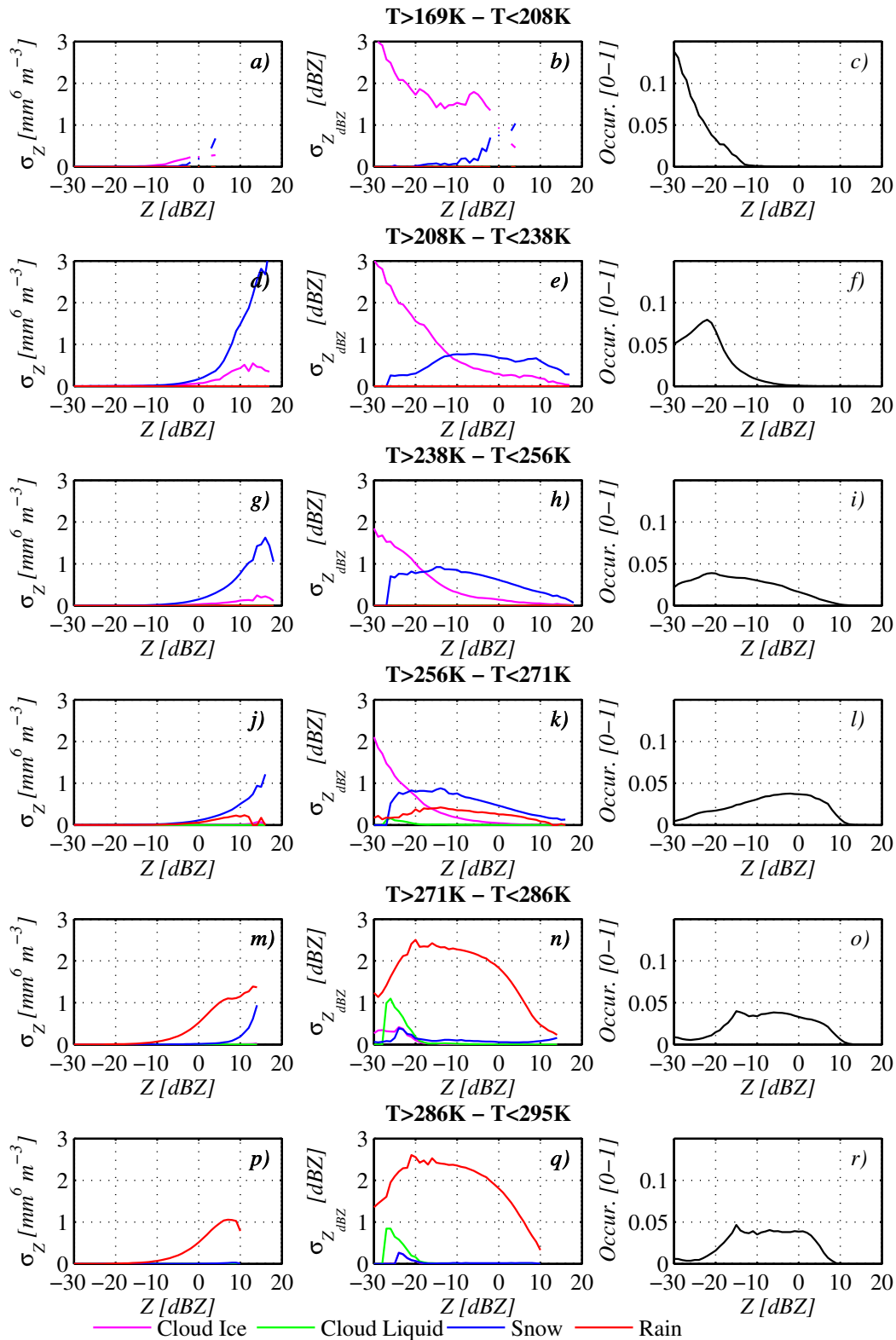

Figure 3.2: Forward modelling uncertainties in ZmVar simulated reflectivity expressed in $mm^6 m^{-3}$ and dBZ (left and central panels, respectively). Curves give the contribution given to uncertainty by each single hydrometeor. Right panels show the occurrence of reflectivity values. Each row refers to the temperature range shown in the title.

Hydro-meteor	PSD parameters	PSD perturbations	Particle shape	Shape Perturbations
Snow	$N(D) = \Phi_{23}(x)\dfrac{M_2^4}{M_3^3}$ $x = D\dfrac{M_2}{M_3}$ $M_n = A(n)M_2^{C(n)}e^{B(n)T_c}$	$\Phi_{23} = \begin{cases} 141.0\,e^{-16.8x} + 102.0\,x^{2.07}e^{-1.1*4.82x} \\ 141.0\,e^{-16.8x} + 102.0\,x^{2.07}e^{-1.05*4.82x} \\ 141.0\,e^{-16.8x} + 102.0\,x^{2.07}e^{-0.95*4.82x} \\ 141.0\,e^{-16.8x} + 102.0\,x^{2.07}e^{-0.9*4.82x} \end{cases}$	Aggregate of columns	-
Rain	0	0	Sphere	-
Cloud Liquid	0	0	Sphere	-
Cloud Ice	0	0	6-bullet rosette	• 4-bullet rosettes • hexagonal columns

Table 3.1: Summary of perturbations to the microphysical setup of ZmVar used to evaluate radar forward modelling uncertainties.

Figure 3.3 shows the total error standard deviation (sum of contributions from all hydrometeors), again separating over temperature ranges. For clarity, values are given as percentage of the corresponding simulated reflectivity. In the same plot, the instrument (native and averaged) CloudSat error noise (black lines) is given as reference. We note that the largest uncertainties (more than 70%) occur when temperature is larger than 0∘. As shown before, this is caused by the uncertainty in liquid precipitation. Instead, at temperatures below freezing monotonically the total error standard deviation decreases as reflectivity increases, with values below 30% already from -15 dBZ.
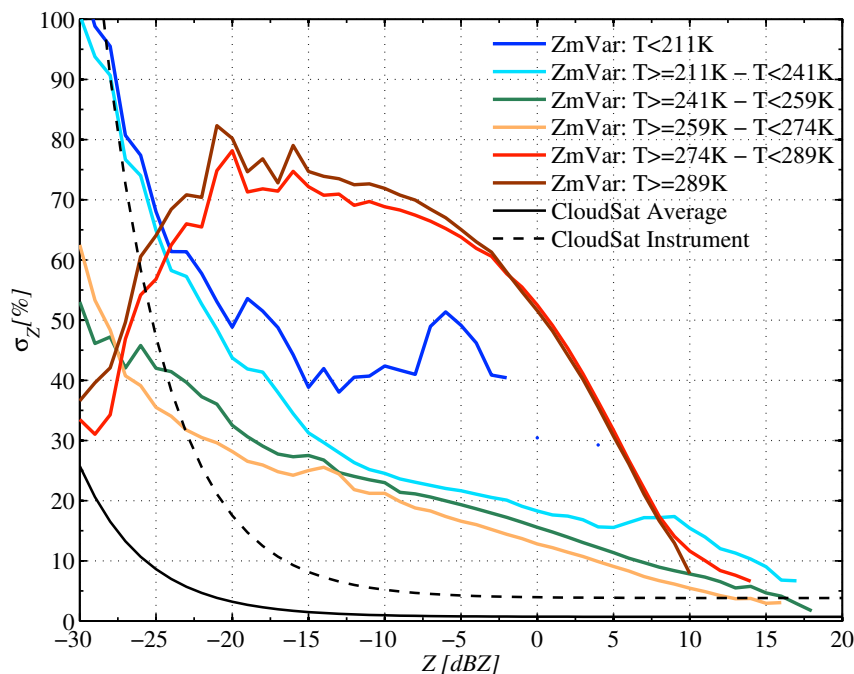

Figure 3.3: Summary of forward modelling uncertainties in ZmVar expressed as percentage of the simulated reflectivity when perturbations of all hydrometeors are taken into account. Each curve refers to the temperature range shown in the legend. For comparison, the CloudSat instrument random error is also given (black lines), considering measurements at the original resolution (dashed) and averaged over 30 samples (solid).

3.2 Representativity errors of radar observations

CloudSat radar observations provide unprecedented vertical views of clouds and precipitation at global scale. However, the lack of scanning capabilities and the rather narrow horizontal field-of-view result in a poor horizontal coverage. This raises the problem of representatitivity error when trying to compare these observations with simulated parameters retrieved by any NWP model (with much larger gridbox size). Based on the work done during the QuARL project (Di Michele et al., 2009, chap.3), the representativity of CloudSat radar observations is estimated. The representativity problem is briefly introduced in Section 3.2.1, while Section 3.2.2 summarizes the methodology used to estimate the systematic and random component of the representativity errors. In Section 3.2.3 an ad hoc measure is introduced, able to take into account flow-dependency. Section 3.2.4 presents the results of the approach for the narrow-FOV space-borne observations of CloudSat radar measurements. Finally, an example of representativity error for two selected scenes of CloudSat observations is shown in Section 3.2.5.

3.2.1 The representativity problem

As mentioned in the beginning of this section, the error due to representativity constitutes a component of the total observation error (besides forward modelling and instrument errors). Prior to any data assimilation process, it is necessary to have some information on the representativeness of these radar observations in order to know which importance and weight should be given to them. How "far" are they from the real state of the atmosphere? Are they a good representation of the "truth"? On the one hand, atmospheric fields, which exhibit an homogeneous and smooth variation with space and time, are likely to show little differences between observations and corresponding model outputs. On the other hand, great discrepancies can appear between model outputs and observations for variables which exhibit strong spatial variations on the sub-grid scale as it is typical for cloud and moisture related fields. In this case, the underlying representativeness of the observations can become a delicate issue.

3.2.2 Methodology

Due to the importance of the representativeness of the observations, a method to define representativity errors has been developed during the previous QuARL project (Di Michele et al., 2009; Stiller, 2010). It was tested and validated on scanning observations (TRMM rain rates and rain fractions as well as MODIS cloud fraction) which have a wider field of view and where "truth" (i.e. better knowledge of the state of the atmosphere) is known. The conclusions and recommendations in Stiller (2010) suggest to apply the method for space-borne lidar and radar observations such as CALIPSO and CloudSat. Thus this method has been adapted to CloudSat radar observations in order to derive an estimate for the representativity error (and bias) resulting from the small footprint of these measurements. An important aspect of these observations is that the estimation of representativity errors must be done at different altitude levels. The method is based on the use of a measure – the Structure Function Maximum (SFM) score – which is flexible enough to derive a flow-dependent error estimate since the magnitude of this error is strongly dependent on the respective weather type. As described in Di Michele et al. (2009), a fully statistical solution of the representativity problem is too complex to be practical and far beyond the scope of the project, and therefore priority was given to a more heuristic approach which includes the generation of synthetic data by stochastic modelling techniques. These techniques are widely used in geostatistics. The approach consists in generating a 2D synthetic dataset sharing some statistical properties of the 1D observations. The sampling error can then be estimated by applying the SFM score to the synthetic data. This score can be computed from the data along the satellite track and is used to differentiate regions according to the magnitude of their representativity error. More precisely, the approach taken is comprised of
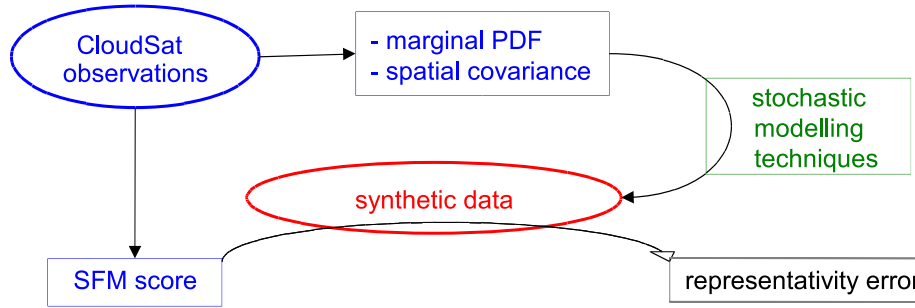
Figure 3.4: Schematic description of the main steps for the retrieval of the representativity error from CloudSat radar observations.

the following steps:

1. Choice of the Structure Function Maximum (SFM) score to discriminate situations whose representativity error is different according to statistical considerations. This statistical measure can be computed from the satellite track at any vertical level (see Sec.3.2.3).

2. Production of synthetic data (employing statistical modelling techniques) that share important statistical properties (marginal PDF and spatial covariance) with the observations of interest (See Sec.3.2.4).

3. Combination of the two previous steps resulting in the application of the SFM score to the synthetic data.

Figure 3.4 illustrates the main steps of the method when applied to CloudSat. From radar observations, the marginal PDF of reflectivity, the respective spatial covariance and the SFM score are retrieved. Through the use of stochastic modelling techniques, synthetic data that share the same statistical properties as the initial observations are generated. Finally, a quasi-empirical relationship is derived when applying the SFM score to the synthetic data in order to get an assessment of the representativity error. A more detailed description of the whole process of obtaining the representativity error for CloudSat observations is provided below.

3.2.3 Measuring flow-dependency with the Structure Function Maximum score

The approach developed in the QuARL project assesses the representativity error based on synthetic data generated with stochastic modelling techniques and relating regions from the satellite track to corresponding regions of the synthetic data by applying a measure. For a random field $q(\vec{x})$ (where $\vec{x}$ is the location), the structure function $\gamma\left(\vec{x}, \vec{h}\right)$ for a displacement vector $\vec{h}$ can be defined as follows:

$$\gamma\left(\vec{x}, \vec{h}\right) = 0.5 \, \mathbb{E}\left\{ \left[q\left(\vec{x} + \frac{\vec{h}}{2}\right) - q\left(\vec{x} - \frac{\vec{h}}{2}\right) \right]^2 \right\} \tag{3.5}$$

where $\mathbb{E}\{\ldots\}$ indicates the ensemble average.

The local estimator $\gamma(x_0, h, L)$ for $\gamma\left(\vec{h}\right)$ can be then computed along the satellite track:

$$\gamma(x_0, h, L) = 0.5 \left\langle \left[q\left(x + \frac{h}{2}\right) - q\left(x - \frac{h}{2}\right) \right]^2 \right\rangle_{x \in I[x_0, L]} \tag{3.6}$$
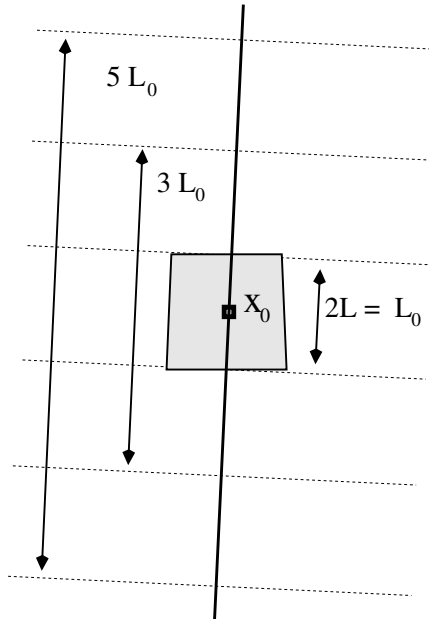
Figure 3.5: Schematic representation of the satellite track (thick line) intersecting a grid box whose centre is x_0 and length is $L_0 = 2L$. Averagings are done with each fixed neighbourhood of rank "k" (with $k = \{0, 1, \ldots, N_{max}\}$). The three closest neighbourhoods are represented here.

by taking the spatial average $\langle \ldots \rangle_{x \in I[x_0, L]}$ over the neighbourhood $I[x_0, L] = [(x_0 - L), (x_0 + L)]$ of size $2L$ around x_0 (x_0 corresponds to the horizontal position of the centre point of the intersect between the satellite track and the model grid box of interest). The corresponding measure γ gives information about the variability at each position x_0. It depends on two length scales h and L. The displacement length h was chosen to be half of the length of a gridbox. Tests with different h tend to indicate that the method is not very sensitive to the particular choice of this parameter (Stiller, 2010).

To construct an error discriminator which comprises information from a larger range of scales, the structure function estimate (Eq. 3.6) was computed for different neighbourhoods $I[x_0, L]$ (i.e. for different values of L) and then the maximum of all the obtained values was taken. More precisely, neighbourhoods $L = k L_0$ (with $k = \{0, 1, \ldots, N_{max}\}$) have been considered, where L_0 is the length of the NWP grid box and N_{max} is the maximal number of considered gridboxes. Using this, the Structure Function Maximum score has been defined as:

$$SFM(x_0, h, L_0) = \max \left[\gamma(x_0, h, (k + 0.5) L_0) \right]_{k = \{0, 1, \ldots, N_{max}\}} \tag{3.7}$$

which turned out to be not too sensitive with respect to N_{max} as long as this number was large enough ($N_{max} = 10$ has been chosen). It should be noted that large score values are usually obtained only from the closer neighbourhoods (i.e. small values of k) while contributions from further neighbourhoods are usually small. They only become relevant enough in Eq.3.7 if the contributions from closer neighbourhoods happen to be small. This reflects the fact that strong variations closer to a gridbox have a larger impact on the representativity error (in a probabilistic sense) than perturbations which are further away. Fig. 3.5 illustrates the different length scales where the SFM calculation is computed.

3.2.4   Generation of CloudSat synthetic data by stochastic modelling

An important part of our statistical approach to estimate the representativity error consists in the generation of 2D synthetic-data that have the same statistical properties as the 1D observations, i.e. (almost) the same marginal probability density function (PDF) and the same covariance function as the satellite observations (it is assumed that the statistical properties are isotropic). The stochastic modelling method used to produce such synthetic data is explained in Di Michele et al. (2009, chap. 3.3). The generation of these synthetic data is made in several steps whereby a major part is to project the non-Gaussian observations (the related cloud and moisture observed fields) onto a Gaussian field, which is then modelled with the help of the method developed in Pardo-Igúzquica and Chica-Olmo (1993, 1994). It is a spectral method which generates the Fourier coefficients of the Gaussian field. More precisely, the method uses the fact that the Fourier transform of a covariance function is the power spectrum, which constrains the Fourier coefficients and leaves only their complex phase to be determined (for Gaussian random functions, the complex phase is homogeneously distributed and can therefore easily be computed by a random number generator). An important condition for generating synthetic data is that the employed covariance function is reasonably smooth. It should be kept in mind that covariances calculated from observations are always a statistical estimate and the convergence at longer distances can become problematic. While the power spectrum from true covariance functions is always positively defined, this property is generally not satisfied for estimated covariances. As positively definite power spectra are essential for generating the synthetic data, the covariances obtained from observations are replaced by a smooth analytic function fitted to the observed covariances. The observed correlation functions are smoothed and fitted by an "optimal" expression, which is composed by the sum of two exponential terms, each preserving the shape of the observed correlation function (one for the short range distance, the other for the long range distance). By this way, the obtained fitted correlation function curves are forced to become 0 in the far long range (fixed to 4000 points) in order to retrieve positive Fourier coefficients, as the definition of the power spectrum requires.

In Di Michele et al. (2009), the validity of the method to retrieve a flow-dependent estimate of the sampling error was tested with 2D observations (MODIS total cloud cover, TRMM rain index and TRMM rain rate) taken from a period of 5 days in the tropical Pacific. In this work, we will use quasi 1D data from CloudSat radar taken from two periods of 1 month (January and July 2007) over the tropical belt ($\pm 20\,°$) and extra-tropics ($30 - 50\,°$N). Since active remote sensing instruments provide the vertical structure of atmospheric components like clouds, the sampling error will be estimated also with respect to altitude. To estimate the sampling error, the CloudSat reflectivity at 94 GHz (GEOPROF_2B product) Z is linearized by changing its units, i.e. using the following transformation:

$$Z \quad \longrightarrow \quad 10^{Z/10}$$
$$[dBZ] \qquad\quad [mm^6 m^{-3}]$$

This transformed reflectivity will be the quantity considered. Some filtering is applied to the original data in order to keep only data between -25 dBZ and 12 dBZ. Note that the minimal detectable value is -27 dBZ for CloudSat radar (Tanelli et al., 2008). The considered parameter therefore ranges then from 0 to $10^{12/10} = 15.85\ mm^6 m^{-3}$. Quality and cloud masks are also taken into account.

Figures 3.6 and 3.7 respectively show the marginal PDF (in logarithmic scale) and the correlation function of reflectivities (in $mm^6 m^{-3}$) coming from CloudSat radar (black) and generated synthetic data (red) in the Tropics ($\pm 20\,°$, left panel) and mid-latitudes North ($30 - 50\,°$N, right panel) at an altitude of 2 km for the period of January 2007. For both geographical areas, these statistical properties of CloudSat reflectivities and generated synthetic reflectivities are in good agreement. The PDF of reflectivity is peaked in the first bin of reflectivity (i.e. small reflectivities) and then the number of samples rapidly decreases as the reflectivity increases. The correlation function shows quite a perfect agreement between observed and synthetic data in the short range distance (below 500 points). It can be noticed that the measured covariance functions become quite noisy (and even negative) for longer distances (beyond 700 points) because of insufficient convergence. Thus for the
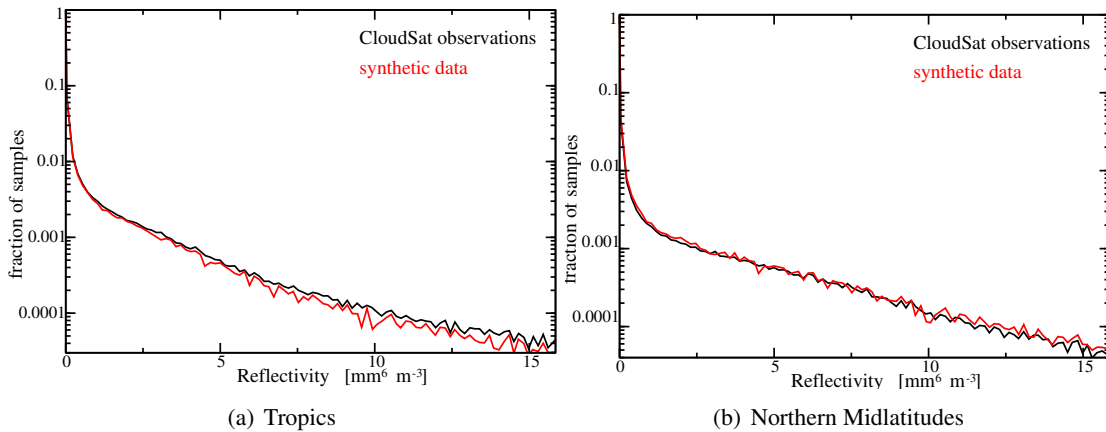
Figure 3.6: *PDF of observed (black) and generated synthetic (red) radar reflectivity in the Tropics ($\pm 20\,^\circ$, left) and in mid-latitudes North ($30-50\,^\circ$N, right), at an altitude of 2 km for the period of July 2007.*
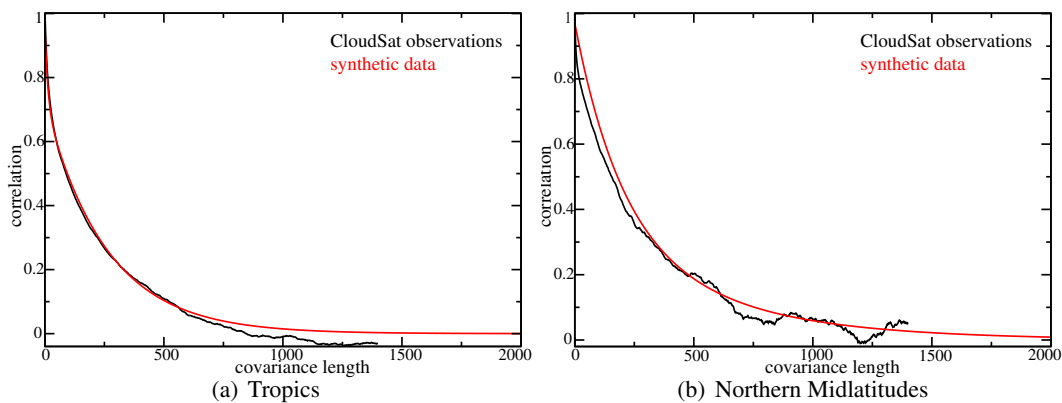

Figure 3.7: *Correlation function of the observed (black) and synthetic (red) radar reflectivity in the Tropics ($\pm 20\,^\circ$, left) and in mid-latitudes North ($30-50\,^\circ$N, right) at an altitude of 2 km for the period of July 2007.*

proper generation of the synthetic data it is essential to use a smooth fitted covariance function curve which decays exponentially to zero in the very far range (Stiller, 2010). Similar plots (not shown) have been computed at other altitudes and similar features between the statistical properties of observed and synthetic reflectivities can be revealed.

In Di Michele et al. (2009) it was shown that the SFM score is a useful tool for differentiating regions according to their error magnitude. As alternative choice, the measured quantity (i.e. reflectivity) itself could also be used to explain the flow dependence of the representativity error. However, the positive aspect of the SFM score is that it uses information far beyond the gridbox where the field is being evaluated.

Figure 3.8 displays the PDF (in logarithmic scale) of the Structure Function Maximum (SFM) score (in $mm^{12}\ m^{-6}$) in the Tropics (left) and in mid-latitudes North (right) at an altitude of 2 km for July 2007 as before. Again, there is very good agreement between the two datasets for the three smallest bins (SFM score $< 2.5\ mm^{12}\ m^{-6}$ which represent about $\sim$85% of the points). For higher scores, discrepancies appear in the shape of the distributions. We note that the slope of the distribution in tropical areas is less steep than the one at mid-latitudes, i.e. the proportion of small SFM scores (i.e. homogeneous regions) is relatively higher in mid-latitudes. Largest SFM scores of observed and synthetic reflectivities were around 180 $mm^{12}\ m^{-6}$ for the observation period. However, values above 50 $mm^{12}\ m^{-6}$ were relatively rare since $\sim$99% of the samples are found to be smaller than 50 $mm^{12}\ m^{-6}$.
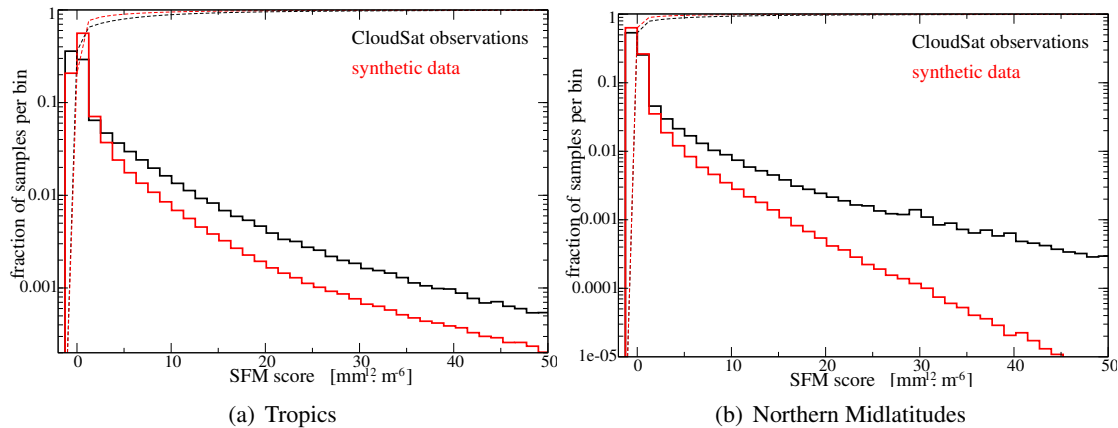
Figure 3.8: PDF of the Structure Function Maximum (SFM) score (in $mm^{12}m^{-6}$) for observed (black) and synthetic (red) radar reflectivity data in the Tropics ($\pm 20°$, left) and in mid-latitudes North ($30-50°N$, right) at an altitude of 2 km for July 2007. The dashed lines represent the cumulative sum of the distribution of the SFM for observed (dashed black) and synthetic (dashed red) data.

As a conclusion, the comparison of the statistical properties of observed and synthetic reflectivities reveals similar features and validates the fact that the synthetic data can be used for the estimate of the sampling error by mean of the SFM score.

3.2.5 Estimating the representativity errors

To assess the sampling error of CloudSat radar reflectivity measurements, synthetic analogue data have been generated by the method described previously while using the statistics retrieved for the two one-month periods (January and July 2007). Averages of the observed reflectivity are computed on line segments. At the same time, averages of the synthetic reflectivities have been computed in the corresponding co-located gridboxes. The line segments are then partitioned into bins of the SFM score, the Root-Mean Square Error (RMSE) is computed as the RMS difference between line and grid box averages of the considered reflectivity in each bin. In this way, the relationship between the SFM score and the sampling error is retrieved.

CloudSat radar measurements give profiles of reflectivity along the satellite track from the surface up to an altitude of 20 km. The relationship between the sampling error and the SFM score is therefore altitude-dependent. This dependence with height is presented in Fig.3.9 for the Tropics and in Fig.3.10 for North mid-latitudes at four different altitudes (2, 6, 10 and 14 km) for July 2007. In the Tropics (Fig.3.9), the RMSE (black curve) increases linearly from 0 to up to 1.5 mm^6 m^{-3} for all altitudes, showing that the employed SFM score is quite effective to discriminate regions with small and large errors. As shown in Fig.3.8, the majority of samples at 2 km ($\sim$99%) correspond to scores below 50 mm^{12} m^{-6} and therefore only scores less than this value have been plotted in Fig.3.9 and Fig.3.10. It could be shown that above 50 mm^{12} m^{-6} the RMSE keeps around $\sim$1.7$\pm$0.2 mm^6 m^{-3} for scores up to 100 mm^{12} m^{-6} and then it becomes very noisy. At North mid-latitudes (Fig.3.10), the statistics appear to be quite similar for altitudes up to 10 km where the RMSE increases slightly up to 1.5-1.7 mm^6 m^{-3}. However, the shape is not necessarily quasi-linear as in the Tropics (up to 6 km). In fact, the RMSE at 2 km varies strongly for scores in the range [0-15] mm^{12} m^{-6} and then small variations occur for higher scores. At an altitude of 10 km, the RMSE shows a small increase for scores smaller than 15 mm^{12} m^{-6}, a linear increase for scores between 15 and 40 mm^{12} m^{-6}, and it stays constant for scores larger than 40 mm^{12} m^{-6}. Finally, error statistics at 14 km become very noisy for SFM values larger than 20 mm^{12} m^{-6}. This is consequence of the few samples at this altitude, because in extra-tropics clear-sky conditions are most likely to occur. For meteorological applications, it is useful to partition the RMSE into bias and variance.

These quantities are plotted in both Fig.3.9 and Fig.3.10. In both figures, the variance follows the shape of the RMSE along the considered range of SFM. In the same figures, the bias slightly decreases almost linearly with the SFM score, but stays quite small (up to -0.5 mm^6 m^{-3}).
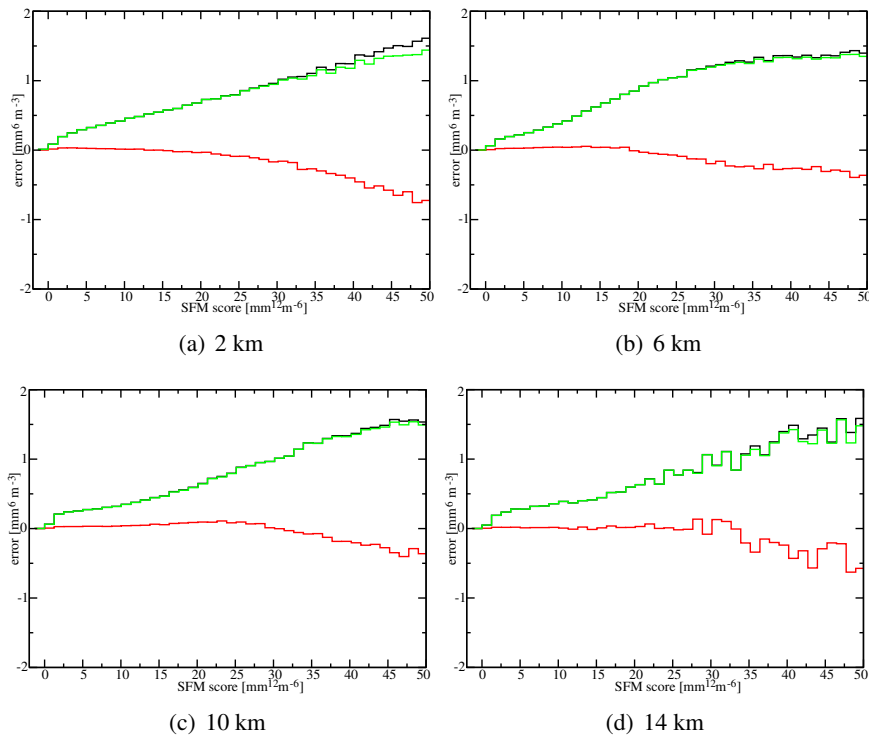

(a) 2 km

(b) 6 km

(c) 10 km

(d) 14 km

Figure 3.9: RMSE (black), bias (red) and variance (green) of CloudSat synthetic data in SFM score bins (in mm^{12}m^{-6}) at an altitude of 2km (a), 6km (b), 10km (c) and 14km (d) for July 2007 in the Tropics ($\pm 20\,^\circ$).

(a) 2 km

(b) 6 km

(c) 10 km

(d) 14 km
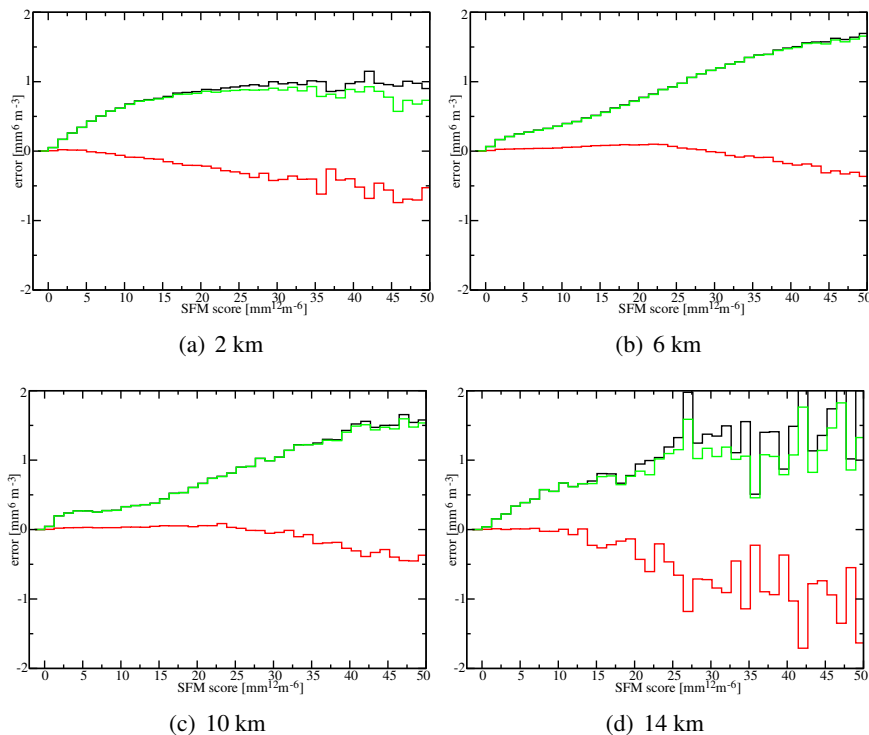
Figure 3.10: Same as Fig.3.9 but for mid-latitudes North (30-50$\,^\circ$).

3.2.6 Application to a case study

Based on the relationships between score and the developed flow-dependent sampling error, it is possible to estimate the representativeness of CloudSat radar measurements. Figures 3.11 and 3.12 show an application of these relationships for a portion of CloudSat observations in the Tropics and in North mid-latitudes respectively (granule 03941 on 23 January 2007). In both figures, the top panel shows CloudSat reflectivity (in dBZ) while the middle panel gives the corresponding simulations using ZmVar. First, the SFM score is computed at different altitudes (2, 6 and 10 km) are computed. Then, using the relationships presented in Fig.3.9 and 3.10, the calculated SFM score are linked to a sampling error. Results are shown in the bottom panels of Figures 3.11 and 3.12 (in mm^6 m^{-3}) for these 3 different altitudes along the track of the satellite. The tropical case study exhibits 3-4 convective cloud towers reaching an altitude of 15 km, while the the mid-latitudes contains a frontal system with cloud tops up to 12 km. Both atmospheric fields are reasonably well simulated by the ZmVar (middle panels): the general patterns are captured, while the intensity or location can slightly differ. The magnitude of the error can be connected to the atmospheric process appearing at the location of observations and simulated by the model: in the frontal system, the sampling error can peak up to 1.5 mm^6 m^{-3} while it can reach up to 4 mm^6 m^{-3} in the case of the convective tropical towers. This result can be linked to the difficulty of the model to simulate the intense convective events because of the limited horizontal resolution.

3.3 Conclusions

Observation errors associated to CloudSat reflectivity measurements have been evaluated considering the different components, i.e. the instrument, forward modelling and spatial representativity. The CloudSat instrument error has been evaluated based on technological considerations. A quantitative estimate of the forward modelling errors has been performed based on the uncertainties in the definition of some main microphysical parameters. The representativity error has been assessed by applying the statistical method developed during the QuARL project to CloudSat data. The method gives a flow-dependent estimate of the sampling error generating synthetic data that share the main important statistical properties (marginal PDF and correlation function) with the observations.
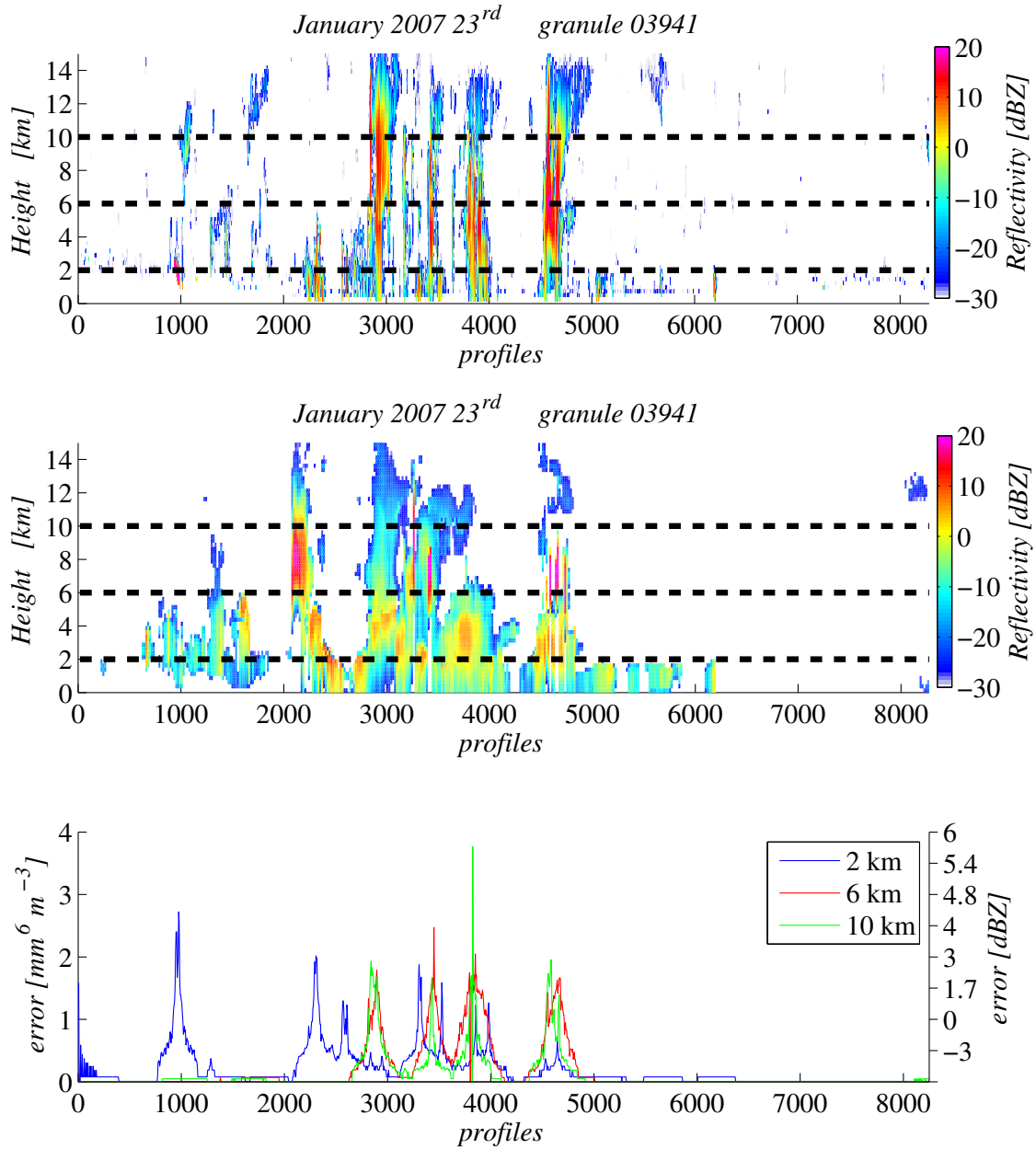
Figure 3.11: Top panel: CloudSat radar reflectivity (in dBZ) observed in the Tropics ($\pm 20°$, left) on 23 January 2007 (granule 03941). Middle panel: corresponding simulated radar reflectivity (in dBZ) retrieved by ZmVar. Low panel: corresponding estimate of the representativity errors of CloudSat reflectivity (in mm^6 m^{-3}) at an altitude of 2 km (blue), 6 km (red) and 10 km (green) represented by horizontal black dashed lines in the top and middle panels.
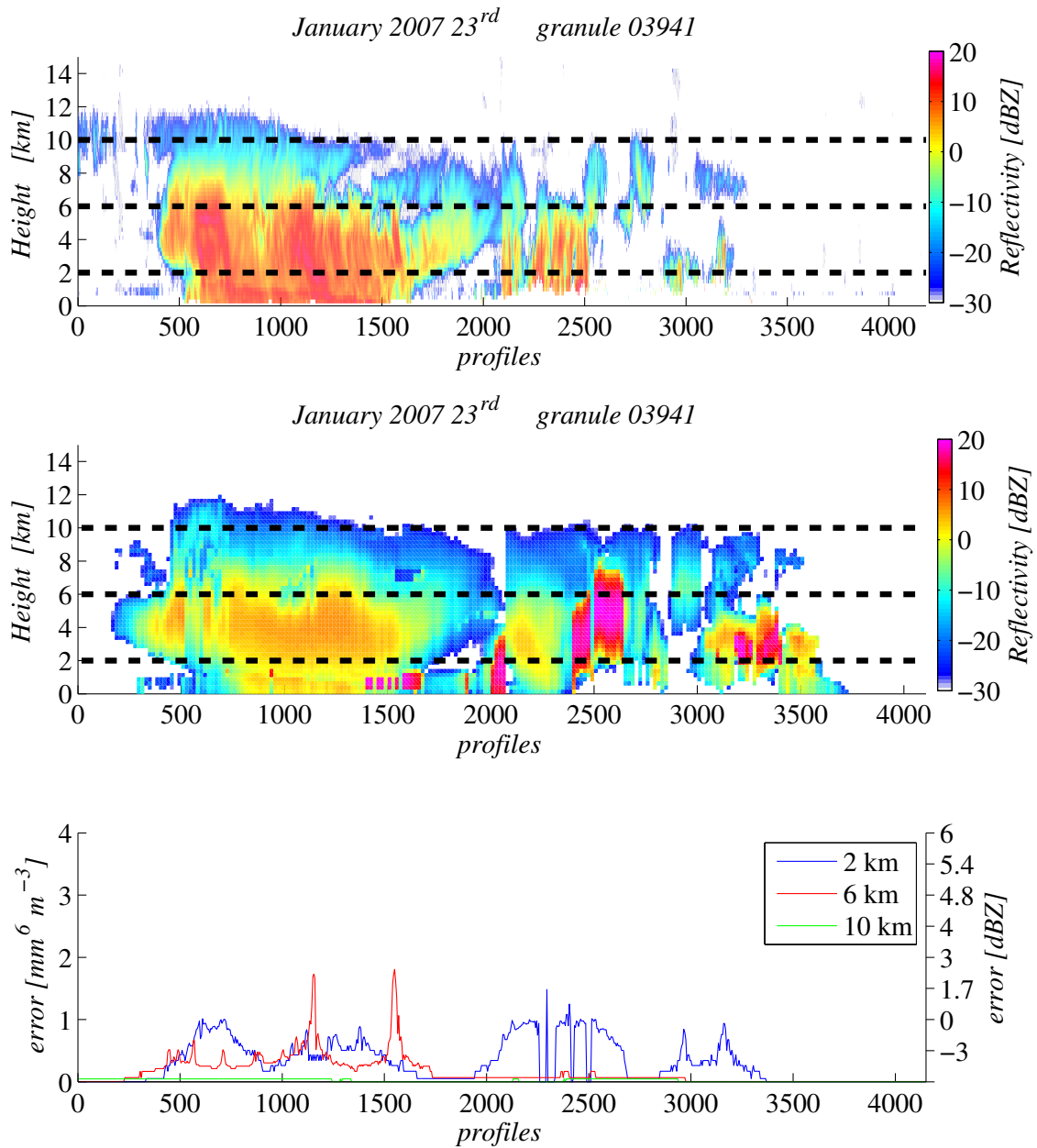
Figure 3.12: Same as in Fig.3.11 but for a portion of the same CloudSat granule in mid-latitudes North (30 ° − 50 °).

4 Development of quality control and bias correction schemes for the radar

Most data assimilation systems require that observations satisfy the following properties:

- Gaussian errors
- Linearity
- Absence of biases

The first two conditions can be achieved by applying a quality control. For the third, a bias correction scheme that guarantees zero-mean statistics for the first-guess (FG) departures is needed. In the following two sections, the techniques of quality control (Section 4.1) and bias correction (Section 4.2) developed for CloudSat observations will be described.

4.1 Quality control of CloudSat observations

An important aspect of any data monitoring and assimilation is the definition of appropriate criteria for quality control of observations. These preliminary checks are necessary in order to:

- Discard measurements flagged as of poor quality or leading to excessive FG departures, i.e. very large differences between observed and corresponding simulated parameters.
- Avoid the assimilation system to perform sub-optimally in cases of large model uncertainties and/or excessive non-linear relationship between observations and corresponding simulated parameters.

A comprehensive evaluation of the agreement between CloudSat and model simulated reflectivities can be done in terms of scatterplots, as shown in Fig. 4.1 and Fig. 4.2 considering cases over ocean during 30 days in January 2007, at Southern mid-latitudes ($30°S$-$60°S$) and in the tropics ($30°S$-$30°N$), respectively. The comparison is done separating observations into six altitude ranges. Above 10 km (panel a), we note that simulations corresponding to CloudSat reflectivities below -20 dBZ (where most the of cases lie) are often too large. On the contrary, above -20 dBZ, simulations underestimate the observations. This feature is common to tropics and mid-latitudes. The underestimation above -20 dBZ becomes more evident between 8 and 10 km (panel b), where the range of the observed reflectivities becomes larger. We tend to believe that this is due to the IFS model rather that to ZmVar forward operator. As a matter of fact, experiments demonstrated that the agreement does not improve changing the parameterization of the ice optical properties in ZmVar (not shown). Between 6 km and 8 km (panel c) the reflectivity range extends to higher values (up to 10 dBZ), indicating the presence of higher ice content and/or larger particles. Above -10 dBZ, we note that there is a good correlation between observations and simulations. However, for lower values, the underestimation/overestimation behaviour seen at higher altitudes is still present. The scatter between simulated and CloudSat reflectivities between 4 km and 6 km (panel d) shows a much better agreement. The underestimation/overestimation feature disappears, while the correlation increases as reflectivity increase without any evident bias. The same behaviour can be seen for observations between 2 km and 4 km (panel e) where there is a good agreement in the majority of cases, especially above 0 dBZ. Below 2 km (panel f), the most evident feature is the systematic overestimation of the lowest reflectivity values. This feature is a consequence of the forecast model producing light rain (drizzle) in cases where only (non-precipitating) cloud liquid is actually observed. The much larger radar signature of rain compared to the one of cloud liquid leads to the evident disagreement in the scatterplot. Figure 4.2 shows that the agreement at levels between 2 km and 8 km in the tropics (panels c, d, e) is considerably worse, while above 10 km and below 2 km plots are similar to the mid-latitudes case. Overall, these scatterplots show that there are cases where simulated reflectivities are too apart from the corresponding observations to be used to build FG departure statistics. A screening procedure for these cases must be therefore defined to retain only the ones with a satisfactory agreement.

The quality control is based on the FG departures, i.e. the difference between CloudSat radar observations and corresponding simulated reflectivities generated by the ZmVar radar forward operator using short-term forecast
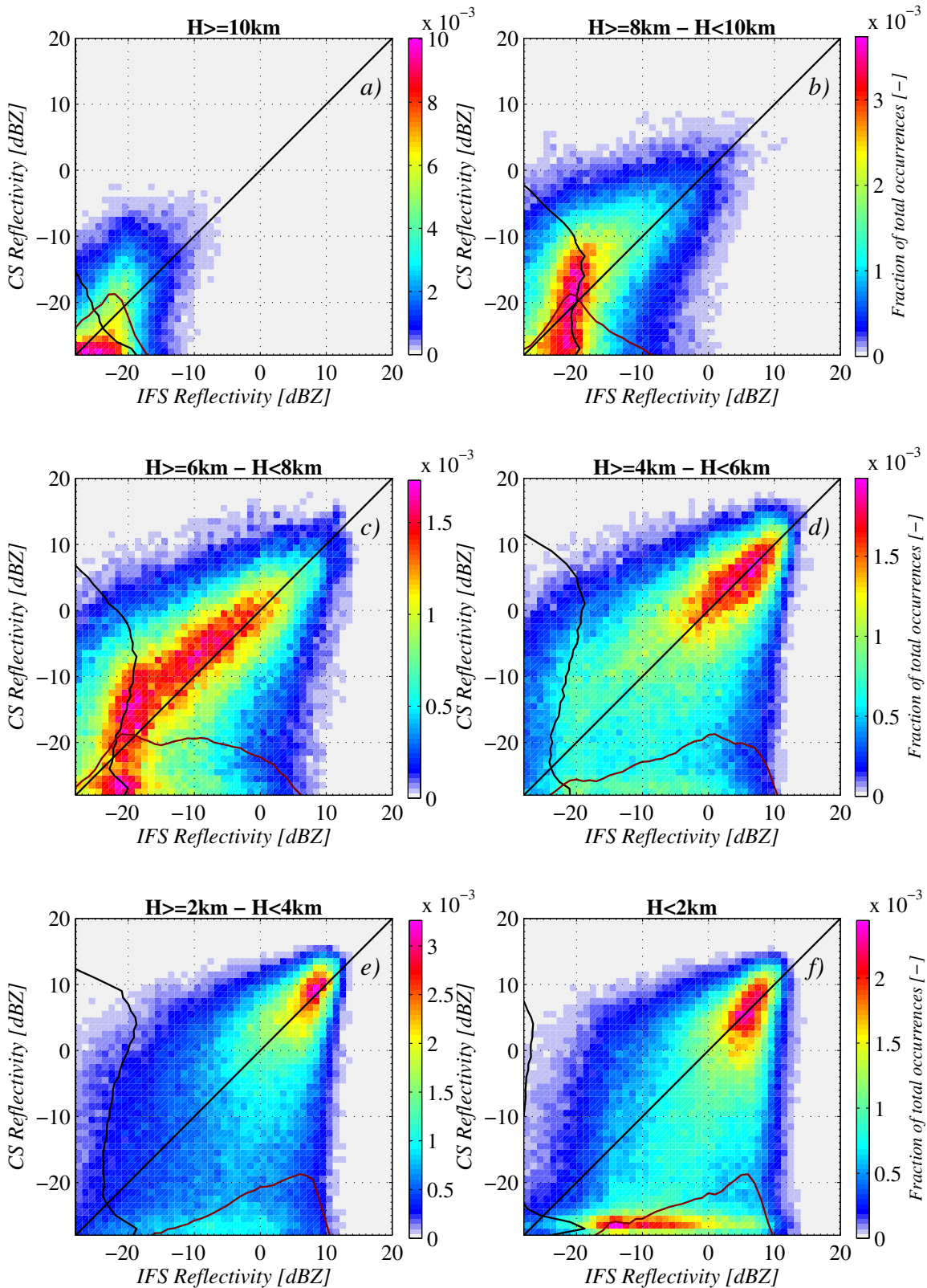
Figure 4.1: Scatter plot between simulated (on abscissa) and CloudSat reflectivity (on ordinate) using observations over ocean matched with the model data for the period of January 2007 between 30°S and 60°S. Curves along axes show the relative occurrence of reflectivity values. Each plot refers to the altitude level indicated in the panel title.
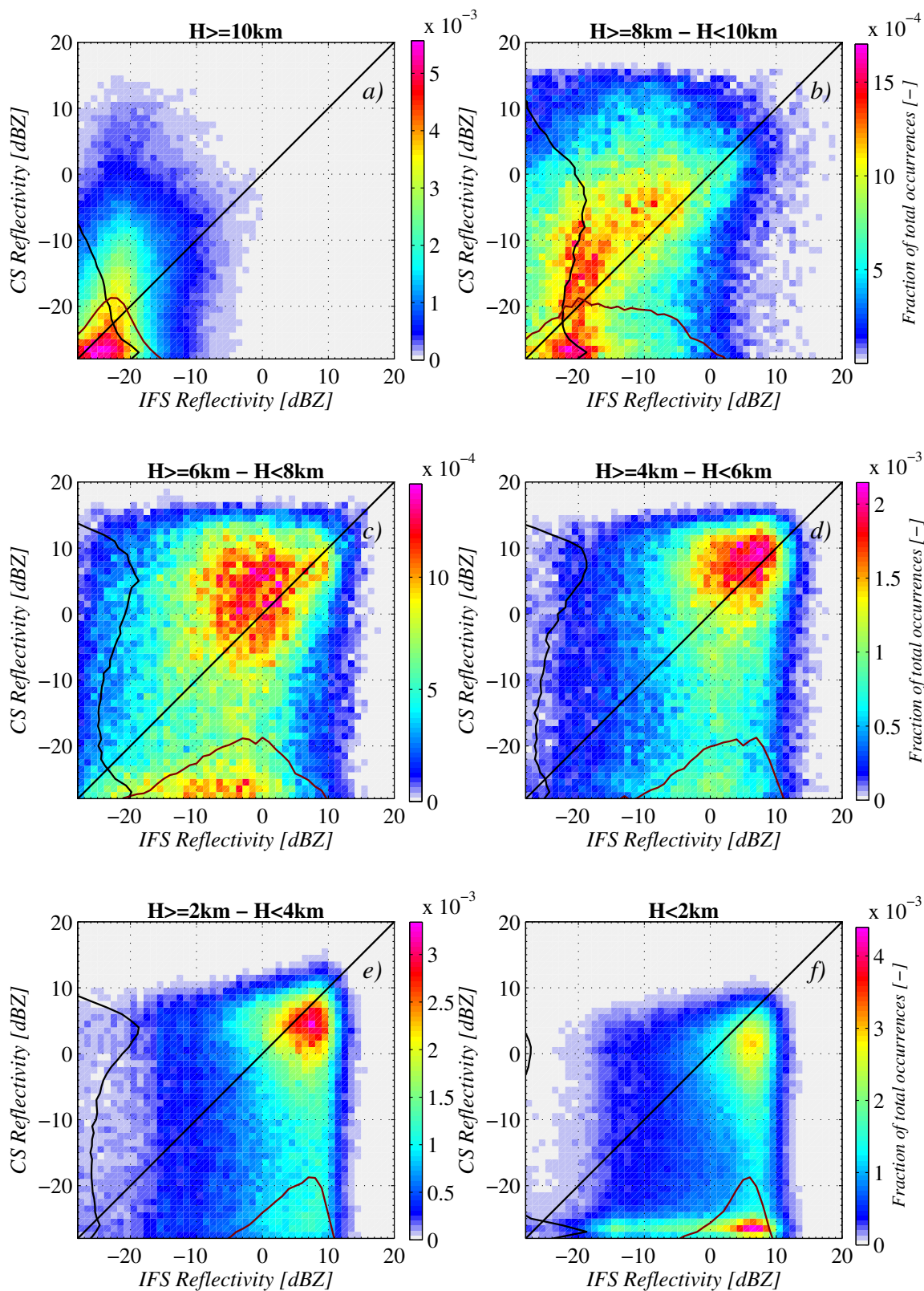
Figure 4.2: Same as Fig. 4.1, but considering observations in the tropics (30°S-30°N).

(background) produced by the ECMWF model. CloudSat reflectivities are made comparable to the background ones averaging them over the model gridbox. Our approach consists in discarding cases where FG departures are larger than 20 dBZ. This threshold has been chosen testing different values in the context of a 1D-Var analysis. The result is a trade off between the needs of retaining as many as possible observations along each cloud profile and at the same time guaranteeing that the FG values are close enough to be able to achieve a successful minimization. As an example, Fig. 4.3 shows a cloud system observed by CloudSat on 1 January 2007 (panel a) and the corresponding reflectivity FG from ZmVar and the ECMWF short-term forecast (panel b). Observations passing and failing the quality control described above are respectively indicated in black and red colours in panel c) of the same figure. By comparison, we notice that the screened out cases represent only a small portion, and that they mostly coincide with the transition regions where only CloudSat or the model gives a reflectivity value.
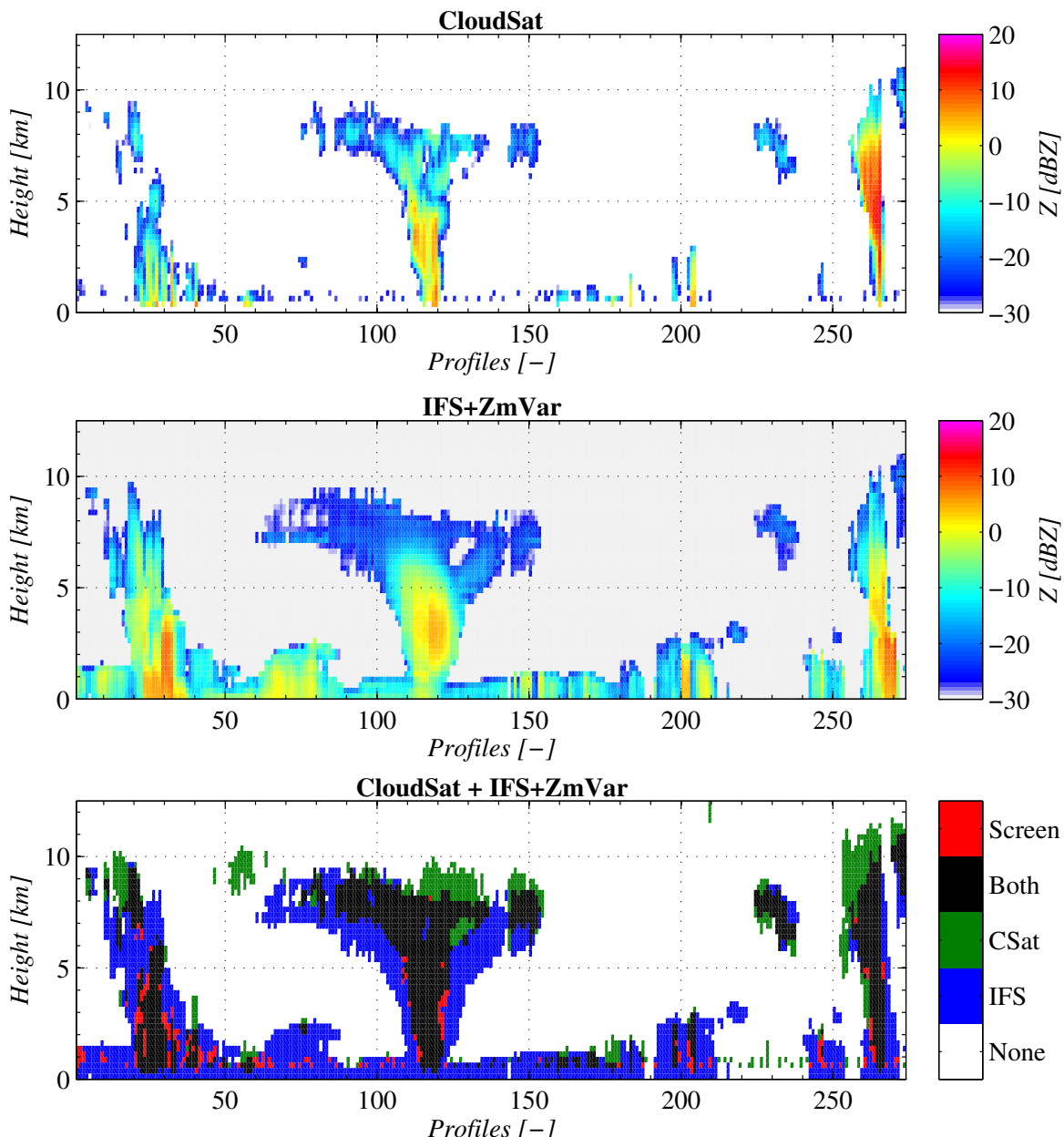

Figure 4.3: *Reflectivity of a cloud system as observed by CloudSat on 1 January 2007. CloudSat observations averaged to the IFS model grid resolution are in panel a). Panel b) displays the corresponding simulated reflectivities from ZmVar and the ECMWF model short-term forecasts. The contingency mask in panel c) highlights (in red) the observations screened out by the quality control.*

The impact of the quality control screening on FG departure statistics is shown in Fig. 4.4 and Fig. 4.5, which respectively consider cases at mid latitudes South (between $30°S$ and $60°S$) and in the tropics ($30°S$-$30°N$). These figures contain six panels, each referring to a particular altitude intervals. Each panel shows histograms of reflectivity FG departures before (blue curves) and after quality control (red curves). Removing the largest FG departures, although does not change the shape of the PDF, slightly reduces the tails and give a more symmetric distribution. The bell-shaped curves show zero-mean Gaussian distributions having the same standard deviation of the reflectivity FG departures. They represent the 'ideal' case and are plotted as a reference. By comparison, the PDFs of FG departures exhibit a shape which is hardly comparable to the Gaussian distribution since there is a prominent peak around zero and the tails are much pronounced. The peculiar shape of that PDF of reflectivity FG departures is consequence of the fact that radar reflectivity can assume values over a range of several order of magnitudes.

Two alternatives to get PDF of the FG departures with a more Gaussian shape have laso been tested. In the first approach, the distribution of reflectivity FG departures is scaled for the observed reflectivity, while in the second approach it is scaled for the range of observed and background reflectivities. Though this second method gives better results, neither of them provides a perfect Gaussian PDFs for the FG departures. In the future, those complex approaches may be revisited, but for monitoring and assimilation studies for this project, a simpler procedure for the quality control screening described above is used.
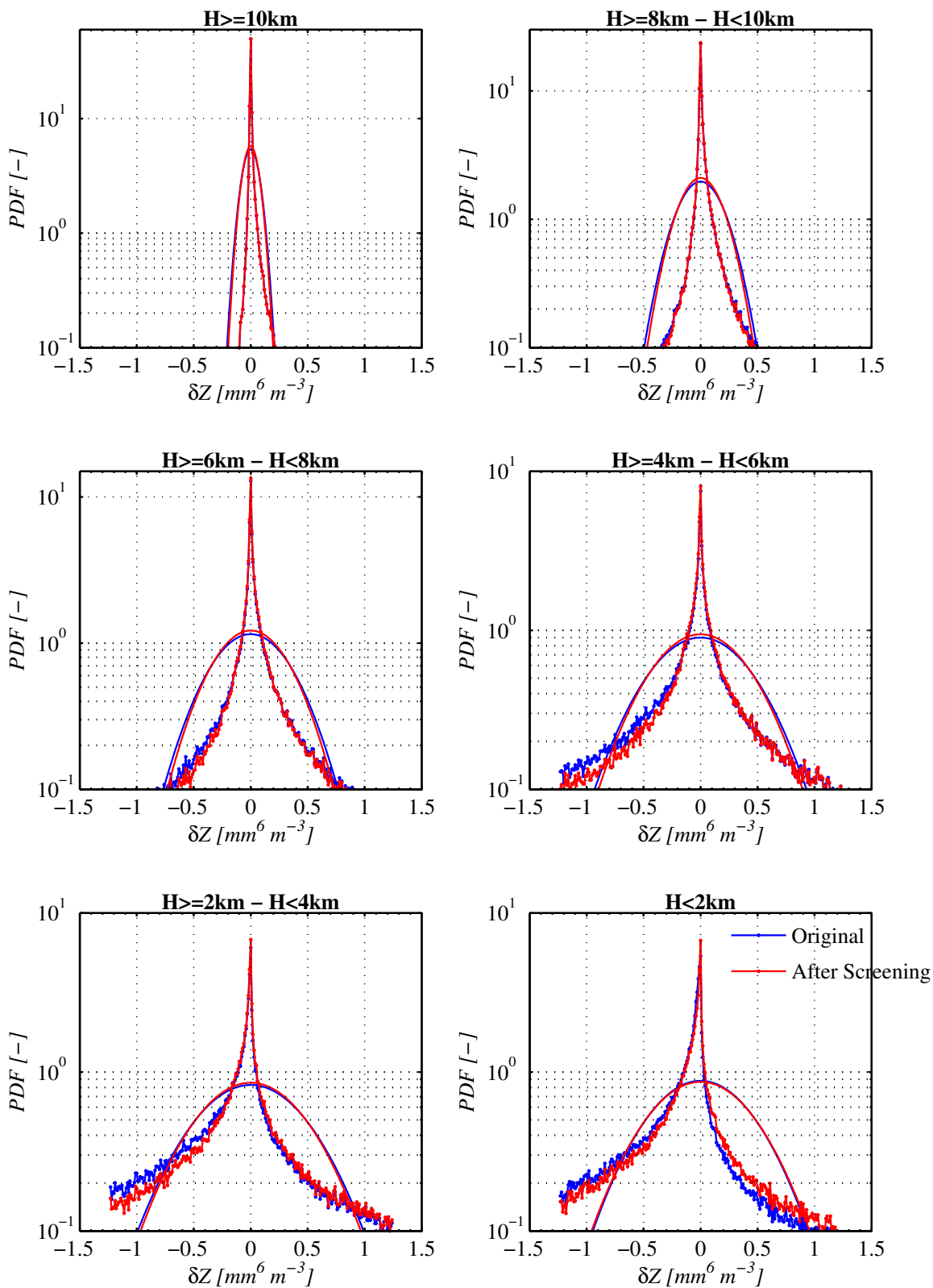
Figure 4.4: PDF of reflectivity FG departures ($mm^6 m^{-3}$) using observations and model data over ocean at mid-latitudes South ($30°S$ - $60°S$) in January 2007. Each plot refers to observations in the altitude range indicated in the panel title. Blue and red curves respectively indicate all cases and to the ones passing the screening. The bell-shaped curves represent a zero-mean Gaussian PDF with the same standard deviation.
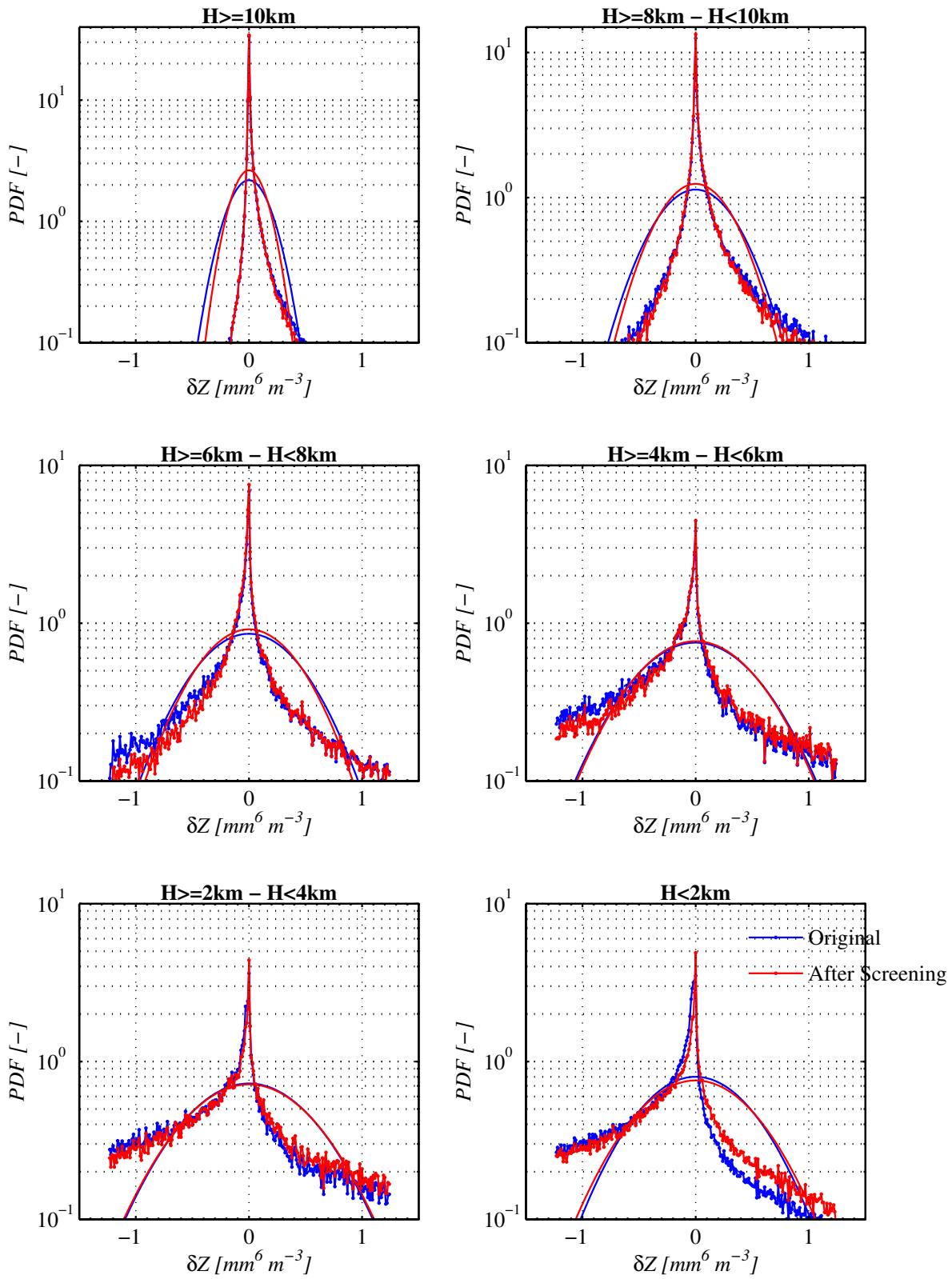
Figure 4.5: *Same as Fig. 4.3, but considering observations in the tropics (30°S-30°N).*

4.2 Bias correction of CloudSat observations

Data assimilation systems are generally affected by biases. They can arise from various reasons like problems with the data, approximations in the observation forward operators, limitations or methodology of the assimilating system. Since assimilation relies on the assumption that both observations and model background are unbiased quantities, it is necessary to quantify these systematic errors and to remove them. This is achieved by applying a bias correction scheme which is able to guarantee zero-mean statistics. The monitoring of observed-minus-background residuals shows evidence of systematic errors in the model or/and the observations. Our goal is to correct the bias in the simulated reflectivity that is generated by the radar forward operator. Indicating these biases with $\mathbf{b}(\mathbf{x})$, Eq.3.1 after bias correction becomes:

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}) - \mathbf{b}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}) - \mathbf{b}(\mathbf{x})) \qquad (4.1)$$

where a term of bias correction is introduced into the observation part of the cost function. In order to build a bias correction scheme, it is important to know and characterize the parameters on which the bias is dependent. Biases $\mathbf{b}(\mathbf{x})$ are estimated using a statistical relationship between predicted reflectivity FG departure and the observed reflectivity. However, the quality control performed for CloudSat radar reflectivity observations (Sec.4.1) revealed a large negative spread which depends on the temperature, i.e. on the water phase and on the size of cloud/rain droplets or ice crystals (weather regimes). For this reason, besides calculating the bias $\mathbf{b}(\mathbf{x})$ as a function of the reflectivity, it is also necessary to consider a variation with the temperature.

Quality-controlled CloudSat reflectivities for a period of one month (January 2007) have been used to get statistics of observation-minus-background departures (i.e. reflectivity FG departures) and to estimate the temperature-dependent bias. Figures 4.6 and 4.7 show the PDF of reflectivity FG departures, expressed in $mm^6 m^{-3}$, considering data over ocean at mid latitudes South and tropics, respectively. Only cases that have passed the step of quality control are taken into account. Curves in different panels separate the results according to the altitude and temperature of the observations. As shown in Fig. 4.4, beside the peaked shape, a main feature of these distributions is their asymmetry. Mean and median of each distribution are compared in each panel. We note that there is always a large discrepancy between the two, making difficult the definition of a robust value for bias correction. This issue can be mitigated using logarithmic units, i.e. converting reflectivity into *dBZ*. This would down-weight the impact of both largest and smallest values in the evaluation of the departures. This is shown in Figs. 4.8 and 4.9, where PDF of FG departures of radar reflectivity expressed in *dBZ* are plotted using the same layout. In addition, curves of Gaussian PDF with same variance are plotted for comparison. We note that the unit transformation gives more regular distributions and, more importantly for bias correction, median and mean have now similar values. Values of bias correction have been then computed based on reflectivity expressed in *dBZ*, stratifying over ranges of temperature and altitude, separately over geographical regions and seasons. Results for winter (December-January-February - DJF) mid latitudes and tropics are shown in Fig. 4.10. The effect of applying these relationships of bias correction on the intercomparison between CloudSat observations and FG reflectivities is shown in Fig. 4.11 and Fig. 4.12, for data relative to January 2007, respectively at mid latitudes South and in the tropics. In both figures the effect is a notable reduction of the offset that, prior to correction, is evident on the reddish areas of the scatterplots, i.e. where the majority of cases lay.
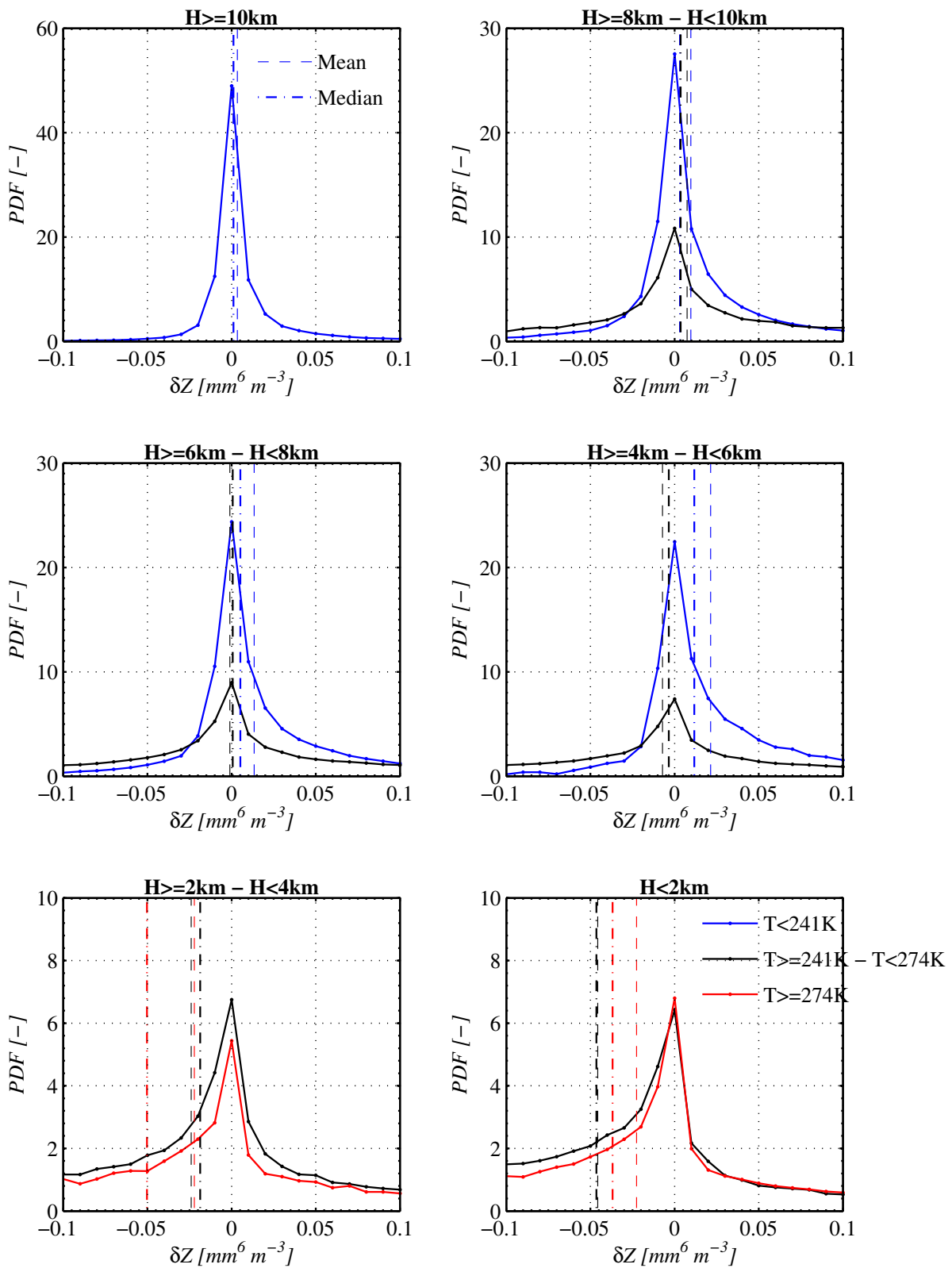
Figure 4.6: *PDF of reflectivity FG departures (expressed in $mm^6 m^{-3}$) using observations and model data over ocean at mid-latitudes South (30°S - 60°S) in January 2007. Only data passing quality control are considered. Each plot refers to observations in the altitude range indicated in the panel title. Different curves indicate the temperature interval shown in the legend. Vertical lines show mean (dashed) and median (dotted-dashed) of the distribution.*
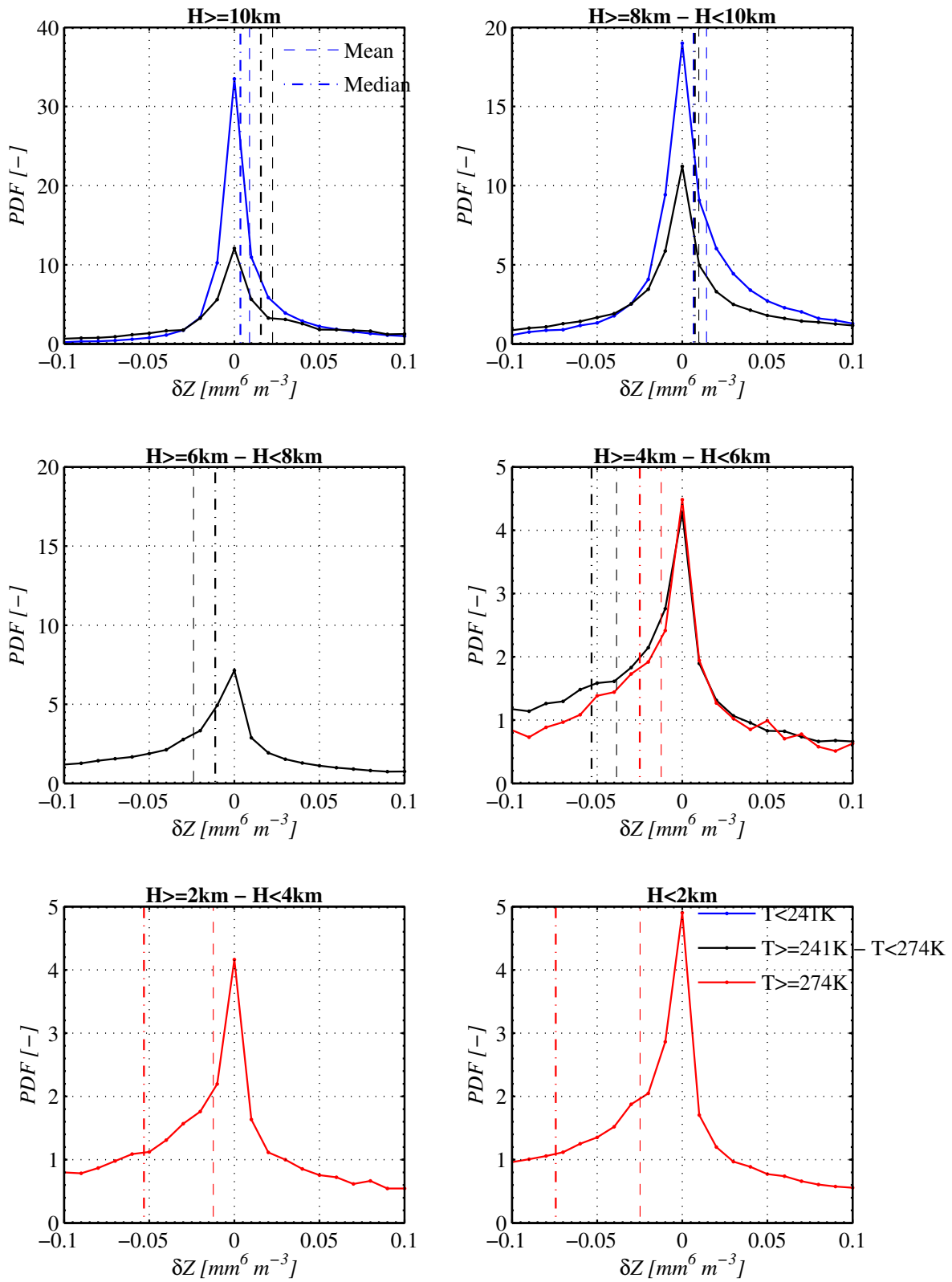
Figure 4.7: Same as Fig. 4.6, but considering observations in the tropics (30°S-30°N).
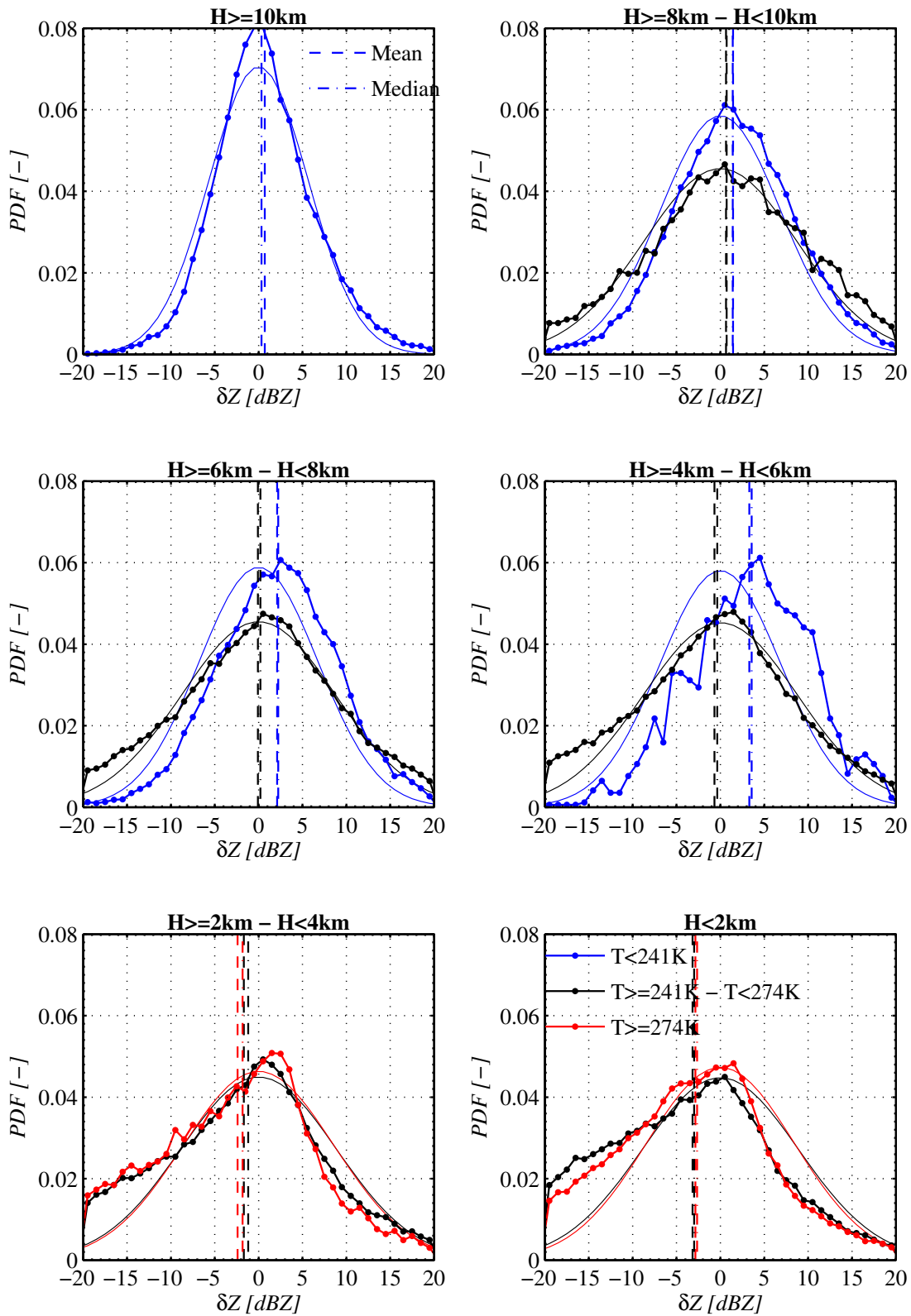
Figure 4.8: Same as Fig. 4.6, but evaluating FG departure of reflectivity expressed in dBZ units.
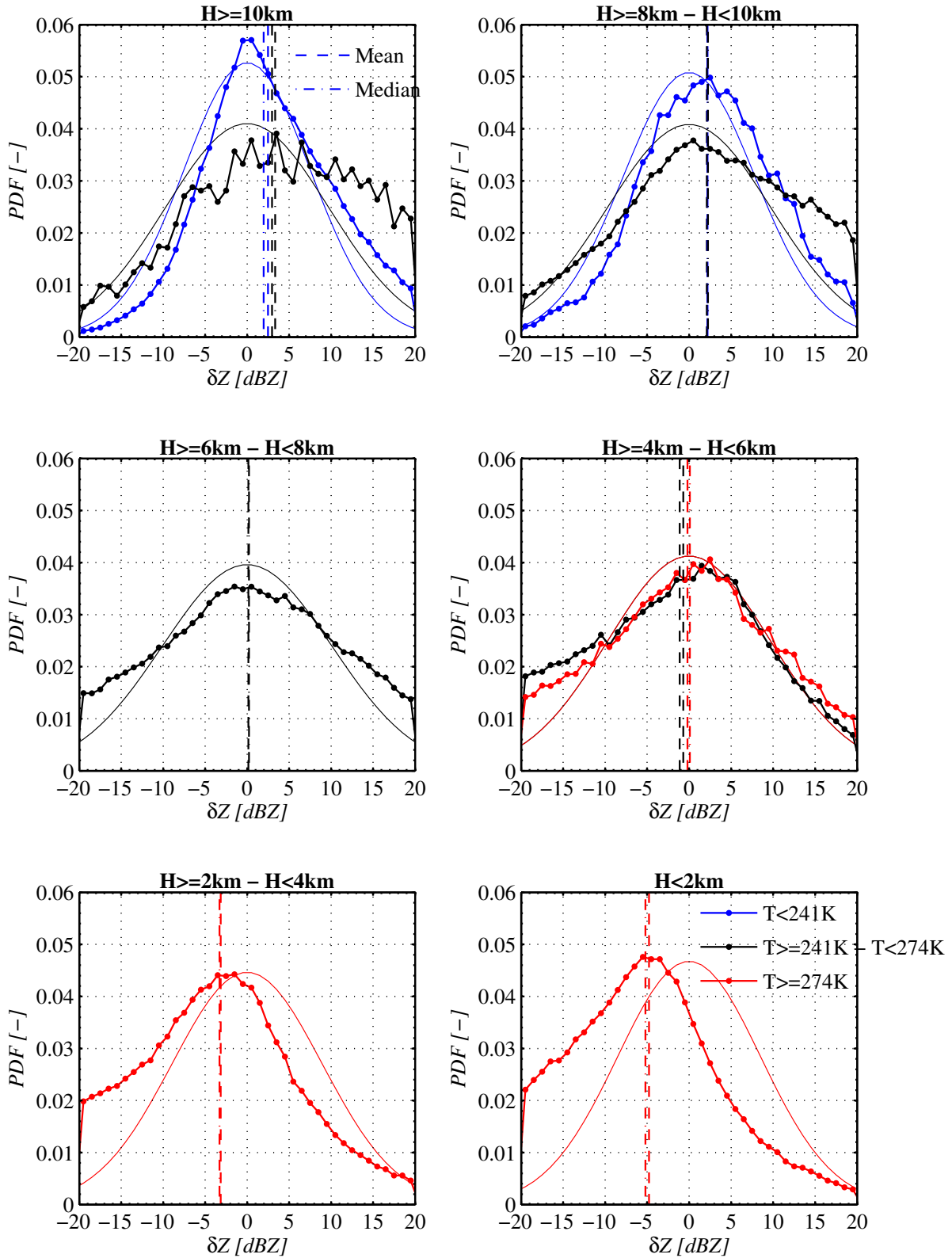
Figure 4.9: Same as Fig. 4.8, but considering observations in the tropics (30°S-30°N).
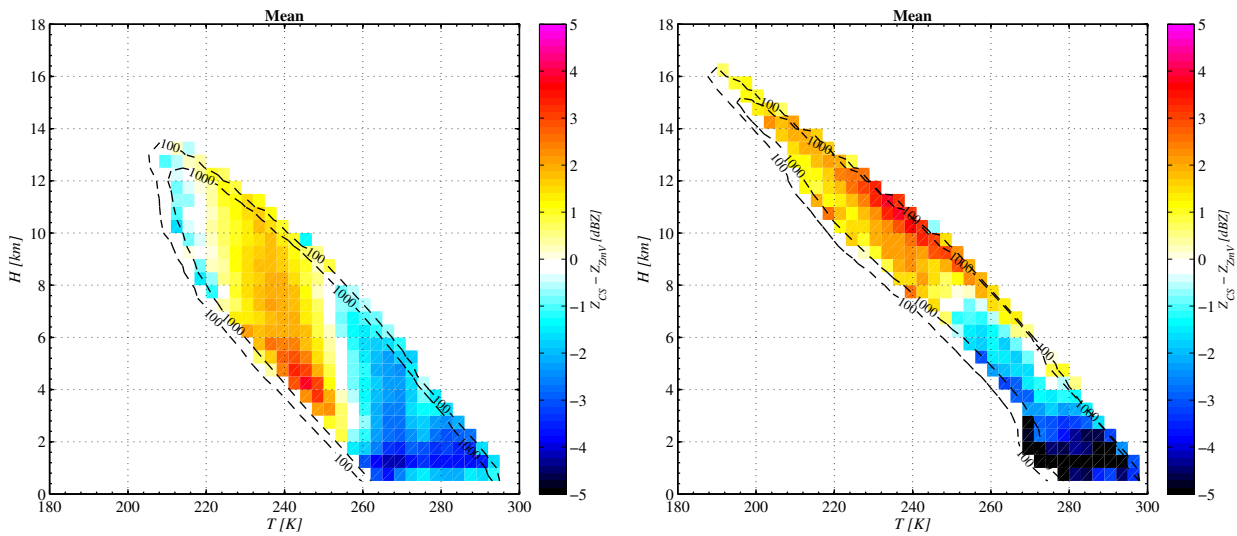
Figure 4.10: Mean FG departures (in dBZ), as function of temperature and altitude location of observations, used for bias correction. Contour lines gives the number of cases in each temperature/altitude bin. Left panel is for data over ocean at mid-latitudes South ($30°S$ - $60°S$) in January 2007. Right panel considers situations in the tropics ($30°S$-$30°N$).
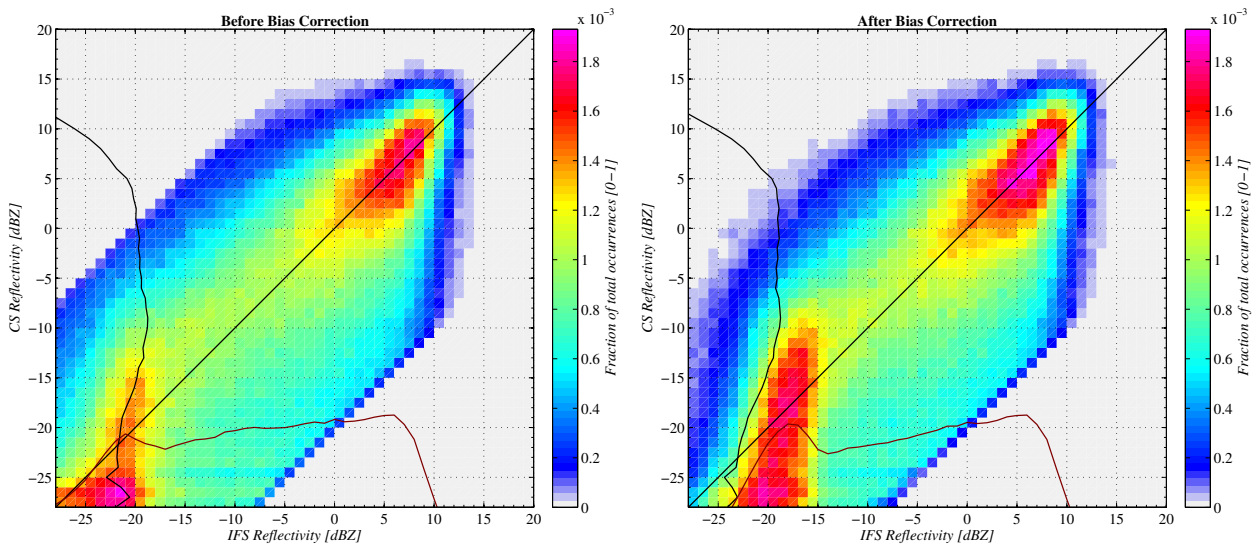

Figure 4.11: Scatter plot between simulated (on abscissa) and CloudSat reflectivity (on ordinate) using observations over ocean matched with model data for the period of January 2007 between $30°S$ and $60°S$. Only data passing quality control are considered. Left panel considers data prior to bias correction, while right panel uses with after applying the bias correction in Fig. 4.10.
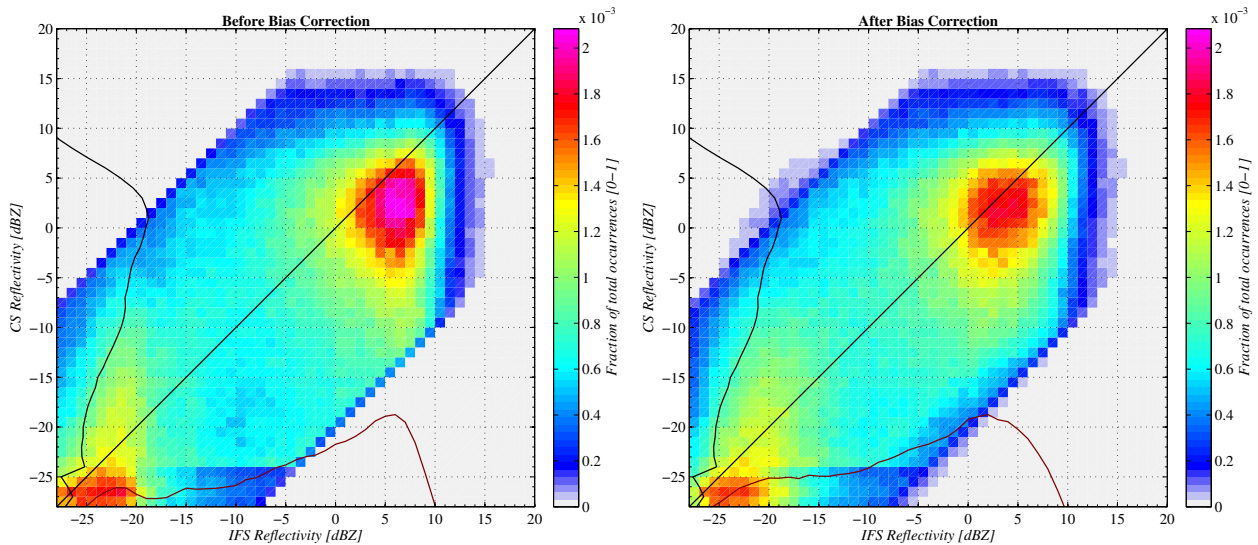
Figure 4.12: Same as Fig. 4.11, but considering observations in the tropics (30°S-30°N).

4.3 Conclusions

The pre-processing that is required prior to the monitoring and assimilation of the future EarthCARE data in the ECMWF model has been implemented and tested based on observed and simulated CloudSat radar reflectivity over a period of one month (January 2007). First, a method has been proposed to control the quality of the data in order to discard observations that would force the assimilation system to perform sub-optimally. The quality control of radar reflectivity measurements showed that the proposed technique is able to retain only data which present a quasi-Gaussian distribution. After quality control, the remaining data have been subject to a bias correction scheme that removes systematic errors. Results show that the method, based on temperature-specific linear regressions derived from statistics of the quality-controlled reflectivity FG departures, is able to significantly reduce the bias.

5   Summary

This report documents the efforts done to prepare for the monitoring/assimilation of EarthCARE radar observations using CloudSat data. The work package focused on a number of areas: improvements to the radar forward operator, estimation of observation errors, developing of a quality control strategy, and setup of a bias correction scheme for CloudSat reflectivities. In each of them, the following results have been achieved:

- The reflectivity forward operator ZmVar has been further developed improving the representation of scattering from frozen particles and implementing a model for multiple scattering.

- Tangent-linear and adjoint versions of the updated reflectivity forward operator, which are required for data assimilation when using a variational technique, have been developed and tested.

- Observation errors for CloudSat have been estimated considering separately the contribution from the instrument, the forward modelling and the spatial representativeness.

- A strategy for quality control of CloudSat observations has been developed that rejects those data that could degrade the performance of the assimilation system.

- Finally, a bias correction scheme has been put in place that is able to minimize the systematic differences between simulated and observed CloudSat reflectivity.

Acknowledgements

List of Acronyms

CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CFAD	Countoured Frequency by Altitude Diagrams
CFMIP	Cloud Feedback Model Intercomparison Project
CloudSat	NASA's cloud radar mission
COSP	CFMIP Observation Simulator Package
CPR	Cloud Profiling Radar
CPU	Central Processing Unit
EarthCARE	Earth, Clouds, Aerosols and Radiation Explorer
ECMWF	European Centre for Medium Range Weather Forecasts
ESA	European Space Agency
FG	First Guess
FOV	Field-of-view
GCM	General (or Global) Circulation Model
GES DISC	Goddard Earth Sciences Data and Information Services Center
IFS	Integrated Forecasting System of ECMWF
ISCCP	International Satellite Cloud Climatology Project
MC	Multi-Column
MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Multiple Scattering
NASA	National Aeronautics and Space Administration
NWP	Numerical Weather Prediction
PDF	Probability density function
PSD	Particle Size Distribution
QuARL	Quantitative Assessment of the operational value of space-borne Radar and Lidar measurements of cloud and aerosol profiles
RMSE	root mean square error
SC	Single-Column
SCOPS	Subgrid Cloud Overlap Profile Sampler
SNR	Signal-to-Noise ratio
SS	Single Scattering
STSE	Support-to-Science-Element
TRMM	Tropical Rainfall Measuring Mission
TDTS	Time Dependent Two Stream (approximation)
Z	Radar reflectivity
ZmVar	Z (reflectivity) Model for Variational assimilation of ECMWF

References

Battaglia, A., T. Augustynek, S. Tanelli, and P. Kollias, 2011: Multiple scattering identification in spaceborne W-band radar measurements of deep convective cores, *J. Geophys. Res.*, **116**(D19), D19201.

Battaglia, A., S. Tanelli, S. Kobayashi, D. Zrnic, R. Hogan, and C. Simmer, 2010: Multiple-scattering in radar systems: A review, *J. Quant. Spectroscopy and Rad. Transfer*, **111**(6), 917–947.

Bauer, P., A. Khain, A. Pokrovsky, R. Meneghini, C. Kummerow, F. Marzano, and J. Baptista, 2000: Combined cloud–microwave radiative transfer modeling of stratiform rainfall, *J. Atmos. Sci.*, **57**(8), 1082–1104.

Benedetti, A. and M. Janisková, 2004: Advances in cloud assimilation at ECMWF using ARM radar data, *Extended abstract for ICCP, Bologna*.

Benedetti, A., P. Lopez, E. Moreau, P. Bauer, and V. Venugopal, 2005: Verification of TMI-adjusted rainfall analyses of tropical cyclones at ECMWF using TRMM precipitation radar, *J. Applied Meteor.*, **44**(11), 1677–1690.

Bissonnette, L., P. Bruscaglioni, A. Ismaelli, G. Zaccanti, A. Cohen, Y. Benayahu, M. Kleiman, S. Egert, C. Flesia, P. Schwendimann, et al., 1995: Lidar multiple scattering from clouds, *Applied Physics B: Lasers and Optics*, **60**(4), 355–362.

Bodas-Salcedo, A., M. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. Klein, Y. Zhang, R. Marchand, J. Haynes, R. Pincus, and V. John, 2011: COSP: satellite simulation software for model assessment, *Bull. Amer. Meteor. Soc.*, **92**(8), 1023–1043.

Brown, P. and P. Francis, 1995: Improved measurements of the ice water content in cirrus using a total-water probe, *J. Ocean. Atmos. Tech.*, **12**(2).

Chandrasekhar, S., 1960: *Radiative transfer*, Dover Pubns.

Delanoë, J., R. Hogan, et al., 2010: Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds, *J. Geophys. Res*, **115**, D00H29.

Di Michele, S., M. Ahlgrimm, R. Forbes, M. Kulie, R. Bennartz, M. Janisková, and P. Bauer, 2012: Iterpreting and evaluation of the ECMWF global model with CloudSat onbservations: ambiguities due to radar reflectivity forwad operator uncertainties, *Q. J. R. Meteorol. Soc.*, **138**, 2047–2065, doi:10.1002/qj.1936.

Di Michele, S., O. Stiller, and R. Forbes, 2009: QuARL WP1000 Report: Forward operator developments - Errors and biases in representativity, Technical report, ECMWF.

Eloranta, E., 1998: Practical model for the calculation of multiply scattered lidar r eturns, *Applied Optics*, **37**(12), 2464–2472.

ESA, 2004: EarthCARE–Earth Clouds, Aerosols and Radiation Explorer, *Reports for mission selection, the six candidate Earth explorer missions, ESA SP-1279, ESA Publications Division c/o ESTEC, Noordwijk, The Netherlands*.

Evans, K. and G. Stephens, 1995: Microwave radiative transfer through clouds composed of realistically shaped ice crystals. Part I: Single scattering properties, *J. Atmos. Sci.*, **52**(11), 2041–2041.

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

Geleyn, J. and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation, *Contrib. Atmos. Phys*, **52**(1), 16.

Haynes, J. M., R. T. Marchand, Z. Luo, A. Bodas-Salcedo, and G. L. Stephens, 2007: A multipurpose radar simulation package: QuickBeam, *Bull. Am. Meteorol. Soc.*, **88**(11), 1723–1727.

Heymsfield, A., S. Lewis, A. Bansemer, J. Iaquinta, L. Miloshevich, M. Kajikawa, C. Twohy, and M. Poellot, 2002: A general approach for deriving the properties of cirrus and stratiform ice cloud particles, *J. Atmos. Sci.*, **59**(1), 3–29.

Hogan, R. and A. Battaglia, 2008: Fast lidar and radar multiple-scattering models. Part II: Wide-angle scattering using the time-dependent two-stream approximation, *J. Atmos. Sci.*, **65**(12), 3636–3651.

Hogan, R., N. Gaussiat, and A. Illingworth, 2005: Stratocumulus liquid water content from dual-wavelength radar, *J. Atmos. Ocean. Tech.*, **22**(8), 1207–1218.

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

Hong, G., 2007: Radar backscattering properties of nonspherical ice crystals at 94 GHz, *J. Geophys. Res*, **112**, D22203.

Hulst, H. and H. Van De Hulst, 1957: *Light scattering by small particles*, Dover Pubns.

Hunt, W., D. Winker, M. Vaughan, K. Powell, P. Lucker, and C. Weimer, 2009: CALIPSO lidar description and performance assessment, *Journal of Atmospheric and Oceanic Technology*, **26**(7), 1214–1228.

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

Janisková, M., 2004: Impact of EarthCARE products on Numerical Weather Prediction, *Contract report to the European Space Agency, 59 pp.*

Janisková, M., O. Stiller, S. Di Michele, R. Forbes, J.-J. Morcrette, M. Ahlgrimm, P. Bauer, and L. Jones, 2010: QuARL - Quantitative Assessment of the Operational Value of Space-Borne Radar and Lidar Measurements of Cloud and Aerosol Profiles, ESA Contract Report on Project 21613/08/NL/CB, 329 pp.

Kollias, P. and B. Albrecht, 2005: Why the melting layer radar reflectivity is not bright at 94 GHz, *Geophys. Res. Lett*, **32**.

Korolev, A., G. Isaac, and J. Hallett, 2000: Ice particle habits in stratiform clouds, *Q. J. R. Meteorol. Soc.*, **126**(569), 2873–2902.

Liebe, H., 1985: An updated model for millimeter wave propagation in moist air, *Radio Science*, **20**(5), 1069–1089.

Liebe, H., G. Hufford, and T. Manabe, 1991: A model for the complex permittivity of water at frequencies below 1 THz, *Int. J. Infrared and Millimeter Waves*, **12**(7), 659–675.

Liebe, H., P. Rosenkranz, and G. Hufford, 1992: Atmospheric 60-GHz oxygen spectrum- New laboratory measurements and line parameters, *J. Quantitative Spectroscopy and Rad. Transfer*, **48**(5), 629–643.

Liou, K., 2002: *An introduction to atmospheric radiation*, volume 84, Academic Press.

Liu, G., 2008: A database of microwave single-scattering properties for nonspherical ice particles, *Bull. Amer. Meteor. Soc*, **89**, 1563–1570.

Lopez, P., A. Benedetti, P. Bauer, M. Janisková, and M. Köhler, 2006: Experimental 2D-Var assimilation of ARM cloud and precipitation observations, *Q. J. R. Meteorol. Soc.*, **132**, 1325–1347.

Marshall, J. and W. Palmer, 1948: The distribution of raindrops with size., *J. Atmos. Sci.*, **5**, 165–166.

Marzano, F., L. Roberti, S. Di Michele, A. Mugnai, and A. Tassa, 2003: Modeling of apparent radar reflectivity due to convective clouds at attenuating wavelengths, *Radio Sci.*, **38**(1), 1002.

Mätzler, C. and U. Wegmüller, 1988: Dielectric properties of fresh-water ice at microwave frequencies, *Journal of Physics D: Applied Physics*, **21**, 1660.

Maxwell Garnett, J., 1904: Colours in metal glasses and metal films, *Trans. R. Soc. London*, **203**, 385–420.

Meador, W. and W. Weaver, 1980: Two-stream approximations to radiative transfer in planetary atmospheres - A unified description of existing methods and a new improvement, *J. Atmos. Sci.*, **37**(3), 630–643.

Mie, G., 1908: Beiträge zur optik trüber medien, speziell kolloidaler metallösungen, *Annalen der Physik*, **330**(3), 377–445.

Miles, N., J. Verlinde, and E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds, *J. Atmos. Sci.*, **57**(2), 295–311.

Pardo-Igúzquica, E. and M. Chica-Olmo, 1993: The Fourier integral method: an efficient spectral method for simulation of random fields, *Mathematical Geology*, **25**(2), 177–217.

Pardo-Igúzquica, E. and M. Chica-Olmo, 1994: SPECSIM: a program for simulating random fields by an improved spectral approach, *Computers & Geosciences*, **20**(4), 597–613.

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.

Stiller, O., 2010: A flow-dependent estimate for the sampling error, *J. Geophys. Res.*, **115**(D22).

Tanelli, S., S. Durden, E. Im, K. Pak, D. Reinke, P. Partain, J. Haynes, and R. Marchand, 2008: CloudSat's Cloud Profiling Radar after two years in orbit: Performance, calibration, and processing, *Geoscience and Remote Sensing, IEEE Transactions on*, **46**(11), 3560–3573.

Tian, L. and J. Curry, 1989: Cloud overlap statistics, *J. Geophys. Res.-Atmospheres*, **94**(D7).

Webb, M., C. Senior, S. Bony, and J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Climate Dynamics*, **17**(12), 905–922.

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.