



# ESA CONTRACT REPORT

Contract Report to the European Space Agency

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

**WP-1200 report: Observation operator  
and observation processing for cloud  
lidar**

October 2012, updated January 2014

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

ESA ESTEC contract 4000102816/11/NL/CT

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

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](mailto:library@ecmwf.int)

©Copyright 2014

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Contract Report to the European Space Agency

---

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

**WP-1200 report: Observation operator and observation  
processing for cloud lidar**

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

*ESA ESTEC contract 4000102816/11/NL/CT*

October 2012, updated January 2014



## ABSTRACT

This work package aims at building the main components needed to handle lidar observations in clouds within the ECMWF monitoring/assimilation system. The results achieved are summarized in this report. The first part describes the lidar forward operator, developed to simulate CALIPSO observations in clouds. The second part deals with the characterization of the observation errors, needed for data assimilation. In case of space-borne lidar (similarly to radar observations), the representativity error is an important component of the observation error. A flow-dependent estimate has been implemented for the CALIPSO lidar following the same approach previously used for the CloudSat radar. The final part of this document describes the development of a quality control and a bias correction scheme able to reduce departures between CALIPSO observations and model first guess to a level suitable for assimilation.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Lidar observation operator</b>	<b>1</b>
2.1	ZmVar lidar observation operator in clouds . . . . .	1
2.2	Microphysical assumptions . . . . .	3
2.2.1	Single particle scattering properties . . . . .	3
2.2.2	Relationships between mass content and bulk hydrometeor scattering properties . . . . .	5
2.2.3	Contoured Frequency by Altitude Displays . . . . .	7
2.3	Inclusion of a model for multiple scattering . . . . .	8
2.3.1	A case study . . . . .	12
2.3.2	Small-angle $\eta$ correction . . . . .	17
2.3.3	Wide-angle multiple scattering . . . . .	17
2.3.4	Computational cost of multiple scattering . . . . .	19
2.4	Summary from operator developments . . . . .	19
<b>3</b>	<b>Lidar observation errors</b>	<b>19</b>
3.1	Forward modelling and instrument errors . . . . .	20
3.2	Representativity errors of lidar observations . . . . .	22
3.2.1	Methodology . . . . .	23
3.2.2	Comparison between observed and synthetic backscatters . . . . .	23
3.2.3	Estimating the representativity errors . . . . .	24
3.2.4	Application to a case study . . . . .	25
3.3	Conclusions . . . . .	27
<b>4</b>	<b>Development of quality control and bias correction schemes for the lidar</b>	<b>29</b>
4.1	Quality control of CALIPSO lidar observations . . . . .	29
4.2	Bias correction of CALIPSO lidar observations . . . . .	35
4.3	Conclusions . . . . .	35
<b>5</b>	<b>Summary</b>	<b>37</b>

## 1 Introduction

The Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE) mission will provide, using a combination of a lidar and a cloud radar, the vertical structure and the horizontal distribution of cloud and aerosol fields on a global scale (ESA 2004). Coincident space-borne lidar and radar observations are currently available thanks to the CloudSat (Stephens et al., 2002) and CALIPSO (Winker et al., 2009) missions, part of the A-Train constellation. 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 have the potential to be assimilated into a Numerical Weather Prediction (NWP) models. The information extracted from these data can therefore improve the initial atmospheric state of General Circulation Models (GCM).

The objective of Support-to-Science-Element (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 European Centre for Medium Range Weather Forecasts (ECMWF) model. The work will lay the foundation of a pre-operational system for the exploitation of radar and lidar observations in data assimilation in the time frame of the EarthCARE mission.

As part of the EarthCARE assimilation project, this report summarizes the initial work that has been done to prepare for the inclusion of the EarthCARE lidar data in the ECMWF monitoring/assimilation system using as testing data lidar observations in clouds from CALIPSO. CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization), the lidar onboard CALIPSO, 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).

The aim of this work package is the development of various components needed to monitor and assimilate CALIOP observations, i.e.: a forward operator, the definition of observation errors and a quality control/bias correction scheme. Section 2 of this report describes the implementation of a forward operator able to simulate cloud backscatter at CALIOP wavelengths and the inclusion of a model for lidar multiple scattering. The definition of lidar observation errors is summarized in Section 3, where forward modelling uncertainties and representativity errors are discussed. Finally, Section 4 describes the schemes developed for the quality control and bias correction needed for the assimilation of CALIOP cloud backscatter.

## 2 Lidar observation operator

The availability of a forward operator able to simulate lidar measurements is a prerequisite for a monitoring/assimilation system that would use observations from the CALIPSO or EarthCARE missions. This section documents the development of such forward operator for lidar backscatter due to clouds.

### 2.1 ZmVar lidar observation operator in clouds

A lidar system works on principles similar to the ones of a radar, the main difference being the operating wavelength in the visible band. Lidar basic variables are therefore the same as for the radar, with the only distinction consisting in the definition of the return signal. In fact, in case of lidars, instead of the radar reflectivity factor, it is customary to use the backscatter coefficient (for hydrometeor type  $h$ , at wavelength  $\lambda$ ),  $\beta_\lambda^h$ , defined as:

$$\beta_\lambda^h = \frac{1}{4\pi} \int_{D_{min}^h}^{D_{max}^h} \sigma_{\text{bsc},\lambda}^h(D) N^h(D) dD \quad (2.1)$$

where  $N^h(D)$  is the particle size distribution (PSD),  $\sigma_{bsc,\lambda}^h(D)$  is the backscattering cross section,  $D_{min}^h$  and  $D_{max}^h$  are the limits for the particle size. For the definition of the other quantities later introduced, the reader is referred to Di Michele et al. (2014) (Observation operator and observation processing for cloud radar). The lidar simulator in clouds has been conveniently developed as an extension of the reflectivity model for variational assimilation (ZmVar), described in Di Michele et al. (2012). Figure 2.1, contains a schematic representation of the different components of ZmVar, highlighting (in red) the ones that have been updated to achieve the simulation of the lidar return in clouds. As for the reflectivity model, inputs to ZmVar are the vertical profiles of cloud and thermodynamic variables generated from the ECMWF model (Module 1). The model data used for this work package were obtained using the 37R3 version of the model (operationally used from November 2011 to June 2012) with a spectral truncation of T799 (approximate corresponding to grid resolution of 25 km) and 91 levels in the vertical up to 80 km altitude. Cloud and precipitation processes are described by prognostic equations for cloud liquid, cloud ice, sub-grid cloud fraction, stratiform rain and snow with a diagnostic parametrization for sub-grid precipitation fraction. This cloud scheme is based on Tiedtke (1993), but it has since been extended and modified significantly Forbes et al. (2011). The most recent versions of the ECMWF model has precipitating rain and snow as additional prognostic variables and also partitions the cloud condensate into separately prognosed cloud liquid and ice variables for an improved representation of the mixed-phase. Microphysical processes are parametrized for nucleation, phase transitions (e.g. condensation, evaporation) and collection processes (e.g. accretion, aggregation) to transfer water between the different categories. Although the fraction of a grid-box that is covered by cloud (the sub-grid cloud fraction) is prognosed in the model with sources and sinks from different processes, the precipitation fraction is a diagnostic variable based on the cloud profile with maximum overlap in the vertical.

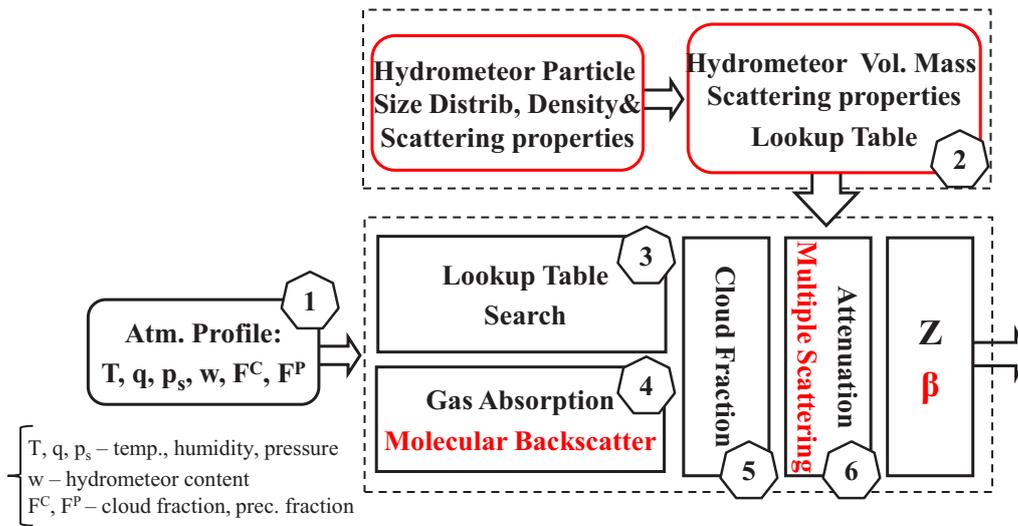


Figure 2.1: Schematic diagram of ZmVar. The components added/modified for lidar simulation are shown in red.

A new look-up table of hydrometeor optical properties has been generated to allow ZmVar to work with the lidar wavelengths (Module 2). A detailed description will be given in the following section. The lookup table search (Module 3) did not need major changes since lidar quantities match the radar ones. At visible wavelengths, the signal is subject to scattering (with negligible absorption) from the molecular components of the atmosphere. Module 4 for lidar prescribes this atmospheric contribution to the backscatter using the following expression (Collis and Russell, 1976):

$$\beta_{\lambda}^{mol} = 5.45 \cdot 10^{-32} \times \frac{p}{KT} \times \left( \frac{\lambda}{0.55} \right)^{-4.09} \quad (2.2)$$

where  $p$  is the atmospheric pressure,  $T$  is temperature,  $K$  is the Boltzmann constant ( $1.38 \cdot 10^{-23} JK^{-1}$ ). The atmospheric extinction can be derived from  $\beta_{\lambda}^{mol}$  assuming the customary value of  $\frac{8\pi}{3}sr$  for the extinction

backscattering ratio of molecular and the air scattering albedo equal to 1.

The treatment of cloud fraction is not related to the type of observations and therefore Module 5 of the lidar simulator is the same as used for the radar. In order to simulate the apparent backscatter as measured by the lidar, the final step is the modelling of the propagation of the backscatter signal along the vertical (Module 6). This involves the evaluation of the signal attenuation and of the multiple scattering contributions, never negligible in a lidar system. A description of the multiple scattering implemented for taking into account the lidar multiple scattering is described in Section 2.3.

## 2.2 Microphysical assumptions

The lookup table of scattering properties is a key component of the ZmVar. During the development of the radar version, big efforts were made in order to setup a table configuration which would lead to a minimization of the discrepancies between simulated reflectivity and corresponding CloudSat observations. Some of those results, extensively described in WP-1100 of this project (Di Michele et al., 2014), have been exploited for the definition of the lidar lookup table. Equations 2.1 and 2.2 in that report show that the bulk scattering/extinction properties are essentially a convolution between the hydrometeor particle size distribution and the (size- and wavelength- dependent) scattering/extinction of the single particle. The PSD assumed for each hydrometeor for the radar are therefore also used for the generation of the lidar table. Noteworthy, the ZmVar lookup table specifies scattering properties relative to six hydrometeors (consistently with the ECMWF forecast model): cloud liquid, cloud ice, large-scale/convective rain and large-scale/convective snow. As for the radar operator, no distinction has been made between the properties of large-scale and convective precipitation (rain and snow). Figure 2.2 shows the shape of PSDs considering a mass content ( $w$ ) of  $0.1 \text{ gm}^{-3}$  for each hydrometeor, while Tab. 2.1 contains their mathematical expressions. For cloud liquid, a lognormal PSD is used which, unlike the exponential, does not change shape with changes in the water content. ZmVar uses the values of median diameter and logarithm width for maritime clouds as reported in Table 3 of Miles et al. (2000). The rain PSD follows the 'normalized' gamma distribution proposed by Illingworth and Blackman (2002). For snow, the parameterization of Field et al. (2007) is used. 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, which is representative of a wider range of meteorological conditions. The PSD of cloud ice also uses the same Field et al. (2007) expression, but in order not to consider any aggregation phenomena for this type of particles, the temperature dependence has been set to a constant value (chosen in  $-70^\circ\text{C}$ ). Figure 2.2 shows the shape of the assumed PSDs considering a mass content ( $w$ ) of  $0.1 \text{ gm}^{-3}$  for each hydrometeor.

### 2.2.1 Single particle scattering properties

For each hydrometeor type ( $h$ ), the fundamental radiative quantities to be known are: backscattering cross section,  $\sigma_{bsc,\lambda}^h(D)$ , extinction cross section,  $\sigma_{ext,\lambda}^h(D)$ , single scattering albedo  $\omega_\lambda^h(D)$  and asymmetry factor  $g_{1,\lambda}^h(D)$  at the given wavelength  $\lambda$  over the range of particle sizes ( $D$ ). For rain and cloud liquid particles, sphericity is a reasonable assumption and therefore the above quantities can be computed using the Mie solution Mie (1908). To this purpose, beside the particle size (or the indeed the size parameter  $D/\lambda$ ), the complex refractive index  $m = m' - i m''$  is required. Results from Pope et al. (1997) have been used to model  $m$  at the wavelengths of the CALIPSO lidar. While sphericity is a convenient approximation for liquid particles, ice particles can assume a variety of shapes depending on their growth environment and on past evolution. Thus a more realistic particle habits for ice (i.e. shapes) must be used, which implies a more complex solution to the problem of finding their radiative properties. In this study, we have used the results of Yang et al. (2000), where the single-scattering properties in the solar spectrum from  $0.2 \mu\text{m}$  to  $5 \mu\text{m}$  have been evaluated for six ice particle habits: plates, columns, hollow columns, planar bullet rosettes with four branches, three-dimensional

Hydrometeor	Particle type	PSD and parameters	Density [ $g/cm^3$ ] $\rho(D) = aD^b$
Cloud Liquid	Sphere	$n(r) = \frac{N_t}{\sqrt{2\pi}(\ln\sigma_g)r} e^{-\frac{\ln^2(r/r_g)}{2(\ln\sigma_g)^2}}$ $\begin{cases} r_g = 30\mu m \\ \sigma_g = 710^{-4} \end{cases}$	$a = 1$ $b = 0$
Cloud Ice	Aggregate of columns	<i>Field et al. (2007), no Temp dependence</i>	$a = 5.615 \cdot 10^{-3}$ $b = -1.1$
Rain LS	Sphere	$N(D) = \frac{0.03N_t D_0^4 \Lambda^{\mu+4}}{\Gamma(\mu+4)} D^\mu e^{-\Lambda D}$ $\begin{cases} N_L = 0.08 cm^{-4} \\ \Lambda = \frac{3.6+\mu}{D_0} \end{cases}$	$a = 1$ $b = 0$
Snow LS	Aggregate of columns	<i>Field et al. (2007)</i>	$a = 5.615 \cdot 10^{-3}$ $b = -1.1$
Rain Conv	Sphere	$N(D) = \frac{0.03N_t D_0^4 \Lambda^{\mu+4}}{\Gamma(\mu+4)} D^\mu e^{-\Lambda D}$ $\begin{cases} N_L = 0.08 cm^{-4} \\ \Lambda = \frac{3.6+\mu}{D_0} \end{cases}$	$a = 1$ $b = 0$
Snow Conv	Aggregate of columns	<i>Field et al. (2007)</i>	$a = 5.615 \cdot 10^{-3}$ $b = -1.1$

Table 2.1: Parametrization of hydrometeor particle properties in ZmVar.

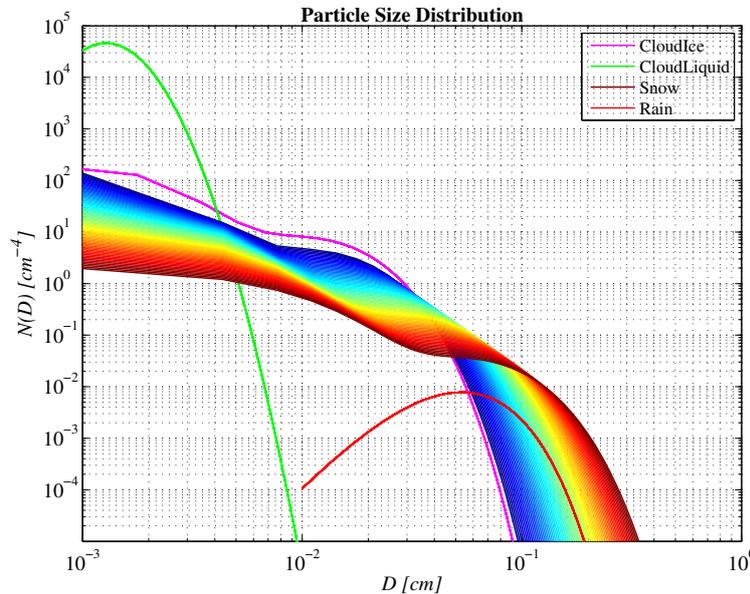


Figure 2.2: Hydrometeor particle size distribution as prescribed in ZmVar lookup table for mass content equal to  $0.1 gm^{-3}$ . Magenta, green, red lines refer to cloud ice, cloud liquid and rain, respectively. Envelopes of rainbowshaded-coloured lines show the temperature dependence of snow from  $-70^\circ C$  (deep bluemoost left) to  $0^\circ C$  (brownmoost right), with a  $1^\circ C$  stepping.

bullet rosettes with six branches, and compact aggregates of solid columns. Calculations were made by a combination of two scattering computational models: the finite-difference time domain method (FDTD) (Yee, 1966) and the improved geometric optics method (IGOM) (Yang and Liou, 1996). We have found that the best agreement with CALIPSO observations (at  $532 \mu m$ ) is obtained assigning the values provided for aggregates to all the three ice categories (cloud ice, large-scale snow and convective snow). This is a reasonable assumption since, as discussed in Korolev et al. (2000), most ice particle size distributions are dominated by irregular shapes. Figure 2.3 shows the single-particle scattering/extinction quantities assigned to each hydrometeor type as function of the size. In panel a), one can note how the cloud and rain backscatter oscillates very rapidly. This can be seen as a result of the interference between reflections from the front and back surface of the spherical droplets, present when the diameter is comparable to the light wavelength. In real clouds there is always a

distribution of sizes of cloud droplets, and these fluctuations in  $\beta$  are smoothed out. As typical of the radiative properties at visible wavelengths, we note that for all hydrometeors the value of extinction efficiency tends to the asymptotic value of 2 (panel b) and that the single scattering albedo is very close to unity (panel c).

Finally, it must be mentioned that, 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. 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:

$$\rho(D) = aD^b \quad (2.3)$$

where  $D$  is interpreted as the maximum dimension of the particle. The values for  $a$  and  $b$  relative to the assumed ice shapes used in ZmVar are given in the last column of Tab. 2.1.

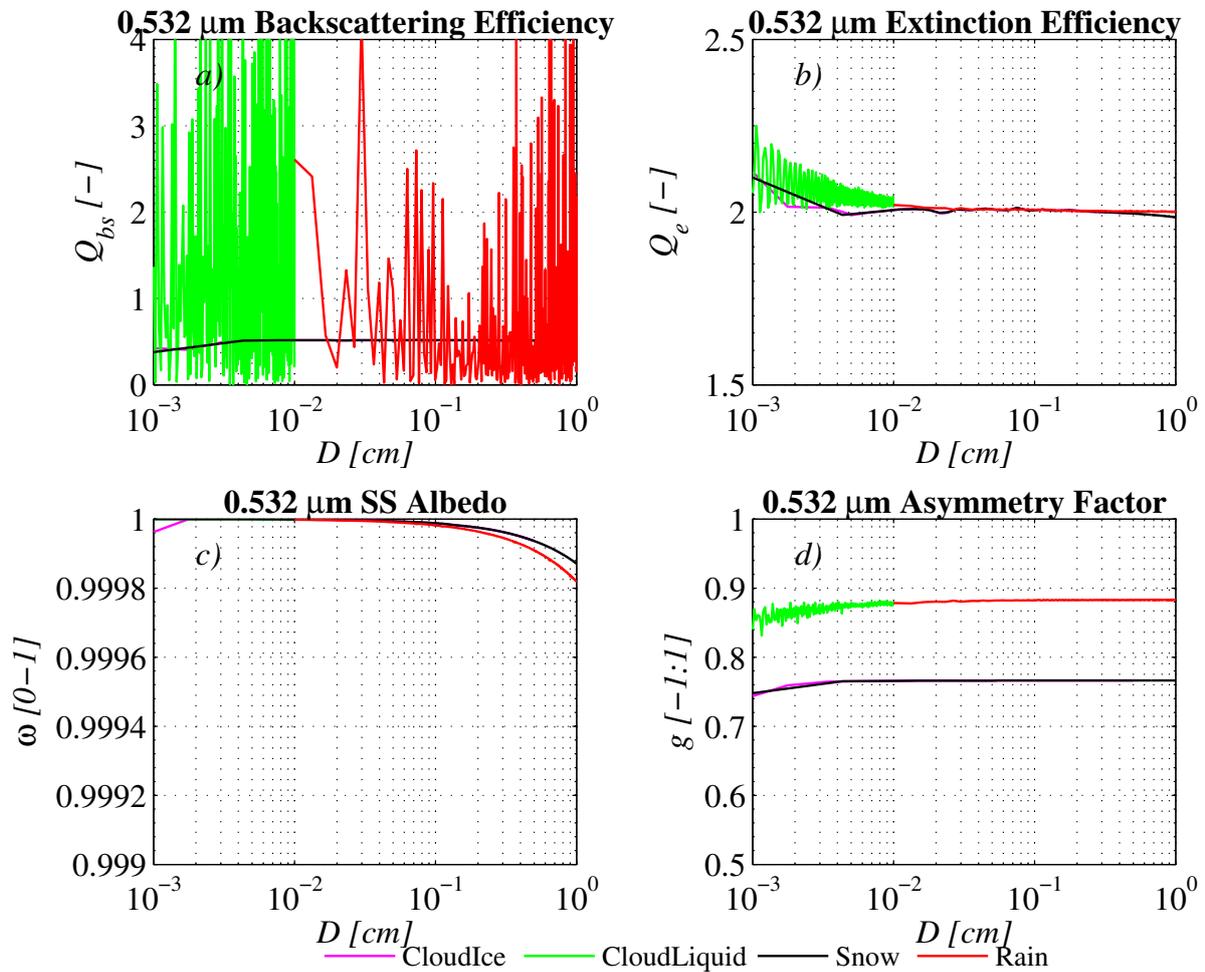


Figure 2.3: Single particle backscattering efficiency (panel a), extinction efficiency (panel b), single scattering albedo (panel c) and asymmetry factor (panel d) at  $0.532 \mu\text{m}$  as function of size, used for each of the hydrometeor categories represented in the ZmVar lookup table.

### 2.2.2 Relationships between mass content and bulk hydrometeor scattering properties

After making the choice for the single particle scattering properties and the PSD,  $\beta_\lambda^h$  of the hydrometeor  $h$  can be evaluated by integrating over the size spectrum according Eq. 2.1. Volumetric  $\sigma_{ext,\lambda}^h$ ,  $\omega_h$ , and  $g^h$  are obtained using similar expressions, as described in WP-1100 (Di Michele et al., 2014). Their values across a range of mass hydrometeor contents are the quantities stored in the lookup table. Figure 2.4 shows the resulting relationships between hydrometeor mass content  $w$  and bulk radiative properties for the  $0.532\mu\text{m}$  CALIOP

channel. Comparing the backscatter coefficient of the hydrometeors (panel a), we note that the cloud liquid can produce values up to one order of magnitude larger than the one of cloud ice. This (well known) result is a consequence of the much larger number of particles of the PSD. For the same reason, the rain represents the lowest backscatter for a given mass content. Cloud ice and snow values lie in between. The volumetric extinction vs.  $w$  (panel b) show similar features, with the only difference being that values of cloud liquid and cloud ice are comparable for  $w < 10^{-3}$ . While the volumetric single scattering albedo (panel c) is always very close to unity (consistently with the one of the corresponding single particle), the volumetric asymmetry factor (panel d) of liquid hydrometeors is significantly larger than the one of frozen particles. This characteristic, as it will be shown in Section 2.3, leads to a much more influence of the multiple scattering phenomenon in warm phase clouds. Bulk hydrometeor scattering properties have been also evaluated for the CALIOP channel at  $1.064 \mu\text{m}$  and the results have been included in the ZmVar lookup table. Figure 2.5 shows the ratio between the quantities of these two channels. At the longer wavelength, rain and cloud liquid have larger backscatter and comparable extinction, while for cloud ice and snow there is a larger extinction and comparable backscatter. It must be noted that the results for frozen particles are not generally valid. As shown in Fig. 2.6, the ratio between single ice particle optical properties at the two wavelengths of interest can be greater or less than unity depending on the shape and, for a given shape, on the size of the particle. Therefore both the assumption of the particle shape and the choice of PSD affect the results in Fig. 2.5.

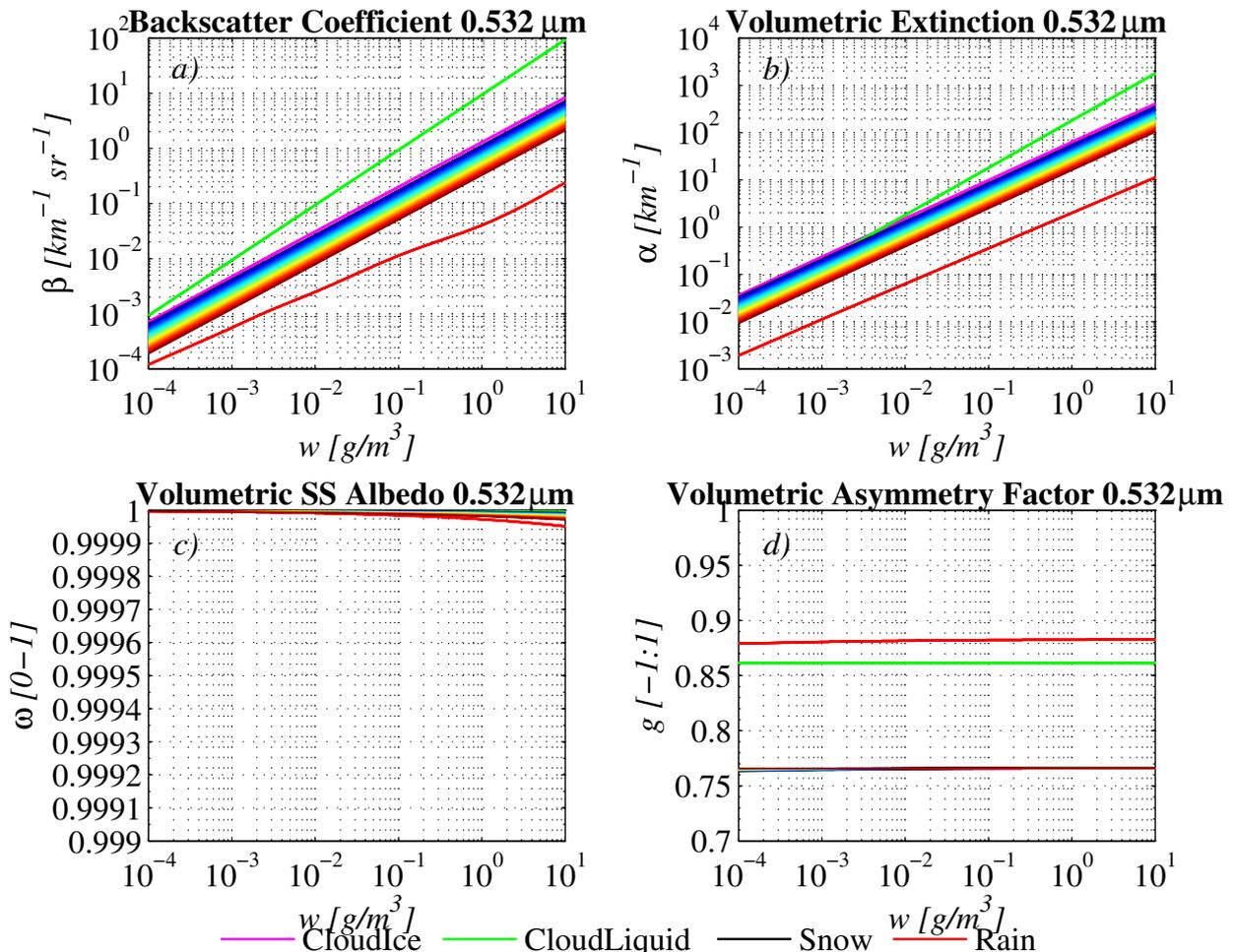


Figure 2.4: Relationships between mass content of each hydrometeor type and corresponding backscatter coefficient (panel a), volumetric extinction (panel b), volumetric single scattering albedo (panel c) and volumetric asymmetry factor (panel d) at 0.532 μm, as prescribed in ZmVar lookup table. Magenta, green, red lines refer to cloud ice, cloud liquid and rain, respectively. Envelopes of rainbow shaded-coloured lines show the temperature dependence of snow from -70°C (deep bluemoost left) to 0°C (brownmoost right), with a 1°C stepping.

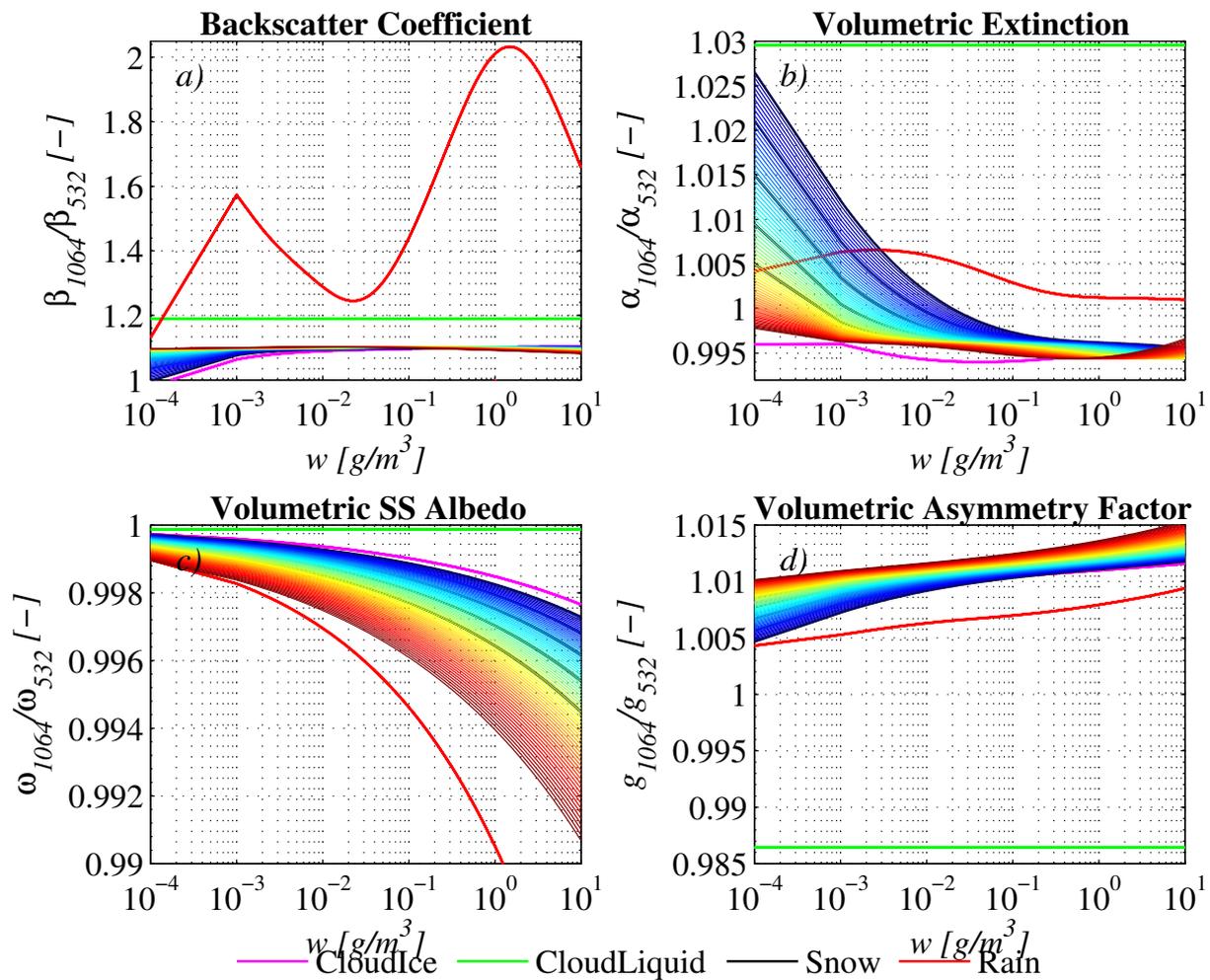


Figure 2.5: As Fig. 2.4, but considering the ratio between backscatter coefficient (panel a), volumetric extinction (panel b), volumetric single scattering albedo (panel c) and volumetric asymmetry factor (panel d) at 1.064  $\mu\text{m}$  and the same quantities at 0.532  $\mu\text{m}$ .

### 2.2.3 Contoured Frequency by Altitude Displays

PDF of reflectivity along height (or temperature), commonly referred as Contoured Frequency by Altitude Displays (CFADs), can provide useful information on the quality of the simulated observations. Figure 2.7 shows these diagrams for the CALIPSO cloud-affected observations for the period of January 2007 (at 0.532  $\mu\text{m}$ ) averaged at model grid resolution (panel a) and the corresponding simulated backscatter from ZmVar (panel b). The shading colours result from the PDFs of backscatter, normalized at each height range. The comparison shows that the pattern is similar for observations and simulations. There are, however, a number of noticeable differences in the intensity, that can be explained by known deficiencies in the representation of clouds in the current version of the forecasts model. Above 15 km, we note an evident underestimation of the ice amount. This problem, occurring mainly in the Tropics, is also highlighted from other types of observations, as for example the ones from the Microwave Limb Sounder (Waliser et al., 2009). At the same altitudes, we also see that the simulated backscatter can assume much larger values (longer tail of the distribution). These occurrences can be traced back to cases of deep convection, probably prescribing an excessive amount of ice on the upper portion of the cloud structure. The second main difference is evident up to 12 km, where simulations show an excessive occurrence of the lowest backscatter values (ranging from  $10^{-4}$  to  $10^{-3}$   $\text{km}^{-1}\text{sr}^{-1}$ ). In this region mixed phase clouds occur and this discrepancy can be explained by the model not producing the right

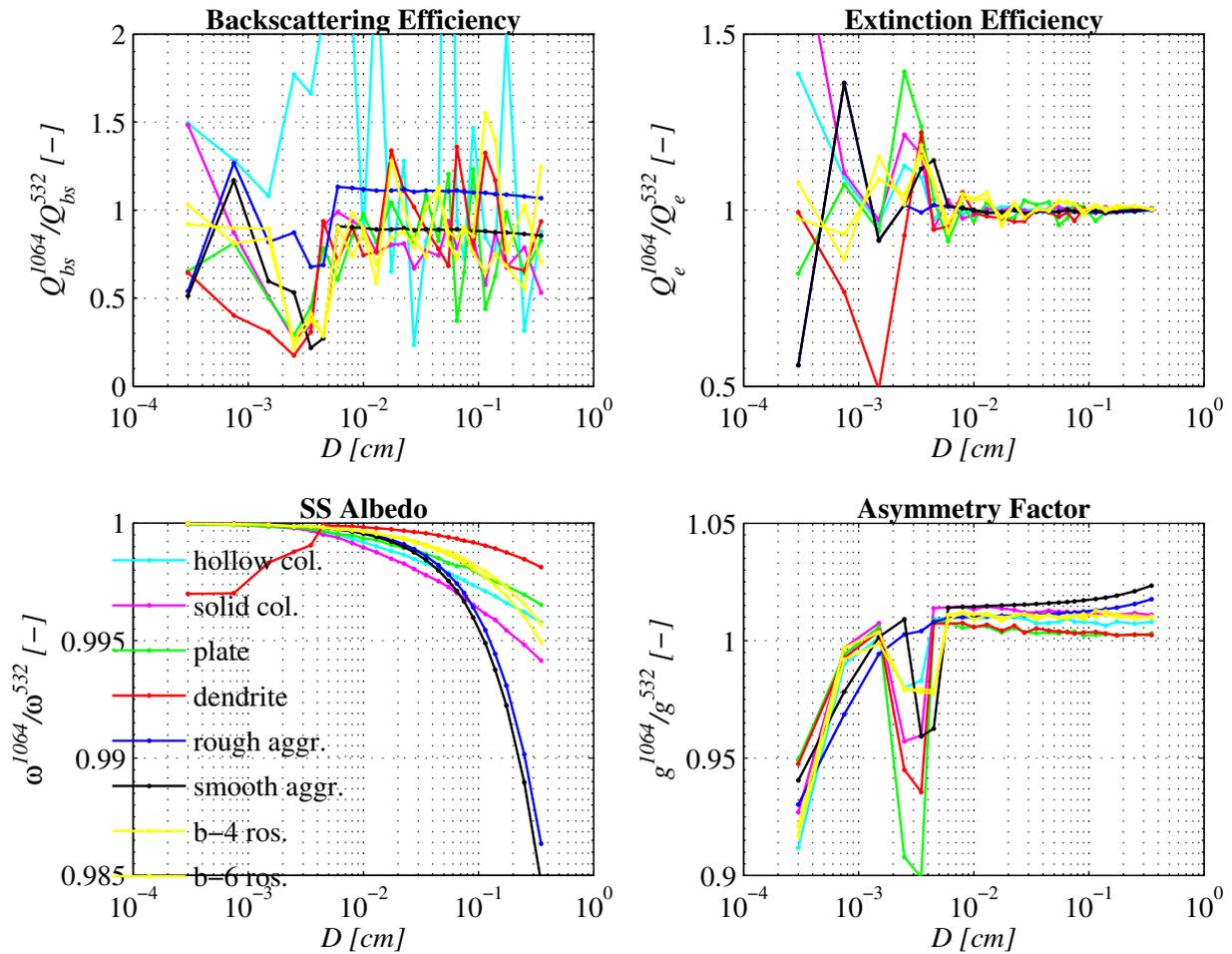


Figure 2.6: As Fig. 2.4, but considering the ratio between single particle backscattering efficiency (panel a), extinction efficiency (panel b), single scattering albedo (panel c) and asymmetry factor (panel d) at  $1.064 \mu\text{m}$  and the same quantities at  $0.532 \mu\text{m}$ .

amount of supercooled cloud liquid. This hypothesis is consistent with the cloud mode (present at altitudes below 7 km) which for CALIOP observations is much more defined than in simulations. Also shown in Fig. 2.7 (panel c), is the fractional occurrence of observations (red line) and simulations (black line) at each altitude range. The two curves are very similar, expressing that overall clouds (and the corresponding backscatter) are correctly simulated, apart from the problems mentioned above. These discrepancies will be revisited in the future, when changes will be made to the cloud parameterization of the model.

### 2.3 Inclusion of a model for multiple scattering

The problem of multiple scattering (MS) experienced by active remote sensors, can be treated considering separately the following two scattering regimes: small angle MS and wide angle MS. WP-1100 of this project (Di Michele et al., 2014) provides definitions and details of the solution for the wide angle MS since radar observations are only affected by this regime. The main scatterers that are important for lidar are hydrometeors, aerosols, and molecules. Lidar observations of clouds are also affected by wide angle MS (and the component of scattering by cloud particles that is not in the forward lobe). However, in addition, they are also affected by small angle MS since only hydrometeor particles are sufficiently large compared to the wavelength to produce a narrow forward lobe in the phase function, while aerosols and molecules can contribute to MS only in the

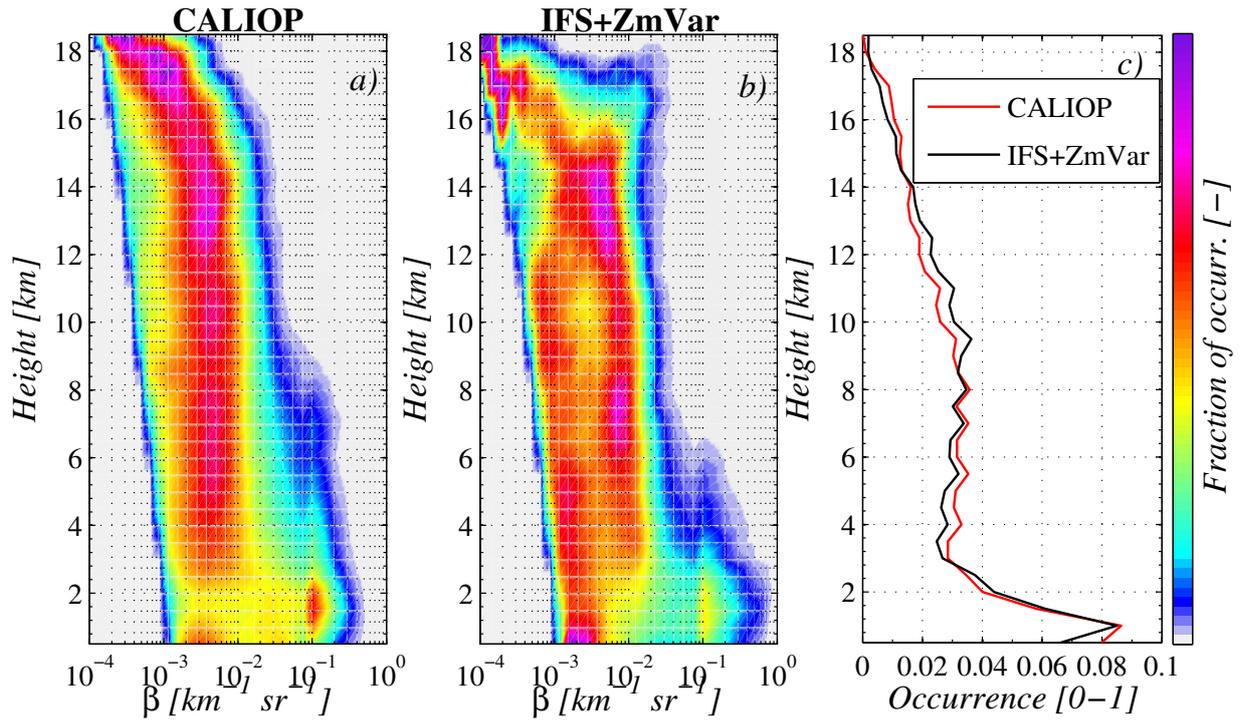


Figure 2.7: Frequency distribution at different altitude levels of lidar backscatter in clouds (at  $0.532\mu\text{m}$ ). Panel (a) shows CALIPSO observations for January 2007 averaged at model resolution, while panel (b) shows the corresponding simulated reflectivities (model grid box mean) from ZmVar. The occurrence of backscatter across altitudes is shown in panel (c).

wide-angle regime.

#### Computing small angle MS using the Photon Variance Covariance method

The field of view of a lidar is such that the small-angle forward-scattered photons may remain within the field of view of the detector and contribute to the apparent backscatter. Because the distance travelled by photons that have only undergone small-angle scattering is approximately the same as unscattered photons, the time of travel can still be converted unambiguously into the distance of the target from the instrument. In single scattering regime, the apparent backscatter  $\hat{\beta}(r)$  measured at a range  $r$  (having a backscatter coefficient  $\beta$  and volumetric extinction  $\alpha$ ) can be written as:

$$\hat{\beta}(r) = \beta(r) e^{-2\delta(r)} \quad (2.4)$$

where  $\delta(r) = \int_0^r \alpha(r') dr'$  is the optical depth of the medium (clouds, aerosols and gases) to range  $r$  from the instrument. The simplest method to model small-angle multiple scattering was proposed by Platt (1973), who introduced a term  $\eta$  in Eq. 2.4 so that:

$$\hat{\beta}_d(r) = \beta(r) e^{-2\eta\delta(r)} \quad (2.5)$$

The value of  $\eta$  can vary between 1 (the single-scattering limit) and  $1/2$  (the wide field-of-view limit). The latter case corresponds to all photons in the forward lobe remaining within the receiver field of view. Thus,  $\eta$  has the effect of reducing the effective optical depth such that these photons are treated as if they had not been scattered at all. If the medium is represented by an N-level profile of extinction values, then the calculation of Eq. 2.5 is  $O(N)$  efficient. However, the problem with this approach is that forward-scattered photons progressively escape from the field of view with increasing distance downstream of the scattering event, implying that  $\eta$  should increase with range.

Unfortunately, there is no satisfactory theory for deriving  $\eta$  and its range dependence for a particular lidar geometry and scattering phase function. [Eloranta \(1998\)](#) took a more rigorous approach, explicitly calculating each scattering order separately. Recently, [Hogan \(2008\)](#) proposed a new method for the modelling of small-angle MS, fast enough to be incorporated in the forward models used for retrieval. Using Eloranta's approach for double scattering, he treated all orders of scattering above the second order together in such a way that the whole algorithm is  $O(N)$  efficient. This was achieved by using a 'photon variance-covariance' (PVC) method, so called because it is based upon an evaluation of the variance and covariance of photons position and direction.

The formulation of the PVC algorithm is simplified by the use of the equivalent-medium theorem, which has been proved to be valid under the small-angle approximation ([Bissonnette, 2005](#)). This theorem states that the backscatter measured in a medium is the same as that from an equivalent hypothetical medium that has twice the extinction and scattering coefficients (but the same phase function) as the true medium on the outward journey but zero extinction and scattering on the return journey. Thus, the two-way problem is transformed into a simpler one-way propagation problem. In common with [Eloranta \(1998\)](#), the following assumptions are made: (i) both the laser divergence and the forward-scattering lobe from a distribution of particles may be represented as Gaussian, (ii) the extra path length of multiply scattered photons may be neglected, (iii) the lidar is monostatic so that the problem has azimuthal symmetry, and (iv) all angles are small enough that  $\sin\theta \approx \theta$ .

Let's consider a laser transmitter that emits a short pulse of total energy  $P_0$  in a Gaussian beam with a  $1/e$  angular half-width of  $\rho_{lr}$ . At a distance  $r$  from the instrument, the energy density of outgoing unscattered photons (in units of  $Jm^{-2}$ ) in the equivalent medium is a function of  $r$  and of the distance perpendicular to the laser axis  $s$ :

$$E_u(r, s) = \frac{P_u(r)}{\pi \overline{s_u^2}(r)} e^{-\frac{s^2}{\overline{s_u^2}(r)}} \quad (2.6)$$

where the variance of the unscattered photons along the lateral distance  $\overline{s_u^2}(r)$  is by definition equal to  $\rho_{lr}^2 r^2$ . Also,  $P_u(r)$  expresses the total unscattered energy at distance  $r$ :

$$P_u(r) = P_0 e^{-2\delta(r)} \quad (2.7)$$

Note that the factor of 2 in the exponent is due to the doubled optical depth of the equivalent medium. The main task of the algorithm is to estimate the distribution of forward-scattered photons as a function of  $r$  and  $s$ . It is convenient to normalize the energies by the energy in the unscattered beam such that  $\hat{P} = P/P_u$ . The most convenient way to calculate the variables describing the forward-scattered distributions is to first calculate the corresponding variables for the full distribution,  $\hat{P}(r)$  and  $\overline{s^2}(r)$ . These variables have exact definitions even though it is not implied that the full distribution has a particular form (i.e. Gaussian). The normalized total energy of the combined distribution is given by:

$$\hat{P}(r) = e^{\int_0^r \alpha^c(r') dr'} \quad (2.8)$$

which expresses the fact that forward scattering by cloud particles has the effect of reducing the effective extinction by a factor  $\alpha^c/2$ , leading to  $\hat{P}$  increasing with range. Note that the use of the equivalent medium theorem leads to a doubling of the  $\alpha^c/2$  factor in the exponent.

As shown in [Hogan \(2008\)](#), the energy and variance of the full photon distribution may be calculated exactly (subject to the assumptions stated above). To do this with  $O(N)$  efficiency, it is necessary to keep track of two other variables: the variance of photon direction  $\overline{\zeta^2}(r)$  and a variable representing the covariance of photon position and direction,  $C^{s\zeta}(r)$ . The following differential equations express how these variables vary with range. Considering the geometry in [Fig. 2.8](#), [Hogan \(2008\)](#) shows that the evolution of  $\overline{\zeta^2}(r)$  is governed by:

$$dC^{s\zeta} = \overline{\zeta^2} dr \quad (2.9)$$

The previous equation shows that the evolution of  $C^{s\zeta}$  depends on the evolution of  $\overline{\zeta^2}$ . The propagation angle of photons within the distribution under consideration is only changed by small-angle forward-scattering events because wide-angle scattered photons are lost from the distribution. As the distribution propagates through the medium, the effect of forward scattering is to increase the variance of the photon propagation angle. Thus, the evolution of  $\overline{\zeta^2}$  can be written as:

$$\overline{\zeta^2}(r) = \Theta^2 \alpha^c dr \quad (2.10)$$

where  $\Theta$  is the standard deviation of the scattering angle. Note that in this equation there has been a cancellation of the factor of 2 (which was due to the use of an equivalent medium) by a factor of 1/2 (which was present because only half of the extinguished energy ends up in the forward lobe). Hogan (2008) shows that

$$\Theta = \frac{\lambda}{\pi a_G} \quad (2.11)$$

where  $a_G$  is the equivalent-area radius of the distribution, i.e.:  $\pi a_G^2 = \langle G \rangle$ , with  $\langle G \rangle$  being the mean cross-sectional area of the scattering particles.

The strength of the method is that arbitrary orders of scattering are represented without having to model each of them explicitly. The use of the differential equations 2.9 and 2.10 with their subsequent integration over a range gate allows the description of the several scattering events of a photon within a single range gate. Details on the estimation of the shape of the forward scattered distribution and on the subsequent calculation of the apparent backscatter can be found in Hogan (2008).

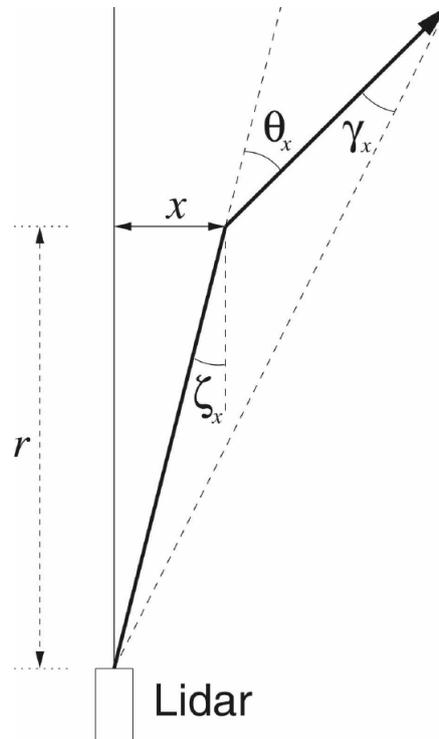


Figure 2.8: Schematic representation of the trajectory of a single photon (thick line) in one of the two directions perpendicular to the lidar axis (thin vertical line). At range  $r$ , the photon has a distance  $x$  from the lidar axis and an angle  $\zeta_x$  with respect to it. It is then scattered by an angle  $\Theta_x$ . If it is subsequently scattered back toward the lidar, the scattering coangle is denoted by  $\gamma_x$  (adapted from Hogan, 2008).

### 2.3.1 A case study

The impact of the MS on the lidar backscatter simulated by ZmVar has been examined on the same case used in WP-1100 (Di Michele et al., 2014) to test the radar MS. The selected situation is a cold front observed by CloudSat and CALIPSO on the 1<sup>st</sup> of January 2007 in the North Atlantic. Figure 2.9 shows CALIPSO observations (granule 2007-01-01T05-19-30ZN) 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 are shown in Fig. 2.10 and Fig. 2.11. 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 2.11 shows that in this portion of the event there is convection activity, generating some solid (convective snow) and liquid precipitation (convective rain). Synthetic cloud backscatter from ZmVar is given in Fig. 2.12, showing both the values on the 20 single subcolumns (top panel) and the average over the model grid boxes (bottom panel). The comparison of the simulated backscatter with the observed ones highlights a remarkable good agreement: the model is able to correctly represent the structure all along the vertical section of the event. There are of course discrepancies, the most noticeable being the less sharp attenuation of the lidar signal along the vertical intensity. This feature can be mainly attributed to the coarser vertical resolution of the model and, of course, to the fact that the simulation does not necessarily exactly reproduce the position of the front structure at the time of satellite overpass.

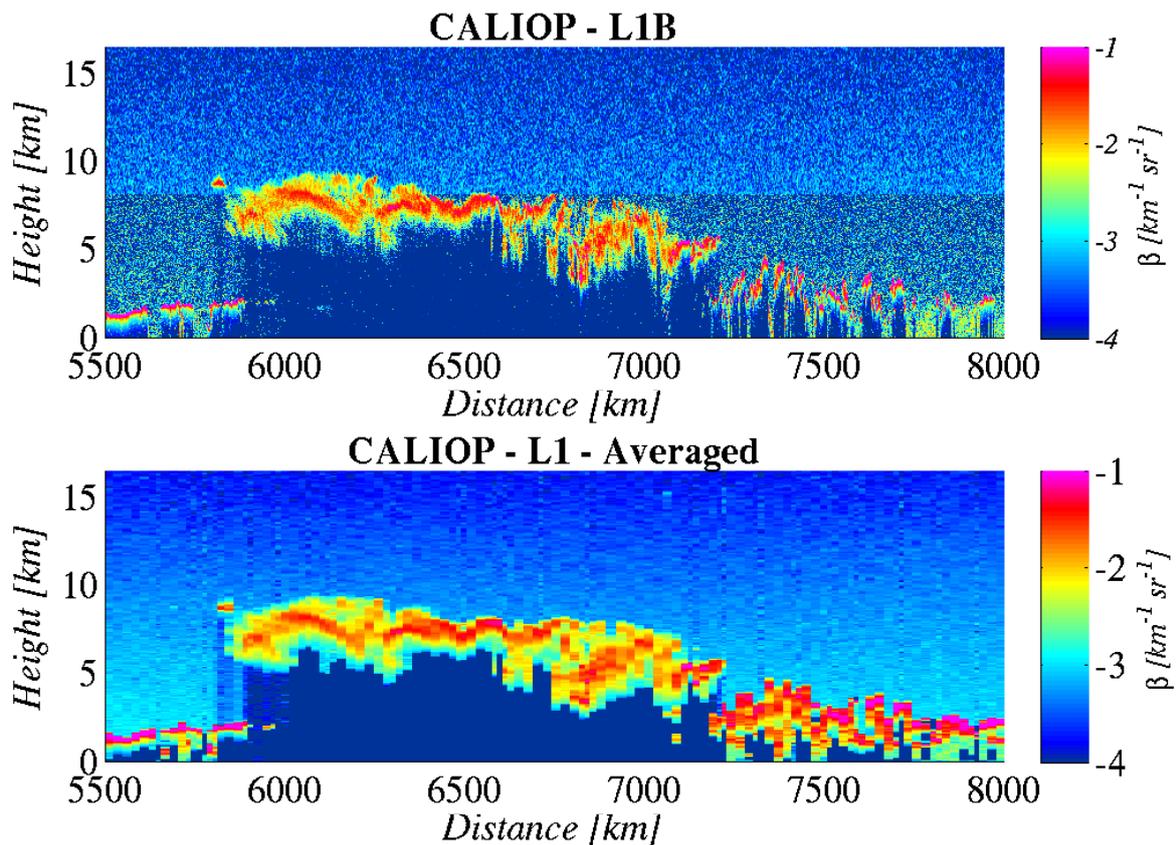


Figure 2.9: 532  $\mu\text{m}$  lidar backscatter from a cloud system as observed by CALIPSO on the 1 January 2007 (panel a). Panel b) shows the same observations after co-locating and averaging to the IFS model resolution.

In order to understand the role of MS on simulated backscatter, the ZmVar has been run again without the wide-angle MS and then switching off both the small-angle and wide-angle MS modules. The backscatter differences between simulations with MS and the ones without are shown in Fig. 2.13. At level of the single subcolumns (top panel) we note a widespread backscatter enhancement. When averaging at the model grid box resolution

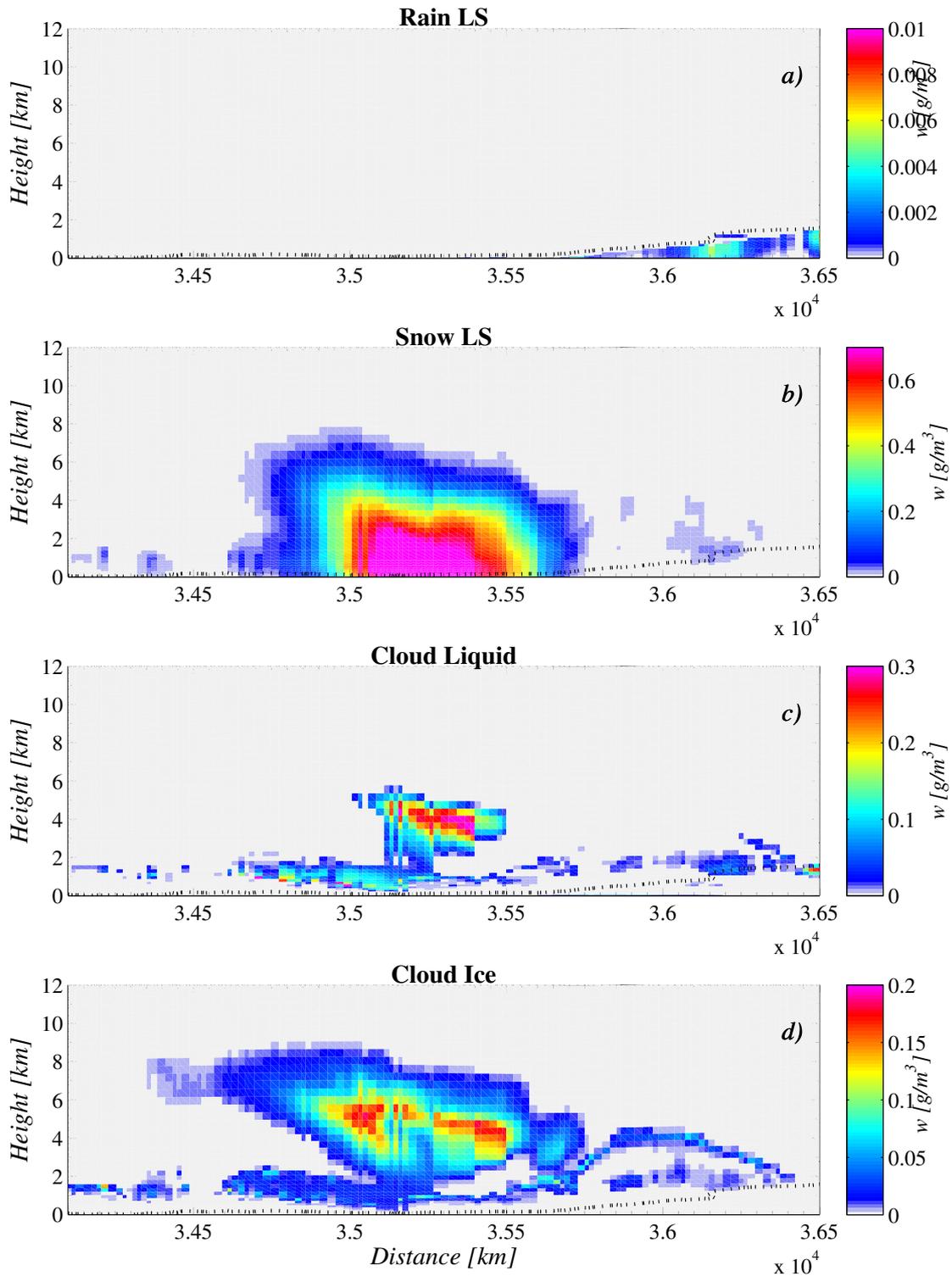


Figure 2.10: Cross section of cloud variables from the IFS model forecast corresponding to the event in Fig. 2.9. From top to bottom: large-scale rain, large-scale snow, cloud liquid, cloud ice. Dashed lines indicate the height of the freezing level.

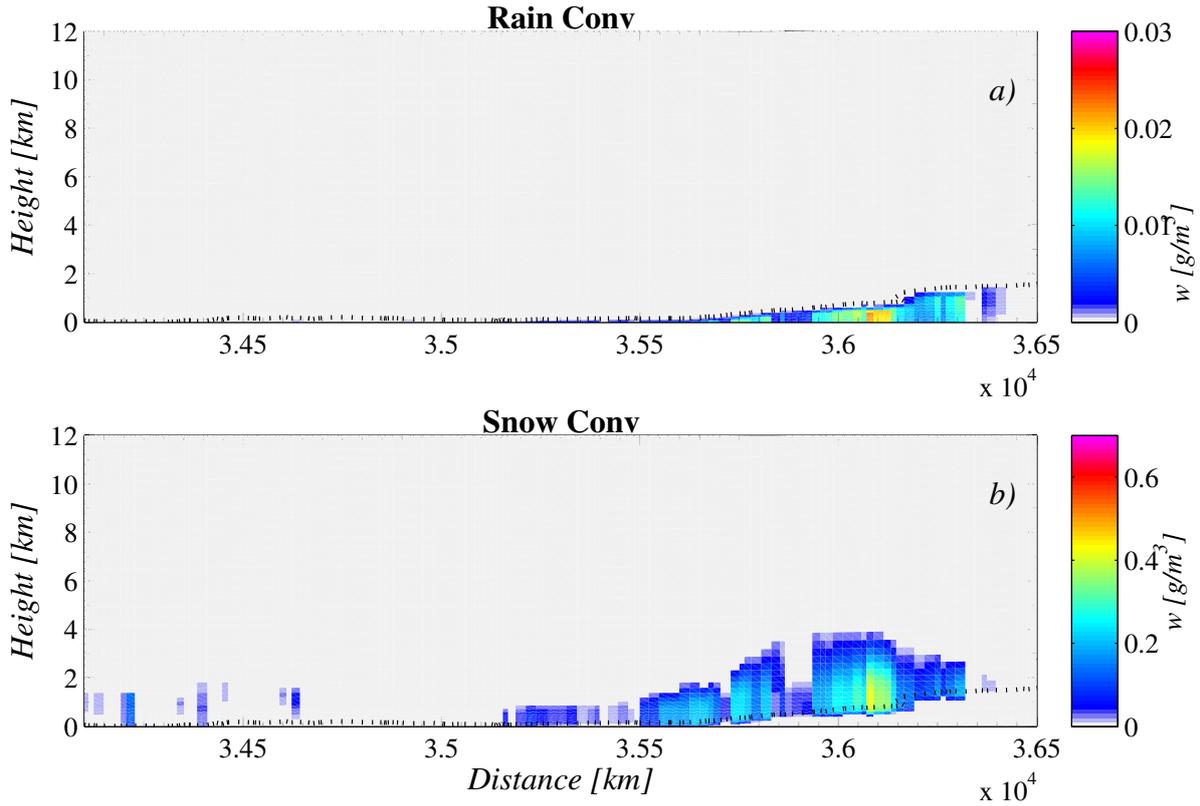


Figure 2.11: As Fig. 2.10, but showing convective rain and convective snow.

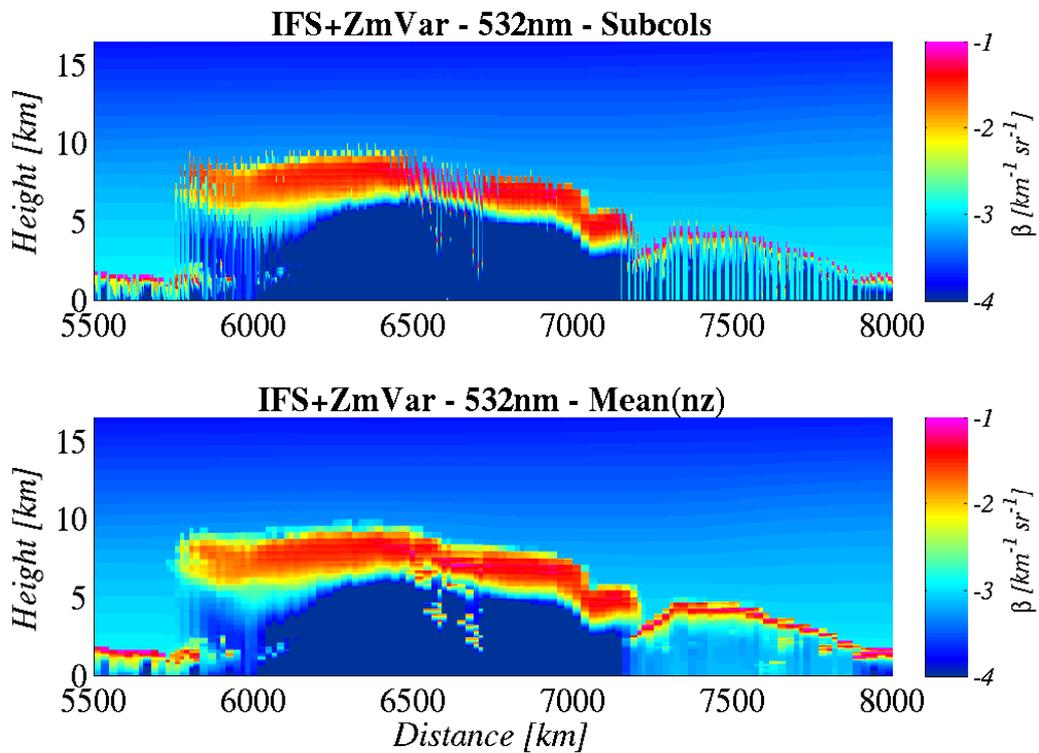


Figure 2.12: ZmVar-simulated 0.532  $\mu\text{m}$  lidar backscatter (including both small-angle and wide-angle MS) relative to the case in Fig. 2.9. Top panel shows the output on each of 20 subcolumns, while bottom panel the grid-box average.

(mid panel), differences have the same qualitative pattern. However, we see few regions where difference is negative. Similarly to the radar, these cases correspond to situations where the pulse stretching occurs (at least along one subcolumn). The additional backscatter generated by MS, when added to the grid box mean, can lead to an overall lower value. This interpretation is confirmed considering the fraction of subcolumns having a backscatter larger than the molecular-only background. The difference between the fractions with and without MS, given in the bottom panel of Fig. 2.13, are indeed positive where differences in averaged backscatter are negative.

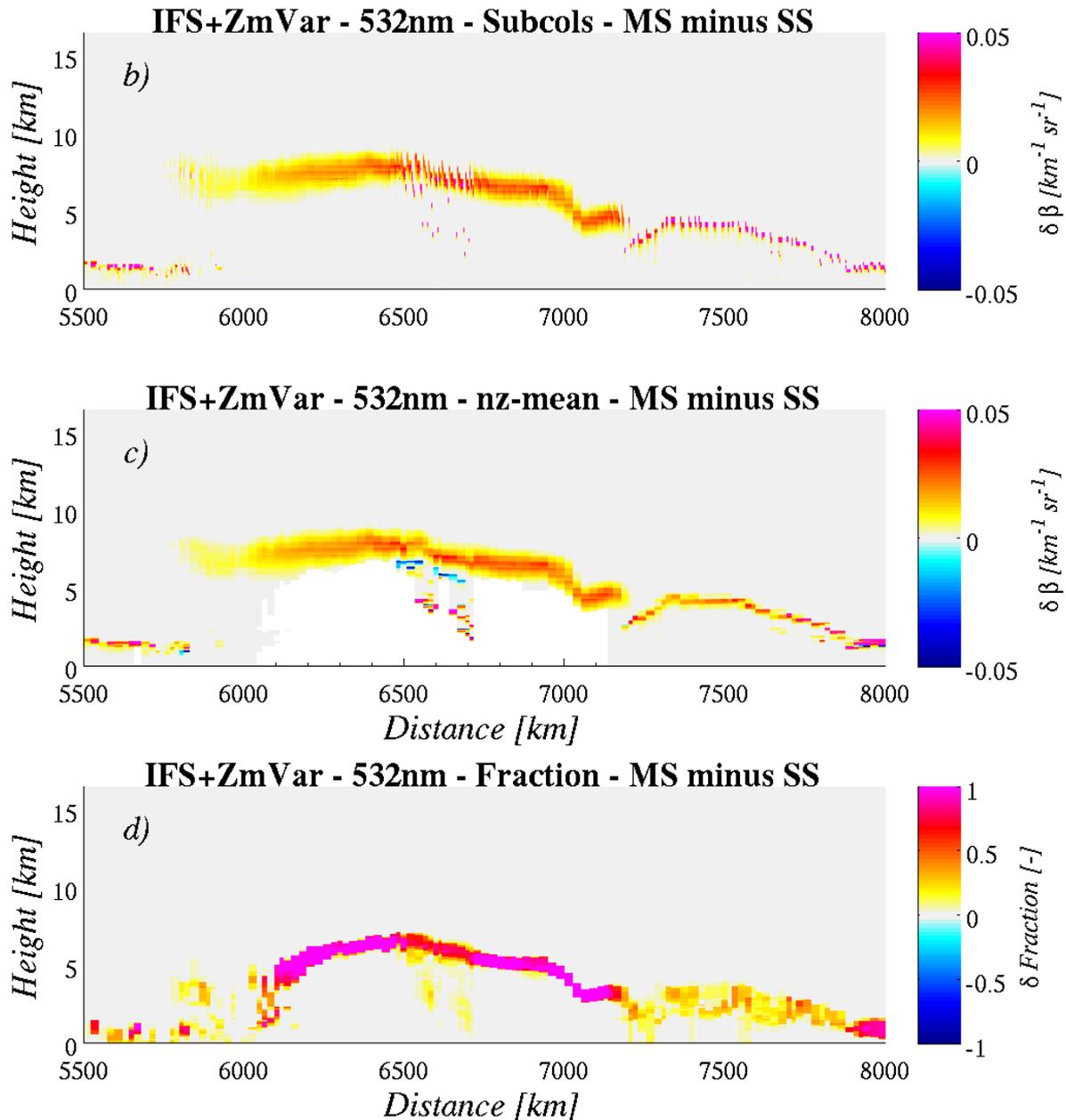


Figure 2.13: Differences using ZmVar with and without MS taken into account. Top panel and mid panels respectively show differences in backscatter coefficient on each subcolumn and on the grid-box average. Bottom panel gives the difference of the grid-box fraction of subcolumns with backscatter larger than the molecular value.

To show the relative role of wide-angle and small-angle MS, a run of ZmVar has been performed where only the small-angle component of the MS has been used. Fig. 2.14 gives the differences between the backscatter

simulated using only the small-angle component of the MS and the same quantity generated without MS. Comparing it to Fig. 2.13 we note that the MS enhancement within the lidar bin is very similar, with the main difference in pattern due to the wide-angle effect of stretching the backscatter signal at levels below the one originating the single scattering signal.

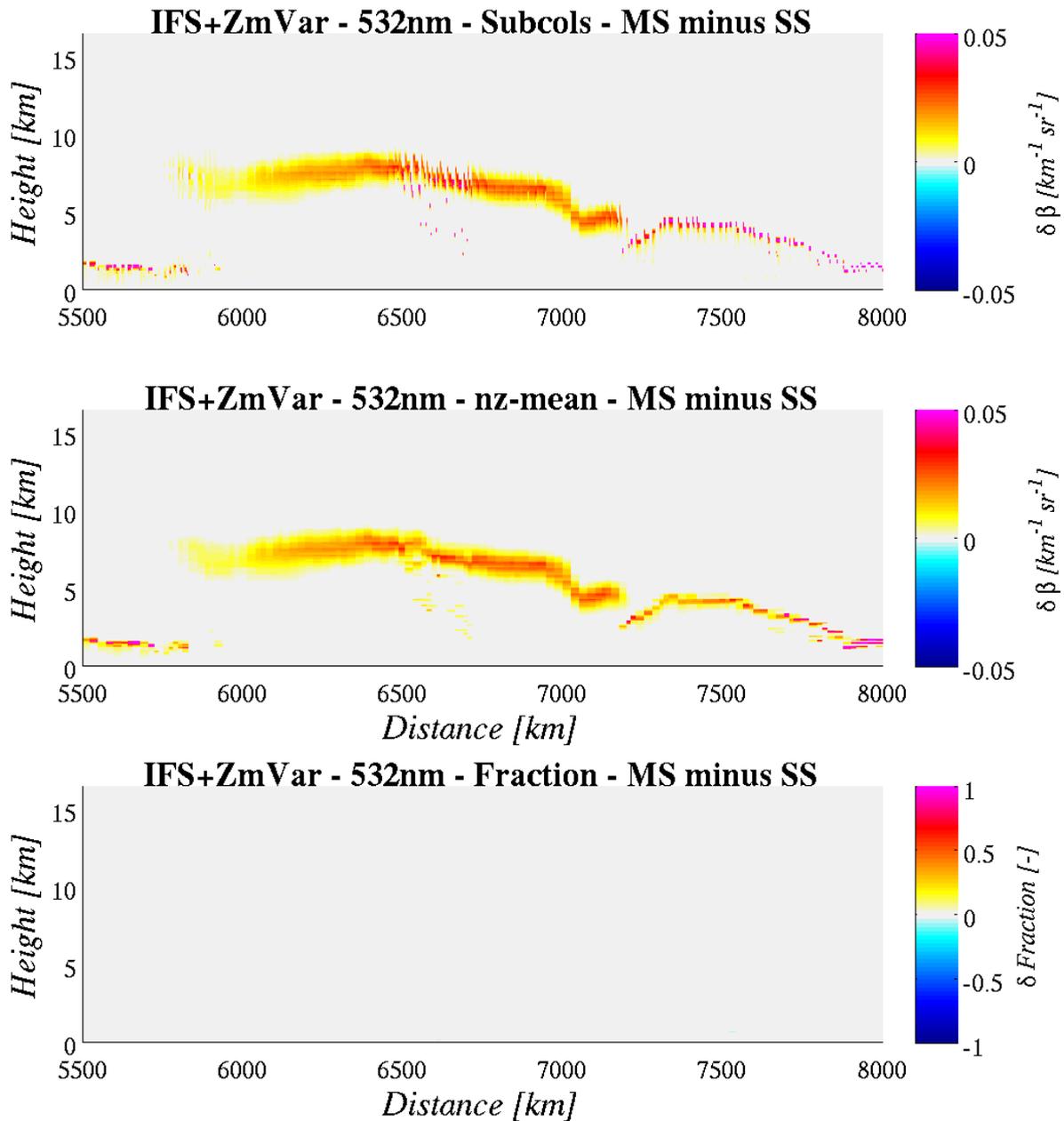


Figure 2.14: As Fig. 2.13, but considering only small-angle (PVC) MS.

### 2.3.2 Small-angle $\eta$ correction

The inclusion of the MS within the ZmVar lidar forward operator allows for an explicit evaluation of the  $\eta$  correction factor of Platt (1973) for the small-angle. This can be done based on Eq. 2.5:

$$\eta(r) = -\frac{1}{2\delta(r)} \ln \left( \frac{\hat{\beta}_d(r)}{\beta(r)} \right) \quad (2.12)$$

where  $\eta$  is the only unknown since  $\delta$ ,  $\beta$  and  $\hat{\beta}_d$  are evaluated by ZmVar. Figure 2.15 shows the occurrence of the retrieved values of  $\eta$  across a range of temperature values. Note that occurrences are normalized by the total number at each temperature interval. Results are presented separately for the cases where the optical depth  $\delta$  is small (i.e.  $\delta < 1$ , panel a) and for those with  $\delta$  larger than one (panel b). When  $\delta < 1$ , the following main features are observed. Below 200 K,  $\eta$  is equal to 0.6 for most of the cases. Between 200 K and 240 K,  $\eta$  gets slighter lower ( $\simeq 0.55$ ), indicating a larger influence of MS at warmer temperatures. Noteworthy, these values are consistent with the ones usually assumed for ice clouds. At temperatures between 240 K and the freezing level,  $\eta$  can have a much wider range of values (up to 0.9). This is a consequence of the presence of supercooled cloud liquid, characterized by a larger asymmetry factor than for ice, i.e. by a wider forward lobe, leading to less efficient small-angle MS (replaced by a more pronounced wide-angle MS). At temperatures warmer than freezing we note the presence of a modal value around 0.8. When considering optically-thicker clouds (panel b) we note a very similar distribution, with differences below 200 K, where  $\eta$  is closer to 0.5 and above freezing, where  $\eta$  has a larger modal value (around 0.65). These findings can help in the choice of the most appropriate value of  $\eta$  in those applications where it is not possible to obtain an exact evaluation of the (small-angle) MS using a method like the PVC due to required computational efficiency (e.g., for variational assimilation).

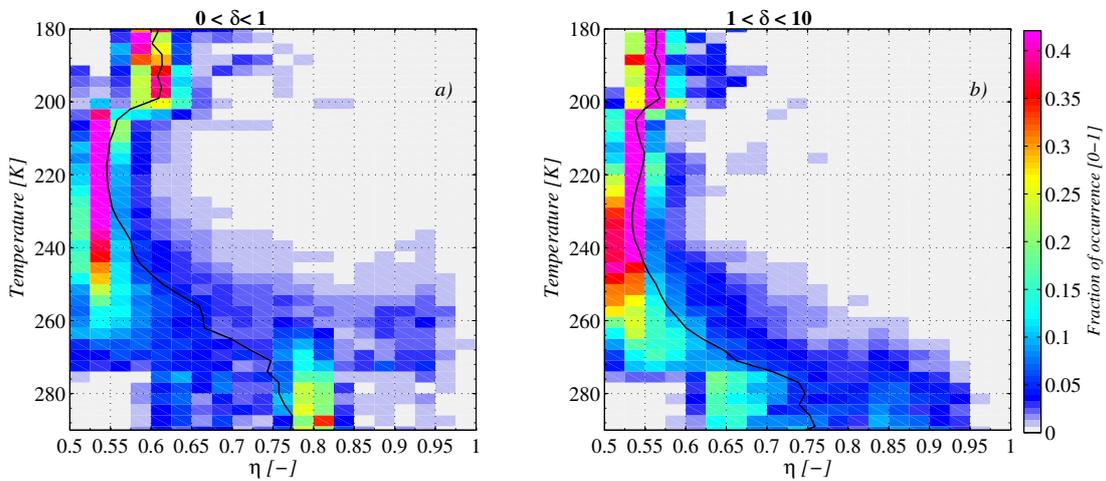


Figure 2.15: Frequency distribution of small-angle correction factor  $\eta$  across a range of temperatures. Each panel contains cases having an optical depth  $\delta$  within the range indicated in the plot title.

### 2.3.3 Wide-angle multiple scattering

Wide-angle scattered photons can be transported outside the field of view or may remain within the field of view and be detected, but with a time delay that makes them appear as being originated at a range beyond the distance to which they actually penetrated. This 'pulse stretching' makes it cumbersome to interpret the signal received since it is not really representative of the associated level. For assimilation, it is therefore reasonable to discard any layers that are contaminated by wide-angle MS, while keeping only the ones at layers above. It is

therefore important to understand the conditions for which the wide-angle MS gives a significant contribution to the measured backscatter. This would allow the definition of appropriate criteria for screening these situations. To do this, we have considered the lidar backscatter (at  $0.532\mu\text{m}$ ) simulated using ZmVar with either both small- and wide-angle modules or only with small-angle. For this, a dataset consisting of approximately 7000 cloud profiles from IFS have been used. For each profile, we have compared the backscatter using both MS components with the one due to small-angle focusing attention at the first layer (from cloud top) affected by the wide-angle MS (if any).

Figure 2.16 gives the fractional increase of the backscatter simulated using both small- and wide-angle MS relative to the one generated considering only the small-angle MS. This quantity is plotted against the corresponding optical depth  $\delta$ , with the hydrometeor mass content in the layer indicated by the dot colours. The left panel of Fig. 2.16 shows that, when considering cold-phase clouds, the wide-angle MS enhances the small-angle by less than 20% most of the times. Figure 2.16 (right panel) presents a similar plot where only warm-phase clouds are considered. In this case, we note the increase (by a factor of 10) in the axis scale due to the much larger impact of wide-angle MS. In fact, the fractional increase can be larger than 50% already for optical depths less than one (or mass contents smaller than  $0.1\text{g m}^{-3}$ ). For larger optical depths, fractions larger than 100% can be obtained because the high attenuation emphasizes this phenomenon. These results confirm that the lidar returns from water phase clouds are not only highly attenuated, but they are also deeply affected by wide-angle MS. The relationships in Figure 2.16 allow the definition of a condition to identify the first (from top) lidar range 'corrupted' by wide-angle MS.

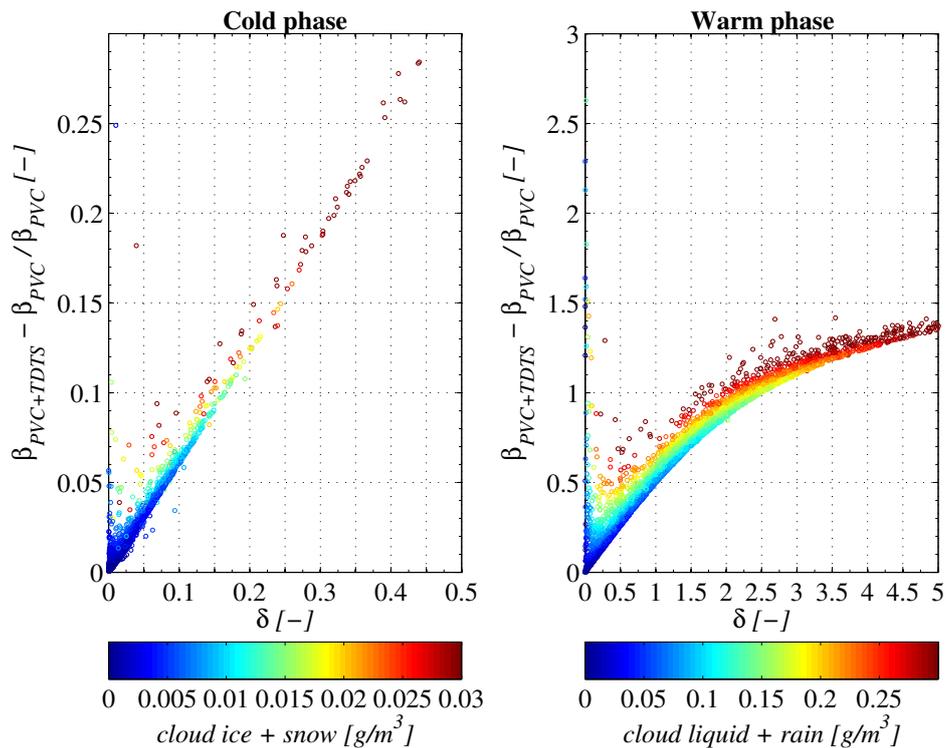


Figure 2.16: Fractional increase of attenuated backscatter (at  $0.532\mu\text{m}$ ) due to wide-angle (TDTS) MS with respect to small-angle only (PVC) as function of optical depth. Plots are based on a dataset consisting of approximately 7000 profiles considering only the first layer (from cloud top) affected by the wide-angle MS. Left panel contains only cold-phase cases while the right panel is for warm-phase cases. Colour of dots indicates the hydrometeor content amount according to the scale below each plot.

### 2.3.4 Computational cost of multiple scattering

The additional calculations required to take into account MS, result in a non-negligible increase in computational cost. This is quantified in Tab. 2.2, where values are relative to a run consisting of 900 profiles (and 20 subcolumns) on a machine with an Intel Duo 3GHz CPU. We note that the most important additional cost (+590%) comes from the process of vertical regridding. The substantial increase is due to the large number of levels (L337) needed to match the lidar vertical resolution (60 m - up to 20 km, consistently with CALIPSO). The regridding is a necessary preliminary step because the MS module works under the hypothesis of atmospheric profiles uniformly spaced along the altitude, while the IFS model values are provided on variable pressure levels. The small-angle MS module adds only 2% to the cost of ZmVar. This is expected since the 'fast' PVC method has time complexity  $O(N)$ . The wide-angle MS is implemented using the time-dependent two stream method (TDTS), which is a  $O(N^2)$  algorithm. Given the larger complexity, it doubles the cost (+133%) when added to the regridding-plus-PVC.

Test	Total (sec)	Incremental difference (%)	Diff. (sec)	Incremental difference (sec)
Single Scattering, No regrid	44	-	-	-
Single Scattering, L337 regrid	305	+590	261	261
MS PVC, Regrid, L337 regrid	312	+2	268	7
MS TDTS, MS PVC, L337 regrid	712	+133	668	400

Table 2.2: Computational cost of the multiple scattering.

## 2.4 Summary from operator developments

The ZmVar forward operator has been extended to simulate lidar observations in clouds starting from the ECMWF model variables. Comparisons in terms of CFADs with data from the CALIPSO lidar showed that simulated backscatter is in reasonable agreement with real observations. Important components of the new version of ZmVar are the modules for the simulation of MS. This new feature has been used to investigate the effects of MS in lidar signal.

## 3 Lidar observation errors

Observation-related errors are represented by the matrix  $\mathbf{R}$ , which contains all errors related to the interpretation of observations within the data assimilation framework. Observation errors include instrument errors, errors in the observation operator, and errors of representation, i.e. the ones resulting from the mismatch of scales in the horizontal or vertical between the observations and the model. Under the hypothesis of uncorrelated errors,  $\mathbf{R}$  can be written as sum of covariances of the error sources mentioned above:

$$\mathbf{R} = \mathbf{E} + \mathbf{F} + \mathbf{O} \quad (3.1)$$

where each matrix represents the following types of errors:

- $\mathbf{O}$  Basic observation error, i.e. the instrumental noise;
- $\mathbf{E}$  Representativity error, i.e. errors from mismatched scales between observation and model;
- $\mathbf{F}$  Forward modelling errors, i.e. errors in the observation operator.

The approach to evaluate the instrument errors arising in a lidar instrument and the ones due to lidar forward modelling will be described in Subsection 3.1. The representativity errors of the CALIPSO lidar will be discussed in Subsection 3.2.

### 3.1 Forward modelling and instrument errors

#### *CALIPSO instrument errors*

Following the work of Liu et al. (2006), the standard deviation  $\Delta\beta$  of random errors (due to shot noise) in the measured lidar backscatter  $\beta$  can be expressed as:

$$\Delta\beta = \left\{ NSF^2\beta + \left(\frac{r^2}{C}\right)^2 [(\Delta V_b)^2 + (\Delta\bar{V}_b)^2] \right\}^{\frac{1}{2}} \quad (3.2)$$

where  $r$  is the distance (in meters) of each lidar gate from the satellite and  $C$  is the lidar calibration constant such that the signal power  $V$  is:  $V = C\beta r^{-2}$ . NSF is the Noise Scale Factor representing the effect of the photomultiplier tube to increase the noise above what would be expected purely from Poisson statistics. All these quantities are included in CALIPSO Level 1B product.

$\Delta V_b$  in Eq. 3.2 is the standard deviation of the background signal power. It can be evaluated using only those ( $N$ ) measurements affected by molecular backscatter (e.g. using the ones above 30 km).  $\Delta\bar{V}_b$  is the standard deviation of the mean background signal and therefore:  $\Delta\bar{V}_b = \Delta V_b / \sqrt{N}$ . It should be noted that when the measured backscatter values are obtained as an average of  $n$  samples onto a lower resolution grid (e.g. 60 m), the standard deviation of random errors for the averaged measurements  $\Delta\bar{\beta}$  is obtained from the original ones  $\Delta\beta$  as:

$$\Delta\bar{\beta} = \frac{1}{n} \sqrt{\sum_{i=1}^n \Delta\beta_i^2} \quad (3.3)$$

#### *Forward modelling errors*

In Section 3.1 of WP-1100 (Di Michele et al., 2014), errors in forward-modelled radar reflectivities were estimated based on ambiguities in the microphysical assumptions. The same method can be employed for the quantification of the errors in the lidar forward operator. Considering the setup shown in Table 2.1 as a baseline, a number of perturbed states can be defined replacing the original PSD or shape/density with possible alternatives. The new configurations can then be used as lookup-tables to run ZmVar on a large-enough dataset of cloud profiles from the IFS. Based on these runs, statistics of forward modelling errors can be obtained across different ranges of backscatter and temperatures.

Figure 3.1 shows the standard deviation of the differences between the perturbed and reference backscatter as function of the latter. Values are given in absolute terms (panels on the left column) and percentage of the reference backscatter (panels on the mid column). Results are shown considering the perturbations of each hydrometeor separately. Results have been stratified across six ranges of temperatures (panels from top to bottom). In addition, panels on the right column gives the relative occurrence of the backscatter value. The main feature is the large uncertainty in backscatter that cloud ice perturbations provide. The percentage standard deviation can reach 60% at colder temperature, gradually decreasing (down to 20%) at warmer ones. We note that, below 238 K, the cloud ice percentage standard deviations increases monotonically, indicating an exponential growth as the backscatter increases.

The overall uncertainty on the backscatter is shown in Fig. 3.2 where, again separating over temperature ranges, the total standard deviation from all hydrometeors are expressed as percentage of the corresponding reference value of backscatter. We note that these values are similar to the one of the cloud ice only, which having much larger standard deviation than other hydrometeors, drives the total. This result emphasise the importance of having a forward operator with an accurate definition of the cloud ice microphysics, especially for the cloud upper portions.

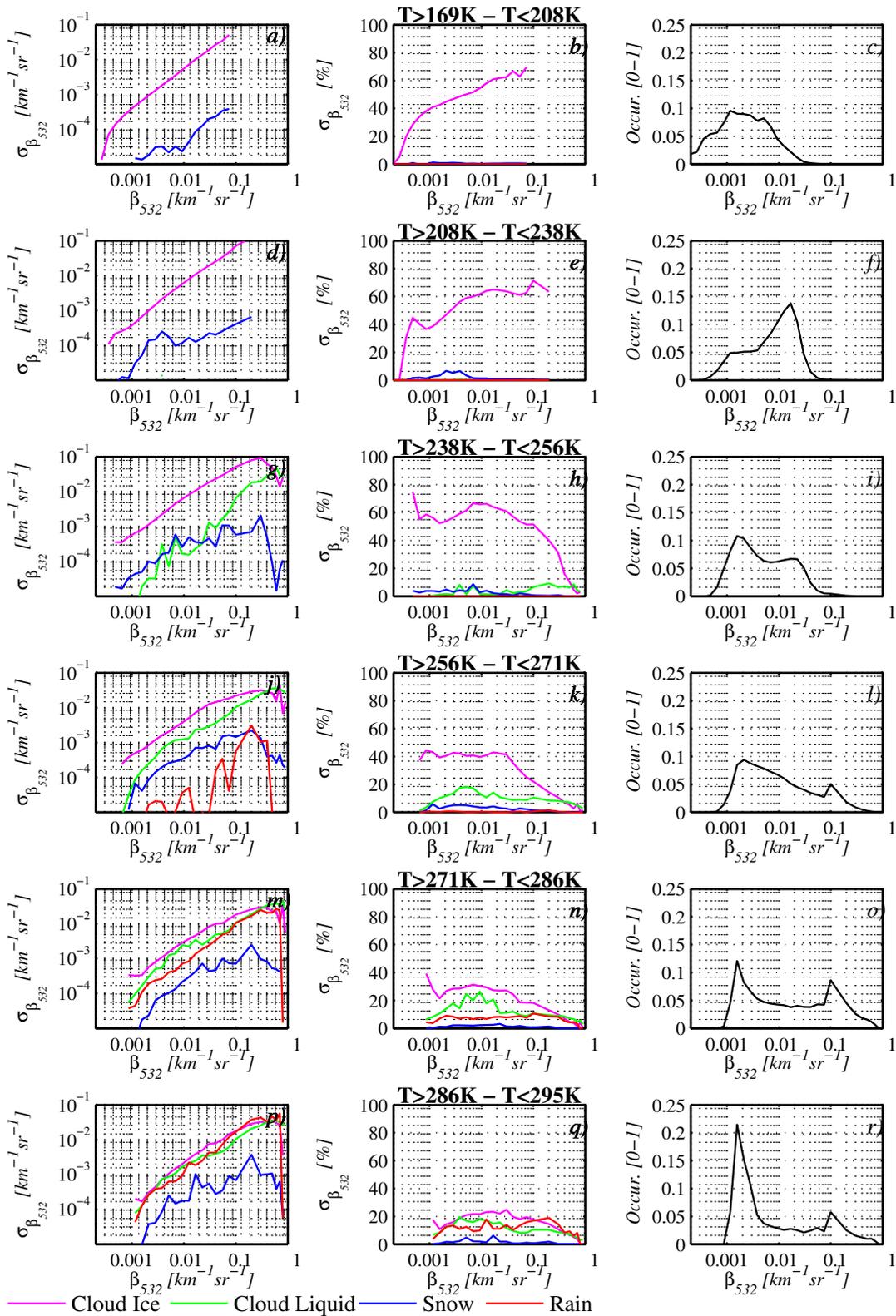


Figure 3.1: Forward modelling uncertainties in ZnVar simulated reflectivity expressed in  $\text{mm}^6\text{m}^{-3}$  and dBZ (left and central panels, respectively). Curves give the contribution given to uncertainty by each single hydrometeor. Right panels show the occurrence of reflectivity values. Each row refers to the temperature range shown in the title.

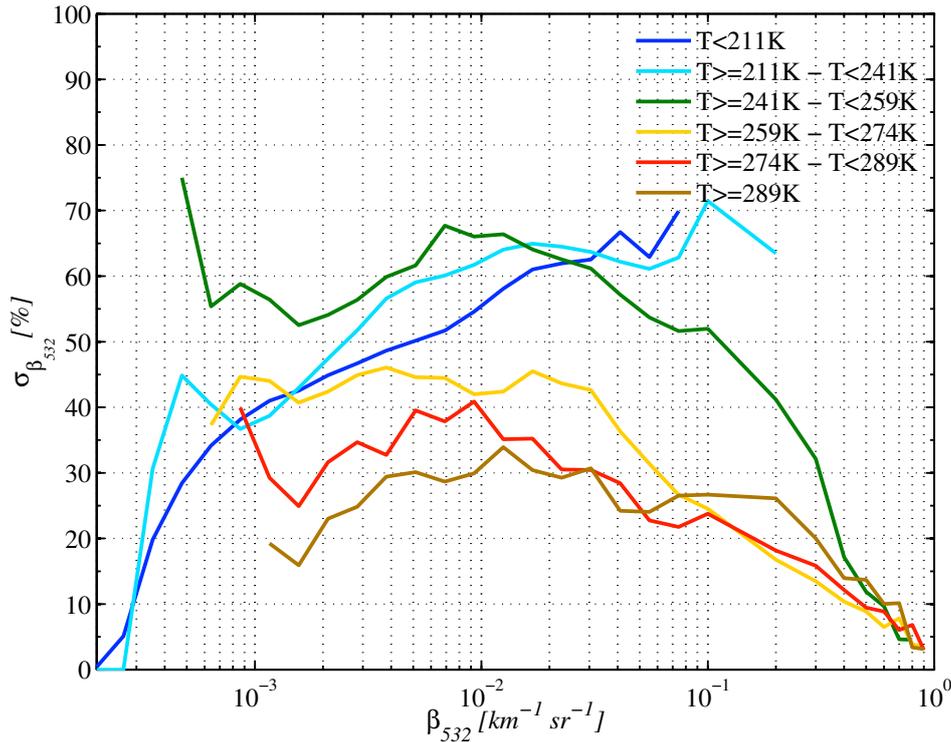


Figure 3.2: Summary of forward modelling uncertainties in ZmVar expressed as percentage of the simulated reflectivity when perturbations of all hydrometeors are taken into account. Each curve refers to the temperature range shown in the legend. For comparison, the CloudSat instrument random error is also given (black lines), considering measurements at the original resolution (dashed) and averaged over 30 samples (solid).

### 3.2 Representativity errors of lidar observations

CALIPSO lidar observations give the magnitude of the signal backscattered by atmospheric hydrometeors and aerosols. Contrary to CloudSat radar measurements, CALIPSO lidar observations can become totally attenuated in the presence of very thick clouds (typically for optical thickness above 3) like in tropical convective towers or frontal systems (i.e., the base of the clouds is not retrieved, as well as their geometrical thickness). All atmospheric features below the level of total attenuation of the lidar signal are therefore unseen.

CALIPSO lidar observations have a footprint of 70 meters, which is very small when compared to the size of the model gridbox (e.g. horizontal resolution with spectral truncation T799 used in this study corresponds approximately to grid resolution of 25 km). The poor horizontal coverage of the measurements raises representativity issues that do not allow a straightforward comparison between these observations and analogue parameters simulated by NWP models. Therefore for space-borne lidar observations it is important to estimate the representativity errors related to them.

The methodology, developed during the QuARL project (Section 3 of [Di Michele et al., 2009](#); [Stiller, 2010](#)) and afterwards applied to estimate the representativity errors from CloudSat radar measurements (Section 3.2 of WP-1100 report, [Di Michele et al., 2014](#)), is also applied to CALIPSO lidar observations. In Section 3.2.1, the main steps of the methodology to assess the representativity errors of CALIPSO lidar observations are explained. Section 3.2.3 presents the results of the approach for the narrow-field of view (FOV) of space-borne CALIPSO lidar observations. Finally, Section 3.2.4 illustrates some examples for the estimate of the representativity error of lidar measurements by selecting two scenes of CALIPSO observations.

### 3.2.1 Methodology

The methodology to estimate the representativity error of observations has been already explained and described in Di Michele et al. (2014) (Section 3.2.2 of WP-1100 report). Therefore only a brief overview is provided here. Here, this methodology is used to assess the representativity error of CALIPSO lidar backscatter measurements. Fig.3.3 summarizes the different steps of the method which is based on the generation of synthetic data which shares very important statistical properties with the original quasi-unidimensional observations. These synthetic data are generated by stochastic modelling techniques. A quasi-empirical relationship is then established between the Structure Function Maximum (SFM) score and the synthetic data to have an estimation of the representativity error.

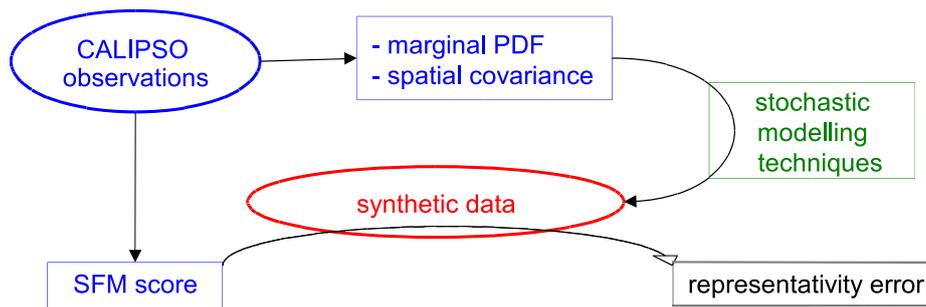


Figure 3.3: Schematic description of the main steps for the retrieval of the representativity error from CALIPSO lidar observations.

The SFM score is a statistical measure comprising information from a larger range of spatial scales that can be computed from the satellite track at any vertical level and is defined as:

$$SFM(x_0, h, L_0) = \max \left[ \gamma(x_0, h, (k + 0.5) L_0) \right]_{k=\{0,1,\dots,N_{max}\}} \quad (3.4)$$

where  $\gamma$  is the structure function of any random field  $q(\vec{x})$  at the location  $x_0$  and for a displacement  $h$ , considering  $k$  neighbourhoods (up to  $N_{max}=10$ ) of size  $L_0$  around  $x_0$  ( $x_0$  corresponds to the horizontal position of the centre point for the intersect between the satellite track and the model grid box of interest).

### 3.2.2 Comparison between observed and synthetic backscatters

One of the important keys of the method is the good agreement between the statistical properties of observed and synthetic backscatter data at every sampled vertical level. In this section, we will use 1 month of CALIPSO data (January 2007) from the product CAL\_LID\_L2\_05kmCPro-Prov over the tropical belt ( $\pm 27^\circ$ ) and extra-tropics ( $30-70^\circ\text{N}$ ). Quality, status and masks are taken into account in order to select only cloud layers.

Figures 3.4 and 3.5 show respectively the marginal probability distribution function (PDF) and the correlation function of observed lidar backscatters (in  $\text{km}^{-1}\text{sr}^{-1}$ , black) and synthetic lidar backscatters (in  $\text{km}^{-1}\text{sr}^{-1}$ , red) for the Tropics ( $\pm 27^\circ$ ) and extra-tropics ( $30-70^\circ\text{N}$ ) at an altitude of 2 km. For both geographical areas, the comparison between the observed and the generated synthetic datasets indicates a good agreement. However, it has to be noted that the PDFs of the synthetic data show always less occurrences for the large backscatters, which implies that the very first bins of backscatter (i.e. the smallest values of backscatter) show more samples in the synthetic data than in the observed data. The comparison between the correlation functions of observed and synthetic backscatters shows a very strong agreement both in the Tropics and in extra-tropics.

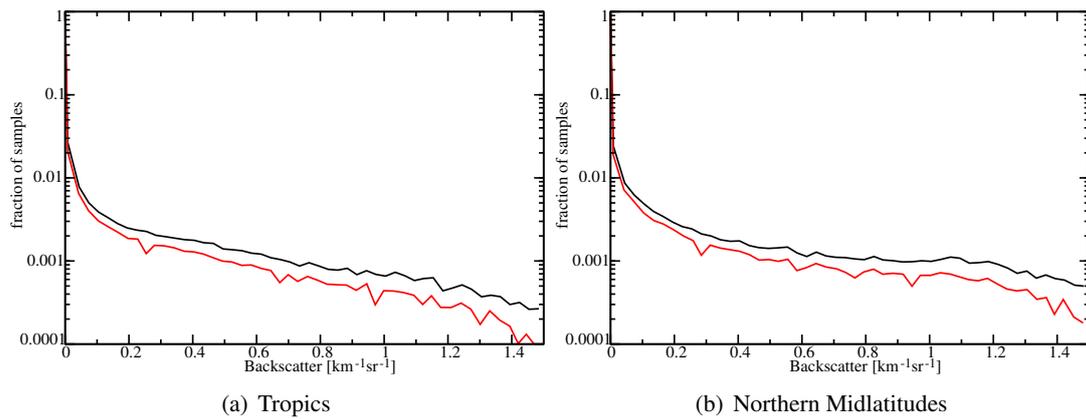


Figure 3.4: PDF of observed (black) and generated synthetic (red) lidar backscatter (in  $\text{km}^{-1}\text{sr}^{-1}$ ) in the Tropics ( $\pm 27^\circ$ , left) and in mid-latitudes North ( $30 - 70^\circ\text{N}$ , right), at an altitude of 2 km for the period of January 2007.

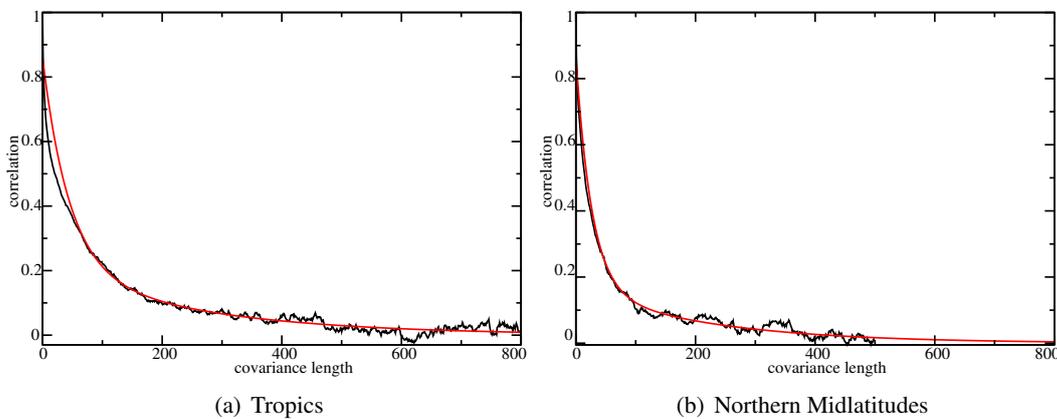


Figure 3.5: Correlation function of the observed (black) and synthetic (red) lidar backscatter in the Tropics ( $\pm 27^\circ$ , left) and in mid-latitudes North ( $30 - 70^\circ\text{N}$ , right) at an altitude of 2 km for the period of January 2007.

Fig.3.6 displays the PDF of the Structure Function Maximum (SFM) score of the observed (black) and synthetic (red) backscatter in the Tropics (left) and extra-tropics (right) at 2 km in January 2007. The comparison between both distributions shows the strength of the method in order to generate synthetic data with the same main statistical properties than the observations as indicated by apparent similarity of the distributions.

To conclude, the comparison of some important statistical properties of observed and synthetic backscatters reveals similar features and validates the fact that the synthetic data can be used to estimate the sampling error by mean of the SFM score.

### 3.2.3 Estimating the representativity errors

To assess the sampling error of CALIPSO lidar backscatter measurements, synthetic analogue data have been generated by the method described previously while using the statistics retrieved for two one-month periods (January and July 2007). Only the results for January 2007 are shown in this study. Averages of the observed backscatter are computed on line segments which corresponds to a portion of the orbit track. At the same time, averages of the synthetic backscatters have been computed in the corresponding co-located gridboxes. The line segments are then partitioned into bins of the SFM score, and the Root-Mean Square Error (RMSE) is

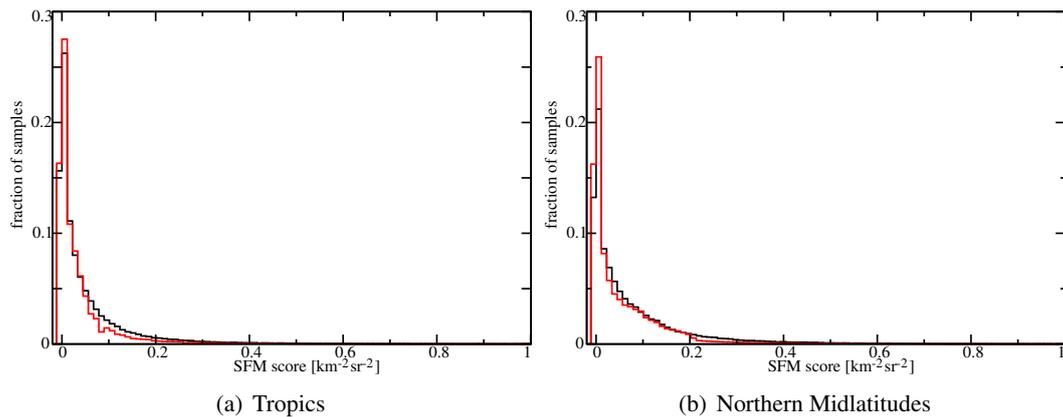


Figure 3.6: PDF of the Structure Function Maximum (SFM) score (in  $\text{km}^{-2}\text{sr}^{-2}$ ) for observed (black) and synthetic (red) lidar backscatter data in the Tropics ( $\pm 27^\circ$ , left) and in mid-latitudes North ( $30 - 70^\circ\text{N}$ , right) at an altitude of 2 km for January 2007.

computed as the RMS difference between line and grid box averages of the considered backscatter in each bin. In this way, the relationship between the SFM score and the sampling error is retrieved.

SFM scores are computed along the satellite track over periods of one month in order to get an empirical relationship between this score and observation-related representativity error for different heights. This dependence with height is presented in Fig.3.7 for the Tropics and in Fig.3.8 for Northern midlatitudes at four different altitudes (2, 6, 10 and 14 km) for January 2007. Each panel shows the RMS error (black), the bias (red) and the standard deviation (green) as a function of the score. It is interesting to notice that the shape and magnitude of each curve do not change so much with latitude. The RMSE curves show almost the same behaviour at 2 and 6 km in both geographical areas with errors reaching  $0.2 \text{ km}^{-1}\text{sr}^{-1}$ . For altitudes corresponding to the mid and high-troposphere, the higher the level, the smaller is the magnitude of the error: RMSE can peak up to  $0.15 \text{ km}^{-1}\text{sr}^{-1}$  at 10 km in Tropics and extra-tropics while it hardly reaches  $0.05 \text{ km}^{-1}\text{sr}^{-1}$  at 14 km in both areas. The high values of RMSE are reached for smaller values of scores at higher altitudes than at lower altitudes. This suggests that there is more homogeneity in cloudy patterns in the high troposphere than in the middle or low troposphere (where, most of the time, the backscatter signal is totally attenuated).

### 3.2.4 Application to a case study

As an illustration of the approach described here above, this part of the report aims at applying the method to CALIPSO lidar observed scenes in order to retrieve their representativity error. Two scenes of CALIPSO lidar measurements from the same observation day (3<sup>rd</sup> January 2007) are selected:

- a portion of the satellite track between  $30^\circ\text{S}$  and  $70^\circ\text{S}$  (between 20:50 and 21:05 UTC), over the Eastern Pacific Ocean.
- a tropical portion of the satellite track between  $27^\circ\text{N}$  and  $27^\circ\text{S}$  (between 10:23 and 10:40 UTC), over Central Pacific Ocean.

Both scenes are represented in the top panel of Figures 3.9 and 3.10 displaying profiles of the backscatter coefficient (in  $\text{km}^{-1}\text{sr}^{-1}$  using logarithmic scale). The representativity error is estimated at 4 different altitudes (2, 6, 10 and 14 km – represented by the horizontal dashed lines on the top panels) and is shown in the bottom panels.

The top panel of Fig.3.9 shows the scene in the mid-latitudes North. The signal is totally attenuated under a certain altitude ranging from 2 to 8 km for profiles between  $\sim 3480$  and  $4180$ . The total attenuation of the signal

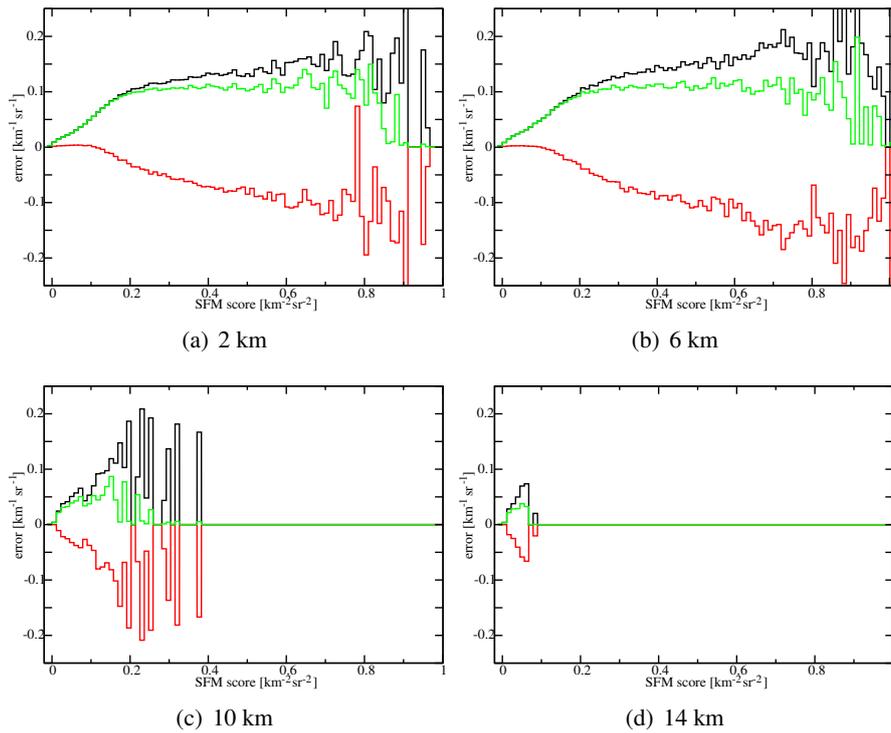


Figure 3.7: RMSE (black), bias (red) and variance (green) of CALIPSO synthetic data in SFM score bins (in  $\text{km}^{-2}\text{sr}^{-2}$ ) at an altitude of 2 km (a), 6 km (b), 10 km (c) and 14 km (d) for January 2007 in the Tropics ( $\pm 27^\circ$ ).

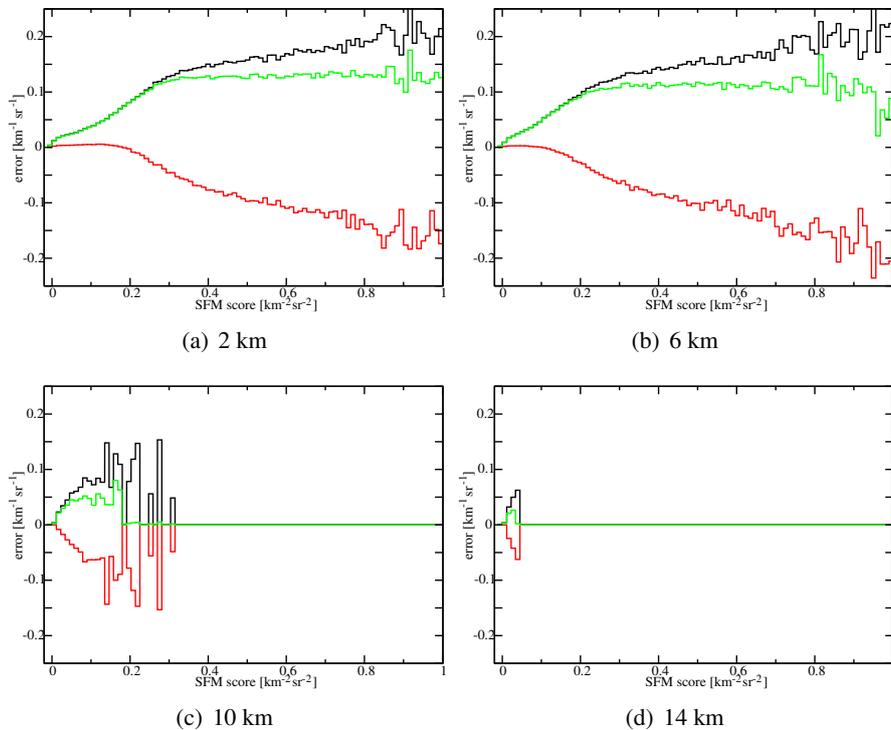


Figure 3.8: Same as Fig.3.7 but for midlatitudes North ( $30 - 70^\circ\text{N}$ ).

is also seen and confirmed in the simulated background (middle panel). The magnitude of the representativity error varies with altitude, with larger variations of the representativity error at lower altitudes: indeed, at an altitude of 2 km, the error ranges from 0.05 to  $0.18 \text{ km}^{-1} \text{ sr}^{-1}$ , while, at 10 and 14 km, the error is constant and smaller ( $\sim 1-4 \cdot 10^{-4} \text{ km}^{-1} \text{ sr}^{-1}$ ).

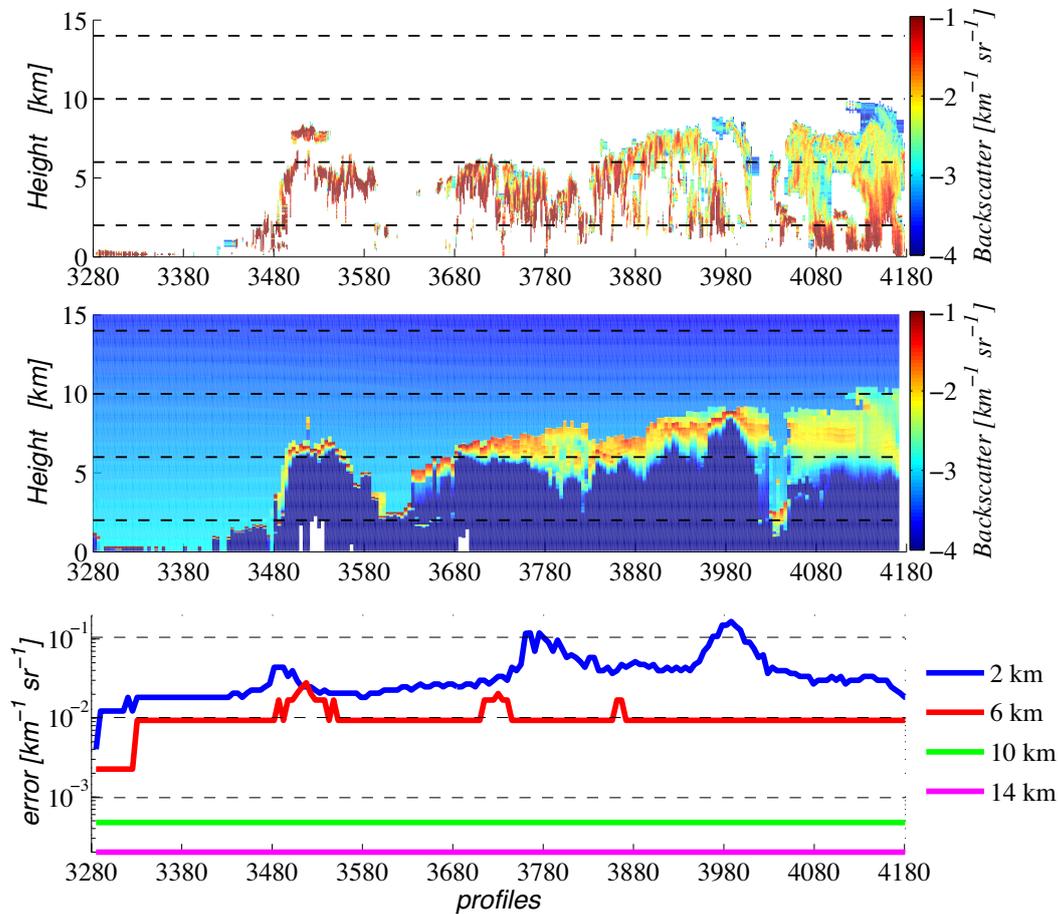


Figure 3.9: Top panel: CALIPSO lidar backscatter (in  $\text{km}^{-1} \text{ sr}^{-1}$ ) observed in the mid-latitudes North ( $30^\circ - 70^\circ$ ) on 3<sup>rd</sup> January 2007 (orbit 20-43-33ZD). Middle panel: corresponding simulated lidar backscatter (in  $\text{km}^{-1} \text{ sr}^{-1}$ ) retrieved by ZmVar. Low panel: corresponding estimate of the representativity errors of CALIPSO backscatter (in  $\text{km}^{-1} \text{ sr}^{-1}$ ) at an altitude of 2 km (blue), 6 km (red), 10 km (green) and 14 km (magenta) represented by horizontal black dashed lines in the top and middle panels.

In the tropical scene displayed in the top panel of Fig.3.10, the magnitude of the representativity error reaches values up to  $0.12 \text{ km}^{-1} \text{ sr}^{-1}$  for an altitude of 2 km and is  $\sim 10$  times smaller for higher altitudes. The sampling error of the high clouds detected at an altitude 10–15km between profiles 2040 and  $\sim 2400$  is around  $0.004 \text{ km}^{-1} \text{ sr}^{-1}$ . The comparison between observed and simulated high ice clouds shows some differences that can be linked to Fig.2.7 where the lack of high ice clouds in the model is illustrated.

### 3.3 Conclusions

Observation errors associated to CALIPSO backscatter measurements have been evaluated considering the different components, i.e. the instrument, forward modelling and spatial representativity. The CALIPSO 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

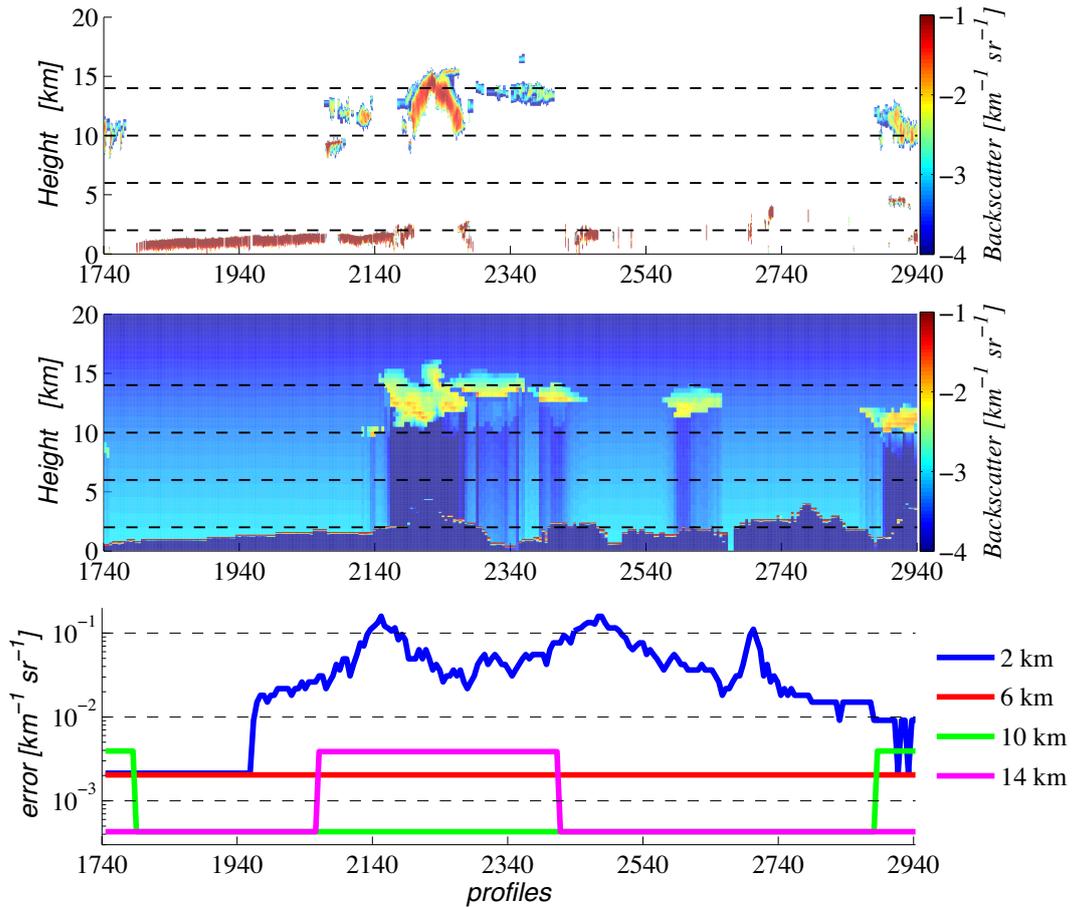


Figure 3.10: Same as in Fig.3.9 but for the CALIPSO track on 3<sup>rd</sup> January 2007 (orbit 10-03-52ZN) in the Tropics ( $\pm 27^\circ$ ).

parameters. The representativity error has been assessed by applying to CALIPSO data the statistical method developed during the QuARL project. 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.

## 4 Development of quality control and bias correction schemes for the lidar

As part of the preparations for the assimilation of lidar data in the ECMWF forecast model, a preprocessing of this new type of observations must be developed in order to have data satisfying the linearity and no-bias conditions.

The data from the CALIPSO lidar have been used to prepare the tools for data processing to be performed prior to assimilation. The first step of data processing consists in a quality control of the data which aim is to discard those observations which could degrade performance of monitoring/assimilation system (Section 4.1). In the second step, the bias correction scheme developed for lidar backscatter and described in Section 4.2 is used. These steps follow concepts similar to the ones presented in Section 4 for CloudSat radar data in WP-1100 report (Di Michele et al., 2014).

### 4.1 Quality control of CALIPSO lidar observations

As done in Subsection 4.1 for CloudSat in WP-1100 report (Di Michele et al., 2014), the conditions for the screening have been chosen based on direct comparison of simulations against observations. Figures 4.1 and 4.2 show the scatterplots of CALIOP measurements (at 532nm) plotted against the simulated cloud backscatter, considering observations at Southern mid-latitudes (30°S-60°S) and in the tropics (30°S-30°N), respectively. Plots are built using a dataset of matched observations and simulations for 30 days in January 2007 using cases over ocean only. As for CloudSat, the comparison is done considering observations separated into six altitude ranges. Figure 4.1 shows that above 10 km (panel a), there is a good correlation all through the backscatter range, despite the logarithmic axes emphasizing the discrepancies below  $5 \cdot 10^{-3} \text{ km}^{-1} \text{ sr}^{-1}$ . The scatter for observations between 8 km and 10 km (panel b) looks similar to the one for cases above 10 km, however we notice a tendency to underestimate observations between  $2 \cdot 10^{-3}$  and  $5 \cdot 10^{-3} \text{ km}^{-1} \text{ sr}^{-1}$  and to overestimate the ones below  $2 \cdot 10^{-3} \text{ km}^{-1} \text{ sr}^{-1}$ . For observations between 6 and 8 km (panel c), the underestimation below  $5 \cdot 10^{-3} \text{ km}^{-1} \text{ sr}^{-1}$  becomes more evident (i.e. occurring more frequently). We notice however that the main bulk of cases lie along the main diagonal. The underestimation becomes a dominant feature between 4 km and 6 km, where simulations rarely reach values above  $5 \cdot 10^{-3} \text{ km}^{-1} \text{ sr}^{-1}$ , greatly deteriorating the correlation seen at altitudes above. Because of the larger presence (and higher amount) of cloud liquid, below 4 km (panels e, f) the observed cloud backscatter often reaches values larger than  $10^{-2} \text{ km}^{-1} \text{ sr}^{-1}$  (and up to  $10^{-1} \text{ km}^{-1} \text{ sr}^{-1}$ ), while simulated values stay lower, therefore heavily underestimating them.

Overall, the underestimation by the simulations represents a common feature, which becomes more and more prominent going from higher to lower altitudes. As part of the activity of forecast model validation, a more accurate analysis of this issue is in progress. The following two main reasons are believed to be the cause:

- a) The lack of a description for the horizontal sub-grid in-homogeneity of the model clouds: the cloud fraction is taken into account by ZmVar using the concept of sub-columns, but these are supposed to have the same cloud amount.
- b) The assumption of the maximum vertical overlap for precipitation may not be the most appropriate.

An inaccurate modelling of these two aspects would both translate into an underestimation of cloud backscatter because layers below are penalized by excessive attenuation. The large difference in resolution between model (around 25 km) and observations (300 m at the original sampling) contribute to a magnification of this issue.

In the tropics, Fig. 4.2 shows a similar behaviour above 4 km (panels a to d). At lower altitudes however, the discrepancies are very large. This is mainly due to the inability of the simulator to correctly represent lidar observation of shallow convection. Sensitivity studies (not shown here) clearly demonstrate that a much higher vertical resolution would be needed to accurately model the lidar returns. In addition, prognostic convective

cloud/precipitation fraction, possibly expressed as PDFs (instead of single values), would alleviate the issues related to the mismatch in the horizontal resolution, providing more realistic values. Clearly, solutions of these problems require efforts that go beyond the scope of this project.

The comparison between simulated and observed cloud backscatter shown in Fig. 4.1 and Fig. 4.2 has been used to define the most appropriate criteria to discard those cases where the agreement is poorest. The quality control put in place for the CALIPSO lidar consists in discarding cases where FG departures are larger than  $0.020 \text{ km}^{-1} \text{ sr}^{-1}$ . In addition, a second condition is applied above 8 km, to discard many cases (more evident in the tropics) where to quite weak observed backscatter values correspond much larger simulated ones. Given the large discrepancies and poor correlation, these cases are screened out without attempting to use them in bias correction scheme or subsequently in data assimilation. Screened is done by rejecting situation above 8 km where either observations or FG are larger than  $0.002 \text{ km}^{-1} \text{ sr}^{-1}$ .

The impact of the quality control screening on FG departure statistics is shown in Fig. 4.3 and Fig. 4.4, which respectively consider cases at mid latitudes South (between  $30^\circ S$  and  $60^\circ S$ ) and in the tropics ( $30^\circ S$ - $30^\circ N$ ). These figures contain six panels, each referring to a particular altitude interval. Each panel shows histograms of reflectivity FG departures before (blue curves) and after quality control (red curves). 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. The screening procedure, noticeably changes the shape of the PDF, reducing the tails and giving a more symmetric distribution. This results in FG departure PDFs which are now resembling Gaussian distributions.

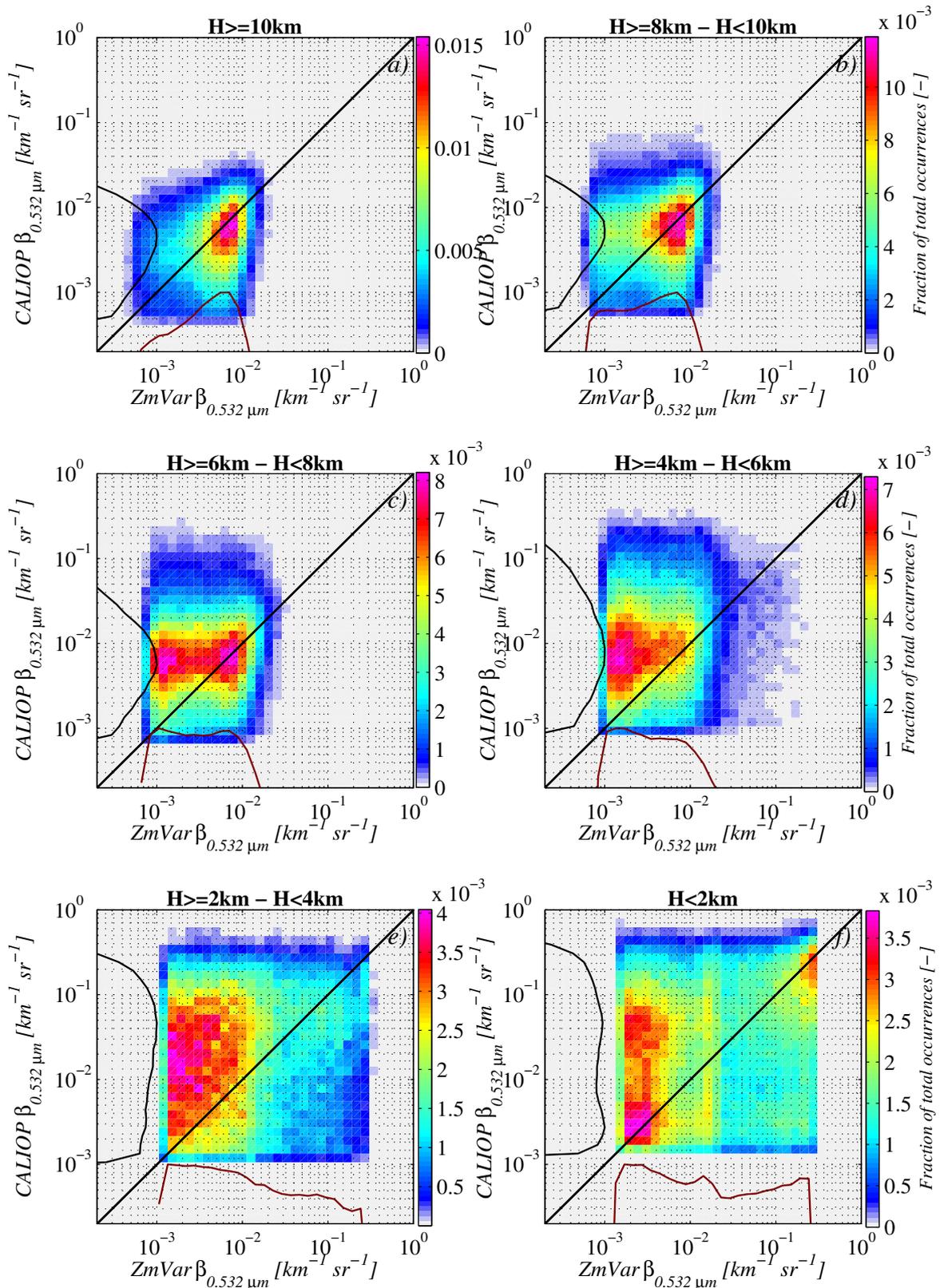


Figure 4.1: Scatter plot between simulated (on abscissa) and CALIOP 0.532  $\mu\text{m}$  cloud backscatter (on ordinate) using observations over ocean matched with model data for the period of January 2007 between 30°S and 60°S. Curves along axes show the relative occurrence of reflectivity values. Each plot refers to the altitude level indicated in the panel title.

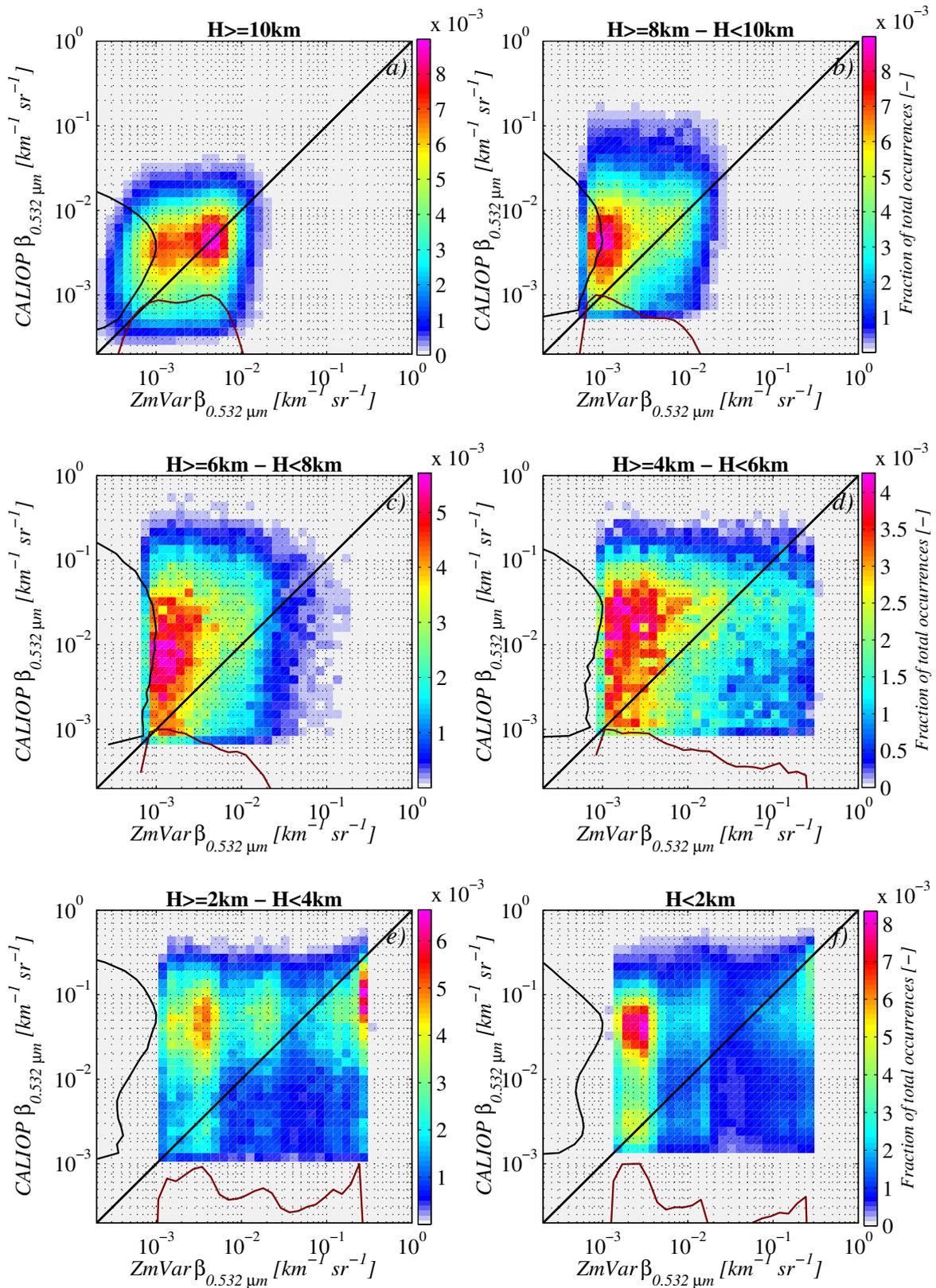


Figure 4.2: Same as 4.1, but considering observations in the tropics (30°S-30°N).

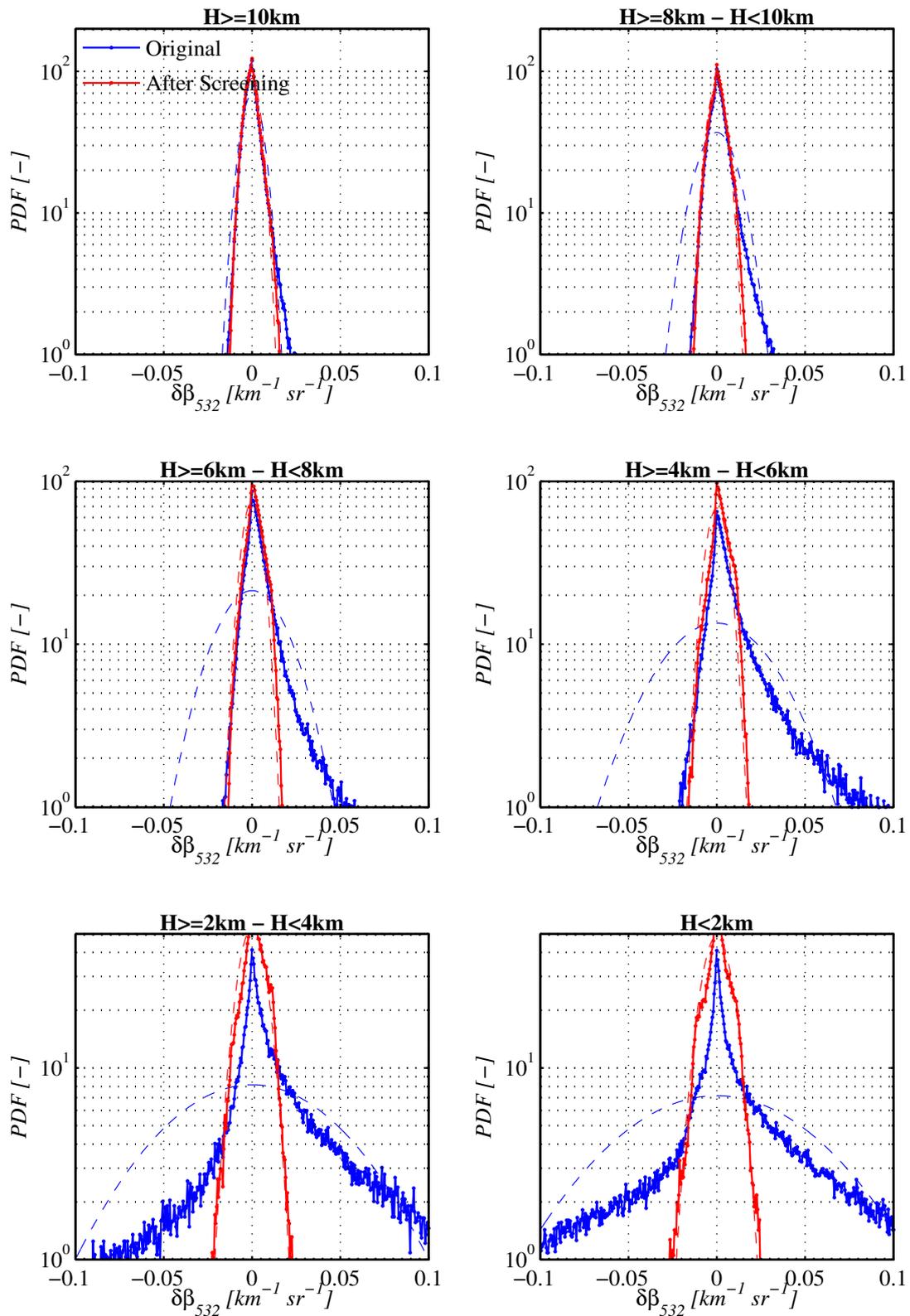


Figure 4.3: PDF of cloud backscatter FG departures (at 0.532 μm) using observations and model data over ocean at mid-latitudes South (30°S - 60°S) in January 2007. Each plot refers to observations in the altitude range indicated in the panel title. Blue and red curves respectively indicate all cases and cases passing the screening, respectively. The bell-shaped curves represent a zero-mean Gaussian PDF with the same standard deviation.

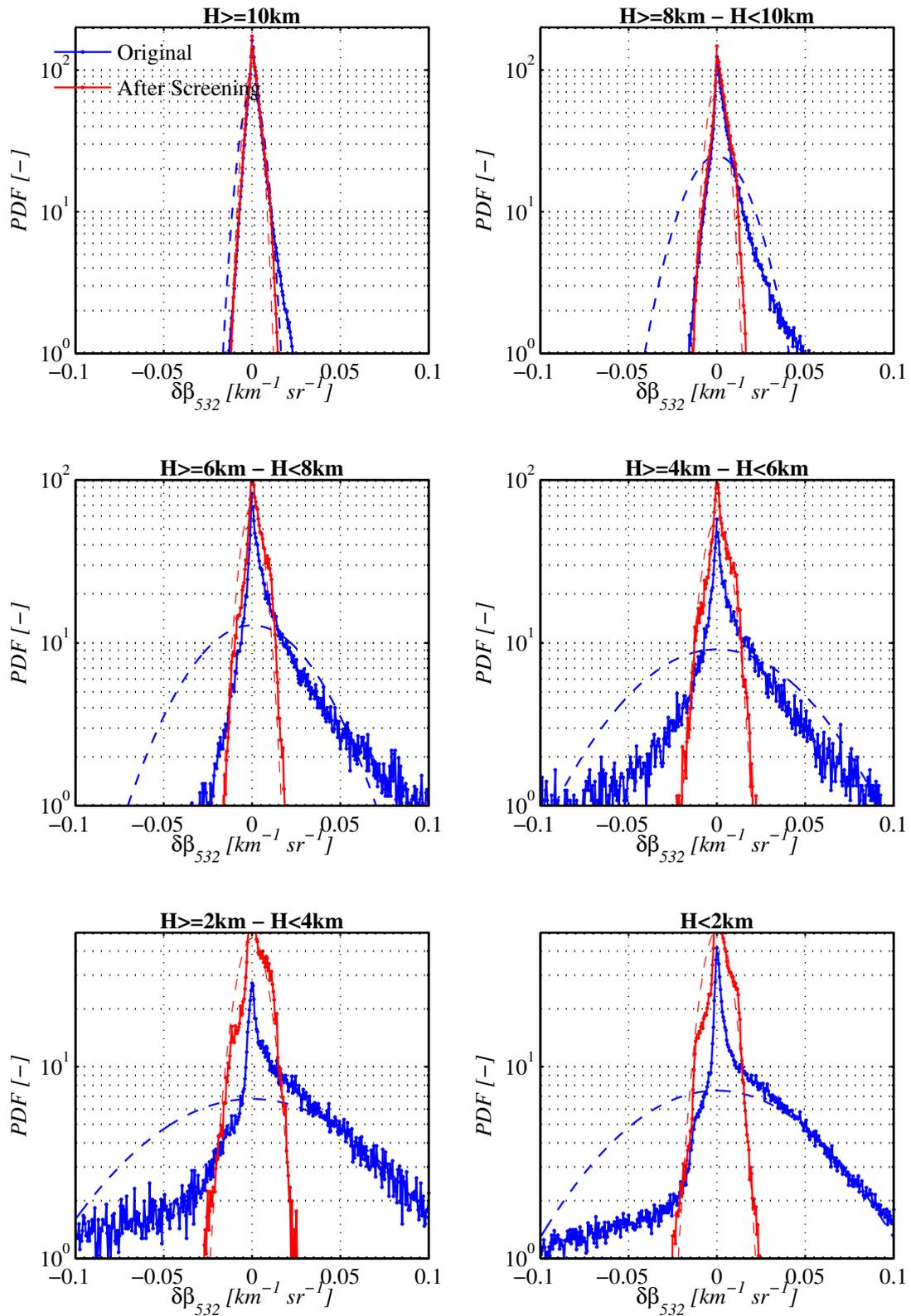


Figure 4.4: Same as Fig. 4.3, but considering observations in the tropics ( $30^\circ\text{S}$ - $30^\circ\text{N}$ ).

## 4.2 Bias correction of CALIPSO lidar observations

As already mentioned in the introduction of Section 4, the processing of lidar backscatter data consists in two consecutive steps: in the first one, a data quality control is used to remove non-optimal and extreme backscatter FG departures. The second step involves the mitigation of any bias present in the statistics of cloud backscatter FG departures. This is a necessary process since the assimilation relies on the assumption that observations minus model background is an unbiased quantity. Therefore, in order to guarantee zero-mean statistics, it is necessary to quantify any systematic errors and to remove them by means of a bias correction scheme.

Similarly to the bias correction of CloudSat radar reflectivity in WP-1100 (Subsection 4.2 in Di Michele et al., 2014), quality-controlled CALIOP cloud backscatter at  $0.532 \mu\text{m}$  for a period of one month (January 2007) have been used to get statistics of observation-minus-background departures and to estimate the entity of a bias. Values of bias correction have been then computed stratifying over ranges of temperature and altitude, separately over geographical regions and seasons. Figure 4.5 show the results for winter (DJF) mid latitudes and tropics. The effect of applying these obtained relationships for bias correction on the intercomparison between CALIOP observations and FG backscatter is shown in Fig. 4.6 and Fig. 4.7, for the period of January 2007 at mid latitudes South and in the tropics, respectively. In both figures the effect of bias correction 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

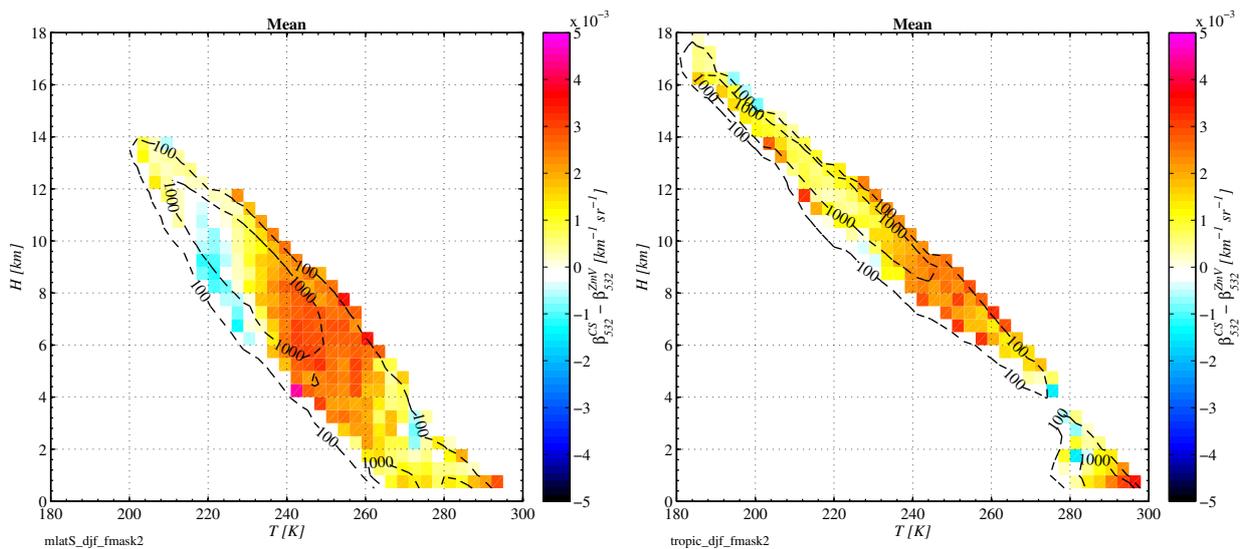


Figure 4.5: Mean FG departures of lidar backscatter in clouds (at  $0.532 \mu\text{m}$ ), as function of temperature and altitude location of observations, used for bias correction. Contour lines give the number of cases in each temperature/altitude bin. Left panel is for data over ocean at mid-latitudes South ( $30^\circ\text{S} - 60^\circ\text{S}$ ) in January 2007. Right panel considers situations in the tropics ( $30^\circ\text{S}-30^\circ\text{N}$ ).

## 4.3 Conclusions

The approaches applied here for quality control of lidar observations gave promising results since applied screening led to more symmetric distribution, closer to the Gaussian shape. Bias correction shows good results in the sense that the average difference between observations and simulations is smaller after applying the bias correction scheme.

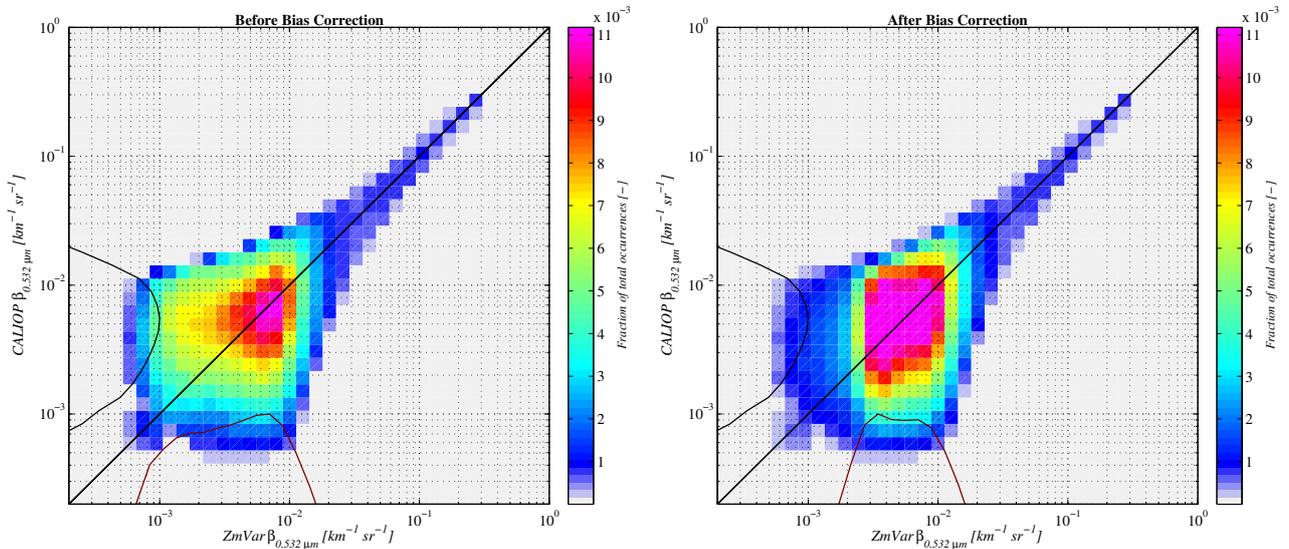


Figure 4.6: Scatter plot between simulated (on abscissa) and CALIOP 0.532  $\mu m$  backscatter in clouds (on ordinate) using observations over ocean matched with model data for the period of January 2007 between 30°S and 60°S. Only data passing quality control are considered. Left panel considers data prior to bias correction, while right panel uses with after applying the bias correction in Fig. 4.5.

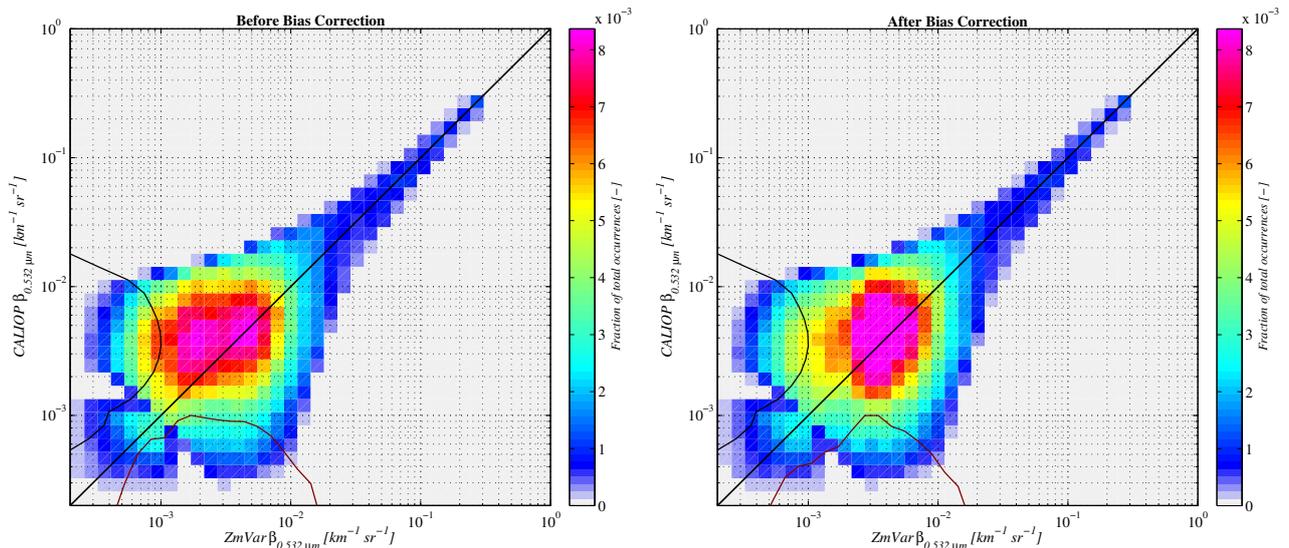


Figure 4.7: Same as Fig. 4.6, but considering observations in the tropics (30°S-30°N).

## 5 Summary

In order to prepare for the monitoring/assimilation of EarthCARE lidar observations using CALIPSO data, the work package focused on development of a lidar forward operator, estimation of observation errors, development of a quality control strategy and a bias correction scheme for the CALIPSO lidar backscatter. The following results have been achieved:

- The reflectivity forward operator ZmVar has been extended to simulate the lidar signal in clouds. The new lidar operator also includes a model to take into account multiple scattering.
- Tangent linear and adjoint versions of the lidar backscatter operator have been developed and tested.
- Observation errors for the CALIPSO lidar have been estimated considering separately the contribution from the instrument, the forward modelling and the spatial representativeness.
- A strategy for quality control of CALIPSO observations has been developed in order to reject those data that cannot be used in the assimilation system.
- Finally, in order to reduce the systematic differences between simulated and observed CALIPSO backscatter, a bias correction scheme has been prepared and tested.

## Acknowledgements

Robin Hogan (University of Reading) is kindly acknowledged for his support with the multiple scattering code. The authors are also grateful to the NASA Langley Research Center - Atmospheric Science Data Center for making the CALIPSO data available. Thanks are due to Anton Beljaars and Stephen English for helpful advice and review of the document.

## List of Acronyms

ATLID	ATmospheric LIDar
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CFAD	Contour Frequency by Altitude Diagrams
CloudSat	NASA's cloud radar mission
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	Global Circulation Model
GES DISC	Goddard Earth Sciences Data and Information Services Center
IFS	Integrated Forecasting System of ECMWF
IGOM	Improved Geometry Optics Method
MC	Multi-Column
MS	Multiple Scattering
NASA	National Aeronautics and Space Administration
NWP	Numerical Weather Prediction
PDF	Probability density function
PSD	Particle Size Distribution
PVC	Photon Variance-Covariance method
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
SFM	Structure Function Maximum
SNR	Signal-to-Noise ratio
SS	Single Scattering
STSE	Support-to-Science-Element
TDTS	Time Dependent Two Stream (approximation)
Z	Radar reflectivity
ZmVar	Z (reflectivity) Model for Variational assimilation of ECMWF

## References

- Bissonnette, L., 2005: Lidar and multiple scattering, *Lidar Range-Resolved Optical Remote Sensing of the Atmosphere*, C. Weitkamp, pp. 43–104.
- Collis, R. and P. Russell, 1976: Lidar measurement of particles and gases by elastic backscattering and differential absorption, *Laser monitoring of the atmosphere*, pp. 71–151.
- Di Michele, S., M. Ahlgrimm, R. Forbes, M. Kulie, R. Bennartz, M. Janisková, and P. Bauer, 2012: Interpretation and evaluation of the ECMWF global model with CloudSat observations: ambiguities due to radar reflectivity forward operator uncertainties, *Q. J. R. Meteorol. Soc.*, **138**, 2047–2065, doi:10.1002/qj.1936.
- Di Michele, S., E. Martins, and M. Janisková, 2014: Observation operator and observation processing for cloud radar, WP-1100 report for the project Support-to-Science-Element STSE Study - EarthCARE Assimilation, 4000102816/11/NL/CT, ECMWF, 59 pp.
- 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 returns, *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.
- 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.
- Forbes, R., A. Tompkins, and A. Untch, 2011: A new prognostic bulk microphysics scheme for the IFS, Technical report, ECMWF Technical Memorandum.
- Hogan, R., 2008: Fast lidar and radar multiple-scattering models. Part I: Small-angle scattering using the photon variance-covariance method, *J. Atmos. Sci.*, **65**(12), 3621–3635.
- Hunt, W., D. Winker, M. Vaughan, K. Powell, P. Lucker, and C. Weimer, 2009: CALIPSO lidar description and performance assessment, *J. Atmos. and Ocean. Tech.*, **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., 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.
- Korolev, A., G. Isaac, and J. Hallett, 2000: Ice particle habits in stratiform clouds, *Q. J. R. Meteorol. Soc.*, **126**(569), 2873–2902.
- Liu, Z., W. Hunt, M. Vaughan, C. Hostetler, M. McGill, K. Powell, D. Winker, and Y. Hu, 2006: Estimating random errors due to shot noise in backscatter lidar observations, *Applied Optics*, **45**(18), 4437–4447.
- 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.
- Platt, C., 1973: Lidar and radiometric observations of cirrus clouds, *J. Atmos. Sci.*, **30**(6), 1191–1204.
- Pope, R., E. Fry, et al., 1997: Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements, *Applied Optics*, **36**(33), 8710–8723.
- Stephens, G., D. Vane, R. Boain, G. Mace, K. Sassen, Z. Wang, A. Illingworth, 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).
- Tiedtke, M., 1993: Representation of clouds in large-scale models, *Mon. Weather Rev.*, **121**(11), 3040–3061.
- Waliser, D., J. Li, C. Woods, R. Austin, J. Bacmeister, J. Chern, A. Del Genio, J. Jiang, Z. Kuang, H. Meng, et al., 2009: Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.*, **114**(D8), D00A21.
- Winker, D., M. Vaughan, A. Omar, Y. Hu, K. Powell, Z. Liu, W. Hunt, and S. Young, 2009: Overview of the CALIPSO mission and CALIOP data processing algorithms, *J. Atmos. and Ocean. Tech.*, **26**(7), 2310–2323.
- Yang, P. and K. Liou, 1996: Geometric-opticsintegral-equation method for light scattering by nonspherical ice crystals, *Applied Optics*, **35**(33), 6568–6584.
- Yang, P., K. Liou, K. Wyser, and D. Mitchell, 2000: Parameterization of the scattering and absorption properties of individual ice crystals, *J. Geophys. Res.-All series*, **105**(D4), 4699–4718.
- Yee, K., 1966: Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *Antennas and Propagation, IEEE Transactions on*, **14**(3), 302–307.