ESA CONTRACT REPORT

Contract Report to the European Space Agency

Expected benefit of wind profiles from the ADM-Aeolus in a data assimilation system

October 2004

Authors: David Tan and Erik Andersson

Final report for ESA contract 15342/01/NL/MM

ESA Study Manager: Paul Ingmann

ECMWF
Shinfield Park
Reading
RG2 9AX
United Kingdom

For additional copies please contact: library@ecmwf.int

Series: ECMWF - ESA Contract Report

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

Contract Report to the European Space Agency

Expected benefit of wind profiles from the ADM-Aeolus in a data assimilation system

Authors: David Tan and Erik Andersson

Final report for ESA contract 15342/01/NL/MM

ESA Study Manager: Paul Ingmann

European Centre for Medium-Range Weather Forecasts

Shinfield Park, Reading, Berkshire, UK

October 2004

ESA STUDY CONTRACT REPORT – SPECIMEN

No ESA Study Contract Report will be accepted unless this sheet is inserted at the beginning of each volume of the Report.

ESA CONTRACT No 15342/01/NL/MM	SUBJECT: Expected Benefit of Wind Profiles from the ADM-Aeolus in a Data Assimilation System		CONTRACTOR ECMWF
* ESA CR( )No	* STAR CODE	No of volumes: 1 This is Volume No 1	CONTRACTOR'S REFERENCE

ABSTRACT:

This Report synthesises the results of the Study "Expected Benefit of Wind Profiles from the ADM-Aeolus in a Data Assimilation System".

Based on an assessment of wind analysis accuracy, the study argues that the main Aeolus impacts are to be expected (i) in the jet streams over the oceans, especially away from main air traffic routes, and in the African/Asian subtropical jet; (ii) in the lower troposphere, e.g. western parts of the N. Pacific and N. Atlantic oceans, if cloud gaps are sufficient; and (iii) in the tropics, where mass-wind balance is weak and hence temperature information is not effective for inferring wind.

These expectations are supported by simulations of the yield and accuracy of ADM-Aeolus data. Observational data from LITE are used to provide realistic profiles of cloud cover as input to such simulations. Profiles of model cloud cover are also used and these are validated against the LITE data. While the overall occurrence of model cloud agress well with LITE-inferred cloud, model cloud cover underestimates the observations by around 20%. The use of model cloud cover for simulating Aeolus data thus results in a modest overestimation of Aeolus penetration depth. Probability distributions of Aeolus instrument error show good agreement, but the worst 10% of errors are underestimated when model cloud cover is used. Overall, the yield and accuracy of simulated Aeolus data are high and compare favourably with radiosondes, with more than 90% of data meeting accuracy requirements in the free troposphere (errors below 2 ms^{-1}). Furthermore they are robust to variations in cloud cover. At low altitudes, two thirds of data from the Mie channel meet accuracy requirements (errors below 1 ms^{-1}) even when substantial cloud is encountered.

Technical developments have been implemented in order to introduce simulated Aeolus data into ECMWF's observation processing and data assimilation system. Standard quality control procedures have been adapted and extended to account for specific features of Aeolus data. In particular, the study found it useful to apply a stricter background quality control to Aeolus data in order to reject a small set of observations with large observation errors. Rejection of these outliers (about 0.6% of the total) results in observation statistics more in keeping with those for conventional wind observations. The study anticipates that the stricter quality control will make it easier to identify unforeseen errors in real Aeolus data.

Assimilation Ensemble experiments have been conducted to assess the expected benefits of Aeolus data in the ECMWF system. These experiments were calibrated by assessing the impact of radiosonde data in the same way. The new technique suggests that the main benefits from Aeolus for analysed wind fields will be found over ocean regions in both hemispheres and in the Tropics. These regions have been identified previously as priority areas for improvement. The benefits seen in analysed fields lead to benefits in forecast fields in regions downstream of the improved analysis regions.

From the encouraging results obtained, it is becoming apparent that the development of high-quality Level-1B and Level-2B processors will be critical for effective exploitation of Aeolus data in NWP applications.

The work described in this report was done under ESA Contract. Responsibility for the contents resides in the author or organisation that prepared it.

Names of authors: David G.H. Tan and Erik Andersson

** NAME OF ESA STUDY MANAGER: Paul INGMANN DIV: EOP-SM - Mission Experts Division DIRECTORATE: EOP - Earth Observation Programmes	** ESA BUDGET HEADING P/I/100.778/600.510/PFI/PM100/01.I15

* Sections to be completed by ESA ** Information to be provided by ESA Study Manager

Abstract

This Report synthesises the results of the Study "Expected Benefit of Wind Profiles from the ADM-Aeolus in a Data Assimilation System".

Based on an assessment of wind analysis accuracy, the study argues that the main Aeolus impacts are to be expected (i) in the jet streams over the oceans, especially away from main air traffic routes, and in the African/Asian subtropical jet; (ii) in the lower troposphere, e.g. western parts of the N. Pacific and N. Atlantic oceans, if cloud gaps are sufficient; and (iii) in the tropics, where mass-wind balance is weak and hence temperature information is not effective for inferring wind.

These expectations are supported by simulations of the yield and accuracy of ADM-Aeolus data. Observational data from LITE are used to provide realistic profiles of cloud cover as input to such simulations. Profiles of model cloud cover are also used and these are validated against the LITE data. While the overall occurrence of model cloud agress well with LITE-inferred cloud, model cloud cover underestimates the observations by around 20%. The use of model cloud cover for simulating Aeolus data thus results in a modest overestimation of Aeolus penetration depth. Probability distributions of Aeolus instrument error show good agreement, but the worst 10% of errors are underestimated when model cloud cover is used. Overall, the yield and accuracy of simulated Aeolus data are high and compare favourably with radiosondes, with more than 90% of data meeting accuracy requirements in the free troposphere (errors below 2 ms^{-1}). Furthermore they are robust to variations in cloud cover. At low altitudes, two thirds of data from the Mie channel meet accuracy requirements (errors below 1 ms^{-1}) even when substantial cloud is encountered.

Technical developments have been implemented in order to introduce simulated Aeolus data into ECMWF's observation processing and data assimilation system. Standard quality control procedures have been adapted and extended to account for specific features of Aeolus data. In particular, the study found it useful to apply a stricter background quality control to Aeolus data in order to reject a small set of observations with large observation errors. Rejection of these outliers (about 0.6% of the total) results in observation statistics more in keeping with those for conventional wind observations. The study anticipates that the stricter quality control will make it easier to identify unforeseen errors in real Aeolus data.

Assimilation Ensemble experiments have been conducted to assess the expected benefits of Aeolus data in the ECMWF system. These experiments were calibrated by assessing the impact of radiosonde data in the same way. The new technique suggests that the main benefits from Aeolus for analysed wind fields will be found over ocean regions in both hemispheres and in the Tropics. These regions have been identified previously as priority areas for improvement. The benefits seen in analysed fields lead to benefits in forecast fields in regions downstream of the improved analysis regions.

From the encouraging results obtained, it is becoming apparent that the development of high-quality Level-1B and Level-2B processors will be critical for effective exploitation of Aeolus data in NWP applications.

Contents

1 Introduction

This document comprises the Final Report for the Study "Expected Benefit of Wind Profiles from the ADM-Aeolus in a Data Assimilation System".

This first section provides a brief introduction to the objectives of the Aeolus Mission and of the current Study. This is followed by an outline of the content of the remainder of this report.

1.1 Objectives of the Aeolus Mission

The Atmospheric Dynamics Mission (ADM-Aeolus) is the second of ESA's Earth Explorer core missions [ESA 1999]. Its objective is to demonstrate the capablity of measuring wind profiles from space using a Doppler wind lidar (DWL) on a polar orbiting platform. ADM-Aeolus is designed to provide high-quality profiles from the surface up to 20 km, the wind information being the horizontal line-of-sight component perpendicular to the satellite track [Stoffelen *et al.* to appear]. Schematics are shown in Figures 1–3. The intention is to launch in October 2007 with a projected lifetime of three years. The Aeolus data products referred to in this Study are described in Appendix A.

The mission objectives and observation requirements have been developed to meet scientific goals of user communities in climate research, atmospheric modelling and numerical weather prediction (NWP). Meteorological observations, particularly those with well-understood and well-quantified error characteristics, are an essential component of NWP. Data assimilation procedures use such observations to derive atmospheric analyses (physically consistent estimates of the atmospheric state). Higher-quality observations are required to improve the accuracy of analyses and associated forecasts. Observations in "meteorologically sensitive" regions [Buizza and Montani 1999] are a priority because these regions have high baroclinicity and thus play an important role in the development and intensification of weather systems [Browning *et al.* 2000]. In these regions, the need for good vertical resolution and the prevalence of cloud pose severe challenges for the global observing system. Active remote sensing by lidars can potentially meet both these challenges, thereby overcoming intrinsic limitations of passive remote sensing techniques.

1.2 Objectives and Rationale of the current Study

In support of the Aeolus mission objectives, ESA funded the current study "Expected Benefit of Wind Profiles from the ADM-Aeolus in a Data Assimilation System", which has been conducted at ECMWF over the period July 2002 to June 2004. The objectives of the study have been to prepare the ECMWF Integrated Forecast System (IFS) to benefit from Aeolus data, and to anticipate potential problems in using the data in an operational data assimilation system such as the IFS. The objectives have been addressed by (i) developing new IFS software to facilitate the assimilation of prototype Level-2B Aeolus data, and (ii) conducting assimilation ensemble experiments to test the new software and to assess the impact of simulated Aeolus data. Interim results have been summarized previously in Technical Notes [Tan and Andersson 2003a,b, 2004]. The results of the study as a whole are synthesized in this Final Report.

The assimilation ensemble experiments assimilate simulated Aeolus data in direct competition with real observational data. The impact of the data is assessed by investigating the spread in the ensemble analyses and in their associated forecasts. The technique offers some advantages over traditional Observing System Simulation Experiments, e.g. [Rohaly and Krishnamurthi 1993; Cardinali *et al.* 1998; Marseille *et al.* 2001; Lord *et al.* 2003], but it is relatively new and has required calibration. This calibration was achieved by assessing the impact of radiosonde data in the same way.

The relevance of the study results depends on the realism of the simulated data. The study has therefore paid

close attention to the estimates of observation errors that are assigned to such data. The modelling assumptions used to derive these estimates must encompass as much as possible the actual processes that will determine the accuracy of future ADM-Aeolus wind measurements. The accuracy of these measurements will depend primarily on the intensity of the backscattered laser light, which in turn depends on the presence and thickness of clouds, and the concentration of aerosol. It is expected that sufficient backscatter will be received from the layers of clear air above clouds, from cloud-top layers, from layers in and below thin clouds, and from layers with sufficient aerosol in the lower parts of the atmosphere. The current study investigates the yield and accuracy of ADM-Aeolus wind profiles through detailed simulations, given realistic cloud distributions and climatological estimates of aerosol concentration. In particular it aims to answer the question: To what extent can the ADM-Aeolus instrument deliver high quality wind data in meteorologically sensitive regions with significant cloudiness?

The simulations in this study thus rely on the provision of realistic cloud cover, cloud water content and on accurate modelling of laser backscatter intensity in clear-sky and cloudy conditions. The study uses cloud data generated by the ECMWF atmospheric model (Autumn-2002 and Winter-2003 versions) at full horizontal resolution (T511, approx 40 km). Simulated backscatter is calculated using the LIPAS (LIdar Performance Analysis Simulator) forward model, developed specifically for the ADM-Aeolus lidar [Veldman *et al.* 1999; Stoffelen *et al.* 2002]. The realism of the simulated backscatter has been verified against actual space-borne lidar backscatter measurements available from the LITE mission [McCormick *et al.* 1993; Winker *et al.* 1996] on the US Space Shuttle in September 1994, provided consistent aerosol backscatter coefficients are employed [Marseille *et al.* submitted]. The study assesses the quality of the model clouds against cloud retrievals, also obtained from the LITE data set.

1.3 Content of the current document

In Section 2, this Report summarizes wind data availability for global NWP and provides estimates of NWP wind accuracy. This serves to identify major weaknesses of the present Global Observing System, in particular with respect to global wind profiles, and to indicate the most likely impacts of Aeolus data.

Section 3 describes the software and datasets used in this study, and includes a validation of model cloud cover. Definitions fundamental for the understanding of data assimilation are also given.

Section 4 contains the study's assessment of the yield and accuracy of simulated Aeolus data, focussing on the sensitivity to cloud cover assumptions.

Section 5 details the technical developments implemented to enable observation processing for Aeolus data at ECMWF; examples are given for the LITE period.

In Section 6 this Report describes Assimilation Ensemble experiments that have been conducted to assess the expected benefits of Aeolus data in the ECMWF system. Examples of analysis and forecast benefits are given.

Section 7 contains the overall conclusions and recommendations of the study, drawing on the earlier Discussion sections (2.3, 4.3, 5.6, 6.5).

2 Available Wind Observations for NWP and Estimates of Wind Analysis Accuracy

2.1 Current availability of wind observations for global NWP

The ECMWF data assimilation system (as at Spring 2003) makes use of more than 1.5 million meteorological observations per 12-hour assimilation cycle [Thépaut and Andersson 2003]. The atmospheric state has many more degrees of freedom so the analysis problem remains severely under-determined [Fisher 2003a]. The coverage and quantity of meteorological observations is monitored routinely at ECMWF; plots updated daily can be found at `http://w3ec2.ecmwf.int/products/forecasts/d/inspect/catalog/Data_Monitoring/DC_used/`. The radiosonde network provides *in situ* observations of wind, temperature and humidity. These are arguably the most valuable source of wind profile information but coverage over ocean regions is poor. Remotely sensed data consist primarily of radiance observations from passive instruments. The vertical resolution of such data is typically several kilometres, which is inadequate for many interesting meteorological situations (examples given below). In addition, infrared sounders are unable to observe below clouds. Radiance observations are most useful for the information they provide on the temperature (hence mass) and humidity fields of the atmosphere. For large-scale extra-tropical dynamics, balance relationships are invoked to infer wind information from radiance observations. However, these relationships have limited validity for the Tropics [Žagar *et al.* submitted], where Coriolis forces are weak, and for smaller-scale extra-tropical dynamics (storms and fronts, with horizontal scales of 100–1000 km and vertical scales less than a few kilometres).

Other types of wind data are available at a limited number of vertical levels. Wind speed at the ocean surface can be inferred from remotely sensed scatterometer data, through relationships between surface wind speed and ocean capillary waves [Isaksen and Janssen 2004]. Tracking of features in cloud imagery can yield atmospheric motion vectors [Schmetz *et al.* 1993] but is subject to errors in height assignment. Wind data from commercial aircraft [Cardinali *et al.* 2003] are abundant at the main cruise altitudes of flight tracks. The profile information available from aircraft ascents and descents is concentrated around airports, so the coverage is far less than what is required for global data assimilation. An active area of research is the exploitation of observations of other meteorological quantities such as cloud cover and precipitation, which may offer some indirect improvement for wind analyses.

In summary, the most important and most challenging requirement for global meteorological analysis remains the observation of wind profiles with high accuracy, global coverage, and good vertical resolution [Baker *et al.* 1995; WMO 2001]. In the remainder of this section are presented examples of wind errors in analyses and short-range forecasts that make use of current metoerological observations. These examples are chosen to substantiate the points made above concerning the regions and spatial scales where better wind data are most needed.

2.2 Estimates of wind errors in global analyses and short-range forecasts

Estimates of wind errors in analyses and short-range forecasts can be obtained through several different means. These include

- Theoretical error estimates produced by the data assimilation procedure [Fisher and Courtier 1995].
- Statistics of observation-minus-background differences (innovations, departures) and of observation-minus-analysis differences (residuals) [Hollingsworth and Lönnberg 1986].
- The amplitude and structure of “key analysis errors”. These are the analysis perturbations that give most reduction in 48-hour forecast errors, and are calculated routinely at ECMWF using the adjoint of the linearized forecast model [Klinker *et al.* 1998].

- The spread of an assimilation ensemble, in which random noise is added to all observations used [Fisher and Andersson 2001].

In this section examples from the latter two approaches are used to illustrate the main areas where better wind observations are required. The focus is on the period Jan–Feb 2003, corresponding to the assimilation ensemble experiments addressed later in this report. Departure statistics and residuals are shown extensively from Section 5 onwards.

Key analysis errors are the perturbations to meteorological analyses that give most improvement to 48-hour forecasts [Klinker *et al.* 1998]. They are based on the adjoint sensitivity of 48-hour forecast error with respect to the forecast's initial condition (the analysis) [Rabier *et al.* 1996]. The method assumes that all of the 48-hour forecast error can be attributed to errors in the initial condition (i.e. zero model error) and that the tangent-linear assumption is valid over the 48-hour range. Climatologies of key analysis errors have been calculated by e.g. Marseille and Bouttier [2001] and used as indication of the atmosphere's sensitive regions where additional observations would be most valuable. Key analysis errors have only been computed routinely at ECMWF since 1998, using the adjoint of the linearized forecast model. As mentioned above, examples are given for Jan–Feb 2003; similar features occur for the LITE period in 1994, see e.g. Tan and Andersson [submitted]

At 250 hPa (Figure 4a) the largest perturbations are found in the jet stream regions, especially over the oceans. At 850 hPa (Figure 4b), the North Atlantic and North Pacific storm tracks feature prominently. The perturbations are small in amplitude, around 0.4 ms^{-1}, but grow rapidly through baroclinic instability mechanisms. Baroclinic instability is largely responsible for the highly tilted structures and small vertical scales evident in key analysis perturbations. Small-scale mixing and condensation processes are also associated with the storm tracks, and result in substantial cloud cover (Figure 4c). Much of this cloud occurs at high altitudes, limiting the effectiveness of passive infrared temperature and humidity sounding [McNally 2002]. DWL has the potential to yield useful data here, all the way down to the cloud top layer, below thin clouds and in gaps between clouds. Indeed, one of the aims of this study is to quantify the frequency and accuracy of Aeolus data in sensitive and cloudy regions.

Figure 5 shows the assimilation ensemble spread for zonal wind at 250 hPa, computed from the "Control" ensemble described in detail in Section 6. The root-mean-square (rms) difference of the 12-hour forecasts is shown, and illustrates the main features of short-range forecast uncertainty in the Autumn 2003 version of the ECMWF forecast system. These same uncertainties are comparable to the background errors employed in data assimilation [Fisher and Andersson 2001], and hence provide some guidance with respect to user requirements for observation errors in new systems such as Aeolus. The largest uncertainties, exceeding 2 ms^{-1}, are found over the oceans in the southern hemisphere and the tropics, as well as in the northern hemisphere storm track regions. Elsewhere, values exceeding 1 ms^{-1} are typical. These spreads may be under-estimating the actual background error. A scaling factor of order 2 should be applied for better agreement with radiosonde departure statistics. The advent of ATOVS data has yielded significant reductions in Southern hemisphere uncertainty [Simmons and Hollingsworth 2002], but there remains a general lack of wind data there.

2.3 Discussion

This section concludes by anticipating developments in Aeolus timeframe (2007–2010). The storm track regions in both hemispheres are highly baroclinic and DWL is one of the few observing systems that promises the vertical resolution (1 km) required to observe important flow structures in these regions. Data from high-resolution passive sounding instruments, e.g. IASI on the European METOP platform, will increase but fundamental problems in the presence of cloud will persist. Thus, the main Aeolus impacts are to be expected

- in the jet streams over the oceans, especially away from main air traffic routes, and in the African/Asian subtropical jet;

- in the lower troposphere, e.g. western parts of the N. Pacific and N. Atlantic oceans, if cloud gaps are sufficient.
- in the tropics, where mass-wind balance is weak and hence temperature information is not effective for inferring wind.

3 Methods, Software Tools and Datasets

As explained in Section 1.2, it is expected that the accuracy of the ADM-Aeolus doppler winds will depend directly on the intensity of the backscattered laser light at the receiver. The signal-to-noise ratio is higher the more photons have been received and detected. This is dependent on instrument characteristics as well as the atmospheric conditions [Marseille and Stoffelen 2003]. The molecular (Rayleigh) backscatter depends on the air pressure, and attenuation by clouds. The aerosol (Mie) backscatter depends on aerosol concentration, presence of clouds and the optical properties of clouds. There can also be a weak dependency on temperature. To estimate the accuracy of Aeolus winds this study therefore needed to simulate the instrument itself, the atmospherc conditions and the effects on the laser beam due to clouds and aerosols. It relied on the LIPAS software (described in Section 3.1 below) to simulate the Aeolus instrument and the atmospheric backscatter. The study used the ECMWF forecast model at full operational resolution to simulate the atmosphere and clouds (Section 3.3), while aerosol distributions were taken from climate. As both clouds and aerosols are uncertain, aspects of the simulation were verified against actual space borne lidar data. For this purpose the study has used data from the LITE instrument flown on the US Space Shuttle for a 10-day period (9–18 September) in 1994.

This is followed by definitions required to understand the treatment of Aeolus observations in ECMWF's 4DVAR data assimilation system (Section 3.4), and the prototype Level-2B Aeolus observations used in this study (Section 3.5, also Appendix A).

3.1 Aeolus simulator "LIPAS"

This study makes extensive use of simulated Aeolus data obtained from LIPAS (LIdar Performance Analysis Simulator). LIPAS was developed at KNMI by A. Stoffelen and G.-J. Marseille [Veldman *et al.* 1999; Stoffelen *et al.* 2002]. It simulates the performance of spaceborne Doppler wind lidars, including Aeolus' current baseline instrument: laser operating at 355 nm, separate Mie and Rayleigh direct detection receiver channels for returns from aerosol/cloud and molecules respectively. The user specifies parameters such as the observation resolution, shot-accumulation length and pulse repetition frequency.

Aeolus instrument performance parameters (see Table 2 of Marseille and Stoffelen [2003]) have been updated to conform with current baseline assumptions (as advised by ESA/ESTEC, September 2002). A 1 km accumulation length is now used (was 3.5 km), telescope diameter is set to 1.5 m (was 1.1 m) and the instrument's optical performance has been reduced slightly. Various sources of instrument error (random and systematic) are modelled according to industry specifications. Other biases, e.g. laser pointing errors and height assignment errors arising from non-uniform aerosol distributions, are now modelled by LIPAS but were not available in earlier ADM studies, e.g. the Observing System Simulation Experiments (OSSE) by Stoffelen and Marseille [1998]; Marseille *et al.* [2001]. Throughout this study, parameter settings have resulted in simulated systematic errors that are virtually identical to previous estimates [ESA 1999], e.g. 0.84 ms^{-1} for the Mie channel at 2 km, 1.3 ms^{-1} for the Rayleigh channel at 10 km.

The main meteorological inputs to LIPAS are profiles of temperature, wind, humidity, cloud cover, cloud liquid water content and cloud ice water content, and are typically supplied from an NWP model. The quality of DWL observations is sensitive to the number of backscattered photons. Aerosol backscatter is therefore another critical input to the LIPAS atmospheric propagation module: the user can (i) select from the "Vaughan

climatology", 5 climatological profiles derived from measurement campaigns at 10.6 μm over the Atlantic during the relatively clean atmospheric period 1988-1990 [Vaughan *et al.* 1998; Veldman *et al.* 1999], (ii) use the mean and standard deviation from the climatology to produce random, spatially uncorrelated, realizations of aerosol backscatter, (iii) select from the "LITE4ADM climatology", 10 climatological profiles derived for the LITE period [Marseille *et al.* submitted] or (iv) provide user-defined profiles.

The main outputs from LIPAS are perfect Aeolus observations at user-specified resolution, together with profiles of estimated random, systematic and representativity error. LIPAS was successfully ported to ECMWF in July 2002 and a number of enhancements have been made during the current Study.

Representativity error is now calculated for a reference model grid length of 50 km rather than 100 km as assumed previously. Two representativity errors are calculated: σ_{P50} for a "point" HLOS observation, which can be regarded as one wind component from a "perfect" radiosonde, i.e. one for which there is no instrument error and only representativity error; and σ_{50} for the Aeolus HLOS observation. This report uses $\hat{\sigma}_M$ and $\hat{\sigma}_R$ to denote the random error arising from instrument effects for the Mie and Rayleigh channels respectively. The corresponding Aeolus observation random errors are defined as $\sigma_M = (\sigma_{50}^2 + \hat{\sigma}_M^2)^{1/2}$ and $\sigma_R = (\sigma_{50}^2 + \hat{\sigma}_R^2)^{1/2}$.

The classification of observation quality has been made stricter. The ratios of Aeolus observation random error to point representativity error, σ_M/σ_{P50} and σ_R/σ_{P50}, are now classified as very good, good, poor or very poor if they fall in the range 0–0.9, 0.9–1.8, 1.8–2.7, 2.7–∞, respectively. The previous classification used the analogous ratios, with a reference grid of 100 km for representativity error, and the ranges 0–1, 1–2, 2–3, 3–∞.

The LITE4ADM climatological aerosol profiles were added as aerosol input options (see above). These profiles imply around 5 times more aerosol backscatter than the Vaughan climatology in the upper troposphere and lower stratosphere because of the presence of aerosol from the 1991 Mt Pinatubo eruption.

Originally, LIPAS assumed a fixed HLOS azimuth angle of 90 degrees. This was generalized to a location-dependent input for the assimilation ensemble simulations (Section 6), in order to match the intended Aeolus configuration, i.e. to have the azimuth perpendicular to the satellite ground track.

Several assumptions affect LIPAS simulation of Aeolus data. Cloud profiles are assumed constant over a horizontal accumulation length (1 km) but vary within a single observation ($\sim$ 50 km) assuming maximum-random overlap. Cloud encounters are computed for a nadir-viewing lidar, whereas Aeolus will use a 35 degree slant angle. The difference results in a shorter optical path and a small over-estimation of probability of cloud-free propagation (at most 5–10% for cloud cover in the range 40–70% [Rapp *et al.* 1973; Yu *et al.* 1986]). Multiple scattering is not modelled but is thought to be a small effect for the Aeolus field-of-view (15 μrad). Droplet size is assumed to vary linearly with pressure. No distinction is made between backscatter from liquid cloud and ice cloud. The shape and size of ice crystals affect multiple scattering and hence backscatter signal strength [Chepfer *et al.* 1999] but the uncertainties are comparable to uncertainties in choice of backscatter-to-extinction ratio. The climatological aerosol backscatter profiles are assumed to have no horizontal variability. A constant backscatter-to-extinction ratio is assumed, the precise value being specified by the user (0.02 sr^{-1} has been used in previous studies and is retained here.)

3.2 The importance of aerosol and clouds

In the simplest model of the Doppler wind lidar observation principle, the HLOS wind component is directly proportional to the shift in frequency between the transmitted photons and the received photons. The quality of an observation depends on the instrument signal-to-noise ratio, which in turn depends on the amplitude of the backscattered signals, i.e. the *number* of received photons. Figure 6 shows photon count sensitivity to different temperature and aerosol backscatter profiles. The temperature profiles are typical zonal mean profiles for northern hemisphere winter (ECMWF operational model, January 2002). The effect of lower temperature is to increase molecular density which results in more molecular backscatter and hence slightly improved signal

quality in the DWL Rayleigh channel. However, photon counts exhibit far greater sensitivity to the assumed aerosol backscatter profile. The 5 profiles from the Vaughan climatology (see Sec. 3.1) vary by an order of magnitude at any given altitude and over several orders of magnitude vertically. While four profiles vary monotonically with height and give photon counts that increase with increasing aerosol backscatter, the upper decile profile has a layer of enhanced aerosol backscatter around 10 km. The corresponding photon counts are large from the top of this layer, but substantial attenuation occurs so that, below 8km, the received signal is much weaker than for the monotonic aerosol profiles. The contrast in observation quality is shown in Figure 7 for the median and upper decile aerosol backscatter profiles. Better quality is obtained from the aerosol layer at 10 km and virtually no useful data between 0 and 1 km due to strong attenuation. In the case of the upper decile aerosol profile, better quality is also obtained between 2 and 6 km. This is an example in which the aerosol backscatter is sufficiently strong to increase the yield and accuracy of data from the Mie channel, outweighing the degraded quality in the Rayleigh channel.

It follows that the specification of aerosol and cloud backscatter warrants careful consideration. In the near future, further valuable data on aerosol and cloud backscatter variability can be expected from satellite missions such as CALIPSO. If successful, these missions will provide profile information that is not available from e.g. MODIS, which nonetheless is providing useful information on horizontal variability. In the absence of such data, the study continued to use the median profile from the Vaughan climatology. This has the advantage that the sensitivity of LIPAS simulations to the cloud cover input is enhanced relative to simulations assuming greater aerosol concentrations.

3.3 Datasets

This study makes extensive use of ECMWF model cloud cover fields. These have been validated against radiance observations, and have well-known deficiencies such as a systematic lack of stratocumulus cloud off the west coast of the continents [Chevallier *et al.* 2001] and limited success in capturing the observed temporal variability in the intertropical convergence zone (ITCZ) [Chevallier and Kelly 2002]. While these studies do not provide much guidance on the impact such deficiencies have on Aeolus simulations, the LITE mission [McCormick *et al.* 1993; Winker *et al.* 1996] provides useful independent data to assess the reasonableness of Aeolus modelling assumptions. In Section 4 this report assesses the yield and accuracy of simulated Aeolus data for the LITE period 9–18 September 1994, and its sensitivity to the assumed cloud cover. To assist in interpreting the results, the report discusses first the study's complementary exercise to validate the model cloud cover against LITE data.

3.3.1 Model fields for the LITE period

A necessary preparation for this study has been to derive high resolution meteorology for the LITE period. Previous model fields are inadequate because of insufficient resolution as well as improvements in physical and dynamical aspects of the ECMWF model. In addition, the ERA-40 re-analysis project [Uppala 1997] has given us better knowledge on how to use the data available at that time. The study has therefore applied the Summer-2002 ECMWF IFS ("Cy25r1") for the LITE period, assimilating all the observations available in ERA-40 to produce the best posssible analyses and forecasts. Relative to ERA-40 assimilations, Cy25r1 uses an improved version of the RTTOV radiative transfer scheme, so new bias corrections for HIRS and SSM/I radiance observations have been generated and employed. The forecast fields from the model have been archived every 2 hours. For comparison with LITE data, these fields are sampled at the time within the range T+12 to T+36 that is closest to the orbit overpasses.

3.3.2 Validation of model cloud cover against LITE cloud retrievals

The study has validated the ECMWF model cloud cover using LITE data from 65 orbits and comprising 12389 profiles of cloud cover. The LITE-inferred cloud cover available at ECMWF was derived elsewhere in two steps. First, a cloud-boundary detection algorithm adapted from [Winker and Vaughan 1994] was applied to raw LITE data from the 532 and 1064 nm channels. Second, the cloud-boundary information was converted to cloud cover at 1 degree horizontal resolution and 31 vertical levels [Miller *et al.* 1999], corresponding to ECMWF's former operational model. (Note that Miller *et al.* [1999] use the term "pseudo-cloud fraction" for the cloud cover product).

It is difficult to quantify the accuracy of the LITE cloud cover product. The thresholds used in the cloud detection were high in the sense that, excluding clouds located underneath dense clouds, any cloud missed would be extremely tenuous (optical depth less than 0.005 say). Subvisible cirrus (optical depth less than 0.03) accounts for around 5–10% of inferred cloud cover [Winker, personal communication]. Miller *et al.* [1999] remark that pulse stretching could lead to an underestimation of cloud base altitude and hence an overestimation of low-level cloud cover.

Figure 8 shows cloud cover inferred from LITE observations for orbit 047 and from the ECMWF model. The model fields are defined on 60 vertical levels so these have been reduced for the purposes of the present comparison to 31 levels which approximately coincide with the LITE product. Where levels have been merged, the maximum-random cloud overlap assumption has been applied [Jakob and Klein 2000]. The level of agreement is typical of other orbits. The main features are

- The occurrence of model cloud cover shows fairly good agreement with LITE-inferred cloud cover. Cloud top height agrees quite well.
- Model cloud cover underestimates LITE-inferred cloud cover by 20% on average. Note that LITE in essence samples each model grid box along a narrow one-dimensional slice. In the context of nadir-pointing cloud radar, analogous sampling issues lead to comparable biases [Astin 1997].
- The model has a systematic lack of low-level cloud. While accurate modelling of stratocumulus cloud is an acknowledged difficulty, discrepancies with LITE may be exaggerated by the effects of pulse stretching (see above).

3.4 Definitions for data assimilation

3.4.1 Formulation of 4DVAR

The principal inputs to data assimilation are a set of observations y and a background estimate x_b of the atmospheric state. The principal output is an analysis x_a that is a physically consistent estimate of the atmospheric state, taking into account the observations, the background estimate, and the error estimates for both entities. The background estimate (also known as the "first-guess") is typically a short-range forecast based on the previous analysis.

The analysis x_a (a vector) minimizes the cost function

$$J(x) = \frac{1}{2}(x-x_b)^T B^{-1}(x-x_b) + \frac{1}{2}(y-Hx)^T R^{-1}(y-Hx) \quad (1)$$

Here, B denotes the error covariance matrix of the background estimate, R the error covariance matrix of the observations, and H the observation operator that is applied to the model state for comparison with the observations. For ease of notation, the formulation presented here incorporates in H any time-integration of the

atmospheric model that is invoked when comparing the atmospheric model with observations taken at different times, as is the case in 4DVAR [Rabier *et al.* 1998, 2000; Mahfouf and Rabier 2000; Klinker *et al.* 2000].

3.4.2 HLOS observation operators

The cost function contributions from uncorrelated sets of observations are additive. For each observation y_k, the calculation of the model equivalent $H_k x$ is typically nonlinear (for example a radiative transfer calculation). However, the situation is much simpler for observations anticipated from the Aeolus mission because Doppler wind lidar observations of horizontal line of sight (HLOS) wind component are a linear function of the wind field. Let $\sigma^2_{o,los}$ denote the error variance of HLOS wind observations. Then the contribution to J from each HLOS observation is

$$J_{los} = \frac{1}{2}\frac{(y_{los} - \mathrm{H}_{los}x)^2}{\sigma^2_{o,los}}, \tag{2}$$

where the HLOS observation operator is given by

$$\mathrm{H}_{los} = -u\sin\psi - v\cos\psi. \tag{3}$$

Here, u and v denote the zonal and meridional wind components, and ψ is the HLOS azimuth angle of the DWL (measured clockwise from North, see Figure 3 and Cardinali *et al.* [1998]; Marseille *et al.* [2001]).

In the incremental formulation of 4DVAR implemented at ECMWF, for each observation operator it is also necessary to have two derived operators, namely the tangent linear operator and its adjoint. Because H_{los} is linear and the HLOS azimuth angle ψ is assumed fixed for a single HLOS observation, these operators are given by

$$\delta\mathrm{H}_{los} = -\delta u\sin\psi - \delta v\cos\psi, \tag{4}$$

and

$$\mathrm{H}^*_{los} = (-\delta y_{los}\sin\psi, -\delta y_{los}\cos\psi)^T. \tag{5}$$

3.4.3 Specification of HLOS error estimates

The error variances of HLOS wind observations, $\sigma^2_{o,los}$, are diagonal elements of the matrix R. To give appropriate weight to the observations in data assimilation, these variances must account for error variance arising from instrument random errors, $\sigma^2_{i,los}$, as well as from representativity errors, $\sigma^2_{r,los}$,

$$\sigma^2_{o,los} = \sigma^2_{i,los} + \sigma^2_{r,los}. \tag{6}$$

Further details are in [Lorenc *et al.* 1992; Stoffelen and Marseille 1998; Tan and Andersson 2003a]. The study has assumed that estimates of observation error variances $\sigma^2_{o,los}$ are reported in the Aeolus data provided as an output of the Level-2B processor.

3.5 Prototype Level-2B Aeolus observations

The basic inputs for Observation Processing at ECMWF are the data available in near-real-time to operational NWP centres. In the case of Aeolus, it is expected that these will be the Level-2B data which have been called “consolidated HLOS wind observations” and are based on the best available estimates of atmospheric state parameters (primarily temperature). This study notes however that NWP centres may in future find it advantageous develop forward operators that would permit assimilation of Level-1 products e.g. Level-1B “unconsolidated” HLOS wind observations or perhaps even Doppler shifts. (Refer to Appendix A).

Error	LIPAS	Data Assimilation
Representativity	σ_{50}	$\sigma_{r,los}$
Instrument	$\hat{\sigma}_M$, $\hat{\sigma}_R$	$\sigma_{i,los}$
Observation	σ_M, σ_R	$\sigma_{o,los}$

Table 1: Correspondence between error notation for LIPAS and for data assimilation.

Pending further details of the Aeolus Payload Data Segment, this study assumes a prototype Level-2B Aeolus product in BUFR format containing the following information for an Aeolus observation profile: the location and time of the observation (latitude, longitude, and time), an instrument identifier (specifying either the Rayleigh or the Mie channel), the line-of-sight azimuth angle, the profile of observed HLOS wind component (ms^{-1}), and the profile of estimated observation error ($\sigma_{o,los}$). The profiles of observed HLOS wind component and estimated observation error both comprise values at 20 pressure levels which are also specified in the BUFR report. Further processing of LIPAS outputs has been developed to provide noisy data in this form. Details of how these are generated for the LITE period are given in Section 5.5, for the assimilation ensembles in Section 6.2. Table 1 shows the correspondence between the error notation for LIPAS (Section 3.1) and that for data assimilation (Section 3.4).

Should future refinements of the Level-2B product use height or geopotential rather than pressure as the vertical coordinate, conversion to pressure coordinate would occur in IFS observation pre-processing. It would be done in the IFS code because it requires access to model fields (surface pressure and temperature).

Should future refinements of the Level-2B product report a wind product that is a layer average over each ADM-Aeolus vertical range bin, this can be accounted for within the HLOS observation operator by averaging the model fields in a corresponding manner.

4 Yield and Accuracy of simulated Aeolus data

Some insight into the yield and accuracy expected of Aeolus data can be derived from examining simulated Aeolus observations [Marseille *et al.* 2001]. Recent work has classified simulated Aeolus data in relation to the accuracy and representativity of radiosonde observations. The performance sensitivities previously investigated include the impact of systematic errors arising from correlations between aerosol variability and wind shear [Stoffelen *et al.* 2002]. This study adopts a similar approach to investigate sensitivity to the treatment of cloud effects in the simulations. Section 4.1 contains the study's assessment of "LITE-based simulations", i.e. LIPAS simulations that use LITE-derived cloud cover. In Section 4.2 these are compared to "model-based" simulations that use the model cloud cover fields validated in Section 3.3. Both the "LITE-based" simulations and the "model-based" simulations use the LITE-period wind-fields described in Section 3.3.1. As mentioned previously, this report shows simulations that assume aerosol backscatter is given by the Vaughan median profile. It is for this comparatively low atmsopheric aerosol loading that sensitivity to the assumed cloud cover is most apparent. In terms of random errors, simulations with higher aerosol loadings indicate that better performance from the Mie channel outweighs Rayleigh channel degradation.

4.1 Yield and accuracy of simulated Aeolus data in the LITE period

For the LIPAS simulations using LITE-retrieved cloud cover, the simulated Aeolus data have been grouped in 5 degree by 5 degree bins. Within each bin, the percentage of simulated data that meet the mission accuracy requirements was calculated. The results are shown in Figures 9, 10 and 11 for altitude ranges 9–10 km, 4–5 km and 0.5–1 km respectively. Contrary to future real ADM-Aeolus data, the simulated data are co-located

	Domain	Rayleigh		Mie		
	N/W/S/E	4–5 km	9–10 km	0.5–1 km	4–5 km	9–10 km
N.Hem	90/-180/20/180	73	91	69	16	22
Tropics	20/-180/-20/180	71	78	66	10	15
S.Hem	-20/-180/-90/180	81	98	58	16	14
N.Atl	65/-70/25/-10	70	89	62	16	26
N.Pac	60/145/25/-130	68	88	57	15	22
E.Asia	60/102.5/25/150	63	90	48	24	26
ITCZ	20/80/0/-80	43	49	49	9	32
S.Atl	-30/-60/-55/60	74	97	59	14	23
S.Pac	-20/-180/-45/-90	81	95	55	19	15

Table 2: Percentage of data meeting mission accuracy requirements.

with LITE cloud cover profiles and hence are confined to 60S–60N. There are also no LITE data available between 60–80E. For comparison purposes the three figures also show r.m.s. of high, medium and low cloud cover (shaded), as taken from the model fields that are alternatives to the LITE-retrievals.

Rayleigh channel yield for 9–10 km exceeds 75% throughout most of the extratropics (Figure 9, upper panel, green markers). High cloud over the North Atlantic and the Southern Hemisphere oceans does not reduce yield significantly at this altitude. In the Tropics, areas of high cloud cover, specifically the ITCZ and the West Pacific warm pool, correlate with reduced Rayleigh channel yield (49%, Table 2). Conversely, good Mie channel yield relies on the presence of cloud (Figure 9, lower panel) at this high altitude.

Figure 10 shows yield at 4–5 km. The Rayleigh channel is still good in cloud-free regions. There is however a systematic reduction in yield in regions where the DWL must penetrate high cloud – see Table 2. For example, Rayleigh channel yield in the North Atlantic is 89% at 9–10 km, reducing to 70% at 4–5 km. Mie channel yield is still linked to the presence of cloud, and complements the Rayleigh channel.

At 0.5–1 km, Figure 11, the Rayleigh channel yields no data that meet the mission requirements. The Mie channel however is expected to perform well when scattering from cloud tops is strong, and this is confirmed e.g. off the west coasts of the continents where stratocumulus cloud is prevalent, as well in the Southern hemisphere storm tracks. Aeosol returns are also expected to yield good Mie channel data at this altitude, and would explain the high yield in cloud-free regions.

4.2 Sensitivity to cloud cover

4.2.1 Quality statistics

Figure 12 shows the percentages of simulated Aeolus data for which the ratio σ_A/σ_{P50} falls into the ranges 0–0.9, 0.9–1.8, 1.8–2.7, 2.7–∞, where $\sigma_A = \min(\sigma_M, \sigma_R)$, assuming LITE-inferred cloud cover. For comparison, equivalent simulations using model cloud cover are shown in Figure 13. Overall the correspondence is good. In the mid-latitudes (upper panels), the simulated data based on model cloud cover overestimate the percentage of “very good” quality (ratio less than 0.9) and underestimate the percentage of “very poor” quality (ratio more than 2.7). This is consistent with the model underestimation of cloud cover. The discrepancies in quality are most pronounced between 0–2 km where the use of model cloud cover underestimates very poor quality data by some 25%. Above 5 km, the percentage of very good data (ratio less than 0.9) is overestimated when model cloud cover is used but the discrepancy is no more than 10%. Similar behaviour is found in the sub-tropics so plots are not shown.

As an alternative measure of Aeolus yield, consider the penetration depth of Aeolus data. The study applies a demanding definition, namely the altitude below which more than 10% of simulated data are classified as very poor. With model cloud cover the penetration depth is 4 km in mid-latitudes. This is more optimistic than the penetration depth associated with LITE-inferred cloud cover (5 km, upper panels of Figures 12 and 13.) As much as 70% of data below the penetration depth remain good.

In comparison to the mid-latitudes, the most significant feature of tropical performance is that Mie channel detection of scattering from cirrus cloud results in 15–20% of Aeolus data being classified as very good between 12 and 16 km. The simulated transmission through thin cirrus remains high so there is no major degradation in quality at lower altitudes. As for mid-latitudes, the results are robust to whether the cloud cover field is LITE-inferred or from the ECMWF model. Even with LITE-inferred cloud cover, the percentage of very poor data between 0 and 2 km is smaller in the subtropics and tropics than in the mid-latitudes (15–30% rather than 20–40%). With LITE-inferred cloud cover, 20% of tropical data below 6 km are very poor but, as for mid-latitudes, this penetration altitude is overestimated with model cloud cover (just 10% very poor down to 0 km.)

4.2.2 Instrument random error profiles

Another perspective on Aeolus performance is provided by examining the probability distributions of simulated Aeolus instrument random error, $\hat{\sigma}_R$ and $\hat{\sigma}_M$ for the Rayleigh and Mie channels respectively. Such distributions contain more detail than the 4 ranges used for quality classification and are of interest for the quality control aspects of data assimilation described later. The upper panel in Figure 14 shows the instrument error distribution for the Aeolus Rayleigh channel. The thick solid curve connects the median error at each altitude and thus represents a “typical” error profile. The spread of the distribution at each level is shown by horizontal lines: solid lines connect the lower and upper quartiles, dashed lines the lower and upper deciles. Above 9 km, cloud effects are small and compact error distributions for the Rayleigh channel are evident. The reduction in errors above 16 km is associated with the change in vertical resolution from 1 km to 2 km. With decreasing altitude, clouds degrade the amplitude of returned lidar signals and result in skewed error distributions with greater probability of large errors. However, median errors decrease with decreasing height (increasing molecular density) until around 2 km where attenuation is too strong.

For reference, the thick dash-dot curve shows the Aeolus mission requirements. These are met comfortably by large amounts of Rayleigh channel data: more than 90% between 9 and 12 km, more than 75% between 4 and 9 km. It can be seen that the Rayleigh channel median error exceeds the mission requirements around 2 km. Conversely, the Mie channel (Figure 14, lower panel) performs well below 2 km, with median errors less than 0.5 ms^{-1} and almost 75% of all errors complying with the mission specification. Median errors in the Mie channel increase sharply above 2 km, exceeding 5 ms^{-1} above 3 km. Nonetheless, the best 10% of Mie channel returns between 2 and 13 km are below 0.5 ms^{-1}. Such cloud top returns may yet prove valuable, but further research is required to assess their representativeness.

4.2.3 Observation random error profiles

To assess the usefulness of Aeolus data in an operational data assimilation environment, it is relevant to examine observation random errors σ_R and σ_M (Figure 15). Because they combine instrument error and representativity error, these observation errors are inversely proportional to the weight that would be given to the observations in data assimilation procedures. They are thus directly comparable to the observation error assigned to radiosonde wind observations, shown as the thick dashed curve. Taking the Rayleigh channel above 2 km and the Mie channel below, the median observation random errors for Aeolus compare favourably with radiosondes. At many altitudes, the Aeolus upper quartile and upper decile are better than radiosondes, implying a high yield.

It is possible to corroborate the findings of Section 4.2.1 by showing the distributions of Mie channel observa-

tion random error in the mid-latitudes and in particular the North Atlantic. Recall that use of model cloud cover results in overestimation of Aeolus data quality. This is manifested in Figure 16, upper panel, by the compact error distribution between 0 and 2 km, 90% of errors below 1.6 ms^{-1}. By constrast, LITE-inferred cloud cover results in more skewed error distributions (lower panel) with around 25% of errors exceeding 1.7 ms^{-1}. A similar increase in skewness is evident in the Rayleigh channel (figure not shown), but the effect is only noticeable in the upper decile. For example, at 4.5 km in the North Atlantic, use of model cloud cover in place of LITE cloud cover changes the Rayleigh channel upper decile observation error from 2.7 ms^{-1} to 4 ms^{-1}. The remainder of the error distribution is remarkably robust, with the corresponding changes in lower decile, median, and quartiles all being less than 1%.

4.3 Discussion

In Section 2, typical wind uncertainties in the ECMWF forecast system were shown to explain why ADM-Aeolus is expected to improve analyses (i) in the jet streams over the oceans, especially away from main air traffic routes, and in the African/Asian subtropical jet; (ii) in the lower troposphere, e.g. western parts of the N. Pacific and N. Atlantic oceans, if cloud gaps are sufficient, and (iii) in the tropics. The yield and accuracy of the simulated Aeolus data illustrated in this Section support these expectations. The simulated observation errors compare favourably with the errors assigned to radiosonde wind observations in an operational weather prediction environment. This provides further evidence that, by meeting the mission observation requirements as currently specified, Aeolus will have a significant impact on operational NWP.

An important aspect of these results has been to assess the impact of different cloud cover assumptions on simulated Aeolus data, focussing on the period 9–18 September 1994. The simulations were found to be robust to large variations in cloud cover, thus reducing previous uncertainties in anticipated performance. The study compared the yield and accuracy of simulated Aeolus data for two cloud cover scenarios: cloud cover is specified either by ECMWF model fields or by a cloud cover product derived from observations made during the Lidar In-space Technology Experiment (LITE). Although model fields underestimate the LITE-inferred cloud cover by 20% on average, Aeolus yield and accuracy under the two scenarios show good agreement in several measures. This is an encouraging result because the assimilation studies presented in Section 6 rely on model cloud cover fields for the simulation of Aeolus data. Overestimation of the penetration depth of Aeolus does arise when model fields are used, but the effect is small. Discrepancies are generally confined to the tails of error distributions. This needs to be taken into account in future work, particularly in developing quality control thresholds.

It should be noted that in the current simulations, the levels of yield and accuracy may be affected by the fact that each simulated observation over 50 km horizontally is the simple arithmetic average of 50 1 km accumulations. If there is sufficient cloud variability within such accumulations, these levels may only be achievable with more sophisticated averaging procedures. The study recommends that this be given due attention in development of the Level-1B and Level-2B processors.

5 Observation Processing for simulated Aeolus data in 4DVAR

In this section the report explains the technical modifications that have been implemented to assimilate simulated Aeolus data at ECMWF (Sections 5.1–5.4). These were tested by assimilating prototype Level-2B Aeolus data (Section 3.5) for the LITE period; examples are shown in Section 5.5.

At ECMWF, meteorological observations received in near-real-time flow through the stages:

Non-IFS observation processing (Bufr2ODB)

Observation Screening (including Background Quality Control)

4d-Variational Assimilation (including Variational Quality Control)

Diagnostic Post-processing

These four stages are described below. It is often convenient to refer to the last three stages as "IFS observation processing". They are distinguished from Non-IFS processing through the use of background information supplied by "first-guess" model fields (Section3.4).

5.1 Non-IFS observation processing (Bufr2ODB)

The role of non-IFS observation processing is to convert observational data from BUFR format to the format used in the Observation Data Base (ODB) – IFS's internal representation of observational data. The principles of non-IFS observation processing in the ECMWF data assimilation system (since Summer 2002) are described in Chapters 1 & 2 of [ECMWF 2003a]. One technical difference introduced in Summer 2003 relates to development of Bufr2ODB, the software to perform non-IFS processsing.

The current Study has introduced prototype Aeolus BUFR data to Bufr2ODB. That is, the information contained in the prototype Level-2B Aeolus data is converted to the internal representation required by the ODB. Within the ODB, Aeolus data are classified as a new subtype of the PILOT observation type because of their similarity to other PILOT subtypes (LAND and SHIP PILOTs and Wind Profilers). Furthermore, code has been introduced that allows a HLOS observation to be recognized as a new known observed variable. The azimuth angle is stored as an auxilliary value.

As mentioned in Section 3.4.3, the study assumes that the prototype Level-2B Aeolus data report the observation error $\sigma_{o,los}$, which accounts for both instrument random error $\sigma_{i,los}$ and representativity error $\sigma_{r,los}$. Should future DWL data report instrument random error only, it would be feasible to obtain observation error by adding representativity error during Non-IFS observation processing, according to eqn. 6.

5.2 Observation Screening (including Background Quality Control)

The purpose of observation screening is to select a quality-controlled set of observations to be used in 4-dimensional variational data assimilation [ECMWF 2003b, Chapter 10]; a more general description is given in [Järvinen and Undén 1997].

It is standard practice to apply a number of "independent" screening tests (independent of other observations) and these have been implemented for prototype Aeolus data. These include preliminary checks on the completeness of the reported observation (the report within the ODB must contain an observed value, location, vertical coordinate, observation time and an observation error, and a background departure must have been computed), a check on whether the data are blacklisted, and background quality control. Background quality control for simulated Aeolus observations is described in detail in the remainder of this section.

The so-called "dependent" screening tests (for removal of duplicated observations, thinning of higher-resolution data, etc) are not currently relevant for Aeolus. Their implementation is a straightforward technical exercise that is best addressed nearer mission launch if at all required.

5.2.1 *Standard Background QC*

In standard Background QC, the normalized background departure is calculated for each observation:

$$\mathscr{B} = \frac{y - Hx_b}{(\sigma_o^2 + \sigma_b^2)^{1/2}}, \tag{7}$$

where σ_b^2 is the background error variance in observation space. Because $\mathscr{B}$ is expected to be a Gaussian random variable (zero mean, unit variance), large values of $\mathscr{B}$ indicate suspect observations (i.e. departures much greater than expected from assumed error characteristics). The value of $\mathscr{B}$ is thus used to set the associated background quality control flag. The flag classifies the observation as "correct", "probably correct", "probably incorrect", or "incorrect". The status of the flag is used in subsequent decisions to accept or reject each observation. The thresholds for the flag classifications depend on the type of observation.

The study has implemented standard Background QC for prototype Aeolus observations. It was necessary to specify the background error variances in observation space, i.e. $\sigma_{b,los}^2$, as input to the data assimilation system. Because the current IFS assumes isotropy in wind errors, the background error variances specified for u and v wind components are equal, $\sigma_{b,wind}^2$ say. It follows that the same specification applies to HLOS wind component observations, and hence it is appropriate to set $\sigma_{b,los}^2 = \sigma_{b,wind}^2$.

Furthermore, in the case of HLOS wind observations, it is appropriate to apply the flag classification thresholds invoked for conventional wind observations. Specifically, using current thresholds, the Background QC flag for HLOS wind observations is set to

Correct for $|\mathscr{B}| < 3$,

Probably correct for $3 < |\mathscr{B}| < 4$,

Probably incorrect for $4 < |\mathscr{B}| < 5$,

Incorrect for $5 < |\mathscr{B}|$.

Only HLOS data classified as "Correct" or "Probably Correct" are accepted as active data for assimilation below.

5.2.2 *Aeolus-specific Background QC*

The study has implemented the following Aeolus-specific extensions to Background Quality Control. Specifically, for the purposes of computing $\mathscr{B}$, the HLOS observation error has been capped as follows:

$$\hat{\sigma}_o = \min\{\sigma_{o,los}, \sigma_o^{max}\}. \tag{8}$$

After various experimentation, the study settled on a constant limit $\sigma_o^{max} = 2.5$ ms^{-1}. The rationale for this extension is illustrated by the examples given in Section 5.5.

The study also experimented with an alternative limit, $\sigma_o^{max} = F\sigma_{b,los}$, with various values of F. These did not prove as effective and are not discussed further.

5.3 4d-Variational Assimilation of DWL observations and Variational Quality Control

The study has implemented the definitions described in Section 3.4 required for assimilation of HLOS wind component observations. In particular, the HLOS observation operator has been coded and implemented in the

forward, tangent linear, and adjoint codes of 4DVAR. The only additional aspect to be reported in this section concerns the implementation of Variational Quality Control (VarQC).

VarQC in ECMWF's 4DVAR system is described by Andersson and Järvinen [1999]. Its main purpose is to account for the possibility of gross errors in observational data, by the use of non-Gaussian observation error statistics. Whereas background quality control is applied prior to the main variational analysis algorithm, VarQC is applied within the variational analysis itself [Ingleby and Lorenc 1993]. VarQC is effective at identifying and rejecting data that are inconsistent with neighbouring data, within the analysis.

In the current ECMWF implementation of 4DVAR, all that is required to implement VarQC for simulated Aeolus observations is the specification of two parameters that determine the assumed distribution of non-Gaussian errors. The assumed distribution is uniform with width D and its height A is the prior probability of gross error [Lorenc and Hammon 1988; Dharssi *et al.* 1992]. These parameters are typically set by examining a large set of real observations. In the absence of real Aeolus data, the study has adopted the parameters for conventional wind observations, namely $D = 10\sigma_o$ and A=6×10^{-5}. These settings will need further examination when real Aeolus data become available.

5.4 Diagnostic post-processing of DWL observations

Simulated DWL observations have been readily incorporated into the standard diagnostic post-processing tools available at ECMWF. The tool used most extensively in the Study is known as `obstat`, and compiles statistics of observation diagnostics in the form of means (arithmetic and rms) and histograms. The observation diagnostics most commonly used are the background departures ($y - Hx_b$) and analysis residuals ($y - Hx_a$). The examples reported below also make use of the normalized background departure $\mathscr{B}$ and the normalized analysis fit $(y - Hx_a)/\sigma_o$. In an ideal data assimilation system (observations uncorrelated, no biases in observation or model) these two quantities both have zero mean and the former has unit variance [Daley 1991].

In addition, impacts of observational data on the assimilation system are routinely examined by a range of other diagnostics including analysis increment differences, forecast errors, forecast verification scores, and case studies. These diagnostics are all available for (simulated) Aeolus data.

5.5 Observation Processing Examples

The examples presented in this section were chosen to test the effectiveness of the study's new code for two standard processing modes anticipated for Aeolus data: passive monitoring and active assimilation. Section 5.5.1 shows the results of passive monitoring of prototype Aeolus data; such monitoring is routine for new data and is expected to be a key component of Aeolus in-orbit commissioning. This is followed by examples of actively assimilating the simulated Aeolus data, with emphasis on the differences between applying standard and Aeolus-specific background quality control (Sections 5.5.2 & 5.5.3). Analysis increments and data usage are addressed in Section 5.5.4. These results were obtained with the assimilation system run at operational resolution (T511, ∼40 km grid). The main conclusions also hold in examples of active assimilation at moderate resolution (T159, ∼125 km). In these observation processing examples, the observation statistics have been compiled for the period 10–19 September 1994 and are for the Rayleigh and Mie channels combined. Separate statistics for the Rayleigh and Mie channels are shown in the assimilation ensemble experiments (Section 6).

All examples are from assimilations performed with release Cy26r1 (ECMWF's operational system April–October 2003) with the extensions for simulated Aeolus data described in Sections 5.1–5.4. The assimilations cover the LITE period in 1994 and utilize the observations available in ERA-40, with the exception of SSM/I (which requires new bias corrections following the change introduced in Cy26r1 to assimiliate SSM/I radiances rather than wind speed and total column water vapour products). The prototype Aeolus data used in the

observation processing examples are derived from LIPAS simulations for the LITE period. These have been used to generate noisy (and biased) realisations of HLOS wind component observations coded in BUFR format (Section 3.5). LIPAS was supplied with input wind profiles and other meteorological fields taken from the same short-range forecasts that were used in Section 4, the same instrument parameters were used, the HLOS azimuth angle was similarly fixed at 90 degrees, and aerosol was specified by the median profile from the "Vaughan" climatology. The LIPAS simulations used in these observation processing examples pre-date those shown in Section 4 in one respect: the cloud cover profiles taken from the LITE product [Miller *et al.* 1999] were only utilized between 0 and 12 km. Each prototype Aeolus observation is reported with the observation time of the corresponding input wind profile. The number of profiles is limited by the available LITE cloud cover profiles (12389) and each profile contains up to 20 vertical levels. The maximum possible yield is thus 247780 HLOS data (100%).

The study makes four other remarks:

1. At each location and altitude, the prototype Aeolus observation is either the Rayleigh or Mie channel observation, whichever has the smaller random error estimate.

2. Error estimates exceeding 50 ms^{-1} are regarded as providing unreliable data and hence observations at such locations are reported as missing data. This accounts for a 4.64% loss from the maximum possible yield. The remaining 95% comprise 69% from Rayleigh channel observations and 26% from Mie channel observations. Including cloud cover above 12 km would increase Mie channel yield by around 5% and reduce Rayleigh channel yield by around 15%.

3. In addition to the systematic errors modelled by LIPAS (i.e. biases arising from instrument hardware and atmospheric inhomogeneities [Stoffelen *et al.* 2002]), the use of short-range forecasts for input winds introduces further bias in the simulated observations relative to the assimilation background fields.

4. In the prototype Level-2B Aeolus data used in observation processing examples, the representativity error used is appropriate for an assimilation system with 50 km horizontal resolution [Tan and Andersson 2003a]. In assimilation experiments at coarser resolution, this is an underestimate, which has the effect of giving too much weight to the simulated Aeolus observations.

5.5.1 Passive monitoring

Figure 17 shows the statistics for all simulated Aeolus observations that have been passively monitored in the assimilation system run at operational resolution (T511, ∼40 km grid, experiment "efh1"). Statistics have been compiled separately for the Northern hemisphere extratropics (20–90N), the Tropics (20S–20N), and the Southern hemisphere extratropics (20–90S), with root-mean-square profiles shown on the left and arithmetic mean (i.e. bias) profiles on the right. Background and analysis departures are shown as the solid and dotted black curves respectively (units of ms^{-1}), while these quantities normalized by their expected errors are shown as the red curves (dimensionless).

For reference, Figure 18 shows the statistics for radiosonde wind observations (zonal component) that have been actively used in the assimilation. By comparison, the simulated Aeolus data have somewhat larger rms errors (4 to 6 ms^{-1}). They also have significantly larger bias, arising from two main sources: (i) simulated instrument bias (around 1.4 and 0.9 ms^{-1} for the Rayleigh and Mie channels respectively) and (ii) from the short-term forecast winds used to simulate Aeolus data (between 1 and 3.5 ms^{-1} at 500 hPa).

Note that Aeolus rms departures in the extratropics increase with decreasing altitude. This behaviour is consistent with poorer signal-to-noise ratio for the Aeolus instrument when thick cloud must be penetrated. In the Tropics, the largest departures are found near the tropopause (around 100 hPa) where large wind speeds and cirrus clouds are encountered.

In spite of the biases present in the background departures for simulated Aeolus data, the rms profiles of normalized background departures (Bennett's diagnostic, Figure 17, solid red curves) are reasonably uniform. This indicates that the data are consistent with the observation errors reported in the simulated observations. They also compare favourably with the equivalent profiles for radiosonde wind data, which implies that they are given comparable weight when actively assimilated (results below).

Figure 19 shows the statistics for simulated Aeolus observations that have passed standard background quality control. The differences from Figure 17 are very slight, which again confirms that the simulated errors are consistent with those reported in the simulated observations. The most notable difference is a reduction in the rms background departure in the Tropics around 100 hPa. This is achieved by retaining just 5981 observations out of 6354 at this altitude (94.1%).

5.5.2 Active assimilation using standard quality control

The purple and green curves in Figure 20 show the results from actively assimilating the Aeolus data when standard quality control is applied. That the assimilation system has drawn to the data is evident from the difference between the background departures and analysis departures (solid purple and dotted purple curves respectively). While such behaviour is generally desirable in the rms profiles (left-hand plots), two further comments are relevant. The first is that the assimilation system has drawn to the biases present in the simulated data (right hand plots, dotted purple values smaller than dashed purple); this is certainly not desirable when the biases are artefacts in the observations rather than a deficiency in the background fields of the assimilating model.

The second point to note is that, especially at low altitudes in the southern hemisphere, the assimilation system has not drawn to the largest departures. This is because the largest departures are associated with low signal-to-noise ratio and large observation error, and hence get relatively little weight. This behaviour is entirely appropriate, and no adverse impact on the performance of the assimilation system is immediately apparent.

Nonetheless, the presence of large rms departures is undesirable because of a different use for these statistics, namely for the identification and monitoring of unforeseen errors in Aeolus data. Should such errors occur in future, it would be rather difficult to identify them in a time-series with a large rms error, such as that arising from standard quality control applied to Aeolus data. The situation would be much improved if quality control yielded Aeolus profiles more in keeping with those for radiosonde wind observations. This motivated the development of Aeolus-specific quality control procedures described in Section 5.2.2. There are illustrated in the following subsection.

5.5.3 Active assimilation using Aeolus-specific quality control

The aim of the Aeolus-specific quality control specified in Section 5.2.2 is to obtain observation statistics more in keeping with those for radiosonde wind observations. Because the statistics arising from standard Background QC are heavily influenced by a small percentage of observations with the largest observation errors (this is most evident in histograms not shown here), the Aeolus-specific procedures were devised to selectively reject those. The study tried various parameter settings and settled for capping the Background QC observation error at 2.5 ms^{-1}. The capping threshold was empirically determined but guided by the typical error assumed for radiosonde wind observations.

The black and red curves in Figure 20 show the results from actively assimilating the Aeolus data when Aeolus-specific quality control is applied. The column labelled "exp−ref" shows the change in the number of Aeolus observations accepted for use by the assimilation system (see also Table 3). The stricter quality control increases the rejection of Aeolus data by a small amount (around 1 percent), but the resulting profiles are much more in

Experiment	efh1	efh0	efgz
Comment	Passive	Standard QC	Aeolus QC
Background QC, accepted (rejected)	233883 (2394) 94.39% (0.97%)	235555 (722) 95.07% (0.30%)	234309 (1968) 94.56% (0.80%)
Used (VarQC-rejected)	not applicable	235454 (101) 95.03% (0.04%)	234216 (93) 94.53% (0.04%)

Table 3: HLOS observation usage at operational resolution. The maximum possible is 236277 (95.36% of 247780).

keeping with those for radiosonde wind observations. The study considers that the benefits are well worthwhile, because it will be easier to identify potentially more damaging errors in future Aeolus data.

5.5.4 Analysis increments and Data usage statistics

The impact of introducing simulated Aeolus data into the ECMWF assimilation system is illustrated by the analysis increments shown in Figure 21. The vectors show the change in the 500 hPa wind analysis that results from assimilating simulated Aeolus profiles along LITE orbits 005 and 006. The largest changes in the wind analysis are found along the orbit tracks: from 80 ° W, 54 ° N to 34 ° W, 20 ° N (orbit 005) and 74 ° W, 40 ° N to 35 ° W, 16 ° S (orbit 006). Because the innovations assimilated here contain contributions from the short-range forecast winds used to simulate Aeolus data, the magnitude of the analysis changes resulting from real Aeolus data may ultimately be smaller than those shown in Figure 21, perhaps by as much as a factor of two. Consequently, the report concentrates here on the overall structure of the changes, which is less likely to be affected.

The overall structure of the wind analysis changes is strongly influenced by the dynamical balance relationships of the assimilation scheme. These balances are an integral part of any assimilation scheme in that they augment the observations, in a dynamically consistent manner, to reduce the under-determinedness of the analysis problem. This augmentation is evident in Figure 21 from the fact that the vector directions of the analysis change bear little relationship to the HLOS azimuth angle. The multivariate nature of the balances also results in a change in the 500 hPa geopotential analysis (shaded), thereby maintaining approximate geostrophic balance in mid-latitudes.

The study has performed analogous assimilations to examine the effectiveness of Aeolus quality control at moderate resolution. At this resolution, Figure 22 shows the statistics for actively used Aeolus data. The contrast between standard QC and Aeolus-specific QC is very similar to that found at operational resolution (Figure 20). The number of data actively used at full resolution is marginally higher (Table 3 versus Table 4). This suggests that the differences in representativity error at the two resolutions are not too important in quality control decisions.

The most encouraging sign from the usage statistics shown in Tables 3 and 4 is the high percentage of simulated Aeolus data that are accepted for active use in data assimilation. Provided real Aeolus data have error characteristics similar to those simulated in this report, more than 94% would be regarded as suitable for assimilation. The results here are not only in accord with those of Section 4, but they are based on a sterner test of the simulated data. Whereas the previous results made off-line comparisons between simulated Aeolus data and current NWP systems, the comparison made here has placed the simulated data in direct competition with real observational data. Both comparisons of course rely on the modelling assumptions made in simulating Aeolus data.

Experiment	efce	efcd	efcc
Comment	Passive	Standard QC	Aeolus QC
Background QC, accepted (rejected)	232159 (4118) 93.70% (1.66%)	235063 (1214) 94.87% (0.49%)	233714 (2563) 94.32% (1.03%)
Used (VarQC-rejected)	not applicable	234989 (74) 94.84% (0.04%)	233648 (66) 94.30% (0.03%)

Table 4: HLOS observation usage at moderate resolution.

5.6 Discussion

Simulated Aeolus data have been introduced into ECMWF's observation processing and data assimilation system. This was achieved by introducing new code in the following areas:

- Non-IFS processing: Prototype Aeolus data, in the form of a Level-2B product in BUFR format, are converted to the representation used by the Observation Data Base within ECMWF's Integrated Forecast System. The code now allows a horizontal line-of-sight wind component observation to be recognized as a new known observed variable.

- Observation Screening/Background Quality Control: Checks for the completeness of each HLOS wind observation have been introduced. In addition to standard background quality control for Aeolus data, stricter Aeolus-specific controls have been developed and implemented. The Aeolus-specific controls reject additionally a small set of HLOS observations with large reported observation errors, in order to assist in identifying and correcting unforeseen deficiencies in real Aeolus data.

- 4d-Variational Assimilation/Variational Quality Control: the HLOS observation operator has been coded and implemented in the forward, tangent linear and adjoint codes of 4DVAR. Variational Quality Control has been implemented for HLOS wind observations, with parameter settings taken from those for conventional wind observations.

- Diagnostic post-processing: HLOS observations have been incorporated into the standard diagnostic post-processing tools available at ECMWF.

The effectiveness of the new code has been demonstrated for prototype Aeolus data that were simulated for the LITE period in September 1994.

The stricter Aeolus-specific quality controls relate to an attractive feature promised by Aeolus data, namely the reporting of an observation error estimate for each HLOS wind observation. They are designed so that statistics of HLOS observations are much more in keeping with those for conventional wind observations. The improved statistics are obtained at a relatively small cost (rejection of an additional 0.6% of Aeolus observations).

Estimates of anticipated Aeolus yield remain high. Around 94% of simulated Aeolus data were accepted for active assimilation in the study's observation processing examples. This re-inforces previous expectations that Aeolus will yield good quality data for NWP applications. Whereas the previous results made off-line comparisons between simulated Aeolus data and current NWP systems, the work reported here use the assimilation system to place the simulated data in direct competition with real observational data.

6 Assimilation Ensemble Method and Data Impact Experiments

In this section, the impact of simulated Aeolus data in the ECMWF forecast system is investigated further using an extension of the assimilation ensemble method. The extended method offers some advantages over traditional Observing System Simulation Experiments, but it is relatively new and has required calibration. This calibration was achieved by assessing the impact of radiosonde data in the same way. Descriptions of the method (Section 6.1) and the simulated Aeolus data used (Section 6.2) are followed by examples (Sections 6.3 and 6.4) and a discussion (Section 6.5).

6.1 The assimilation ensemble method for data impact assessment

The original motivation for the assimilation ensemble method stems from the need to estimate background errors for use in the ECMWF assimilation system [Fisher 2003b]. For this purpose, the method consists of generating an ensemble of analyses and associated short-range forecasts. Each ensemble member is an independent run of the ECMWF 4d-Var analysis/forecast system. The observations supplied to the analysis system differ between ensemble members, through the addition of random perturbations consistent with the observation errors assumed in the analysis system. That is, the observation perturbations are spatially uncorrelated and each is drawn from a Gaussian distribution with standard deviation specified by the observation error of the observation under consideration.

The underlying premise of the method is that differences in contemporaneous short-range forecasts, arising from different ensemble members, are a useful surrogate for background errors. Similarly, that differences in contemporaneous analyses are a useful surrogate for analysis errors. The method has been used to construct the background error statistics used in the ECMWF variational analysis since October 1999 [Fisher and Andersson 2001], and to formulate a new humidity analysis [Hólm *et al.* 2002]. The ensemble data have also been used for Aeolus-related studies of wind errors in the ECMWF system [Tan and Andersson 2003a] and of wind-mass balance relationships in the tropics [Žagar *et al.* 2004].

Here, the assimilation ensemble method is extended to assess the impact of different observation types (simulated Aeolus data and, for calibration purposes, radiosondes). This has required the generation of new ensembles that differ in the observations made available to the ensemble method. Four ensembles have been generated and each ensemble consists of 4 independent members; each member is run for the period 10th January 2003 to 28th February 2003. The four ensembles differ in the observations that are perturbed and assimilated as follows:

Control All data used in the 2004 ECMWF operational system,

DWL Control with simulated Aeolus data added,

NoSondes Control with radiosondes and wind profilers removed, and

DWL-NoSondes NoSondes with simulated Aeolus data added.

Statistics, such as the spread in a particular ensemble, are compiled for the period 16th January to 28th February (to disregard common initial conditions). The four members in a particular ensemble give rise to three independent difference fields for use in the spread calculations. A beneficial observation impact should correspond to a reduction in ensemble spread. The report concentrates on the impact of radiosondes (Control vs NoSondes) and the impact of adding Aeolus data (DWL vs Control/NoSondes). Comparisons involving the DWL-NoSondes ensemble give alternative impact measures. Removal of radiosonde observations in ensembles NoSondes and DWL-NoSondes extends to radiosonde temperature and humidity observations as well as wind observations.

The approach of removing real observation types (or adding simulated ones) in the assimilation ensemble method resembles the approach taken in traditional Observing System (Simulation) Experiments (OSEs/OSSEs). There are two perceived advantages of the assimilation ensemble method over OSSEs for data impact studies. The first is that, for existing observation types, the ensemble method perturbs real observations and thus eliminates the need to simulate such observations from a common nature run; the nature run is only required for simulating new observation types such as Aeolus. (This has made it feasible to include AIRS radiances in the study's experiments; by contrast previous OSSEs conducted for Aeolus were unable to include TOVS data [Cardinali *et al.* 1998; Marseille *et al.* 2001]). The second is that the data impact measures in the ensemble method are based on relative differences such as ensemble spread, and are thus thought to be less sensitive than OSSE impact measures to the uncertainties surrounding the treatment of observation/simulation bias. The study acknowledges however that the assimilation ensemble method for data impact assessment is entirely new and that many aspects of its performance remain un-explored.

6.2 Prototype Level-2B Aeolus observations for the ensemble period

The baseline set of simulated Aeolus observations, to which perturbations were added in the analysis ensemble experiments, was generated using the LIPAS simulator. The procedure used improved on that used for the observation processing examples (Section 5.5) in several respects. An orbit simulator was used to specify the orbit locations and HLOS azimuth angles of the planned Aeolus orbit. Met Office analysed winds were sampled at each location to provide input to LIPAS. Because the Met Office analysed winds were available at six-hourly intervals, each was sampled at the orbit locations for a six hour period (approximately 750 profiles, Figure 2).

In previous sections, LIPAS simulations used an observed cloud cover product derived from the LITE mission (9–18 September 1994), whereas for the assimilation ensembles it was necessary to rely on model cloud cover fields from the period Jan–Feb 2003. These and other meteorological inputs to LIPAS, principally profiles of temperature and cloud water content, were obtained from the ECMWF operational archive. The observation quality simulated by LIPAS for the ensemble period is shown in Figure 23. The agreement with the LITE period quality using observed cloud (Figure 12) is rather good.

A representativity error for 150 km input winds (the resolution available from the Met Office) was also incorporated in the baseline dataset in both the observed values and their reported observation errors. One consequence is that the simulated DWL data receive less weight in the assimilation ensemble experiments than might be expected for real Aeolus data.

Before examining the assimilation ensembles, it is useful to consider a reference assimilation in which the baseline set of simulated Aeolus observations is assimilated in addition to all operational data. The reference assimilation and all of the assimilation ensembles were generated with 12-hour 4d-Var using Cycle 26r3 of the ECMWF system (operational from Autumn 2003) run at T159 resolution (approx 120 km grid). Observation statistics are shown in Figures 24 & 25. The assimilation system has diagnosed a bias of approximately 0.9 ms^{-1} for Mie channel observations and 1.4 ms^{-1} for Rayleigh channel observations, in agreement with the parameter settings adopted in LIPAS from Aeolus Phase A studies.

It is useful to compare the data counts for DWL observations with those for TEMP and PILOT. The figures given in the centre column of Figures 26 & 27 should be doubled to allow for observations of both wind components. Radiosondes are of course heavily weighted towards the Northern Hemisphere; by contrast DWL observations outnumber radiosonde observations in the Southern Hemisphere and the Tropics.

While it is easy to obtain realistic data counts for simulated Aeolus data, it is more difficult to simulate the anticipated quality in terms of rms errors. For example, it is expected that Aeolus data will be of comparable quality to radiosondes. This is manifested in the baseline dataset for the Mie channel, for which rms background departures are comparable to TEMPs and PILOTs, but Rayleigh channel departures are greater by around 0.5 to 1.0 ms^{-1}. Two aspects of the simulations that contribute to this are the simulated bias and 150 km

representativity error. The study thus believes that the current simulations are conservative regarding DWL quality.

6.3 Analysis impact of radiosondes and DWL data

The impacts of radiosonde and simulated Aeolus data are most apparent in terms of analysis spread, so these are described first. Short-range forecast impact (T+12, corresponding to background errors) are in the following subsection.

Figure 28 (upper panel) shows the difference in the spread of analysed zonal wind at 500 hPa between the ensembles Control and NoSondes. Negative values correspond to a reduction in spread due to the addition of radiosondes, and hence green and blue areas indicate regions where radiosondes have a beneficial impact on the assimilations. The patchy appearance and the presence of positive values are consequences of the finite statistical sample. Nonetheless the most coherent spatial features correlate well with the coverage of the radiosonde network and so a meaningful interpretation of the diagnostic is possible.

In the Northern Hemisphere it is not surprising to find the main benefits correspond to the regions with good coverage by radiosondes and wind-profilers: North America, Europe and Japan. In the tropics, where wind analyses are less certain, the benefits are somewhat bigger in magnitude. They are still confined to regions observed by the radiosonde network (but of course the assimilation system and its background error covariances play a role in spreading the benefits to surrounding regions). While radiosonde impact shows a similar geographical distribution at other altitudes, it is interesting to note the reduction in radiosonde impact over North America at 200 hPa (Figure 28, lower panel). This can be attributed to the abundance of aircraft data at this altitude. Other illustrations of the vertical structure of data impact are given in the next subsection. Overall, the radiosonde impacts in the assimilation ensemble technique have many features in common with traditional OSE assessments, e.g. [Cress and Wergen 2001], and thus provide some justification for using the new technique to investigate data impacts. Consequently the report now turns to the impact of DWL data.

The benefits of adding DWL data to the current observing system are shown in Figure 29 (difference between ensembles DWL and Control). As for radiosonde impact, the finite statistical sample results in a rather noisy appearance. The effect is exacerbated by the distributed nature of the simulated orbit locations. Nonetheless, it is apparent that from a sufficiently coarse-grain perspective that ensemble spread is reduced when DWL data are introduced. At 500 hPa (upper panel) the benefits are principally over regions where wind observations are most lacking: the southern hemisphere and tropical oceans, the North Atlantic and the North Pacific (especially beyond the influence of Japanese wind profilers). At 200 hPa (lower panel), the benefits of DWL in the Tropics are more widespread than those of radiosondes, consistent with the comprehensive geographical coverage of the Aeolus orbit.

Maps of forecast impact are similar, but with interpretation made more difficult by the effects of advection and error propogation, and hence are not shown. Instead, area-averaged profiles of analysis and forecast impact are presented in the next section.

6.4 Profiles of analysis and forecast impact

Area-averaged profiles provide further perspectives on the impacts of DWL and radiosonde data. For ease of comparison, the ensemble differences are taken relative to the NoSonde ensemble, i.e. DWL vs NoSondes and Control vs NoSondes. In Figure 30 these are the green and red curves respectively (with more negative values representing a beneficial impact), and the additional benefit of DWL relative to the current observing system is indicated by the area between these curves.

The results for analysis impact are shown in Figure 30(a–c) for the North Atlantic, the Tropics and the Southern

Hemisphere. The plots confirm that DWL is beneficial, or at least neutral, throughout the depth of the free troposphere. The largest DWL benefit is in the Tropics above 200 hPa as remarked above.

Short-range forecast impact is shown in Figure 30(d–f). The results are for 12-hour forecasts because these are relevant for background errors in the assimilation system, and also because of the link between improvements in short-range forecasts and longer-range forecasts. It is clear that the reduction in analysis spread is carried over into the short-term forecasts. This is confirmed by maps (not shown) with the forecast benefit regions being found downstream of the analysis benefit regions.

As a final example, Figure 31 shows short-range forecast ensemble spread, averaged over the Tropics, in absolute terms for the Control, DWL and NoSondes ensembles. A calibration factor of order 2 is required to obtain values commensurate with actual wind errors in NWP models (e.g. Section 2). It is evident nonetheless that the impact of simulated Aeolus data compares favourably with radiosonde data. Apart from the small differences between the DWL and Control profiles between 1000 and 850 hPa, all other differences are found to be significant according to Student's T-test ($p < 0.0007$).

6.5 Discussion

The study has demonstrated the feasibility of investigating data impacts through the assimilation ensemble technique. The original technique has been extended and applied to the period Jan–Feb 2003. The main benefits from simulated Aeolus data for analysed wind fields were found over ocean regions in both hemispheres and in the Tropics. These regions have been identified previously as priority areas for improvement. The benefits seen in analysed fields led to benefits in forecast fields as well, throughout the depth of the troposphere. The novelty of the extended technique meant that calibration was required; this was achieved through assessing radiosonde impact in the same way.

Given the positive results obtained to date, the Study believes that further development of the technique would be worthwhile. In particular, the realism of the ensemble members would be increased by modelling correlated error perturbations for real observations, and there is scope for corroborating the findings of this study through other diagnostics such as information content [Fisher 2003a] and self-sensitivity [Cardinali *et al.* to appear].

It is possible to argue for a reduction in the representativity error attributed to simulated data, particularly when the nature run for the simulated data is independent of the assimilating system. This has not been pursued in the current study and so the measures of Aeolus impact shown here can be regarded as conservative.

Specific to Aeolus, it would be interesting to separate the impact of Mie and Rayleigh channels. Some insight can be obtained through further analysis of the existing ensembles, particularly where the data counts are very different, but it may also be useful to run further ensembles with data from one or other channel removed.

It is well-known that, in order to obtain quantitative agreement with other measures of background error, the assimilation ensemble technique requires re-scaling of the ensemble forecast spread. By analogy the study expects the spread differences, reported in this section and attributed to radiosonde and DWL data impact, to require re-scaling in order to match traditional (i.e. OSE and OSSE) measures of data impact.

7 Overall Conclusions and Recommendations

Interim conclusions are given in the Discussion sections (2.3, 4.3, 5.6, 6.5). There are consolidated here and further recommendations are made.

7.1 Overall Conclusions

Provision of high-quality observations of the global wind field remains an important objective for advancing numerical weather prediction. In Section 2 this report has shown typical wind uncertainties in the ECMWF forecast system, reflecting in part the deficiencies in the current global observing system. The observation requirements defined for the Aeolus mission suggest that the data should significantly improve NWP. In providing wind profiles with high vertical resolution, Aeolus is expected to improve analyses (i) in the jet streams over the oceans, especially away from main air traffic routes, and in the African/Asian subtropical jet; (ii) in the lower troposphere, e.g. western parts of the N. Pacific and N. Atlantic oceans, if cloud gaps are sufficient; and (iii) in the tropics.

Simulations of Aeolus yield and accuracy (Section 4) support these expectations. Simulated Aeolus observation errors compare favourably with the errors assigned to radiosonde wind observations in an operational weather prediction environment. This provides further evidence that, by meeting the mission observation requirements as currently specified, Aeolus will have a significant impact on operational NWP. The Aeolus simulations are found to be robust to large variations in cloud cover, thus reducing previous uncertainties in anticipated performance. The impact of different cloud cover assumptions on simulated Aeolus data was assessed by focussing on the period 9–18 September 1994 and comparing the yield and accuracy of simulated Aeolus data for two cloud cover scenarios: cloud cover is specified either by ECMWF model fields or by a cloud cover product derived from observations made during the Lidar In-space Technology Experiment (LITE). Although model fields underestimate the LITE-inferred cloud cover by 20% on average, Aeolus yield and accuracy under the two scenarios show good agreement in several measures. Overestimation of the penetration depth of Aeolus does arise when model fields are used, but the effect is small. Discrepancies are generally confined to the tails of error distributions. This needs to be taken into account in future work, particularly in developing quality control thresholds.

Simulated Aeolus data have been introduced into ECMWF's observation processing and data assimilation system. This was achieved by introducing new code in the following areas: (i) Non-IFS processing, (ii) Observation Screening/Background Quality Control, (iii) 4d-Variational Assimilation/Variational Quality Control (iv) Diagnostic post-processing. Details are given in Section 5, together with examples to demonstrate the effectiveness of the new code. In addition to standard background quality control for Aeolus data, stricter Aeolus-specific controls have been developed and implemented. The Aeolus-specific controls reject additionally a small set of HLOS observations with large reported observation errors, in order to assist in identifying and correcting unforeseen deficiencies in real Aeolus data. The stricter Aeolus-specific quality controls relate to an attractive feature promised by Aeolus data, namely the reporting of an observation error estimate for each HLOS wind observation. They are designed so that statistics of HLOS observations are much more in keeping with those for conventional wind observations. The improved statistics are obtained at a relatively small cost (rejection of an additional 0.6% of Aeolus observations). Estimates of anticipated Aeolus yield remain high. Around 94% of simulated Aeolus data were accepted for active assimilation in observation processing tests. This re-inforces previous expectations that Aeolus will yield good quality data for NWP applications. Whereas the previous results made off-line comparisons between simulated Aeolus data and current NWP systems, the study's observation processing tests have placed the simulated data in direct competition with real observational data.

Simulated Aeolus data have been assimilated in Assimilation Ensemble experiments (Section 6). The original technique has been extended to permit the data impacts to be assessed. The novelty of the extended technique meant that calibration was required; this was achieved through radiosonde denial experiments in the assimilation ensemble technique. The new technique suggests that the main benefits from Aeolus for analysed wind fields will be found over ocean regions in both hemispheres and in the Tropics. These regions have been identified previously as priority areas for improvement. The benefits seen in analysed fields lead to benefits in forecast fields as well, throughout the depth of the troposphere. The ensemble approach has involved assimila-

tion of simulated Aeolus Level-2B data in many 6-week assimilations with each ensemble member assimilating differently perturbed observations. The treatment of Level-2B data developed in this study has proved robust in all cases, and provides a good foundation for the development of Level-2C products.

7.2 General Discussion and Recommendations

It is becoming apparent that the major issues for successful NWP exploitation of Aeolus data are a) the ability to produce high-quality Level-2B data in near-real-time, and b) good characterization of observation and background ("model first-guess") errors within the assimilation system.

Achieving (a) is closely linked with development of Level-1B and Level-2B operational processors. The study recommends that this development pays attention to characterizing anticipated Aeolus errors, including bias and the effects of cloud inhomogeneity on vertical and horizontal scales of 1 km or less. For example, in the current simulations, the levels of yield and accuracy obtained may be affected by the fact that each simulated observation over 50 km horizontally is the simple arithmetic average of 50 1 km accumulations. If there is sufficient cloud variability within such accumulations, such levels might only be achievable by more sophisticated averaging procedures. The study recommends that this be given due attention, partciularly when aircraft campaign data become available for testing processor algorithms.

The work of Žagar *et al.* [2004] contributes to (b) especially in the Tropics. It also argues convincingly for the need to develop an approach to data assimilation that combines tropical and mid-latitude dynamics and produces balanced increments globally; the study recommends further research in this area.

The study recommends further development of the assimilation ensemble technique as discussed in the previous section. The technique offers a tractable alternative to OSSEs that would be useful during further preparation for Aeolus and for other missions.

The study conclusions and recommendations are now re-iterated in bullet-point form:

C1: ADM-Aeolus will primarily benefit global NWP by improving wind analyses in the jet streams, over the oceans and in the Tropics, provided the data meet previously stated mission requirements on observation quality.

C2: Assimilation of prototype Aeolus observations, incorporating Aeolus-specific quality control, has been developed & tested in this study and is a suitable starting point for the development of Level-2C products.

C3: For effective exploitation of Aeolus data in NWP applications, the development of high-quality Level-1B and Level-2B processors is of major importance. In particular, their treatment of cloud effects will have a critical impact on mission products.

R1: Develop Level-1B and Level-2B processors that ensure Aeolus data meet mission observation requirements, paying particular attention to cloud inhomogeneity on vertical and horizontal scales of 1 km or less.

R2: Conduct further impact studies, coordinated with development of Level-1B/2B processors and the availability of aircraft campaign data.

Acknowledgements

The authors thank the Met Office for granting DT a secondment to work on this study. Ad Stoffelen and Gert-Jan Marseille assisted with LIPAS, as did Steve Miller, Dave Winker and Frederic Chevallier with LITE-inferred cloud cover. Richard Barnes and Kelvyn Robertson provided the Met Office wind fields used in

this study, and Nedjeljka Žagar the code to compute azimuth angles from orbit locations. Milan Dragosavac, Sami Saarinen and Drasko Vasiljevic advised on technical aspects of this work, and Lars Isaksen incorporated Aeolus observations into `obstat`. The work has benefitted from many useful discussions with colleagues in ECMWF's Research Department, the ADM-Aeolus Mission Advisory Group, and the Aeolus project team.

A Glossary (Data Products Contents)

A list of Aeolus data products is given by the current issue of the ADM-Aeolus Data Products Contents Guidelines [ESA 2004]. The Guidelines also identify the primary parameters generated by the on-ground processing chains and included in the various data structures. The main definitions relating to Level-1B and Level-2B data are repeated here for completeness.

Level-1B "geo-located, fully calibrated observational data".

An "engineering level" product containing fully processed Aeolus atmospheric scene and calibration data, engineering calibration applied.

Measurement data to include

- On-board reference time ("time stamp", UTC) of actual observation
- Geolocation of line-of-sight target and viewing direction in topocentric coordinates
- Profiles of signal strength and LOS Doppler shift
- HLOS wind profiles for individual measurements or groups thereof (i.e. on the 1 to 10 km horizontal scale) based on standard pressure and temperature profiles

Level-2B "Aeolus consolidated HLOS wind observations"

A "geophysical level" product containing HLOS wind profiles using correct atmospheric state parameters (i.e. best-estimates of temperature and pressure as available from NWP models).

Measurement data to include

- On-board reference time (UTC) for actual measurement cycle
- Spacecraft, LOS target and viewing direction parameters as in Level-1B product
- HLOS wind profiles using "correct" atmospheric state parameters
- Fully processed profile quality information (variance/covariance matrices, others) as required as input to wind assimilation processing. Estimates shall be based on Level-1B Product Confidence Data and external geophysical information
- Atmospheric variabilities, gradients etc over Basic Repeat Cycle measurement

References

Andersson, E. and H. Järvinen, 1999: Variational quality control. *QJRMS* **125**:697-722.

Astin, I., 1997: Sampling errors and bias in satellite-derived fractional cloud cover estimates from exponential and deterministic cloud fields as a consequence of instrument pixel size and number. *J. Atmos. Oceanic. Technol.* **14**:1146–1156.

Baker, W. E., G. D. Emmitt, F. Robertson, R. M. Atlas, J. E. Molinari, D. A. Bowdle, J. Paegle, R. M. Hardesty, R. T. Menzies, T. N. Krishnamurthi, R. A. Brown, M. J. Post, J. R. Anderson, A. C. Lorenc and J. McElroy, 1985: Lidar-measured winds from space: A key component for weather and climate prediction. *Bull. Amer. Meteorol. Soc.* **76**:869–888.

Browning, K. A., A. J. Thorpe, A. Montani, D. Parsons, M. Griffiths, P. Panag and E. M. Dicks, 2000: Interactions of tropopause depressions with an ex-tropical cyclone and sensitivity of forecasts to analysis errors. *Mon. Wea. Rev.* **128**:2734–2755.

Buizza, R. and A. Montani, 1999: Targeting observations using singular vectors. *J. Atmos. Sci.* **56**:2965–2985.

Cardinali, C., J. Pailleux and J.-N. Thepaut, 1998: Use of simulated satellite Doppler wind lidar data in NWP: An impact study. *CNRM, Météo France,* 56pp.

Cardinali, C., L. Isaksen and E. Andersson, 2003: Use and impact of automated aircraft data in a global 4D-Var data assimilation system. *Mon. Wea. Rev.* **131**:1865-1877.

Cardinali, C., S. Pezzulli and E. Andersson, : Influence matrix diagnostic of a data assimilation system. *QJRMS accepted*.

Chevallier, F., P. Bauer, G. Kelly, C. Jakob and T. McNally, 2001: Model Clouds over Oceans as Seen from Space: Comparison with HIRS/2 and MSU radiances. *J. Climate* **14**:4216–4229.

Chevallier, F. and G. Kelly, 2002: Model Clouds over Oceans as Seen from Space: Comparison with Geostationary Imagery in the 11-μm Window Channel. *Mon. Wea. Rev.* **130**:712–722.

Chepfer, H., J. Pelon, G. Brogniez, C. Flamant, V. Trouillet and P.H. Flamant, 1999: Impact of cirrus cloud ice crystal shape and size on multiple scattering effects: application to spaceborne and airborne backscatter lidar measurements during LITE mission and E LITE campaign. *GRL* **26**:2203–2206.

Cress, A. and W. Wergen, 2001: Impact of profiles observations on the German Weather Service's NWP system. *Meteor. Zeit.* **10**:91–101.

Daley, R., 1991: Atmospheric data analysis. Cambridge:CUP, 457pp.

Dharssi, I., A. Lorenc and N.B. Ingleby, 1992: Treatment of gross errors using maximum probability theory. *QJRMS* **118**:1017–1036.

European Centre for Medium-Range Weather Forecasts, 2003: IFS Documentation. Part I: Observation Processing (Cycle CY25R1). **ECMWF** (Peter White *ed.*).

European Centre for Medium-Range Weather Forecasts, 2003: IFS Documentation. Part II: Data Assimilation (Cycle CY25R1). **ECMWF** (Peter White *ed.*).

European Space Agency, 1999: The Four Candidate Earth Explorer Core Missions – Atmospheric Dynamics Mission. **ESA SP-1233(4)**, 157 pp.

European Space Agency, 2004: ADM-Aeolus Data Products Contents Guidelines. **ESA AE-TN-ESA-SY-007** (issue 1A, 30 pp).

Fisher, M., 2003a: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Tech.Memo.* **397**, 18 pp.

Fisher, M., 2003b: Background error covariance modelling. In *Proceedings of ECMWF Seminar "Recent developments in data assimilation for atmosphere and ocean, 8–12 September 2003"*, Reading, UK, pp.45–64

Fisher, M. and E. Andersson, 2001: Developments in 4D-Var and Kalman Filtering. *ECMWF Tech.Memo.* **347**, 38 pp.

Fisher, M. and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Tech.Memo.* **220**, 26 pp.

Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus* **38A**:111–136.

Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J-F. Mahfouf, A.J. Simmons and J-N. Thepaut, 2002: Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans. *ECMWF Tech.Memo.* **383**, 55 pp.

Ingleby, N.B. and A. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *QJRMS* **119**:1195–1225.

Isaksen, L. and P. A. E. M. Janssen, 2004: Impact of ERS scatterometer winds in ECMWF's assimilation system. *QJRMS* **130**:1793–1814.

Järvinen, H. and P.Undén, 1997: Observation screening and background quality control in the ECMWF 3D-Var data assimilation system. *ECMWF Tech.Memo.* **236**, 33 pp.

Järvinen, H., E. Andersson and F. Bouttier, 1999: Varational assimilation of time sequences of surface observations with serially correlated errors. *Tellus* **51A**:469-488.

Jakob, C. and S.A. Klein, 2000: A parametrization of the effects of cloud and precipitation overlap for use in general-circulation models. *QJRMS* **126**:2525–2544.

Klinker, E., F. Rabier and R. Gelaro, 1998: Estimation of key analysis errors using the adjoint technique. *QJRMS* **124**:1909–1933.

Klinker, E., F. Rabier, G. Kelly and J.-F. Mahfouf, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *QJRMS* **126**:1191–1215.

Lord, S.J, M. Masutani, J. S. Woollen, J. C. Derber, R. E. Kistler, T. J. Kleespies, H. Sun, G. D. Emmitt, S. A. Wood, S. Greco, J. Terry and R. Atlas, 2003: Observing systems simulation experiments using the NCEP data assimilation system (2003-7IOS). AMS 7th Symposium on Integrated Observing Systems: The Water Cycle, 9-13 February 2003, Long Beach, CA.

Lorenc, A. and O. Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *QJRMS* **114**:515–543.

Lorenc, A., R.J. Graham, I. Dharssi, B. Macpherson, N.B. Ingleby and R.W. Lunnon, 1992: Study of Preparation for the Use of Doppler Wind Lidar information in Meteorological Assimilation Systems. Final Report, ESA Study Contract 9063/90/HGE-I.

Mahfouf J.-F. and F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part II: Experimental results with improved physics. *QJRMS* **126**:1171–1190.

Marseille, G. J. and F. Bouttier, 2001: Climatologies of sensitive areas for short-term forecast errors over Europe – EUMETNET-EUCOS study. *ECMWF Tech.Memo.* **334**, 50 pp.

Marseille, G.J. and A. Stoffelen, 2003: Simulation of wind profiles from a space-borne Doppler wind lidar. *QJRMS* **129**:3079–3098.

Marseille, G.J., A. Stoffelen, F. Bouttier, C. Cardinali, S. de Haan and D. Vasiljevic, 2001: Impact assessment of a doppler wind lidar in space on atmospheric analyses and numerical weather prediction. *KNMI Scientific Report* **WR 2001-03**, 56pp.

Marseille, G.J., A. Stoffelen and A. van Lammeran: A novel method for aerosol backscatter retrieval from LITE data. *J. Geophys. Res.*, submitted.

McCormick, M. P., D. M. Winker, E. V. Browell, J. A. Coakley, C. S. Gardner, R. M. Hoff, G. S. Kent, S. H. Melfi, R. T. Menzies, C. M. R. Platt, D. A. Randall and J. A. Reagan, 1993: Scientific investigations planned for the Lidar In-Space Technology Experiment (LITE). *Bull. Amer. Meteorol. Soc.* **74**:205–214.

McNally, A.P., 2002: A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *QJRMS* **128**:2551–2556.

Miller, S.D., G.L. Stephens and A.C.M. Beljaars, 1999: A Validation Survey of the ECMWF Prognostic Cloud Scheme using LITE. *GRL* **26**:1417–1420.

Rabier, F., E. Klinker, P. Courtier and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *QJRMS* **122**:121–150.

Rabier, F., J.-N. Thépaut and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *QJRMS* **124**:1861–1887.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *QJRMS* **126**:1143–1170.

Rapp, R.R., C. Schutz and E. Rodriguez, 1973: Cloud-Free Line-of-Sight Calculations. *J. Appl. Meteorol.* **12**:484–493.

Rohaly, G.D. and T.N. Krishnamurthi, 1993: An Observing System Simulation Experiment for the Laser Atmospheric Wind Sounder (LAWS). *J. Appl. Meteorol.* **32**:1453–1471.

Schmetz, J., K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch and L. van De Berg, 1993: Operational cloud-motion winds from Meteosat infrared images. *J. Appl. Meteorol.* **32**:1206–1225.

Simmons, A. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.* **128**:647–687.

Stoffelen, A. and G.J. Marseille, 1998: Study on the Utility of a Doppler Wind Lidar for Numerical Weather Prediction and Climate. **ESA-CR(P)-4198**, 56pp+apps+figs.

Stoffelen, A., M. Håkansson and G.J. Marseille, 2002: Measurement Error and Correlation Impact on the Atmospheric Dynamics Mission: Report on Tasks 3 and 4. *Study Report* ESA Contract **RFQ/3-9992/01/NL/MM**, 75pp.

Stoffelen, A., J. Pailleux, E. Källén, J. M. Vaughan, L. Isaksen, P. Flamant, W. Wergen, E. Andersson, H. Schyberg, A. Culoma, R. Meynart, M. Endemann and P. Ingmann,: The European Atmospheric Dynamics Mission ADM-Aeolus for Global Wind Field Measurement. *Bull. Amer. Meteorol. Soc., accepted.*

Tan, D.G.H. and E. Andersson, 2003a: Yield of simulated ADM-Aeolus measurements. *Technical Note 1, ESA Contract 15342*, Study manager Dr P. Ingmann.

Tan, D.G.H. and E. Andersson, 2003b: Treatment of simulated ADM-Aeolus observations in the ECMWF 4DVAR system. *Technical Note 2, ESA Contract 15342*, Study manager Dr P. Ingmann.

Tan, D.G.H. and E. Andersson, 2004: Assimilation of simulated ADM-Aeolus observations in analysis ensemble experiments. *Technical Note 3, ESA Contract 15342*, Study manager Dr P. Ingmann.

Tan, D.G.H. and E. Andersson,: Simulation of the yield and accuracy of wind profile measurements from the Atmospheric Dynamics Mission (ADM-Aeolus). *QJRMS submitted*.

Thépaut, J.-N. and E. Andersson, 2003: Assimilation of high-resolution satellite data. *ECMWF Newsletter* **97**:6–11.

Uppala, S., 1997: Observing System Performance in ERA. *ECMWF Re-Analysis Final Report Series* **3**, 261pp.

Vaughan, J.M., N.J. Geddes, P.H. Flamant and C. Flesia, 1998: Establishment of a backscatter coefficient and atmospheric database. *Study Report* **ESA-CR12510**.

Veldman, S.M., H.A. Knobbout, A. Stoffelen, G.J. Marseille and E.A. Kuijpers, 1999: Lidar Performance Analysis Simulator (LIPAS). Part 8: Summary. *Study Report* ESA Contract **12718/98/NL/GD**, 50pp.

Winker, D.M. and M.A. Vaughan, 1994: Vertical distribution of clouds over Hampton, Virginia observed by lidar under the ECLIPS and FIRE ETO programs. *Atmos. Res.* **34**:117–133.

Winker, D.M., R.H. Couch and M.P. McCormick, 1996: An overview of LITE: NASA's Lidar in-space Technology Experiment. *Proc. IEEE* **84**:164–180.

World Meteorological Organization, 2001: Statement of Guidance regarding how well Satellite and In Situ Sensor Capabilities meet WMO User Requirements in Several Application Areas. *WMO Satellite Reports SAT-26* **WMO/TD No. 1052**.

Yu, S., K.E. Case and J. Chernick, 1986: Methodologies and comparisons for Lund's two methods for calculating probability of Cloud-Free Line-of-Sight. *J. Climate and Appl. Meteorol.* **25**:389–397.

Žagar, N., E. Andersson and M. Fisher, 2004: Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *ECMWF Tech.Memo.* **437**, 28 pp.

Žagar, N., N. Gustafsson and E. Källén,: Variational data assimilation in the tropics: the impact of background error constraint. *Q. J. R. Meteorol. Soc.*, submitted.

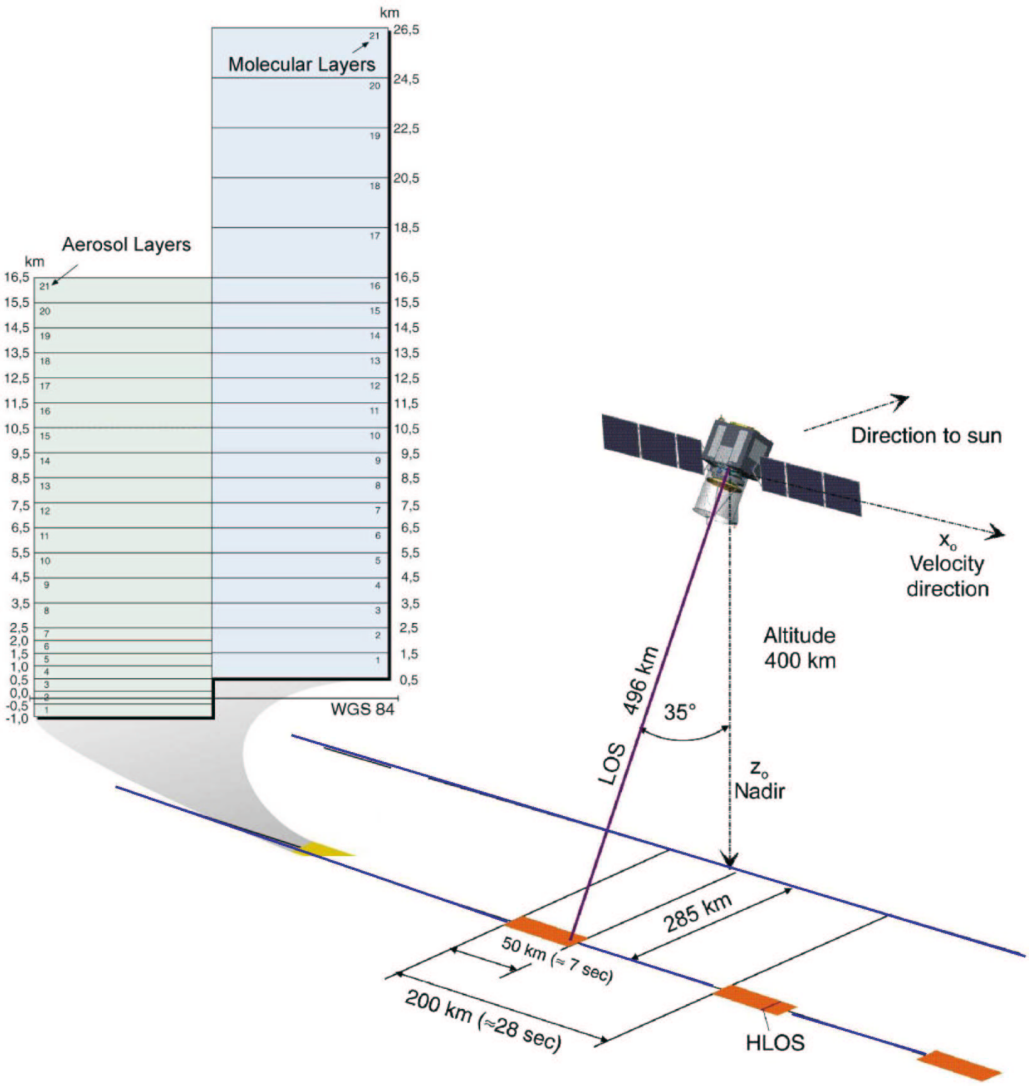


Figure 1: Vertical resolution and viewing geometry of the Aeolus satellite. (Source: ESA website http://www.esa.int)

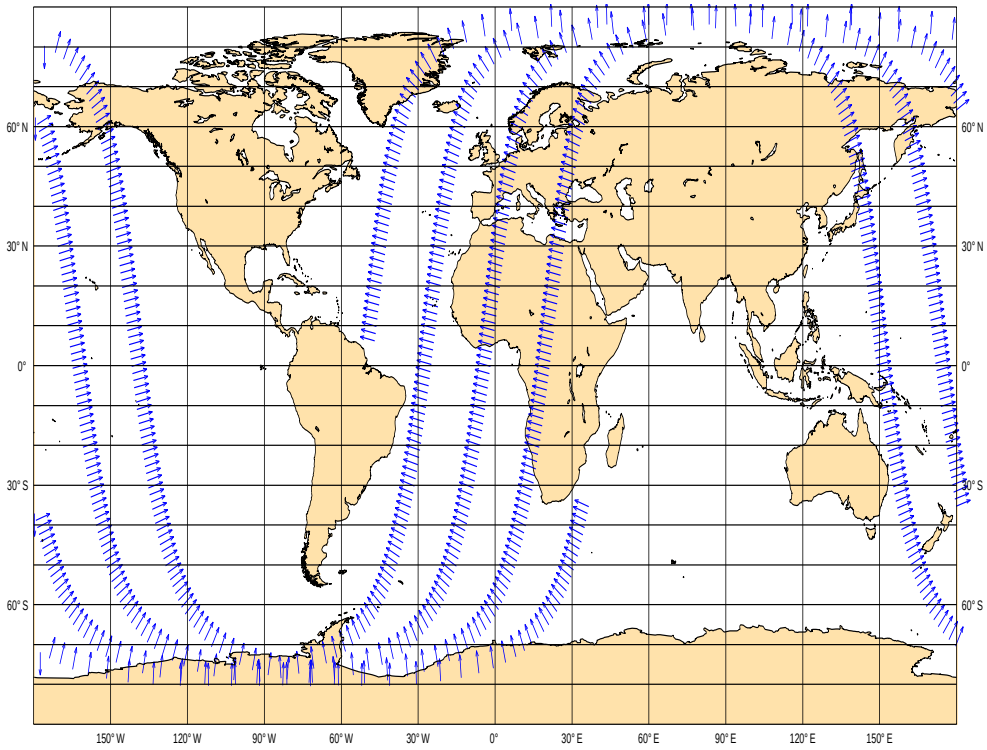


Figure 2: Simulated orbit coverage and HLOS direction for a 6-hour period.

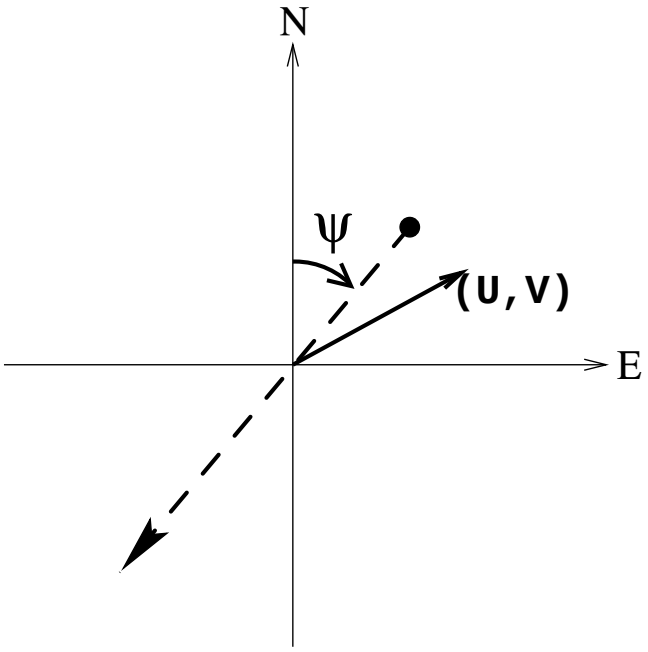


Figure 3: Geometry for horizontal line-of-sight wind observation. The Doppler wind lidar direction (dashed line) makes an azimuth angle ψ with geographical north. For the wind vector components (u, v) drawn in this illustration, the HLOS wind component is a negative value (filled circle). For actual ADM-Aeolus data, ψ is $270\,^\circ \pm 7\,^\circ$ in the Tropics with larger deviations from the E-W direction at higher latitudes.

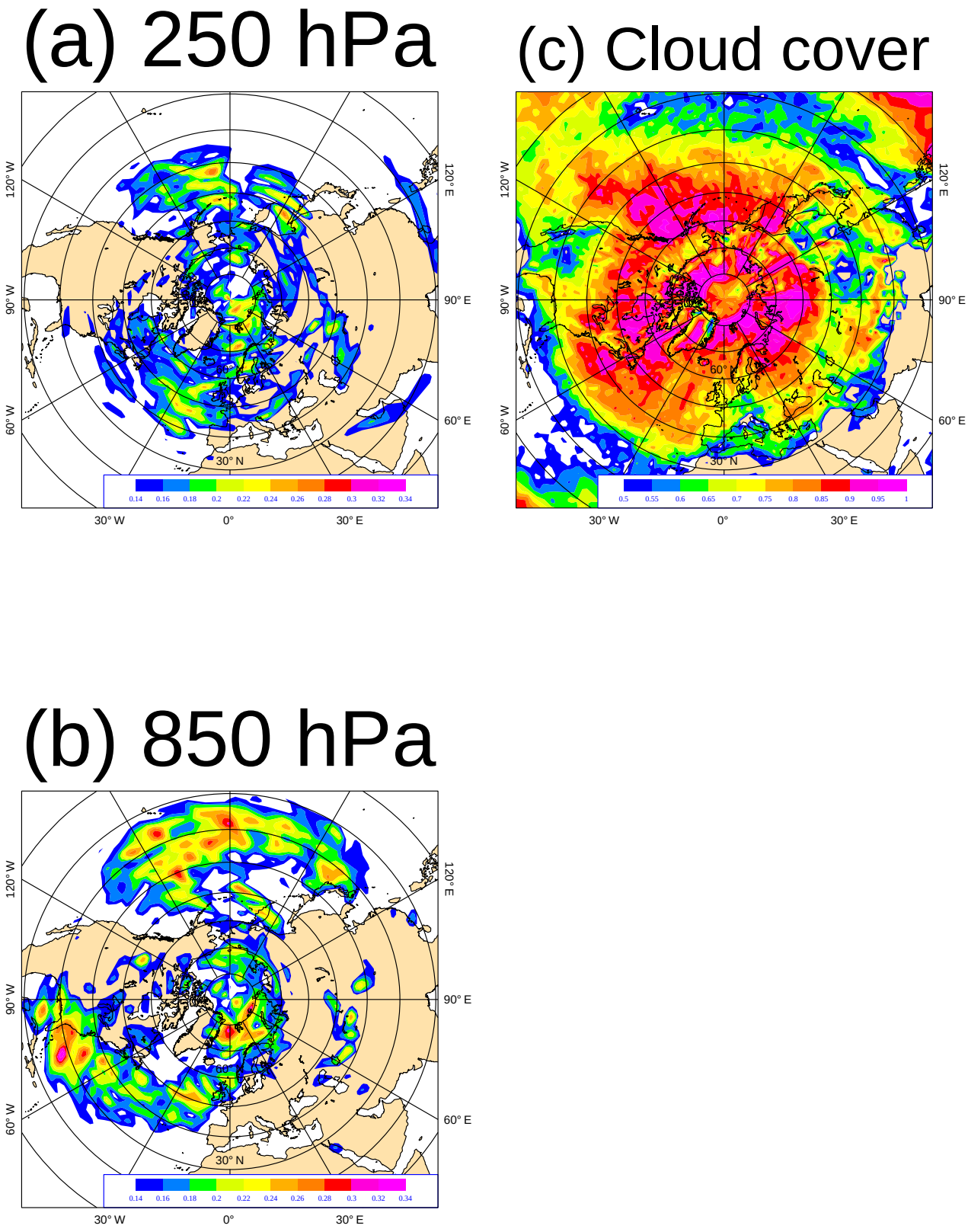


Figure 4: (a,b) R.m.s. of key analysis error for zonal wind component in ms^{-1} (shaded, see legend), for the period 16 Jan to 28 Feb 2003. (c) Total cloud cover for the same period.

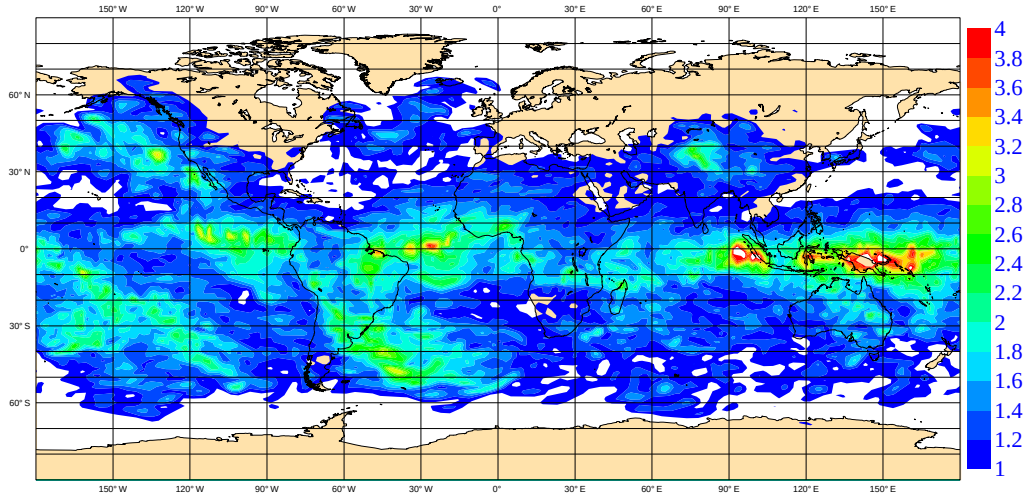


Figure 5: Accuracy of zonal wind (in ms^{-1}) at 250 hPa estimated with a data assimilation ensemble, Jan–Feb 2003. The spread between 12-hour forecasts is indicative of background error for the ECMWF data assimilation system (the version used Autumn 2003). The absolute magnitude is uncertain and requires calibration, e.g. through comparison with actual observation-minus-background departures.

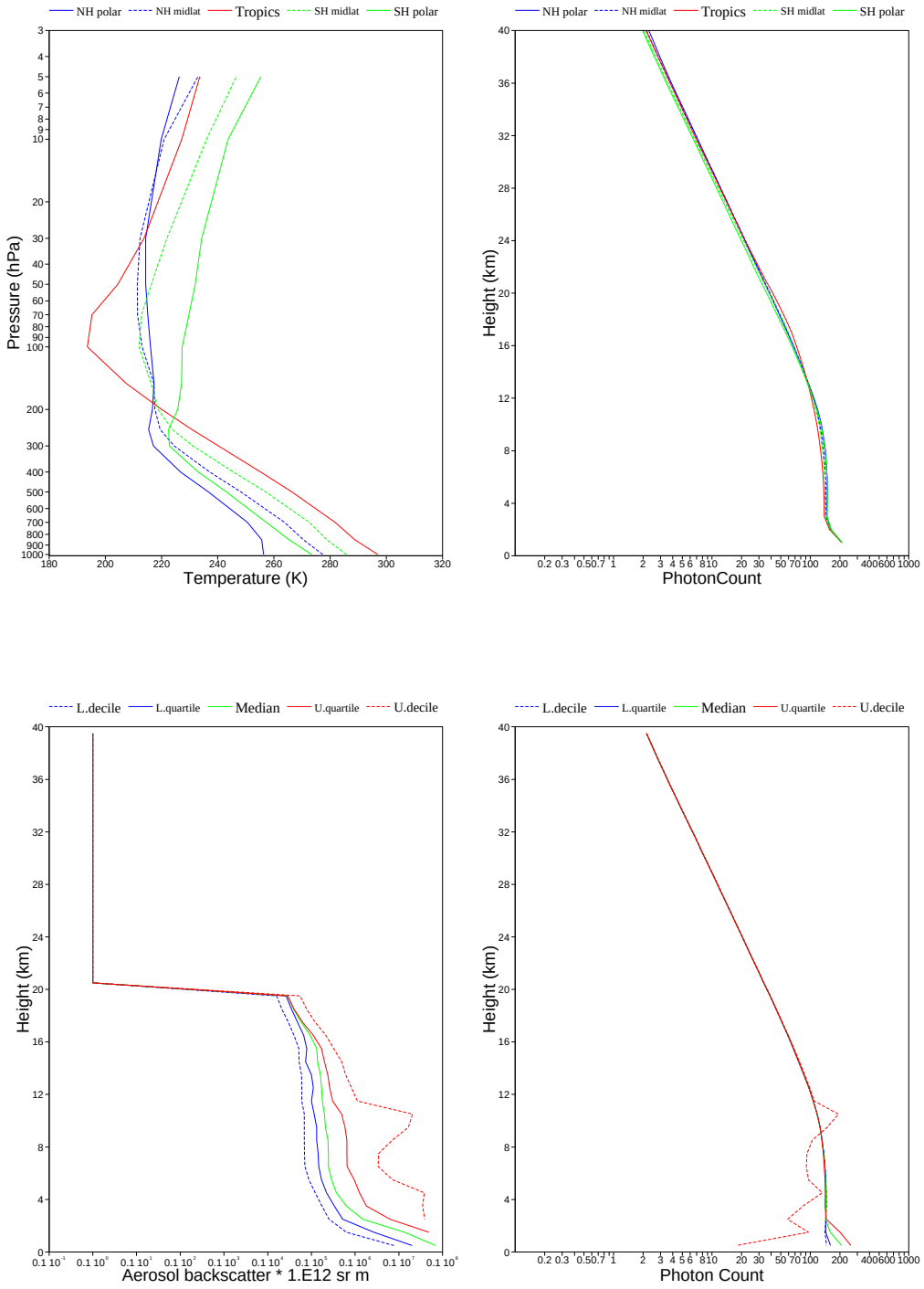


Figure 6: Input profiles for temperature (upper left) and aerosol backscatter (lower left) and corresponding photon counts (right panels) produced by LIPAS simulations of the ADM-Aeolus instrument.

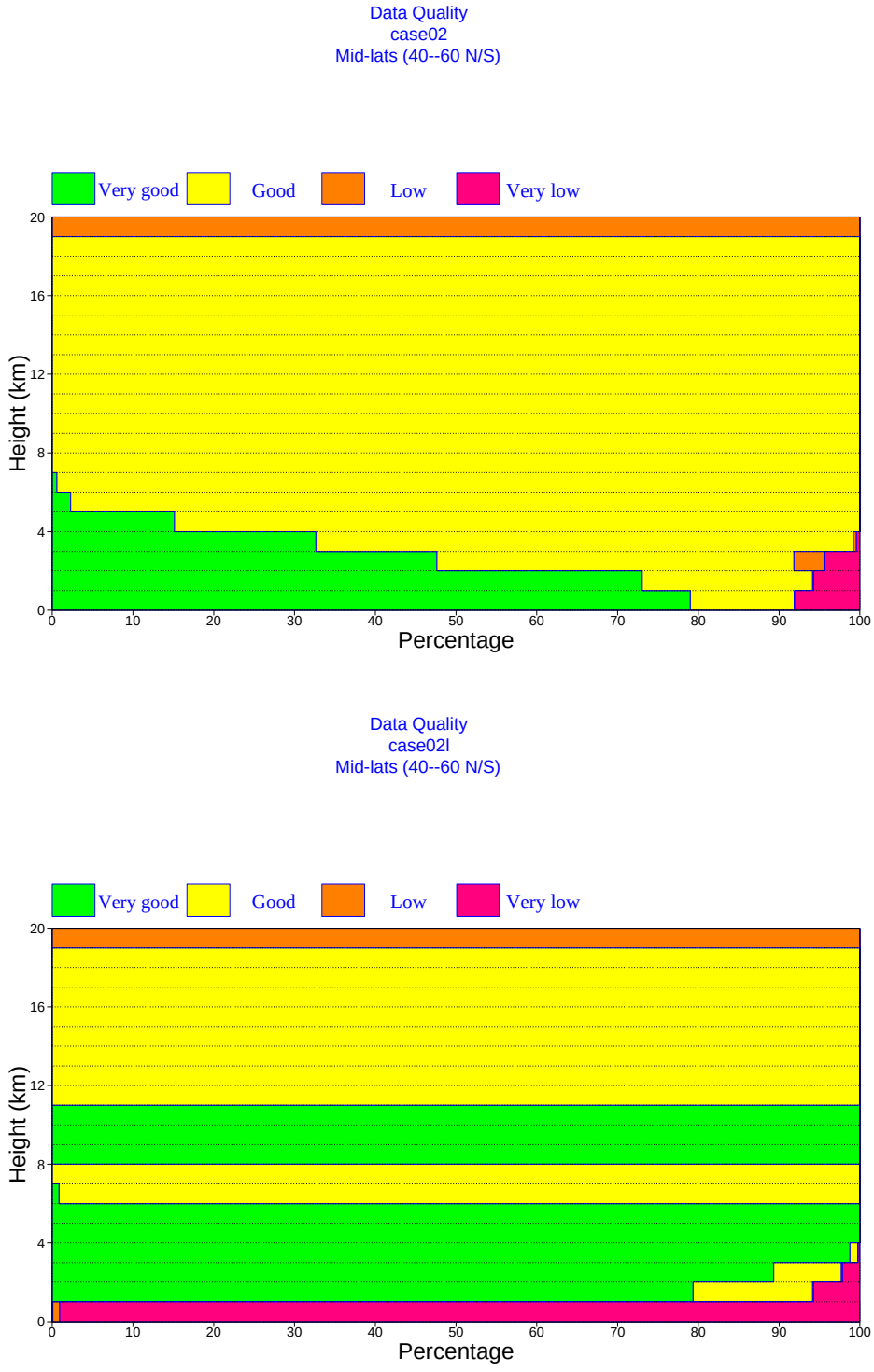


Figure 7: Statistics of DWL observation quality in 4 classes labeled very good, good, low and very low (see legend). 1400 observations have been simulated with median backscatter (upper panel) and upper decile backscatter (lower panel).

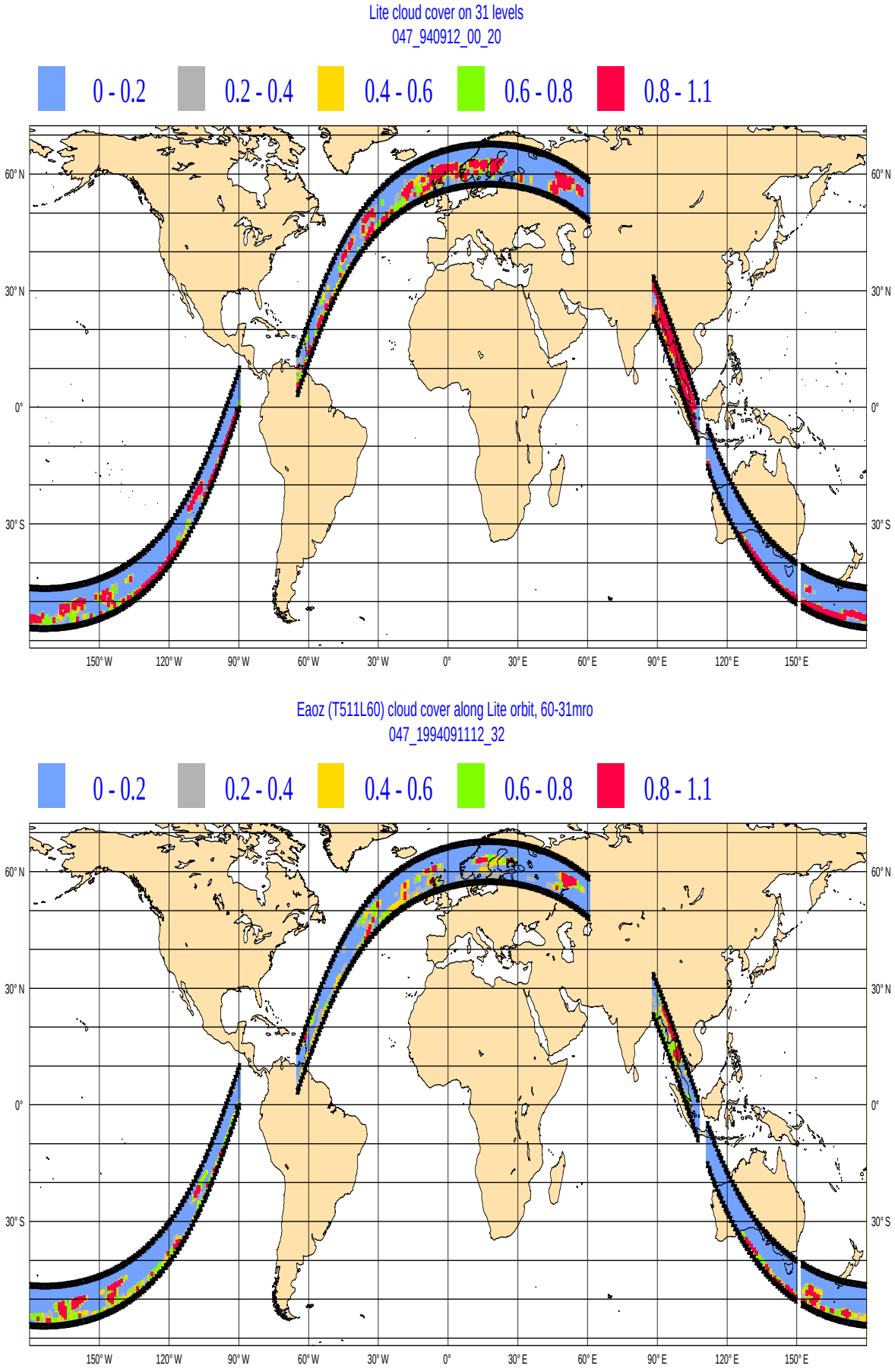


Figure 8: Profiles of cloud cover along LITE orbit 047 (12 Sep. 1994) from LITE retrievals (upper panel) and the ECMWF model (lower panel). The lower black curve is the satellite track, the upper black curve is approximately 10 hPa. The vertical scale in the boundary layer is exaggerated.

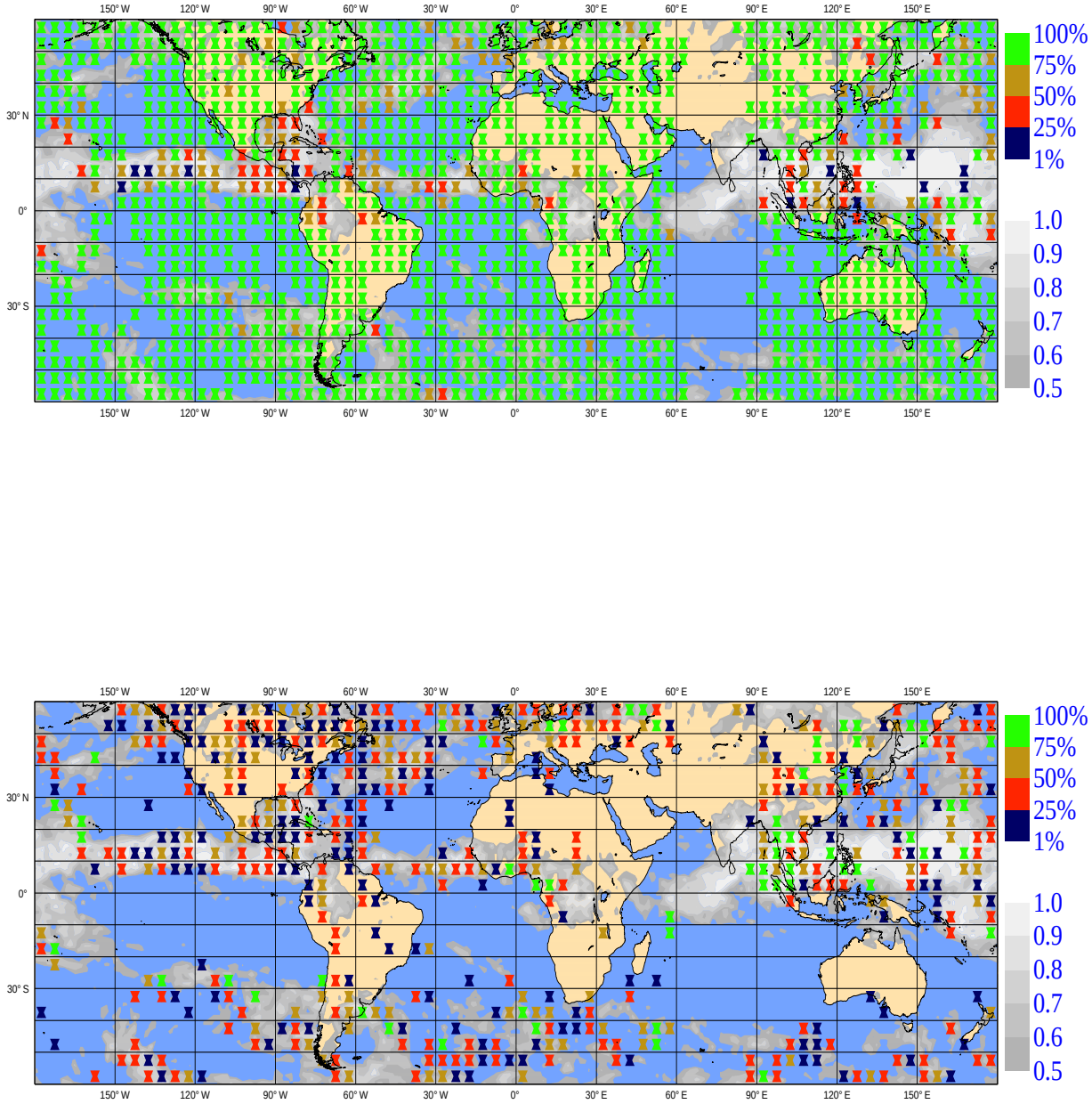


Figure 9: Simulated Aeolus yield at 9–10 km in terms of percentage of good data in 5x5 degree boxes represented by coloured markers (see legend), based on simulations of all available LITE orbits, for Rayleigh channel (upper) and Mie channel (lower). Shading shows rms of ECMWF high cloud cover within the LITE period.

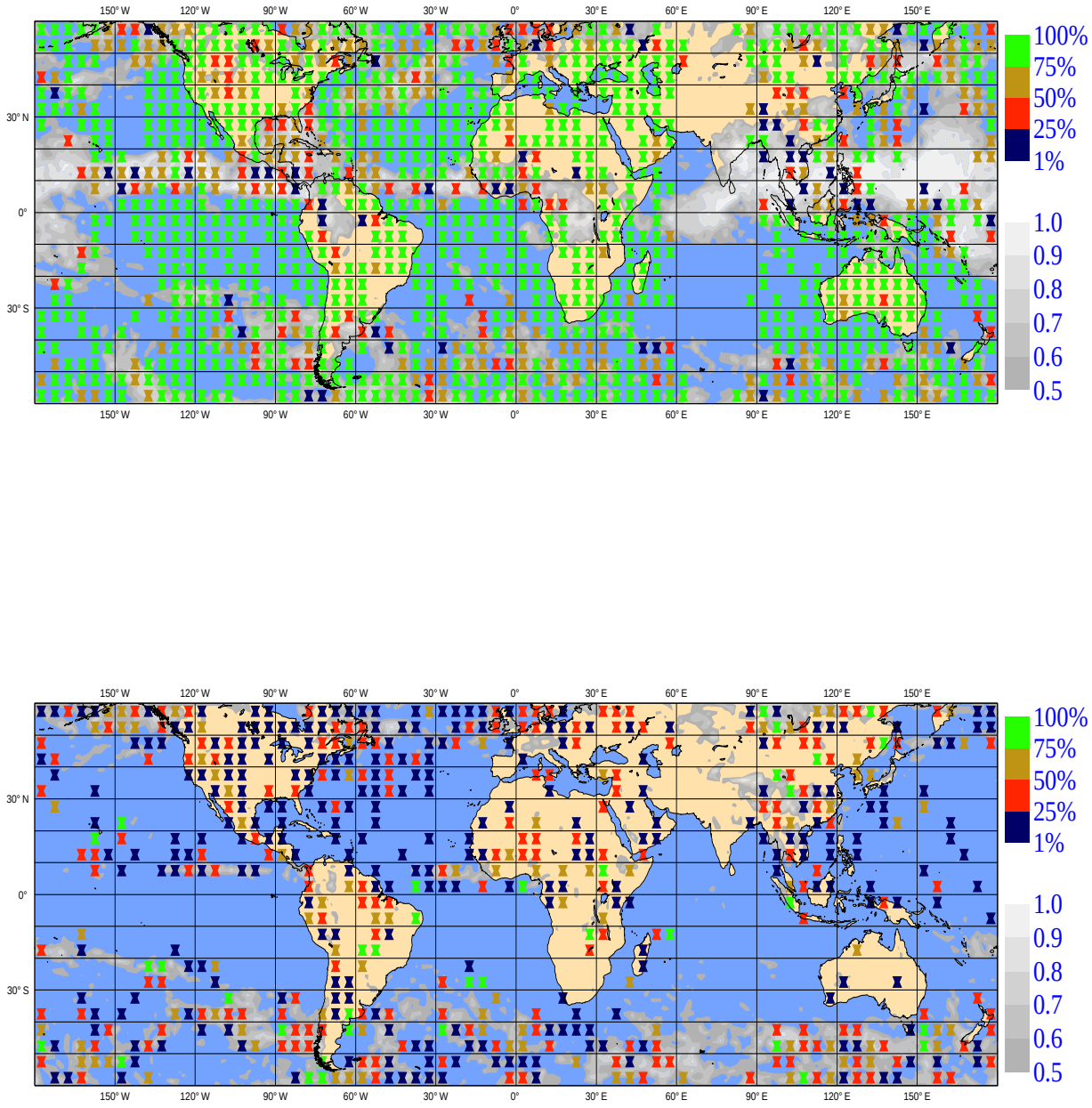


Figure 10: Simulated Aeolus yield at 4–5 km. Upper: Rayleigh channel/high cloud cover, lower: Mie channel/medium cloud cover. Otherwise like Fig. 9

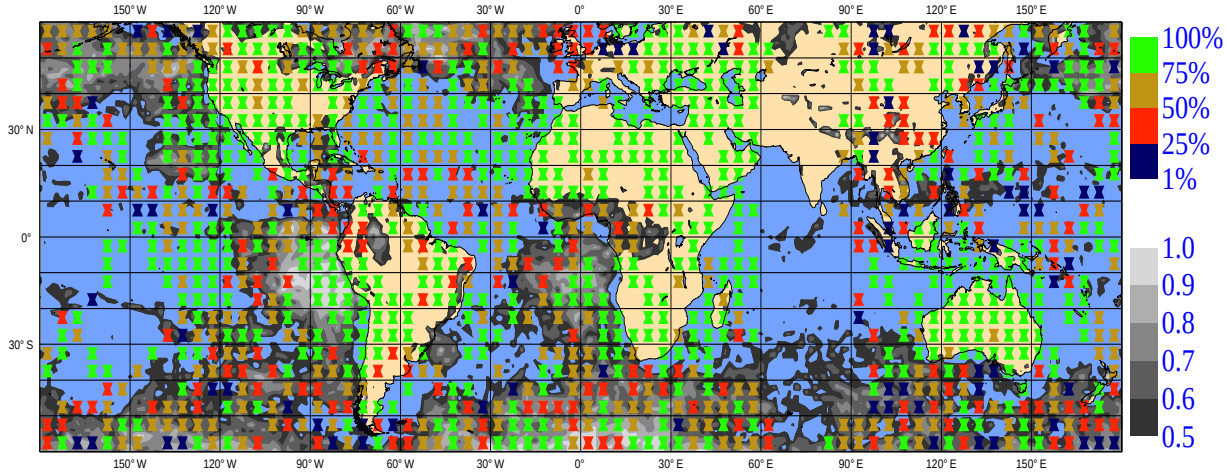


Figure 11: Simulated Aeolus yield at 0.5–1 km for Mie channel. Shading shows rms of ECMWF low cloud cover. Otherwise like Fig. 9

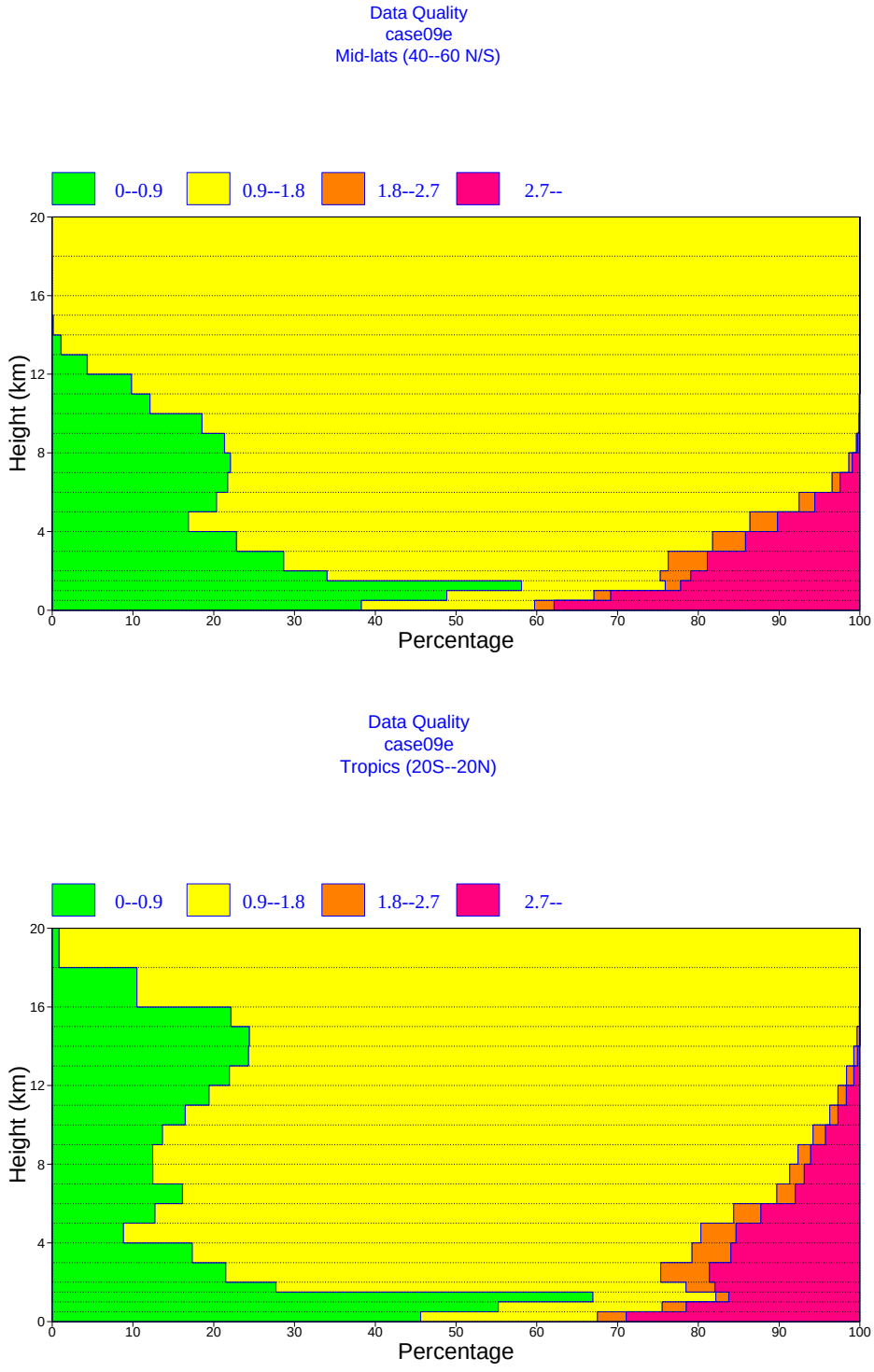


Figure 12: Statistics of DWL observation quality using observed cloud cover for the LITE period in mid-latitudes (upper panel) and the tropics (lower panel).

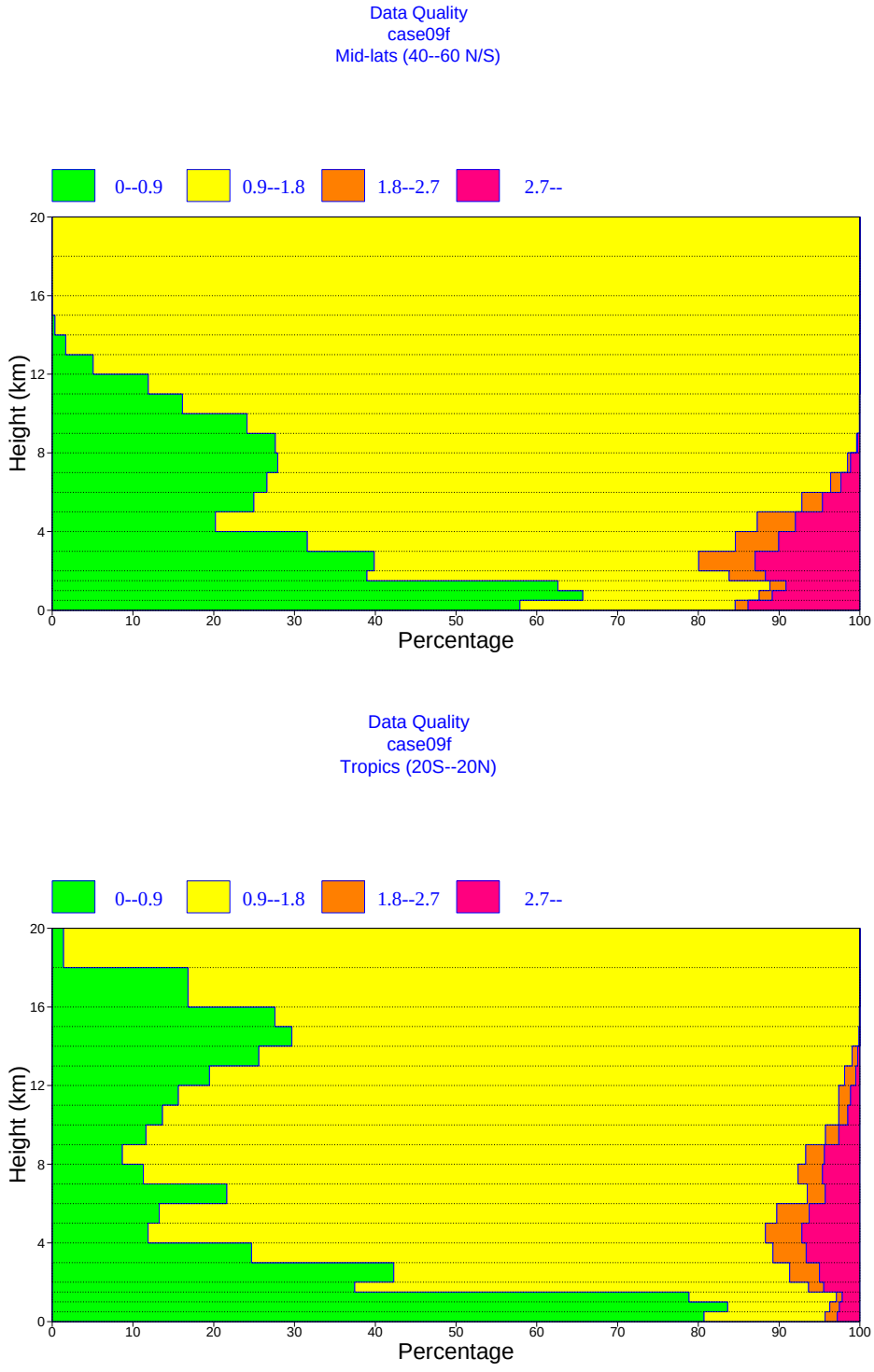
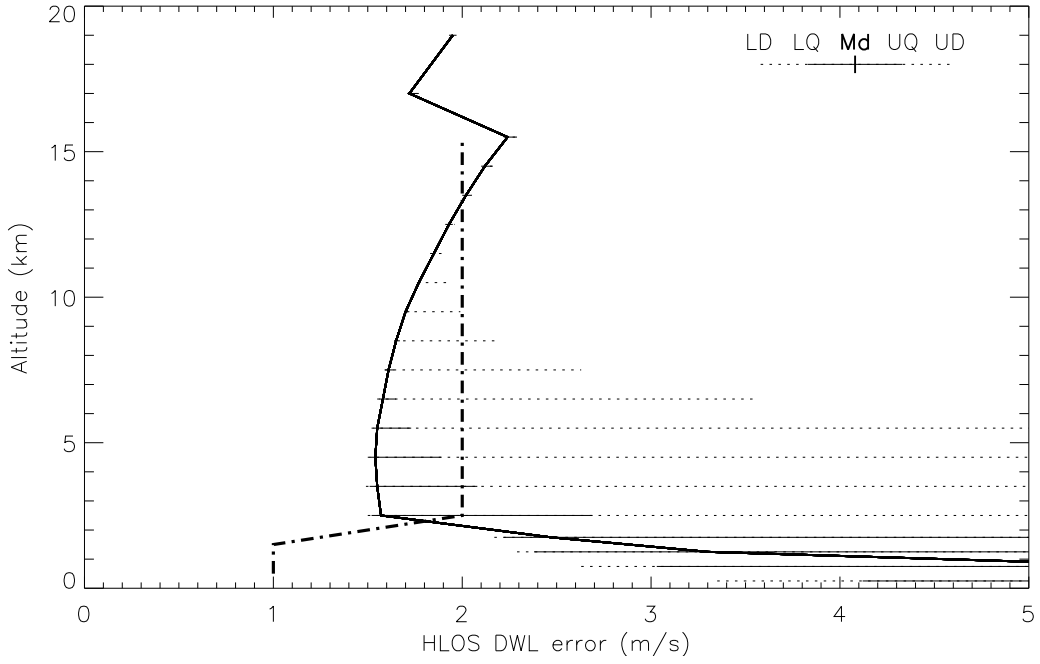


Figure 13: Statistics of DWL observation quality (defined in text) using ECMWF model cloud cover for the LITE period. 5633 observations have been simulated in the mid-latitudes (upper panel) and 3177 in the Tropics (lower panel).

(a)



(b)

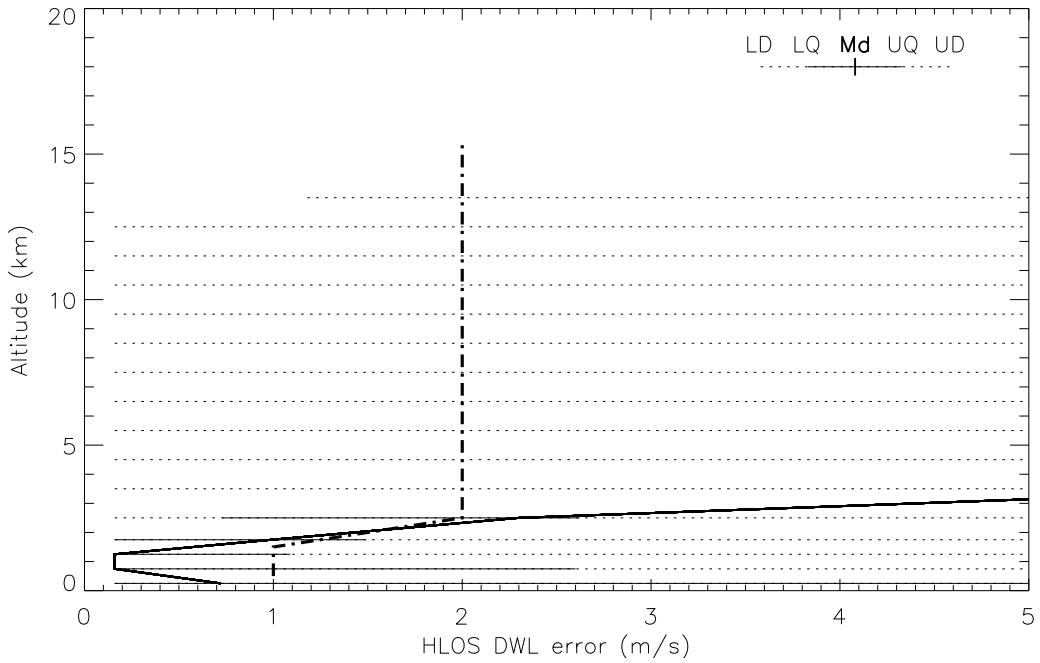
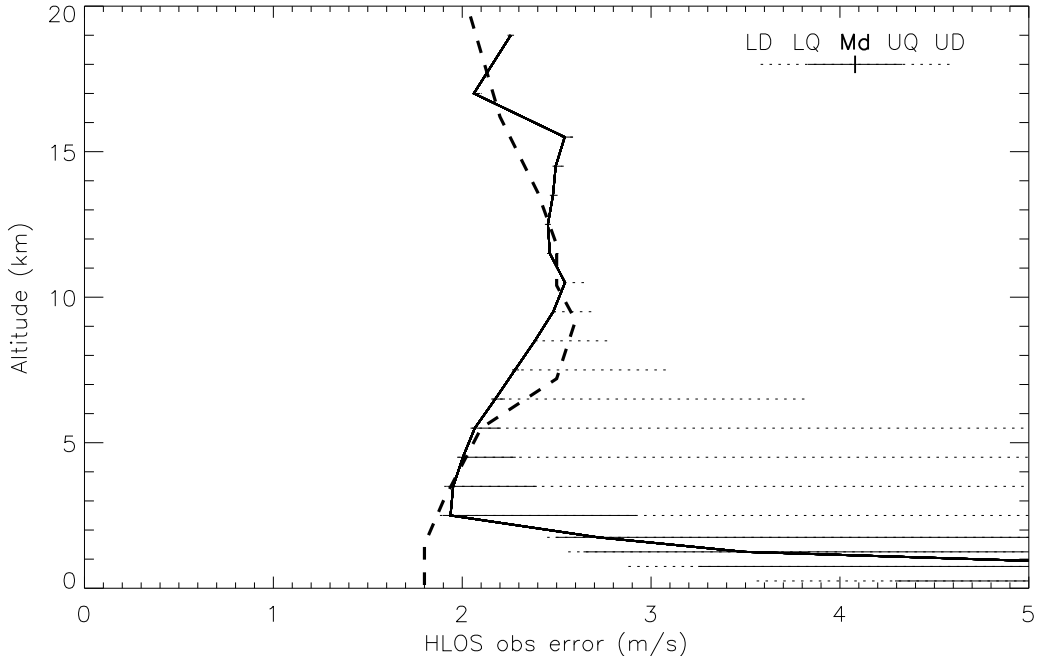


Figure 14: Distribution of simulated Aeolus instrument random error (in ms^{-1}) using LITE-inferred cloud cover. (a) Rayleigh channel $\hat{\sigma}_R$, (b) Mie channel $\hat{\sigma}_M$. Dash-dot shows the mission-specified requirement.

(a)



(b)

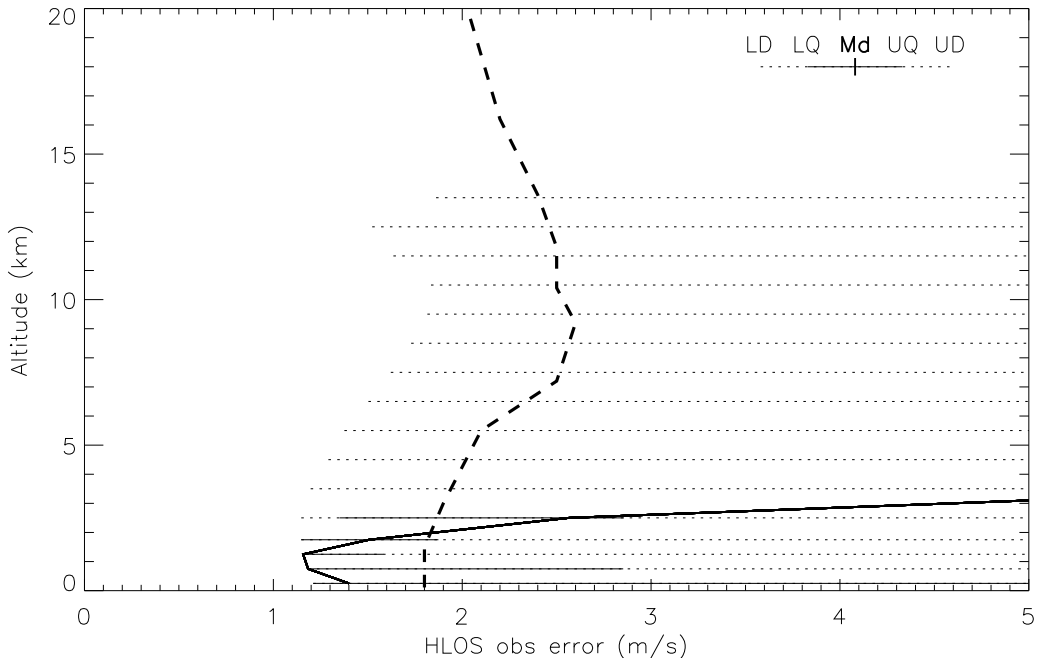
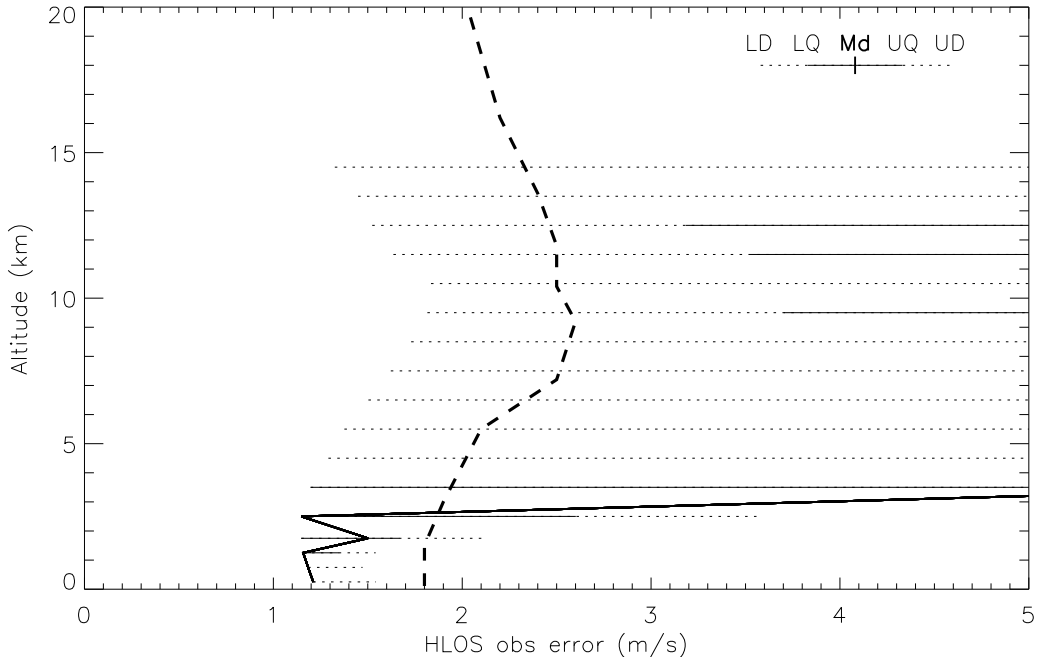


Figure 15: Distribution of simulated Aeolus observation random error (in ms^{-1}) using LITE-inferred cloud cover. (a) Rayleigh channel σ_R, (b) Mie channel σ_M. Dashed shows operationally assigned radiosonde errors.

(a)



(b)

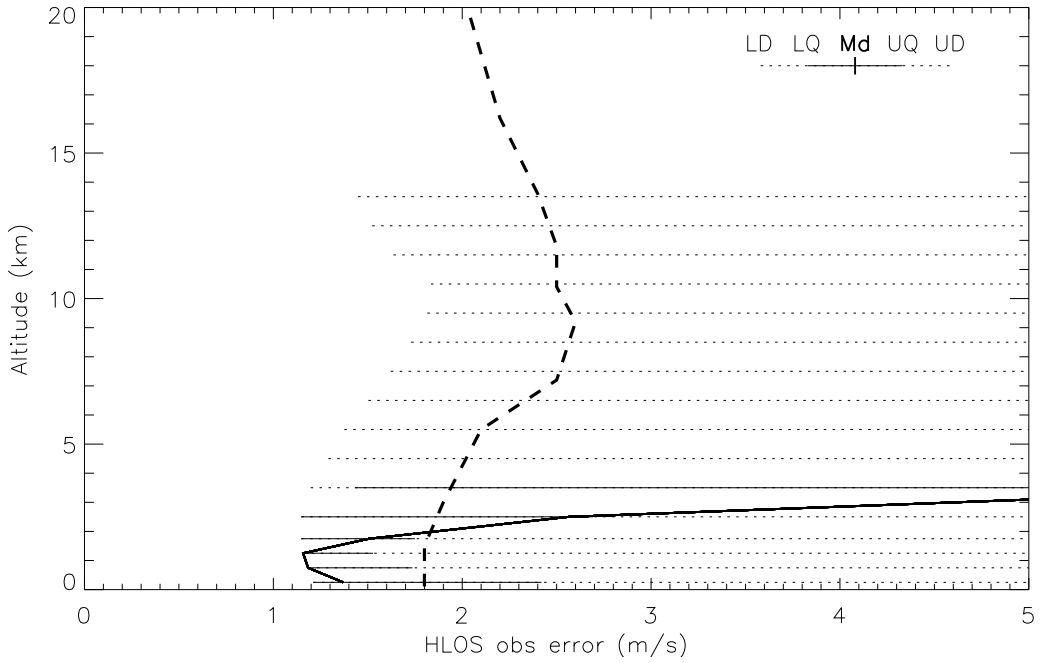


Figure 16: Distribution of simulated Aeolus observation random error, Mie channel (σ_M, ms^{-1}), North Atlantic. (a) using model cloud cover, (b) using LITE-inferred cloud cover. Dashed shows operationally assigned radiosonde errors.

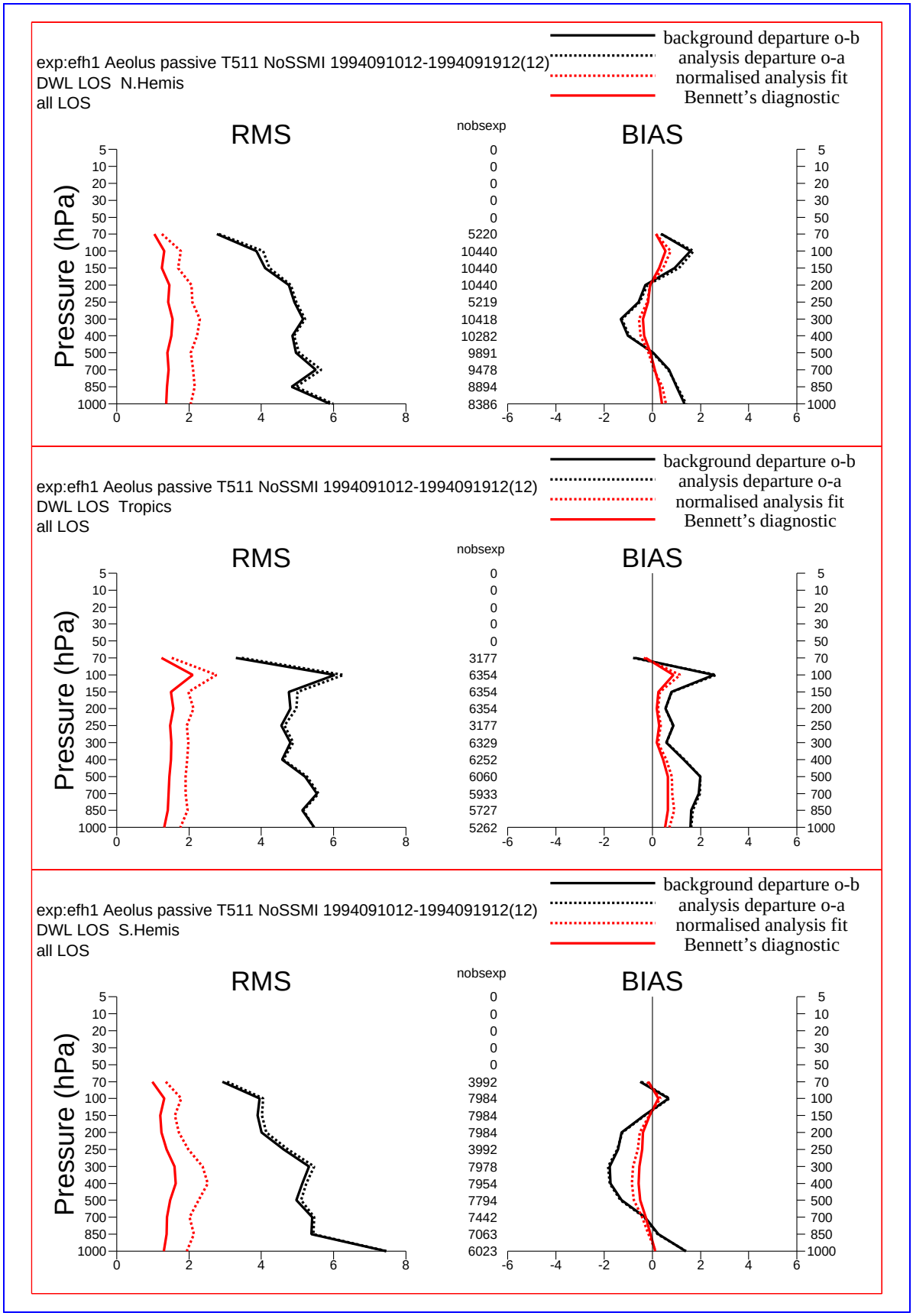


Figure 17: Observation statistics for passive monitoring of all Aeolus observations simulated for the LITE period. The assimilating model used has operational resolution (T511, ∼40 km). Line styles are explained in the legends.

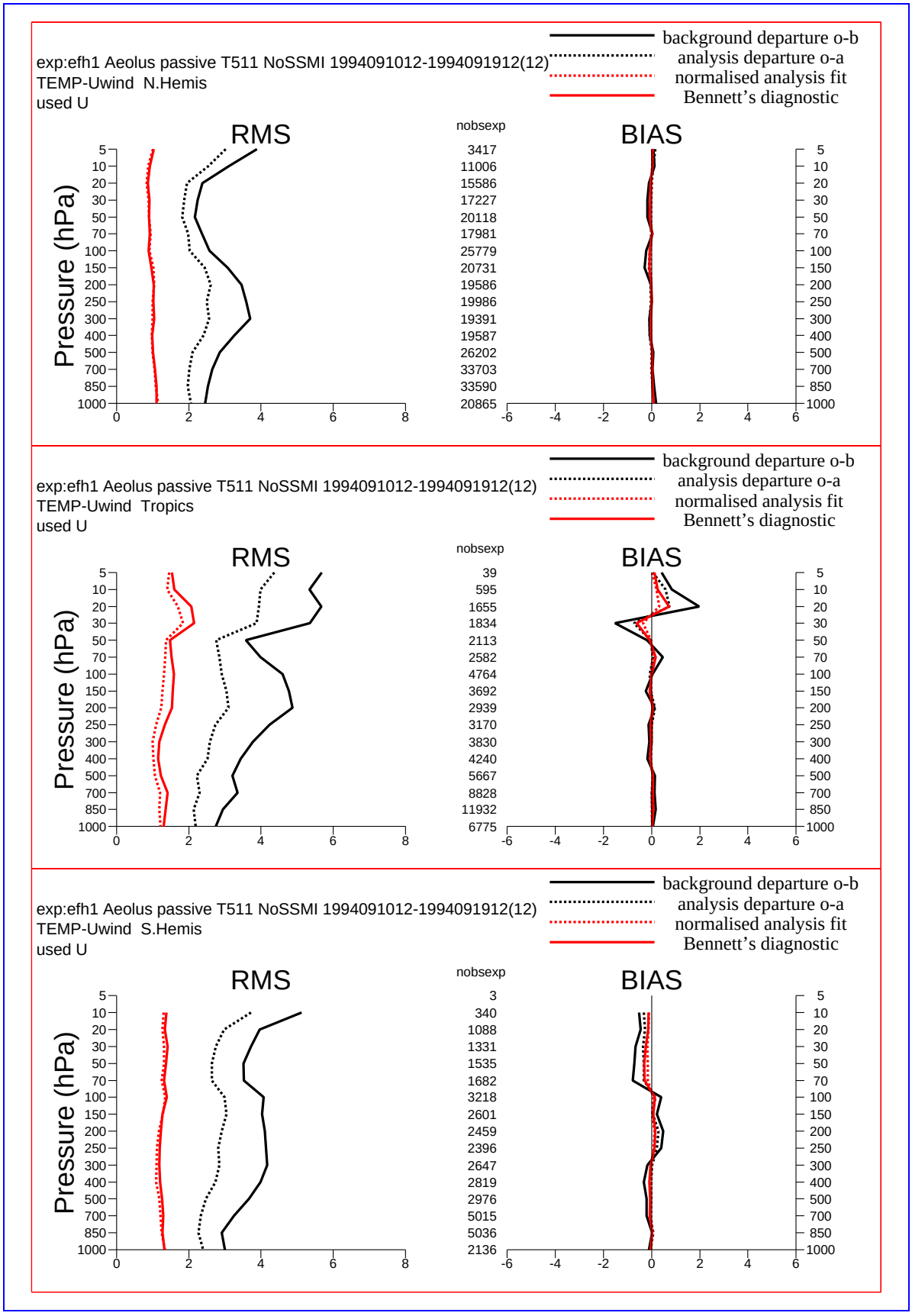


Figure 18: As for Figure 17 but showing active radiosonde zonal wind observations.

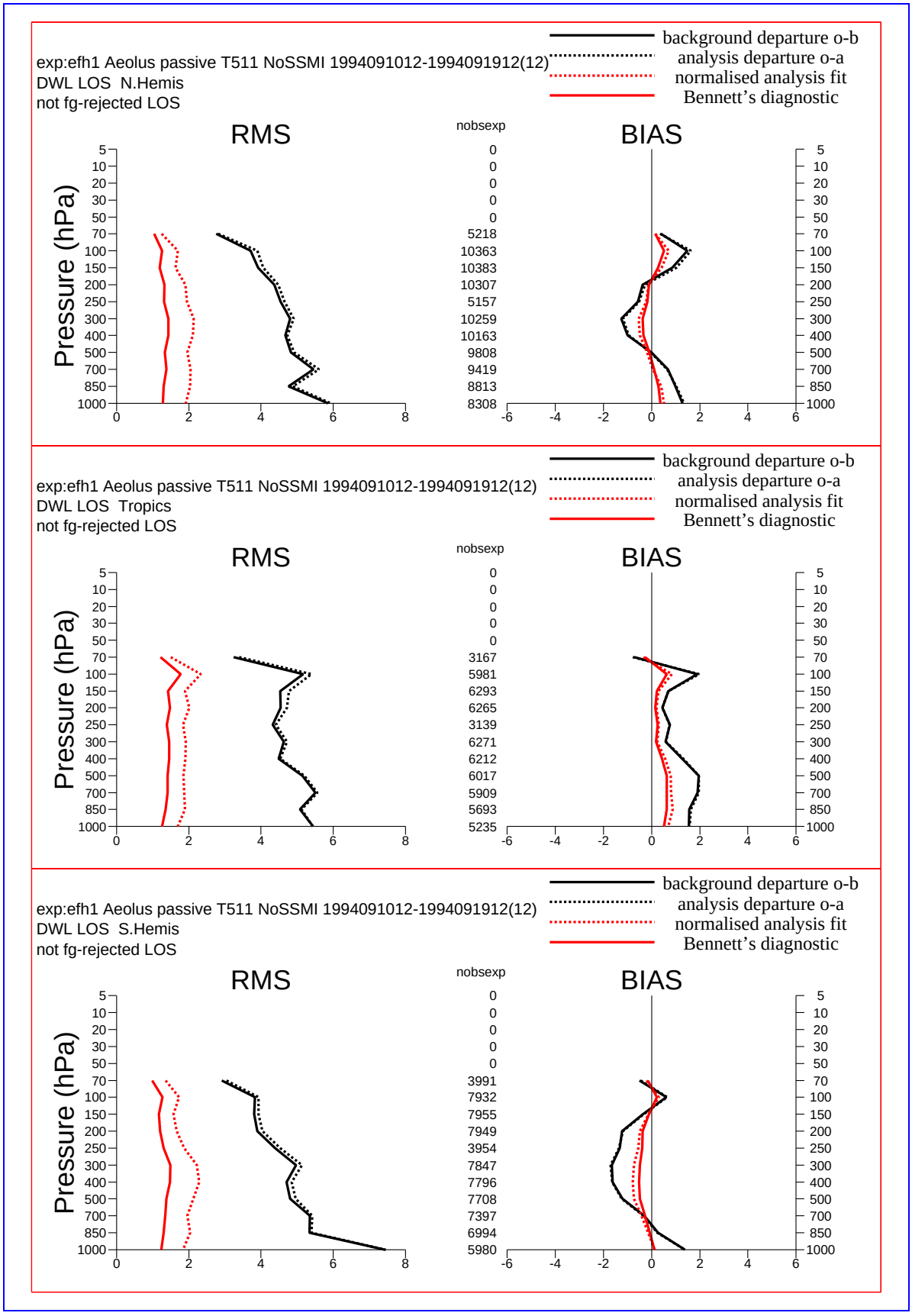


Figure 19: As for Figure 17 but only showing Aeolus data that pass standard background quality control.

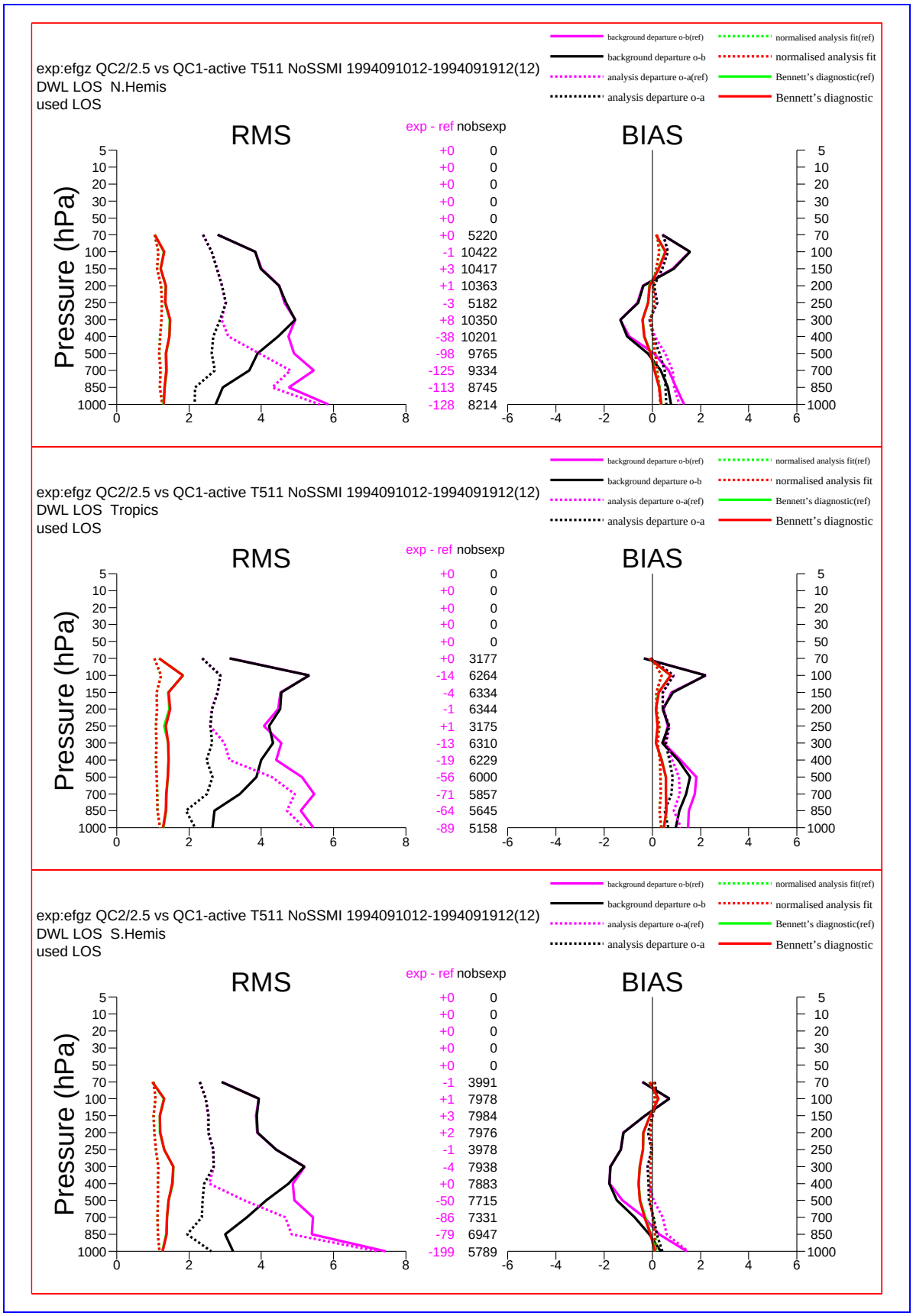


Figure 20: As for Figure 19 but showing active Aeolus data with standard QC (purple and green curves) and Aeolus-specific QC (black and red curves). The green curves are largely indistinguishable from their red counterparts. Similarly, purple is indistinguishable from black at higher altitudes.

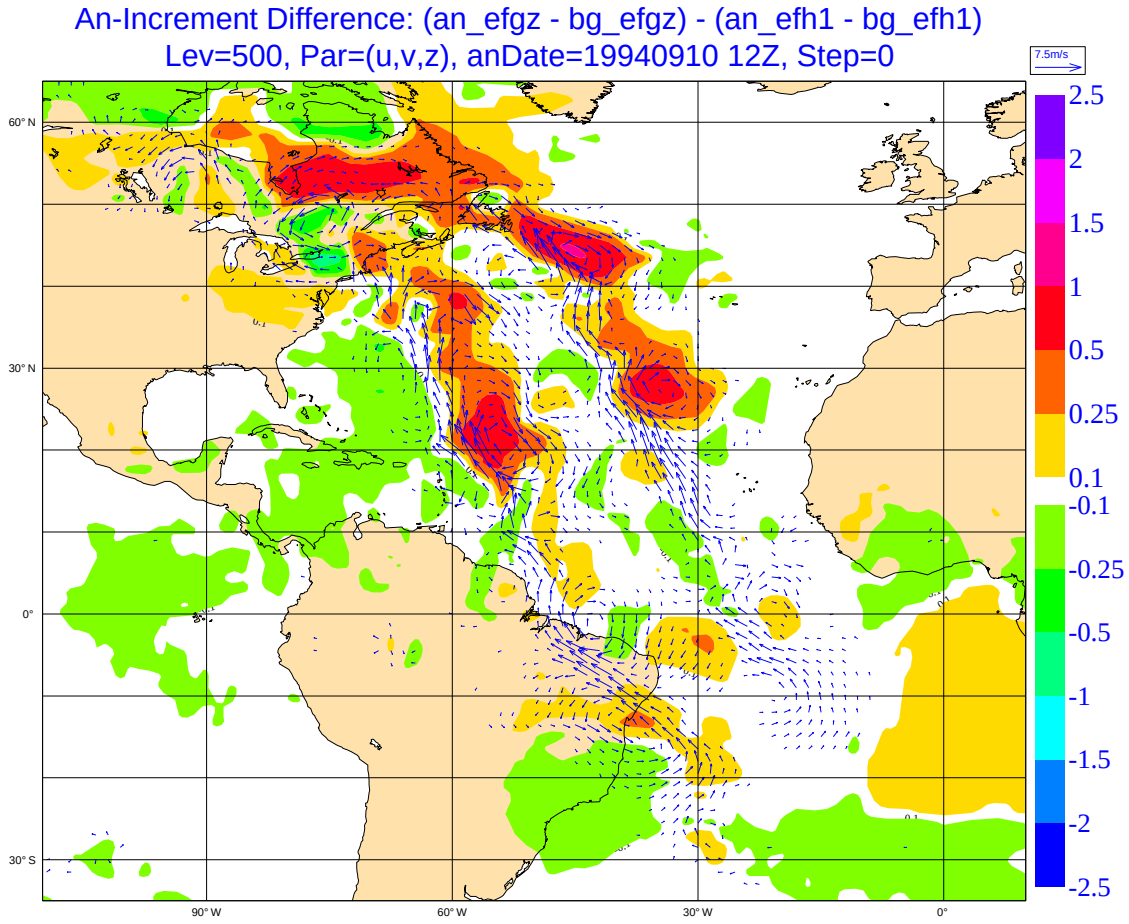


Figure 21: Analysis increments at 500 hPa from simulated Aeolus data. Shading shows geopotential $\times g^{-1}$ (units of dm).

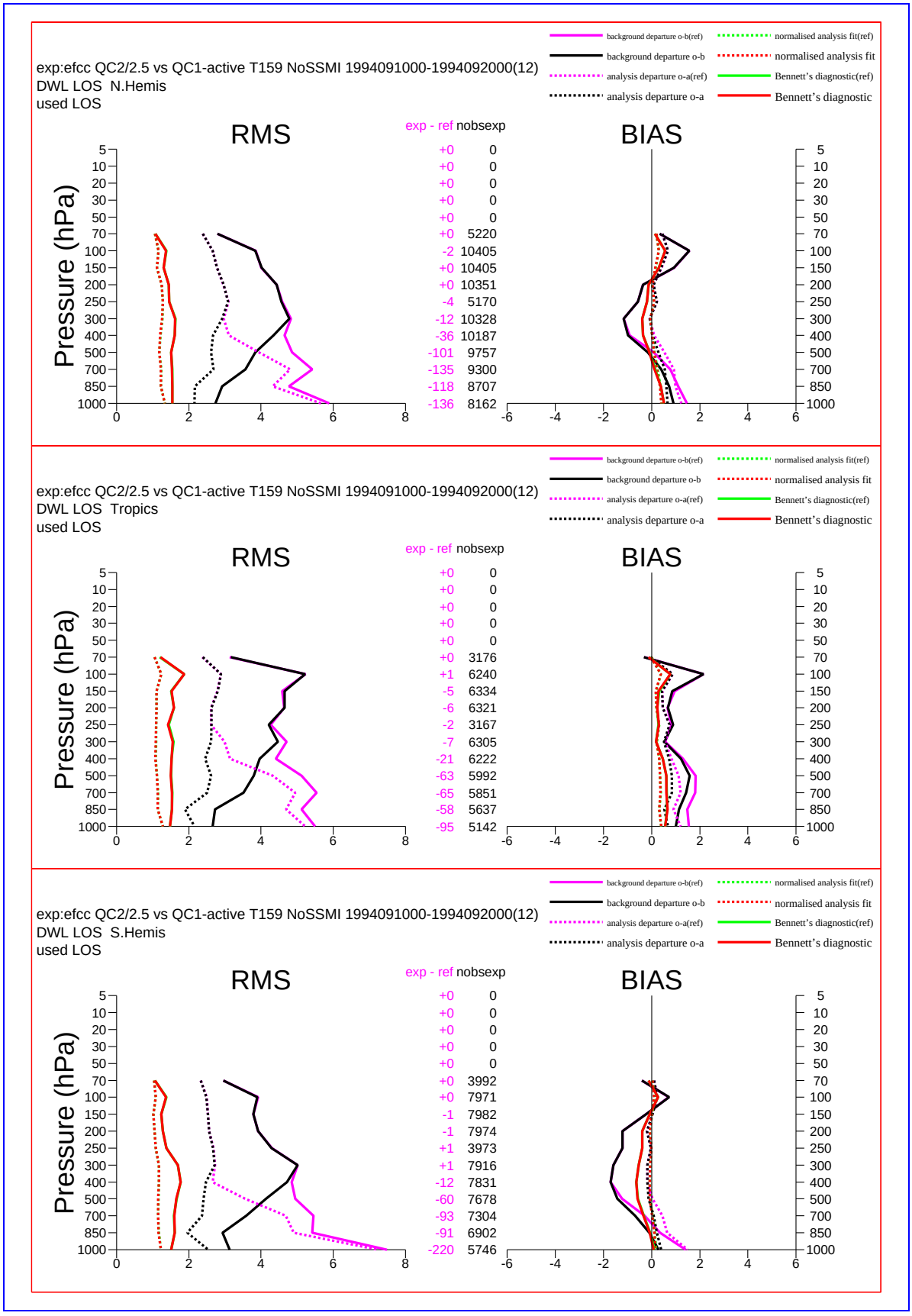


Figure 22: As for Figure 20 but at moderate resolution (T159, ∼125 km).

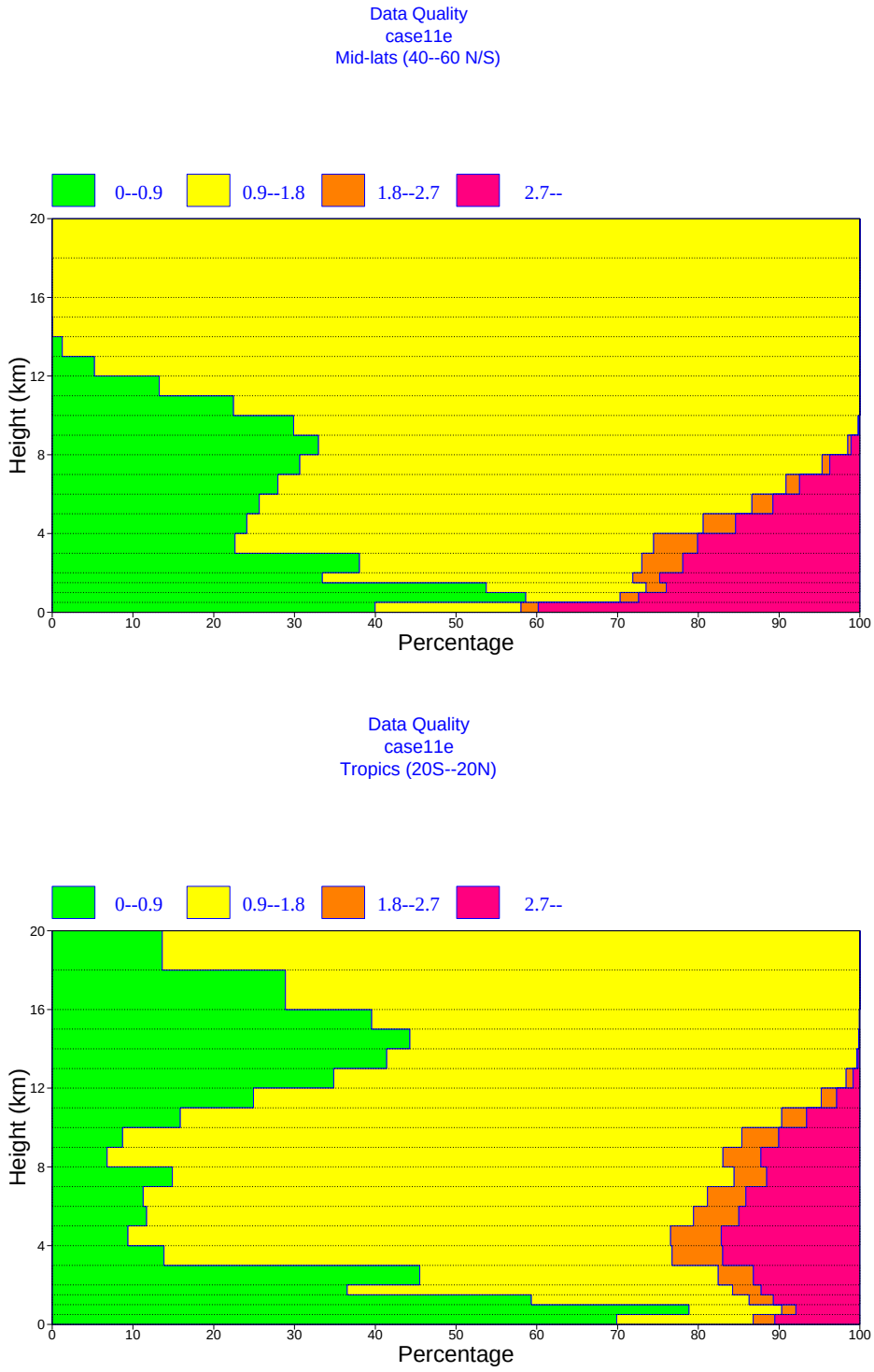


Figure 23: Like Figure 12 but using ECMWF cloud cover for the assimilation ensemble period Jan-Feb 2003.

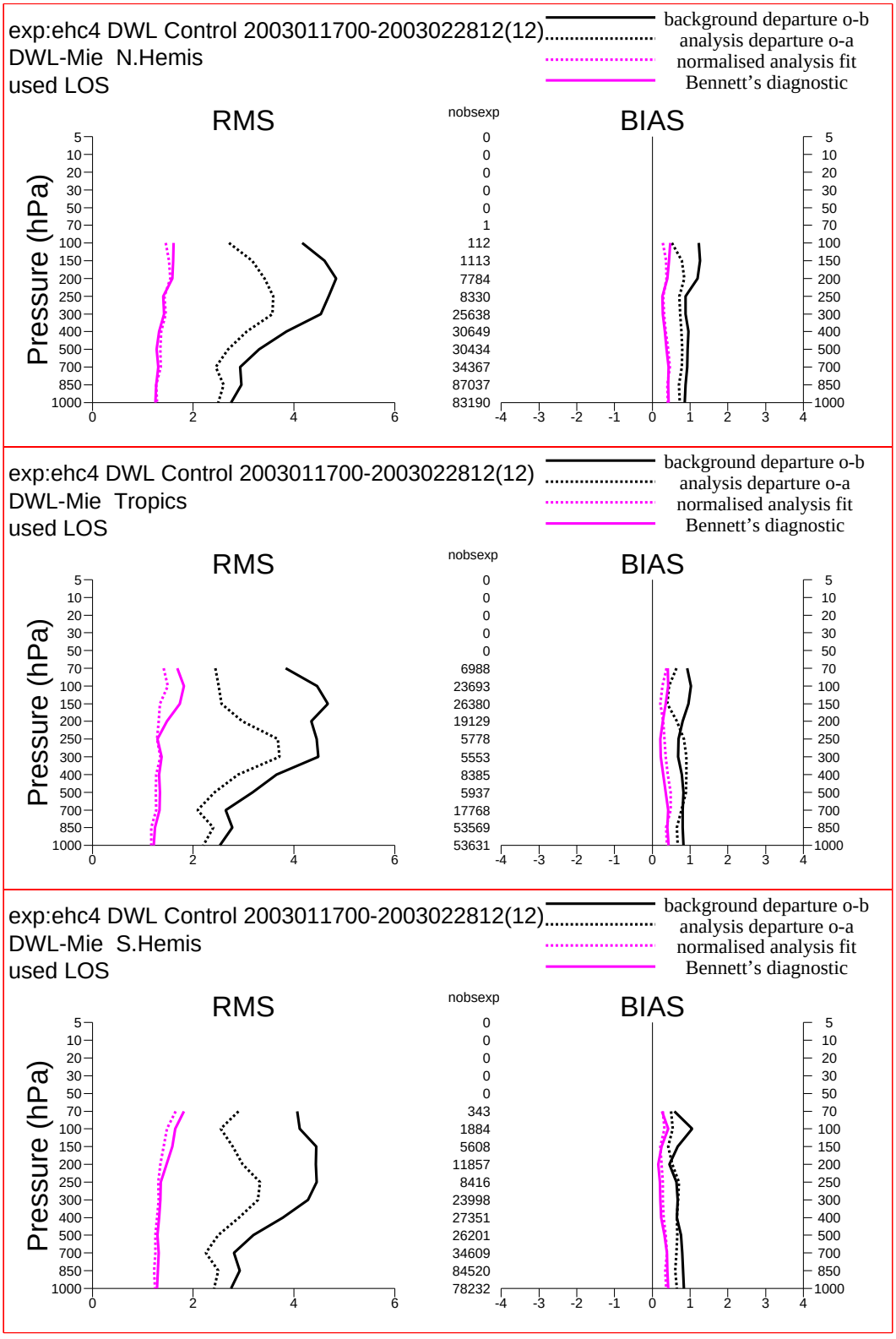


Figure 24: Mie channel observations statistics (simulated Aeolus data for the assimilation ensemble period).

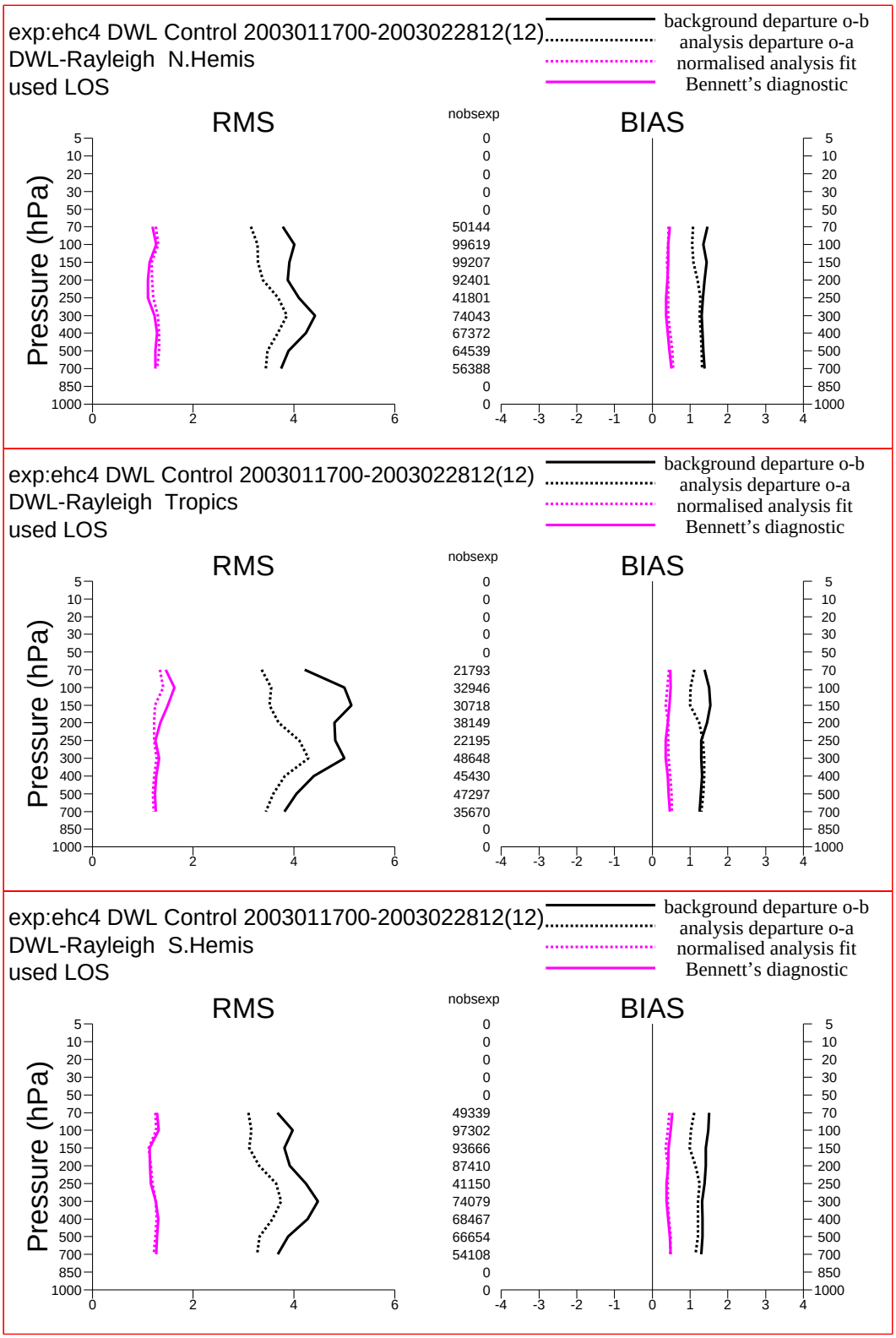


Figure 25: As for Figure 24 but for the Rayleigh channel.

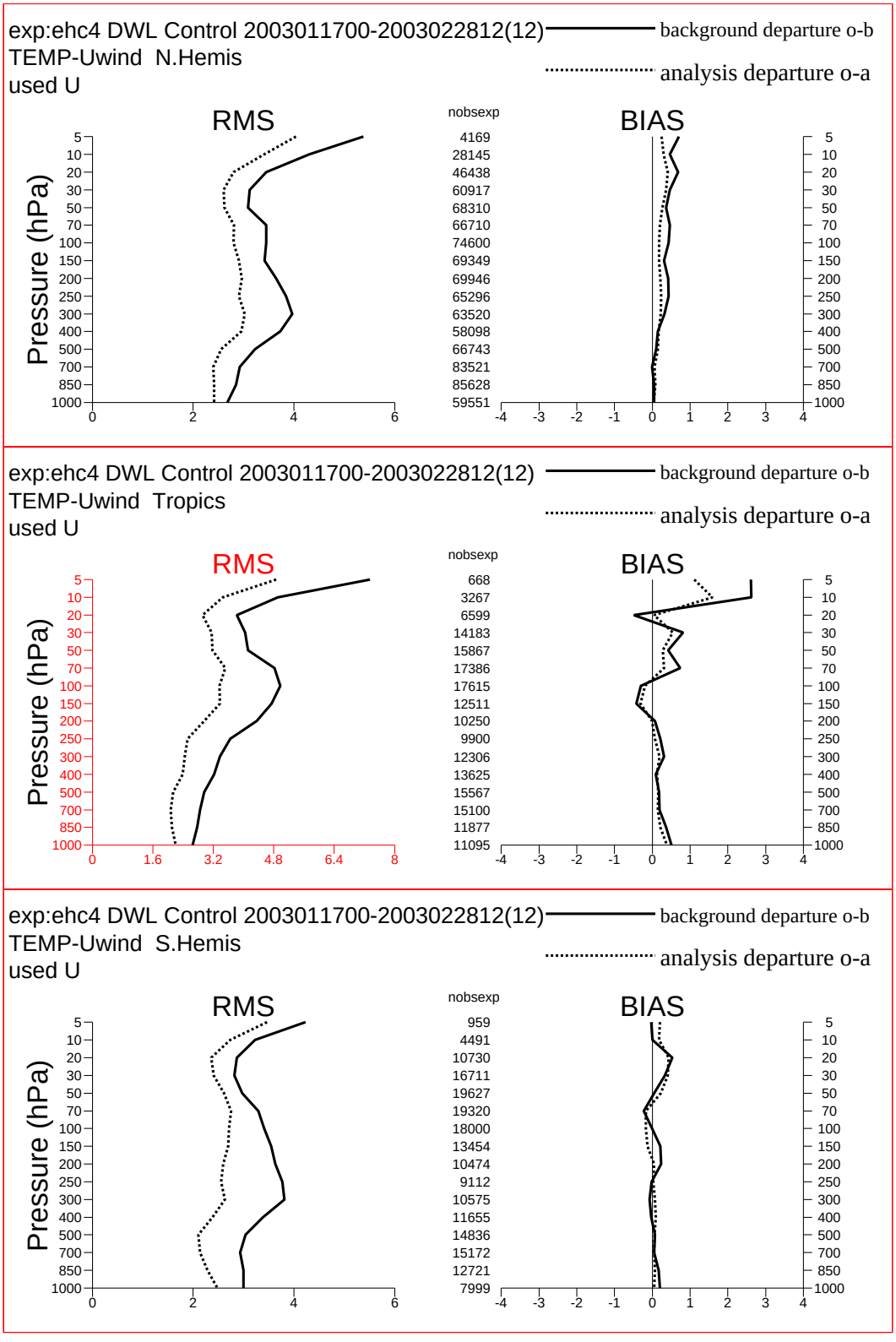


Figure 26: As for Figure 25 but for zonal wind observations from TEMPs.

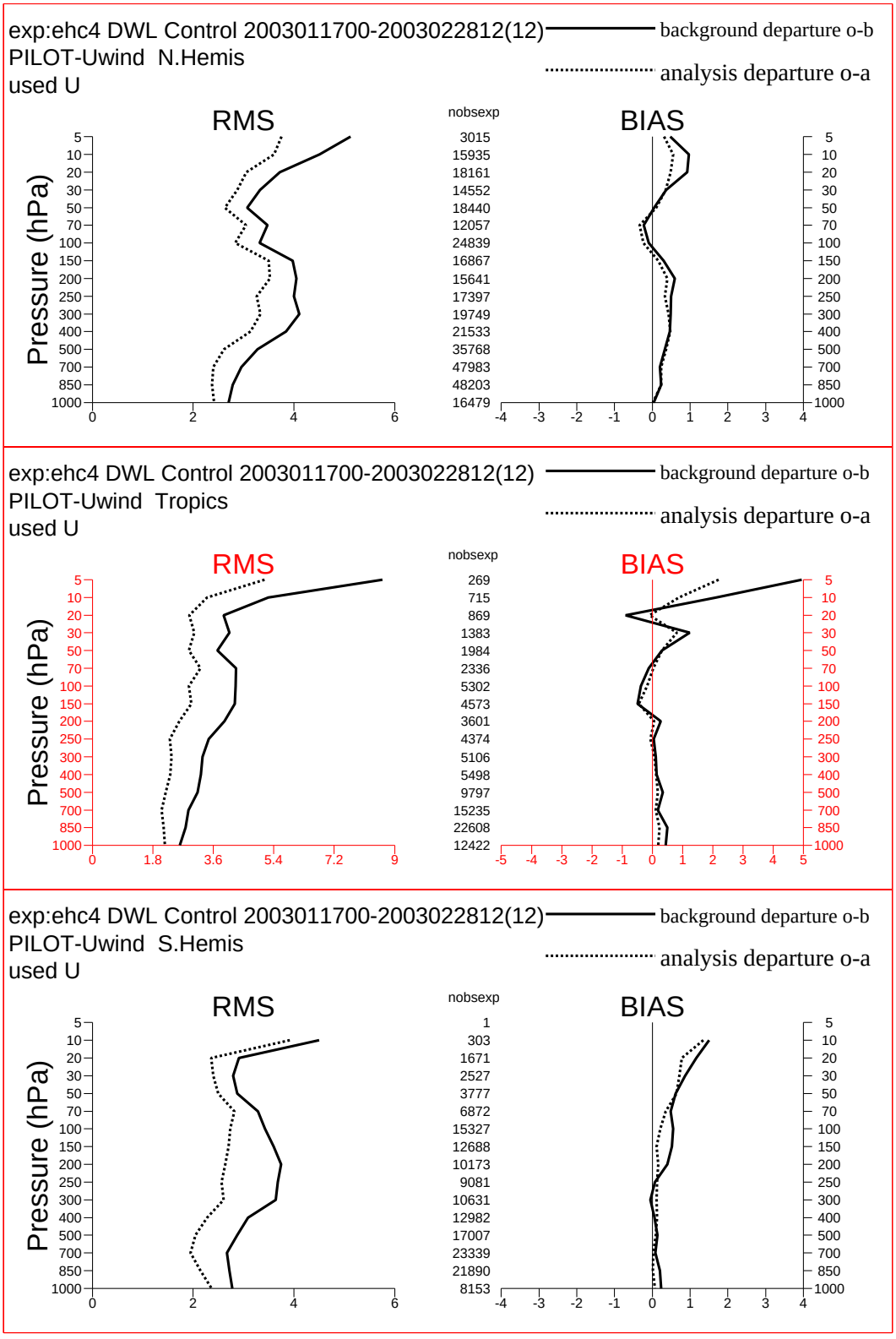


Figure 27: As for Figure 26 but from PILOTs.

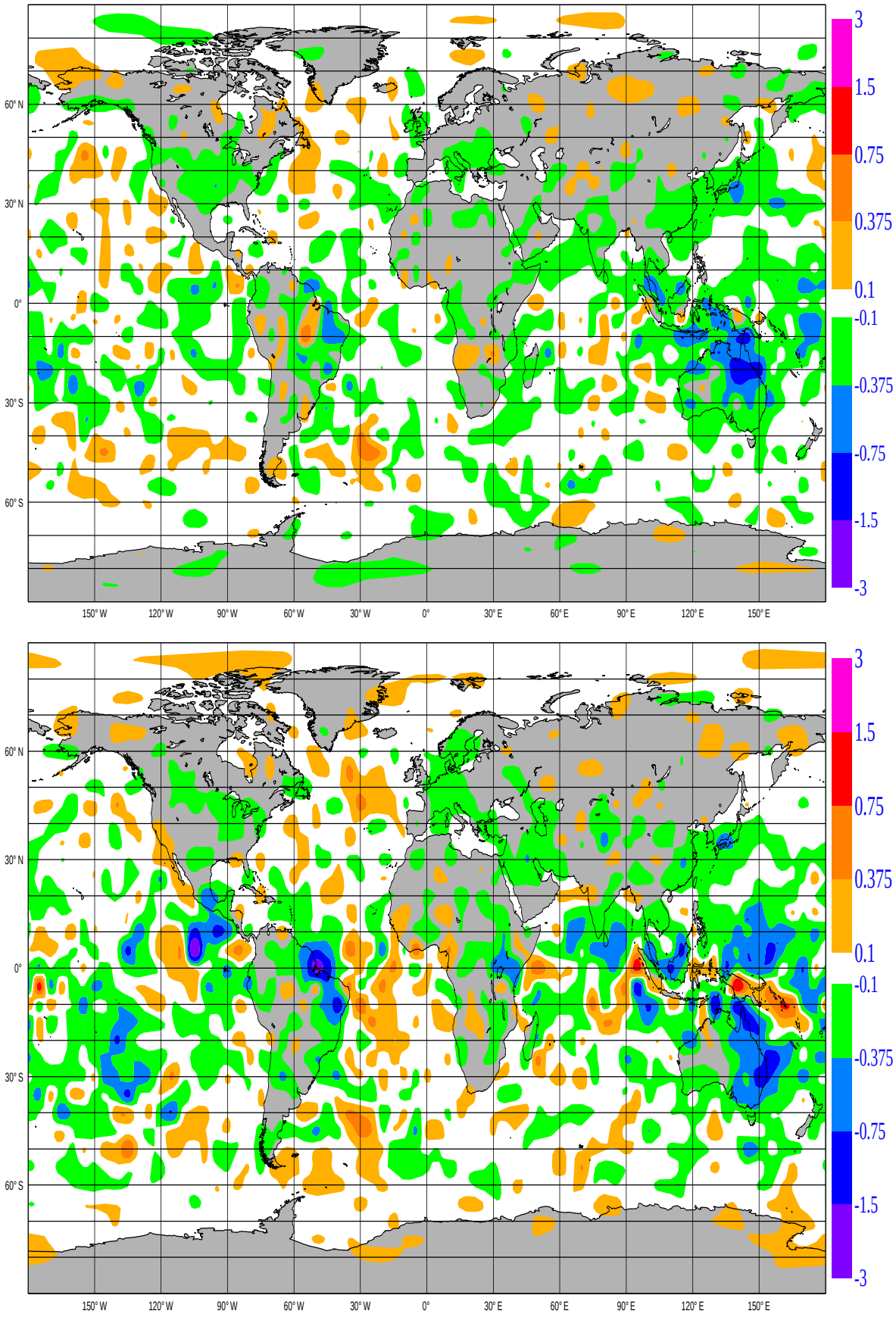


Figure 28: Radiosonde impact on assimilation ensemble spread at 500 hPa and 200 hPa (upper and lower panels). Benefits are indicated by negative values – a reduction in spread.

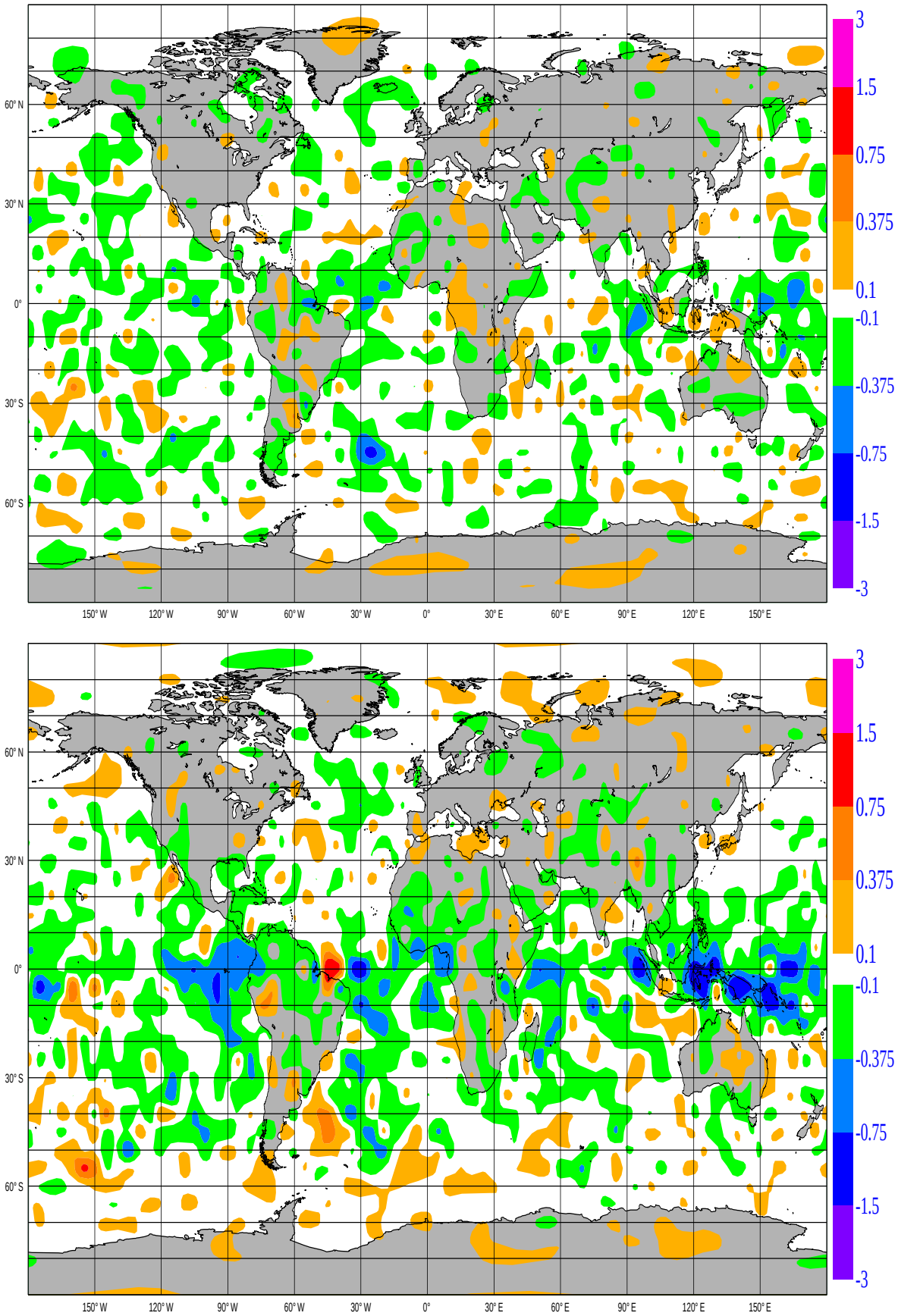


Figure 29: As for Figure 28 but for addition of simulated Aeolus data.

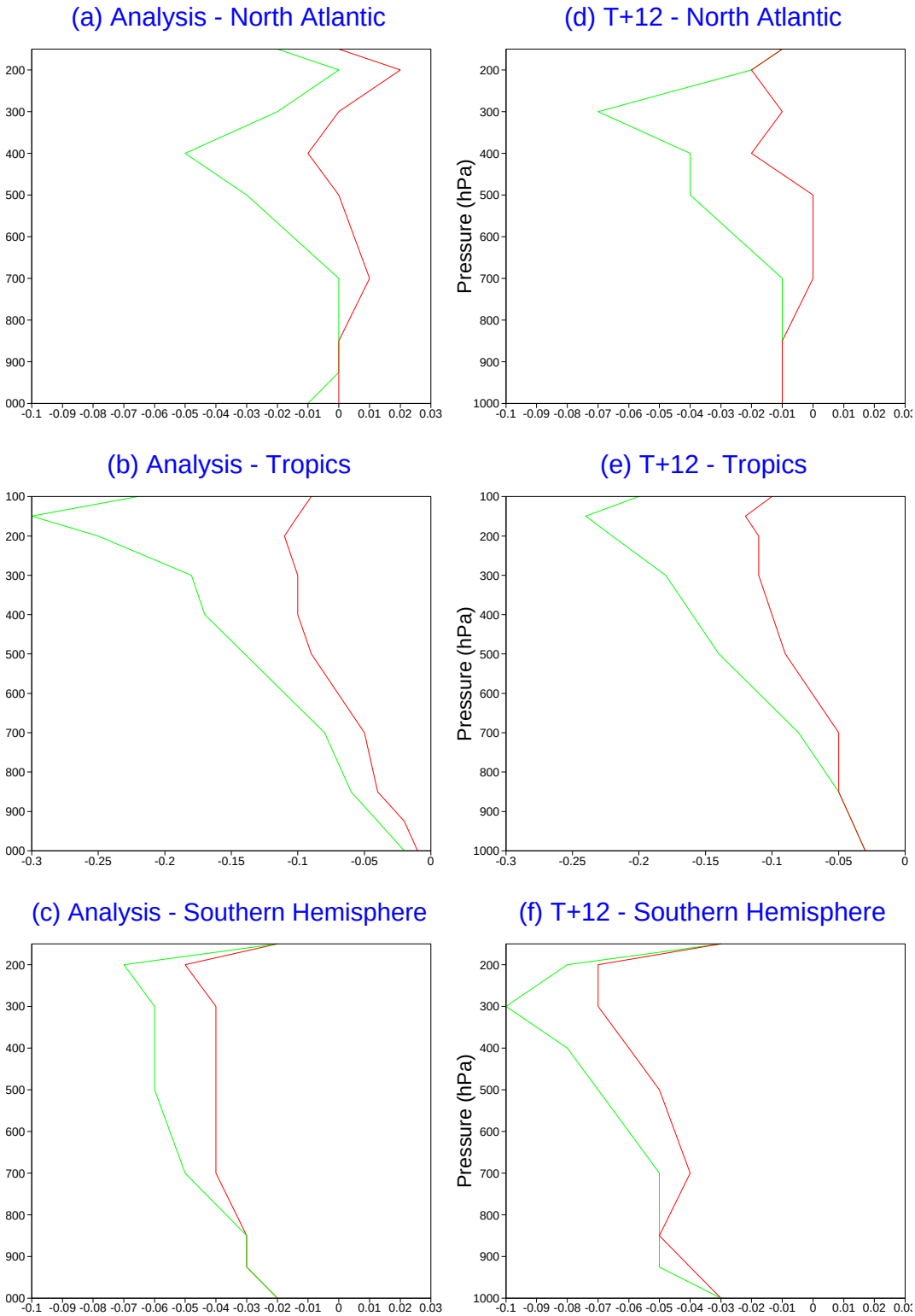


Figure 30: Area-integrated change in ensemble spread for zonal wind. Radiosonde impact in red, radiosonde+DWL in green. Positive values are an artefact of the finite statistical sample.

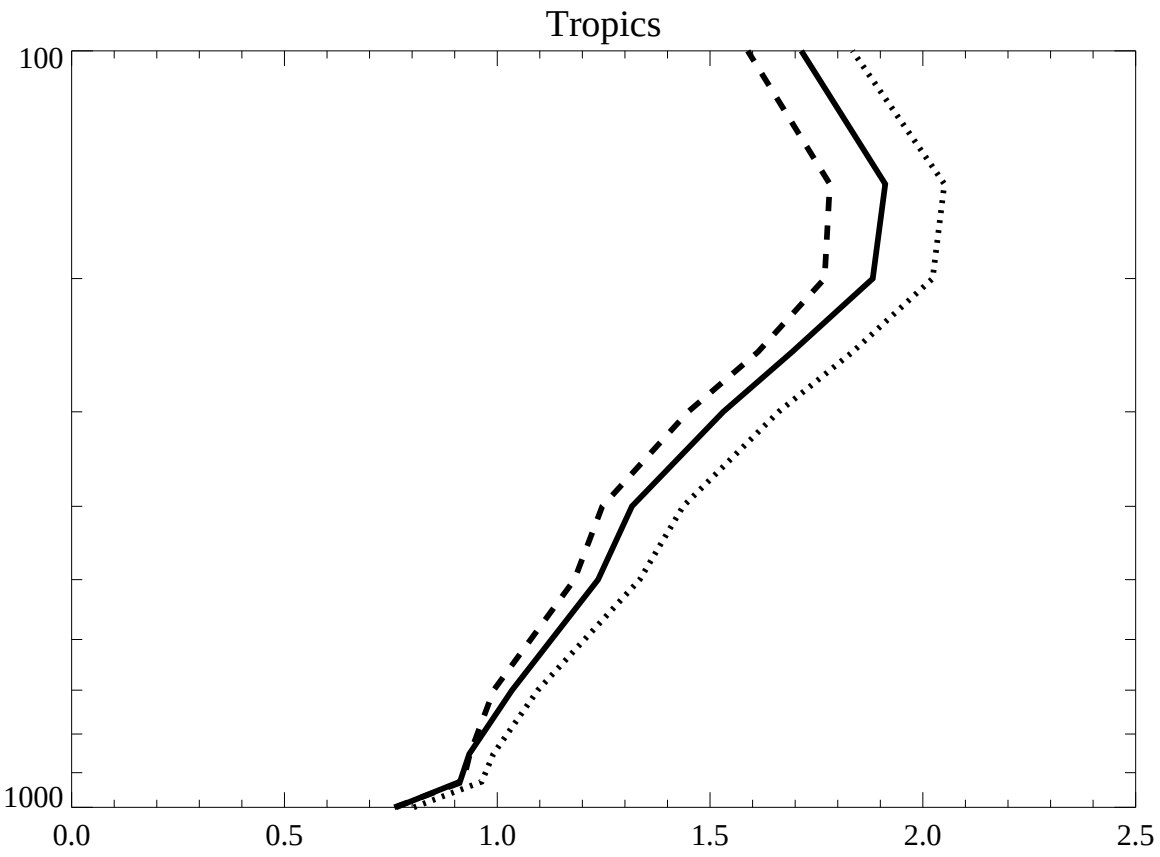


Figure 31: Ensemble spread for zonal wind (horizontal axis, m/s) in the Tropics. Vertical axis: pressure (hPa). Solid curve: Control ensemble (2004 observing system), dashed: DWL added, dotted: radiosondes/wind-profilers removed.