SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Aeolus' impact estimation for different sampling scenarios using EDA experiments
Computer Project Account:	spsekoer
Start Year - End Year :	2012 - 2012
Principal Investigator(s)	Heiner Körnich Linda Megner
Affiliation/Address:	Department of Meteorology Stockholm University SE-10691 Stockholm Sweden Heiner Körnich is since May 2012 affiliated at: SMHI SE-60176 Norrköping Sweden
Other Researchers (Name/Affiliation):	Lars Isaksen, David G. H. Tan, Andras Horanyi (ECMWF) Gert-Jan Marseille (KNMI)

The following should cover the entire project duration.

Summary of project objectives

(10 lines max)

The Earth Explorer Atmospheric Dynamics Mission (Aeolus) is a unique space-borne Doppler wind lidar that will yield vertical profiles of line-of-sight wind observations with global coverage . The launch of the polar-orbiting satellite is currently planned for 2014. This special project contributes to the ESA study ”Vertical and Horizontal Aeolus Measurement Positioning” (VHAMP) that consists of collaboration between KNMI, the department of meteorology at Stockholm University (MISU) and Met.No. Specifically, in the proposed project, it is examined:

1. How reliable is the EDA tool for impact estimation of observing systems?
2. Does the recent change in instrument design still ensure mission impact?
3. How do different spatial sampling scenarios of the satellite affect Aeolus' impact on NWP

Summary of problems encountered

(If you encountered any problems of a more technical nature, please describe them here.)

- The set-up of the EDA experiments required intense support from Carsten Maass. Owing to his almost immediate help, the set-up was not a problem for us.
- In the beginning of the project, the tool obstat was not available from the outside, but fortunately this changed towards the end of the project.
- As the Aeolus satellite is still under development at ESA, postponements occur when technical difficulties are encountered. These problems have also an effect on this special project, as new instrument characterisations have a direct impact on the simulated Aeolus observations produced for our experiments. Recently, an inconsistency in the instrument sensitivity of the so-called Mie channel was found when calculated with different tools. This caused a halt in our experiments in the middle of the project. Despite this delay we were able to finish the special project during 2012.

Experience with the Special Project framework

(Please let us know about your experience with administrative aspects like the application procedure, progress reporting etc.)

The email reports on the used project resources are useful. The reporting procedure was not entirely clear or well documented. An automatic email would be useful.

Summary of results

We have used the EDA method to investigate the impact of the Aeolus data on NWP quality for different laser configurations; the original Aeolus burst-mode (Aeolus BM 110 mJ), the corresponding continuous-mode configuration (Aeolus CM 110 mJ) and the CM configuration with the laser power reduced to 80 mJ (Aeolus CM 80 mJ). According to the EDA method it is shown that the impact on the forecast quality, regardless of laser configuration, is of the same order as the impact of the entire fleet of radiosondes and wind profilers. The differences in the impact of the observations on NWP between the various laser operational settings, both on a global average and on a regional basis, are minor. For all settings improved forecast quality seen in the tropics and the ocean regions in the troposphere and the lower stratosphere. Forecast improvement is also indicated in the winter polar region at higher altitudes. The fact that the EDA indicates a substantially larger impact of Aeolus data in the tropics than in the extra-tropics, may be connected to that our experimental setup, with a relatively coarse model resolution, better observes the impact on the large scales; large-scale wind in the extra-tropics is determined by geostrophic balance. Thus, it is possible that the Aeolus impact in the extra-tropics will be larger in a model with higher resolution, such as the running IFS model version when the Aeolus data becomes available. Moreover, the localized positive Aeolus impact in the tropics indicates that these are sensitive regions where additional observations are beneficial for an improved description of the flow.

The impact reduction from CM 110 mJ to CM 80 mJ is surprisingly small, given that the data in the latter scenario is degraded both in quality (13 % larger errors) and in quantity (approximately 10% less observations). This subtle change in impact might result from the good impact on large-scales which is still very accurately sampled for both laser configurations, at least relative to our model resolution. However, this lack of benefit may be due to the coarse resolution of our model. The running model of 2015, when the Aeolus data is to become available, will have better resolution, and thus it is possible that system will benefit substantially more from the small-scale observations.

The EDA method is a relatively new method for impact assessment, and it has never been fully evaluated. The first part of this project was therefore aimed at conducting an evaluation of the EDA method, with focus on identified issues that may jeopardize the trustworthiness of EDA estimates. One such issue was the assumed linearity between the magnitude of the perturbations added to the observations and the resulting EDA spread. This assumption is part of the theory on which the EDA method is founded, and is thus crucial for EDA impact assessment. The linearity assumption was tested by conducting two experiments with reduced magnitude of the perturbations and assessing if the resulting EDA spread was reduced in a linear fashion. It was shown that there is indeed a linear relationship and that it is valid throughout the domain. Another issue that was investigated was if the added perturbations would cause the quality control system to reject the perturbed observations. This would cause the EDA spread not to increase as expected with an underestimate of the EDA forecast error as a result. It was shown that almost all observation systems had not been rejected to any larger extent, thus the vast majority of the perturbed observations had passed the quality control system. The exceptions were the atmospheric motion winds and the limb observations, of which 10% and 5% of the perturbed observations had been rejected. Since these two observing systems do not constitute a significant part of the total observations, this issue is not of great concern as long as the observations that one is assessing the impact of are not too unrealistic. However, this issue could play a significant role if one is assessing impact of unrealistic observations, as we shall discuss soon.

During the project other short-comings of the EDA method has been brought into light. In particular the EDA methods inability to observe a bias of an observing system may be problematic. The LIPAS data set indeed is contaminated by such a bias, in that it is generated by the UKMO model, which state will differ slightly from that of the IFS model. This is likely the explanation to the negative impact of the LIPAS data set that was observed by OSE analysis in an independent study at ECMWF (personal communication). Since we are not interested in this bias of the UKMO model, it does, in this particular case, not cause a problem, and is even advantageous. However, it is obviously a short-coming of the EDA method that such a bias cannot be estimated, and it should be noted that a possible unknown bias of the Aeolus data, which would be highly detrimental for the NWP forecast quality, can not be determined in this way. Another short-coming of the EDA method is that the rejection of data by the quality control system that will cause the EDA spread to saturate when the observations become unrealistic, that is to far from the model state. However, if this happens to as large of an extent in the baseline experiment as in the Aeolus experiment, then the spread of the two experiments will be reduced by the same rate, or factor. The impact estimate would thus be decreased by this factor. However, since the impact has been scaled to yield the same magnitude as the OSE estimate, such a factor has already been implicitly handled. This means that the scaled impact estimate will not be affected by the saturation of the ensemble, as long as the saturation effect is moderate and the perturbed observations are rejected to the same degree in the two experiments we are comparing. Table 1 shows that the rejection rate of the baseline and the Aeolus 80mJ is roughly the same. Hence, the saturation issue is not a great concern for the Aeolus impact assessment.

	Baseline experiment	Aeolus CM 80 mJ
Land, Ship	0.055	0.064
Aircraft	0.024	0.023
Atmos. motion winds	8.8	8.4
Buoys	0.093	0.17
Radiosondes	0.21	0.30
Balloons, Profilers	0.085	0.21
Satellite soundings	2.6	2.8
Scatterometer	5.5	5.5
Limb observations	0.006	0.025

Table 1: Percentage of extra loss of different types of observations for perturbed EDA member compared to the unperturbed member for the 31st of January.

In summary the EDA spread reduction is a valuable metric for observation impact assessment, but a conclusive impact assessment should also include observation rejection analysis and an estimation of the bias of the observing system. Since the rejection rate of the Aeolus experiments and the reference experiment were found to be moderate and of similar magnitude the rejection caused by the quality control system does not seem to be of great concern for the Aeolus impact assessment. The bias of the Aeolus system, should such exist, has not been, and cannot be, assessed with the EDA technique. The coarse resolution of our data assimilation system, as compared to the system that will be running when Aeolus data becomes available, further means that it is limited to assess observation impact on the large atmospheric spatial scales. This is a strong limitation, since Aeolus winds are expected to be most beneficial for NWP on the sub-synoptic scales. Thus it is possible that we have underestimated the impact of Aeolus in the regions where meso-scales are more important, such as the midlatitudes.

The related ESA report is attached at the end of the report.

List of publications/reports from the project with complete references

- Megner, L., H. Körnich, L. Isaksen, D. G. H. Tan and A. Horanyi, 2013: Evaluating the linearity assumption of the Ensemble Data Assimilations technique for observing-system-impact assessment. Submitted to Quart. J. Meteorol. Soc.
- Körnich, H., L. Megner, L. Isaksen, G.-J. Marseille, 2012: Preparing for the Atmospheric Dynamics Mission Aeolus using ensemble data assimilation. Poster at the Annual Meeting of the Swedish Space Researcher Collaboration group (SRS) on 13-14 March, 2012 in Stockholm, Sweden.
- Körnich, H., L. Megner, L. Isaksen, G.-J. Marseille, 2012: Preparing for the Atmospheric Dynamics Mission Aeolus using ensemble data assimilation. Poster at the 22nd ALADIN workshop / HIRLAM All-Staff Meeting 2012 on 7-10 May 2012 in Marrakech, Marrokko.
- Megner, L., L. Isaksen, D. Tan, and A. Horanyi, 2012: Assessment of the EDA (Ensemble of data assimilation) technique as a tool for estimating the uncertainty of a prediction and the impact of observations. AGU Fall meeting 2012, December 2012, San Francisco, USA.
- Megner, L., and H. Körnich, 2013: Forecast quality impact assessment of Aeolus wind observations. ESA-report AE-TN-MISU-VHAMP-010_v0.4. Attached to this document.

Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new Special Project.)

During the project, more questions were raised how the EDA method is comparable to an OSE or OSSE experiment and how the EDA method can be used to provide impact estimates of observing systems. More research is needed here.
Generally, a very strong impact was found for wind observations. It is however not clear how high-resolution observations contribute to the forecast quality. This becomes especially interesting with the increasing resolution of the forecast model and reduced decorrelation lengths in the background error covariance matrix. This question is also important for the integration length of the Aeolus observation.

AE-TN-MISU-VHAMP-010_v0.4

Forecast quality impact assessment of Aeolus wind observations

Linda Megner, Heiner Körnich
MISU, SMHI

Change Log

Date	Author	Comment
10/03/2012	L. Megner	First version 0.1
20/10/2012	L. Megner	v 0.2. Contains the comparison of BM, CM 110 mJ and CM 80 mJ laser scenarios. The EDA evaluation from draft 0.1 has not yet ben rewritten into .doc format and is therefore not yet included in this document. For the theory evaluation of EDA see version 0.1.
20/02/2013 19/03/2013	L. Megner, H.Körnich L. Megner, H.Körnich	v 0.3 First version of the full report v 0.4 Extended discussion of data usage in section 4.3 and section 6.

Contents

1 Introduction

The Global Observing System lacks accurate global wind information. This gap will be filled by the globally distributed vertical profiles of horizontal wind by the Aeolus mission. The observations will be provided by the ALADIN instrument, a Doppler-wind lidar on board the Aeolus satellite. The Aeolus mission is expected to demonstrate the impact of these data on weather forecasting.

The VHAMP - "Vertical and Horizontal Aeolus Measurement Positioning" project is a follow up to the original VAMP - "Vertical Aeolus Measurement Positioning" project that had the aim to give recommendations for the operation of Aeolus with regard to spatial sampling, in particular for the positioning of the vertical measurement bins, to provide maximum mission benefit. As the Aeolus project proceeded, the specifications of the ALADIN laser changed. The original plan was to run the laser in a so called "burst-mode" with 110 mJ energy [A1], where the laser would be measuring in bursts every 200km. Due to problems with the laser it was decided to measure continuously (so called continuous mode, "CM") instead. The VHAMP project was launched, to investigate also the horizontal positioning and to evaluate the Aeolus observation impact also in CM mode.

As the project proceeded, it was suggested to reduce the power of laser to 80 mJ, at least during the first period of the mission in order to prolong the laser lifetime. The VHAMP project was extended to include studies of the effect of this on the expected quality of the wind measurements and on the impact of the observations in NWP systems. This meant that not as many resources could be used for the original goal, which means that only 2 different sample scenarios will be investigated.

This document reports on VHAMP task 5, which is the impact assessment of the Aeolus CM observation scenarios on NWP assimilation system. Our tool to analyze the impact of the system is the ensemble of data assimilations ensemble technique (EDA). The first step of task 5, is to analyze the representativity of the EDA ensemble spread as a tool for assessing impact of observing system. In section 4 we therefore evaluate the EDA impact estimate by investigating the data usage, and by comparing the EDA estimates of forecast error and impact to those of the more traditional OSE technique. In section 5 we report of the impact of the different laser specifications and sampling scenarios on NWP quality.

The NWP system used for the 4D-var data assimilation is the European Centre for Medium-Range Weather Forecasts, ECMWF's, integrated forecast system (IFS) Cycle 35 and Release 2. The models spectral resolution is T399 (which equals an approximate model grid resolution of 50 km) and the 91 vertical levels reach from the ground to 0.1 hPa (around 65 km).

2 Documents

Both applicable and reference documents are provided below.

Applicable documents

	Document	Issue	Title
[A1]	*AE_TN_KNMI_VAMP_FR*	1.0	Vertical Aeolus Measurement Positioning Final Report, 9 November 2010
[A2]	*AE_TN_KNMI_VHAMP_009*		Simulation of Aeolus CM wind observations and recommendations for Aeolus sampling

Reference documents

1. Arnold, C.P. Jr., D. C. H. (1986) Observing-Systems Simulation Experiments: Past, Present, and Future. . *Bull. Amer. Meteor. Soc.* ,
2. Tan, D. G. H., Andersson, E., Fisher, M., & Isaksen, L. (2007) Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM–Aeolus wind profiling mission. *Quarterly Journal of the Royal Meteorological Society* 133, 381--390.
3. Fisher, M. (2003) Background error covariance modelling. *ECMWF Seminar on Recent developments in data Assimilation for Atmosphere and Ocean* , 45-–63.
4. Zagar, N., Andersson, E., & Fisher, M. (2005) Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *Quarterly Journal of the Royal Meteorological Society* 131, 987--1011.
5. Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., & Raynaud, L. (2010) Ensemble of data assimilations at ECMWF.
6. Leutbecher, M., Jung, T., & Rodwell, M. (2009) Revision of the stochastic physics scheme.
7. Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., & Weisheimer, A. (2009) Stochastic parametrization and model uncertainty.
8. Shutts, G., Leutbecher, M., Weisheimer, A., Stockdale, T., Isaksen, L., & Bonavita, M. (2011) Representing model uncertainty: Stochastic parametrizations at ECMWF. *ECMWF Newsletter* 129, 19-24.
9. Andersson, E. & Sato, Y. (2012) Outcome and recommendations of the WMO Workshop on the Impact of Various Observing Systems on NWP.
10. Bonavita, M. (2012) Ensemble of data assimilations and uncertainty estimation. *ECMWF Seminar on Data assimilation for atmosphere and ocean* , 135--160.
11. Cress, A. & Wergen, W. (2001) Impact of profile observations on the German Weather Service's NWP system. *Meteorologische Zeitschrift* 10, 91--101.

12. Megner, L. (2013) Evaluating the linearity assumption of the Ensemble Data Assimilations technique for observing-system-impact assessment. *Quarterly Journal of the Royal Meteorological Society* , to be sumbitted.

3 Acronyms

ADM	ESA's Core Earth Explorer Atmospheric Dynamics Mission
BM	Burst-Mode
CM	Continuous-pulsed Mode
DWL	Doppler wind lidar
ECMWF	European Centre for Medium-range Weather Forecasts
EDA	Ensemble Data Assimilation
ESA	European Space Agency
HLOS	Horizontal Line of Sight
LIDAR	Light Detection And Ranging
LIPAS	Lidar Performance Analysis Simulator
LOS	Line-of-Sight
NWP	Numerical Weather Prediction
OSE	Observing System Experiment
OSSE	Observing System Simulation Experiment
RMS	Root-Mean-Square
TN	Technical Note
UKMO	United Kingdom Met-Office
VAMP	Vertical Aeolus measurement Positioning
VHAMP	Vertical and Horizontal Aeolus measurement Positioning

4 Evaluation of the EDA method for Impact Assessment

4.1 Introduction

In order to examine future observing-systems an OSSE (observing system simulation experiment) is the usual choice [1]. However such an endeavor has two disadvantages; it is computationally costly and it relies on a representation of the true state, e.g. a nature run from an "independent" model, which typically generates an unwanted bias. Alternatively, the data assimilation ensemble technique (EDA) can be employed for impact assessment [2]. The EDA is based on the fact that input and output fields of the assimilation cycles contain errors. The input consists of surface fields (e.g. sea surface temperature), the observations, and the background field. Output fields are the analysis and the forecast. If the observations and surface fields were perturbed externally in an ensemble of data assimilation systems and if these perturbations were chosen with the statistical characteristics of the true input errors, then the spread of the ensemble analysis and of the ensemble forecast will have the statistical characteristics of the analysis and forecast error [3] [4]. This means that the ensemble spread may be used to estimate the forecast error. For more information on the EDA method see [A1], [4]. For the presented experiments, only observations and surface fields that are an outcome of the model, like surface temperatures and soil moisture, were perturbed at each assimilation time step, while constant surface fields as the roughness length were kept fixed in all ensemble members. As the subsequent analysis uses the output forecast from the preceding cycle, these perturbations propagate to the background field and to the proceeding cycles, see [4] [5]. The method thus ensures that the background states are implicitly perturbed to reach a steady-state level after a 3-7 day spin-up period. This means there is no need for explicit background error perturbations. The model uncertainties were represented by the SPPT (stochastic physics) scheme [6] [7] [8] that perturbs parameterization tendencies at every time step of the model integration. The magnitude of the perturbation of the stochastic physics is reduced above 100 hPa and reaches zero at 50 hPa [6].

The EDA technique can be used to estimate the impact of an observation type by running two sets of ensembles, one with the observation type and one without. The forecast ensemble spread will differ for these two sets, since the observation type in question will help to constrain the forecast. The spread difference will then provide an estimate for the impact of this observation. The impact of future observing systems can also be assessed by simulating artificial observations from short-range

forecasts, which is what will be done here. Note here that the EDA impact estimate is based on analysis of the difference between perturbed ensemble members, which means that a potential bias that affects all the ensemble members will not be seen. The spread therefore represents the standard deviation of the forecast error, not the RMSE, which includes also the bias.
The theory of the EDA method was covered in the [A1]. Here we will evaluate the EDA method, by comparing it to a traditional OSE method, by investigating the loss in data usage because of the error enhancement and by verifying the linearity of the EDA system.

4.2 Comparison of the EDA method and the traditional OSE method

4.2.1 Introduction to the EDA and the OSE technique

Here we compare the impact estimate of the EDA method to that of a traditional OSE technique. Impact will be measured in both cases as changes in forecast error. We choose to study the impact of adding radiosondes to the observation system, since these are important wind measurements that have a substantial impact on the forecast. We thus run two separate experiments. One run with all available observations (the "base line" experiment) and one run where the observations from radiosondes and wind profilers have been removed from the data set (the "denial" experiment). By comparing these experiments we can estimate the impact of adding the observations from radiosondes and wind profilers. This impact will be assessed both using the EDA technique and the standard OSE technique. The experiments were run for 31 days (January 2007), out of which the first 6 days are considered a spin-up period and thus not used in the analysis. Each experiment consists of an ensemble of 10 perturbed members and one unperturbed member.

In the EDA method the impact of an observing system (here radio sondes and wind profilers) is estimated by analyzing the difference in the ensemble spread for the two EDA experiments (one with and one without radio sondes and profilers).
The squared spread of the EDA ensemble experiment for a forecast length f at a certain time t is defined as

$$\sigma^2_{EDA,\,t} = \frac{1}{N}\sum_{n=1}^{N}\left(x^f_{n,t} - \left\langle x^f_t \right\rangle\right)^2 ,$$

where the index n loops over the number of ensemble members (N=10) and the angle brackets denote the ensemble mean. The temporal average is thus

$$\sigma^2_{EDA} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}\left(x^f_{n,t} - \left\langle x^f_t \right\rangle\right)^2 ,$$

where the index t loops over all initial forecast times with a total number of T=52, i. e. 26 dates from January 6 to 31, 2007 with forecasts being initiated each day at 00 and 12 UTC.

The average impact of a specific observation system can now be estimated as the scaled spread reduction between one EDA experiment with this observation system (the base line experiment) and one without (denial):

$$I_{EDA} = \alpha\left(\sigma_{EDA\ baseline} - \sigma_{EDA\ denial}\right) .$$

The factor α is needed because of the well-known known problems with insufficient spread in forecast ensembles due to sub-optimal model uncertainty and observation error specifications [2]{Bonavita2012}. We will estimate this factor in section □.4
In the OSE method the impact is determined as the decrease of the variance of the forecast error. The forecast error in a certain parameter is defined as the forecasted value of that parameter x minus the true value of the parameter x_{true}. The variance of the forecast error is therefore computed as

$$\sigma_{OSE}^2 = \frac{1}{T}\sum_{t=1}^{T}[(x_t^f - x_{true,t}) - \overline{(x_t^f - x_{true,t})}]^2 ,$$

where the over-line denotes the time mean.
The precise true state x_{true} of a parameter in the atmosphere is never known. In lack of a real true state the OSE method uses the best available representation, namely the analysis of the full system where all available observations have been used, i. e. the base line experiment, so that the equation becomes

$$\sigma_{OSE}^2 \approx \frac{1}{T}\sum_{n=1}^{T}[(x_t^f - x_{ba,t}) - \overline{(x_t^f - x_{ba,t})}]^2$$

where the superscript "*ba*" is short for "baseline analysis".
Using the formula above for the OSE error, the estimated OSE impact can then be calculated as

$$I_{OSE} = \sigma_{OSE\,d} - \sigma_{OSE\,b} = \sqrt{\frac{1}{T}\sum_{n=1}^{T}[(x_{d,t}^f - x_{ba,t}) - \overline{(x_{d,t}^f - x_{ba,t})}]^2} - \sqrt{\frac{1}{T}\sum_{n=1}^{T}[(x_{b,t}^f - x_{ba,t}) - \overline{(x_{b,t}^f - x_{ba,t})}]^2}$$

where "d" denotes the radiosonde denial experiment and "b" denotes the baseline experiment.
It is important to realize that the use of the baseline experiment's analysis as the "true state" causes a systematic difference in the evaluation between the two experiments; the forecast errors of the baseline experiment (which uses its own analysis as true state) will be reduced compared to the ones for the denial experiment (which does not use its own analysis as the true state). In other words the use of the baseline analysis favors the baseline experiment with respect to the denial one. Thus the OSE method tends to overestimate of the impact of an observing system, in this case the radiosondes. The use of the baseline experiment was, despite this issue, recently recommended over the alternative of self-verification (using the analysis of each experiment) [9].

4.2.2 Forecast error estimate

Given that the observations and surface fields of each ensemble member have been randomly perturbed with perturbations of the same statistical characteristics as the true input errors, then the spread of the ensemble analysis and of the ensemble forecast will have the statistical characteristics of the analysis and forecast error [3] [4]. Theoretically this implies [4] that σEDA should equal σ_{OSE}. However, due to sub-optimal model uncertainty and observation error specifications, the spread of the ensemble is always smaller than the theory predicts. This problem of insufficient spread in forecast ensembles is one of the reasons why a calibration that inflates EDA spread is required [10].

In this section the main purpose is therefore to examine the spatial distribution of the EDA forecast error estimate compared to that of the OSE. In section 4.2Comparison of the EDA method and the traditional OSE method.4 we will proceed to determine an approximate scale factor with which the EDA estimate should be multiplied to correspond to the OSE estimate.

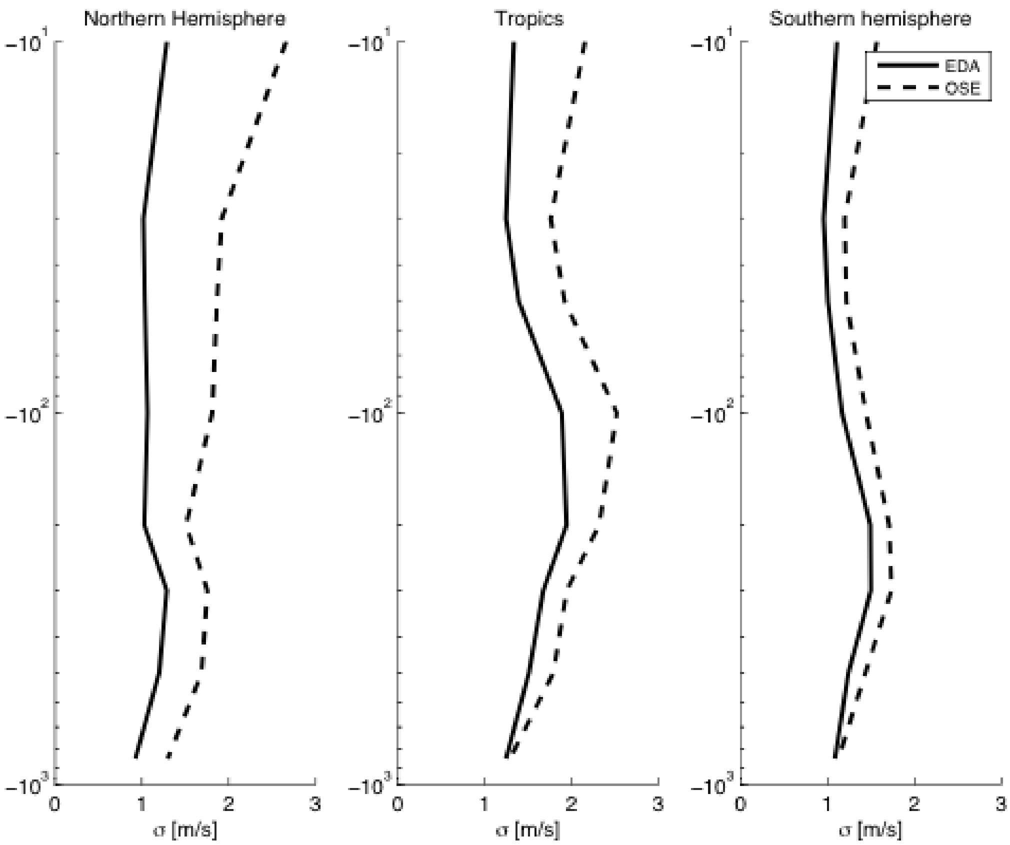


Figure 1: Comparisons between the 12 h zonal wind forecast errors estimated by the EDA and OSE technique, respectively. The left-hand panel shows the Northern hemisphere extra-tropics (90N-30N), the middle panel shows the tropics (30N-30S), and the right-hand panel shows the Southern Hemisphere extra-tropics (30 S-90 S).

Figure 1 shows the average zonal wind ensemble forecast spread σ_{EDA} and the OSE error estimate σ_{OSE} as functions of pressure for the Northern and Southern Hemisphere extra-Tropics and the Tropics. As expected, because of the problem of insufficient ensemble spread, σ_{EDA} is always smaller than σ_{OSE}. In the northern stratosphere the difference between σ_{EDA} and σ_{OSE} becomes larger. In Figure 2 the geographical distribution of the two estimates for different altitudes is compared. In order to visually simplify the comparison of the right- and left-hand panels varying color scales have been used. Here too the difference of the two estimates in the northern stratosphere is clearly visible. Nevertheless, the spatial correlation (given as "C" in the figure) between the right- and left-hand panels is fairly high; between 69 and 87%, with a drop to 63% in the middle stratosphere. This means that the two approaches agree rather well on the spatial distribution of the forecast quality, at least in the troposphere and up to the lower stratosphere.

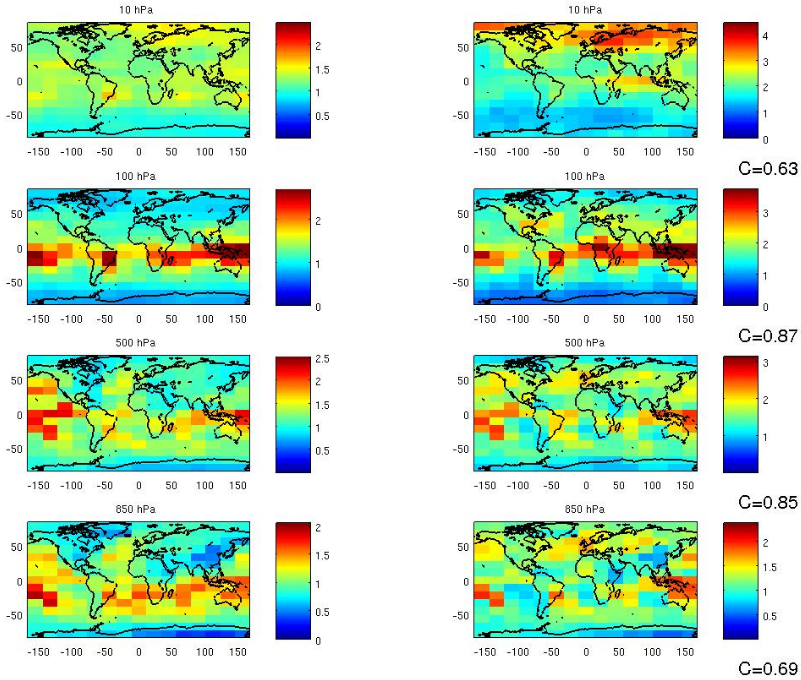


Figure 2: Comparison between the 12 h zonal wind forecast error estimated by the EDA (left-hand panels) and OSE technique (right-hand panels), respectively. The unit of the color scales is m/s. Note that the color scale differs between the plots in order to visualize the spatial similarities. The value C is the spatial correlation between the right- and left-hand panels.

There could be several reasons behind the larger differences in the stratosphere, especially the northern hemisphere. One is that the stochastic physics based perturbations used in the EDA ensemble are reduced at higher altitudes, as described in Section \ref{Experimental setup}. Another possibility is that the observation errors in the stratosphere might be underestimated. A third possible explanation is that the small random perturbations give insufficient perturbations at larger scales. This would mean that the large-scale Rossby waves that propagate to the stratosphere are less perturbed than they should be, which would lead to insufficient spread in the stratosphere. Since the easterlies in the summer hemisphere do not allow planetary waves to propagate into the stratosphere, this would explain the observed asymmetry between the Northern/winter and Southern/summer Hemispheres. Moreover, the large-scale planetary wave perturbations might need a longer time to spin-up, as they are not baroclinically unstable and not as quickly triggered by the observation perturbation as the shorter baroclinic waves.

In general it is clear that the EDA and the OSE approach give similar spatial pictures of the forecast quality but the magnitude of the EDA forecast estimate is less than that of the OSE. Although the OSE method has its own short-comings, it is the standard method and its magnitude should be of the right order. Hence, the EDA spread is too small and there is indeed a need to calibrate or scale the EDA spread to obtain a realistic magnitude of the forecast error estimate.

4.2.3 Impact estimate

We will now compare the denial experiment where the observations from radiosondes, wind profilers and pilots have been removed from the assimilation to the baseline experiment where all data is used. We will use the two different techniques to assess the impact of adding these sets of observations and compare the results. Figure 3 shows the impact as estimated by the uncalibrated EDA technique (I_{EDA} with α=1) on the left hand panels and by the OSE technique (I_{OSE}) on the right hand panels. Both methods suggest that the impact of adding radiosondes is positive, as one would expect. The spatial correlations between the two impact estimates are fair, between 39% and 70%. The declining correspondence in the troposphere is explained by the small impact, in this region, where the airplane measurements compensate for the lack of radio sonde observations [11]. The small impact means that the signal to noise ratio is low, and thus we cannot expect an excellent correspondence of the two methods. In agreement with the results of section Comparison of the EDA method and the traditional OSE method4.2.2 the EDA method suggests a much lower impact than the OSE method (note the difference in scale in the plots) which again stresses the need to scale the EDA estimate.

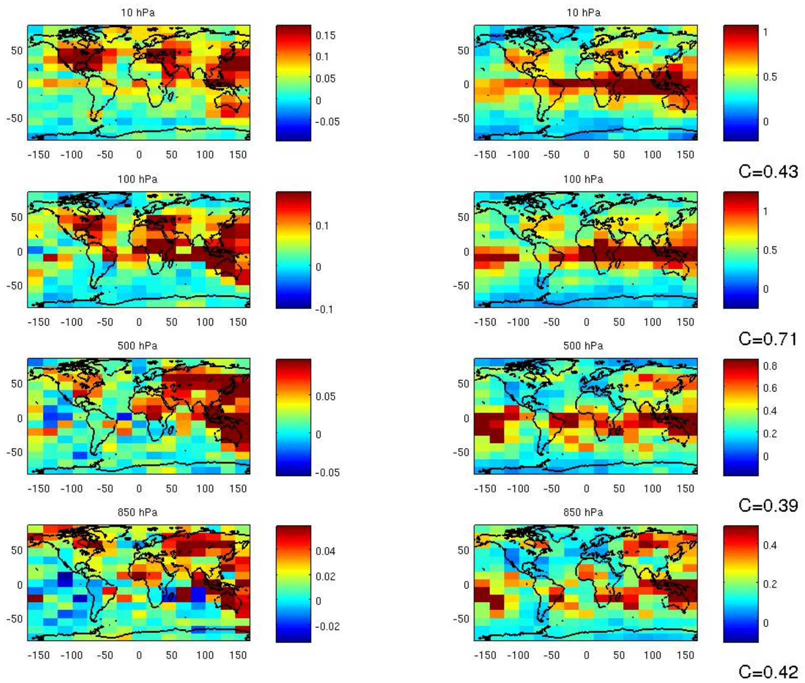


Figure 3: The 12 h forecast zonal wind impact (m/s) of adding radio sondes and profiles as estimated by EDA (left-hand panels) and by OSE (right-hand side panels). Note that the color scale differs between the plots to visualize the spatial correspondence. The value C is the spatial correlation between the right and left hand side plot.

4.2.4 Scaling the EDA method to the OSE method

From the results in the previous sections it is apparent that there is a need to scale the EDA spread to obtain a realistic interpretation of the magnitude of the error and impact estimates. In this section we will estimate this scaling factor. To obtain a true calibration factor it is necessary to have an unbiased estimate of the true state, which we do not have. We will therefore settle to calibrate the EDA method to the OSE result simply by dividing σ_{OSE} and σ_{EDA}. As have been discussed in section 4.2.1 the OSE method has its own problems, which means that neither the OSE estimate is perfect. However, the aim here is to get an approximate estimate of the Aeolus impact, and for this a scaling to the OSE estimate should suffice. Hence, in Figure 4 the global average scaling factor α is obtained by dividing σ_{OSE} and σ_{EDA} is shown. It is clear that this factor increases with altitude from 1.2 at 850 hPa to 1.7 at 1 hPa .

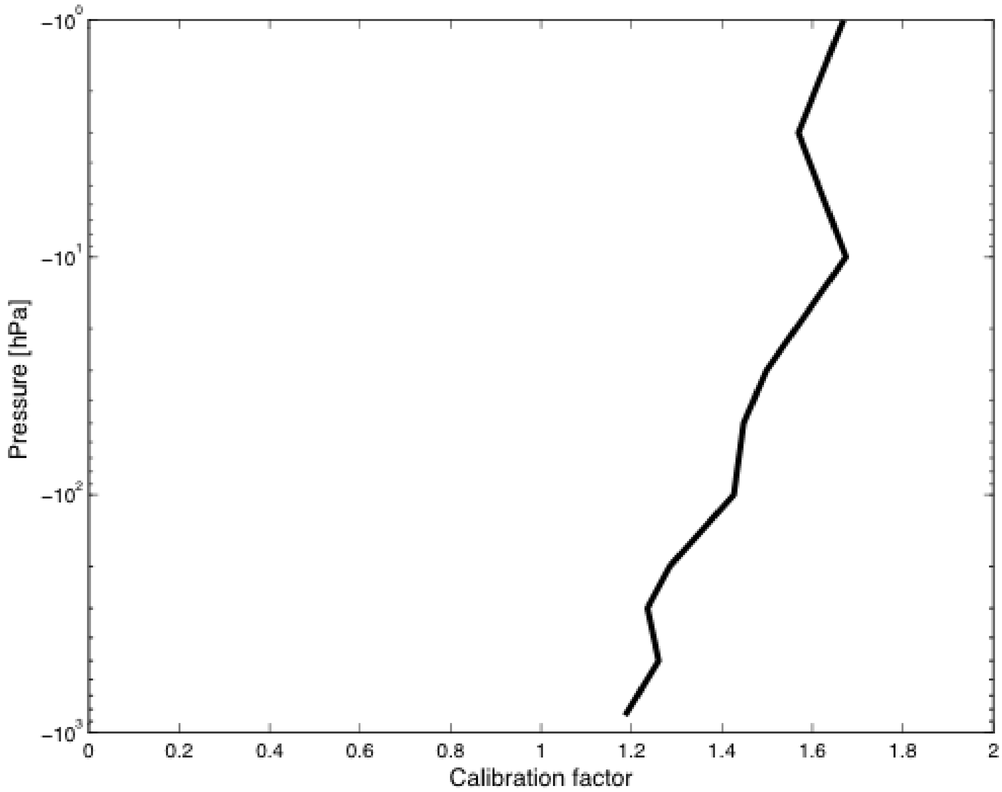


Figure 4: The global average scaling factor between EDA and OSE as a function of pressure.

4.3 Change in data usage due to error amplification

When adding perturbations to the observations there is always a risk that the value of the new perturbed observation falls outside the cut of what is consider a "reasonable observation", and thus is removed by the quality control system that is part of the data assimilation procedure. Obviously, if large perturbations are added, then this would happen often, and we would end up with a considerably reduced number of observations in the system. By having too large perturbations, the end result may in fact even be a reduction of EDA spread, since there are not enough observations left to significantly spread the ensemble. We therefore investigate the to what extent this reduction of data usage has happened in our experiment. As can be seen in table 1 the only observation types that have been significantly are the limb observations and the atmospheric motion winds. These losses have been concluded not to be critical [A1]. Moreover, as will be discussed in section 6, for observation impact

Comment [Linda Meg1]: Anne-Grete, You said that this had been concluded in PM of last summer, is this the document you meant? Can you, or somebody else, add something of why? Or point out where in the report it is mentioned. You also asked if these measurements were referring to GRAS or MIPAS. I don't know how to check that. Why do you ask? Anyway. We are still working a little on the dats usage, as why we have lost so many satellite observations and what this means.

assessment, the actual loss may not be problematic, as long as the loss is of similar magnitude in both experiments, that is the one with and the one without the observations in question.

Observation type	Percentage loss day 1	Percentage loss day 15	Percentage loss Day 31
Land, Ship	0	0	0
Aircraft	0	0	0
Atmos. motion winds	7.2	8.6	9.0
Buoys	0	0	0
Radiosondes	0	0	0
Balloons, Profilers	0	0	0
Satellite soundings	0.2	0.1	0.1
Limb observations	5.0	5.2	5.5

Table 1: Percentage of extra loss of different types of observations for perturbed EDA member compared to the unperturbed member.

4.4 Linearity of the system

The forecast error estimate above was achieved by perturbing the observations and the stochastic physics in the model with perturbations δη and δξ, respectively. In the theory that the EDA technique is based upon [4] it is assumed that these perturbations are small enough so that one can assume a linear relationship between the magnitude of the perturbation and the resulting EDA ensemble spread, $\sigma_{EDA.}$ Here we will investigate if this is a valid assumption. We will do this by reducing the perturbations on the observations and investigate the impact on the total EDA spread.
If this linear relationship holds true, one can write the variance of the EDA ensemble as [12]

$$\sigma_{EDA}^2 = K_1^2\sigma_\eta^2 + K_2^2\sigma_\xi^2 \ ,$$

where K_1 and K_2 are linear operators.
In order to test the linearity assumption we conduct two more experiments, where the observation perturbations δη were reduced to 0.5 and 0.75 times the original magnitude of the perturbations. Due to computational costs we only ran these experiments for 15 days instead of a full month. Again the first 6 days are considered a spin-up period and thus not used in the diagnostics. We can write the expected variances for our three experiments as

$$\sigma_{EDA1}^2 = K_1^2\sigma_\eta^2 + K_2^2\sigma_\xi^2$$

$$\sigma_{EDA0.75}^2 = 0.75^2 K_1^2\sigma_\eta^2 + K_2^2\sigma_\xi^2$$

$$\sigma_{EDA0.5}^2 = 0.5^2 K_1^2\sigma_\eta^2 + K_2^2\sigma_\xi^2$$

We will now express $\sigma^2_{EDA0.75}$ in terms of σ^2_{EDA1} and $\sigma^2_{EDA0.5}$ and thereby obtain an expression for the expected $\sigma^2_{EDA0.75}$, which should be valid if the assumed linear relationship holds true. We will denote this expected $\sigma^2_{EDA0.75}$ by σ2EDA0.75E. By inserting the expression for σ^2_{EDA1} into that of $\sigma^2_{EDA0.75}$ one obtains

$$\sigma_{EDA0.75E}^2 = 0.75^2 K_1^2\sigma_\eta^2 + \sigma_{EDA1}^2 - K_1^2\sigma_\eta^2 = (0.75^2 - 1) K_1^2\sigma_\eta^2 + \sigma_{EDA1}^2$$

Subtracting $\sigma^2_{EDA0.5}$ from σ^2_{EDA1} yields

$$\sigma^2_{EDA1} - \sigma^2_{EDA0.5} = 0.75 K_1^2 \sigma_\eta^2$$

which, inserted into the expression for $\sigma^2_{EDA0.75E}$ above, gives

$$\sigma^2_{EDA0.75E} = (0.75^2 - 1)/0.75(\sigma^2_{EDA1} - \sigma^2_{EDA0.5}) + \sigma^2_{EDA1} = 0.58\,\sigma^2_{EDA0.5} + 0.42\,\sigma^2_{EDA1}$$

Figure 5 shows the EDA spread obtained from the three experiments. As expected, the EDA spread is reduced with decreasing perturbations at all altitudes. The dashed black line represents the expected $\sigma_{EDA0.75E}$. As seen this is very close to the observed $\sigma_{EDA0.75}$. The linearity assumption thus proves valid on a global scale in both the tropo- and stratosphere.

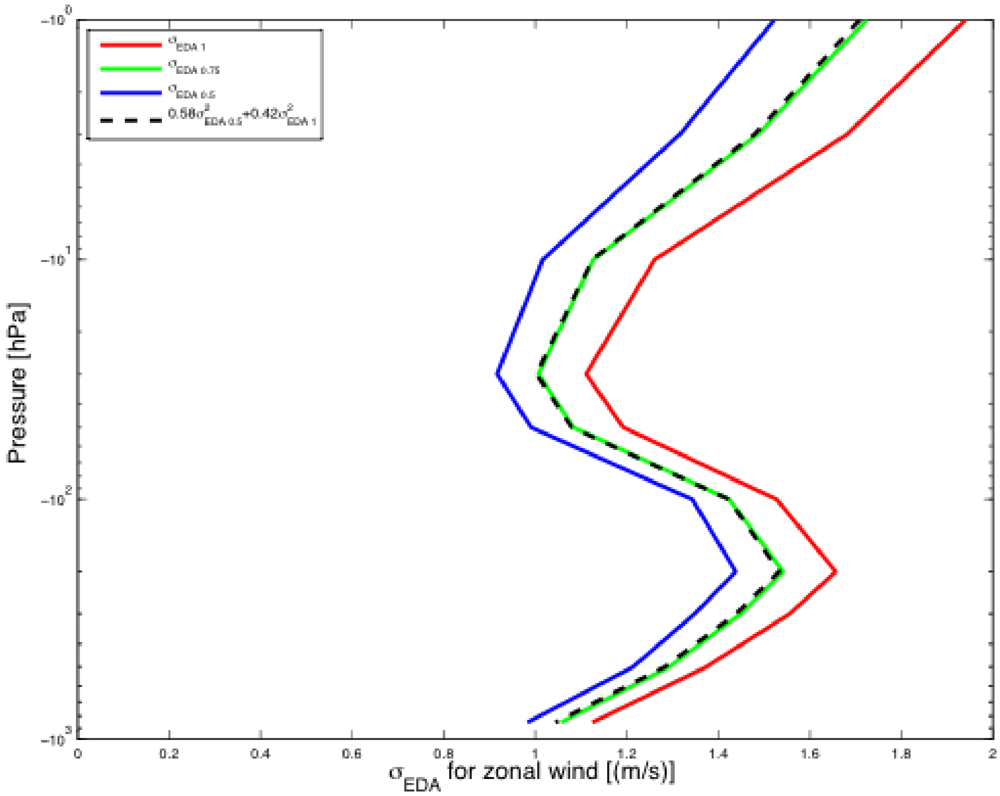


Figure 5: The color-coded lines show the EDA spread in the 12-hour forecast obtained from the experiments with 1, 0.75 and 0.5 times the original perturbation of the observations. The dashed black line shows the expected result for the spread with 0.75 times the perturbation if the linear relationship holds.

Figure 6 shows the spatial distribution of EDA spread for the experiment with full perturbations (left column) and halved (middle column) perturbations of the observations. To check how well the linearity assumption holds in different parts of the domain, we calculate the square root of the difference between $\sigma^2_{EDA0.75E}$ and $\sigma^2_{EDA0.75}$, cf. the right hand panels of Figure 6. We observe that the magnitude of this difference is very small (note the differing color scales) and that the regional dependency has been more or less removed so that mostly random fluctuations remain. There is hence a linear relationship between the perturbation and the EDA spread, not only in the global average as shown in Figure 5, but also in every part of the globe.

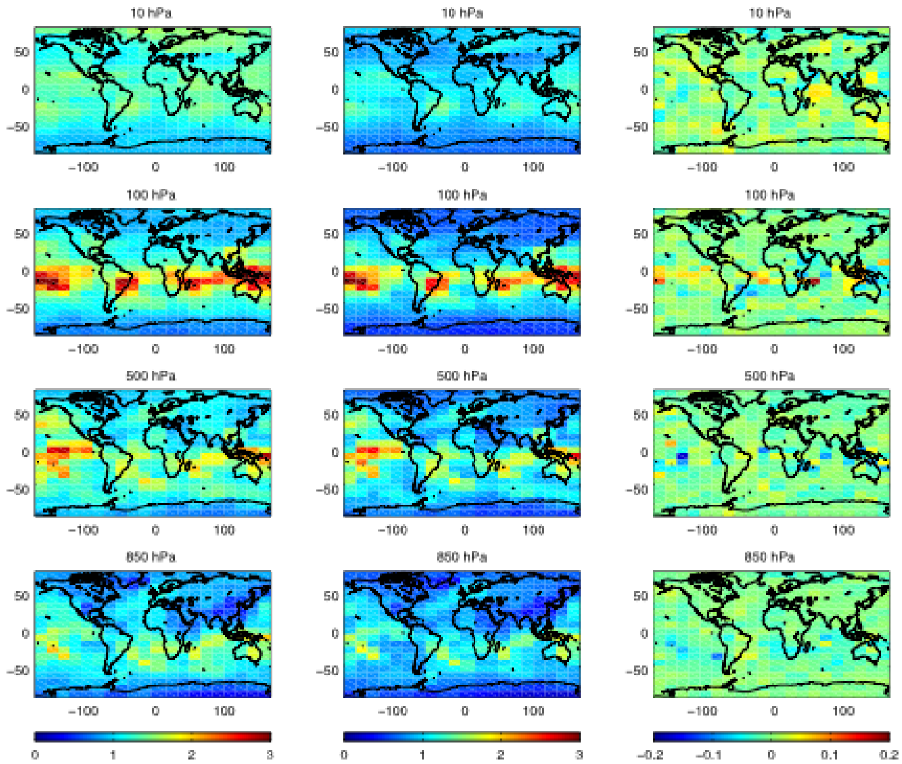


Figure 6: The 12-hour EDA spread in zonal wind. The left-hand column shows the reference case with original perturbations and the middle column the EDA spread when the perturbations have been halved. The right column shows the difference between the expected and measured spread 0.75 times the perturbations. It is clear that this difference is very small and mostly consists of random noise (note the different color scale used for the right-hand column).

5 Impact assessment for different laser modes and energies

5.1 Introduction

As discussed in the introduction to this note the technical specifications for the ALADIN laser has changed during the Aeolus project. In this section we use the EDA method to evaluate the effect that these changes have had on the expected impact of the Aeolus observations on NWP systems. In the same way as in earlier studies we use simulated winds from the LIPAS tool [A2] as observational input to the NWP system.

The vertical sampling scenario adopted in this section is the so-called wvm_tr_zwc1_1km, which has ground-calibration capability (Mie bins reaching the surface) and focuses on tropical cirrus events with the highest Mie bin at 18.5 km altitude. The reasons for selecting this scenario for our calibration runs are covered in [A2].

It was found in [A1] that a conclusive impact assessment of Aeolus with EDA requires simulated winds along the complete orbit rather than along half orbits. The method used to complete the Aeolus orbit is described in [A2]. In order to compare the impact of Aeolus continuous mode data w.r.t. burst mode data, a new calibration run has been performed using simulated BM observations of whole orbits including day-time. This BM calibration run will here be denoted "BM 110 mJ" and has the same specification as the continuous mode run "CM 110 mJ", apart from the switch from BM to CM. In order to reduce the risk for laser-induced damage in the ALADIN instrument, it was suggested to reduce the maximum laser peak energy from 110 to 80 mJ, at least for the first part of the mission. Based on this suggestion it was then decided to run a calibration run also for this lower energy scenario here denoted "CM 80mJ". Table 2 shows the specifications of all laser specification scenarios. The first three scenarios (BM 110 mJ, CM 110 mJ and CM 80 mJ) represent different laser configurations and will be analyzed in section 5.2. The fourth scenario in the table (CM 250 km) investigates the impact of increasing the data integration length from 87 km to 250 km and will be discussed in section 5.3Sampling strategy experiment.

	BM 110 mJ	CM 110 mJ	CM 80 mJ	CM 250 km
laser peak energy (mJ)	110	110	80	80
pulse repetition frequency (Hz)	100	50	50	50
maximum accumulation length (km)	50	84	84	250
Observation separation distance (km)	200	0	0	0

Table 2: LIDAR instrument specifications used for the various experiments.

5.2 Comparison of the NWP impact for different laser scenarios

Here we will compare the impact of the different laser scenarios on the NWP system. For this purpose we have performed 3 experiments: BM 110 mJ, CM 110 mJ and CM 250 km. The specifications for these experiments are found in table 2. We will compare the effect of the different laser specifications on the impact an in turn compare the magnitude of this impact to that of observations from radiosondes and profilers.

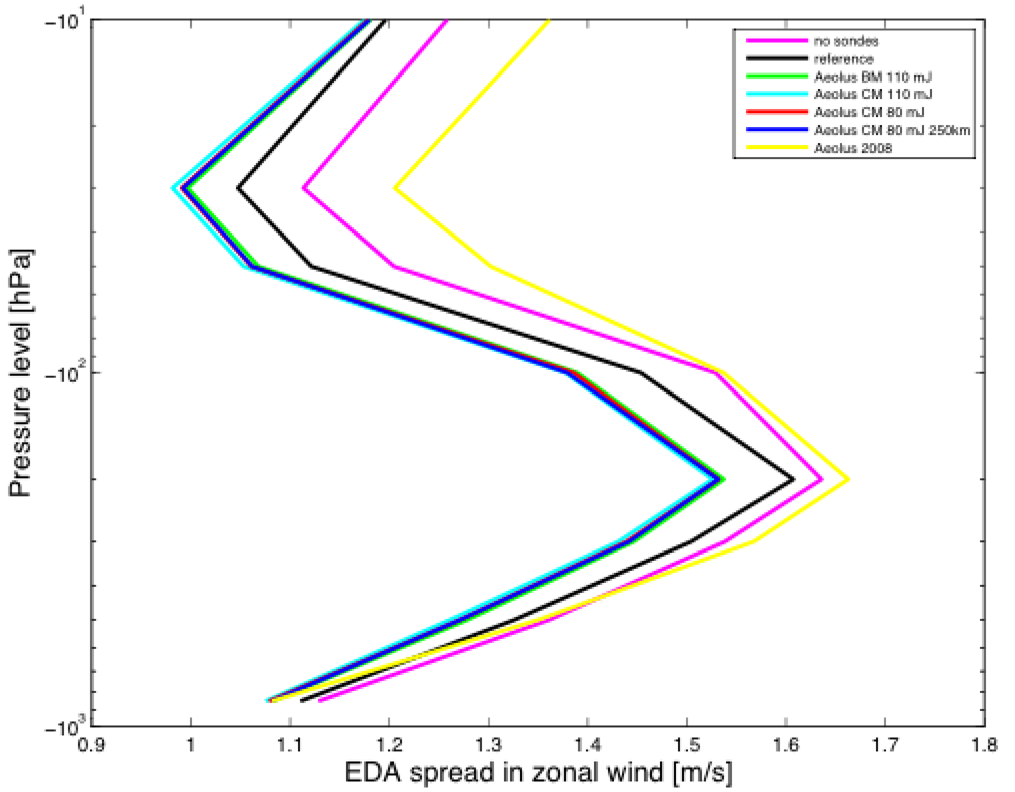
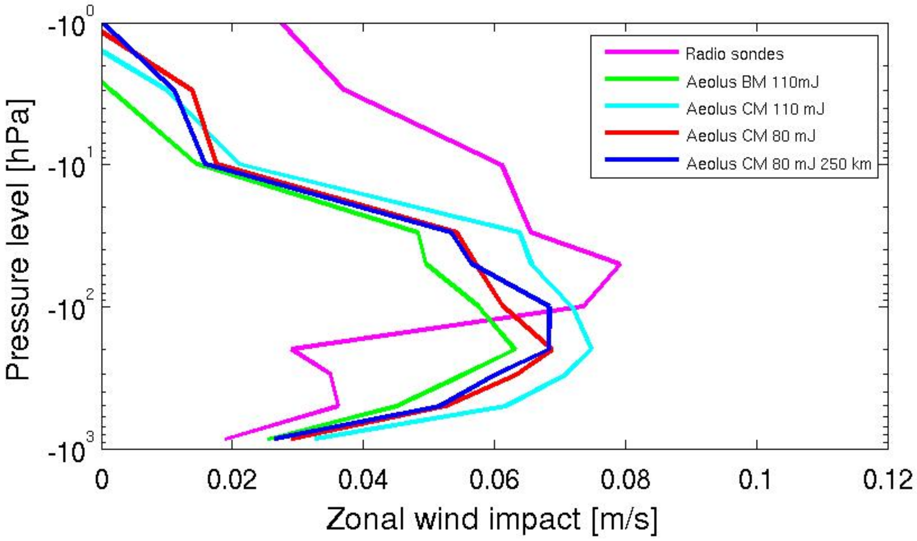


Figure 7: Global average EDA 12 h zonal forecast spread for the different experiments.

Figure 7 shows the global average EDA spread for the 12 h zonal wind forecast, which can be used as an estimate for the forecast error (see section 4) obtained with the different laser scenarios described in section 5.1. (The dark blue and yellow line will be discussed in section 5.3 and 5.4, respectively.) Also shown is the EDA spread for a reference experiment without Aeolus wind observations and for the experiment with neither radiosondes, pilots and wind profilers nor Aeolus wind data, denoted the "no sonde" experiment. As expected, this experiment has increased spread as compared to the reference experiment and all the Aeolus experiments have reduced spread. All the experiments with Aeolus observations have a rather similar EDA spread. To better be able to see the differences between the experiments we calculate the impact of the added observations.
The impact of an added observation type is in section Evaluation of the EDA method for Impact Assessment defined as

$I_{observation\ type} = \alpha(\sigma_{without\ observation\ type} - \sigma_{with\ observation\ type})$, where α refers to the scaling factor between OSE and EDA.



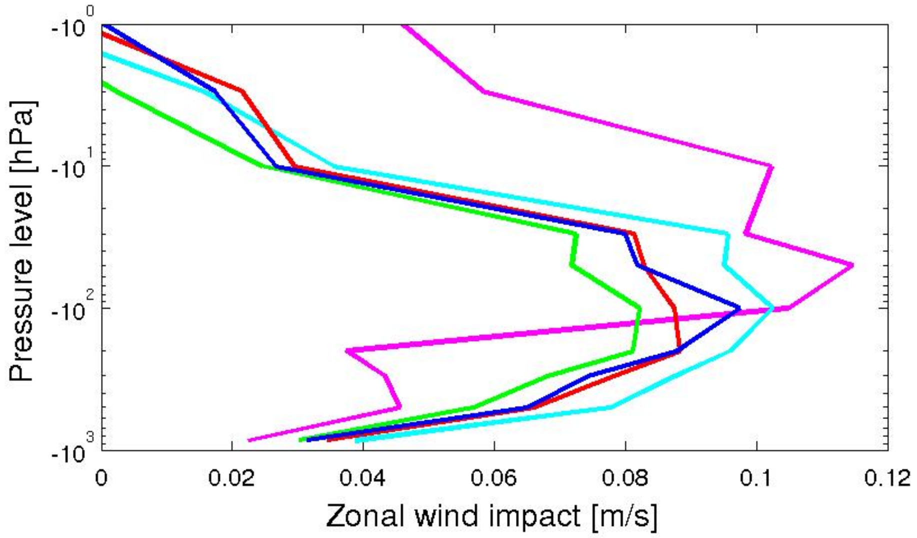


Figure 8: Non-scaled (top) and scaled (bottom) EDA impact in 12 hour forecast quality for the different experiments

Figure 8 shows the impact of the observation from different laser specifications on the NWP forecast, the upper panel show the unscaled EDA spread reduction (α =1) and the lower panel show the result with the altitude-dependent scaling factor determined in section Comparison of the EDA method and

the traditional OSE method.4 (figure 3). Note that the impacts of the Aeolus experiments are calculated as the EDA spread of the reference experiment minus the spread of the Aeolus experiment. The impact of the no sonde experiment on the other hand is calculated as the EDA spread of the denial experiment minus the spread of the reference experiment.
We first note that the impact of adding radiosondes and wind profilers is roughly of the same order as the impact of the Aeolus data in the part of the domain where LIPAS data is available (the Aeolus does not provide observations higher than 30 km which is clearly visible in the impact drop at 10 hPa). This is true independently of Aeolus laser instrument specifications. The BM 110 mJ, CM 110 mJ and CM 80 mJ experiments all show similar impact, and the impact reduction from 110 mJ to 80 mJ is small. The data counts and the observational error on the 31st of January for these two experiments are shown in table 2. In both these aspects the CM 80 mJ experiment is degraded. We will discuss this further below.

	BM 110 mJ	CM 110 mJ	CM 80 mJ	CM 250km
Data Count Aeolus 31st of January	55691	100003	91357	37832
Observation error Aeolus 31st of January [m/s]	2.19	2.13	2.44	1.53

Table 3: Data quantity and error different experiments.

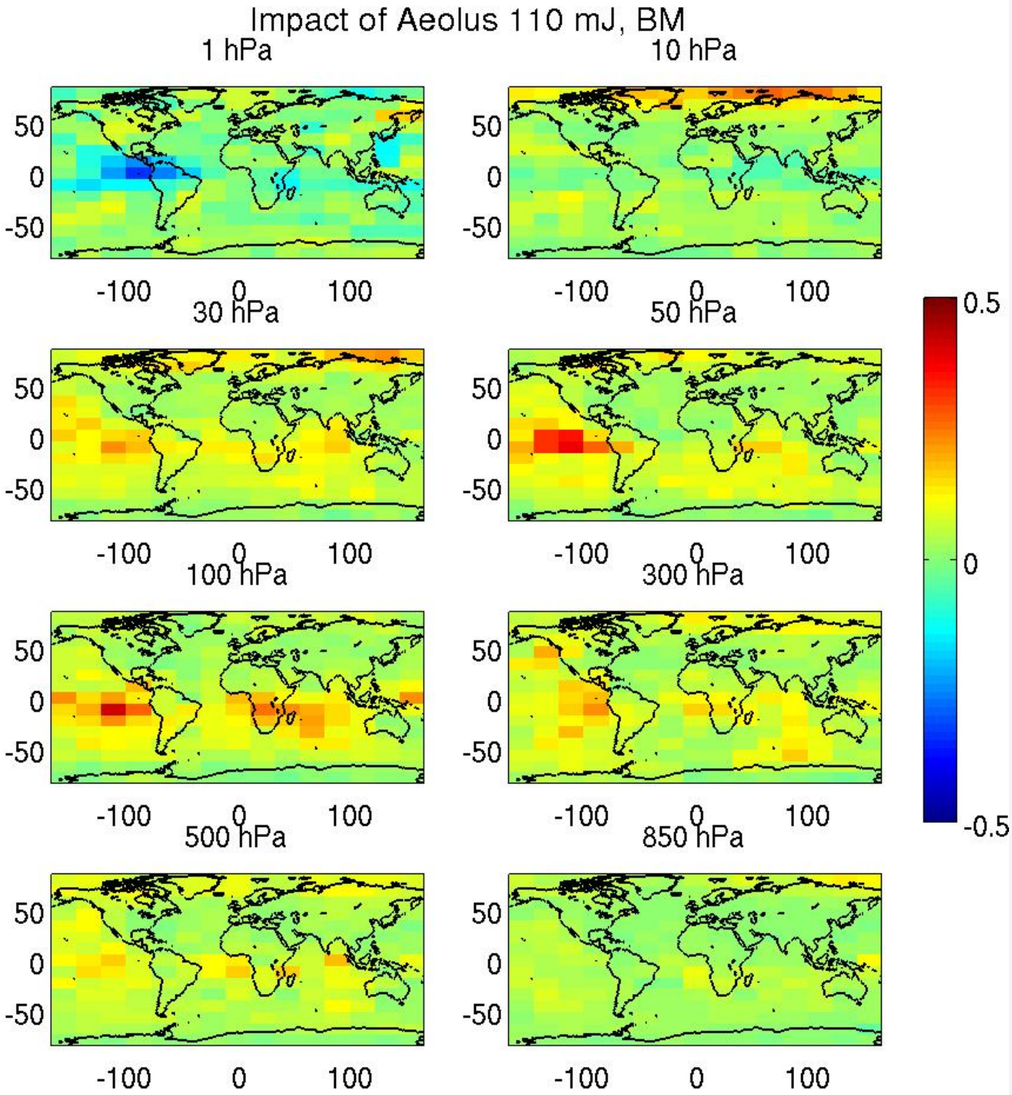


Figure 9: Impact on 12 h zonal wind forecast quality in m/s for the Aeolus BM 110 mJ scenario.

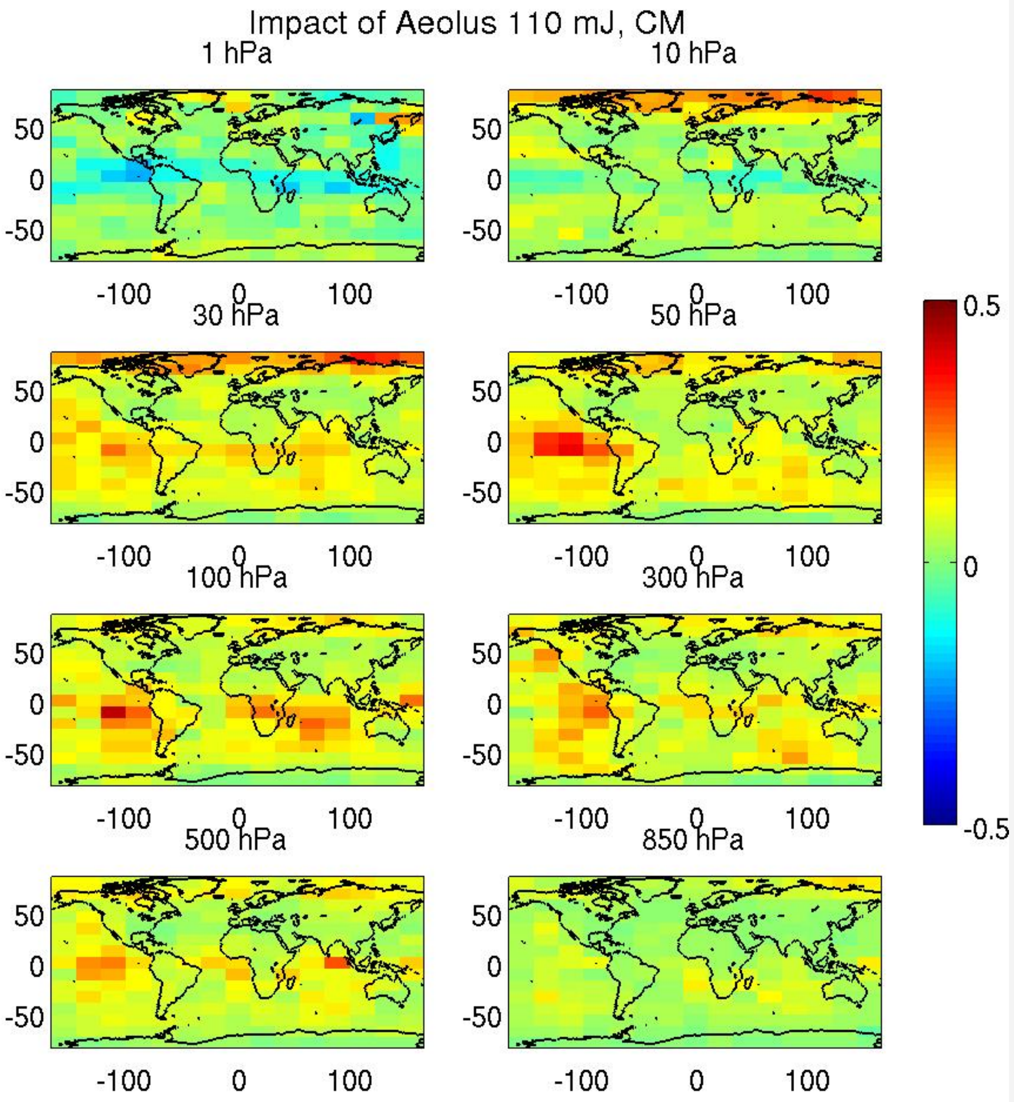


Figure 10: Impact on 12 h zonal wind forecast quality in m/s for the Aeolus CM 110 mJ scenario.

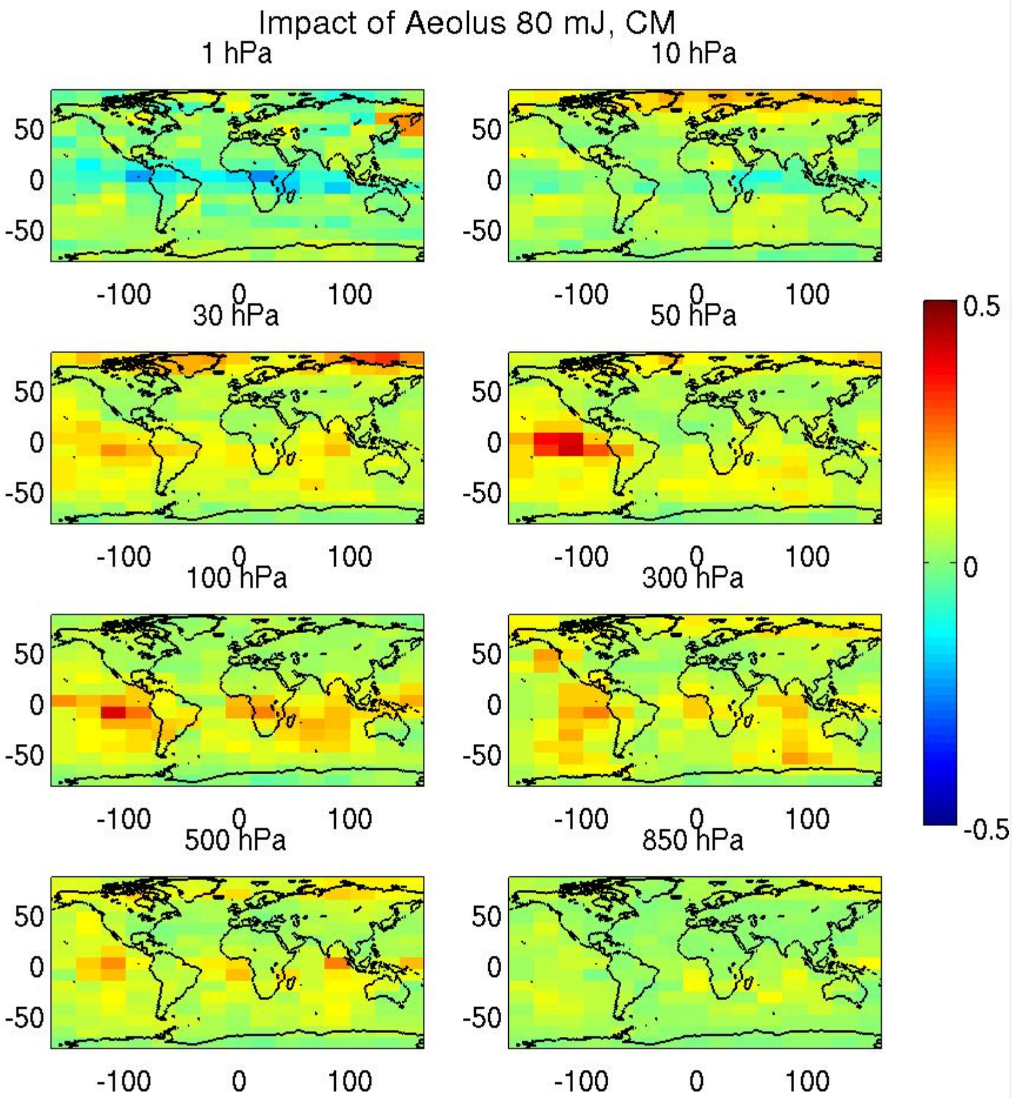


Figure 11: Impact on 12 h zonal wind forecast quality in m/s for the Aeolus CM 80 mJ scenario.

Figure 9 to Figure 11 show the spatial distribution of the impact for the various laser specifications. Here too, there are only minor differences between the experiments. The main impact for all experiments is seen in the ocean regions in the troposphere and the lower stratosphere as well as in the winter polar region.
To better understand the minor differences between the experiments we investigate the observation statistics for the LIPAS doppler winds. This is shown in Figure 12 to Figure 14. The left-hand panels show the standard deviation of the analysis departure (i.e. observation minus the analysis) as dashed lines, and the standard deviation of the background departure (i.e. observation minus background) as solid lines. The right-hand panels show the bias of the analysis departure and the background departure.

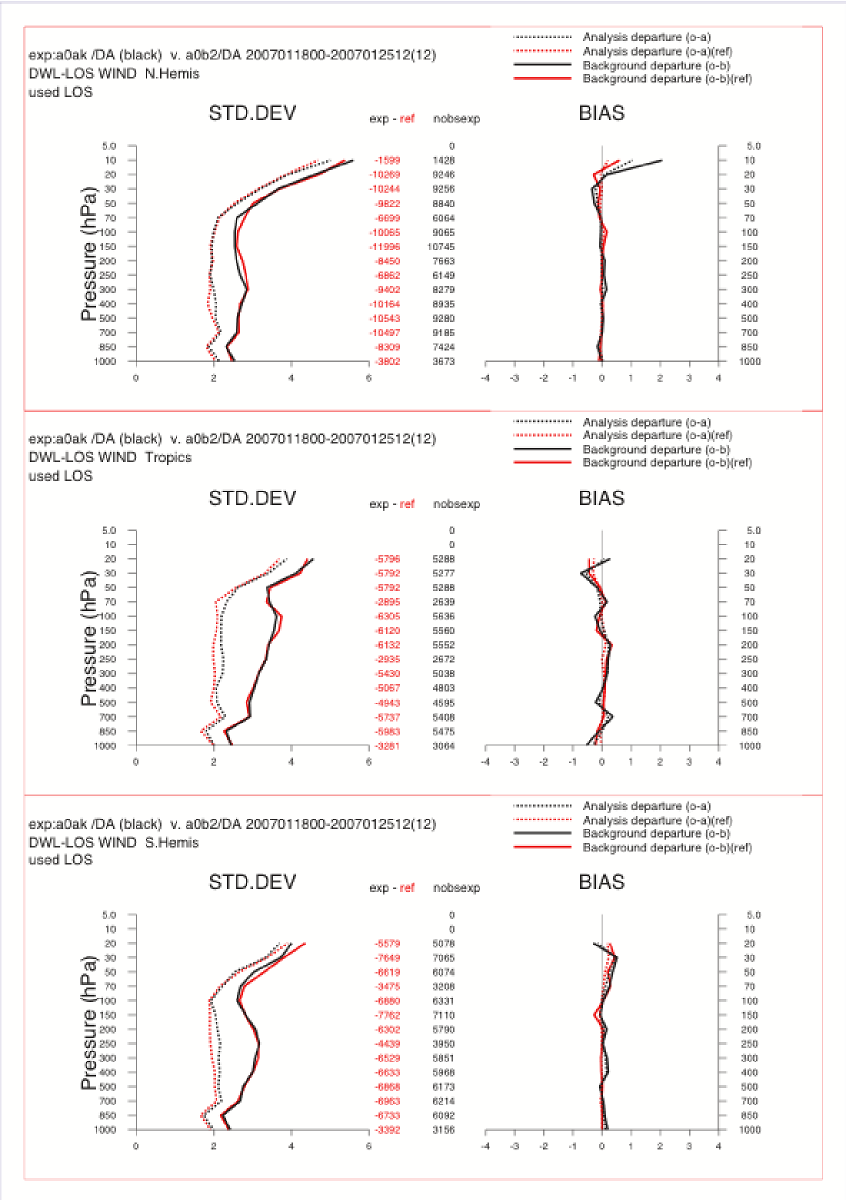


Figure 12: Comparison of the observation departures from the background (solid lines) and the analysis (dashed lines) of the LIPAS Doppler winds for the Aeolus BM 110 mJ (in black, a0ak) and CM 110mJ (in red, a0b2). "nobsexp" is the number of used observations in BM (a0ak) and the "exp-ref" shows the difference of the number of used observations in BM (a0ak) and CM (a0b2). Negative red numbers thus indicate that the CM experiment had more observations.

In Figure 12 the statistics of the doppler winds of Aeolus 110 mJ BM (denoted “a0ak”, in black) is compared to those of Aeolus 110 mJ CM (denoted “a0b2”, in red). The bias of the analysis and background departure of the two experiments are almost the same. The standard deviations of the background departures of the two experiments are also very similar. However, the standard deviation of the analysis departure is smaller for the CM experiment, especially in the tropics and the southern hemisphere. This is due to the larger number of observations in the CM experiment (see table 3, or the numbers in Figure 9), giving more weight in the data assimilation and thus forcing the analysis towards the observations with a reduced analysis departure as a result. Thus, the EDA method (Figure 7) suggests that CM is slightly more beneficial for the forecast quality than the BM, but from the background departures are of the same magnitude of the two experiments. It appears rather that the BM experiment (black in Figure 12) yields slightly smaller background departures. The fact that the impact difference between BM and CM suggested by the EDA method is not visible in the background departure may be connected to the fact that the LIPAS data are not real data, but inferred data from the UKMO winds. It is not obvious that better information about the state of the UKMO model will improve the ECMWF model. This difference thus highlights an important difference between the EDA method of studying the impact of observation and the more traditional method of investigating the departures. The EDA method, which is based on the analysis of differences between ensemble members, as explained in the introduction, cannot observe the bias that would occur if the errors of the observations are not randomly distributed around the true state. As an example, if we had whole observing system that a certain day had a bias of 0.1 m/s, then this would not be observed with the EDA method, but it would be notable in the background and analysis departures. Note that this bias would not show up as a bias in the left hand panel of Figure 12 unless it was consistent throughout the time period. Here, we are not interested in the bias that the state of the UKMO model adds to our system, since the real Aeolus data will not have this bias. However, it is important to note this limitation of the EDA method which we will discuss further in section 5.4.

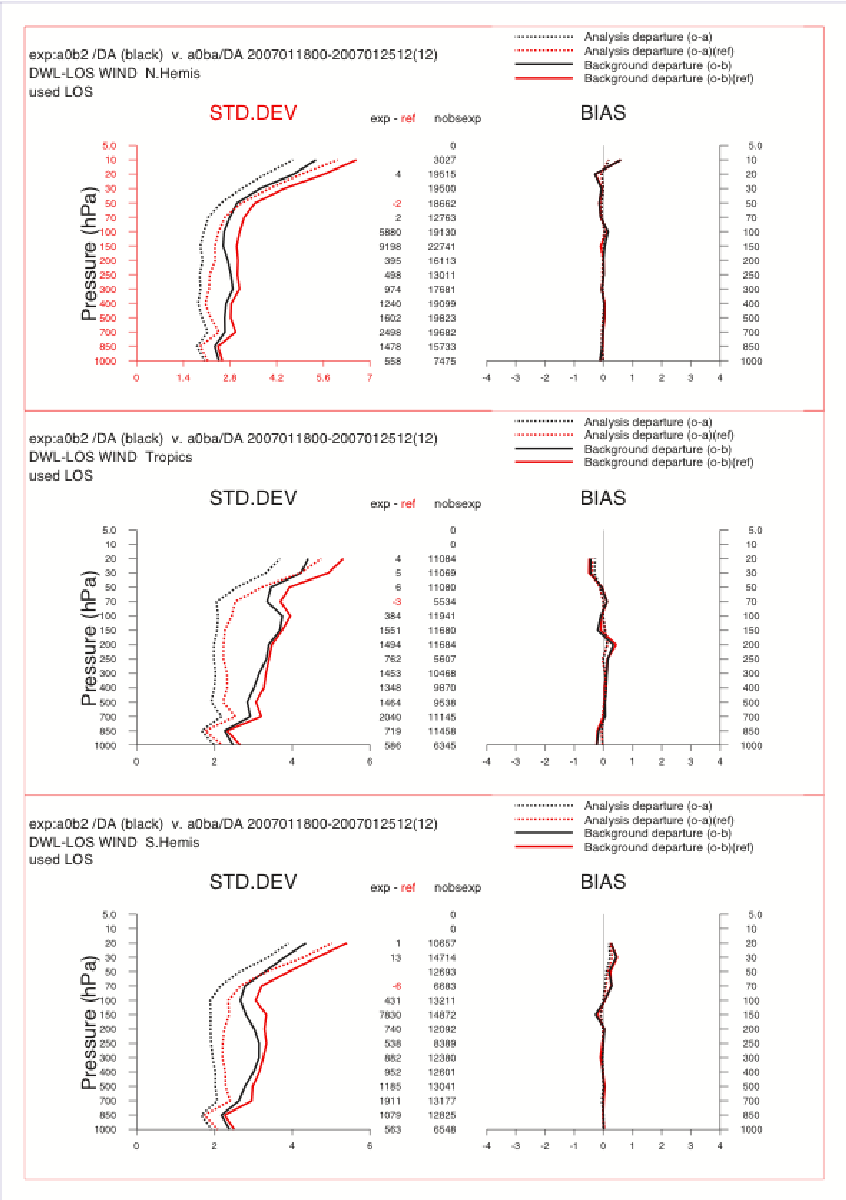


Figure 13: Comparison of the observation departures from the background (solid lines) and the analysis (dashed lines) of the LIPAS Doppler winds for the Aeolus CM 110 mJ (in black, a0b2) and CM 80 mJ (in red, a0ba). "nobsexp" is the number of used observations in CM (a0b2) and the "exp-ref" shows the difference of the number of used observations in CM 110 mJ (a0ak) and CM 80 mJ (a0b2). Black numbers thus indicate that the CM 110 mJ experiment has more observations.

We will now compare the two CM experiments with different laser power. In Figure 13 the statistics of the doppler winds of Aeolus 110 mJ CM (still denoted "a0b2" but now in black) is compared to those of Aeolus 80 mJ CM (denoted "a0ba", in red). We first note that the quantity of the observations has been reduced with approximately 10% in general but up to almost half at 150 hPa in the extra-tropics (see the numbers in the figure and in table 3). Again, the biases of the two experiments are very similar. The background and analysis departures have increased in the 80 mJ scenario, which is associated with the larger error of the observations (table 3). Despite the degradation in both quality and quantity the EDA impact of the CM 80 mJ on the NWP is almost as large as the CM 110 mJ. From Figures 9 to 11 we conclude that the impact in our experiments results from large-scale patches mainly at tropical and subtropical latitudes. Assumingly, these patches are likely to be related to large-scale flow features in the atmosphere. These features are then sampled by numerous Aeolus observations, reducing the actual observational error by the square root of the number of observations. In this manner both the experiments CM 80 mJ and CM 110 mJ would provide large-scale wind observations with relatively small error, which results in comparable large impact for both experiments.

5.3 Sampling strategy experiment

One of the aims with the VHAMP was to investigate how and to best sample the atmosphere to get maximum impact on the forecast quality. As the project continued it was decided that it was more important to focus on the impact of the different laser specifications and on evaluating the EDA method, at the cost of less effort for investigating the sampling strategies. In particular, this meant that only one experiment with different horizontal resolution was conducted. For this experiment, denoted CM 250 km, we use an approximately threefold longer integration length of 250 km (see table 2). The global average forecast impact of this experiment (the dark blue line in figure 7) is almost exactly the same as the one with an integration length of 87 km (the dark red in figure 7), that is the EDA technique suggests that both integration lengths are equally beneficial. Comparing figure Figure 11 and Figure 15 it is clear that there is no substantial differences on the regions of the impact either. Figure 14 shows the statistics of the doppler winds of Aeolus CM 250km (denoted "a0bp", in black) comparing to those of Aeolus CM 80 mJ (still denoted "a0ba", in red). From the reduction of background and analysis departure when going to a longer integration length it is clear that the error of the individual observations has decreased by averaging over a longer interval. However, the observation quantity has been reduced by a factor of two or even three in some regions (see table 3 and the numbers in Figure 14). The EDA results thus indicate that the net result of the decreased quantity and the improved quality is zero. It is somewhat surprising that the information of the small scales that 87 km integration length provides is not more beneficial for the forecast quality. However, it should be pointed out that this lack of benefit may be due to the coarse resolution of our model. The employed model resolution of T399 is equivalent to a grid distance of about 50 km. However the effective model resolution will lie around 8-times this value, ie. 400 km. From this point of view, the employed model does not see different observations, when using approximately 3 observation with 87 km spacing or 1 observation at 250 km. Furthermore, the same argument as above applies concerning the large-scale nature of the impact structures (see also Fig. 15). Again relatively coarse observations can capture these structures. The running model of 2015, when the Aeolus data is to become available, will have better resolution and a shorter horizontal de-correlation length of the B-matrix. It is possible that this will increase the impact of small-scale observations. Thus the current setup may underestimate the impact of small-scale observations.

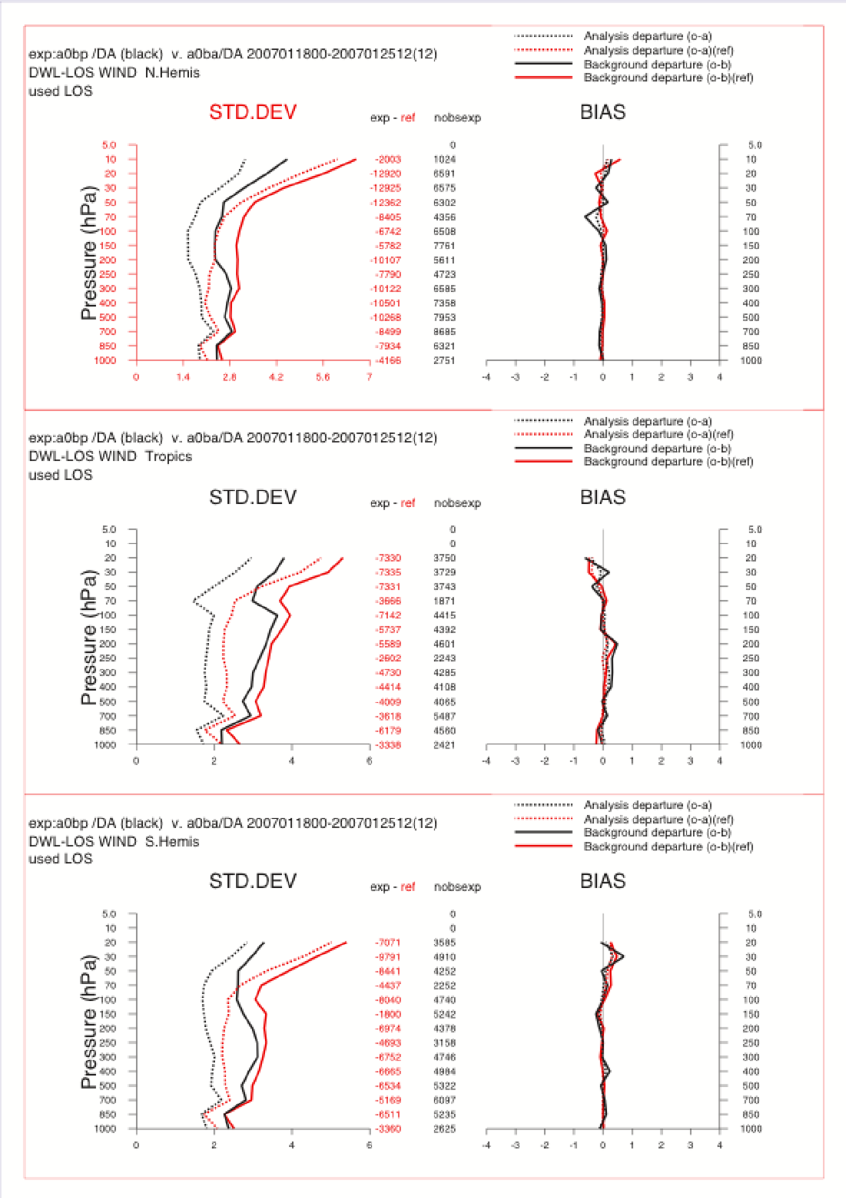


Figure 14: Comparison of the observation departures from the background (solid lines) and the analysis (dashed lines) of the LIPAS Doppler winds for the Aeolus with 250 km and 87 km integration length, in black (a0bp) and read (a0ba), respectively. "nobsexp" is the number of observations the 250 km case and the "exp-ref" is the difference of the number of observations in the 250 km and 87 km cases. Black numbers thus show that the 87km integration length gives more observations.

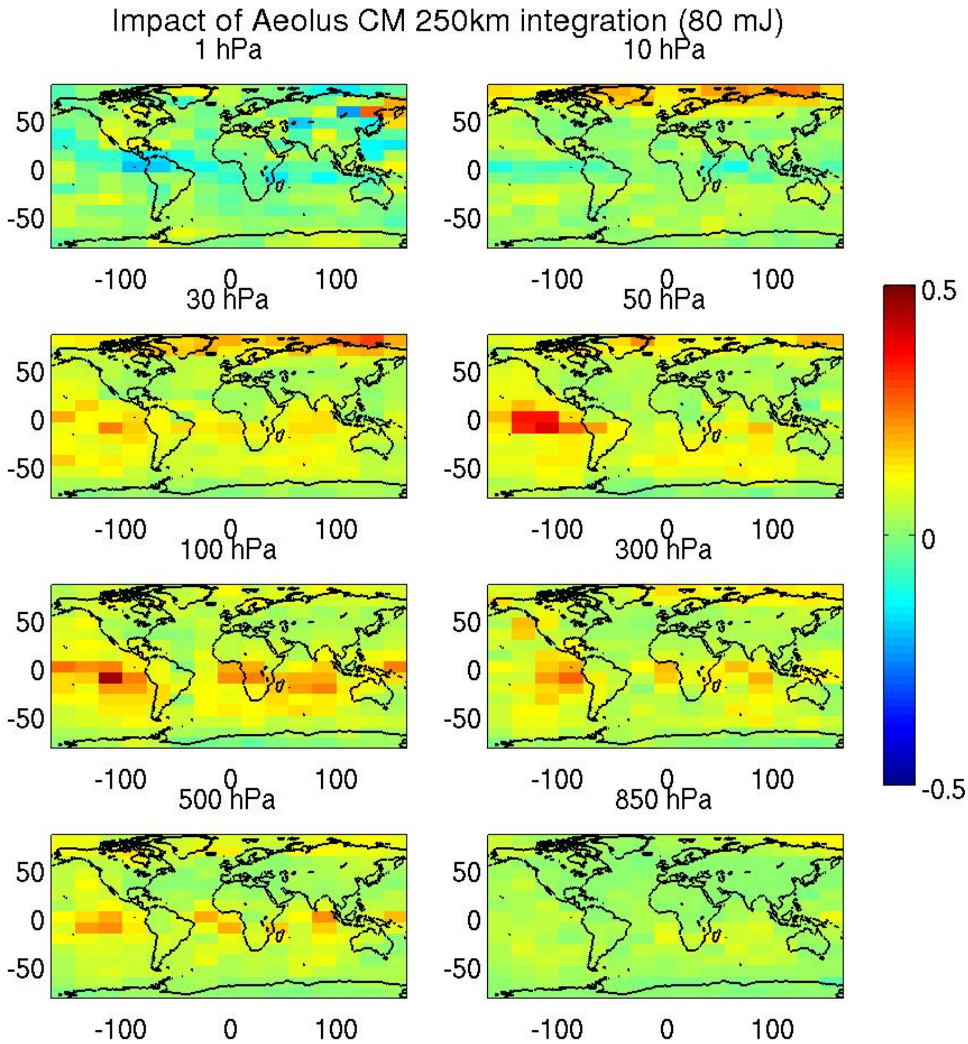


Figure 15: Impact on 12 h zonal wind forecast quality in m/s for the Aeolus CM 80 mJ with 250 km data integration length.

5.4 Time-mismatch experiment

As discussed in sections 1.2 and 5.3 the EDA method has limitations, one being that it cannot observe a potential bias, or systematic error, that is introduced by an observing system. In this case, where

LIPAS data was generated using UKMO winds, such a systematic error is likely to occur, since the state of the UKMO model generally is slightly different from that of the IFS system. Since the real Aeolus data will not be dependent on the UKMO model this, by the EDA method undetectable systematic error is, in this particular case, not interesting for our purpose, in this particular case. However, if we want to compare to other impact estimates, such as the OSE technique, it poses a problem. Recent OSE studies at ECMWF of the impact of LIPAS observations actually showed negative impact on the forecast quality. This could be explained by the systematic error. The observations will all be biased towards the state of the UKMO model. Including them in the data assimilation will therefore increase the magnitude of the departures from the simulated true state of the atmosphere, since the latter state is generated by the IFS model. Nevertheless, despite this explanation to why the OSE method may give an erroneous answer, it is not reassuring that the EDA and the OSE method show such different results, and the question of how the EDA method would respond to erroneous data was posed.

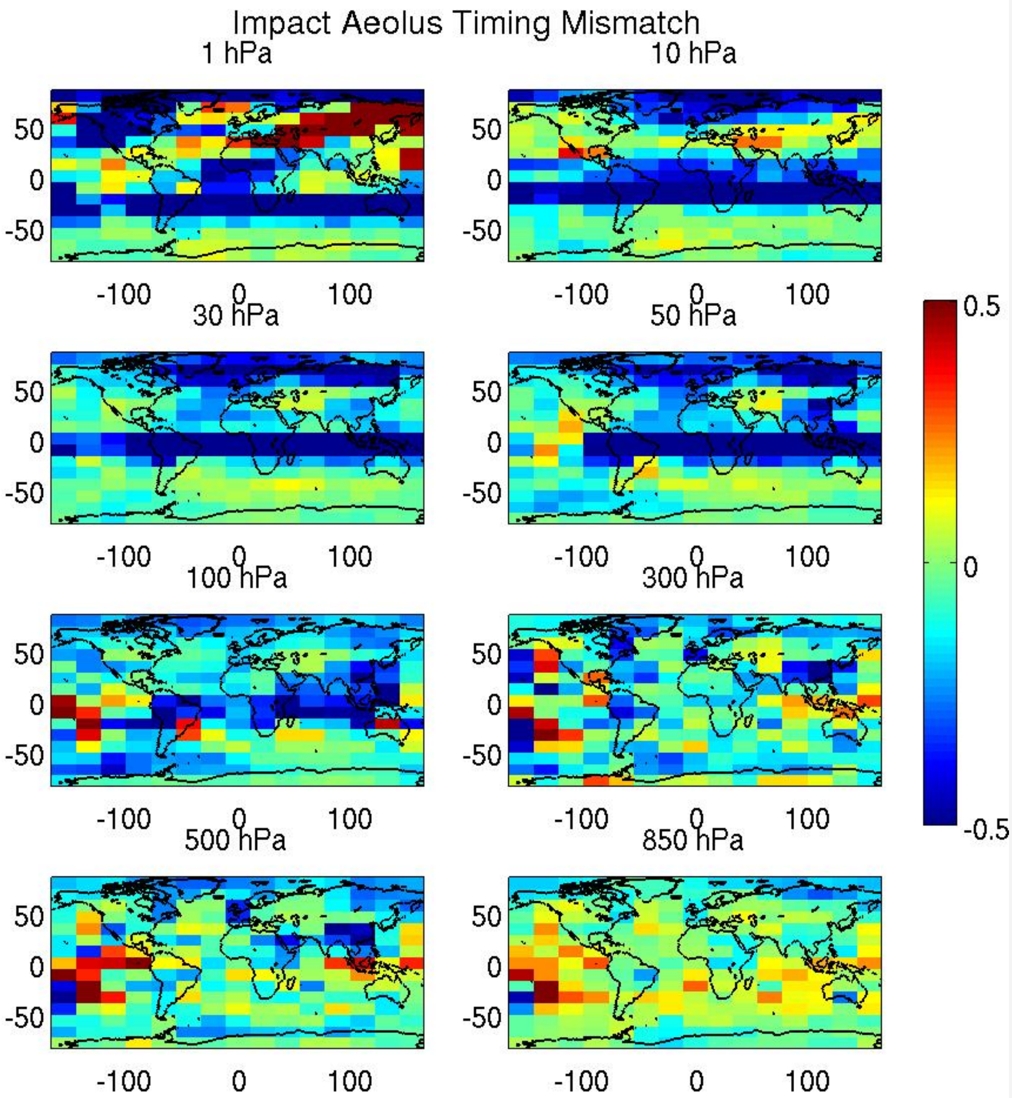


Figure 16: Impact on 12 h zonal wind forecast quality in m/s when assimilated the LIPAS CM 80 mJ winds of January 2007 into the model dates of January 2008.

To investigate this question a time-mismatch experiment was conducted. This experiment, denoted Aeolus 2008, assimilates Aeolus January 2007 winds into the model atmosphere of January 2008. A complicating issue with this experiment is that erroneous data will be rejected by the quality control system of the data assimilation. If the majority of erroneous data is rejected then this will not allow us to study the EDA response erroneous data. Table 4 shows a comparison between the data used in the Aeolus experiment with the correct timing (Aeolus CM 80mJ, a0ba) and the experiment with the timing-mismatch (a0bx). Roughly about 30 % of the data in the Northern hemisphere and 10-20% of the data in the Tropics is removed by the quality control system. In the Southern hemisphere less than 10% of the data has been removed at high altitudes but as much as 40% at lower altitudes. Thus there is substantial rejection of the erroneous winds, as expected. Yet, the number of accepted winds is still high given the time mismatch. The Huber norm used in the cost function minimization procedure reduces the weight of bad (far away from the model background) observations. However, closely spaced observations that are far away from the background, but that are consistent still get weight in the analysis with the Huber norm, thus enabling to draw the model state away from the truth in particular in otherwise data void regions. Despite that a relatively high number of erroneous winds have passed the quality control system, the loss of the rejected data will still cause a smaller EDA spread than if all data had been accepted. Hence the EDA will underestimate of the forecast error and the negative impact. In other word, the quality control system will cause the EDA spread to become saturated once we attempt to assimilate very poor data. The quality control can be regarded as a negative feedback to the spread-generating observations in the EDA. Moreover the rejection rates may be problematic that different ensemble members reject different observations, which would falsify the initial construction of the EDA estimate as described in the introduction, since we cannot guarantee that the perturbations of the used observations is equal to the covariance of the observation error.

hPa	a0ba NH	a0bx NH	a0ba TR	a0bx TR	a0ba SH	a0bx SH
5	0	0	0	0	0	0
10	3027	1992	0	0	0	0
20	19511	13421	11080	9103	10656	10563
30	19500	13752	11064	8731	14701	14352
50	18664	14268	11074	7017	12693	11931
70	12761	9971	5537	4774	6689	5727
100	13250	10206	11557	9189	12780	9734
150	13543	10180	10129	7660	7042	4314
200	15718	10363	10190	8156	11352	6415
250	12513	7208	4845	4076	7851	4195
300	16707	9280	9015	7795	11498	6447
400	17859	10314	8522	7612	11649	6999
500	18221	11918	8074	7385	11856	7696
700	17184	12534	9105	8570	11266	8073
850	14255	9320	10739	9841	11746	7498
1000	6917	4494	5759	5328	5985	3744

Table 4: Data usage in the experiment with correctly timed Aeolus data (a0ba) and in the time-mismatch experiment (a0bx).

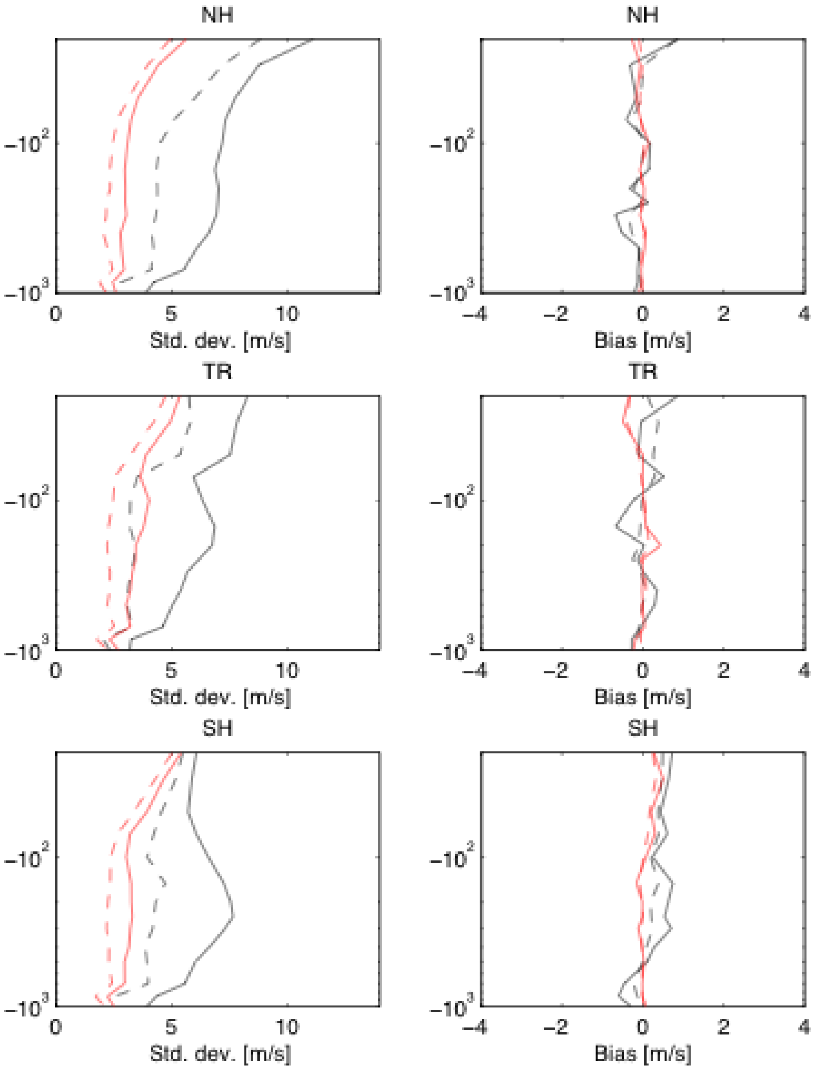


Figure 17: Comparison of the observation departures from the background (solid lines) and the analysis (dashed lines) of the LIPAS Doppler winds for the Aeolus CM 80 mJ winds at the correct timing (in red) and used for the wrong year (in black).

Figure 17 compares the observation statistics of the two experiments. The one with correct timing (a0ba) is shown in red and the time-mismatch (a0bx) in black. The solid lines refer to the background departures (o-b) and the dashed to the analysis departures (o-a). The bias (right-hand panels) of the time-mismatch experiment is larger as can be expected since the atmospheric state of January 2008 is different to that of January 2007 and consistent observations still get weight in the analysis as explained above. The standard deviations of both the analysis and background departures are greatly

increased in the time-mismatch experiment showing the poor quality Aeolus winds used in the assimilation.
The global mean EDA spread of the time-mismatch experiment is shown as the yellow line in Figure 7. The spread has increased substantially throughout the free troposphere and stratosphere, as expected, except below 850 hPa, which is surprising. To better understand the latter, Figure 16 displays the impact at a number of pressure levels for the time-mismatch experiment; red/blue-colored means EDA spread reduction/increase, i.e., positive/negative observation impact. There is noisy positive/negative impact at 500 and 850 hPa. Although, the mean impact at 850 hPa is positive, the noisy appearance does not put much confidence in the added value of this observation set. The noisiness of the impact might result from an undersampling of the large-scale flow features, which normally lead to positive impact. By chance, the time-mismatch experiment creates large-scale positive and negative impact. At higher altitudes the negative impact becomes more consistent and prominent. On average, negative impacts are found at all altitudes above 500 hPa with large positive/negative impacts locally. The other experiments, that is the different Aeolus scenarios and the radiosonde denial experiment (Figure 9 to Figure 11 and Figure 15) all showed consistent positive to neutral impact. The increase in EDA spread is relatively mild (approximately twice that of the radio sonde denial experiment) considering that the observations should be heftily erroneous. This may be explained by the, for the EDA method, invisible bias or systematic error, when the entire ensemble follows the bad observations. The increase of ensemble spread may then be limited, but the bias (the model states draws away from the true atmospheric state) will increase.
In summary, it appears that the EDA method has handled the erroneous data as well as can be expected, given the a) quality control system that will cause the EDA spread to saturate when the observations become too unrealistic and b) that the bias that several observations, all are consistent with an erroneous state of the atmosphere, will introduced, is unseen.

6 Discussion and Conclusions

We have used the EDA method to investigate the impact of the Aeolus data on NWP quality for different laser configurations; the original Aeolus BM (Aeolus BM 110 mJ), the corresponding CM configuration (Aeolus CM 110 mJ) and the CM configuration with the laser power reduced to 80 mJ (Aeolus CM 80 mJ). The method shows that the impact on the forecast quality, regardless of laser configuration, is of the same order as the impact of the entire fleet of radiosondes and wind profilers. The differences in the impact of the observations on NWP between the various laser operational settings, both on a global average and on a regional basis, are minor. For all settings improved forecast quality seen in the tropics and the ocean regions in the troposphere and the lower stratosphere. Forecast improvement is also indicated in the winter polar region at higher altitudes. The fact that the EDA indicates a substantially larger impact of Aeolus data in the tropics than in the extra-tropics, may be connected to that our experimental setup, with a relatively coarse model resolution, better observes the impact on the large scales; large-scale wind in the extra-tropics is determined by geostrophic balance. Thus, it is possible that the Aeolus impact in the extra-tropics will be larger in a model with higher resolution, such as the running IFS model version when the Aeolus data becomes available. Moreover, the localized positive Aeolus impact in the tropics indicates that these are sensitive regions where additional observations are beneficial for an improved description of the flow.
The impact reduction from CM 110 mJ to CM 80 mJ is surprisingly small, given that the data in the latter scenario is degraded both in quality (13 % larger errors) and in quantity (approximately 10% less observations). This subtle change in impact might result from the good impact on large-scales which is still very accurately sampled for both laser configurations, at least relative to our model resolution. However, this lack of benefit may be due to the coarse resolution of our model. The running model of 2015, when the Aeolus data is to become available, will have better resolution, and thus it is possible that system will benefit substantially more from the small-scale observations.
The EDA method is a relatively new method for impact assessment, and it has never been fully evaluated. The first part of this project was therefore aimed at conducting an evaluation of the EDA

method, with focus on identified issues that may jeopardize the trustworthiness of EDA estimates. One such issue was the assumed linearity between the magnitude of the perturbations added to the observations and the resulting EDA spread. This assumption is part of the theory on which the EDA method is founded, and is thus crucial for EDA impact assessment. The linearity assumption was tested by conducting two experiments with reduced magnitude of the perturbations and assessing if the resulting EDA spread was reduced in a linear fashion. It was shown that there is indeed a linear relationship and that it is valid throughout the domain. Another issue that was investigated was if the added perturbations would cause the quality control system to reject the perturbed observations. This would cause the EDA spread not to increase as expected with an underestimate of the EDA forecast error as a result. It was shown that almost all observation systems had not been rejected to any larger extent, thus the vast majority of the perturbed observations had passed the quality control system. The exceptions were the atmospheric motion winds and the limb observations, of which 10% and 5% of the perturbed observations had been rejected. Since these two observing systems do not constitute a significant part of the total observations, this issue is not of great concern as long as the observations that one is assessing the impact of are not too unrealistic. However, this issue could play a significant role if one is assessing impact of unrealistic observations, as we shall discuss soon.

During the project other short-comings of the EDA method has been brought into light. In particular the EDA methods inability to observe a bias of an observing system may be problematic. The LIPAS data set indeed is contaminated by such a bias, in that it is generated by the UKMO model, which state will differ slightly from that of the IFS model. This is likely the explanation to the negative impact of the LIPAS data set that was observed by OSE analysis in an independent study at ECMWF (personal communication). Since we are not interested in this bias of the UKMO model, it does, in this particular case, not cause a problem, and is even advantageous. However, it is obviously a short-coming of the EDA method that such a bias cannot be estimated, and it should be noted that a possible unknown bias of the Aeolus data, which would be highly detrimental for the NWP forecast quality, can not be determined in this way. Another short-coming of the EDA method is that the rejection of data by the quality control system that will cause the EDA spread to saturate when the observations become unrealistic, that is to far from the model state. However, if this happens to as large of an extent in the baseline experiment as in the Aeolus experiment, then the spread of the two experiments will be reduced by the same rate, or factor. The impact estimate would thus be decreased by this factor. However, since the impact has been scaled to yield the same magnitude as the OSE estimate, such a factor has already been implicitly handled. This means that the scaled impact estimate will not be affected by the saturation of the ensemble, as long as the saturation effect is moderate and the perturbed observations are rejected to the same degree in the two experiments we are comparing. Table 5 shows that the rejection rate of the baseline and the Aeolus 80mJ is roughly the same. Hence, the saturation issue is not a great concern for the Aeolus impact assessment.

	Baseline experiment	Aeolus CM 80 mJ
Land, Ship	0.055	0.064
Aircraft	0.024	0.023
Atmos. motion winds	8.8	8.4
Buoys	0.093	0.17
Radiosondes	0.21	0.30
Balloons, Profilers	0.085	0.21
Satellite soundings	2.6	2.8
Scatterometer	5.5	5.5
Limb observations	0.006	0.025

Table 5: Percentage of extra loss of different types of observations for perturbed EDA member compared to the unperturbed member for the 31st of January.

In summary the EDA spread reduction is a valuable metric for observation impact assessment, but a conclusive impact assessment should also include observation rejection analysis and an estimation of the bias of the observing system. Since the rejection rate of the Aeolus experiments and the reference experiment were found to be moderate and of similar magnitude the rejection caused by the quality control system does not seem to be of great concern for the Aeolus impact assessment. The bias of the Aeolus system, should such exist, has not been, and cannot be, assessed with the EDA technique. The coarse resolution of our data assimilation system, as compared to the system that will be running when Aeolus data becomes available, further means that it is limited to assess observation impact on the large atmospheric spatial scales. This is a strong limitation, since Aeolus winds are expected to be most beneficial for NWP on the sub-synoptic scales. Thus it is possible that we have underestimated the impact of Aeolus in the regions where meso-scales are more important, such as the midlatitudes.