We thank the reviewer for his/her extensive review and very useful comments to improve the paper. We did a major revision as recommended, reducing the paper size by about 25%. A detailed point by point reply (in blue) is provided hereafter.

Major comments

comment 1

Although the manuscript claims to assess the Aeolus measurement performance in heterogeneous conditions, a major portion of the manuscript (ch. 3 and 4) is dedicated to a rather lengthy description of the construction of backscatter profiles from relative humidity profiles from radiosonde data, which was published before by the author in Zhang et al. (2010). The manuscript should focus much more on the implications for Aeolus as stated in the title of the paper. Thus I would recommend to drastically shortening ch. 3 and 4, e.g.

- Ch. 3.1 describes different methods of cloud layer detection in radiosonde relative humidity data, which is mainly presented in Zhang et al. 2010, and does not need to be repeated here. The assumptions for applying it to high-resolution data could be discussed in 1- 2 paragraphs. I would recommend removing Fig.2, because a comparison of relative humidity RH from radiosonde with RH from ECMWF does not provide any insights for the topic of the paper.

We agree. Section 3.1 has been shortened substantially by referring to previous papers to avoid repetition. For completeness we describe the Zhang2010 method with changes for application to the different climate zone of De Bilt Fig. 2 has been removed.

- Ch. 3.2 reviews the well-known relationship between relative humidity and lidar backscatter coefficient with numerous references. For the purpose of this paper (effect on Aeolus performance) this should be shortened to the applied method and basic assumptions for the used parameters and limited to the relevant references.

We agree. Section 3.2 has been shortened substantially as well. The main equations (24)-(28) remain with references to the used parameters in these equations.

- Ch. 4.1.1, 4.1.2, 4.1.3: I do not understand the purpose to compare both the WR95 and Zhang2010 method for a limited dataset. To my opinion it would be sufficient to show results from Zhang2010 method for the purpose of this paper. Obviously no significant difference between the two methods exists and the method is only applied to a very limited dataset (1 site, 1 profile per day).

We agree with the reviewer's comment. Although we made a thorough comparison between the WR95 and Zhang2010 method, we agree that the superiority of the latter was already demonstrated in Zhang et al. (2010). We remove the WR95 results from the figures (previous Fig. 3, Fig. 5, and Fig. 6) and text and instead mention (at the end of the Sect. 4.1) that:

Finally we mention that a thorough intercomparison between the Zhang2010 and the method from Wang and Rossow (1995) for the De Bilt radiosonde, confirmed earlier conclusions from Zhang et al. (2010) of better results obtained from their method (not shown)

- Ch. 4.2.1: It is not valid to apply the growth factor to backscatter coefficient and lidar ratio in case of clouds, as presented in Fig. 9. This scaling is only applicable for the aerosol layers. Thus only Figure 9 (lower right) does make sense, but shows also the weakness of the method by using climatological values for clouds with sharp gradients at the borders.

We have assumed the climatological aerosol profile throughout the troposphere, also in cloudy conditions, i.e., inside atmospheric scenes with clouds we assume a mixture of aerosols and cloud droplets. Since aerosols act as condensation nuclei, we have overestimated the aerosol density in clouds in Fig. 9. This is corrected in the updated figure (now Fig. 13). Aerosol backscatter in the lower left panel is set to zero for cloud presence. This correction had no impact on the conclusions since backscatter from clouds is orders of magnitude larger than from aerosols.

We disagree with the reviewer that sharp boundaries at the cloud borders, by using climatological values, are a weakness of method. Backscatter gradients are large at cloud top and bottom (see for instance scenes from CALIPSO). Maybe the reviewer is confused and mixes backscatter and extinction with total attenuated backscatter and extinction. The former are shown in Fig. 9 (now Fig. 13), not the latter. Indeed attenuated backscatter is now added to the bottom right panel of Fig. 13; the cyan line represents attenuated total backscatter coefficient ($m^{-1}sr^{-1}$) from molecule, aerosol and cloud.

comment 2

The authors claim that high variability is introduced in the atmospheric database by high-resolution radiosonde data, but in fact they are using mainly climatological input for cloud and aerosol backscatter, which is slightly modified by radiosonde RH profiles. I see two major weaknesses in their methodology:

- The high-resolution radiosonde data is only used for cloud-layer detection, and then climatological values for cloud-backscatter and extinction are used from Vaughan (2002). This introduces artificially sharp gradients of backscatter and extinction at the cloud boundaries, and no variability of backscatter/extinction within the cloud. The variability of backscatter/extinction within clouds is underestimated with this method. This drawback and its consequences are not discussed in the manuscript. What is the advantage of using this dataset wrt vertical cloud variability instead of the database constructed by CALIPSO observations as presented by Marseille et al. (2011)?

High resolution radiosonde data is used for both **cloud-layer detection** and **wind**.

Above we discussed the sharp gradients of cloud backscatter and extinction also observed in Fig. 4 of Marseille et al. (2011). We agree that the backscatter/extinction variability within clouds is underestimated. Fig. 4 of Marseille et al. (2011) clearly shows backscatter variations inside clouds.

However, quantifying backscatter/extinction variability inside water clouds has limited value for Aeolus assessment; the UV lidar cannot penetrate water clouds and only observes the cloud top.

For ice clouds (e.g. cirrus) the situation is different indeed. Fig. 4 of Marseille et al. (2011) shows large variability in cirrus clouds. However, this variability is different for different clouds. But co-located radiosonde and CALIPSO data are rare as mentioned in section 4.1.2, so CALIPSO data are generally not available for simulating backscatter/extinction variability in ice clouds.

Any alternative choice is then artificial. For instance, Table 1 of Marseille et al. (2011) gives statistics of backscatter/extinction for tropical cirrus, but no evidence that the same numbers apply for mid-latitude cirrus and no cirrus backscatter realizations.

Advanced cirrus simulation following Hogan and Kew (2005) is outside the scope of this paper.

In conclusion, any choice of introducing backscatter/extinction variability inside clouds is artificial and we decided to take the simplest choice of constant backscatter, indeed underestimating variability. Note however that the attenuated backscatter inside the cloud is non-constant, such that wind-shear over the bin may still cause substantial wind error.

- Similarly the method of constructing aerosol backscatter profiles is based on a climatological profile, which is modified by the relative humidity profile of the radiosonde (eq. 25). I doubt that eq. 25 can be used for that purpose, because the hygroscopic growth factor is based on a concept of modifying a backscatter profile with a reference RH (e.g. eq 22), but eq. 25 uses a climatological profile, which was obtained not for a reference RH but includes a whole range of RH values. Also the resulting backscatter profiles in Fig. 10 show much less variability (low to high quartile) than the ones measured by the UV lidar. So I do not understand, how the authors could conclude that the aerosol backscatter profiles are representative (p. 21).

The motivation above for clouds also applies for aerosol backscatter profile simulation. Generally, no CALIPSO data are available. Any choice is then artificial, but to motivate our choice, we validated the climatological profile, corrected based on RH, with real UV lidar backscatter data. Most important for Aeolus wind quality is the aerosol backscatter variability inside Aeolus bins, rather than the backscatter magnitude. Fig 11 (now Fig. 12) is thus of more relevance in that respect than Fig. 10 (now Fig. 11), the latter provided for completeness.

The red curves in both panels of Fig. 11 (now Fig. 12) are not too different. So, the reviewer is right that the aerosol backscatter profiles are less representative for Aeolus wind assessment. But as stated clearly in the text: "*the aerosol backscatter VARIABILITY simulated from radiosonde observations is representative for real atmospheric scenes ....*"

Because of these 2 limitations, I do not consider the proposed method as being better appropriate for simulating atmospheric variability rather than the method already discussed in Marseille et al. (2011), especially for the variability in the aerosol/cloud backscatter and extinction.

As said above, the Marseille et al. (2011) method is not applicable to radiosondes because of lacking co-located radiosonde-CALIPSO data. The main drawback of Marseille et al. (2011) is a strong underestimation of wind variability by the ECMWF model. The high-resolution radiosonde data solves this issue. Marseille et al. (2011) indeed more realistically simulates backscatter/extinction, although the variability within Aeolus bins is well simulated by both Marseille et al. (2011) and the underlying study.

comment 3

Although some validation of the atmospheric database is presented in chapter 4, statistics of the atmospheric database (as in Marseille et al. 2011) for median, quartiles the profiles and their gradients are not presented in the manuscript. So it is difficult to assess the content of chapter 5 (influence on Aeolus performance), if the content of the database is not well described (e.g. histograms, percentile profiles). In addition the authors should investigate and discuss in more detail the thickness of the cloud layers in comparison to the range-gate dimensions, which was shown of relevance in chapter 2. Does the database contain clouds, which are thinner than the range gate dimensions? What is the thickness of the radiosonde detected clouds in comparison to the Cloudsat/CALIPSO detected clouds (e.g. as presented in Fig. 5 for cloud base/top). Is the cloud thickness distribution such, that a cloud is contained in several range gates, and only the uppermost range gate is influenced by a cloud, which is not fully covering the range gate?

We agree with the reviewer that database statistics provides additional information to the Aeolus mission. We add some about the vertical distribution of cloud layers as found from radiosondes in section 4.1.

comment 4

Within Ch. 2 the analyses of the influence of the molecular backscatter and extinction profile in combination with a vertical gradient in wind speed is not studied as a separate case, which is a serious limitation of the manuscript. This will lead already without any clouds to a systematic (because of the non-random profile) error in the Rayleigh wind. Equations should be presented in ch. 2 and also a possible correction of this error could be discussed, e.g. by a modification of the height assignment of the Rayleigh winds. This error in Rayleigh wind or height assignment is also obvious in Fig. 14 for altitudes above 10 km were no clouds are present in the database (as discussed ch. 5 on p. 1427). I would strongly suggest including a theoretical analysis in ch. 2, deriving a correction for that, and discussing the residual error after this correction in ch. 4.

This is a very good point. Section 2 has been extended with a theoretical analysis of the impact of the non-constant attenuated molecular backscatter inside Aeolus bins and the impact on Aeolus wind quality in aerosol/cloud sparse atmospheric regions (most of the atmosphere!) and possible correction schemes. A new section 2.1 was introduced with very useful results for improving the Aeolus level-2 processor.

comment 5

The author claim at several places of the manuscript, that the variability of the backscatter profiles leads to systematic biases, which are larger than the systematic error requirement for the Aeolus mission of 0.4 m/s. I do not agree with the authors that these errors can be considered as systematic errors, and thus can be compared to the systematic error requirement for the Aeolus mission. I consider an error as systematic, if it shows up over a significant portion of a dataset (e.g. time series over a portion of the orbit) as an under- or overestimation of the true wind speed. A prominent example would be the systematic error of AMV´s due to a systematically wrong height assignment. But this is different for a lidar, because the variability of the atmosphere within one range gate can be randomly distributed (as the authors assume correctly in ch. 2.1, 2.2.) ; the thickness of the backscatter layer, the location of the backscatter layer and the wind speed gradient could be randomly distributed within one range gate. Thus this could lead to a random over- or underestimation of the wind speed for several observations. Thus in my view the investigated error leads to an increase in the random error due to atmospheric variability. In addition I recommend mentioning the view, that the errors could be considered also as height assignment error rather than a wind error, as introduced by the weighting function in eq. 1

We disagree with the reviewer when considering clouds that will cause systematic errors in our view.

It is true that the location of a cloud layer can be randomly distributed when considering different cloud scenes. But for a particular cloud scene the cloud layer may be well fixed and may extend over long distances (many Aeolus observations). For instance low-level stratus clouds with approximate constant cloud top height give a constant wind error (thus systematic) in bins covering the cloud top and that my extend over long distances (thousands of kilometers). Such errors are known to be detrimental for NWP.

Also (tropical) cirrus cloud fields may extend over long distances (see CALIPSO), giving rise to similar systematic errors as described above, but depending on the backscatter variability and wind-shear over the cloud.

Minor comments

Ch. 1:, p. 1397: The statement about the ECMWF effective model resolution and the provided numbers of 15-20 is quite strong and controversy, compared to the factor of 7 given by Skamarock (2004). It should be discussed, how the effective model resolution is obtained/defined leading to a factor of 15-20 or the sentence should be re-formulated.

Skamarock (2004) considers the meso-scale non-hydrostatic WRF model only. The situation is different for global model. Vogelzang (2012) and an ESA Technical Note to motivate the number of 15-20 valid for ECMWF. These references are added in the text.

Ch. 2: It would be nice to provide a figure with a schematic sketch of both situations analysed in ch. 2.1 and 2.2 with all the notations, e.g. z_l, z_0, z_c, l, and assumed profiles for backscatter and wind profile.

Instead of a figure, we added a table with typical values for aerosol and cloud layers.

Ch. 2: It should be noted that equation 2 and the subsequent analyses includes the simplifying assumption that the backscatter signal can be separated by the instrument perfectly in a particle component and a molecular component. The instrument senses the total backscatter, which consists of β_p+ β_m, but eq. (2) and the analysis in ch. 2 assumes that the Mie channel is only containing signal from β_p (with no signal from β_m) and the Rayleigh channel contains signal only from β_m (with no signal from β_p). This is a simplified assumption wrt the behaviour of the real Aeolus instrument (see e.g. discussion in Dabas et al. 2008, Tellus).

We agree that the theoretical analysis in section 2 assumes perfect separation of the Mie and Rayleigh signal. The level-2A processor (discussed by Dabas et al., 2008) is developed in particular to do so, but it remains to be seen how this processor will perform when dealing with real data. Also, we consider algorithms to separate Mie and Rayleigh signals as part of the L2Bp optical properties code, under development.

Ch. 2.1, p. 1400: The definition of $U_p^M=u_0+\alpha z_c$, requires that z_c is provided in coordinates such that $z_c=z_0$ at the bottom of the range gate with wind speed u_0. z_c is not the absolute height (AGL).

The reviewer is right that z_c is relative to the bin bottom at z_0. This is confusing in the text and corrected.

Ch. 2.1., eq. (9): The subscript “m” is missing in the second term of ε.

That's right.

Ch. 2.1., p. 1402, discussion of Eq. (10): It is stated that the Rayleigh channel error grows quadratically with the cloud layer transmission τ_c, although eq. (10) contains τc2 in the enumerator and denominator. This statement is only valid for small values of τ_c^2.

Indeed, the statement is true for small τ_c^2 only. The text has been updated.

Ch. 2.1, p.1402: A literature reference should be given for the used value for the vertical wind shear of 0.004 s^{-1}.

Reference to Hakansson for 0.004 s^{-1} shear have been included. The reference is as follows:

Håkansson, M.: Winds, shear and turbulence in atmospheric observations and models, Doctoral thesis, Department of Meteorology, Stockholm University, ISBN 91-7265-497-X, 2002.

Ch. 3, p. 1406: It is only stated that the used radiosonde is of high-quality, whatever this means. Instead numbers for accuracy and precision should be given for the used parameters.

High quality means within the requirements defined by WMO. The following table is the result of the Vaisala Radiosonde RS92's performance as measured in the WMO intercomparison in 2010 (Yangjiang, China).

Vaisala	Score	Accuracy limit for score*
Temperature, Night, Height < 16 km	5	0.3 °C
Temperature, Night, Height > 16 km	5	0.6 °C
Temperature, Day, Height < 16 km	5	0.3 °C
Temperature, Day, Height > 16 km	5	0.6 °C
Protection for Evap. cooling errors	Yes	
Humidity, T > -40 °C, Night	5	3 %RH
Humidity, T > -40 °C, Day	5	3 %RH
Humidity, T < -40 °C, Night	5	5 %RH
Humidity, T < -40 °C, Day	4.5	5 ... 10 %RH
Height, P > 100 hPa	5	10 m
Height, P < 100 hPa	5	20 m
Pressure, P > 100 hPa	5	1 ... 0.3 hPa
Pressure, P < 100 hPa	5	0.1 ... 0.04 hPa
Wind, troposphere	5	0.5 m/s
Wind, stratosphere	5	0.5 m/s
2 years in operation	Yes	
Used in	global	

The reference is:
http://www.vaisala.com/Vaisala%20Documents/White%20Papers/WEA-MET-WMO-Test-White_Paper-B211129EN-D-LOW.pdf

Ch.3, p. 1406: It is stated that only 12 UTC radiosondes were used. What was the reason for not using the 0 UTC radiosondes, which could have extended the dataset also to night-time observations?

00 UTC radiosondes were not available from the database.

Ch. 3: The use of the symbol RH (for Relative Humidity) in this chapter is at some places misleading (e.g. eq. 23), because actually the value of the scattering cross section σ(RH) or the backscatter coefficient β(RH) is meant.

The use of symbol RH has been checked.

Ch. 3.2: It is not mentioned, which climatological profile (median or other profile) was used from Vaughan et al. 1998.

We used the median profile from Vaughan (1998). This has been added to the text.

Ch. 3.2: It is not mentioned in the text, which values are used for backscatter and extinction for a water cloud from Table 3, which contains several types. Is there another distinction using the altitude of the detected cloud?

We select the backscatter and extinction for a water cloud (distinguished by temperature from ice cloud) by using the altitude of the detected cloud.

Ch. 4.2: The type of the UV lidar should be specified in the text - commercial?

The lidar system is indeed a commercial system. Back in 2007, it was a ALS-300 system manufactured by Leosphere. Added in the text, now.

Ch. 4.2.2, Fig. 10, 11: As stated in the text, the UV lidar observations are influenced by the telescope overlap below 300 m. From Fig. 10 it seems that the backscatter profiles are even influenced above 500 m from the overlap. As no valid backscatter profiles can be obtained below 300/500 m, I would remove those altitudes from Fig. 10/11 (right) for the UV lidar.

Based on the fitting of an overlap model in the case of low aerosol loading conditions it was determined that the lidar overlap was more than 95% complete at 300 meters. An empirical overlap correction (Guerrero-Rascado et al. 2010) has been applied to the data to extend the range of valid data to 100 meters. Below 100m the scattering ratio (R) was fixed to a constant value. To avoid problems with this reference, we only plot the data above 100 meters.

The reference: Guerrero-Rascado, J. L., Costa, M. J., Bortoli, D., Silva, A. M, Lyamani, H., Alados-Arboledas, L.: Infrared lidar overlap function: an experimental determination. Optics Express, 18, 20350-20359, 2010.

Ch. 6, Conclusion and Ch. 4.1.3: I consider the statement in the conclusion about comparison of vertical profile of cloud coverage from radiosonde database compared to ECMWF rather strong, because the radiosonde database is limited to 1 site (de Bilt) and 1 time (12 UTC. Such a general statement about ECMWF model cloud coverage as presented in the conclusion is not supported by this study.

The statement on 2007 ECMWF model cloud bias was found earlier by Houchi (PhD thesis) on a substantial larger dataset and similar conclusions were found in this study.

Ch. 6, p. 24:It is argued that the atmospheric database shows factor 4 higher backscatter than the UV lidar below 700 m. Due to the overlap issue and the presented profiles in Fig. 10 (right), I would not consider the UV lidar measurements reliable below 500 m (or even 700 m).

That is right, we use an overlap model from Guerrero-Rascado et al. (2010) to solve this problem, and we have modified the two figures (Fig. 10 (Fig. 11 now) and Fig. 11 (Fig.12 now)). We also removed the part of profiles below 100 m in Fig. 11.