Authors response to Anonymous Referee #1 (Received and published: 18 July 2018)

The paper by Zentek et al. describes the use of a scanning lidar for ship-borne wind measurements without a motion stabilisation platform. The authors used data collected with an external Altitude Heading Reference System to correct for the ship's pitch and roll after the measurement campaign. The presented technique and the statistical comparison of the lidar wind measurements to radio soundings as well as to ship measurements is important for the scientific community due to the clear need for wind measurements over the oceans – especially in the polar regions. Such measurements are important for a better understanding of atmospheric processes in the maritime environment. The paper is suitable for publication in AMT and can be published after minor revision.

Major comment:

- The lidar measurements that have been corrected for the ship's pitch and roll after the measurements are performed consist of profiles that are the average of 12 to 15 seconds of individual rays for the PS96 campaign and 1.5 seconds for the PS85 campaign. The movement of the ship during these averaging periods introduces horizontal wind components into the vertical wind. This is an important source of error and should be discussed in the paper. How does the proposed methodology account for movements during the time needed to obtain the averaged profiles that are later motion corrected?
We address this issue in section 2.3.1 "Ship motion correction" (page 4, line 23ff & Fig.3) but never stated clearly what we address and just refer to it as "the error" (page 4, line 25). We corrected this
Before:

> During PS96 the averaging time of a single ray was typically 12–15 seconds, so that we corrected each single measurement with the mean value over the averaging time. All measurements that have a standard deviation of roll or pitch angle larger than 0.5° or yaw angle larger than 2° over this averaging time were excluded from the analysis in order to reduce the error.
> [… other text …]
> For a data point in 1 km distance from the lidar a change of elevation from 75° to 75.5° (25° to 25.5°) causes a difference in height of 2 m (8 m) and a horizontal wind speed error of less than 3.3% (0.4%). This is acceptable as we will later interpolate over height intervals of 50 m.

Now:

> During PS96 the averaging time of a single ray was typically 12–15 seconds, so that we corrected each single measurement with the mean value over the averaging time. <u>This introduces an error whenever the ship's angle and thus the lidar angle changes during this averaging time. In order to reduce the error,</u> all measurements that have a standard deviation of roll or pitch angle larger than 0.5° or yaw angle larger than 2° over this averaging time were excluded from the analysis. <u>Correcting the direction of the lidar measurement by the mean roll and pitch angle during the averaging time should already cause most of the error to average out, as it measures partly too much and partly too less wind speed. But even if this is not the case, for</u> a data point in 1 km distance from the lidar a change of elevation from 75° to 75.5° (25° to 25.5°) causes a difference in height of 2 m (8 m) and <u>the resulting</u> horizontal wind speed error <u>is</u> less than 3.3% (0.4%). This is acceptable as we will later interpolate over height intervals of 50 m <u>and only evaluate the horizontal wind in our paper</u>.

Other comments:

- Page 5, line 12: Can you really assume horizontal homogeneous wind fields? The elevation changes during the scan.
We assume the homogeneity for a fixed height and only take data points from that height (page 5, line 10f). A different elevation as such is no problem. For example the wind velocity could still be computed if there are 3 different measurements with elevations of 50, 60 and 70° and azimuths of 20, 30, and 40°
The error resulting from the change of elevation (due to the ships movement) during each single ray was discussed in section 2.3.1.

- Page 6 line 30: Doppler velocity due to horizontal wind speed is less than 26 % at this elevation. Is that still true if you correct pitch and roll after the measurements were taken? Your elevation is not stable at 75° due to the ship's motion.
Yes, a Doppler velocity of 10 m s^-1 for an elevation of 73/75/77° would result in 34/39/44 m s^-1 horizontal wind speed. These values can be safely considered to be unrealistic for our conditions.

- Page 7 line 9-10: What are the reasons for the different SNR thresholds for the two campaigns? Could it be the different averaging times of the rays? The elevation is not stable during the measurements and you get different horizontal wind components into your vertical wind component. With a longer averaging time the effect might be enhanced.
The reviewer is partially correct. As stated on page 6 line 15, the value for a SNR threshold can vary depending on the instrument specific performance (detector noise) and the variability of atmospheric conditions within the measured volume. We think that the main reason is not the influence of elevation but the averaging time itself. We changed the passage explaining this.
Before:

> "Additionally due to the different averaging time for each ray during PS85 and PS96 (1.5 vs 12–15 sec), the PS96 data contains less noise and thus it makes sense to choose a different SNR threshold for each data set." (page 6, line 25-27)

After:

> "Additionally due to the longer averaging time for each ray during PS96 (12–15 sec) than during PS85 (1.5 sec), the PS96 data allow for a lower SNR threshold compared to the PS85 data, because averaging over a longer period given the same SNR results in better data. Thus, it makes sense to choose a less strict SNR threshold for the PS96 data set to make both data sets more comparable."

- Table 2 and Table 3: Similar to previous comment the statistics for the PS85 campaign with a shorter averaging time are better than for PS96 with a longer averaging time. What is the reason for this?
We think our data sets are too small to reach any definite conclusion or even to be reasonably certain that one data set / scanning technique is really better than the other. We think it is possible, that external sources like a bias in weather condition during one cruise could be reason enough to cause this differences. One reason to include Table 3 in this paper was to show that a different reasonable analysis-configuration would lead to different statistics. For example the computed RMSD for PS96 can change from 0.9 to 0.7 m s^-1 making it the same as for PS85.

- Page 8 line 19/20: Could the higher bias be explained by not having a horizontally homogeneous wind field? You only correct for the elevation and azimuth but you cannot correct for the horizontal wind component being present in the vertical wind component.
We see no reason why this would lead to a positive bias in wind direction for both cruises and for both comparisons (radio sounding and anemometer).

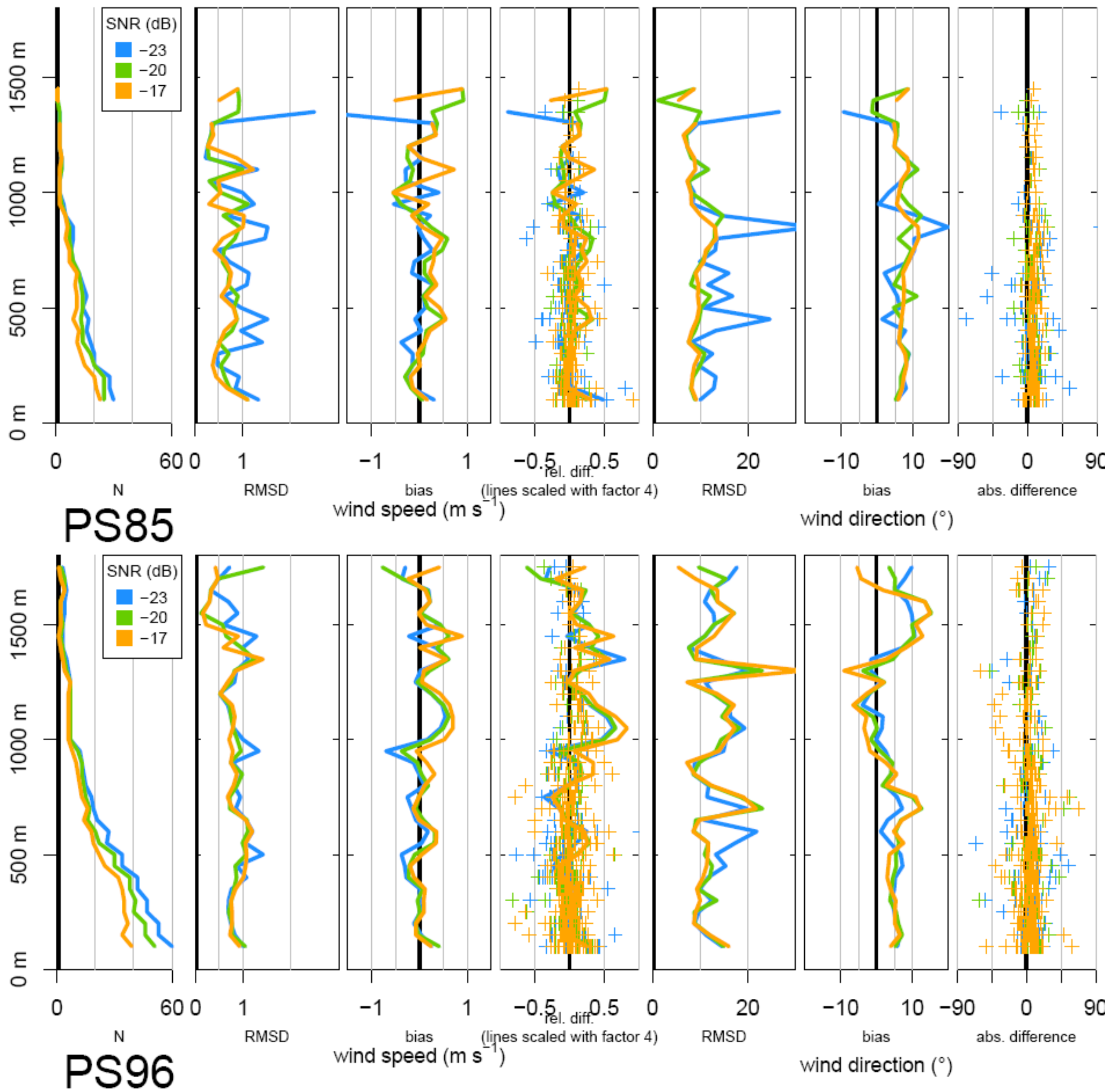
- Figure 6: please add a plot for the relative difference between lidar and radio soundings by height for wind speed and wind direction.

We do not know of any definition for relative differences of wind directions. (Just labeling the axis differently by dividing by 180°?)

We plotted the relative difference for wind speed and absolute difference for wind direction with the symbol “+” for each single case. We also plotted the mean relative difference for wind speed as lines scaled with a factor of 4.

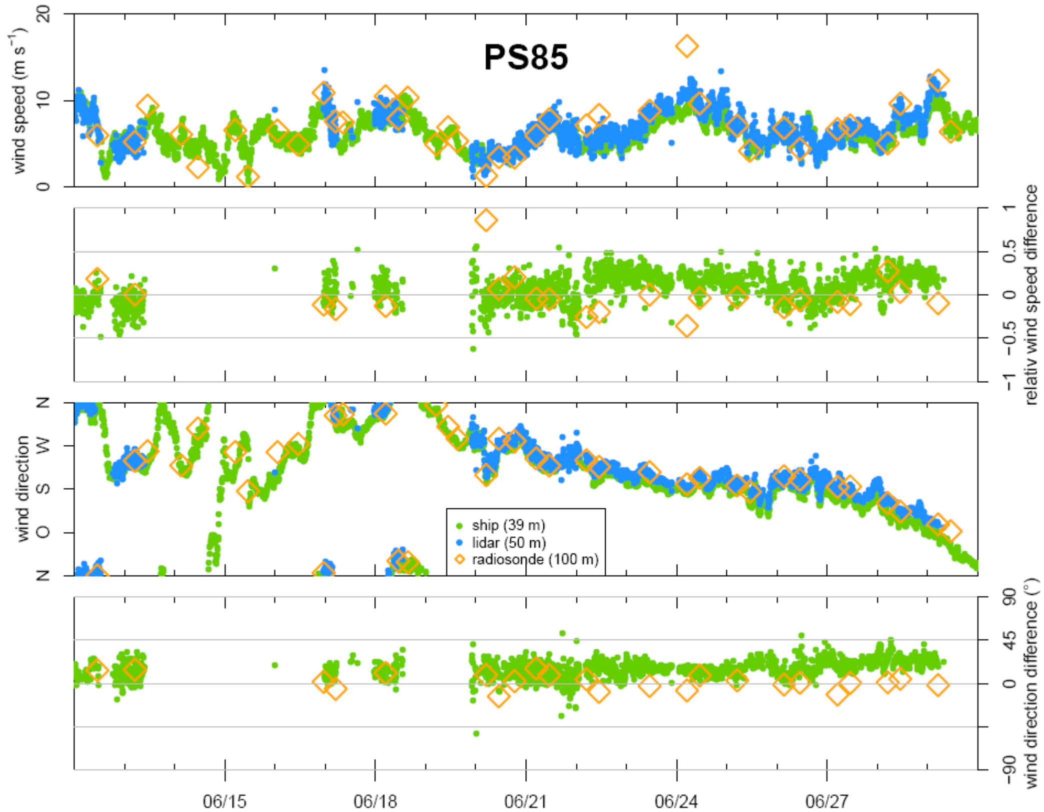
We don´t think that the benefit is high enough to add this additional plots to the paper

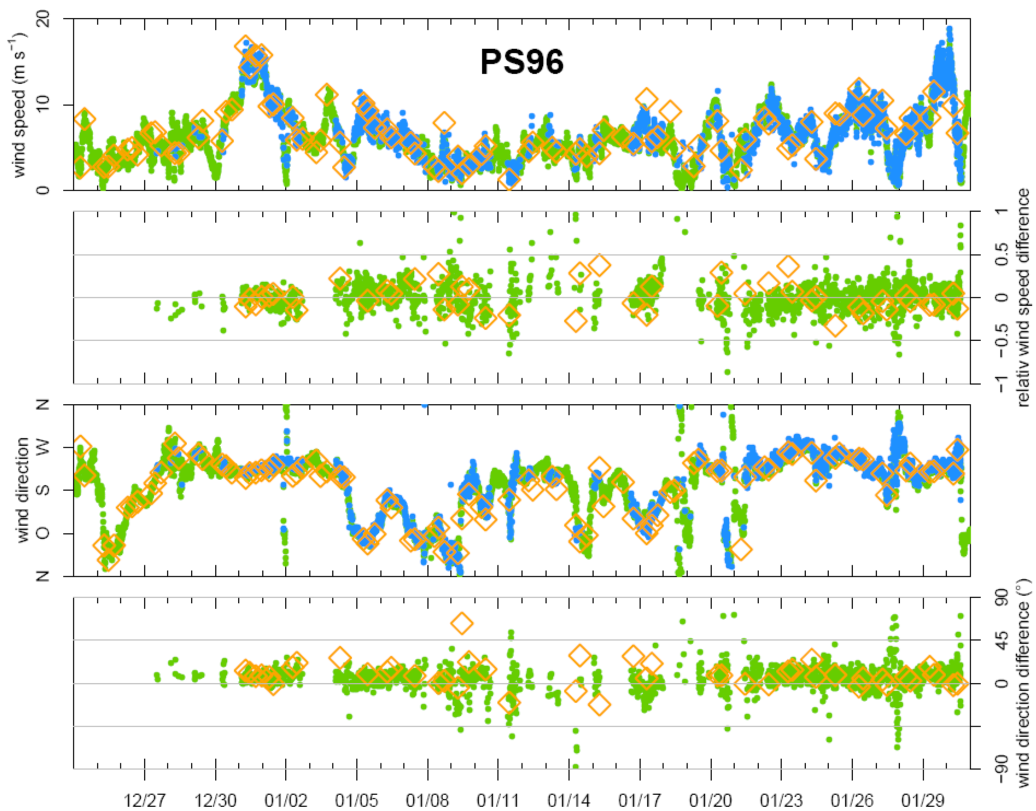
(But we change the Figure 6 by adding a little space between the RMSD and bias subplots.)



- Figure 7: Please add a plot for relative difference for the comparison of wind speed and wind direction for lidar and radio soundings as well as for lidar and ship anemometer.

We added the relative differences for wind speed and absolute difference for wind direction. We split the old figure into two separate once.





- Figure 9: It looks like the lower SNR values between 300 and 600 m Figure 8 (bottom) have more influence on the wind direction than the wind speed. What would be the reason?

The higher scatter of lidar wind directions between 400 and 700 m in Fig.9 are more likely associated with lower wind speeds.

Authors response to Anonymous Referee #2 (Received and published: 3 August 2018)

The manuscript (Zentek et al.) explains how a commercial Doppler lidar (HALO Streamline) is operated on RV Polarstern. The results are compared to standard measurements of radiosondes and sonic anemometers onboard. The lidar was operated during two campaigns in the Arctic and Antarctic. Such measurements in the changing Arctic regions specifically and on the sea, in general, are of importance as those places lack such data. The manuscript focuses on the technical aspects of operating the HALO and analyzing the datasets from the ship. The wind profiles measured from the Doppler lidar agree well with the other sensors, as shown in many other studies before.

Although the steps taken to derive wind profiles are fine the authors should write more explicitly what is new (approaches or findings) compared to other similar measurements.
We think that is already mentioned in the abstract: 1) first ship-based lidar measurements in the Antarctic, 2) assessing the quality of wind profiles of a non-motion stabilized Doppler lidar operated on an icebreaker, 3) empirical SNR method. We repeat this information now in the conclusion section in the following way:

> …high-frequency Attitude Heading Reference System. This is the first time that a wind lidar was operated on an icebreaker in the Antarctic. A processing chain including quality control tests with a new empirical SNR threshold method and an error quantification is presented.

One general statement of the manuscript seems to be that the active stabilization of the Doppler lidar is not required as shown as in Achtert et al. It should be explicitly stated in the conclusions that this is probably true only for measurements of horizontal winds with the VAD technique. Measurements in PPI scanning mode configurations, or even more importantly turbulence and sedimentation-speed measurements of clouds and ABL vertical-wind measurements are very strongly influenced by the motion of the ship.
Before:

> In conclusion, the results of the postprocessing of non-motion stabilized lidar data achieve comparable good quality as the motion-stabilized lidar study of Achtert et al. (2015).

Now:

> In conclusion, the results of the postprocessing of non-motion stabilized lidar data achieve comparable good quality as the motion-stabilized lidar study of Achtert et al. (2015). <u>As our study focuses only on horizontal winds it should be noted that the influence on vertical wind and turbulence measurements is higher and was not evaluated. The need of a motion stabilized lidar for those measurements could be very important.</u>

I found it a bit confusing to see three different things called "NOISE" in the manuscript. There is the signal detection noise in the SNR (which also determines the SNR threshold), then there is the error of the line-of-sight wind estimator (peak-finding accuracy or similar, which is connected to the Cramer-Rao Lower Bound theorem), there are outliers if the wind estimator fails, and finally there is the error of the VAD result on the final horizontal wind by non-perfect compensation of the ship's attitude. These quantities should be differentiated more carefully in the manuscript.
We agree and tried our best to distinguish between error, SNR and empirical noise. Apart from our revisions due to the minor comments from the reviewers (e.g. "P6L26-27") we searched the paper for the words "noise", "SNR" and "error" and made the following changes:

Page 1 line 13

Change: “SNR” to “SNR threshold”

Page 5 line 6

Before:

> First a signal-to-noise ratio (SNR) threshold was chosen and all data points within one ray with a worse SNR were removed.

Now:

> First a signal-to-noise ratio (SNR) threshold was chosen and all data points within one ray with a worse SNR were removed. The SNR is a value given in the lidar output for each scanned Doppler velocity value. It is separate from the empirical noise defined in section 2.3.3 as well as from the “noisy influence” due to other error sources like uncertainties related to the ships movement.

Page 7 line 1

Before:

> Data points outside this range can be regarded as noise. This condition is used to find a SNR threshold in a three-step procedure. First, we look at the overall frequency distribution of measured Doppler velocities (Fig. 5, top). We assume that the data mainly consists of two parts: the noise (homogenous along all wind speeds; top to bottom) and the wind signal (relatively homogenous along the signal intensity or SNR; left to right). Signal intensity is defined as SNR+1. All points above 10 m s−1 or below -10 m s−1 are taken to construct a noise distribution as a function of intensity using the mean value (Fig. 5, bottom). We call this the empirical noise. We call this the empirical noise. In the second step, we take the ratio of the empirical noise and the mean of the measured Doppler velocities for each intensity, which results in an empirical noise fraction (plotted as solid line in Fig. 5, bottom). The noise fraction is close to zero for high intensities and starts to increase rapidly at different SNR values for both data sets.

Now:

> Data points outside this range can be regarded as wrong (or empirical noise). This condition is used to find a SNR threshold in a three-step procedure. First, we look at the overall frequency distribution of measured Doppler velocities (Fig. 5, top). We assume that the data mainly consists of two parts: the empirical noise (homogenous along all wind speeds; top to bottom) and the wind signal (relatively homogenous along the signal intensity or SNR; left to right). Signal intensity is defined as SNR+1. All points above 10 m s−1 or below -10 m s−1 are taken to construct an empirical noise distribution as a function of intensity using the mean value (Fig. 5, bottom). ~~We call this the empirical noise.~~ In the second step, we take the ratio of the empirical noise and the mean of the measured Doppler velocities for each intensity, which results in an empirical noise fraction (plotted as solid line in Fig. 5, bottom). The empirical noise fraction is close to zero for high intensities and starts to increase rapidly at different SNR values for both data sets.

Page 9 line 2

Change: “noisy pixels” to “outliers”

Page 10 line 2

Change: “SNR” to “SNR threshold”

Fig 8 and 10

Change: “SNR” to “SNR threshold”

Minor comments:

-The abstract is too long and can be shortened. It shouldn't include a motivation and lengthy formulations. Just the facts in a very condensed form.
We removed the first two (motivating) sentences of the abstract, but think that the other formulations are adequate.
Abstract:

> ~~Profiles of wind speed and direction at high spatial and temporal resolution are fundamental meteorological quantities for studies of the atmospheric boundary layer. Ship-based Doppler lidar measurements can contribute to fill the data gap over oceans particularly in polar regions.~~ In the present study a non-motion stabilized scanning Doppler lidar was operated […]

–RMSD is not explained
Changed in the abstract from "RMSD" to "root-mean-square deviation"
Added in with the first appearance in the text "root-mean-square deviation (RMSD)"

-P1L19: the abbreviation AOI is not really needed as it is never used again in the manuscript. (It could be mentioned that AOI interactions are strongly related to turbulent processes in the ABL which can be observed with a lidar, too. Even though turbulence parameters and not measured here.)
We changed the two occurrences of "AOI" to atmosphere-ocean-ice.

-P2: Most of the introduction/literature review deals with the specifics of the HALO lidar and not for Doppler lidars in general. This should be mentioned or revised.
This is now mentioned.
Before:

> In synergy with additional remote sensing instruments measuring the temperature profile, the turbulent mixing conditions in the ABL can be described at high temporal and vertical resolution of 10 min and 10 m, respectively (Brooks et al., 2017).
> In this study we analyze data from a scanning Doppler lidar on board of RV Polarstern in the Arctic (June 2014) and Antarctic (December–January 2015/2016).

Now:

> In synergy with additional remote sensing instruments measuring the temperature profile, the turbulent mixing conditions in the ABL can be described at high temporal and vertical resolution of 10 min and 10 m, respectively (Brooks et al., 2017). Note that our literature research was focus on lidars similar to our own, thus it is likely biased towards lidars from the same manufacturer.
> In this study we analyze data from a scanning Doppler lidar on board of RV Polarstern in the Arctic (June 2014) and Antarctic (December–January 2015/2016).

-P3L21: Could you please describe the HALO configuration in more detail? How can there be a 3m range resolution? I assume that the laser pulse is much longer? The effect of overlapping gates and non-independent measurements at those range gates should be mentioned. A bit is seen in Fig. 4, but the explanation could be more specific for the zigzag lines
Our usage of the term "gate" was misleading/inconsistent. The HALO lets one choose the gate length by selecting in multiples of 6 m. In the software you choose 2*n points per gate, where each point increases the range length by 3 m. We changed our phrasing making it also consistent with Table 1.
Before:

> One ray is divided into gates of 3 m length and the measured Doppler velocity is representative for six gates (18 m). During PS85 those six gates were non-overlapping,

thus measurements were available every 18 m. During PS96 the six gates were overlapping, thus measurements were available every 3 m.

Now:

One ray is divided into sections of 3 m length and one measured Doppler velocity is representative for gate length of six sections (18 m). During PS85 those six sections were non-overlapping, thus measurements were available every 18 m. During PS96 the six sections were overlapping, thus measurements were available every 3 m. But the measurements with overlapping sections are not independent as they are computed based on partially same data.

-P4L34: ..."if data quality is not of importance". Better mention the errors with and without correction. This could help a reader to evaluate the effects of a/no stabilization better.

We said: "**high** data quality". We think this is already given in the sentence following "For example, […] causes […] wind speed error of less than 13% […]"

-P5L6: How it the SNR defined for the HALO? This is important to follow the upcoming discussion about the thresholds. Some people also derive the CNR to be more correct. I would like to see a bit more discrimination between those terms here or at P6L17.

We would like to stay with SNR, which is used in many other studies. We added the following sentences near P6L17 to explain how the SNR is computed:

The background noise is usually measured at least once a day and at most every hour. For this, the scanning head is turned away from the sky towards the lidar casing and measures the signal while sending no pulses out. Thus the background noise can vary with time and operating conditions and can be different for different HALO instruments. To compute the SNR this signal strength of the background noise is subtracted from the signal strength of the measurement and afterwards divided by the signal strength of the background noise. If the signal during a measurement is lower than during the background noise scan, it can therefore cause a negative SNR. In general, more background noise scans were performed during PS85, but we didn´t investigate the background noise further.

-P5L20: Could you please find formula signs (one character) for wind speed and direction other than dd and ff?

We changed it to "v_h = sqrt(u^2+v^2)" and "phi_h" for wind direction

-P5L25: 1st: "has a fixed elevation ANGLE" And 2nd: Is that really true for a ship?

In our setup not. But with a stabilizing platform it would be.

Before:

Assuming that the lidar remains stationary and has a fixed elevation angle θ, the equation further simplifies to

Now:

Assuming that the lidar remains stationary and has a fixed elevation angle θ (which is not the case in our setup), the equation further simplifies to

-P6L17: Should you even expect that the SNR threshold is the same for every of the HALO instrument? I'm not sure that these thresholds can be really compared. Maybe the laser power/pulse length/DAQ bit resolution is different? Again, it depends, how the dB's are defined here.

Yes comparability is a big issue. As we show in section 2.3.3 with Fig.5, even the same lidar with another averaging time demands a different SNR threshold. Also we do not know of any common definition of a SNR threshold so a comparison is already problematic here. We just gave a short overview of values in literature and assumed that the common definition of dB in the context of a lidar is 10log_10(x).

-P6L22: Is it really true, that the spectra or ACF cannot be stored? One should always try to store the spectra (at least for a while) in low-signal regimes like the Arctic so that later post-averaging is possible to increase the SNR.
Correction / Clarification: Yes, it is possible, but we didn´t do it. We add this as a recommendation to the end of the conclusions.
Before:

> This is necessary during the measurements, since raw data of single pulses are not stored and no postprocessing is possible.

Now:

> This is necessary during the measurements, since raw data of single pulses were not stored and thus no postprocessing is possible.

End of conclusions:

> For conditions with low backscatter due to the low aerosol concentration as it is typical for the polar regions, a possibility to optimize the averaging time of the lidar would be the storage of the of the raw data (spectra) for post-processing.

-P6L26-27: If the PS96 data contain less noise then I would expect that the SNR is higher. How can this mean, that you need a different SNR threshold? This is a bit counterintuitive. Except the so-called "SNR" is the signal and not an actual SNR?
The other Reviewer although raised this issue. We added the following clarification
----------------------8<-------snip------------------------
Before:

> “Additionally due to the different averaging time for each ray during PS85 and PS96 (1.5 vs 12–15 sec), the PS96 data contains less noise and thus it makes sense to choose a different SNR threshold for each data set.” (page 6, line 25-27)

After:

> “Additionally due to the longer averaging time for each ray during PS96 (12–15 sec) than PS85 (1.5 sec), the PS96 data allow for a lower SNR threshold compared to the PS85 data, because averaging over a longer period given the same SNR results in better data. Thus it makes sense to choose a less strict SNR threshold for the PS96 data set to make both data sets more comparable.”

----------------------8<-------snap------------------------

-P6L27-30: Please explain why the vertical wind is around 0. This depends on the averaging period, on the precision of the angle of vertical stare (especially on a ship), and sometimes even on the synoptic situation (stationary waves, etc). In fact, you can determine the noise of your LOS wind by looking at the difference between the Autocovariance function at 0 and at the first leg. Or by evaluating the high-frequency tail of the wind power spectrum. And this could be done for different SNR thresholds.
“The vertical wind is around 0 m s^{-1}” just means it is close to zero. Päschke et al. (2015) found that this is valid for quiescent atmospheric conditions. Measured vertical velocities exceeding 5 m s^{-1} can be safely considered as noise for the conditions of our measurements. Since we used the 75° elevation measurements (instead of vertical stares), we assumed LOS velocities of +- 10 m s^{-1} unrealistic.
We added the following clarification
Before:

Knowing that these had to be around 0 m s−1, the influence of noise could be evaluated. We follow a similar approach and evaluated the Doppler velocity from all individual rays for VAD scans with an elevation of 75°

Now:

Knowing that vertical velocities are close to zero, Päschke et al. (2015) could evaluate the influence of noise from vertical stares for quiescent atmospheric conditions. As we did not have a stabilizing platform, the evaluation of the vertical stares is not possible because of the influence of horizontal wind on the signal. To circumnavigate this problem, we follow a similar approach and evaluated the Doppler velocity from all individual rays for VAD scans with an elevation of 75°

-P7L1-12: this paragraph is a bit hard to understand. Can you explain why you do not use the goodness of the VAD fit to determine when a VAD delivered good and bad results?

We computed the goodness of the VAD fit, but we did not include this in our analysis so far. Of course using the goodness of the VAD fit is also a valid error measure of the wind retrieval, which includes also the inhomogeneity of the wind field.

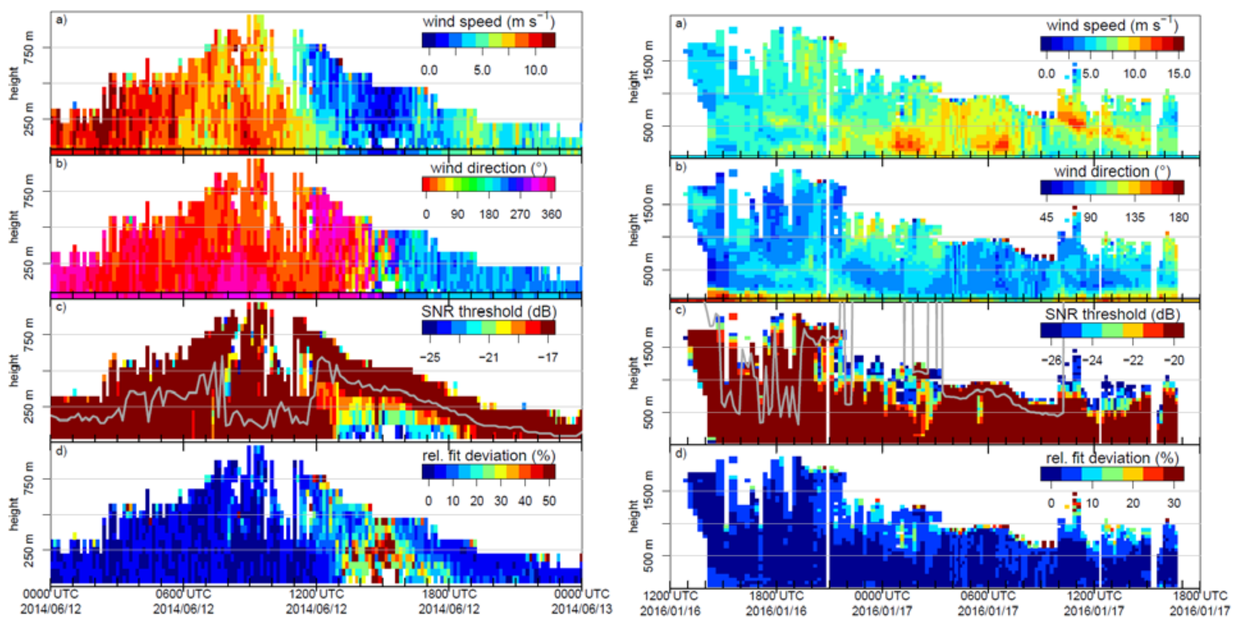
We added the goodness of the fit in the theory section (page 6 line 5)

As the system of equations is only solved approximately for a given a solution (u*,v*,w*) we can define a measure for the goodness of the fit. Päschke et al. (2015) define the coefficient of determination. We define the fit deviation in our paper as:

<<<formel>>> || Matrix * (u*,v*,w*) – Vector ||index2

For the purpose of comparing the fit deviation, only scans with the same elevation should be used. It should also be noted that measuring a non-homogenous or non-stationary wind field would result in a larger fit deviation value.

We added the fit deviation to Figure 8 and 10 and changed bottom/middle/top references to a/b/c/d in the text



New Figure Description:

Lidar wind speed (a) and direction (b) for -25 dB SNR threshold for the 12 June 2014 (location see PS85 in Fig. 1). Colors below the black line (40 m) show the wind measurements of RV Polarstern (anemometer). The plot (c) presents the SNR thresholds that would allow for a wind calculation. The grey line is the cloud base from ceilometer measurements of RV Polarstern. The relative fit deviation (fit deviation divided by wind speed) is shown in (d). Values outside the colour range are plotted with the highest colour. Only scans with a 75° elevation where used.

We added the following text to page 9 line 4:

> The fit deviation (Fig. CITE, d) can help with this decision, but notice that the high relative fit deviation in the afternoon stems mostly from the low wind speeds

-P7L5: It is mentioned that the VAD results are averaged for 20 min. But since it is possible that there are outliers in those VAD results shouldn't the median be a better indicator here?

We agree that there are more options for the post-processing. This could include also a filtering of outliers in each vertical profile. In Figure 5 and Table 3 we want to distinguish the good/bad influence of choosing a different SNR threshold or other configurations and thus there is also an influence of outliers.

-P7-8L20: The bias of 10° for the wind direction is quite unexpected. One would assume that radiosondes and the lidar both use GPS? Or is there a magnetic compass involved somewhere which might show a bias in Polar regions? It is a bit unsatisfactory that the reason for this bias remains unclear here. What is the VAISALA specification for their wind-direction bias?
The bias is between 3 and 7° for the radiosonde comparison (not: 10°).
Yes, the radiosondes use GPS. The lidar wind direction is computed relative to the ship and is modified by the data from the ship navigation system (as far as we know this is a combination of magnetic compass and GPS). We contacted the people maintaining the ship navigation system to determine if there could be a bias (either with the system or along the way of data transmission and data base up-/download), but that seemed not to be the case.
The reference system for the ship navigation system and the radio sounding is true north (not magnetic north).
Vaisala states an uncertainty of 2° for their radiosondes (standard deviation of differences in twin soundings, wind speed above 3 m/s)
[we noticed the link for the Vaisala reference was not shown correctly and changed it]

-P8L31: here it should be -17dB.
Yes. We searched our paper for "dB" and also added missing minus signs in some other cases.

-P9L5-15: The occurrence of these 3 fast LLJs is interesting. Can you give any explanation of the processes? Usually, in these latitudes, the typical Ekman oscillation time should be 12 hrs. So there must be another cause for these LLJs. Maybe you could reference some work on Arctic LLJ (e.g. Jacobson, ACP, 2013).
There are different mechanisms for the formation of LLJs: 1) decoupling of layers in the SBL resulting in supergeostrophic winds by an inertial oscillation. 2) baroclinity causing vertical wind shear, 3) katabatic winds, 4) topographic channeling and local density flows, 5) LLJ due to a change in surface friction e.g. from rough sea ice to smooth water surfaces. We did not want to go into too much detail of dynamical processes for our case study. The LLJs on 17 January 2016 were measured during the passage of a synoptic front. In addition, the ship was located in a polynya the lee of a huge iceberg (A23A, size about 60kmx80km), which causes low-level baroclinicity. So baroclinicity seems to be the main reason.

We add the following text to P9L15:

> The dynamics of the LLJs were not studied in detail. They occurred during the passage of a synoptic front, when the ship operated in a polynya the lee of a huge iceberg (A23A, size about 60kmx80km). Baroclinicity is therefore a likely reason for the LLJs. While LLJs caused by inertial oscillations are frequent in the Weddell Sea during winter (Andreas et al. 2000), the observed jets during PS96 are comparable to the situation of the summertime Arctic Ocean, where Jakobson et al. (2013) find mostly baroclinic jets associated with transient cyclones.

Andreas, E. L., Claffy, K. J., and Makshtas, A. P. 2000. Low-level atmospheric jets and inversions over the western Weddell Sea, Boundary-layer meteorology, 97, 459–486. https://doi.org/10.1023/A:1002793831076

Jakobson, L., Vihma, T., Jakobson, E., Palo, T., Männik, A., and Jaagus, J.: Low-level jet characteristics over the Arctic Ocean in spring and summer, Atmos. Chem. Phys., 13, 11089-11099 , https://doi.org/10.5194/acp-13-11089-2013, 2013.

-P9L17: A bit general comment: I believe the topic of the paper should not be: The radiosondes are correct and let's see how well the lidar can be verified by this. A Doppler lidar by its design is one of the best methods to measure wind. In the end, it just depends on how precise one can measure frequencies. It is more a question of how much errors are generated by a moving platform like a ship. And what are the best scan strategies on a ship? It would be nice to get some answer to these questions, as other researchers might profit from this.

Before:

> We presented a verification of wind speed profiles measured by a wind lidar during two cruises of the research vessel Polarstern in the Arctic and Antarctic. The lidar was not motion-stabilized, but ship motions and the ship's orientation were measured by the ship's navigation system and by a high-frequency Attitude Heading Reference System. The wind calculation is based on VAD scans with eight directions (rays), thus there is a high oversampling which allows for additional quality tests.
> […]
> The comparison with the routine wind measurements of the ship at 40 m height yields a larger data set and a similar bias and RMSD.

Now:

> We present a verification of wind speed profiles measured by a wind lidar without a stabilizing platform during two cruises of the research vessel Polarstern in the Arctic and Antarctic. The ~~lidar was not motion-stabilized, but~~ ship motions and the ship's orientation were measured by the ship's navigation system and by a high-frequency Attitude Heading Reference System. The wind calculation is generally based on VAD scans with eight directions (rays) at an elevation angle of 75° (an elevation of 85° was discarded after a short test period), thus there is a high oversampling which allows for additional quality tests. Wind retrievals from scans at multiple elevation angles elevation angles (25, 50 and 75°) slightly improve the quality of the wind profile, but take more time. The low aerosol concentrations in polar regions result in a low backscatter. As a strategy to optimize the backscatter signal for these conditions the adjustment of telescope focal length of the lidar and the averaging time is useful.
> […]
> The comparison with the routine wind measurements of the ship at 40 m height yields a larger data set and a similar bias and RMSD. The choice of a longer averaging time is preferred as it allows to reduce the SNR threshold and thus increases the amount of data. For longer averaging times the influence of the ship's movement can be higher, but this effect is small in our case because the ship operated mainly in sea ice where wave heights are relatively small.

-P9L27: "The RMSD is 10° but we also find 5°". This sentence needs to be revised.

Before:

> Overall, the radiosonde comparisons yield similar results as found in in Achtert et al. (2015) using as motion-stabilized lidar. The wind speed bias is very small (0.1 m s−1) for our standard data processing and the RMSD is about 1 m s−1. **For wind direction, the RMSD is about 10°, but we also find a bias of 5°.** In conclusion, the results of the

postprocessing of non-motion stabilized lidar data achieve comparable good quality as the motion-stabilized lidar study of Achtert et al. (2015).

Now:

Overall, the radiosonde comparisons yield similar results as found in in Achtert et al. (2015) using as motion-stabilized lidar. The wind speed bias is very small (0.1 m s−1) for our standard data processing and the RMSD is about 1 m s−1. **For wind direction, the RMSD is about 10°, which is comparable to other studies. The mean bias between radiosondes and lidar is about 5°. This is higher than the value of 2° found by Achtert et al. (2015), who find a bias of 5°only at higher levels, which is explained by the drift of the radiosonde and the resulting decrease in collocation of the measurements. Overall** the results of the postprocessing of non-motion stabilized lidar data achieve comparable good quality as the motion-stabilized lidar study of Achtert et al. (2015).

-P10L1: "turning the wind perpendicular to the wind". Something is wrong here.

Before:

Turning the wind perpendicular to the wind is desirable.

Now:

Turning the ship perpendicular to the wind is desirable.

-Fig.3: Can this also be done for the YAW angle? It seems the wind direction error bias is highest. So probably one should have a look at this angle, too.

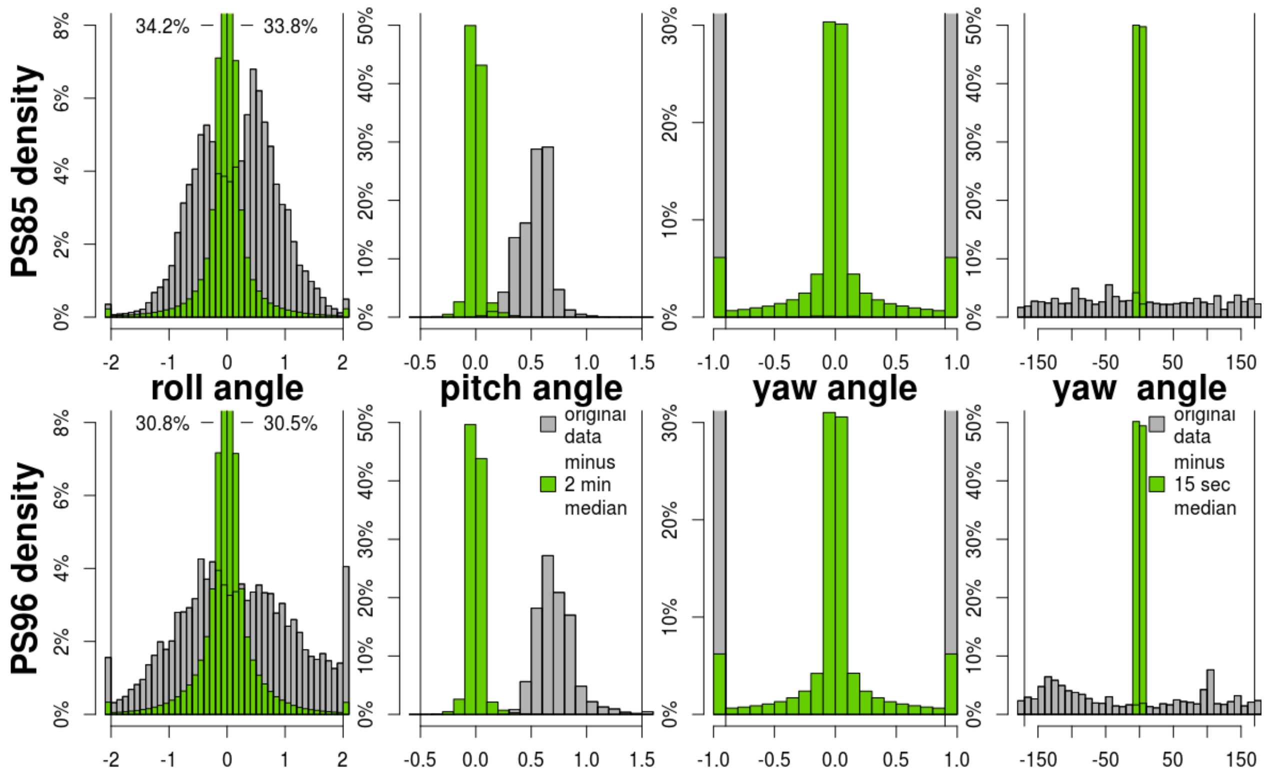
Fig. 3 is focused on the roll and pitch angle during “in sea ice” condition. And shows that high frequency data is not that important but a low frequency data set already does good work.
The yaw angle is another issue not focused on polar regions or motion-stabilizing.
The 2 min median subtraction gives no relevant information in the yaw case. So we plotted 15 second median subtraction (appropriate for our average time).
We think the benefit of such a plot is too small as it just gives an approximation of how much data is removed by the sd(yaw) < 2° criterion.
We instead added this information on page 4 line 25ff

Before:

It should be noted that the correction and filtering process causes almost no loss of data, as the ship’s movement even during ice breaking conditions generally does not result in high-frequency changes of roll and pitch (except some cases of ramming).

Now:

It should be noted that the correction and filtering process causes almost no loss of data. Only in 6% of the time, the standard deviation of the yaw angle over 15 seconds is larger than 2° and the ship’s movement even during ice breaking conditions generally does not result in high-frequency changes of roll and pitch (except some cases of ramming).



-Fig.5: If Intensity is SNR+1, why does the axis start at 0.99?

The way the lidar software computes the SNR (subtracting a background noise from the signal) can cause negative values for the signal (and thus a negative SNR). This question should be resolved by our changes concerning comment -P5L6

-Fig.7: It would be nice to include a correlation plot.

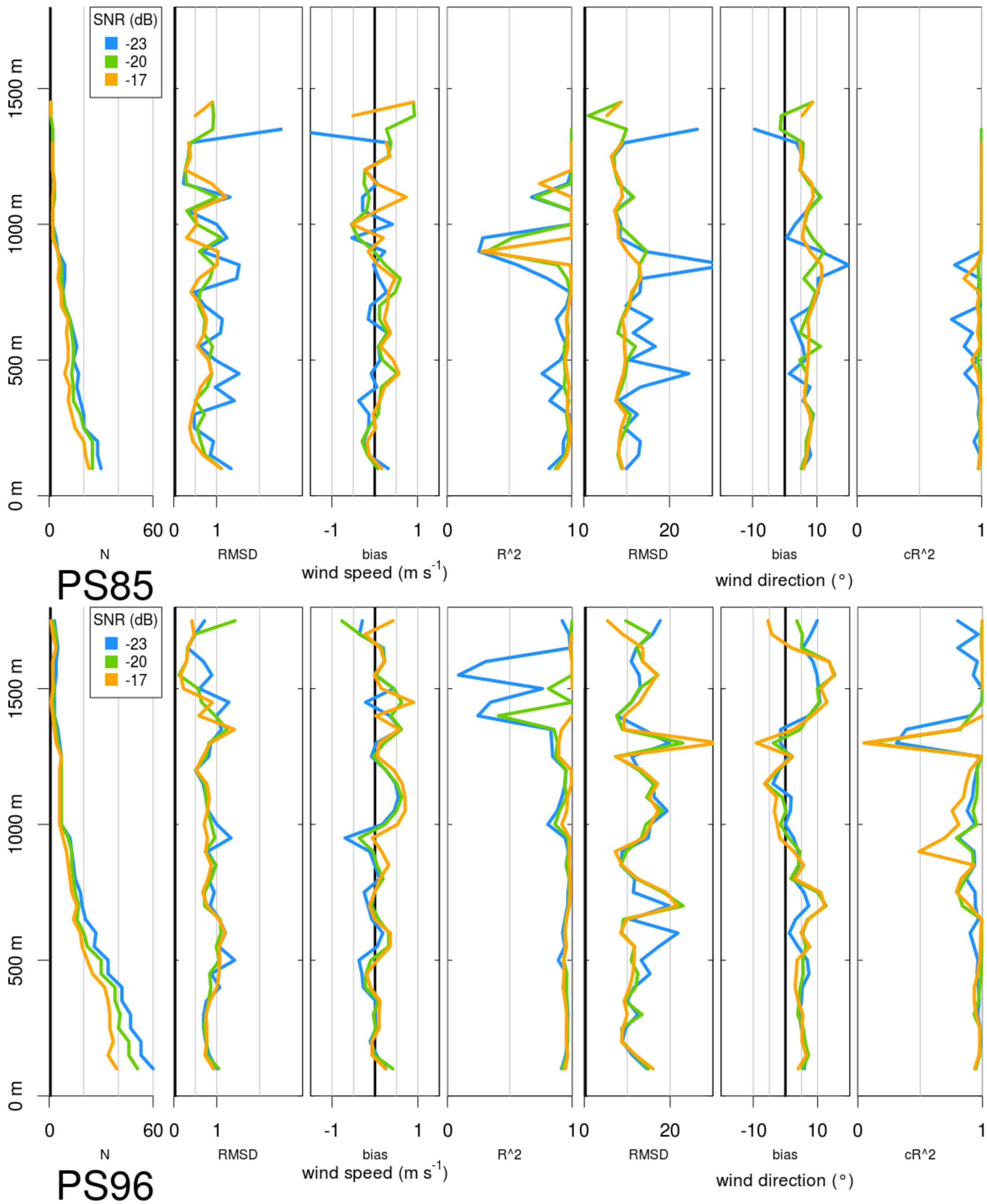
We don´t think the plot gives additional information. It is close to 1 except where there is not enough data with the same characteristics as the RMSD and bias. The numbers can better be seen in Table 2 and 3. We added the following sentence on page 7 line 24

Before:

> Furthermore, a systematic dependence on height is not present. At heights above 1000 m the sample size is relatively small and differences between different SNR thresholds are not robust.

Now:

> Furthermore, a systematic dependence on height is not present. <u>We also check for a height dependence of the correlation (not shown), but there was none present.</u> At heights above 1000 m the sample size is relatively small and differences between different SNR thresholds are not robust.



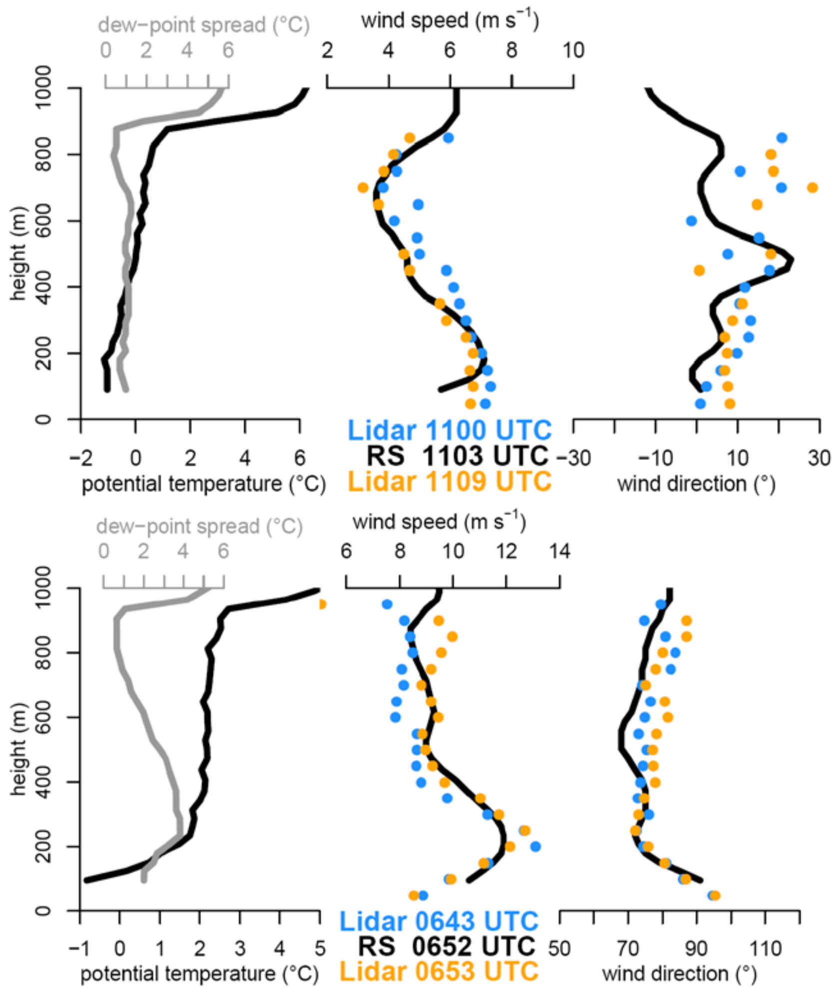
-Fig.8: What is seen on the y-axis? Height? If it is height, how can there be measurements several hundred meters above the cloud base? Please explain in the text.

Yes its height, we added the label.

We checked the backscatter data of the lidar found that it was not as high as for example after 1200 UTC. We appended the following explanation to the section 4.1 PS85 - Arctic 2014/06/12

"Note that the height difference between lidar and ceilometer from 0800 to 1200 UTC is likely due to a thin layer of low clouds that the lidar could partially penetrate."

-Fig.9: Theta can be omitted and the x-axis can be annotated "POT.TEMPERATURE (°C)". Also, the date can be omitted since it is mentioned in the figure caption.
All axis could be relabelled… so we did that. And added to the Figure description: "25, 50 and 75° elevation scans where used.



-Table 1: The SNR threshold is -20dB, but in the text, it is written -18.2dB is from the manufacturer. So what means "default"?
We just to give an approximate number in the table, but we agree that it is more confusing than helpful. We removed the row in Table 1
"Threshold for signal-to-noise ratio (SNR) : variable (default -20 dB)".

Analysis of the performance of a ship-borne scanning wind lidar in the Arctic and Antarctic

Rolf Zentek, Svenja H. E. Kohnemann, and Günther Heinemann

Department of Environmental Meteorology, University of Trier, Germany

Correspondence: Rolf Zentek (zentek@uni-trier.de)

Abstract. ~~Profiles of wind speed and direction at high spatial and temporal resolution are fundamental meteorological quantities for studies of the atmospheric boundary layer. Ship-based Doppler lidar measurements can contribute to fill the data gap over oceans particularly in polar regions.~~ In the present study a non-motion stabilized scanning Doppler lidar was operated on board of RV *Polarstern* in the Arctic (June 2014) and Antarctic (December–January 2015/2016). This is the first time that such a system measured on an icebreaker in the Antarctic. A method for a motion correction of the data in the post-processing is presented. The wind calculation is based on vertical azimuth display (VAD) scans with eight directions that pass a quality control. Additionally a method for an empirical signal-to-noise ratio (SNR) threshold is presented, which can be calculated for individual measurement setups. Lidar wind profiles are compared to total of about 120 radiosonde profiles and also to wind measurements of the ship.

The performance of the lidar measurements in comparison with radio soundings shows generally small ~~RMSD~~ root-mean-square deviation (bias) for wind speed of around 1 m s^{-1} (0.1 m s^{-1}) and for wind direction of around 12° (6°). The postprocessing of the non-motion stabilized data shows a comparable good quality as studies with motion stabilized systems.

Two case studies show that a flexible change of SNR threshold can be beneficial for special situations. Further the studies reveal that short-lived Low-Level Jets in the atmospheric boundary layer can be captured by lidar measurements with a high temporal resolution in contrast to routine radio soundings. The present study shows that a non-motion stabilized Doppler lidar can be operated successfully on an icebreaker. It presents a processing chain including quality control tests and error quantification, which is useful for further measurement campaigns.

1 Introduction

Changes in the Arctic and Antarctic climate system are strongly related to atmosphere-ocean-ice ~~(AOI)~~ interactions and feedbacks between the atmospheric boundary layer and the free atmosphere. Hence, the knowledge about the state of the atmospheric boundary layer (ABL) is crucial for the understanding of ~~AOI~~ atmosphere-ocean-ice processes, atmospheric transports, air pollution processes and the verification and improvement of numerical weather forecast and climate models for polar regions. Profiles of wind speed and direction at high spatial and temporal resolution are fundamental meteorological quantities for ABL studies. While at mid-latitudes the ABL is studied using tall towers and ground-based remote sensing instruments such as lidar, radar or sodar at several observatories, these measurements are rare or absent in the Arctic and Antarctic. Thus radiosondes are generally the main source to measure quantities of the ABL in the polar regions. Since the radiosonde stations are primarily located over land, there are huge data gaps over the ocean. Furthermore, the temporal resolution of radio soundings is generally of the order of a couple of hours. Over the polar oceans, only few research vessels provide radio soundings, which are very valuable to improve the initial conditions for numerical weather forecasts and for reanalyses (e.g., Dee et al., 2011), but are insufficient for detailed studies of boundary layer processes.

Ship-based Doppler lidar measurements are a possibility to fill the gap of radio soundings over oceans, since they provide wind profiles with a high spatial and temporal resolution (Tucker et al., 2009; Achtert et al., 2015). In addition, Doppler wind lidar measurements allow for the determination of the turbulence structure of the ABL (Banta et al., 2006; Pichugina et al., 2012; Kumer et al., 2016). If two Doppler lidars are available, techniques like the 'virtual tower' can be applied (Calhoun et al., 2006; Damian et al., 2014). In synergy with additional remote sensing instruments measuring the temperature profile, the turbulent mixing conditions in the ABL can be described at high temporal and vertical resolution of 10 min and 10 m, respectively (Brooks et al., 2017). Note that our literature research was focus on lidars similar to our own, thus it is likely biased towards lidars from the same manufacturer.

In this study we analyze data from a scanning Doppler lidar on board of RV *Polarstern* in the Arctic (June 2014) and Antarctic (December–January 2015/2016). There are two important aspects measuring with a Doppler lidar on board of a moving ship in polar regions: a) the ship's movement requires data corrections regarding its orientation and b) the adaptation of lidar measurement settings and analysis configuration for conditions with low backscatter due to the low aerosol concentration. Some studies present measurement campaigns dealing with challenge a) (e.g., Pichugina et al., 2012; Tucker et al., 2009; Achtert et al., 2015). All of them use a motion-stabilization platform to remove the effects of ship motion. We present a different option to deal with the varying orientation of the ship. The adaptation of measurement settings for the polar environment (challenge b)) is less documented. The goal of these adaptions is the improvement of the signal-to-noise ratio (SNR). Hirsikko et al. (2014) recommend the use of an optimized telescope focal length of the lidar and the increase of the integration time for measurements in Finland. The main goal of the present paper is the assessment of the wind lidar performance in comparison with radiosondes on the German icebreaker *Polarstern*. A similar study was made by Achtert et al. (2015), who used a motion-stabilized scanning wind lidar during a cruise of the Swedish icebreaker ODEN in the Arctic in 2014 (Tjernström et al., 2014). Their three-month campaign started immediately after our Arctic campaign in 2014. No ship-based measurement campaign of a Doppler wind lidar is known for the Antarctic. The combination of the measurement framework and the presented comprehensive analysis of the settings serve as basis for improvements in further data collections. The outline of the paper is as follows: In Section 2 an overview of the measurement campaigns and the data processing is given. Section 3 presents the results for intercomparisons of lidar data with radiosondes and ship wind measurements. Two case studies are shown in Section 4. A summary and conclusions are given in Section 5.

2 Measurements and data processing

The measurements were performed during the two *Polarstern* cruises PS85 and PS96 of the Alfred-Wegener Institute Bremerhaven (Germany). The cruises with approximate sea ice conditions during the measuring periods are shown in Figure 1. PS85 took place in the Arctic from the 06th June till 03rd July 2014 and PS96 in the Antarctic from the 06th December 2015 till 14th February 2016. Lidar measurements were taken for a period of 18 days (12th till 29th July) during PS85 and for 38 days (24th December till 30th January) during PS96. *Polarstern* is the German research icebreaker with a length of 118 m and a weight of 17300 tons (Fig. 2). The typical cruise speed is 12 knots.

2.1 Doppler wind lidar

The instrument is a "Halo-Photonics Streamline" wind lidar, which is a scanner and can operate with a maximum range of 10 km, but was used only for a range up to 3600 m due to the low aerosol concentration (Table 1). The lidar was installed on the port (starboard) side of the ship during PS85 (PS96) approximately 20 m above the waterline (see Figure 2). Besides the lidar, an external Altitude Heading Reference System (AHRS; XSENS MTi-G-700-GPS/INS) was installed for higher frequency (sampled with up to 400 Hz) recordings of the ship's pitch and roll, in addition to lower frequency (1 Hz) navigation data from the ship's internal systems.

A variety of different scanning programs were used: vertical azimuth display (VAD), horizontal stare in two or three directions, range-height indicator (RHI) and vertical stare. In the present paper we will focus on the VAD measurements that allow the computation of vertical profiles of horizontal wind speed. One VAD scan is composed of eight rays with fixed elevation and different azimuth (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). During PS96 we changed the elevation from 85° to 75° after three days. The averaging time for each ray was usually 12–15 seconds. During PS85 the averaging time for each ray was only 1.5 seconds but azimuth-circles were done at 25°, 50° and 75° elevation. For the analysis we will either use only the 75° or all 25°, 50° and 75° elevations. To make them comparable in case of using all three elevations, we will count the $3 \times 8 = 24$ rays as one VAD. One ray is divided into ~~gates~~ sections of 3 m length and ~~the~~ one measured Doppler velocity is representative for ~~six gates~~ gate length of six sections (18 m). During PS85 those six ~~gates~~ sections were non-overlapping, thus measurements were available every 18 m. During PS96 the six ~~gates~~ sections were overlapping, thus measurements were available every 3 m. But the measurements with overlapping sections are not independent as they are computed based on partially same data. VAD wind profiles are typically available every 15 minutes and a whole VAD scan required about 2 minutes for PS96. Photos of the weather condition were taken manually for special situations during PS85 and automatically with a GoPro (with constant power connection) every minute during PS96.

2.2 Radiosondes

Radiosondes at *Polarstern* (König-Langlo, 2014a, 2016a) were usually launched twice a day at 05 and 11 UTC during PS85 (39 radiosondes over the 18 days) and 07 and 11 UTC during PS96 (70 radiosondes over the 38 days). Radiosondes of the type Vaisala RS92 (Vaisala, 2013) were used. The measurement uncertainty for wind is specified as 0.15 m s^{-1} for speed and 2° for direction. For the intercomparison of lidar wind profiles with the radiosonde profiles additional aspects apart from instrumental errors have to be considered. As shown below, the vertical range of the lidar is generally limited to the height of the ABL of a few hundred meters. When the ship is cruising, the radiosondes are launched close to the ship's superstructure and are affected by the turbulent wake of the ship. The radiosonde also needs time to accelerate to the ambient wind speed after launch, and exhibits strong pendulum motions during this phase. This results in a strong noise in the raw wind data, and a low-pass filter is applied, resulting in a reduced vertical resolution (estimated as about 200 m by Päschke et al. (2015)). As documented by Achert et al. (2015) for the RV Oden, the ship superstructure modifies the mean flow depending on flow direction. The largest effect occurs for a relative wind along the ship's axis. For these conditions, the disturbance decreases with height and is estimated as smaller than 2% for horizontal wind speeds at altitudes above 75 m. For a flow being perpendicular to the ship this effect reduces to 2% also below 75 m. A study of Berry et al. (2001) for RV *Polarstern* shows that the largest flow distortion for the ship orientated into the wind occurs as a wind decrease up to 30% in the lee of the main superstructure in the lowest 50 m (where the radiosonde is launched).

2.3 Analysis of the lidar data

The wind analysis consists of different steps. First we look at the influence and correction of the ship's motions. In the second part we describe our data processing method and computation of horizontal winds. In the third part we discuss our choice of the signal to noise ratio threshold.

2.3.1 Ship motion correction

The main difficulty in receiving reliable wind data results from the movements of the ship. The ship's velocity and orientation and their changes influence the directions of the lidar's outgoing and incoming rays. Therefore the ship's velocity and orientation angles are the two main factors for the correction of the measured data. During both cruises PS86 and PS96, the ship was moving with more than 1 m s^{-1} about 50% of the time. The lidar was aligned with the ship by eye as best as possible (deviations of the yaw angle between lidar and the ship are discussed later in the results section). Measured ship data from the scientific navigational platform are taken to correct each single lidar measurement by the ship's speed and roll-pitch-yaw angles. The resolution of these data is 1 Hz. The correction for the ship's roll and pitch movements can be avoided by using a motion-stabilizing platform (Achtert et al., 2015). We had no such platform, but additionally to the ship's 1 Hz navigation data, we recorded roll and pitch movements also at high-frequency (up to 400 Hz) by the AHRS that was attached to the lidar. The AHRS data were used to determine the roll and pitch offset between the AHRS (resp. lidar) reference system and the ships reference system. During PS96 the averaging time of a single ray was typically 12–15 seconds, so that we corrected each single measurement with the mean value over the averaging time. ~~All~~ This introduces an error whenever the ship's angle and thus the lidar angle changes during this averaging time. In order to reduce the error, all measurements that have a standard deviation of roll or pitch angle larger than 0.5° or yaw angle larger than 2° over this averaging time were excluded from the analysis~~in order to reduce the error.~~. Correcting the direction of the lidar measurement by the mean roll and pitch angle during the averaging time should already cause most of the error to average out, as it measures partly too much and partly too less wind speed. But even if this is not the case, for a data point in 1 km distance from the lidar a change of elevation from 75° to 75.5° (25° to 25.5°) causes a difference in height of 2 m (8 m) and the resulting horizontal wind speed error is less than 3.3% (0.4%). This is acceptable as we will later interpolate over height intervals of 50 m and only evaluate the horizontal wind in our paper. It should be noted that the correction and filtering process causes almost no loss of data~~, as the~~. Only in 6% of the time, the standard deviation of the yaw angle over 15 seconds is larger than 2° and the ship's movement even during ice breaking conditions generally does not result in high-frequency changes of roll and pitch (except some cases of ramming). The important part is in fact the low-frequency change in roll and pitch (e.g. pumping water from one tank to another, changing cargo) that gets corrected. This can be seen by subtracting a 2-minute running median from the roll and pitch data (Fig. 3). The remaining angles are within -0.1° and 0.1° in 60–70% of the time. ~~For a data point in 1 km distance from the lidar a change of elevation from 75° to 75.5° (25° to 25.5°) causes a difference in height of 2 m (8 m) and a horizontal wind speed error of less than 3.3% (0.4%). This is acceptable as we will later interpolate over height intervals of 50 m.~~ Without roll and pitch correction, values amount to -2° to 2° for roll (for 95% of the cases) and 0° to 1.5° for pitch. Therefore a setup without any roll or pitch correction at all would still provide usable data, if a high data quality is not of importance. For example, for a data point in 1 km distance from the lidar a change of elevation from 75° to 77° (25° to 27°) causes a difference in height of 8 m (31 m) and horizontal wind speed error of less than 13% (17%). We also corrected for the influence of the angular velocity of roll pitch and yaw, but it was found to be negligible. For PS96 (PS85) the correction due to angular velocity was less than 0.2 m s^{-1} for 99.7% (99.9%) of the time and never greater than 0.5 m s^{-1}.

2.3.2 Data processing

First a signal-to-noise ratio (SNR) threshold was chosen and all data points within one ray with a worse SNR were removed. The SNR is a value given in the lidar output for each scanned Doppler velocity value. It is separate from the empirical noise defined in section 2.3.3 as well as from the "noisy influence" due to other error sources like uncertainties related to the ships movement. The background noise is usually measured at least once a day and at most every hour. For this, the scanning head is turned away from the sky towards the lidar casing and measures the signal while sending no pulses out. Thus the background noise can vary with time and operating conditions and can be different for different HALO instruments. To compute the SNR this signal strength of the background noise is subtracted from the signal strength of the measurement and afterwards divided by the signal strength of the background noise. If the signal during a measurement is lower than during the background noise scan, it can therefore cause a negative SNR. In general, more background noise scans were performed during PS85, but we didn´t investigate the background noise further.

Furthermore the first data points near the lidar were removed (approx. the first 30 m) as these measurements are often affected by the outgoing pulse. Then each single ray was segmented into bins of 100 m. For each bin, outliers (radial velocity > 3 × standard deviation) were removed. If less than 50% of the data remained or if the standard deviation of the radial velocity of remaining data in the bin was greater than 3 m s^{-1}, the whole bin was removed.

To compute a vertical profile of horizontal wind speed from a complete VAD we first divided all data points into layers of different heights. A thickness of 50 m was chosen for each layer for the radiosonde comparison, but thicknesses down to 10 m were tested as well. We used the standard assumption for VAD processing that the wind field is horizontally homogenous in each layer. The general approach for the processing of VAD scans is the calculation of the 3D wind by finding the solution to a system of equations. There are two common perspectives on their definition. The first perspective operates in the (local) Cartesian coordinate system “East, North, Up”, where wind is described by the components (u, v, w) and the direction of the lidar beam (normalized radius vector (x_L, y_L, z_L)). Each measured Doppler velocity d (negative, if wind is blowing towards the lidar) satisfies the following linear equation

$$d = x_L \cdot u + y_L \cdot v + z_L \cdot w \tag{1}$$

The second perspective describes wind with as horizontal wind speed and direction and the vertical component (~~ff =~~ $v_h = \sqrt{u^2+v^2}$ horizontal wind speed, ~~dd =~~ ϕ_h wind direction, w). The Doppler velocity is then a function of the scanning directions in polar coordinates (ϕ = azimuth, θ = elevation).

$$d = \cos(\phi - \cancel{dd}\phi_h - \pi) \cdot \cancel{ff}v_h \cdot \cos(\theta) + \sin(\theta) \cdot w \tag{2}$$

As equation 1 can be transformed into equation 2, they are equivalent (see appendix). Assuming that the lidar remains stationary and has a fixed elevation angle θ (which is not the case in our setup), the equation further simplifies to

$$\frac{d}{c_1} = \cos(\phi - \cancel{dd}\phi_h - \pi) \cdot \cancel{ff}v_h + w \cdot c_2 \tag{3}$$

with the constants $c_1 = \cos(\theta)$ and $c_2 = \tan(\theta)$. Wind speed and direction can then be determined by a cosine fit for all avalaible scan directions. Although the equations 2 and 3 are more intuitive, and our lidar software already uses the parameters elevation and azimuth, we found it is easier to work in a Cartesian coordinate system to apply corrections and thus choose equation 1. Since we have eight rays per VAD (and more than one measurement per ray in each layer) we get a system of linear equations. Given a measured set of Doppler velocities d_i $(i = 1, ..., n)$ in directions (x_i, y_i, z_i) (east, north, up) the wind speed (u, v, w) can be calculated by solving the overdetermined system

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \dots & \dots & \dots \\ x_n & y_n & z_n \end{pmatrix} \times \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix} \tag{4}$$

using the least squares method. To ensure the quality of the data we added the condition that at least six out of eight azimuth angles had data (that was not removed), thus at least measurements in a sector of 270° were available.

As the system of equations is only solved approximately for a given a solution (u^*, v^*, w^*) we can define a measure for the goodness of the fit. Päschke et al. (2015) define the coefficient of determination. We define the fit deviation in our paper as:

$$\left\| \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \dots & \dots & \dots \\ x_n & y_n & z_n \end{pmatrix} \times \begin{pmatrix} u^* \\ v^* \\ w^* \end{pmatrix} - \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix} \right\|_2 \tag{5}$$

For the purpose of comparing the fit deviation, only scans with the same elevation should be used. It should also be noted that measuring a non-homogenous or non-stationary wind field would result in a larger fit deviation value.

In Figure 4 we show the amount of computed wind speed / direction data from VAD scans for different SNR thresholds. The increase of computed data stagnates around -30 dB. A further decrease of the SNR threshold adds only data that is thrown out again by the "100m bin method" or for other reasons. One can also see the zig-zag artefact that is produced by this 100m-bin combined with computing winds every 50 m. It its more dominant in case of PS96 as the measurement was taken every 3 m while the measurements for PS85 were taken every 18 m. The benefit of using additional scans with 25° and 50° elevation for PS85 can be seen for the lowest 750 m, if a higher SNR threshold is chosen. The choice of SNR threshold for this paper is explained in the next section.

2.3.3 Choice of signal-to-noise ratio thresholds

SNR-based thresholds for the separation between reliable and unreliable data points are a common technique for lidar data processing (Päschke et al., 2015; Pearson et al., 2009; Frehlich and Yadlowsky, 1994; Barlow et al., 2011). This value can vary depending on the instrument specific performance (detector noise) and the variability of atmospheric conditions within the measured volume. The recommendation of the manufacturer for the lidar is -18.2 dB. However, Päschke et al. (2015) showed that this value is rather conservative and reduces the amount of data by up to 40% (between -20 dB and -18.2 dB). Hirsikko et al. (2014) use a threshold of -21 dB and state that -25 dB could still suitable for horizontal wind measurements. Pearson et al. (2009) find experimentally a ~~threshold SNR~~ SNR threshold for reliable data of ~~23~~ -23 dB. The potential SNR threshold was already considered during our measurements by adjusting telescope focal length of the lidar and the integration time (following the recommendations Hirsikko et al. (2014)). This is necessary during the measurements, since raw data of single pulses ~~are~~ were not stored and thus no postprocessing is possible. Figure 4 shows the sensitivity of available data for PS85 and PS96 on the SNR threshold. We find a similar reduction as Päschke et al. (2015). A rule of thumb for our measurements seems to be increasing the SNR threshold by 1 dB results in a (relative) loss of 5–10% of the data. Additionally due to the different averaging time for each ray during PS85 and PS96 (1.5 vs 12–15 sec), the PS96 data ~~contains less noise and thus~~ allow for a lower SNR threshold compared to the PS85 data, because averaging over a longer period given the same SNR results in better data. Thus, it makes sense to choose a ~~different~~ less strict SNR threshold for ~~each data set~~ the PS96 data set to make both data sets more comparable. Päschke et al. (2015) checked the measured wind speed of vertical stares. Knowing that ~~these had to be around 0 m s^{-1},~~ vertical velocities are close to zero, Päschke et al. (2015) could evaluate the influence of noise ~~could be evaluated. We~~ from vertical stares for quiescent atmospheric conditions. As we did not have a stabilizing platform, the evaluation of the vertical stares is not possible because of the influence of horizontal wind on the signal. To circumnavigate this problem, we follow a similar approach and evaluated the Doppler velocity from all individual rays for VAD scans with an elevation of 75° (only the first data points near the lidar were removed; see subsection data processing). Since the Doppler velocity due to horizontal wind speed is less than 26% at this elevation, the range of realistic Doppler velocities should be ± 10 m s^{-1}. Data points outside this range can be regarded as ~~noise~~wrong (or empirical noise). This condition is used to find a SNR threshold in a three-step procedure. First, we look at the overall frequency distribution of measured Doppler velocities (Fig.

5, top). We assume that the data mainly consists of two parts: the empirical noise (homogenous along all wind speeds; top to bottom) and the wind signal (relatively homogenous along the signal intensity or SNR; left to right). Signal intensity is defined as SNR+1. All points above 10 m s^{-1} or below -10 m s^{-1} are taken to construct ~~a~~ an empirical noise distribution as a function of intensity using the mean value (Fig. 5, bottom). ~~We call this the empirical noise.~~ In the second step, we take the ratio of the empirical noise and the mean of the measured Doppler velocities for each intensity, which results in an empirical noise fraction (plotted as solid line in Fig. 5, bottom). The empirical noise fraction is close to zero for high intensities and starts to increase rapidly at different SNR values for both data sets. We choose a SNR threshold (step three) of -17 dB for PS85 and -20 dB for PS96. This empirical SNR threshold results in about 14%/26% of usable raw data for PS85/96. Comparing this to the resulting VAD percentages 14%/21% (Fig. 4) it should be noted that the decrease for PS96 comes mostly from the restriction sd(yaw) <2° and sd(roll/pitch)<0.5°. Without this condition, the computed VAD percentage is 25%.

3 Results

A verification of the lidar wind data is presented in the following by comparisons with radiosondes and ship measurements. For the statistics of wind direction the absolute values of the differences are adjusted to be smaller than 180° to avoid the discontinuity at northerly directions (e.g. a difference of 270° becomes -90°). For the correlation of wind direction we used the correlation coefficient for angular variables (Jammalamadaka and Sarma, 1988). Radiosonde data was interpolated linearly with height to match the lidar data. Lidar wind speed and direction was first computed for every VAD and then averaged over a 20 min interval centered around the launch time (plus 100 sec) of the radiosonde (100 sec after the start the radiosonde is at a height of around 500 m). We excluded all data points with wind speed < 0.5 m s^{-1} for the statistics of wind direction, but this condition was only met during PS96 and only for up to six data points at different heights/times. Figure 6 shows the calculated ~~RMSD~~ root-mean-square deviation (RMSD) and bias by height for different SNR thresholds. While ~~23~~ -23 dB leads to some larger differences particularly for PS85, our empirical thresholds of ~~20 dB and 17~~ -20 dB and -17 dB are found to be reasonable. Furthermore, a systematic dependence on height is not present. We also check for a height dependence of the correlation (not shown), but there was none present. At heights above 1000 m the sample size is relatively small and differences between different SNR thresholds are not robust.

The overall statistics of the radiosonde comparisons is shown in Table 2. Although our data set is smaller than that of Achtert et al. (2015) we find similar results (RMSD for wind speed around 1 m s^{-1} and wind direction around 10°) except for our larger bias in wind direction. This bias persists even when applying a stricter condition for the allowed standard deviation of yaw angle during the measuring/averaging time (last row in Table 2). The bias for the wind speed is very small.

In order to quantify the impact of changes in our standard data processing, the effects of changing the layer thickness and changing the averaging time around the radiosonde launch were investigated. Table 3 summarizes the ranges of these effects RMSD, bias and R^2. None of these changes had any relevant influence. We also computed the 95% confidence interval bounds for the bias for wind direction which was found to be only up to 1° higher/lower than the biases given in Tables 2 and 3.

As mentioned above, our results are similar to Achert et al. (2015), who used a motion stabilized platform and found mean bias for wind of 0.3 m s^{-1}, and a mean standard deviation of 1.1 m s^{-1} and 12° for wind speed and direction, respectively. However, our bias for wind direction is larger than the value of 2° found by Achert et al. (2015). As described in the section 2, the lidar was aligned with the ship's axis only by eye. We tried to estimate the yaw offset by checking the correlation of the roll and pitch 1 Hz data from the AHRS (resp. lidar) and the ship navigation system. By assuming a yaw offset and correcting the roll and pitch angles, we determined the peak of the correlation. The results depend largely on the chosen time window and scattered between -5° to 5°. Overall, this can explain a lidar yaw offset of around -0.5° for PS85 and +1° for PS96, leaving the question of the observed 5° and 7° bias compared to radio soundings. To investigate this further we compared the winds measured on the crow's nest of the ship (König-Langlo, 2014b, 2016b). There are two anemometers (2D-sonic anemometers, one at each side, König-Langlo et al. (2006)) mounted at a height of around 39 m above sea level. The first usable data points of the lidar measurements are at approximately 50 m height. Comparing the wind direction measured by the lidar in 50 m with wind direction in 60 to 200 m, we found an overall linear increase (decrease) of wind direction with height during PS85 (PS96). Assuming this change of wind direction is also present between the 39 m anemometer and the lidar data (approx. 50–75 m) this could lead to a slight positive (negative) bias during PS85 (PS96) of about 1°. An overview is shown in Figure ~~?? and~~ 7 and 8 and the statistics computed for this comparison are shown in Table 4.

Overall lidar and ship (anemometer) measurements agree well. In case of PS96 the comparison of the lidar to the ship anemometers suggests that the determined bias compared to radio soundings is also present. However, the anemometers are also disturbed by the ship's superstructure depending on wind direction (see section 2). One obvious explanation for the bias would be a misalignment of the lidar with respect to the ship. As an offset of 6° should be visible by eye and is not confirmed by the analysis of the inclinometer correlation, the reason for the bias is still unclear.

4 Case Studies

In the following, we present two case studies. The first one focuses on the choice of the SNR threshold and the second one underlines the added value of lidar measurements compared to standard ship anemometer and radio sounding data.

4.1 PS85 - Arctic 2014/06/12

The beginning of the 12 June 2014 starts with wind speeds around 8.5 m s^{-1} and wind from N-NW (Fig. 9) By midday, the wind decreases down to approx. 2 m s^{-1} and the direction changes almost by 180°, thus from S-SW now. Weather charts for this day show that *Polarstern* was navigating through a synoptic high pressure ridge, which causes the measured wind changes.

The radiosonde wind profile at 1103 UTC agrees well with the lidar wind profiles at 1100 and 1109 UTC (Fig. 10), and the lidar data agree also with the ship wind measurements (Fig. 9). The potential temperature profile shows an almost neutral stratification with high humidity topped by a strong inversion at 900 m. The plot for the SNR (Fig. 9, ~~bottom~~c) shows that with the conservative SNR threshold determined by the method presented in this study (~~17dB~~ -17 dB for PS85) the wind speed decreases in the afternoon would only be partially detected. However, this decrease below 250 m seems to be highly realistic in comparison with the ship measurements. Extending the SNR threshold to -20 dB or -23 dB yields overall reasonable results, but adds also some ~~noisy pixels~~ outliers particularly at the top height of the measurements. The presented method for determining a conservative SNR threshold seems to distinguish well between reliable and unreliable data. However, for specific cases it does make sense to check manually, if the limit can be extended to gain reliable data. The fit deviation (Fig. 9, d) can help with this decision, but notice that the high relative fit deviation in the afternoon stems mostly from the low wind speeds. Note that the height difference between lidar and ceilometer from 0800 to 1200 UTC is likely due to a thin layer of low clouds that the lidar could partially penetrate.

4.2 PS96 - Antarctic 2016/01/16 – 2016/01/17

The second case study is located in the Antarctic during PS96 (Fig. 11). It is chosen because it presents a situation of a stable boundary layer (SBL) with low-level jets (LLJs). The first LLJ was measured close after midnight at the 17 January 2016 between 0030 and 0230 UTC, and a second LLJ a few hours later between 0530 and 0730UTC, and the third LLJ between 1000 and 1130 UTC (Fig. 11, ~~top~~a). The LLJ wind speeds reached a maximum of up to 14 m s^{-1} at a height of 200 m (Fig. 11, ~~top~~a). Three radio soundings are available at 16 January 1700 UTC, and for 17 January at 0700 and 1200 UTC. Only the profile at 0652 UTC on 17 January captured one of the LLJs (Fig. 12). The radiosonde profile agrees well with the lidar winds. The LLJ is located at the top of a surface inversion, and is associated with a strong directional shear in the lowest 200 m. It has to be noted that the ship was orientated perpendicular to the wind for this radiosonde launch, so that the ship's influence on the radiosonde winds was minimized for this LLJ situation. The short duration and fast developments of the LLJs illustrate the benefit of vertical wind profiles with high temporal resolution. The dynamics of the LLJs were not studied in detail. They occurred during the passage of a synoptic front, when the ship operated in a polynya the lee of a huge iceberg (A23A, size about 60 km x 80 km). Baroclinicity is therefore a likely reason for the LLJs. While LLJs caused by inertial oscillations are frequent in the Weddell Sea during winter (Andreas et al., 2000), the observed jets during PS96 are comparable to the situation of the summertime Arctic Ocean, where Jakobson et al. (2013) find mostly baroclinic jets associated with transient cyclones.

5 Conclusions

We presented a verification of wind speed profiles measured by a wind lidar without a stabilizing platform during two cruises of the research vessel *Polarstern* in the Arctic and Antarctic. The ~~lidar was not motion-stabilized, but~~ ship motions and the ship's orientation were measured by the ship's navigation system and by a high-frequency Attitude Heading Reference System. This is the first time that a wind lidar was operated on an icebreaker in the Antarctic. A processing chain including quality control tests with a new empirical SNR threshold method and an error quantification is presented. The wind calculation is generally based on VAD scans with eight directions (rays) at an elevation angle of 75° (an elevation of 85° was discarded after a short test period), thus there is a high oversampling which allows for additional quality tests. Wind retrievals from scans at multiple elevation angles elevation angles (25, 50 and 75°) slightly improve the quality of the wind profile, but take more time. The low aerosol concentrations in polar regions result in a low backscatter. As a strategy to optimize the backscatter signal for these conditions the adjustment of telescope focal length of the lidar and the averaging time is useful. We present a processing chain for the data, which includes a quality control for each ray and a method for deriving an empirical SNR threshold. This threshold can be calculated for individual measurements setups (e.g. different number of rays, averaging time), and robust thresholds of -17 dB and -20 dB are found for the Arctic and Antarctic cruise, respectively. Due to the oversampling, an error estimation of the lidar winds can be made, which can be used as additional quality criteria. The lidar wind profiles were compared with the routine meteorological measurements of the ship and radiosonde data. Overall, the radiosonde comparisons yield similar results as found in in Achtert et al. (2015) using as motion-stabilized lidar. The wind speed bias is very small (0.1 m s^{-1}) for our standard data processing and the RMSD is about 1 m s^{-1}. For wind direction, the RMSD is about 10°, ~~but we also~~ which is comparable to other studies. The mean bias between radiosondes and lidar is about 5°. This is higher than the value of 2° found by Achtert et al. (2015), who find a bias of 5° ~~. In conclusion, the~~ only at higher levels, which is explained by the drift of the radiosonde and the resulting decrease in collocation of the measurements. Overall the results of the postprocessing of non-motion stabilized lidar data achieve comparable good quality as the motion-stabilized lidar study of Achtert et al. (2015). ~~The~~ As our study focuses only on horizontal winds it should be noted that the influence on vertical wind and turbulence measurements is higher and was not evaluated. The need of a motion stabilized lidar for those measurements could be very important. The comparison with the routine wind measurements of the ship at 40 m height yields a larger data set and a similar bias and RMSD. The choice of a longer averaging time is preferred as it allows to reduce the SNR threshold and thus increases the amount of data. For longer averaging times the influence of the ship's movement can be higher, but this effect is small in our case because the ship operated mainly in sea ice where wave heights are relatively small. It has also to be considered, that the wind field around the ship is influenced by the ship's superstructure, particularly if the ship is orientated into the wind. As this often occurs for radiosonde launches during the ship cruise, the lowest 50 m of the radiosonde wind profile should not be used for these situations. Turning the ~~wind~~ ship perpendicular to the wind is desirable. The two case studies show that for special situations a flexible change of the SNR threshold can be beneficial, and that ABL phenomena like short-lived LLJs are generally not captured by the routine radio soundings. The lidar with a high temporal resolution of 10–15 min can detect these phenomena, and would be ideally combined with a temperature profiler with a similar resolution. Alternatively, the lidar measurements can guide dedicated radiosonde launches during future campaigns, since e.g. LLJs can be detected in real-time with the lidar. For conditions with low backscatter due to the low aerosol concentration as it is typical for the polar regions, a possibility to optimize the averaging time of the lidar would be the storage of the of the raw data (spectra) for post-processing.

Appendix A

Given a measured Doppler velocities d (negative if wind is blowing towards the lidar) in normalized directions (x, y, z) (east, north, up) and the wind speed (u, v, w) we have the following equation:

$$d = x \cdot u + y \cdot v + z \cdot w \tag{A1}$$

Transforming the wind (u,v,w) to (~~ff =~~ $v_h = \sqrt{u^2+v^2}$ horizontal wind speed, ~~dd =~~ ϕ_h wind direction, w) with ~~ff = dd = 0~~ $v_h = \phi_h = 0$ if $u = v = 0$ we get

$$d = x \cdot \left(\cos(-\mathrm{dd}\phi_h - \frac{\pi}{2}) \cdot \mathrm{ff}v_h\right) + y \cdot \left(\sin(-\mathrm{dd}\phi_h - \frac{\pi}{2}) \cdot \mathrm{ff}v_h\right) + z \cdot w \tag{A2}$$

Transforming the direction (x,y,z) to (θ = elevation angle, ϕ = azimuth angle starting north and turning clockwise) with $\phi = 0$ if $\theta = \pm 90° = \pm\frac{\pi}{2}$ we get

$$d = \left(\cos(-\phi + \frac{\pi}{2}) \cdot \cos(\theta)\right) \cdot \left(\cos(-\mathrm{dd}\phi_h - \frac{\pi}{2}) \cdot \mathrm{ff}v_h\right) + \left(\sin(-\phi + \frac{\pi}{2}) \cdot \cos(\theta)\right) \cdot \left(\sin(-\mathrm{dd}\phi_h - \frac{\pi}{2}) \cdot \mathrm{ff}v_h\right) + \sin(\theta) \cdot w \tag{A3}$$

Simplifying we get

$$d = \left(\cos(-\phi + \frac{\pi}{2}) \cdot \cos(-\mathrm{dd}\phi_h - \frac{\pi}{2}) + \sin(-\phi + \frac{\pi}{2}) \cdot \sin(-\mathrm{dd}\phi_h - \frac{\pi}{2})\right) \cdot \mathrm{ff}v_h \cdot \cos(\theta) + \sin(\theta) \cdot w \tag{A4}$$

Using the trigonometric formula $\cos(a-b) = \cos(a)\cdot\cos(b) + \sin(a)\cdot\sin(b)$ we get

$$d = \left(\cos(-\phi + \frac{\pi}{2} + \mathrm{dd}\phi_h + \frac{\pi}{2})\right) \cdot \mathrm{ff}v_h \cdot \cos(\theta) + \sin(\theta) \cdot w \tag{A5}$$

Simplifying we get

$$d = \cos(\phi - \mathrm{dd}\phi_h - \pi) \cdot \mathrm{ff}v_h \cdot \cos(\theta) + \sin(\theta) \cdot w \tag{A6}$$

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The measurements were performed during the two *Polarstern* cruises PS85 and PS96 funded by the Alfred-Wegner-Institute under Polarstern grants AWI_PS85_01 and AWI_PS96_03. The research was funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung – BMBF) as part of the project ’Laptev-Sea Transdrift’ under grant 03G0833D and by the SPP 1158 ‘Antarctic research’ of the DFG (Deutsche Forschungsgemeinschaft) under grant HE 2740/19.

From all the software that was used we would like to highlight R, RStudio and the R-packages “doParallel” and “data.table”. We thank our colleague Clemens Drüe for many useful discussions and our colleage Uwe Baltes for help with the cruise preparations.

References

Achtert, P., Brooks, I. M., Brooks, B. J., Moat, B. I., Prytherch, J., Persson, P. O. G., and Tjernström, M.: Measurement of wind profiles by motion-stabilised ship-borne Doppler lidar, Atmospheric Measurement Techniques, 8, 4993–5007, https://doi.org/10.5194/amt-8-4993-2015, 2015.

Andreas, E. L., Claffy, K. J., and Makshtas, A. P.: Low-Level Atmospheric Jets And Inversions Over The Western Weddell Sea, Boundary-Layer Meteorology, 97, 459–486, https://doi.org/10.1023/a:1002793831076, 2000.

Banta, R. M., Pichugina, Y. L., and Brewer, W. A.: Turbulent Velocity-Variance Profiles in the Stable Boundary Layer Generated by a Nocturnal Low-Level Jet, Journal of the Atmospheric Sciences, 63, 2700–2719, https://doi.org/10.1175/jas3776.1, 2006.

Barlow, J. F., Dunbar, T. M., Nemitz, E. G., Wood, C. R., Gallagher, M. W., Davies, F., O'Connor, E., and Harrison, R. M.: Boundary layer dynamics over London, UK, as observed using Doppler lidar during REPARTEE-II, Atmospheric Chemistry and Physics, 11, 2111–2125, https://doi.org/10.5194/acp-11-2111-2011, 2011.

Berry, D. I., Moat, B. I., and Yelland, M. J.: Airflow distortion at instrument sites on the FS Polarstern, techreport, Southampton, Southampton Oceanography Centre, 36pp. (Southampton Oceanography Centre Internal Document, 69), 2001.

Brooks, I. M., Tjernström, M., Persson, P. O. G., Shupe, M. D., Atkinson, R. A., Canut, G., Birch, C. E., Mauritsen, T., Sedlar, J., and Brooks, B. J.: The Turbulent Structure of the Arctic Summer Boundary Layer During The Arctic Summer Cloud-Ocean Study, Journal of Geophysical Research: Atmospheres, 122, 9685–9704, https://doi.org/10.1002/2017jd027234, 2017.

Calhoun, R., Heap, R., Princevac, M., Newsom, R., Fernando, H., and Ligon, D.: Virtual Towers Using Coherent Doppler Lidar during the Joint Urban 2003 Dispersion Experiment, Journal of Applied Meteorology and Climatology, 45, 1116–1126, https://doi.org/10.1175/jam2391.1, 2006.

Damian, T., Wieser, A., Träumner, K., Corsmeier, U., and Kottmeier, C.: Nocturnal Low-level Jet Evolution in a Broad Valley Observed by Dual Doppler Lidar, Meteorologische Zeitschrift, 23, 305–313, https://doi.org/10.1127/0941-2948/2014/0543, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Frehlich, R. G. and Yadlowsky, M. J.: Performance of Mean-Frequency Estimators for Doppler Radar and Lidar, Journal of Atmospheric and Oceanic Technology, 11, 1217–1230, https://doi.org/10.1175/1520-0426(1994)011<1217:pomfef>2.0.co;2, 1994.

Hirsikko, A., O'Connor, E. J., Komppula, M., Korhonen, K., Pfüller, A., Giannakaki, E., Wood, C. R., Bauer-Pfundstein, M., Poikonen, A., Karppinen, T., Lonka, H., Kurri, M., Heinonen, J., Moisseev, D., Asmi, E., Aaltonen, V., Nordbo, A., Rodriguez, E., Lihavainen, H., Laaksonen, A., Lehtinen, K. E. J., Laurila, T., Petäjä, T., Kulmala, M., and Viisanen, Y.: Observing wind, aerosol particles, cloud and precipitation: Finland's new ground-based remote-sensing network, Atmospheric Measurement Techniques, 7, 1351–1375, https://doi.org/10.5194/amt-7-1351-2014, 2014.

Jakobson, L., Vihma, T., Jakobson, E., Palo, T., Männik, A., and Jaagus, J.: Low-level jet characteristics over the Arctic Ocean in spring and summer, Atmospheric Chemistry and Physics, 13, 11 089–11 099, https://doi.org/10.5194/acp-13-11089-2013, 2013.

Jammalamadaka, S. and Sarma, Y.: A correlation coefficient for angular variables, Statistical Theory and Data Analysis 2, 349pp, North Holland, 1988.

König-Langlo, G.: Upper air soundings during POLARSTERN cruise PS85 (ARK-XXVIII/2), https://doi.org/10.1594/PANGAEA.844803, https://doi.pangaea.de/10.1594/PANGAEA.844803, 2014a.

König-Langlo, G.: Continuous meteorological surface measurement during POLARSTERN cruise PS85 (ARK-XXVIII/2), https://doi.org/10.1594/PANGAEA.839962, https://doi.pangaea.de/10.1594/PANGAEA.839962, 2014b.

König-Langlo, G.: Upper air soundings during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN), https://doi.org/10.1594/PANGAEA.861658, https://doi.pangaea.de/10.1594/PANGAEA.861658, 2016a.

König-Langlo, G.: Continuous meteorological surface measurement during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN), https://doi.org/10.1594/PANGAEA.861441, https://doi.pangaea.de/10.1594/PANGAEA.861441, 2016b.

König-Langlo, G., Loose, B., and Bräuer, B.: 25 Years of Polarstern Meteorology, WDC-MARE Reports, 4, 1–137, http://dx.doi.org/10.2312/wdc-mare.2006.4, 2006.

Kumer, V.-M., Reuder, J., Dorninger, M., Zauner, R., and Grubišić, V.: Turbulent kinetic energy estimates from profiling wind LiDAR measurements and their potential for wind energy applications, Renewable Energy, 99, 898–910, https://doi.org/10.1016/j.renene.2016.07.014, 2016.

Pearson, G., Davies, F., and Collier, C.: An Analysis of the Performance of the UFAM Pulsed Doppler Lidar for Observing the Boundary Layer, Journal of Atmospheric and Oceanic Technology, 26, 240–250, https://doi.org/10.1175/2008jtecha1128.1, 2009.

Pichugina, Y. L., Banta, R. M., Brewer, W. A., Sandberg, S. P., and Hardesty, R. M.: Doppler Lidar–Based Wind-Profile Measurement System for Offshore Wind-Energy and Other Marine Boundary Layer Applications, Journal of Applied Meteorology and Climatology, 51, 327–349, https://doi.org/10.1175/jamc-d-11-040.1, 2012.

Päschke, E., Leinweber, R., and Lehmann, V.: An assessment of the performance of a 1.5 µm Doppler lidar for operational vertical wind profiling based on a 1-year trial, Atmospheric Measurement Techniques, 8, 2251–2266, https://doi.org/10.5194/amt-8-2251-2015, 2015.

Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-E 89-GHz channels, Journal of Geophysical Research, 113, https://doi.org/10.1029/2005jc003384, 2008.

Tjernström, M., Leck, C., Birch, C. E., Bottenheim, J. W., Brooks, B. J., Brooks, I. M., Bäcklin, L., Chang, R. Y.-W., de Leeuw, G., Liberto, L. D., de la Rosa, S., Granath, E., Graus, M., Hansel, A., Heintzenberg, J., Held, A., Hind, A., Johnston, P., Knulst, J., Martin, M., Matrai, P. A., Mauritsen, T., Müller, M., Norris, S. J., Orellana, M. V., Orsini, D. A., Paatero, J., Persson, P. O. G., Gao, Q., Rauschenberg, C., Ristovski, Z., Sedlar, J., Shupe, M. D., Sierau, B., Sirevaag, A., Sjogren, S., Stetzer, O., Swietlicki, E., Szczodrak, M., Vaattovaara, P., Wahlberg, N., Westberg, M., and Wheeler, C. R.: The Arctic Summer Cloud Ocean Study (ASCOS): overview and experimental design, Atmospheric Chemistry and Physics, 14, 2823–2869, https://doi.org/10.5194/acp-14-2823-2014, 2014.

Tucker, S. C., Senff, C. J., Weickmann, A. M., Brewer, W. A., Banta, R. M., Sandberg, S. P., Law, D. C., and Hardesty, R. M.: Doppler Lidar Estimation of Mixing Height Using Turbulence, Shear, and Aerosol Profiles, Journal of Atmospheric and Oceanic Technology, 26, 673–688, https://doi.org/10.1175/2008jtecha1157.1, 2009.

Vaisala: Radiosonde RS92-SGP, data sheet, available at: https://www.vaisala.com/sites/default/files/documents/RS92SGP-Datasheet-B210358EN-F-LOW.pdf, last access: 2 January 2018, 2013.

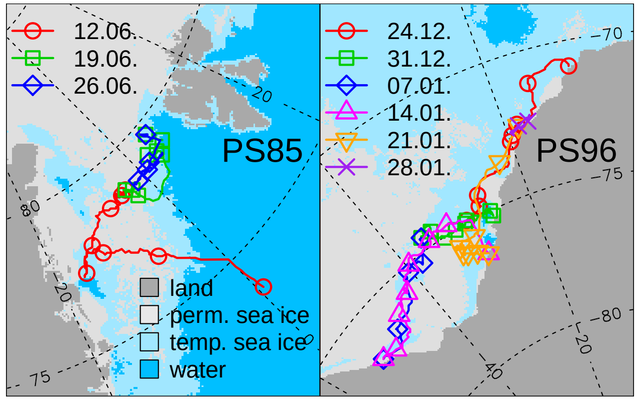


Figure 1. Cruise track of *Polarstern* during PS85 (left) and PS96 (right) with different colors for every week (symbol mark every day 0000 UTC). Beside land (dark gray) and water (dark blue) sea ice concentration (>15%) during the measuring period are shown: present every day (light gray) and present at least one day (light blue). Sea ice concentration taken from AMSR2 (Spreen et al., 2008)

Figure 2. Position of the lidar on the RV *Polarstern*

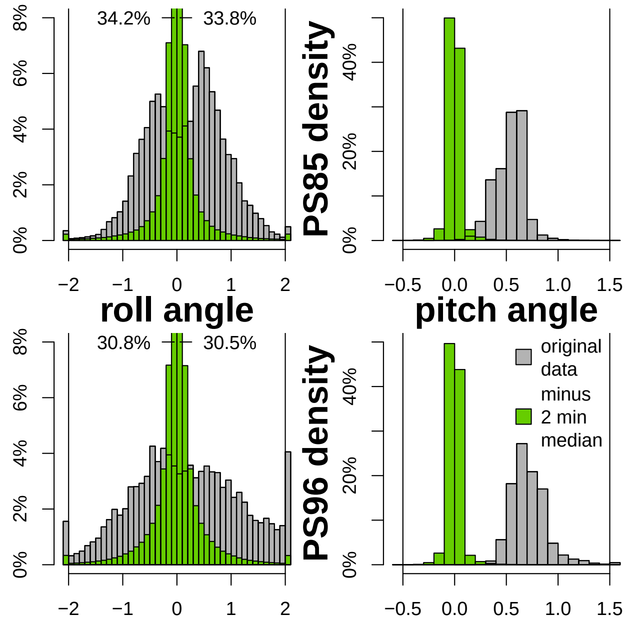


Figure 3. Frequency distribution of the ship angle (gray) and ship angle minus a 2-min running median (green) during the measurement time.

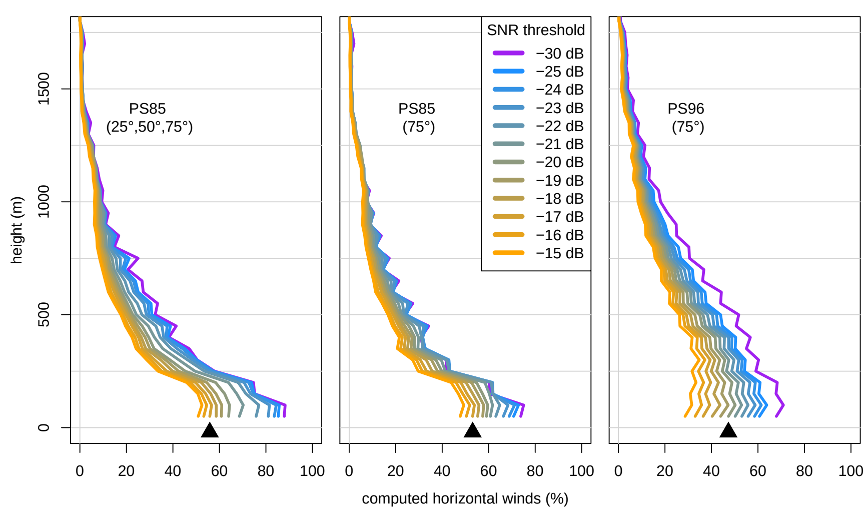


Figure 4. Percentage of wind calculations (speed/direction) from VAD scans as a function of SNR threshold at different heights during PS85 using all elevations (left), only elevation of 75° (middle) and PS96 (right). Overall number of VADs for PS85/PS96 was 3552/4250. The black triangle indicates the chosen SNR threshold based on Figure 5. For the height between 0 to 1750 m this chosen threshold results in 15% (PS85, all elevations) / 14% (PS85, only 75°) / 21% (PS96) computed horizontal winds.

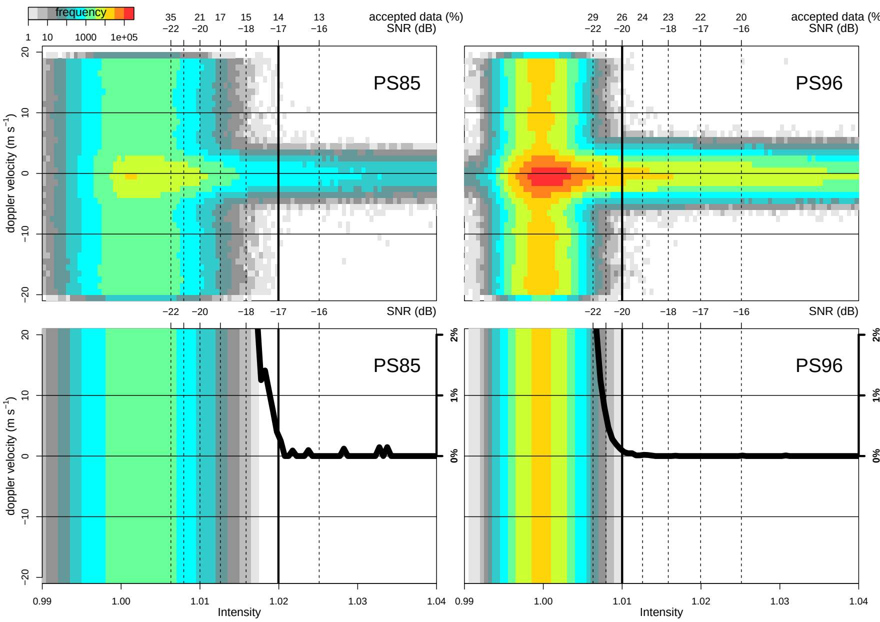


Figure 5. Top row: Frequency of Doppler velocities of VAD scans with 75° elevation depending on the intensity/SNR and for PS85 (left) and PS96 (right). Bottom row: Empirical noise computed as the mean for points above 10 m s^{-1} or below -10 m s^{-1}. The solid black line shows the ratio of empirical noise and all measured data (top) at each intensity/SNR. On the top axis it is also noted how much data would be accepted if the respective (minimal) SNR would be chosen.

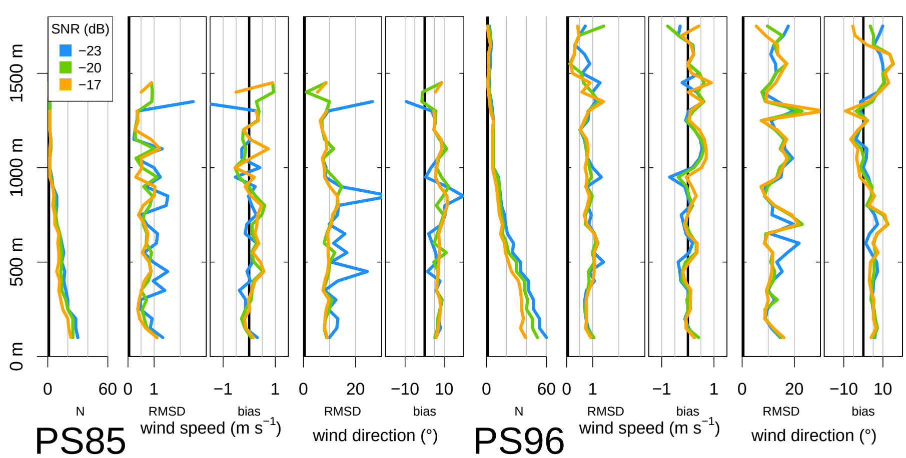


Figure 6. RMSD, bias and number of used radio soundings (N) by height of wind speed and direction for PS85 (left) and PS96 (right). Different colors show different SNR thresholds (-23 dB blue, -20 dB green, -17 dB orange). Only scans with an elevation of 75° were used.

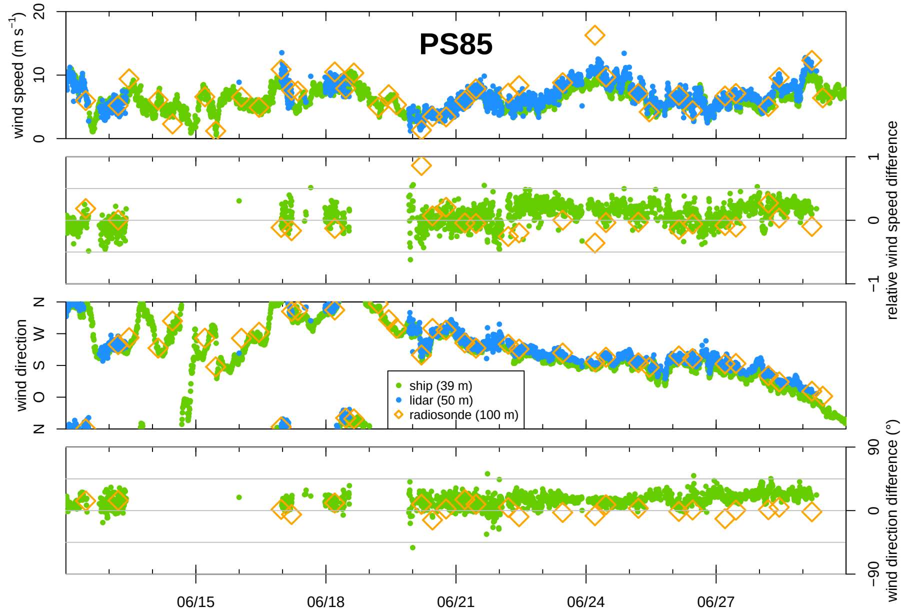


Figure 7. Comparison of wind speed ~~(ff)~~ and wind direction ~~(dd)~~ between lidar at 50 m height (blue) and ship anemometer (green) for PS85~~(top) and PS96 (bottom)~~. Radiosonde winds at 100 m are marked (orange diamond) for reference. The (relative) difference is computed as "lidar - anemometer" (divided by "anemometer"); respectively radiosounde. Only scans with an elevation of 75° were used.

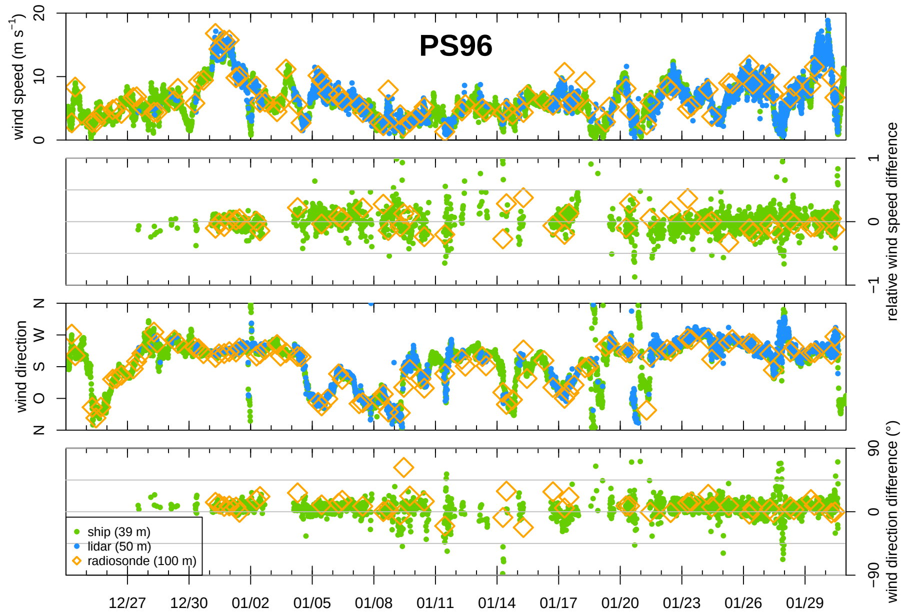


Figure 8. ~~Lidar wind speed (top) and direction (middle)~~ As Figure 7 but for ~~-23 dB SNR threshold for the 12 June 2014~~ PS96 (~~location see PS85 in Fig. 1~~Antarctic)~~. Colors below the black line (40 m) show the wind measurements of RV *Polarstern* (anemometer). The bottom plot presents the SNR thresholds that would allow for a wind calculation. Grey line is the cloud base from ceilometer measurements of RV *Polarstern*.~~

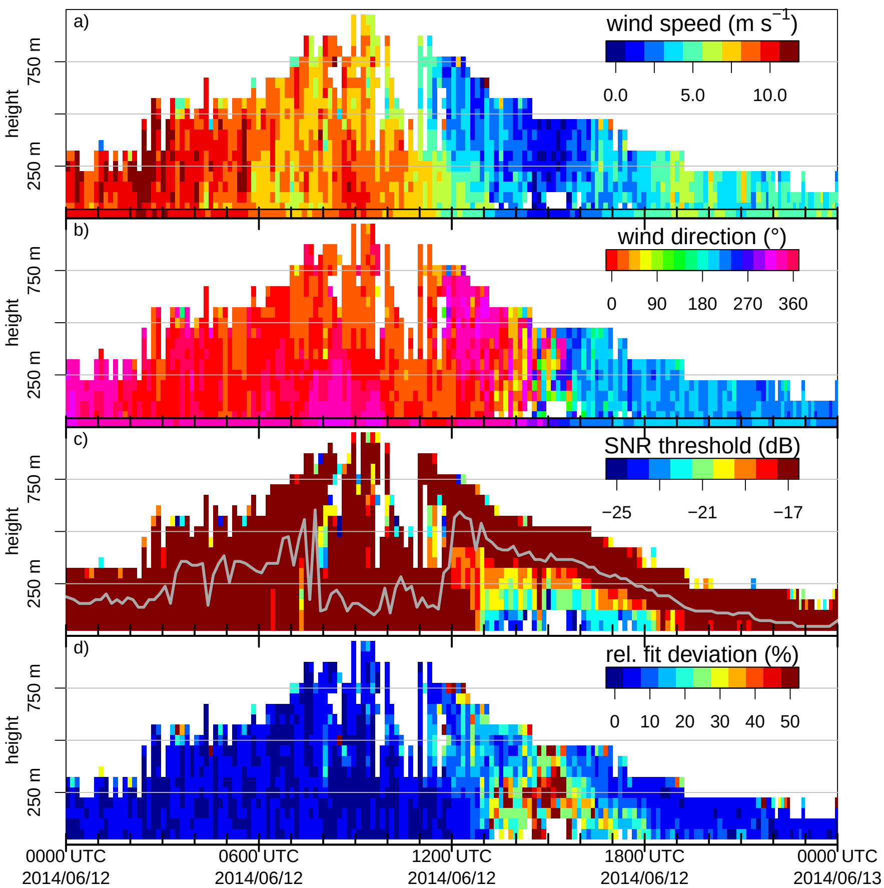


Figure 9. Lidar wind speed (a) and direction (b) for -25 dB SNR threshold for the 12 June 2014 (location see PS85 in Fig. 1). Colors below the black line (40 m) show the wind measurements of RV *Polarstern* (anemometer).The plot c) presents the SNR thresholds that would allow for a wind calculation. The grey line is the cloud base from ceilometer measurements of RV Polarstern. The relative fit deviation (fit deviation divided by wind speed) is shown in d). Values outside the colour range are plotted with the highest colour. Only scans with a 75° elevation where used

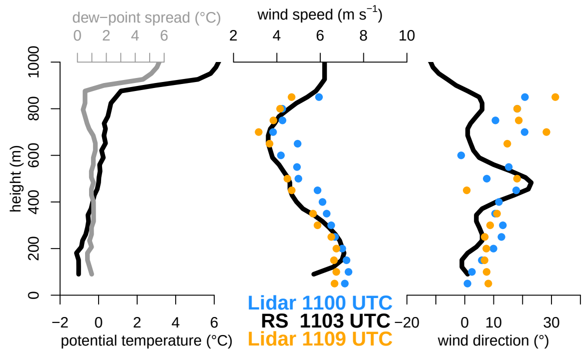


Figure 10. Vertical profiles of potential temperature~~(theta)~~, dew-point spread~~(T-Td)~~, wind speed ~~(ff)~~ and direction ~~(dd)~~ of radiosondes vs lidar wind speed and direction for around 1100 UTC 12 June 2016. A SNR threshold of -23 dB ~~was~~ and elevations of 25, 50 and 75° were used.

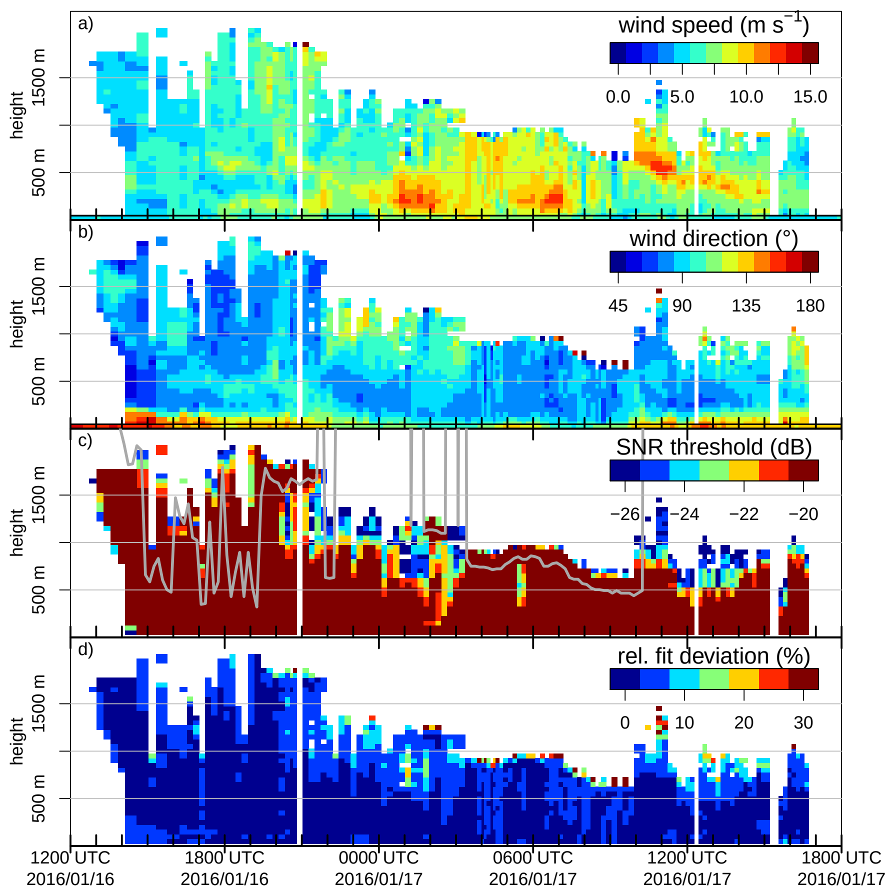


Figure 11. As Figure 9 but for 16 and 17 January 2016 (Antarctic, PS96) and with a -26 dB SNR threshold.

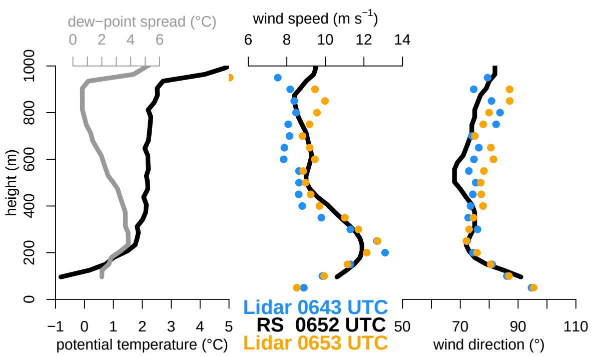


Figure 12. As Figure 10 but for the LLJ around 0700 UTC 17 January 2016 (PS96). A SNR threshold of -26 dB was used.

Table 1. Characteristics of the lidar measurements.

wavelength	1.5 μm (eye-safe, class 1m)
Gate length	18 m
Points per gate	6 (overlapping for PS96)
Band width	$\pm$19.4 m s^{-1}
Resolution	0.038 m s^{-1}
~~Threshold for signal-to-noise ratio (SNR) variable (default -20 dB)~~Measurement error	ca. 0.1 m s^{-1} (depending on SNR)
Pulse rate	10 kHz
Beam range	30–3600 m
Beam focus	variable (300–1800 m)
Averaging time	variable (1–30 s)
Scanning horizontal	0° to 360°
Scanning vertical	-15° to 90°

Table 2. Statistics for all available lidar data compared to radio soundings. M indicates the number of used radio soundings. N indicates the number of compared measurements (N is lower for the wind direction because up to six cases with wind speed < 0.5 m s^{-1} are removed). PS85 computed for -17 dB SNR threshold with only 75° elevation scans (first column, as shown in Fig. 6) and with all 25°, 50° and 75° (second column). PS96 computed for ~~20~~ -20 dB SNR threshold with default case (standard deviation of yaw angle below 2° for each ray; third column, as shown in Fig. 6) and a stricter case (standard deviation of yaw angle below 0.5° for each ray; fourth column). aR is the correlation coefficient for angular variables.

			wind speed in m s^{-1}			wind direction in deg		
	M	N	RMSD	bias	R^2	RMSD	bias	aR^2
PS85 (VAD with 75°)	28	216	0.7	0.1	0.95	9	7	0.99
PS85 (25,50,75°)	28	227	0.7	0.0	0.95	6	3	0.99
PS96 (2° yaw-sd)	58	574	0.9	0.1	0.95	13	5	0.96
PS96 (0.5° yaw-sd)	49	502	0.8	0.0	0.96	13	5	0.95

Table 3. Statistics as in Table 2, but showing the range of the statistic variables for different computations. These includes all possible combination of the following two (default marked with *): (1) the thickness of layers and thus the interpolation in height of lidar data [10, 20, 30, 40, 50* m] and (2) the time range of used lidar measurements around the radio sounding measurement (100 s after start) [± 5, 10*, 15, 30 min].

			wind speed in m s^{-1}			wind direction in deg		
	M	N	RMSD	bias	R^2	RMSD	bias	aR^2
PS85 (VAD with 75°)	27 – 28	192 – 489	0.7	0.1 – 0.2	0.94 – 0.96	8 – 9	7 – 8	0.99
PS85 (25,50,75°)	28	209 – 508	0.6 – 0.7	-0.1 – 0.0	0.95 – 0.97	6	3	0.99
PS96 (2° yaw-sd)	39 – 60	368 – 1391	0.7 – 0.9	0.0 – 0.1	0.95 – 0.96	13 – 16	5 – 6	0.95 – 0.97
PS96 (0.5° yaw-sd)	32 – 51	315 – 1226	0.7 – 0.8	0.0 – 0.1	0.96	13 – 17	5 – 6	0.94 – 0.96

Table 4. Statistics for computed lidar data points compared ship anemometer (39 m); standard case 50 m (representing approx. 50–75 m); N indicates the number of compared measurements (N is lower for the wind direction because up to 1 case with wind speed < 0.5 m s^{-1} is removed). aR is the correlation coefficient for angular variables.

		wind speed in m s^{-1}			wind direction in deg		
	N	RMSD	bias	R^2	RMSD	bias	aR^2
PS85 (VAD with 75°)	1886	1.1	0.5	0.76	19	15	0.95
PS85 (25,50,75°)	1984	0.6	0.0	0.87	11	7	0.98
PS96	2010	1.0	0.0	0.93	14	4	0.94