Response to Reviewer 3 Comments:
Blowing snow detection from ground-based ceilometers: application to East Antarctica

Alexandra Gossart[1], Niels Souverijns[1], Irina V. Gorodetskaya[2,1], Stef Lhermitte[3,1], Jan T.M. Lenaerts[4,1,5], Jan H. Schween[6], Alexander Mangold[7], Quentin Laffineur[7], and Nicole P.M. van Lipzig[1]

[1]Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium
[2]Centre for Environmental and Marine Sciences, Department of Physics, University of Aveiro, Aveiro, Portugal
[3]Department of Geosciences and Remote Sensing, Delft University of Technology, Delft, the Netherlands
[4]Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, The Netherlands
[5]Departement of Atmospheric and Oceanic Sciences, University of Colorado, Boulder CO, USA
[6]Institute of Geophysics and Meteorology, Koeln University, Koeln, Germany
[7]Royal Meteorological Institute of Belgium, Brussels, Belgium

For clarifying our answers to the referees' comments, the following scheme is used: comments of the referees are denoted in **bold**, our answers are denoted in black and quotes from the revised text are in *italic*. Please note that reference to figures in the answer refer to the original manuscript, or to the improved figure displayed in the Response document. Figures referenced in the italic text are relative to the new manuscript.

General comments:

1. The evaluation of the performance of the detection algorithm is based on the comparison with human observations at Neumayer III. The different classes (occurrence or not) may be unbalanced (much more cases without blowing snow than with, as suggested on l.14, p.13) requires more robust statistics than the one used. There is a lot of literature about what criteria can be employed for such confusion matrices. See for instance Allouche et al. (2006). I suggest the authors to sue such commonly used statistics (e.g. Cohen's kappa, true skill statistics) for the evaluation of the algorithm. The estimated depth is not really evaluated, what would be needed to do so.

Thank you for this excellent remark. I have used the statistics indicated in Allouche et al. (2006) (see answer to Question 6 below). Regarding the depth of the layer measured by the ceilometer, it is likely underestimated, due to the attenuation of the signal. A way to evaluate the layer height, is to compare height measured by the BSD to satellite detections (data from Palm et al. (2011)), when concurrent. This is part of an ongoing work and not included in this paper.

2. The statistics derived from the outcome of the blowing snow detection algorithm are informative and relevant, but they could be more complete, by including data and analysis about the inter- and intra-event variability of the blowing

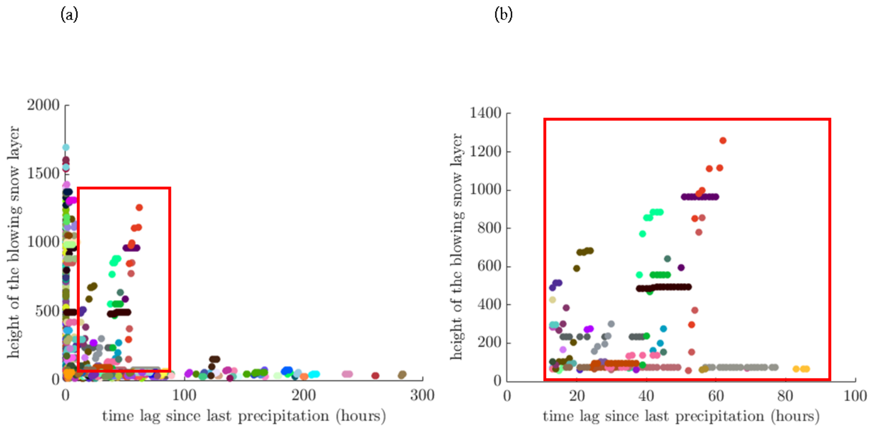

Figure 1. Inter- and intra variability of blowing snow layer height. Each color represents a distinct blowing snow event.

snow occurrence and depth.

We have performed this analysis on the blowing snow depth, and time versus last precipitation (Figs. 11 and 13 in the original manuscript), but the results showed no real inter variability between the events: Fig.11 was similar. The intra-variability in the layer height versus time since last precipitation (Fig. 1 above) shows two types of events: a majority of blowing snow layers of stable height, and a few events display an increase/decrease in layer heights. Given the limited amount of data and the focus of the paper, we decided to leave this out of the manuscript.

1 Question 1: P.7, l.28: the choice of smoothing the signal over 1 h should be better justified (why 1 h and not 30 min or 2 h?). The typical variability of the BS layer features should be commented (if there is a lot of dynamics within 1 h, one may loose relevant information by smoothing over 1 h).

The reason the running mean was set to 1 hour was to (1) smooth out effects of turbulence, but (2) mainly to get rid of the periodic fluctuation in the signal due to the heater switching on and off. 30 mins is not enough to smooth this out (see figure 2 below), and one hour was chosen since it is the lowest time period at which the heater artifact was substantially smoothed out. This was indeed not mentioned in the original manuscript.

We average every 15s- β_{att} profile over one hour using a running mean, to create mean attenuated backscatter profiles at every time step and avoid the variability due to turbulence and hardware noise. Figure 4. shows the resulting β_{att} at 09:30 UTC, based on the average of 240 profiles (120 preceding and 120 following 09:30 UTC). An additional reason for the integration of the signal over longer time periods, is that it improves the signal to noise ratio (SNR). No additional SNR correction is

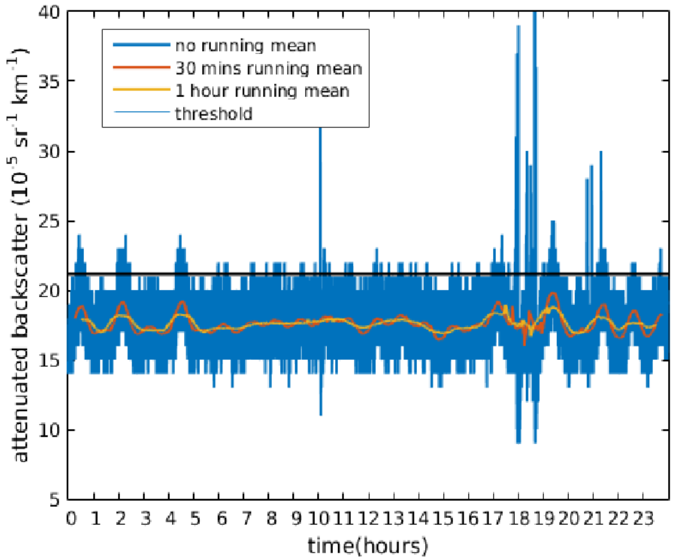

Figure 2. analysis of the periodic fluctuation visible in the ceilometer signal. Second range bin, 06.02.2013, clear sky day.

performed on the raw data, as we found that a temporal SNR higher than 0.3 would remove parts of the blowing snow signal (Gorodetskaya et al., 2015).

There are two sources of noise and artifacts affecting the ceilometer backscatter signal: the hardware of the Vaisala ceilometers, and the internal processing of the data (Kotthaus et al., 2016). Firstly, a heater is incorporated in the device to stabilize the laser temperature in cold environments. This heater is placed close to the laser transmitter and the periodic turning on (when a minimum temperature is reached by the instrument) and off (when the laser temperature is high enough) of the heater introduces a small periodic variation in the stability of the emitted signal (and therefore of the detected signal). This effect is stronger in the first range bins, closest to the device, and the hourly running mean enables to smooth out most of this signal variation.

2 Question 2:P.8, l.1: “SNR higher than 0.3": I guess it is expressed in dB. If so, it should be clearly mentioned.

The SNR is the signal-to-noise ratio, and is uniteless. It is calculated at each height range bin *j* at time step *i* as :

$$SNR_{i,j} = \frac{\overline{\beta_{i,j}}}{\sqrt{\frac{1}{2M}\sum_{k=-M}^{+M}(\beta_{i+k,j} - \bar{\beta_{i,j}})^2}} \tag{1}$$

which is the ratio of the temporal mean $\bar{\beta}_{i,j}$ and standard deviation of the attenuated backscatter over ± M time steps around time step i and range bin j (Van Tricht et al., 2014).

No additional SNR correction is performed on the raw data, as we found that a temporal SNR higher than 0.3 would remove parts of the blowing snow signal (Gorodetskaya et al., 2015).

3 Question 3: P.10, Fig.5: I probably missed something, but I do not understand why the backscatter signal from BS+precip is so much smaller (below 100 m alt) than the one from BS only (red vs blue). I would expect the two signals to sum up somehow... Or is the concentration in BS particles much smaller when there is precip? If so, what could be the explanations?

Figure 5 (Fig. 6 in the new manuscript) is indeed not clear, as we chose one day (24.04.2016) where the different typical events occur. During that day, blowing snow accompanied with cloud/precipitation unfortunately occurred at the end of the blowing snow event. Hence, the lower intensities and the resemblance to the pure precipitation profile. We have therefore selected another day (10.02.2014) to illustrate the pure precipitation, and the mixed event. In this new figure (Fig.3 below, Fig.6 in the new manuscript), the intensity of the profile in the lowermost bins is clearly indicative of blowing snow (red line), and the increase around the 15th bin indicates the presence of clouds/precipitations (arrow). In the case of precipitation/cloud without blowing snow, this low-level decrease is absent, and the increase around the 17th bin reflects the presence of a cloud/precipitation.

4 Question 4: P.10, l.19: related question: it is written “The precipitation intensity might cover the blowing snow signal", which I find confusing with the curves in Fig.5 (for the lower altitudes). To be clarified...

We know that most of the time, blowing snow happens together with storms and intense precipitation (the snowflakes rebound on the ground and are displaced by strong winds). Hence, in some cases the signal intensity is not decreasing with height, and the profile criterion is not met. In those cases, blowing snow during very intense events was discarded by the algorithm. Fig 4 (below) presents a case of blowing snow around 23:00 UTC (intense coloration). From the profiles for 23:00 UTC it is clear that the signal of the cloud eclispes that of blowing snow, even though the threshold is largely exceeded in the second bin, and the decrease in bins 2-3 is visible. As the intensity of the signal in bins 3 to 7 is larger than in the second bin, the BSD algorithm does not detect blowing snow. We therefore adapted the method, applied in the new version of the manuscript. This improved algorithm limits the decrease of the profile to cases where the precipitation signal is not intense. In cases of signal in the second bin exceding $1000 \cdot 10^{-5} \cdot \mathrm{km}^{-1} \cdot \mathrm{sr}^{-1}$ (threshold adapted from (Gorodetskaya et al., 2015)), the algorithm reports a heavy precipitation event mixed with blowing snow in all cases. A chart presenting the blowing snow detection algorithm has been added to the paper (Fig. 5 below, Fig.5 in the new manuscript).

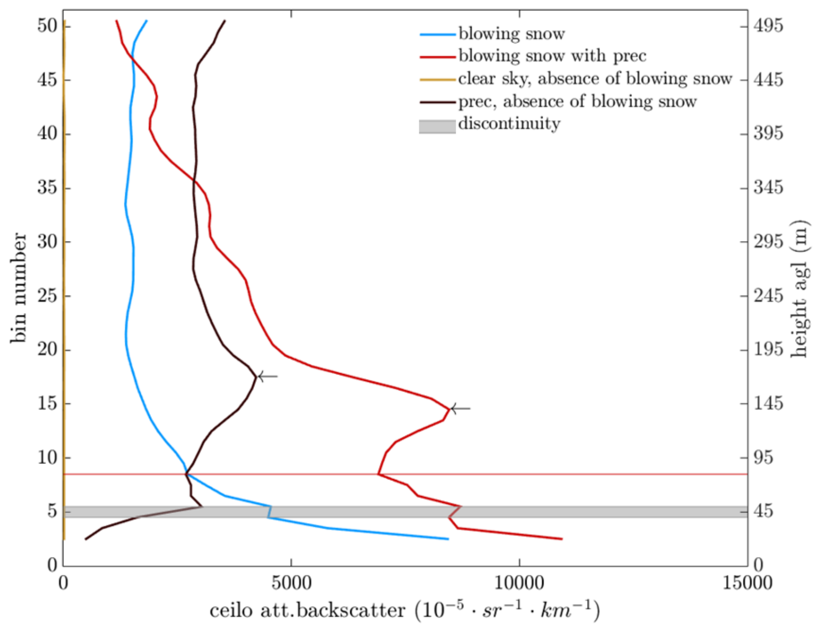

Figure 3. Different types of profiles relevant for blowing snow measured by the ceilometer at PE station: blue line - typical blowing snow signal with no precipitation nor clouds (24-04-2016); red line - blowing snow overlaid by precipitation (10-02-2016); black line - precipitation with no blowing snow (10-02-2014); yellow line - near-zero signal for clear sky conditions (24-04-2016). The height above ground is indicated on the right axis and the corresponding bin number on the left axis. All profiles exclude the lowermost bin, and start at the second bin (15 m agl.). The grey lines represent the discontinuity between bins 4 and 5 (35-45 m). The arrows indicate the presence of precipitation.

The algorithm therefore investigates the shape of the profile in order to detect blowing snow. A condition is set, that a blowing snow profile implies that the mean of the overlying bins 3 to 7 (25 to 65 m*) must be lower than the signal in the second range bin (15* m*). In this way, the discontinuity, as described in section 3.1. (visible in Figures 1 and 6 between 35 and 45 m), is not affecting our retrievals. In order to detect blowing snow occurring during clouds or precipitation, the profile shape is analyzed to identify a second increases in the signal intensity above the 7th bin (65 m height). A clear differentiation between clouds or precipitation cannot be made on the basis of the ceilometer alone, but the presence of clouds and/or precipitation can be identified. This analysis is carried out for both blowing snow and non blowing snow measurements. The information retrieved from the Micro Rain Radar (hourly precipitation rates) is collocated to ceilometer blowing snow detection, to determine the time (in hours) since the last precipitation event at PE station.*

Inherent to this profile-based method, the detection of blowing snow during precipitating events is limited to cases when the blowing snow signal is preserved close to the ground. In case of precipitation associated with storms, there is always blowing snow due to the high wind displacing the snow, and no distinction between precipitation and blowing snow is possible, as the ceilometer signal is entirely attenuated near the surface (Gorodetskaya et al., 2015), it is not possible to get signal in

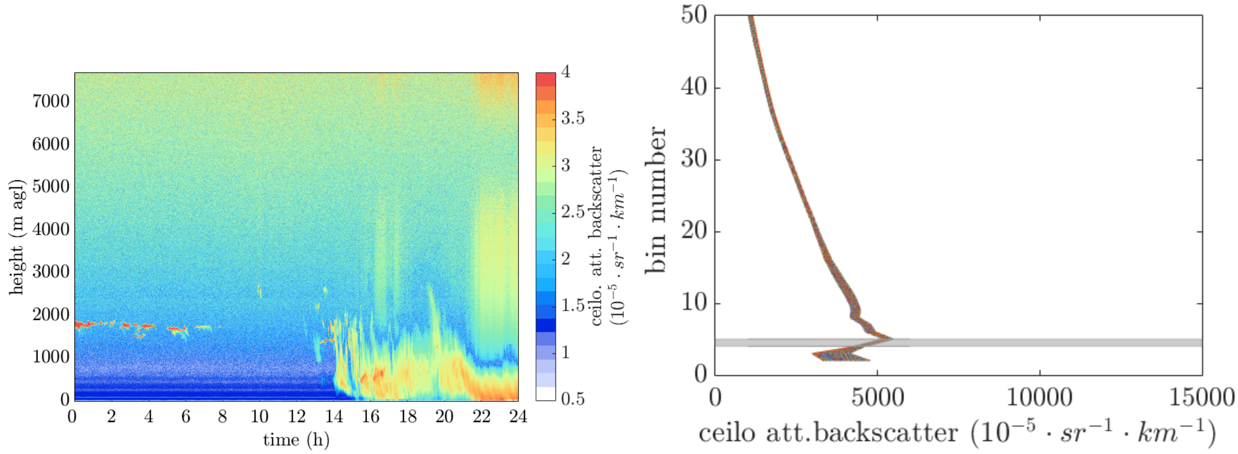

Figure 4. Quicklook presenting a mixed blowing snow event : clouds and precipitation occurring together with blowing snow around 23:00 UTC (left), and ceilometer attenuated backscatter profiles (intensity versus height) around 23:00 UTC (right).

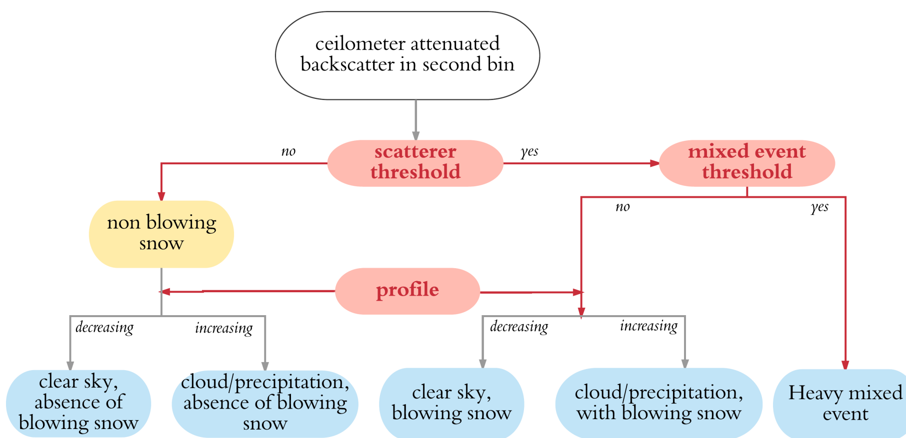

Figure 5. Chart of the blowing snow detection method

the overlying bins, and the profile of the backscatter intensity might not decrease upwards. Such intense precipitating events mixed with snowfall are identified as having a second bin signal higher than 1000 $\cdot 10^{-5} \cdot \mathrm{km}^{-1} \cdot \mathrm{sr}^{-1}$ *(threshold adapted from Gorodetskaya et al. (2015)). In those cases, the events are classified as a mixed blowing snow event, and the profile analysis is eluded by the algorithm.*

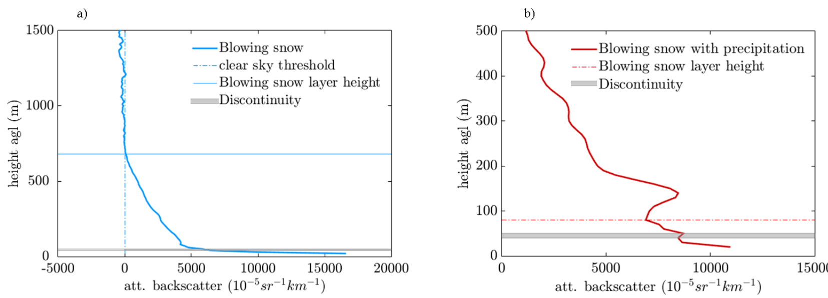

Figure 6. Determination of the height of the layer by the BSD algorithm. (a) in case of a clear sky blowing snow profile, the height of the layer is attained when the backscatter intensity reaches the clear sky threshold. (b) in case of precipitation, the height of the blowing snow layer is reached when the intensity of the backscatter signal re-increases.

5 Question 5: P.11, l.6-7: about the estimation of the top of the BS layer: it would help the reader to indicate in Fig.5 where is this limit. And how reliable would be the outcome in case of virgas?

Indeed. The graphs have been added (Fig. 6 above) to the Supplements (Fig. S3 in the new manuscript)

In our case, no distinction is made between clouds, precipitation and virga. this means that any re-increase in the profile is treated as the presence of clouds and/or precipitation.

If there is precipitation or a cloud/virga during the blowing snow event, the range gate at which the profile increases again is the top of the blowing snow layer, and the base of the cloud and/or precipitation layer (around the 7th bin in Fig.5, for the black and the red profiles). Layer height definition is illustrated in Fig. S3 in the supplements.

6 Question 6: P.12, l.7-14: better metrics could be computed to evaluate the performance of the algorithm, see General comments above.

Indeed, thank you for this excellent remark. I have used the statitics indicated in Allouche et al. (2006).

The results (Table 1 and table S4, in supplements) show a very good match in the blowing snow detection (Table. 1) and the optimum, 78 %, is reached for events classified as (heavy) blowing snow with or without precipitation. The lowest match (70 %) is found when all blowing and drifting snow is taken into account: the number of visually detected events strongly increases since more categories are included, whereas the number of detections by the BSD algorithm is fixed. In 21 % of the time, the visual observer reports something (blowing or drifting snow) that is not detected by the BSD algorithm. This is related to the fact that the ceilometer points upwards and is elevated at a height of 12-15 m *above the surface, which prevents it from detecting shallow layers of drifting snow.*

Table 1. Detection numbers and scores of the different categories of observations. The first 4 columns give N BS_{both}- stands for blowing snow detected by both the algorithm and the visual observations, N BS_{none} - when both methods agree that there is no blowing snow, N BS_{ceilo} and N BS_{vis} - represent detections by the algorithm and the observer only, respectively (the corresponding percentages are presented in table S4, in the supplement). The four last columns give the scores. B stands for blowing and D for drifting snow. The total number of measurements is 10584.

	N BS_{both}	N BS_{none}	N BS_{ceilo}	N BS_{vis}	accuracy	sensitivity	specificity	TSS
B and D snow, with or without prec	2404	5170	972	2308	0.70	0.51	0.84	0.35
B and D snow, without prec	992	6578	2373	897	0.70	0.52	0.73	0.26
heavy B snow, without prec	378	7406	2998	72	0.72	0.84	0.71	0.55
all B snow, without prec	822	6993	2554	485	0.72	0.63	0.73	0.36
all B snow, with or without prec	1856	6665	1520	813	0.78	0.69	0.81	0.51
heavy blowing snow, with or without prec	1114	7249	2262	229	0.77	0.83	0.76	0.59

A fraction of 84 % visually observed heavy blowing snow events is detected by the BSD algorithm:

$$\frac{NBS_{both}}{NBS_{both}+NBS_{vis}} \tag{2}$$

In this case, we consider visual observations reported as 'heavy blowing snow' only. For 95 % of the N BS_{ceilo} events not reported as 'heavy blowing snow' by the observer, intensities of the backscatter signal are below 1000 $\cdot 10^{-5} \cdot \mathrm{km}^{-1} \cdot \mathrm{sr}^{-1}$; it is therefore likely that those events are classified as 'slight' or 'moderate' by the visual observer instead of being considered heavy. For the N BS_{vis}, 54 % do not attain the threshold indicating the presence of scatterers and in 46 % of the cases the ceilometer attenuated backscatter profile does not decrease with height.

We also compare the skills of the BSD algorithm to different evaluation metrics (Allouche et al., 2006) (the equations for each of the metrics are presented in Supplements). The accuracy, highest for the category collecting all blowing snow events, is the proportion of correctly detected events. To take into account omission errors, sensitivity is used and the best score is attained by both heavy blowing snow categories. Specificity reflects commission errors, and the categories encompassing most events (blowing and drifting snow) perform best. Finally, since the N_{BSnone} is larger than the other categories, the matches are likely biased, and we therefore use the true skill statistics (TSS, (Allouche et al., 2006)), which is a method enabling to measure the overall accuracy and correct for the accuracy expecting to occur by chance, which also accounts for commission and ommission errors, while being independent of prevalence in the data. TSS statistics range from -1 to +1, where values under zero indicate no better performance than random, and the closest the result to 1, the better the agreement. This metric clearly indicates that heavy blowing snow is the best detected category of events.

7 Question 7: P.14, l.3: Which statistical tests have been used to check if there are “statistically significant differences"?

By means of a t-test, at the 95 % significance level. The difference for the other variables is not significant, meaning that blowing snow or non blowing snow conditions give similar distributions for these variables. However, due to the comments received, this section has been removed from the new version of the manuscript, together with Figs. 8 and 9.

8 Question 8: P.15, l.16: a minimum of description should be provided about the clustering method employed, so the reader does not have to check the reference to know what type of clustering method has been used for instance...

Some information has been added. However, since the ceilometer profile classification is also useful to discriminate between the different types of events, less emphasis is put on the cluster analysis.

The near surface atmosphere changes, associated with blowing snow events, are investigated for both stations, and detailed means and standard deviation are displayed in Table S6 and S7, in supplements. We investigate how blowing snow hourly means relate to weather regimes, derived from the hierarchical cluster analysis using PE AWS data following Gorodetskaya et al. (2013), which defines the weather regimes at PE station: "cold katabatic", "warm synoptic", and "transitional synoptic". The cold katabatic regime is characterized by slower wind speeds and lower humidity, reduced incoming long wave radiation, a slight surface pressure increase, and a substantial temperature inversion. Warm synoptic conditions involve higher wind speeds and specific humidity, strongly positive anomalies of incoming long wave radiation. The surface pressure is slightly lower, and the temperature inversion is strongly reduced than during average conditions. Finally, average wind speeds, humidity and incoming long wave radiation, as well as slightly lower surface pressure are observed during the transitional regime, when the situation evolves from synoptic to katabatic or the other way around (Gorodetskaya et al., 2013).

9 Question 9: P.16, Fig.10: the font size of the text in the figure should be increased.

The font size has been adapted accordingly (see Fig.7 below).

10 Question 10: P.16, l.14: I do not understand why the number of observations would decrease... The ceilometer is collecting data every 15s, no? The explanation should be clarified.

The sentence was wrongly phrased. "Observations" has been replaced by "measurements". The number of "observations/measurements" decreases because it is rare that there is no precipitation for a dozen of days continuously.

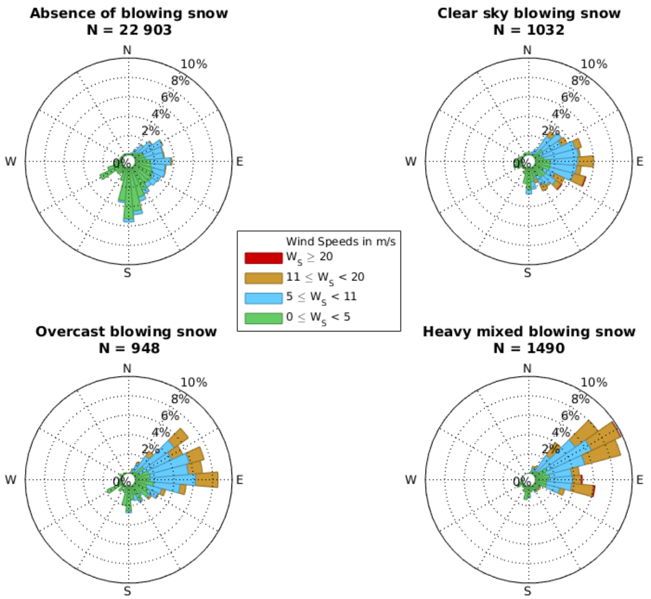

Figure 7. Wind roses presenting the wind direction for all blowing snow, blowing snow with or without precipitation , and non-blowing snow conditions at Princess Elisabeth station.

This is, however, not so obvious if we normalize the distribution of blowing snow events taking into account the total number of measurements within each time lag after precipitation [...]. A possible explanation is that there are less measurements as we go in time, and that blowing snow occurred during those measurements.

11 Question 11: P.18, Fig.13: are these distributions for the two stations (Neumayer and PE) or only one location?

This is for PE station only, since the time since the last precipitation event is only available using the micro-rain radar. The text and legend have been adapted accordingly.

We further tested the hypothesis that the height of the blowing snow layer is related to wind speed at PE station. While there is no correlation, (also found by Mahesh et al. (2003)), the height of the blowing snow layer is related to the time since last precipitation (Fig. 13). The height of the blowing snow layer can reach up to 1000 m *within 24 to 48* h *after precipitation, and 95 % of the blowing snow layers thicker than 500* m *occur shortly after the last precipitation event at PE. Blowing snow events taking place much later after the precipitation event are limited to a vertical extend lower than 100* m *thick.*

and legend :

Scatter plot of the time since last precipitation event versus height of the blowing snow layer at PE station. Each point represents a blowing snow event. The colorbar represent the data density (number of observations divided by the entire sample size).

12 Question 12. P.18, l.17: “commission errors" is repeated twice.

"Commission error" was stated twice, and should only appear once. The second mention should have been "ommission error". In our case, a commission error is a BSD detection that is not reported by the visual observer. It is similar to a "false alarm", but since we do not consider visual observations as ground truth, but as another means of measuring blowing snow, we chose the omission/commission terms. The omission error refers to missing a blowing snow occurrence that is reported by the visual observer.

Furthermore, the hourly time filtering applied leads to commission errors (events detected by the algorithm, but not reported by the visual observations) and ommission errors (short-lived events are likely removed from the running mean).

13 Question 13. P.19, l.9: I guess the same algorithm could be applied to lidar systems, no?

Indeed, the algorithm could be adapted to lidars. However, we developed the algorithm for existing instruments at PE station, which is not equipped with a lidar. Moreover, such instruments are more expensive, and less widespread as ceilometers (low-cost networks).

References

Allouche, O., Tsoar, A. and Kadmon, R.: Assessing the accuracy of species distribution models: prevalence, kappa and the true skills statistics (TSS), Journal of applied Ecology, 43, 1223–1232, 2006.

Gorodetskaya I. V., Kneifel, S., Maahn, S., Van Tricht, K., Thiery, W., Schween, J. H., Mangold, A., Crewell, S., and van Lipzig, N. P. M.: Cloud and precipitation properties from ground-based remote-sensing instruments in East Antarctica, The Cryosphere, 9, 286–304, 2015.

Gorodetskaya, I. V., van Lipzig, N. P. M., van den Broeke, M. R., Mangold, A., Boot, W., and Reijmer, C. H.: Meteorological regimes and accumulation patterns at Utsteinen, Dronning Maud Land, East Antarctica: Analysis of two contrasting years, J. Geophys. Res., 118, 4, 1700–1715, 2014.

Kotthaus, S., O'Connor, E., Muenkel, C., Charlton-Perez, C., Gabey, M. G, and Grimmond, C. S. B.: Recommendations for processing atmospheric attenuated backscatter profiles from Vaisala CL31 Ceilometers, Atmos. Meas. Tech. Discuss., 9.8, 3769–3791, 2016.

Mahesh, A., Eager, R., Campbell,J. R., and Spirnhirne, J.D.: Observations of blowing snow at South Pole, Journal of Geophysical Research, 108 (D22), 4707, 2003.

Palm, S. P., Yang, Y. Spirnhirne, J. D., and Marshak, A.: Satellite remote sensing of blowing snow properties over Antarctica, J. Geophys. Res., 116, D16123, 2011.

Van Tricht, K., Gorodetskaya, I.V., Lhermitte, S., Turner, D.D., Schween, J.H., and van Lipzig,N.P.M.: An improved algorithm for polar clouds-base detection by ceilometer over the ice sheets, Atmos.Meas.Tech., 7, 1153–1167, 2014.