Manuscript: "Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study"
Authors. M. A. Wolff, K. Isaksen, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan

Reply to Eckhard Lanzinger (referee)

Dear Dr. Lanzinger,

Many thanks for your positive review. We are happy to read, that you describe our work as valuable and of interest for the hydrological community. We address your comments individually below.

Both chapters [*chapters 1 and 2*] need some minor lingual corrections (e.g. word order).

Before sending the revised manuscript, we will ask a native-speaker for a complete language-review to help improving the language.

In chapter 3 on p. 10051 it is stated that "no significant differences occurred" in the qualitative analysis of 10 min and 60 min events. Is there any proof of this statement that could be shown? Theoretically there could be a difference, especially when temperature and wind conditions are changing or highly variable (see also p. 10066, line 13 ff.). A 60 min event represents an average over this time period. The relation between catching ratio and wind speed or temperature should be influenced if this relation is non-linear. If this effect is negligible in reality it should be explained why (small non-linearity? Small variations in temperature or wind speed at this site?).

Attached is a figure that shows all 10-min events. It is comparable to Figure 4 with catch ratio for different wind speeds during the 2011-12 winter. Note that this figure has slightly different temperature classes, and a forward scatter instrument (type Vaisala PWD 21) was used for the classification rain, sleet and snow (see section 3.3.3 (Precipitation type) in the manuscript for information and discussion). In addition, no filtering was used, comparable to Figure 3a.

Overall, the same pattern appears as in Figure 4: clear differences are seen for precipitation classified as "cold" snow (snow below -1), mixed precipitation (e.g. snow above -1 and sleet) and rain. The under-catch has a pronounced relation to temperature and a non-linear relation to wind speed. For solid precipitation the slope of the catch ratio subsides remarkably and stabilizes at ca. 20 % at 7 - 8 ms^{-1}. Thus the differences between 10 min and 60 min events seem to be very small. Because of not delivering any additional information, we decided to not overload the manuscript with this figure.

As mentioned in section 3.2 (Data filtering), variations of both, temperature and wind speed, during an event were evaluated. Events with a standard deviation smaller than 0.2 C during the event period are shown in panel e in Figure 3. It seemed most natural to weigh wind speed variations with the mean wind speed: The maximum ratio between the standard deviation of wind speed and the mean wind speed was set to 0.2 for the events shown in panel f in Figure 3.

These filter methods for temperature and wind speed variations, did not improve the catch ratio dataset from Haukeliseter, which can explain the small differences between the 10min and 60 min events.

*60 min events were also chosen to meet the criteria of easy application of the continuous adjustment function at operational weather stations only equipped with basic sensors. This will also be written explicitly and highlighted in the main text at the end of section **3.1.1.***

In section 3.2 the description of fig. 3, panel d) and e) has to be checked. There is a discrepancy between the text and the caption. From my understanding of the caption, panel f and h show all points for temperatures below -2 C, but in panel h the temperature and wind speed filter was not applied. If this is correct, it could be mentioned to support the argument why none of the filters were finally used.

We realize that both the text and caption was not made totally clear and we will change this in the revised version. At P10053, L9-10 we will write: "Panels g and h are based on the same data as in panel c, but are filtered for temperature < 0 _C and temperature < −2 _C, respectively." And for L12-16 the correct references to panels are d and e, respectively (not e and f). In addition we include a new sentence for panel f in the text: "The same filters are applied in panel f as for panel e, however only events with mean temperature below −2 C are shown."

Lingual correction in Section 3.3.2 at line 16: data was "divided" into … classes.

That will be changed in the revised manuscript.

The plots in figure 8 are very small and might be hard to look at when printed on paper. I suggest plotting only one larger graph with the resulting functions for some different temperature ranges in different colors, similar to fig. 4. If this graph tends to be overloaded, I would suggest reducing the depicted temperature classes. Despite the discussion of the residuals later on I was wondering if some explanations could be given, why the scatter of data is so different between certain temperature ranges. I also wonder, why the data points for the same temperature classes are different in fig. 8 a) and b). Shouldn't they be the same?

We agree that the panels of figure 8 became very small when displayed on a single page in landscape. We decided to not combine all temperature classes in one figure (similar to figure 4), because the graph indeed gets overloaded. We suggest to divide the figure into two (a and b) and have them printed on different pages. That would allow at least twice the size for all numbers/labels and symbols and ensures readability also for the printed version.
The difference in the amount of data points are due to the fact that a lot of gauge-height wind data were neglected because of the turbulence-affected wind measurements (see section 3.1.2).

As the article is about the correction of measured precipitation data I was looking for a graph and/ or table where the resulting improvement after correction could be seen. Eventually this could be a contour plot like fig. 10 showing the deviations from the reference with applied correction. Maybe a "before" and "after" plot in this contour style could give a good impression of where the problem is and how well it is solved.

At the time of preparing the paper, we only had very few processed data beyond those which were already used as input for the derivation of the adjustment function. Evaluating the validity of the adjustment function as well as the connected improvement by its application with these data would

not have been an independent test. Furthermore, a thorough evaluation and quantification would preferably require a detailed study of very different individual events, longer periods of various lengths and also includes data from other sites, easily filling another full-size manuscript.
We do indeed hope that part of this work can be done within the WMO SPICE effort, which gives the brilliant opportunity of numerous similar equipped sizes all around the world.

Now, one completed winter further, we did some preliminary checks on the newest data to see the effect of the application of the transfer function to independent data. We will include the results of two checked events of March 2014, representing a snow and a mixed precipitation event, respectively. Please, see the table below, which also contains the results of application of the transfer function on already analyzed data (thus not independent data) from longer periods in March 2011 and March 2012.

It is important to note, that these results are not from a thorough analysis and can't be used to exactly quantify the improvement connected to the application of the presented adjustment function, but they might give an indication of its effect. In all four cases a significant improvement could be achieved. Differences between the adjusted precipitation amount and the precipitation measured inside the DFIR are both positive and negative, which might indicate that the remaining differences are actually representing the noise/uncertainty of the method. For the two cases where the original difference was 32%, the adjusted precipitation differed less than +/-10% from the DFIR measurement. The remaining differences after adjustment of the two cases with the larger original differences (52% and 74%) are 20% and 16%, respectively.

Period	**Temp (hourly averages)**	**Wind (hourly averages)**	**Observed accum. DFIR**	**Observed accum.**	**Corrected accum**	**Difference before**	**Difference after**	**Improve-ment**
03/2011 30 days	*-25°C- +5°C*	*On average 5-15 m/s, >20 m/s for some events*	*78.8*	*53.2 (X1)*	*80.5*	*25.6 (32%)*	*-1.7 (-2%)*	*30%*
03/2012 20 days	*-10°C- +7°C*	*5 – 25 m/s*	*29.3*	*14.0 (X1)*	*23.6*	*15.3 (52%)*	*5.7 (20%)*	*32%*
19.-20.3.2014 37 hours	*-2°C- +3°C*	*6-13 m/s*	*20.7*	*14.0 (X2)*	*19.2*	*6.7 (32%)*	*1.5 (7%)*	*25%*
21.-22.3.2014 27 hours	*<-2°C*	*8-15 m/s*	*14.6*	*3.8 (X2)*	*17.0*	*10.8 (74%)*	*-2.4 (-16%)*	*57%*

On p. 10067, line 25 it is said that the uncertainty of the correction cannot be properly derived. An unknown distribution of residuals and a missing specification of regression noise are mentioned as the reason. Could eventually GUM provide a solution, because in this framework uncertainties can be derived without knowing the actual distribution of noise (by using a rectangular distribution)?

We are not familiar with the GUM framework. From the description it seems to be a non-parametric classic statistic methodology which must be powered by some mixture of rectangular distributions. As the object is to search for the distributional aspect of the regression (and the inverse regression) also for extreme outcomes, we fear this methodology will be far from representative in these extreme parts. One needs a distributional representation that is both complicated enough to catch detectable features in the data (heavy tails) and at the same time not over-complicated (i.e. to have many degrees of freedom). If we are to continue using Bayesian statistics, it is also important to use methodology that centered on distribution, not on estimation technique. As GUM seems to imply an *estimation technique-driven non-parametric version of regression, it is both in danger of being over-complex and not amenable to Bayesian analysis. We think a better way is to look for other known distributions first (such as the t-distribution) and check if these are sufficient to describe the data.*

Please note that in response to the comments from Kochendorfer some relevant parts of section 5.3 will be expanded. The t-distribution will now explicitly be mentioned as an option for further study.

Chapter 5 and 6 need more detailed lingual corrections by a native speaker (which I am not).
We will check and correct the language of these chapters when performing a complete language-review with a native-speaker as mentioned earlier.

Best regards,

Mareile Wolff and Co-authors.