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 John Kochendorfer (referee)

Dear Dr. Kochendorfer,

Many thanks for your review and the provided comments on our paper. We are very pleased that you value the scientific relevance of our research. Please find the answer to your specific comments below:

The writing needs improvement. […]. Specific suggestions are given below in the Technical Corrections portions of this comment, but they are not comprehensive. A complete review of the manuscript for such issues needs to be performed by the authors.

We appreciate the time you spent correcting the language. We'll follow your technical corrections and revise the manuscript accordingly. We will also ask a native-speaker for a complete language-review before sending the revised manuscript.

Although references are provided, more care and detail should be provided describing the wind speed and air temperature measurements. These measurements are central to the manuscript and merit description within the manuscript itself.

The following paragraph, describing shortly the temperature and wind measurements, will be added in section 2.1 (Setting).

> *[…] as one of currently 20 host-sites worldwide.*
> *Additionally, measurements of numerous other meteorological parameters are performed to support the analysis of the precipitation data. Air-temperature is measured with a pt100 element (1/10 DIN) protected by a standard Norwegian radiation screen, installed at gauge height on a tower close to the DFIR.*
> *Wind is measured by different sensors at several places around the measurement site. Standard 10 m wind measurements are performed at the tower close to the DFIR with an ultrasonic wind sensor from Gill (Windobserver II with extended heating). Three wind sensors are directly mounted to the precipitation gauges for measuring wind at gauge height (Windobserver II at precipitation gauge inside DFIR, and Young Wind Monitor SE at the two closest precipitation sensors (X1 and X2, see layout)). In 2013, a Thies Ultrasonic Wind Anemometer 3 D was installed on a separate mast at 4.5 m (gauge height) to allow measurements undisturbed by the precipitation sensor installations (see section 3.1.2).*
> *A complete list of the instruments and an evaluation of the homogeneity of the test-site can be found in Wolff et al. (2010, 2013).*

The chosen function (eq. 9) is quite complex, making it difficult to test on new datasets without employing sophisticated statistical tools. Are there simpler functions that also performed well? In the selection process was any consideration given to the number of parameters or the complexity of the function?

Deriving the functional form objectively was one of the main methodological issues in the paper. Setting the balance between a parsimonious model and good fit to the available data is always a problematic issue. Often, this key procedure is overlooked in studies.

Having decided the general characteristic of the R-V-T relationship by the half bell function and the associated temperature dependent parameter functions (with up to three levels), we set out to choose the most appropriate model of the 81 possible combinations. Note that we formulated the two-level parameter functions as simple as possible with four parameters describing only the major attributes: the two levels, the value of transition and the fuzziness of this. For model selection, we applied the Bayesian Model Likelihood (BML), which is described in section 3.4.2, first to third paragraph. In our case it was as mentioned 81 competing models, and no one was favoured a priori, at least not directly using explicit prior information. The Bayesian machinery focused in on a model that was far from the most complex one, considering that the most advanced version of Eq. (7) contains 24 unknown parameters. Note that the simplest model is also represented and tested by the BML framework, but yielding low posterior probability. The simplest model function is the one with constant values for (τ, β, θ). Furthermore, within this model form, a version seen in the literature where $\tau = 0$ and $\beta = 1$ is in fact also included.

Note also that eq. 7 appears extra formidable since we here describe the $\tau(T)$ explicitly, which appears two places in the equation.

On the question whether any assumptions on the numbers of parameters or functional complexity was given in the model selection process, one might answer: "Yes, we applied the BML". Bayesian model comparison has the so-called Occam's razor built into it, as described for instance in Jefferys et al. (1991): "…a hypothesis with fewer adjustable parameters will automatically have an enhanced posterior probability, due to the fact that the predictions it makes are sharper…"

We will make some adjustments in the manuscript to highlight the model selection:

> *Line 10, page 10058: Replace "Furthermore, it is assumed that the properties of Eq. (4) vary with temperature" with "Furthermore, it is assumed that each of the characteristics of Eq. (4) can vary with temperature. But for each property we also consider whether it might be constant for all temperatures."*
>
> *Line 12, page 10060: Add "It is known that using Bayesian model comparison, more parsimonious models are preferred to more complicated ones (see for instance Jefferys et al., 1991).*
>
> *Line 8, page 10071:*
> *Jefferys, W. H., and Berger, J. O. (1991). "Ockham's Razor and Bayesian Statistics" (Preprint available as "Sharpening Occam's Razor on a Bayesian Strop"). American Scientist* ***80****: 64–72.*

As the authors note themselves in Sec. 5.3, pg 10067, ln. 21-23, the function developed in the manuscript must be inverted before being applied to actual precipitation data (eq. 11 and 12 are raised to the power of -1). With the stated goal of creating a function available to correct measured precipitation, why was the transfer function fit to catch efficiency (PM/PT) data, rather than the inverted correction factor (PT/PM) data? Are there any significant differences between the two approaches that might affect the form of the chosen function or the uncertainty of the correction?

Due to tradition, we considered the classical catch efficiency. This is also in line with former studies, so that comparisons could be made easily. It is also worth mentioning that p_M is the measured precipitation and p_T was considered to be the true precipitation, implying that it was natural to consider p_M as a noisy function of the true precipitation rather than vice versa. It also made the process of making models more intuitive. Unfortunately, this approach turned out to make inference and uncertainty analysis unreliable.

A clear scope of the study was to express the measured response as a function of the true one. We believe that this is addressed in the paper. But it also seemed fair to point out that care is needed when in using this relationship in an inverted form, as we stated in section 5.3. A multivariate model would have served this purpose better than a regression formula. This would permit the modeling to proceed as before, except that not only measured response as a function of true response, but also true response would be associated with a distribution. Such a multivariate distribution could then be used both for expressing the distribution of measured response as a function of true response and vice versa. We will now mention this in section 5.3 (see next comment), but we also note that this was beyond the scope of the analysis.

In Sec. 5.3, pg 10067, ln. 23-26 the authors state that the results of the uncertainty analysis were unreliable. This is disappointing, especially considering the significant statistical expertise that is demonstrated in the manuscript. Could the uncertainty be better quantified (in mm and/or percent of hourly precipitation) by applying the correction to the actual PM data and comparing it to the standard PT results? I realize this may be beyond the scope of the current manuscript, but would a bootsrapping-type technique be appropriate for quantifying correction uncertainties using the available dataset?

The bootstrap is used in classical statistics for analysing the effect of parameter uncertainty in prediction. Firstly, bootstrap methods have no equivalents in the Bayesian perspective. Here all inference is based on the posterior distribution, which in our case is formulated through the MCMC samples. Secondly, we saw from these samples that parameter uncertainty contributed little to the total uncertainty of the response (the catch ratio R). We saw quite clearly that the uncertainty of the individual measurements, the error term in the regression, dominated the net effect. This fact proved to contain more complexity than our modelling framework (formulating p_M/p_T as the response and T and V as determinants in the regression) allowed for. To shed some more light on the distributional aspects we would like to insert some more text in section 5.3 (Regression noise and uncertainty analysis), p. 10067, line 27 :

> *[…], makes uncertainty analysis about the true precipitation estimate substantially unreliable at this stage.*
>
> *"While it may be sensible to model the distributional properties of the measured precipitation as a function of wind speed, temperature and the true precipitation, the objective is to use this to predict true precipitation given wind speed, temperature and measured precipitation, like formulated in in Eq. (11) and Eq. (12). These formulas however,*

only relate to estimates of R and the measured precipitation, and do not consider the distributional aspects. A multivariate model for true and measured precipitation would allow for expressing one as a distribution of the other, whether true or measured precipitation was of interest. It is also worth noting that even with a normally distributed R as the denominator, the resulting distribution will be the rather unfamiliar reciprocal normal distribution, which is heavy-tailed and bimodal. The bimodality might not be a problem as long as we require a positive R, but the tails are so heavy that the expectation are not available, making it difficult to evaluate bias. If a distribution with heavier tails was considered for the noise terms, such as the t-distribution, even more inflated tails can be expected. Medians are however preserved during monotone transformations, which should make Eq. (11) and Eq. (12) valid as median estimates."

In order to correct operational precipitation measurements, operational anemometers must be installed so that are unaffected by nearby obstructions to the wind (such as wind shields). This important point is clearly inferred from your work, but it should be stated explicitly.

We will mention that in a separate paragraph in the conclusions, section 6:

Analyses showed the importance of good data quality for successfully retrieving and applying the adjustment functions. Some of the wind measurements at Haukeliseter could be shown to be highly influenced by nearby installations, which had a negative impact on the analysis. Before having installed a less disturbed wind sensor at gauge height, an extended number of data had to be refused for the analysis. It is therefore highly recommended to use only wind measurements from sensors installed separately and undisturbed when applying the adjustment functions on precipitation measurements.

Best regards,

Mareile Wolff and co-authors.