1) Point-by-point response to the reviews

We thank the reviewers and the editor for their thorough and useful reviews.

Response to General Comments from Eva Mekis

As of possibilities for improvements, the site descriptions should be more consistent and coordinated. The use of minimum 0.25 mm threshold is acceptable, but debatable (can be lowered, as low as 0.1 mm). For the Norwegian site further description of pre-processing method would help the understanding of the results. The time period used for the analysis is missing from the description part. The individual gauge and the combined series analysis suggested to be separated into individual tables. For the individual gauge transfer function development the US-SA and NOR-SA analysis can also be added. This additional result can also be compared with the modified coefficients using the merged ALL-SA dataset. The use of "universal" transfer function is a bit ambitious and misleading in this paper, since beside the US gauges, only one additional gauge was added to the analysis representing somewhat different climate.

Authors' response: We agree with the bulk of Mekis' general comments and will change the manuscript accordingly. The site descriptions will be made more consistent, the methods used to handle the Norwegian measurements will be more thoroughly described, and the 30-min time period used for the US analysis will be described elsewhere in addition to the "Data analysis and event selection" section where it is currently described. In addition, we approve of the suggestion to create separate tables describing the combined transfer function and the individual site transfer functions and to include results within these tables describing single Alter transfer functions developed from both sites individually. We also agree that the use of the word 'universal' overstates the additional value of using two sites, and we will change the entire manuscript accordingly.

Mekis rightly commented that the use of a 0.25 mm minimum threshold is somewhat arbitrary. For example, another statistic could be used instead of the standard error to select this threshold. In addition, the threshold determined using the standard error is sensitive to both the size of the dataset being analyzed and the amount of noise in the measurements. The resultant threshold may therefore change with the length of the measurement campaign, the frequency of precipitation, the site, the gauge, and the shield. However for the sake of consistency, in the current manuscript the same threshold was selected for all the gauges evaluated. It is also true that the standard error for a threshold of 0.1 mm was quite similar to a threshold of 0.25 mm, and it is difficult to discern a significant change in the standard error in this region on Figure 3 of the manuscript under discussion. However as shown in the modified Figure 1R (included at the end of this document), where the scale of Panel c of Figure 3 has been altered to better resolve the minimum in the standard error, the standard error was indeed lower at 0.25 mm than at 0.1 mm.

Response to Specific Comments from Eva Mekis

ABSTRACT

- It is well written. I would only argue on the use of the words “remove” in the bias analysis. In 7 out of the 8 SA verification cases (3 cases at gage height and 4 cases at 10 m wind speed) the biases was reduced (or decreased) and not removed.

Authors’ response: Thank you. We agree with the comments from Mekis, and the abstract will be reworded accordingly.

CHAPTER 1

- Line 4 and 9: If possible, add more recent references here
- Line 11: Add Mekis and Vincent, 2011 reference (see at the end)
- Correct the reference: Goodison et al, 1998 (see at the end)

Authors’ response: We will add more recent references, add Mekis and Vincent, 2011, and change the year in Goodinson et al, 1997 to 1998.

- Line 28: “various measuring sites” should not be used here since in the paper only two sites are included (representing two climate types only)

Authors’ response: We will change this to “two different sites”.

CHAPTER 2

The title “Methods” of this chapter is incorrect, misleading. It includes data (metadata) of the sites, applied gauges beside the methodology.

Authors’ response: We will add a separate chapter called, “Site Descriptions”, but we believe that the section describing precipitation gauges and shields should remain within “Methods”, because the measurements recorded and the gauges used can be considered part of the methods.

CHAPTER 2.1

- Add a small map with site locations

Authors’ response: A map will be added.

- Level of details in describing the two sites should be consistent, coordinated and synchronized: MARSHALL: Add reference gauge description; add any reference where this site is already described in details and HAUKELISETER: Add typical snow regime

Authors’ response: We will make the description of the two sites more consistent.

- Line 12-14 sentence: Add reference to this statement

Authors' response: A reference to Figure 5 will be added to describe the results regarding replica gauges. The analysis supporting the statement regarding the effects of wind direction on catch efficiency was never published, so we can only reference it as, "unpublished". However we can briefly summarize the results of this analysis.

CHAPTER 2.2

- Consider adding a figure: The descriptions and all the technical details is a bit "dry" using numbers only, the comparison of the shields (for an outsider) would be easier with a sketch of the shield types.

Authors' response: Thank you. With the same intent, in the third review the editor Michael Earle asked us to include photos of the different shields below, so we will probably do this instead of including sketches, which have already been published in Rasmussen et al. (2012). In addition, a reference to Rasmussen et al. (2012) will be added.

- Identify the sites (US and/or NOR) where the given shields were installed and used in this chapter (will correspond later to tables).

Authors' response: Thank you. This is a good suggestion, which will be easily accommodated.

CHAPTER 2.3

- Line 25: Use the same notations U4.5 vs Ugh

Authors' response: We will replace U_{gh} with $U_{4.5}$.

- Identify the air temperature and wind instruments for the NOR site as well (similarly to 2.3.1), please include more details

Authors' response: We will include more details describing the measurements from the NOR site.

CHAPTER 2.4

- Figure 2 reference is missing, it is probably belong to here

Authors' response: Thank you! We will add a Figure 2 reference.

CHAPTER 2.5.1

- The period used for transfer function development is completely missing from the C3 paper (it is mentioned as "several years" in the introduction) – please include here (or somewhere in chapter 2).

Authors' response: Thank you. We will add the time periods describing measurements from both sites.

- For NOR the short summary of the method from Wolff (2015) paper would be beneficial, help the understanding. Important to know how was the 3-wire input used, or any further sensor was included (wetness sensor, PWD or any other overlapping observation).

Authors' response: Thank you. We will add a summary of Wolff et al.'s (2015) analysis methods.

CHAPTER 2.5.2

- The use of 0.25 mm lower limit for the standard (reference) gauge is debatable based on Figure 3, the SE minimum value is around 0.1 mm.

Authors' response: See Figure 1R near the end of this document, and also the response above to the general comments from Mekis.

CHAPTER 2.5.3

- Figure 4: SA data points should be better identified (period of observation, number of measurements by sites)

Authors' response: The figure caption will be augmented to include the number of measurements per site and the periods of observation.

CHAPTER 2.6.1

- Is there any NOR pair to be included?

Authors' response: We'll include a similar comparison plot for a pair of single Alter gauges from Norway. Figure 2R below shows the available measurements. There were some issues that affected the comparison of the two single Alter gauges, but we can describe these in more detail in the revised manuscript.

CHAPTER 3.2

- Table 2 and 3 are hard to read, lines are broken, difficult to find the related numbers

Authors' response: We will improve the tables. In order to fit all the values on one line and make the tables easier to read, the transfer function coefficients and the resultant RMSE and biases will be separated into two separate tables. Page breaks will also be added before the tables where necessary.

- The TF function development results and the validation is confused in Table 2-3. From Table 2 & 3 the US SA and NOR SA fitted values are missing, that should be part of these tables. Separate (or last line) should be used the verification values – RMSE and bias values computed from the SA-ALL coefficients. It would also add the possibility to compare the RMSE and Bias before/after values.

Authors' response: We will include the transfer function coefficients and the resultant RMSE and biases in separate tables. We will also produce site-specific transfer functions and include separate tables for the combined and individual site results.

- Line 14: decreased the bias, not removed; it is especially true for the NOR site

Authors' response: Thank you. We will change the wording accordingly.

- Line 25-26: easier to follow from the table, if % values are not rounded

Authors' response: Thank you. We weren't actually rounding, but describing the general results rather than the results of a specific correction type. These numbers weren't directly available in the tables, as they were the average of different results presented in the tables. We can clear this up by choosing a specific transfer function as an example, rather than trying to summarize the results of all the different correction types (eg. U_{gh}, U_{10m}, Sig, Exp).

- Lines 31-32: goes back to my original point – the algorithm at the NOR site used to create the output from the 3 wires in missing from section 2.5.1. The difference can be due to different pre-processing algorithms as well.

Authors' response: We will describe the pre-processing algorithms in the methods section.

- P13: Line 9: US SA N = 1156 (not 843).

Authors' response: Thank you. We will correct this.

- P13: Gauge height and 10 m TF analysis should be separated (in lines 6-12 the values refer to Table 2, then suddenly the conclusion is for both Table 2 and 3).

Authors' response: Thank you. We will change accordingly.

- P13 / line 15-16: Wind speed observations were available from several gauges (from 2.3.1 first paragraph), so independent gauge height wind speed measurements could have been used. Next sentence is meaningless, since the gauge height and 10m wind input are dependent (derived from each other).

Authors' response: Mekis rightly points out that the gauge height wind speeds were not actually measured at gauge height. For the transfer functions derived from only one site it is therefore true that similarities between the resultant gauge height and 10 m wind speed corrections are meaningless. We noted this on page 13, Ln 15-16.

However, at the NOR and US sites the relationships between the gauge height wind speed and the 10 m wind speed were significantly different from each other; the gauge height wind speed at the NOR site was estimated as 93% of the 10 m wind speed and the gauge height at the US site was estimated as 72% of the 10 m wind speed. Assuming that catch efficiency is more closely associated with the wind speed at gauge height, one would therefore expect the errors and biases to be more significant when using the 10 m winds to determine catch efficiency at two such sites. We could not determine this using the actual gauge height wind speeds for the transfer function development because differences in samples sizes caused by the limited uncompromised gauge height wind speed measurements actually made it more difficult to discern the effects of the two different measurement heights. We used the admittedly somewhat synthetic gauge height wind speed measurements because we wanted prove unequivocally that wind speed height affects the application of such corrections. We maintain that it is notable that we were unable to prove this.

- P14 / lines 1-6: Small variations are not as important as bringing the total closer to reality (in the context of the water budget) – this can be further highlighted.
Authors' response: Thank you. We will change accordingly.

CHAPTER 4.1
- Line 9-10: this statement is in contrast with the fact, that for this study the US gauge height wind speed was derived from 10 m; in spite of the fact, that it was available at 1.5, 2 and 3 meters. The use of 10 m wind for the combined dataset is understandable, but not necessary for the individual analysis. Additional study including different wind sources could have been completed for the US installations.
Authors' response: The discussion that Mekis objects to is based mainly on first principles, rather than our measurements. We will clarify this in the revised manuscript. Or we can certainly remove this section if the editor requests, but despite the fact that we did not use the actual gauge height wind speeds we feel we can nevertheless recommend using "the gauge height wind speed when a gauge height wind speed is available or an approximation of the gauge height wind speed". The reality is that at the NOR and US sites, for the reasons discussed above in response to Mekis' comment on P13 / line 15-16, use of the actual gauge height and near gauge height wind speed measurements to correct precipitation measurements resulted in larger errors than use of the 10 m height winds. This may in fact be a common problem due to the challenges of recording unobstructed gauge height wind speeds, but the problem is certainly not universal, nor do we believe that the actual US or NOR gauge-height wind speed measurements should be used to draw general conclusions about the accuracy of transfer functions developed using the actual gauge height wind speeds.

Additional studies were performed on the use of the different wind speed measurements, and the approach we adopted was as a result of these studies. We actually spent quite a bit of time contemplating this issue and testing different methods to evaluate the effects of different wind speed measurement heights on the resultant transfer functions. A large part of the reason we adopted the approach developed here is because we believed it would be more widely applicable, with 10 m height winds more widely available and often more representative of a monitoring site as a whole than gauge height wind speeds. The approach we developed demonstrates a practical and defensible method for adjusting the 10 m wind speeds down to the gauge height to correct precipitation measurements, and we developed this approach both to test the sensitivity of the corrections to different gauge heights and because we hoped the technique will be useful to others.

We will augment the discussion of wind speed height to clarify this.

CHAPTER 4.2

- As I mentioned earlier, this “universal” transfer function is not universal enough to justify the use of this word here, since the analysis representing only two different climates.

Authors’ response: Thank you. We will change this here and throughout the text.

- The discussion of the site-specific analysis perhaps can be replaced by the more general climate-specific analysis.

Authors’ response: Thank you. We will change accordingly, rewording this section to focus more on climate than site.

CHAPTER 5

- Line 19: The end of the sentence was not clear, a suggestion to improve: “and for various gauge/shield combinations”

Authors’ response: Thank you. We will change accordingly.

- Line 21: Verb missing at the end: Wolff et al. (2015) is presented.

Authors’ response: The active verb in this sentence is “performed”, but we can still improve the way this sentence is written.

REFERENCES:

- Line 16: Precipitation

Authors’ response: We confirmed with Mekis in a personal communication that ‘precipitation’ is spelled correctly, and there is no need to correct this.

- Please correct the reference: Goodison, B.E., P.Y.T. Louie, D. Yang (1998): WMO Solid Precipitation Measurement Intercomparison – Final Report. Instruments and Observing Methods Report No. 67, WMO/TD-No. 872, World Meteorological Organization, 212 p.

Authors’ response: Thank you. We will correct the reference.

- Nitu et al, 2016: it was a presentation – can’t be referenced.

Authors’ response: We will inquire with HESS regarding their guidelines, and remove if necessary.

- Add: Mekis, E. and L.A. Vincent, 2011: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. Atmosphere-Ocean, 49(2), 163-177.

Authors’ response: Thank you. We will change accordingly.

OTHER:

- Figures should be numbered following the appearance – Figure 1 only referenced toward the end of the paper

Authors' response: Thank you. We will change accordingly, either by referencing Figure 1 earlier in the manscript or by changing the order of the figures.

Response to General Comments from Samuel Buisan

My big comment is that more gauges at Norwegian site could have clearly helped in the analysis. Another possibility would be to include in the study other sites to support the concepts of "Universal Transfer Function" and "various measuring sites". This could be the objective of another related article.

Authors' response: Thank you. It is certainly true that more gauges at the Norwegian site would have been useful and that additional sites would help make the results more universal. The work presented here was performed before the WMO-SPICE measurements were available, and similar techniques will be used on the WMO-SPICE measurements in subsequent publications. These will include more sites and more gauges. While the measurements presented in this manuscript would certainly benefit from the inclusion of more gauges and sites, the work nevertheless demonstrates new techniques (such as the Exp. function), new gauge types (such as the double Alters and the SDFIR), and for one shield type combines automated measurements to create a two-site transfer function for the first time.

More description of Norwegian site is needed in comparison with the description of US site.

Authors' response: Thank you. Mekis provided the same recommendation in her review, and we will augment the manuscript accordingly.

Response to Specific Comments from Samuel Buisan

Page 2, lines 20-21: provide a citation for "The collection of rain…"

Authors' response: This is a good suggestion. We will provide a citation.

Page 3, line 10: Why it was "unexpected"?

Authors' response: We will reword and remove the word "unexpected".

Page 4, site descriptions: In NOR site you use snow depth and in US site snowfall accumulation. Please use the same variables for consistency.

Authors' response: Thank you. We will change the text accordingly.

Page 4, site descriptions: A figure with location of the sites, pictures and layout of the sites could be useful.

Authors' response: Thank you. Mekis also suggested that we include a map. We will do this, and include example photos of the shields as well.

Page 4, line 19: "several gauges…" why these gauges were not included in the analysis? Problems with dataset?

Authors' response: This is a good question. Yes, the other two single Alter gauges and an unshielded gauge at NOR were too noisy to use. These problems were resolved eventually, but the measurements recorded within the time period included in the manuscript were too limited to be very useful. In addition, the different positions of the SA-gauges relative to the DFIR required the filtering of very different wind direction angles to allow for uninfluenced measurements – in total there were too few episodes to compare directly. This is apparent in the new panel (Figure 2R, included below for reference) that will be added to Figure 5 in response to one of Mekis' comments. Only 103 60-min periods of precipitation were available for comparison of the two gauges.

Page 5, line 5: I don't understand the term "porosity".

Authors' response: Porosity describes how porous something is, or the percent of a material that consists of holes. For a wind shield, porosity is the percent of the shield that is open, allowing air to pass through; air can pass though 50% of a DFIR fence (the other 50% is blocked by wood). This word is fairly common in English, and trying to describe the concept without it might cause confusion, but we will include more explanation in the text.

Page 6, lines 15-17: "The 10 m wind speed" it seems more a conclusion than a description of the site. In addition, it is confusing to me the sentence " 10 m wind produced more accurate precipitations corrections" because later in the discussion section it is recommended to use wind speed at the gauge height.

Authors' response: Thank you. This is indeed confusing, and we will clarify in the manuscript. Also please refer to the explanation of this topic included in the responses to Mekis' comments. In the methods section we must explain how the wind speeds were used to estimate the gauge height wind speeds, so we maintain that it is also appropriate to describe the reason for adopting this approach within the methods section. However we will clarify that the approach was adopted based on "preliminary analysis". In addition, the section describing our recommended use of the gauge height winds will be altered to explain more clearly that the recommendation was based solely on first principles and differences in gauge height or changes in snow depth, rather than our actual measurements. The Wind Speed for Transfer Functions discussion will also be expanded to include a more balanced description of the difficulties of recording accurate gauge height wind speeds, the advantages of estimating the gauge height winds using the 10 m winds, and the fact that our measurements did not actually indicate any advantages to using the estimated gauge height winds over the 10 m height winds.

Page 7, line 6: Where the present weather detector was located? Inside DFIR? Maybe you should reference presentation on non-catchment technologies at TECO 2016 where the limits of these instruments were presented. Why data from present weather detector from NOR site was not used in the analysis? It would be really useful and also consistent to have the same Figure 2 (US site) for NOR site.
Authors' response: The present weather detector was located in the open. This detail will be added to the manuscript. We will also add a reference describing the limits of these sensors, but the TECO presentation didn't actually include weather type results, as it was focused on precipitation intensity. The NOR weather type analysis was previously included in Wolff et al. (2015), and we will add this reference to the manuscript.

Page 8, lines 19-20: please consider to describe briefly the methods for NOR site.
Authors' response: Thank you. We will summarize the NOR event selection methods here.

Page 9, lines 10 -13: please refer to figure 5 to state that "SDFIR data was the most comparable to DFIR" or add a citation.
Authors' response: Thank you. We will add a reference to Table 1.

Page 11, lines 10 – 13: I don't understand
Authors' response: The 10-fold cross validation required iterative fitting of the transfer function to different sub-selections of the available measurements; we fit the sigmoid function to 90% of the available SDFIR measurements ten different times. Because the SDFIR catch efficiency was fairly linearly related to wind speed, it was difficult to obtain a fit using the sigmoid function; the curve fitting software frequently failed to converge on a single SDFIR curve. It was possible to constrain the sigmoid coefficients within certain bounds and thereby produce a fit, but the constraints that worked on one set of SDFIR measurements didn't always work on another. Because of this we simply used one sigmoid curve fit to the entire population of measurements, and we validated the fit on the same population. It is important to note that this deviation from our preferred method is not significant; the error estimates produced using cross-validation were typically very similar to the error estimates produced by circularly fitting and testing the function on the full population of measurements.

Page 13, lines 6 -13: Did you test the results using the same number of events for both sites? It would be a good experiment to see if the results were independent of the number of events used. Another explanation for the bias could be the wide range of wind speeds at NOR site making more difficult to obtain an accurate transfer function for NOR site.
Authors' response: In response to this comment we performed a quick test by using only one out of every four US single Alter measurements. The resultant single Alter dataset included 352 NOR measurements and 292 US measurements. Using the gauge height winds and the Exp. transfer function as an example, there were only small differences between the resultant transfer function and the original transfer function created using all of the available US and NOR measurements. Likewise the site-specific errors were not changed by the omission of 3 out of every four US measurements. The NOR results were

almost identical, with a change in the RMSE from 0.45 mm to 0.45 mm and a change in the bias from -0.04 mm to -0.05 mm. The US results were also not significantly changed, with an increase in the RMSE from 0.15 mm to 0.17 mm, and a change in the bias from 0.00 mm to -0.01 mm. The authors suggest that such differences are not significant. We also believe that performing such an analysis and documenting it within the manuscript would not significantly increase the usefulness of the manuscript.

However as a result of this extra analysis prompted by Buisan's suggestion, the sentence suggesting that differences in the number of events affected the results will be removed.

Considering the results of this new analysis, Buisan is correct that the high wind speeds at the NOR site likely played a more important role in the magnitude of the transfer function errors than the number of available measurements. However even at low wind speeds the differences between the corrected NOR measurements and the DFIR-shielded measurements were larger from NOR than from US.

Page 14, lines 21-23: Do you think that wind adjustment is more difficult at NOR site because of higher wind speeds and turbulence? Any study related to this? Consider more discussion and citations

Authors' response: It was indeed more difficult to correct the NOR site because the winds were higher there. We will include some discussion of this in the manuscript.

Page 14, lines 25-33: Consider including that "10 m is standard according WMO guidelines and widely used in National Weather Services"

Authors' response: Thank you. We will include this.

Page 14, line 31: "snow depth… campaign maximum of 2 m…" it is confusing to me because previously it was written that snow depth reaches 3 m.

Authors' response: The maximum snow depth during the period studied in this paper (Feb2011-Apr2013) was indeed 1.97 m as written in the commented line. However, in other years snow depths of 3 m are often reached. Because of this, the precipitation gauges at NOR were mounted at the unusual height of 4.5 m. This is mentioned in Section 2.1 (Site Description, page 4, line 23), but we will further clarify this in Section 2.1.

Response to General Comments from Michael Earle

The application of the approach to the determination of 'universal transfer functions' is perhaps a bit lofty given the limited number of sites considered in the analysis, but the concept is novel and forward-thinking, and sets the stage well for broader implementation within the context of WMO-SPICE.

Authors' response: Thank you, and we will remove the 'universal' descriptor we over-ambitiously used to describe the transfer function.

Bigger things

1. The paper would benefit greatly from a more detailed description of the data analysis and event selection approach used for data from the Norwegian site in Sections 2.5.1 and 2.5.2. The approach used for the US site is clearly outlined, and the selection of 30 minute event intervals is justified by the authors. Meanwhile, a reference to the approach used for the Norwegian data is provided, with no further description of the approach, and no real justification for using 60 minute intervals instead of 30 min intervals in this case. Perhaps these points are well-articulated in the reference provided, but some description of the approach is required here. This also applies to the discussion of thresholds for precipitation events – it is unclear how thresholds for the reference and weighing gauges under test were determined and implemented for the Norwegian site data/events.

Authors' response: Thank you. We will add more description of the methods used on the Norwegian site.

2. Of greater concern are the implications of using different approaches for the different site data sets, with one approach using 30 minute time intervals, and the other using 60 minute intervals. In Sections 3.1 and 3.2, results for each site are presented and compared without mention of the caveat that the precipitation event datasets for each site were generated using different approaches that cover different time periods. a) It would make a far stronger case for comparison if both datasets were generated in the same way, and events covered the same time interval. Is it possible to apply the same approach to data from both sites? I would suggest this as a means of strengthening the results, or at least of validating that the specific approach employed does not significantly impact the results. Using the same time interval for events from both sites would also help to address bias observed in the results obtained using the combined transfer functions. In Section 3.2, this bias was related to the different number of precipitation events from the US and Norwegian sites that are used in the analysis. Reducing the time interval for the Norwegian events would likely increase the number of events, providing a means of testing this hypothesis. Moreover, combined transfer functions determined using events generated in the same manner, and covering the same time interval, would provide a much firmer step toward 'universal' transfer functions. b) Recognizing that going back and re-processing the data may be beyond the scope of this work, this paper is still an important contribution in its current state, and is suitable for publication provided the authors explicitly note the differences in how the events were generated as a complicating factor in the comparison of results from different sites, and in the determination of combined

transfer functions. A discussion of the expected impacts of using different event selection approaches and time intervals on the results and comparison would also be necessary. I feel that approach (a) will strengthen the results of this paper significantly relative to approach (b); however, approach (b) has strong ‘teaser trailer potential’ in terms of establishing the foundation for the work that will be done using the WMO-SPICE dataset, which includes precipitation events determined in the same way, using data that were processed in the same way, and cover the same time interval, from several different field sites/climate regimes.

Authors’ response: For this study, already pre-existing data sets from two different sites were used. The NOR-dataset was prepared for the study by Wolff et al. (2015) in which a transfer function based on measurements from only one site was derived. Wolff et al. (2015) describe the generation of 10 min and 1 h datasets, and a qualitative comparison of those NOR datasets did not reveal any significant differences. The more detailed Wolff et al. (2015) analysis was then performed on the 1 h data set because that time interval was similar to the operational measurement frequency in Norway. Both 30 and 60 minute time periods are short enough that representative averages of temperature, wind speed, and precipitation type can be calculated, and there is no other reason to believe that the choice of time interval would bias the resultant catch efficiencies. Based on the findings of Wolff et al. (2015) and also first principals, we did not expect significant differences between the 30 and 60 minute catch efficiencies.

In addition, as described in the abstract, the current study is meant as a conceptual study for the development and test of a transfer function based on datasets from multiple sites. The authors therefore focused on developing, presenting, and discussing their methods by applying them to existing datasets from two sites. The development of a common approach to data processing data and event selection is a major task within WMO-SPICE. Application of the methods presented in the current manuscript to more standardized datasets from several WMO-SPICE sites is currently underway, and these results will be published in the WMO-final report and associated publications. We will add some clarification regarding these points to Sections 2.5.2 and 4.2.

Also from Michael Earle: Smaller things (bold text denotes recommended additions/substitutions)

P1, L16: ‘high-quality’ is a subjective qualifier, and should be removed.

Authors’ response: “high-quality” will be replaced with “all-weather”.

P1, L20 (and throughout): ‘Altar’ should be replaced with ‘Alter.’

Authors’ response: Thank you!

P1, L28: write out ‘World Meteorological Organization’ and include full project acronym in parentheses, ‘(WMO-SPICE)’.

Authors' response: Thank you. We will correct this.

P2, L5-6: why are these changes in precipitation expected?

Authors' response: Such changes are predicted by climate models. We will change "expected" to "predicted" and reference accordingly.

P2, L17: ….increases as wind **speed** increases.

Authors' response: Thank you. We will correct this.

P2, L18: collection efficiency is mentioned for the first time here, but is not defined until page 9. It needs to be defined here, perhaps just in general.

Authors' response: Thank you. We will correct this.

P2, L23: I don't think it's critical to note who designed the wind shields.

Authors' response: The word, "scientists" will be removed.

Section 1 (general): pictures of the different shield types should be included, and will go a long way toward clarifying the (highly detailed) descriptions in the text.

Authors' response: We will add photos of the different shield types.

P3, L17: …a **Geonor** gauge ('Geonor' is capitalized several times throughout the paper… why?)

Authors' response: Thank you. "GEONOR" will be replaced with "Geonor".

P3, L18-19: include ice crystal habit among factors noted.

Authors' response: We will include ice crystal habit.

P3, L26: 'more robust results' is another subjective qualifier… I get what you're trying to say, but I'm not sure that it's necessary.

Authors' response: We will remove the phrase, 'more robust results'.

P3, L26-32: the structure and formulation is a bit odd here. I recommend starting with 'These results include:', and then listing the different aspects.

Authors' response: Thank you. We will re-write the sentence more clearly.

P4, L12-14: have these results been published elsewhere?

Authors' response: No, these analyses were performed on the Marshall measurements but never published. We didn't think they were significant enough to merit explicit inclusion.

P4, L21: the site was homogeneous in what sense?

Authors' response: We will add a more detailed description of the homogeneity of the site. Precipitation and wind measurements recorded before the DFIR shield was installed using two similar sets of instruments were in very good agreement, and indicated sufficient homogeneity of the location.

Section 2.2: note the oil, antifreeze used for gauges at each test site.

Authors' response: We will do this.

P5, L3: the term 'porosity' should be defined.

Authors' response: We will clarify what is meant by a porosity of 50%.

P5, L8: here you note the lath length for the small DFIR; what is the lath length for the standard or 'tall' DFIR?

Authors' response: We will include this detail.

Section 2.2: here, also, some demonstrative photos would go a long way toward clarifying the text description.

Authors' response: We will add some photos.

Section 2.3.2: the description for the Norwegian site is significantly lacking relative to that for the US site and should be expanded (e.g. specific sensors used for ancillary measurements).

Authors' response: We will augment the Norwegian site description.

P7, L27-28: I would consider removing the part starting with ', and any temperature threshold chosen…', and reformulating the similar statement on P8, L1-2. (Otherwise, it is redundant.)

Authors' response: Thank you! We will correct this redundancy.

P8, L23-24: without including a demonstrative plot, the statement regarding the 'gap in spectra describing atmospheric motions' doesn't mean much to those less familiar with surface layer dynamics; I would consider removing this statement.

Authors' response: We will remove the statement.

P9, L1-2: here you finally define catch efficiency, '…described as the ratio between…'; the ratio of what?

Authors' response: Thank you. We will add the word "precipitation".

P9, L16-22: What do you mean by 'unbiased transfer functions'? It is also not clear to me why the test gauge being able to measure more than 0.25 mm in 30 min necessitates the use of a lower threshold for test gauges. Can you please clarify or reformulate this statement? It is important to emphasize the key role of wind shield porosity in defining the thresholds for gauges under test.

Authors' response: Our goal was to develop transfer functions based on representative precipitation measurements, rather than a biased sub-selection of measurements. Because a threshold was needed for the standard and also because measurements from both the standard and the gauge under test were subject to a significant amount of random variation, this was actually somewhat challenging. To further clarify, in the example given the standard (not the gauge under test as stated by Earle in the comment above) erroneously measures more than 0.25 mm. If many such events are included in the analysis the measurements from the gauge under test will be erroneously biased low (as a result of the standard being erroneously biased high), and the resultant transfer function will therefore typically over-correct the gauge under test. Because of this, in order to handle measurements from both the DFIR and the gauge under test without biasing the results, the gauge under test also requires a minimum threshold. We will explain this better in the manuscript, and clarify by rephrasing "standard gauge/shield combination" as, "DFIR-shielded precipitation gauge".

P9, L30: if you're not planning to describe Bayesian analysis in any way, I don't think it's necessary to note this here.

Authors' response: The phrase, "using Bayesian analysis" will be removed.

P10, L5: Start a new paragraph beginning with 'We also propose…

Authors' response: A new paragraph break will be added in addition to the suggested phrase.

P11, L1: I would consider changing the wording to '….to what degree a **single** transfer function…' at this point and then noting the implications for 'universal' transfer functions within the context of the results and discussion.

Authors' response: Thank you. We will reword accordingly, "… to what degree a single two-site transfer function was valid for each of the two sites". This section will also require some rewording to accommodate the fact that the results will also include individual transfer functions for each of the SA sites.

P11, L21: 'size' seems like an odd descriptor for the different shields (e.g. double-Alters with different slat shape/mobility are effectively the same size, but have different porosities); perhaps 'porosity' is a better term?

Authors' response: With the exception of the Belfort double Alter, size was correlated with the efficacy of the wind shields. Porosity on the other hand does not describe the relationship between the SDFIR, double Altar, and single Altar catch efficiencies. We maintain that the relationship between shield size and shield efficacy is “generally” true, however we concede that it would be more accurate if we restated it as “size or efficacy”.

P11, L23: I feel like you need to justify why you would expect this, and so propose removing ‘As expected’ from the statement.

Authors’ response: We will remove ‘As expected’.

General note: some inconsistency with use of RMSE as singular/plural throughout. Also, I prefer ‘RMSE values’ to ‘RMSEs’ for plural use, but that’s not critical.

Authors’ response: Thank you. We will check for consistency and replace RMSEs with RMSE values.

P12, L24: change ‘efficacy’ of shield to ‘porosity.’

Authors’ response: Thank you, but we maintain that the word ‘efficacy’ is appropriate here. Note that ‘efficacy’ is not the same word as ‘efficiency’. The purpose of a wind shield is to reduce precipitation measurement errors. A shield associated with smaller errors is therefore more effective; efficacy can be determined directly by the resultant errors. Regarding porosity see above. For example, the Belfort double Alter is less porous than an SDFIR or a DFIR, but not as effective.

Section 4.2 (P15): it is stated that the Norwegian precipitation data are noisier and that the site is much windier. Is it possible to qualify these statements with some statistics? (Mean wind speed during precip events at each site, for example.)

Authors’ response: Thank you. This is a good suggestion. We will include statistics comparing the sites such as the mean wind speeds during precipitation events, and we can also calculate errors and correlations from the single Alter gauges at both sites in similar wind speeds.

P15, L31: ‘ephemeral’ is a really great word. No comment/suggestion here, just respect.

Authors’ response: Thank you!

P16, L1-2: the statement that precipitations ‘must be standardized’ comes off as a bit strong/preachy, and could be softened by adding ‘…standardized to the extent possible…’. It may not be possible to use the same approach in developing countries, for example, or different gauge types may be better suited to different climate regions.

Authors’ response: Thank you. We will rephrase ‘must be standardized’ and clarify the intent of this paragraph, which was related to the way the measurements are corrected rather than how the measurements are recorded.

Tables 1-3: I’m going to be blunt – your tables aren’t really visually appealing. It doesn’t change the content, of course, but they would look much nicer within the manuscript if they were cleaned up a bit.

Authors' response: Thank you for being candid. A detailed description of the changes we will make to the tables is included in the response to Mekis. The tables will all be reformatted, and Tables 2 and 3 will each be replaced with three tables, allowing for all the results for a given shield type to be displayed on a single line.

Figures 1 and 7: the impact of the application of corrections to the sample dataset is mitigated by showing the corrected and uncorrected data in separate figures. I would consider changing Figure 7 to Figure 1b. Not only would this clearly demonstrate the impact of the corrections, showing this early in the paper may serve as a 'sneak preview' to entice readers to continue.

Authors' response: We agree that the large separation between Figure 1 and Figure 7 makes it difficult to compare the two, and will include Figure 7 in a separate panel of Figure 1, or alternatively make Figure 1 part of Figure 7.

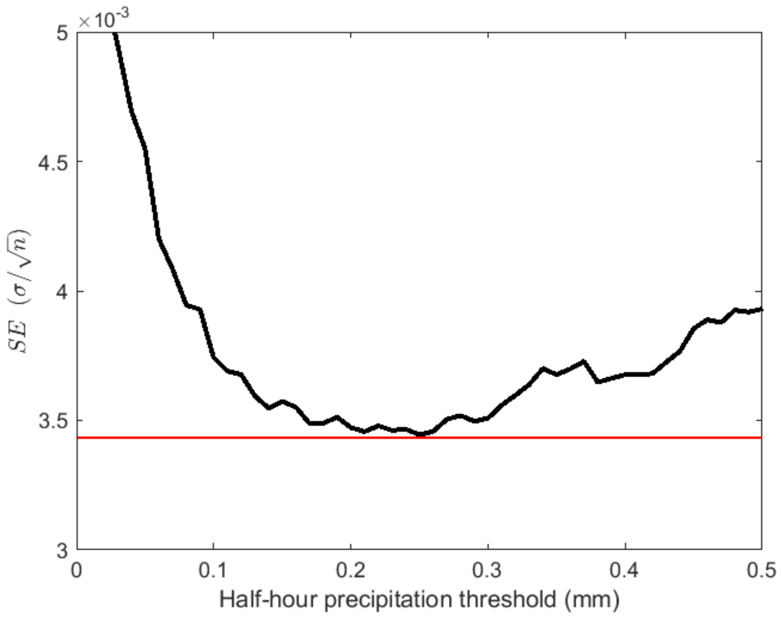


Figure 1R: The same plot shown in Figure 3 panel c in the manuscript under discussion, with the ranges of the x- and y-axes altered. The standard error (SE $= \sigma/\sqrt{n}$) of the transfer function is shown (black line) in addition to a horizontal line (red) placed to help locate the minimum in the standard error.

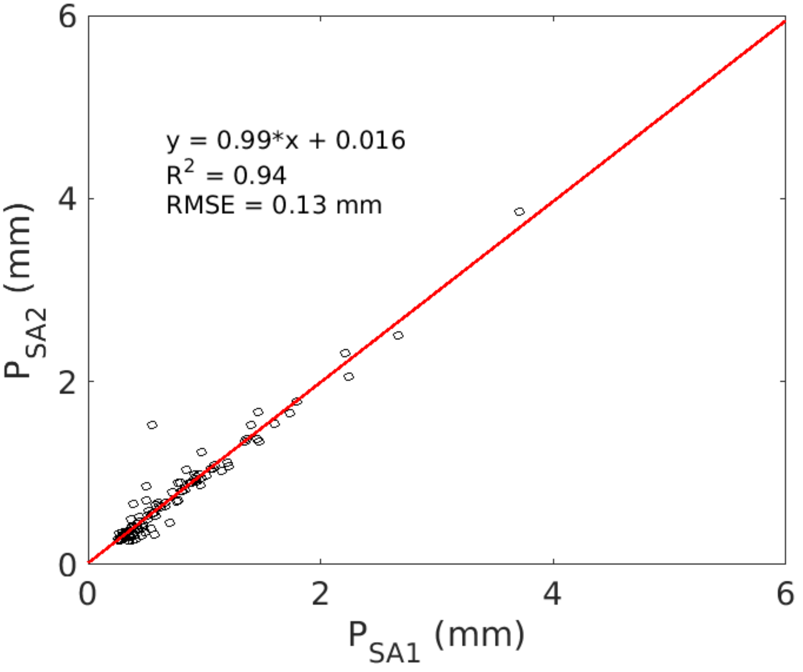


Figure 2R: The same plot shown in Figure 3 panel c in the manuscript under discussion, with the ranges of the x- and y-axes altered. The standard error ($SE = \sigma/\sqrt{n}$) of the transfer function is shown (black line) in addition to a horizontal line (red) placed to help locate the minimum in the standard error.

2) List of all relevant changes made in the manuscript

Abstract:

1. “high-quality” changed to "all-weather”
2. “Altar” changed to “Alter”, also throughout the manuscript.
3. “Universal” removed, also throughout the manuscript.
4. “Removed” changed to “decreased”.

Introduction:

1. “Expected” changed to “predicted” and reference added.
2. “Wind” changed to “wind speed”.
3. Collection efficiency defined.
4. The word “scientists” was removed.
5. Goodinson et al. (1998) reference corrected here and throughout the manuscript.
6. Removed “As expected”.
7. “Ice crystal habit” added to the list of factors affecting catch efficiency.
8. “More robust results” removed and the following sentence was re-written as a list.

Site Descriptions:

1. Site Descriptions moved from Methods section into their own section.
2. References to the new Fig. 1, which includes photos of the shields and a site map, added throughout the Site Description and Methods section.
3. US and NOR site descriptions made more consistent.
4. Relationship between catch efficiency and wind direction at the US site described in more detail.
5. References to past publications on the US site added.
6. Explained in more detail that at the NOR site snow depth only reached 1.97 m during the study period.
7. Summarized the Wolff et. Al (2010) NOR site homogeneity results.

Methods:

1. Added a sub-section describing the period of measurement included in the manuscript from both sites.
2. Changed “GEONOR” to “Geonor” here and throughout the manuscript.
3. Described the oil and anti-freeze used at both sites.
4. Added lath length to the DFIR description.
5. Described out usage of porosity using the DFIR as an example.
6. Included a reference to the drawing of several shields that is included in Rasmussen et. Al (2012)

7. “Based on preliminary analysis” added to the description of the 10 m wind speed vs. the gauge height wind speed transfer functions at the US site.
8. Descriptions of the NOR ancillary measurements added.
9. Description of NOR wind obstructions added.
10. US Present Weather Detector described as “unshielded”.
11. A reference describing the shortcomings of precipitation type measurements was added.
12. Reference to Fig. 2 added.
13. Reference to Wolff et. al (2015) precipitation type analysis added.
14. Redundant description of “arbitrary” temperature thresholds removed.
15. Clarified that 30-min precipitation measurements were used for the US site.
16. Description of the 60-min event selection methods used at the NOR site added.
17. Discussion of the spectral gap removed.
18. Description of 10-min vs. 60-min catch efficiencies from the NOR site added.
19. Description of catch efficiency clarified.
20. Reference to Table 1 added to the explanation of the similarity of SDFIR and DFIR measurements.
21. The concept of unbiased transfer functions and representative measurements was described in more detail in the justification for a minimum threshold for gauges under test.
22. The term “Bayesian analysis” was removed.
23. Description of the two single Alter shields used for measurement uncertainty analysis and Fig. 5 was added.
24. The Uncertainty of Transfer Functions section was rearranged to accommodate the addition of site-specific single Alter transfer functions.

Results:

1. “RMSEs” replaced with “RMSE values” throughout the manuscript.
2. Added “or efficacy”.
3. Removed, “As expected”.
4. The “Transfer functions and uncertainty” section was rearranged and augmented to accommodate the addition of site-specific single Alter transfer function results. Also references to the tables were updated to reflect the division of each of original transfer function table into several separate tables.
5. Our hypothesis that the transfer functions were biased towards the US measurements because they were more numerous than the NOR measurements was removed from the text. The analysis performed in response to Buisan’s suggestion that we try using the same number of events proved that this hypothesis was ultimately incorrect.
6. The text was changed to include a description of the 1-day time series of uncorrected precipitation measurements, which now appears later in the manuscript than in the original manuscript to accommodate the suggestion that it be merged with the corrected time series of the same event.

7. The importance of bias correction vs. uncorrected variability for water budgets was added.

Discussion:

1. The Wind Speed for Transfer Functions section was rewritten to include a clearer and more balanced discussion of wind speed measurement height.
2. The role of climate was emphasized throughout the Two-Site Transfer Function (previously Universal Transfer Function) section.
3. Statistics were also added describing mean wind speeds from both sites and the results of additional test performed on datasets filtered for similar wind speed distributions.
4. Again the hypothesis that the greater number of US measurements affected the NOR bias was removed.
5. A ‘teaser’ sentence describing the improved common analysis performed on all WMO-SPICE sites was added.

Conclusions:

1. Replaced, “widely-used shields” with “common gauge/shield combinations”.

Tables and Figures:

1. The transfer function coefficients and the resultant RMSE values and biases were separated into different tables. Also site-specific transfer functions for the single Alter gauges were provided, and the two-site RMSE values and biases for NOR and US sites were presented in a separate table.
2. A figure describing the location of the sites and photos of the sites was added (Fig. 1).
3. A description of the number of measurements and the time periods was added to the Fig. 4 caption.
4. Figure 5 was augmented to include a pair of single Alter gauges from the NOR site.
5. Figure 7 now includes both the uncorrected and the corrected measurements in two separate panels.

3) Marked-up manuscript version

The Quantification and Correction of Wind-Induced Precipitation Measurement Errors

John Kochendorfer[1], Roy Rasmussen[2], Mareile Wolff[3], C. Bruce Baker[1], Mark E. Hall[1], Tilden Meyers[1], Scott Landolt[2], Al Jachcik[2], Ketil Isaksen[3], Ragnar Brækkan[3], and Ronald Leeper[4,5]

[1]ARL/Atmospheric Turbulence and Diffusion Division, National Oceanic and Atmospheric Association, Oak Ridge, TN, 37830, US
[2]National Centers for Atmospheric Research, Boulder, 80305, US
[3]Norwegian Meteorological Institute, Oslo, 0313, Norway
[4] N Carolina State Univ., Cooperative Inst of Climate and Satellites, Asheville, 28801 US
[5] National Center for Environmental Information, National Oceanic and Atmospheric Association, Asheville, 28801 US

Correspondence to: John Kochendorfer (john.kochendorfer@noaa.gov)

Abstract

Hydrologic measurements are becoming increasingly important for both the short and long term management of water resources. Of all the terms in the hydrologic budget, precipitation is typically the most important input. However, measurements of precipitation are still subject to large errors and biases. For example, an high quality butall-weather unshielded weighing precipitation gauge designed for the measurement of solid precipitation can collect less than 50% of the actual amount of solid precipitation when wind speeds exceed 5 m s^{-1}. Using results from two different precipitation testbeds, such errors have been assessed for unshielded weighing gauges and for four of the most common windshields currently in use. Functions used to correct wind-induced undercatch were developed and tested. In addition, corrections for the single AlterAltar weighing gauge were developed using the combined results of two separate sites, one of which was in Norway and other in the US. In general the results indicate that corrections described as a function of air temperature and wind speed effectively remove correct the undercatch bias that affects such precipitation measurements. In addition, a single 'universal' function developed for the single AlterAltar gauges effectively removed effectively decreased the bias at both sites, with the bias at the US site improved from -12% to 0%, and the bias at the Norwegian site improved from -27% to -3%. These correction functions require only wind speed and air temperature, and were developed for use in national and local precipitation networks, hydrological monitoring, roadway and airport safety work, and climate change research. The techniques used to develop and test these transfer functions at more than one site can also be used for other more comprehensive studies, such as the World Meteorological OrganizationWMO Solid Precipitation Intercomparison Experiment (WMO-SPICE).

1 Introduction

Precipitation measurements are used by policy makers, hydrologists, farmers, and watershed managers to quantify and allocate the water available for society's needs. Precipitation measurements are necessary for public safety in areas as diverse as avalanche control, flood forecasting, roadway safety, and aircraft de-icing operations. Precipitation measurements are also used to evaluate radar-based estimates of rainfall, to monitor climate change, and help improve climate and weather models. More specifically, monitoring changes in the frequency, intensity, duration and phase of precipitation is critical for current and future climate research (Barnett et al., 2005; Blunden and Arndt, 2016; Trenberth, 2011; Trenberth et al., 2003). Although precipitation has been monitored for many centuries, the need to accurately measure precipitation will become even more important in the future, as changes in precipitation are ~~expected~~ predicted to be complex and variable with location, requiring robust and accurate precipitation measurements (Trenberth, 2011). Despite this critical need, precipitation observations are still beset with significant biases and errors (eg. Adam and Lettenmaier, 2003; Førland and Hanssen-Bauer, 2000; Groisman and Legates, 1994; Scaff et al., 2015; Vose et al., 2014; Yang et al., 2005).

Solid precipitation is particularly difficult to measure accurately, and biases between wintertime precipitation measurements made using different technologies, different measurement networks, or across different regions can be larger than 50% (Mekis and Vincent, 2011; Rasmussen et al., 2012; Yang et al., 1999). Previous studies have identified wind effects as one of the primary causes for snow undercatch (eg. Folland, 1988). This is due to two important factors: 1) the relatively slow fall velocity of snow, and 2) the creation of flow distortions of similar magnitude to the snow fall velocity by the gauge itself as air flows past it. These two factors cause a snowflake trajectory to be significantly deflected by the airflow past the gauge. In particular, the updraft at the leading edge of the gauge can lead to an upward deflection of snowflake trajectories, causing snow to miss the gauge orifice and not be measured. The flow distortion around the gauge increases as the wind speed increases, while snowflake terminal velocity remains the same, causing more snowflakes to be deflected around the gauge. Thus, the amount of precipitation caught in a precipitation gauge relative to the reference or actual amount of precipitation, (also referred to as the collection efficiency) ~~of snow by a gauge~~ decreases with increasing wind speed. The collection of rain in a weighing gauge suffers from this same problem (eg. Duchon and Essenberg, 2001), but to a much lesser extent due to the order of magnitude higher fall velocity of raindrops, allowing them to be subject to only minimal trajectory deflection due to flow distortions around the gauge.

To mitigate this deflection and undercatch of snow by weighing gauges, ~~scientists have designed~~ wind shields ~~to~~ are used around~~surround~~ the gauge with the goal of slowing the oncoming flow, and lessening the resulting flow distortions around the gauge. Prominent among these is the Alter shield (Alter, 1937). This shield consists of vertical slats of ~40 cm length suspended from a circular rim ~100 cm in diameter surrounding the gauge. The flow speed is indeed decreased (Rasmussen et al., 2012), and the flow distortion around the gauge reduced leading to an increase in snow collection. This shield is widely used around the world to improve the measurement of snow and rain.

To quantify the impact of the Alter shield and other types of shields (Alter, 1937; Nipher, 1878; Yang et al., 1995), the World Meteorological Organization (WMO) sponsored a solid precipitation measurement program for manual weighing gauge in the early 1990's (Goodison et al., 1998).(Goodison et al., 1998). A significant result of this study was the creation of a reference standard gauge for snowfall. This was determined to be a precipitation gauge embedded within a carefully pruned bush without leaves. The height of the bush was the same level as the orifice of the gauge. A secondary reference was established as the Double Fence Intercomparison Reference (DFIR, Groisman et al., 1991; Yang, 2014). This allowed field sites to establish a standard for snow measurement using a DFIR without having to grow and prune a bush system. A key characteristic of the reference was that the snowfall rate did not significantly depend on the magnitude of the wind.

The 1990's WMO study used both the bush and DFIR as references to compare to a variety of gauges with different shielding (Goodison et al., 1998).(Goodison et al., 1998). A key result from this study was that the collection efficiency of the gauge was primarily determined by the type of wind shield used (Yang et al., 1999). Rasmussen et al. (2012) compiled a review of snowfall measurements to date that confirmed this result for automatic weighing gauges and other wind shield types.

These studies confirmed that the weighing gauge collection efficiency for snow as compared to the above-mentioned reference gauge systems decreased with increasing wind speed, depending on the type of the shield. An unexpected resultIn addition, past studies have shown that was that the collection efficiency for a given wind speed and wind shield/gauge type variedcan vary significantly (Yang et al., 1999, Rasmussen et al., 2012). While functions describing the decrease in collection efficiency could be derived, there was often as much variability at a given wind speed as across the range of wind speeds (Yang et al., 1999). These results revealed that wind speed is not the only factor that impacts the trajectory of a snowflake past a gauge. Since the fall speed of a snowflake is often close to the magnitude of the flow distortion, an obvious candidate was the fall speed of the snow. Using numeric modelling, Theriault et al. (2012) showed that the difference between wet and dry snow fall speeds can lead to significant changes in the collection efficiency of snow by a GENOR weighing gauge with and without an Alter shield. Other factors include airflow turbulence, time dependence of the flow past a gauge (Colli et al., 2015), snow size distribution (Theriault et al., 2012), ice crystal habit, and snow density (Colli et al., 2015).

As a result, the performance of weighing snow gauges with various types of wind shields may be expected to vary by climate regions. The previous WMO Solid Precipitation Intercomparison (Goodison et al., 1998) examined solid precipitation from manual gauges at a limited number of field sites.(Goodison et al., 1998) examined solid precipitation from manual gauges at a limited number of field sites. The more recent WMO Solid Precipitation Intercomparison Experiment (SPICE) expanded the number of climate regimes covered, and used automatic gauges instead of manual (Nitu et al., 2016). This paper examines results from two sites participating in this experiment, and includes several years of measurements that pre-date the WMO experiment for more robust results. These results include: 1) the assessment of various weighing gauge/shield combinations across these sites, and: 21) a description of the dependence of snow collection efficiency as a function of wind speed as well as its uncertainty at the two differentvarious measuring sites with characteristic climatological conditions, 32) identifies athe functional form for the collection efficiency/wind speed relationship (referred to as a “transfer function”), 43) an assessment ofes whether each climatological site requires a different transfer function or if a universal single multi-site function can be used to reduce wind induced snow undercatch, and 54) quantification ofes the expected uncertainty of the correction as a function of gauge/shield type and climatological conditions.

2 Methods

2.1 Site dDescriptions

The US field site is located just outside of Boulder, Colorado along the eastern slopes of the Colorado Front Range (Fig. 1 a). The site resides on top of the Marshall Mesa at 39.949° north, 105.195° west, and is ~1740 m above sea level. Prairie grasses and low scrub are the primary flora found at the site (Fig. 1 d and e), and its proximity to the mountains makes it an ideal location for studying upslope snowfall events. Although tThe US site experiences only intermittent snow on the ground and the snow depth is typically less than 0.5 m, it typically receives a total of ~200 cm of snowfall throughout the course of the winter months. Snowfall at the site typically occurs from October through April but can occasionally occur as early as September or as late as June. The site is relatively flat and the lack of trees, large buildings and other obstacles allow for uninterrupted wind flow around the gauges. The site has several Double Fence Intercomparison Reference (DFIR) shields. There was a single Alter shield with a weighing precipitation gauge within the DFIR used for this study. All of the gauges included in this study were mounted with their inlets approximately 1.9 m above the ground. The layout was designed to minimize the effects of the gauges and their wind shields on each other. The site generally stretches out perpendicular to the wind direction that prevails during snow storms, and special care was taken not to put larger shields upwind of unshielded and smaller-shielded gauges. Comparison of replica single- and double-Alter shielded gauges from differing locations at the field site (Fig. 5) and a lack of any significant relationship between wind direction and catch efficiency significant wind direction effects on catch efficiency(eg. for the single Alter gauge $R^2 = 0.001$) confirmed the fact that wind direction did not play an important role in catch efficiency at the site. The site is also described in detail in past publications (Rasmussen et al., 2012; Rasmussen et al., 1999).

The Norwegian test site (Haukeliseter) is situated at 59.812° N, 7.214° E, and is at 991 m a.s.l. on a plateau in an alpine region in southwestern Norway (Fig. 1). Snowfall at the site typically occurs October through May. Annual ~~maximum~~ snow depth often ~~reaches~~ reaches up to~ 3 m, although during the time period included in the present study it only reached 1.97 m. The site has one Double Fence Intercomparison Reference (DFIR) surrounding an automated weighing gauge within a single Alter shield (SA). The DFIR and several gauges, most of them within single Alter shields, ~~are~~ were installed in two lines perpendicular to the main wind directions (east-southeast and west-northwest). An evaluation of two similar sets of precipitation and wind measurements recorded at different locations ~~throughout the~~at the site during February - May 2010 indicated that the site was homogeneous (Wolff et al., 2010)~~.~~, with comparison of the two sets of measurements resulting in a correlation coefficient of 0.94, a standard deviation of 0.07 mm (8.2%), and similar average event precipitation from both locations (0.89 mm and 0.87 mm). ~~(Wolff et al., 2010).~~ The gauges at the Norwegian test site ~~are~~ were all mounted ~~at~~ 4.5 m ~~altitude~~ above the ground level in order to mitigate the effects of blowing snow and to allow for the increasing snow depth through the season. ~~Annual maximum snow depth reaches ~3 m.~~ The site is also described in more detail in previous publications (Wolff et al., 2013; Wolff et al., 2015).

3.0 Methods

3.1 Measurement period

The US measurements used here span from Jan 01, 2009 through March 07, 2014, and include all seasons. The NOR measurements used here are identical to those used in Wolff et al. (2013), and were recorded only during winter periods as follows: Feb 01, 2011 – Apr 30, 2011; Nov 01, 2011 – Apr 30, 2012, and Feb 01, 2013 – May 31, 2013.

3.2~~1~~~~2.2~~ Precipitation g~~G~~auges and s~~S~~hields

To reduce potential sources of uncertainty, all of the precipitation measurements presented here were recorded using the same model weighing precipitation gauge (3-wire T200B, Geonor Inc., Oslo, Norway) and all of the gauge inlets were heated using the same type of inlet heaters (described in NOAA Technical Note NCDC No. USCRN-04-01). Both the outer and inner tubes of the ~~GEONOR~~Geonor were heated to prevent snow melted at the orifice from re-freezing while it drips into the collection bucket. The inlet heaters were activated only when the inlet temperature and the air temperature were both < 2 °C. At the US site, 2 litres of antifreeze (60% Methanol, 40% Propylene Glycol) and 0.4 litres of hydraulic oil (Lubriplate Minus 70) were added to every gauge to prevent freezing and evaporation. At the NOR site, 5 litres of Methanol, 3.3 litres of Ethylene Glycol, and 0.4 litres of hydraulic oil (Hydraway HVXA 15LT) were used.

~~T~~Present at both the NOR and US sites, the DFIR shield has the largest footprint of any of the shields, and consists of three concentric shields. The outer two shields are octagonal in ~~design~~design, and are made out of 1.5 m tall wood lath, with the outer shield having a diameter of twelve meters and the middle shield having a diameter of four meters. The DFIR shield has a porosity of 50%, with 50% of its surface area open allowing air to pass though (the other 50% of its surface area is blocked by the wood lath), and both the outer and middle shields are perpendicular to the ground. For the third innermost shield, an Alter-style shield of standard size and configuration is used. The DFIR shield is described in more detail in the first WMO Solid Precipitation Intercomparison (Goodison et al., 1998).~~.~~

The 2/3 scale version of the DFIR, hereafter referred to as the small DFIR (SDFIR), was designed for the US Climate Reference Network program and ~~tested at the US site~~was installed at only the US site. The SDFIR laths are 1.2 meters long, and the diameter of the outer shield is eight meters in diameter and the middle shield is 2.6 meters in diameter. Additionally, the middle shield height is 10 cm lower than the outer shield. A standard diameter Alter shield is used as the innermost shield, which is 10 cm lower than the middle shield and located at the same height as the gauge orifice inlet.

Alter shields, hereafter referred to as single Alter shields (SA) to avoid confusion with the double Alter shields, were installed at both the NOR (foreground of Fig. 1 b) and US sites. The single Alter shield ~~hereafter referred to as single Alter shields (SA) to avoid confusion with the double Alter shields,~~ consists of metal laths about 40 centimeters in length (though some versions of the Alter use slightly longer laths that are 46 centimeters in length). The laths on the SA shield are typically attached near the top to a circular ring, 1.2 meters in diameter, and allowed to move freely in the wind. The double Alter shield (DA, Fig. 1 d) is a variation of the single Alter shield and has two concentric shields instead of one (Rasmussen et al., 2001). This shield consists of a standard 1.2 meters diameter single Alter shield surrounded by an additional outer ring of laths measuring two meters in diameter. Like the single Alter, the laths on both rings are approximately 40 cm in length, secured only at the top, allowing them to move freely at the bottom. A sketch of the DFIR, single Alter shields, and double Alter shields and a description of their effects on the wind speed at the gauge inlet is available in Rasmussen et al. (2012).~~(Rasmussen et al.)~~

The Belfort double Alter shield was only present at the US site (Fig. 1 e), and is a modified version of the standard double Alter shield. The diameter of the inner shield is 1.2 meters and the laths are 46 cm long. The diameter of the outer shield is 2.4 meters and the laths are 61 cm long. Unlike the standard single and double Alter shield laths, these laths don't taper at the bottom and are only allowed to swing inwards or outwards at a maximum 45-degree angle. The Belfort double Alter shield is also only approximately 30% porous, which is significantly less than the ~50% porous double Alter shield.

~~2.~~3.3~~2~~ Other m~~M~~easurements

~~2.~~3.3~~2~~.1 US Site

At the US testbed the air temperature was measured using fan-aspirated (Model 076B Radiation Shield, Met One Instruments, Grants Pass, OR, US) platinum resistance thermometers (Thermometrics, Northridge, CA, US) mounted at a height of 1.5 m. Three wetness sensors (Model DRD11A, Vaisala, Helsinki, Norway) also mounted at a height of 1.5 m were used to independently detect precipitation. Wind speed was measured at 1.5 m using a cup anemometer (Model 014A Wind Speed Sensor, Met One Instruments, Grants Pass, OR, US), at 2.0 and 3.0 m using propeller anemometers (Model 05103 Wind Monitor, RM Young, Traverse City, MI), and at 10 m using both a propeller anemometer (Model 05103 Wind Monitor, RM Young) and a two-dimensional sonic anemometer (Model 86004 Ultrasonic Anemometer, RM Young). The two anemometers at 10 m were found to interfere with each other due to wind shadowing when winds were from the north or the south, and a composite 10 m wind speed was therefore produced using the ultrasonic anemometer measurements to replace the propeller anemometer measurements when winds were from the north (wind directions < 30° or > 340°). Likewise, as identified by plotting the ratio of the measured wind speed to the 10 m wind speed as a function of wind direction, the lower wind speeds were found to be subject to interference at some wind directions. Using 30-min mean wind speeds measured in a clear sector (wind direction < 30°) from all measurement locations, the roughness length (z_0 = 0.01 m) and displacement height (d = 0.4 m) were determined based on the log wind profile (Thom, 1975):

$$U_z \approx \ln\left[{(z-d)}/{z_0}\right] \quad (1)$$

where U_z is the wind speed (U) at a height z (U_z). Using the same relationship, the wind speed at the gauge height of 1.9 m was estimated be equal to U_{10m}×0.71. Only wind speed measurements recorded during precipitation events were used to develop this relationship, minimizing the neglected effects of stability on the wind profile in the typically overcast and near-neutral surface layer conditions associated with precipitation. Due to the effects of near-field obstructions on the near-surface wind speed measurements, the 10 m wind speed measurement was used to estimate the gauge-height wind speed measurement using this method throughout the study. Based on preliminary analysis, t~~T~~he 10 m wind speed was more generally representative of the wind speed affecting all of the gauges throughout the site, and it produced more accurate precipitation corrections than the gauge-height wind speed. Errors in the method used to estimate the gauge height wind speed from the 10 m wind speed were evaluated using the mean half-hour 2 m wind speed measurements recorded during precipitation events when the recorded wind speed was greater than 1 m/s from unobstructed wind directions (wind direction~~WD~~ < 30 deg). The 2 m wind speeds were compared to the gauge height (1.9 m) wind speed estimated using the log profile, resulting in an RMSE of 0.4 ms^{-1} (10.1%) and a bias of -0.12 ms^{-1} (-3.1%).

~~2.~~3.3~~2~~.2~~.2~~ Norwegian s~~S~~ite

At the Norwegian site (hereafter NOR), air temperature was measured with a 100 Ω platinum resistance thermometer within a standard Norwegian radiation screen installed at gauge height on a tower near the DFIR. A precipitation detector (Yes/No

Precipitation Monitor, Thies Clima, Göttingen, Germany) was also mounted at gauge height and used to record the presence/absence of precipitation. The primary wind speed was measured at a height of 10 m with a heated two-dimensional sonic anemometer (WindObserver II, Gill Instruments, Lymington, UK~~At the Norwegian site (hereafter NOR), u~~). In addition, a propeller anemometer (Wind Monitor, R.M. Young, Traverse City, US) was mounted at gauge height, but was found to be obstructed by a neighbouring single Alter shield from some wind directions. Unobstructed (wind direction > 240°) 10 m and gauge height (4.5 m) wind speed measurements recorded during precipitation events were used to determine the relationship between the gauge height and 10 m wind speeds:

$U_{4.5m}$ = 0.93×U_{10m}, R^2 = 0.99, RMSE = 0.54 m s^{-1}, and the wind speed at 10 m was used to predict the gauge-height wind speed from all wind directions. Following Wolff et al. (2015), precipitation events were also screened for wind directions associated with shadowing between gauges and shields (wind direction < 240° and wind direction > 355°), and were excluded from the analysis.

3.4~~32.4~~ Precipitation type

Transfer functions have commonly been developed separately for snow, mixed precipitation, and rain (eg. Goodison et al., 1998; Yang et al., 2005). Other proposed classification schemes also include differentiation between wet and dry snow. In the past, manual observations of precipitation type were recorded and used to develop such transfer functions, but modern automated measurement networks now rarely include such manual measurements. Airport weather stations often include precipitation type measurements, but hydrological, meteorological, and climate stations do not typically include precipitation type measurements, and defensible methods to correct wind-induced errors without precipitation-type measurements are therefore needed.

At the US testbed, precipitation type was determined using an unshielded present weather detector (Vaisala PWD22, Helsinki, Finland). Half hour increments of rain, mixed, and snow were identified using more than 15 min of any of these precipitation types as detected from the present weather detector measurements recorded every minute. For rain and snow, less than 5 min total of any other precipitation type were allowed to occur. These identifiers were used only for the analysis of precipitation type shown in ~~Figures~~Fig. 2, 3 and 5, but not for the development of transfer functions. This is because such sensors are not without errors in their determination of precipitation type (Merenti-Välimäki et al., 2001; Sheppard and Joe, 2000), and also because air temperature is more widely available than precipitation type sensors for estimating precipitation type and applying transfer functions.

At the US site, below -2.5 ˚C more than 95% of the precipitation in every 1 ˚C bin was classified as snow using the present weather detector. Above 2.5 ˚C, more than 95% of the precipitation was classified as rain (Fig. 2). These thresholds will of course change depending upon the climate of a given site, and the 95% threshold could also be adjusted to suit the needs of a given study. However, based on these temperature thresholds, the precipitation type that was most sensitive to classification methods was mixed precipitation. The present weather detector classified only 5% of the available half hours of precipitation as mixed, whereas 19% of the precipitation occurred between -2.5 ˚C and 2.5 ˚C. A significant amount of rain and snow may therefore be misclassified as mixed precipitation when temperature thresholds are used to predict precipitation type, assuming that the present weather detector accurately identified precipitation type. Wolff et al. (2015) performed a similar analysis of the precipitation type at the NOR site, using their Present Weather Detector (Vaisala PWD21 , Helsinki, Finland).

An alternative to using temperature thresholds to differentiate between different precipitation types is to use a continuous function of both air temperature and wind speed to define the catch efficiency (Wolff et al., 2015). Although the functions produced using such methods are more complex than a relationship between wind speed and catch efficiency for a single precipitation type, such an approach is arguably more convenient because only one equation is needed to determine catch efficiency for all conditions. More importantly, a continuous function of temperature may more accurately represent reality, as catch efficiency varies continuously with air temperature, especially near 0 °C. The transition from liquid precipitation to dry snow is continuous, without any well-defined step changes~~, and any temperature threshold chosen to differentiate between different precipitation types will inevitably be somewhat arbitrary~~. In addition, a continuous function of air temperature and wind speed eases the comparison of catch efficiency results from different sites with potentially different climates. This in turn aides in the evaluation of the uncertainty inherent in a~~ ‘universal’~~ transfer function ~~describing ~~that describes more than one site.

Following Wolff et al. (2015), for the transfer functions developed here we use a continuous f(T_{air}, U) as air temperature measurements are more universally available than precipitation type, different precipitation type detectors do not always agree on the precipitation type (Merenti-Valimaki et al., 2001; Wong, 2012), and air temperature thresholds used to separate different precipitation types must be chosen somewhat arbitrarily.

3.54~~2.5~~ Transfer f~~F~~unction d~~D~~evelopment

3.54~~2.5~~.1 Data analysis and event selection

For the US results, the US Climate Reference Network precipitation algorithm was used to determine 5-minute accumulations from all the 3-wire Geonor precipitation gauges (Leeper et al., 2015). The algorithm relied upon wetness sensor measurements to detect periods of precipitation, and it calculated the average accumulation of the three wires by inversely weighting the individual wire accumulations using the variance of the individual depths over the last 3 hours. This was done to lessen the contribution of noisier wires to the total gauge depth, and thereby decrease the amount of noise in the precipitation measurements. The algorithm was modified for the purposes of this study by increasing the 5-minute precipitation resolution from 0.1 mm to 0.01 mm. The 5-minute accumulations were then summed into half-hour periods, with these half-hour precipitation measurements used for all of the subsequent US analysis.

In addition at the US site, more than 10 minutes total of any type of precipitation as identified by the present weather detector had to occur within each half-hour to be included in the transfer function analyses. Half-hour increments with unrealistic air temperatures or wind speeds were also excluded from the transfer function analyses. For example, half-hour increments with 10 m anemometer measurements affected by ice accumulation were identified by comparing the 10 m propeller anemometer measurements with the 1.5 m cup anemometer measurements.

~~For the NOR results the methods described in Wolff et al. (2015) were used for the selection of 60-min periods included in the analysis.~~For the NOR results the methods presented in Wolff et al. (2015) were used to calculate precipitation accumulation and select 60-min precipitation periods for analysis. Every minute, gauge depths were determined by averaging the output of the three wires of each sensor. As an additional noise filter, 10 min running averages were calculated for the precipitation gauges and the optical precipitation detector. For the next step, periods with continuous and clear precipitation signals were selected following the following criteria:

a) The precipitation detector signalled precipitation for at least eight out of ten minutes
b) The accumulation was greater than 0.1 mm per 10 minutes or greater than 1 mm for events that lasted longer than 100 min.

The resulting precipitation periods were of different lengths and were divided into sets of 10 min or 60 min events for the study by Wolff et al. (2015). No significant differences between the 10 min and the 60 min catch efficiencies were detected in the qualitative analysis performed on those data sets (Wolff et al., 2015), and following (Wolff et al., 2015) the transfer function development and testing was performed using the 60 min precipitation events.

3.54~~2.5~~.2 Selection of precipitation threshold~~.~~

For the US site, 30-minutes was selected as the most suitable time interval for the creation of transfer functions, as 30-minutes ~~is near the centre of the gap in spectra describing atmospheric motions within the surface layer (eg. Stull, 1988), allowing~~ allows averages of air temperature and wind speed measurements that are both representative of the field site as a whole and also relatively stationary (eg. Stull, 1988). A longer time period would be subject to increased mesoscale, synoptic, and diurnal changes in precipitation type, wind speed, and air temperature. Under quiescent conditions a shorter averaging period would approach turbulent time scales, where the presence or absence of an individual eddy would affect the results, making the results less representative of the entire site. At the NOR site, 60-minute periods of precipitation were used following Wolff et al. (2015), with many of the same arguments supporting the 30-minute period equally valid for 60-

minutes. Qualitative analyses by Wolff et al. (2015) on 10- and 60-minute datasets did not reveal any significant differences between those time intervals. Therefore the 60-minute time period, which is similar to the operational measurement frequency in Norway, was chosen for further analysis by Wolff et al. (2015).

Because catch errors are best described using catch efficiency, described as the ratio between precipitation accumulated in a gauge under test and the~~a~~ standard precipitation accumulation ~~gauge~~ ($CE = {P_{UT}}/{P_{DFIR}}$, where CE is catch efficiency, P_{UT} is the accumulation of precipitation from a gauge under test, and P_{DFIR} is the accumulated DFIR precipitation used as the standard), a minimum threshold is necessary to constrain errors in the denominator of this ratio. Using all the 30-minute periods of snow measured within the SDFIR and DFIR gauges from the US site, a minimum 30-minute precipitation amount of 0.25 mm in the DFIR gauge was found to provide a good balance between reducing the effects of measurement noise while simultaneously maintaining a large sample size of events. This was examined by iteratively increasing the DFIR precipitation threshold from zero in 0.01 mm steps, and calculating a simple linear transfer function for each threshold. The number of 30-minute events (n) and standard deviation (σ) of the CE model were estimated for every 0.01 mm increase in threshold, and a minimum in the standard error ($SE = {\sigma}/{\sqrt{n}}$) was encountered at 0.25 mm (Fig. 3). For this threshold test the SDFIR gauge data was used to develop transfer functions, as it was the most comparable to the DFIR gauge (Table 1) and was described well by a simple linear transfer function for snow events. In reality, some variability occurs in this 'ideal' threshold based on the specific gauge/shield under test, the sample size, precipitation type, and the amount of noise in the measurements. However for the sake of simplicity and consistency the same 0.25 mm reference threshold was used for all of the ~~gauges~~precipitation measurements- included in this study.

Using the original 30-minute SDFIR precipitation measurements, which were generally nearly equal to the DFIR measurements, we discovered that a minimum threshold for the gauge under test was also ~~necessary~~ required to help select ~~produce unbiased transfer functions~~representative precipitation measurements, rather than a biased sub-selection of measurements. This was because both the gauge under test and the standard DFIR gauge were affected by random measurement error and random spatial variability in precipitation; the ~~standard gauge/shield~~DFIR gauge ~~combination~~ is capable of measuring more than 0.25 mm of precipitation in a 30-minute period even when the actual site-average rate is below this threshold. Because many solid precipitation events occur near the 0.25 mm threshold, many such events may be included in the analysis. Without the use of a minimum threshold for a gauge under test, the measurements from a gauge under test may therefore be erroneously biased low (as a result of the DFIR being erroneously biased high), and the resultant transfer function may over-correct the gauge under test. The minimum thresholds for the gauges under test were determined from Eq. 2, using the entire multi-year datasets available.

$$THOLD_{UT} = \frac{median(P_{UT})}{median(P_{DFIR})} 0.25\,mm \qquad (2)$$

Where $THOLD_{UT}$ is the threshold of the gauge under test, P_{UT} is the 30-minute accumulation from the gauge under test, and P_{DFIR} is the 30-minute DFIR accumulation. The resultant thresholds were 0.18 mm for the unshielded gauge, 0.20 mm for the single Alter gauges, 0.21 mm for the double Alter gauges, 0.22 mm for the Belfort double Alter gauge, and 0.25 mm for the SDFIR gauge.

3.542.5.3 Choice of transfer function model

Wolff et al. (2015) tested many sigmoidal type transfer functions for the determination of catch efficiency from a single Alter gauge as a function of T_{air} and U. The equation they selected using Bayesian analysis was sigmoidal both in respect to its response to T_{air} and U. It is included here for reference, as it is used throughout this study:

$$CE = \left[1 - \tau_1 - (\tau_2 - \tau_1)\frac{e^{\left(\frac{T_{air}-T_\tau}{s_\tau}\right)}}{1+e^{\left(\frac{T_{air}-T_\tau}{s_\tau}\right)}} e^{-\left(\frac{U}{\theta}\right)^{\beta}}\right] + \tau_1 + (\tau_2 - \tau_1)\frac{e^{\left(\frac{T_{air}-T_\tau}{s_\tau}\right)}}{1+e^{\left(\frac{T_{air}-T_\tau}{s_\tau}\right)}}, \qquad (3)$$

where T_{air} is the air temperature, U is the wind speed, and $\tau_1, \tau_2, T_\tau, s_\tau, \theta, \beta,$, are coefficients fit to the data described in more detail in Wolff et al (2015). An example of the sigmoid function fit to SA CE measurements from both NOR and US is shown in FigureFig. 4.

We also propose an alternative function developed for the sake of simplicity, which responds with an exponential decrease in CE to wind speed, and with a simple sigmoid ($\tan^{-1}$) response to T_{air}:

$$CE = e^{-a(U)(1-[\tan^{-1}(b(T_{air}))+c])}, \qquad (4)$$

where a, b, and c are coefficients fit to the data. The form of both of these functions follows the same form presented by others, with catch efficiency rather than correction factor determined, such that the inverse (CE^{-1}) must be used to correct actual precipitation data ($P_{DFIR} = P_{UT}/CE$). Comparisons of corrections based on correction factors (P_{DFIR}/P_{UT}) and catch efficiency (P_{UT}/P_{DFIR}) revealed no significant differences, so CE is used in the present study because it enables comparison with past experiments. However it is difficult to interpret CE error estimates (eg. Wolff et al., 2015), so special care has been taken in the present study to estimate transfer function uncertainties and biases that are relevant to the measurement of precipitation, rather than the measurement of CE.

3.652.6 Uncertainty aAssessment

3.652.6.1 Measurement uUncertainty

The repeatability or random error of the 30-minute precipitation measurements can be estimated using the three four different sets of replicate or near-replicate sets of identical gauge-shield combinations available at the US and NOR testbeds. This type of analysis includes both random gauge measurement uncertainty and uncertainty caused by the spatial variability in precipitation occurring across a field site. At the US site there were two SA-shielded gauges and two DA-shielded gauges recording precipitation measurements throughout the field study. There was also one full-sized DFIR gauge and SDFIR-shielded gauge, which were similar enough in their catch to be considered identical for the purposes of estimating measurement uncertainty. At the NOR site there were two pairs of near-identical SA-shielded gauges, but the two gauges were on opposite sides of the DFIR. This limited the amount of data available for comparison between them, as additional screening for wind directions was necessary. In addition, two different heating systems were used on the NOR gauges, with the primary gauge used for the transfer function analysis configured with the NOAA Climate Reference Network heating system, and the secondary SA gauge using the Geonor heater system. Figure 5 shows the results of the comparison of these ~~three~~ four sets of identical or near-identical gauge-shield combinations. The root mean square errors (RMSE) were calculated from differences between the identical gauges, and were < 0.1 mm. These results include only snowfall, with precipitation type at the US site determined using the present weather detector by classifying 30-min periods with more than 15 min of snow and less than 5 min of other precipitation types as snow, and precipitation type at the NOR site determined using the air temperature, with snowfall identified as T_{air} < -2 °C.~~.~~

3.65~~2.6~~.2 Uncertainty of t~~T~~ransfer f~~F~~unctions

Uncertainty in the transfer functions was estimated by applying the transfer functions to the different gauges under test and then comparing the results to the standard DFIR precipitation. ~~To assess the SA NOR and the SA US transfer function uncertainty, the transfer function developed using the combined results from both sites was applied at the individual sites. This was done to assess to what degree a 'universal' transfer function was valid for the two sites. For the other US gauges, in order t~~To maintain some independence between the data used to develop the transfer function and the data used to test the transfer function, the uncertainty of the transfer functions was estimated using a 10-fold cross-validation. This model validation technique randomly separated the available measurements of a given shield type into 10 equally sized groups, determined the transfer function using 90% of the data, tested the transfer function on the remaining 10% of the data, and was repeated for all 10 groups. The coefficients describing the transfer function were based on the entire data set, but the uncertainty estimates were based on this 10-fold cross-validation. The uncertainty estimated from 10-fold cross validation and the uncertainty estimated by circularly developing and testing the transfer function on identical data were very similar, but the 10-fold validation was used where possible for more defensible estimates of the transfer function uncertainty. For the SDFIR-shielded gauge under test it proved impossible to constrain the sigmoid function for all 10 iterations of the model fitting, so the cross-validation was not used and the uncertainty and the transfer function were circularly determined using exactly the same data. To assess the uncertainty of the two-site transfer function developed for the combined SA NOR and SA US measurements, the transfer function was applied to the entire dataset using the 10-fold cross validation, and it was also applied to the two sites individually. The site-specific results were calculated to assess to what degree a single transfer function was valid for the two individual sites. Because the dataset used to create the transfer function was split into two groups for this two-site validation, 10-fold cross-validation was not used.

4~~3~~ Results~~ and Discussion~~

4~~3~~.1 Shielding e~~E~~rrors and ~~B~~biases

Errors in uncorrected precipitation measurements were estimated by comparing the DFIR-shielded precipitation measurements to the other shielded and unshielded measurements. Based on the differences between the 60-min DFIR accumulation and the single Alter (SA) accumulation at the NOR site, root mean square errors (RMSE) and biases were calculated (Table 1, NOR SA). The RMSE of the NOR SA measurements was 0.64 mm (51.6%) and the bias was -0.34 mm (27.1%), indicating that errors in the uncorrected data were significant. At the US site the 30-min precipitation measurements were similarly used to calculate ~~RMSEs~~RMSE values and biases (Table 1), with the ~~RMSEs~~RMSE values and biases generally decreasing in absolute magnitude as the size or efficacy of the shield increased. For example the absolute magnitudes of the unshielded gauge (US UN) RMSE (0.30 mm or 28.6%) and bias (-0.17 mm or -16.2%) at the US site were much larger than the SDFIR (US SDFIR) RMSE (0.14 mm or 14.7%) and bias (-0.03 mm or -3.6%). ~~As expected, t~~The RMSE and bias for the combined SA dataset (All SA) that included US and NOR measurements fell between the US SA and the NOR SA results. In addition to quantifying the errors occurring with the different shields at the sites, these uncorrected results also serve to provide some prospective for the corrections that were developed and applied to the precipitation measurements. ~~Error~~ RMSE and bias estimates reported in both mm and percent further demonstrate that a relatively small error or bias reported in mm can actually be quite significant in terms of percent. This is due to the fact that many of the 30 or 60 minute precipitation measurements, particularly for snow, included less than 0.5 mm of accumulation.

4~~3~~.2 Transfer f~~F~~unctions and u~~U~~ncertainty

From the US site, the resultant precipitation catch efficiencies were described as a function of T_{air} and U using both the sigmoid function (Eq. 3) and exponential function (Eq. 4) for the unshielded (UN), single Alter (SA), double Alter (DA), Belfort double Alter (BDA), and small DFIR (SDFIR) Geonor T-200B precipitation gauges. These transfer functions were developed for both the gauge height wind speed (Table 2) and the 10 m height wind speeds (Table 5). In addition to creating transfer functions for the individual NOR and US site SA measurements, the single alter (SA) results from the NOR and US sites were combined and used to create two-site exponential (Exp) and sigmoid (Sig) transfer functions (labelled as 'All SA' in Table 2~~3~~ and 5~~4~~).

The RMSE and bias in the corrected measurements were then determined for every transfer function and its corresponding precipitation measurements for the gauge height wind speeds (Tables 3) and the 10 m height wind speeds (Table 6). This includes t~~T~~he two-site SA transfer functions ~~developed using all the data were used to determine the~~and the associated errors ~~uncertainty~~ ~~in~~ determined from the combined dataset ('All SA' in Table 3~~2~~ and 6~~3~~). In addition, by applying the SA All transfer functions individually to the NOR and US SA measurements, RMSE values and biases were estimated separately for the US and NOR SA measurements using the two-site transfer functions (Tables 4 and 7). This was done ~~by individually~~

applying transfer functions developed from SA data to the two separate sites and calculating the uncertainty and biases (shown in NOR SA and US SA in Tables 2 and 3). This approach was chosen to evaluate site-biases and the effects of different climates on the transfer functions. Results are described separately for the gauge-height wind speeds (Table 2) and the 10 m wind speeds (Table 3). For example, for the SA gauges, at the US site the two-site gauge height wind speed Exp transfer function reduced the RMSE from 23.6% (Table 1) to 15.4% (Table 42) and improved the bias from -11.7% (Table 1) to -0.21% (Table 42). Likewise the NOR SA RMSE was decreased from 51.6% (Table 1) to 36.49% (Table 42) and the bias was improved from -27.1% (Table 1) to -4.13.3% (Table 2). This indicates that the two-site transfer function effectively removed reduced the bias at both sites. The NOR SA and US SA transfer functions determined separately for the two sites did not perform significantly better at the individual sites (Table 3 and 6) than the two-site transfer function (Table 4 and 7), with the only notable improvements from the individual site transfer functions in the NOR biases. For example, using the Exp function for the gauge height wind speeds, the individual site transfer function improved the bias to -1.1%, whereas the two-site transfer function only improved it to -4.1%.

. The RMSE values of the corrected results reflects the significant residual variability in the corrected catch efficiency. It is worth noting that the RMSE of even the corrected SDFIR measurements was greater than 0.1 mm, indicating such uncorrected errors may be due to random measurement error and site variability rather than crystal type and wind speed; even a gauge that is relatively unaffected by wind speed with respect to the DFIR was subject to such errors..

Generally, differences between the Exp and Sig functions were quite small, indicating that the Exp function can be used as a simpler alternative to the Sig function developed by Wolff et al. (2015). For the US site, the transfer function RMSE values were was about 0.15 mm (15%) for all the gauges irrespective of whether the sigmoid function or the exponential function was used. The errors in the uncorrected data (Table 1) were generally much larger than the errors in the corrected data, with the errors in the uncorrected data dependent upon the efficacy of the shield. For example, using the gauge height wind speed transfer functions the corrected SDFIR gauge RMSE (0.13 mm or ~14.2%) and bias (-0.01 mm or -1.6%) shown in both Table 2 and Table 3 were only slightly better than the uncorrected error RMSE (0.14 mm or 14.75%) and bias (-0.03 mm or -3.64%) shown in Table 1, whereas application of the transfer function resulted in a much more significant improvement in the unshielded and SA gauge error, (UN and SA, in Tables 1, 32, and 63).

TFor the combined NOR and US SA transfer functions, the RMSE values were much larger for the NOR SA measurements than the US SA measurements. This was due primarily to a generally more noisy gauge at the NOR site than at the US site. Random measurement error from these vibrating-wire weighing gauges can be reduced via trial and error by rotating and remounting the vibrating wires within the gauge and also by mounting the shield separately from the gauge, but such noise can vary significantly from gauge to gauge and even from wire to wire within a gauge equipped with redundant vibrating wires. The NOR site is also windier than the US site and the undercatch is generally larger, and because of this one can expect the RMSE to increase. A well-shielded gauge that requires less correction will be less affected by variability in precipitation type and crystal habit, for example. In addition blowing snow may have increased the RMSE values of the NOR results, with 2.6% of the events occurring wind speeds greater than 15 ms^{-1}.

The bias found for the NOR SA gauge corrected using the two-site transfer function is perhaps however more notable, as the bias should be relatively unaffected by random measurement noise. For example, tThe NOR SA measurements corrected using the gauge-height transfer function had a larger bias (approximately -0.05 mm or -4% as the mean of the Exp and Sig function results) than the US SA gauge (0.00 mm or ~0%) possibly because there were more US SA events (N = 843) than NOR SA events (N = 352), and the resultant transfer function was therefore weighted more towards the US SA catch efficiency (Table 2 and 3). More importantly, tThis NOR SA bias indicates that small site-biases exist that can effect such 'universal' multi-site transfer functions, and such a bias has been quantified here using automated gauges for the first time.

For the gauges at the US site, there was no difference in the RMSE or bias between the 10 m wind speed and the gauge-height wind speed transfer functions. However, this is neither surprising nor noteworthy as the gauge-height wind speed was estimated based on the 10 m wind speed. It is notable however that the RMSE for the combined US and NOR SA gauge results was also not significantly affected by the choice of wind speed measurement height (Table 42 and 73). These RMSE data values indicate that although the gauge heights at the US site and the NOR site were significantly different, there was no significant loss in accuracy when transfer functions were created for both sites using the 10 m wind speed. There was however a more negative bias in the 10 m wind speed *WS* NOR SA results (eg. Sig bias = -: 0.10 mm, or -7.8%) than for the gauge height wind speed *WS* NOR SA results (eg. Sig bias := -0.06 mm, or -5.2%), indicating that there may be a small advantage to using the gauge height wind speed in preference to the 10 m wind speed for the development and application of precipitation transfer functions.

To demonstrate both the importance and the limitations of the transfer function corrections, the 30-min uncorrected (Fig 6 a) and corrected (Fig 6 b) SA snow (T_{air} < -2.5 °C) measurements were compared to the DFIR precipitation. The uncorrected SA snow measurements are subject to significant errors as a result of variability in wind speed and crystal type, with dense, wet, warm snow and low-wind speed snow less affected by shielding than cold, dry, light and windy snow. In addition, the corrected SA snow measurements reveal the effects of gauge noise, the spatial variability of precipitation, and also variability in crystal habit that are inadequately captured by air temperature. For example, at a given temperature variability in crystal type has a significant effect on hydrometeor fall velocity, drag coefficient, and the resultant relationship between wind speed and *CE* (Colli et al., 2015; Theriault et al., 2012).

To further demonstrate the necessity of the transfer functions and the effects of errors and variability in the transfer functions, an example event is shown Fig. 7 a. 1One min accumulations during thise Figure 1 March, 2013 event were corrected using the appropriate transfer function and the mean 1-min T_{air} and *U* (Fig. 7 b). This ‘typical’ event was in fact somewhat atypical, with *CE* lower than predicted by the transfer functions. This is another example of the results shown in ~~Figure~~Fig. 6 b, with some periods significantly overcorrected and others significantly under-corrected by transfer functions. These types of events serve as a good example of why it is always preferable to make the most accurate measurement possible, and only rely upon corrections when absolutely necessary. In comparison with the uncorrected precipitation values shown in ~~Figure~~Fig. 7 a~~1~~ however, ~~Figure~~Fig. 7 b also demonstrates the necessity of such transfer functions. For example, with a standard DFIR accumulation of 22.5 mm for the entire event, the UN accumulation was improved from 8.5 mm (38% of the DFIR) to 17.8 mm (79% of the DFIR). When precipitation measurements are used to help describe water budgets, the variability in the corrected measurements demonstrated in Fig. 6 and 7 may be relatively unimportant relative to the improvement in the bias, or the total amount of precipitation.

5~~4~~ Discussion

5~~4~~.1 Wind s~~S~~peed for t~~T~~ransfer f~~F~~unctions

Wind speed measurements at the gauge-height and the standard 10 m measurement height both have advantages. The 10 m height wind speed is widely used by national weather services~~available~~, it is designated as the standard wind speed measurement height by the WMO, and measurements at this height are also less likely to be affected by obstacles such as towers and precipitation gauge shields. The effects of obstacles at gauge height were apparent in both the NOR and US sites, for example. However the catch efficiency of a shielded or unshielded gauge is more closely linked to the wind speed at the gauge height. If, for example, the wind speed at gauge height is affected by the changing height of the snowpack or by vegetation or other obstacles this will affect the relationship between the 10 m and gauge height wind speed, and potentially lead to additional sources of error. ~~It is worth noting however that at the NOR site, where snow depth during this measurement campaign was a maximum of almost 2 m, changing snow depth did not significantly affect the relationship between the 10 m wind speed and the gauge height wind speed.~~

I~~However i~~n the present study, where ~~an estimated gauge height wind speed and~~ a 10 m wind speed and a gauge height wind speed derived from the 10 m wind speed were both used to create two-site transfer functions for gauges at significantly different heights, the ~~uncertainty (~~RMSE~~)~~ of the combined NOR and US transfer function showed no significant change when tested on the US, NOR, or combined datasets (Table 4 and 9). The only notable change caused by the use of the 10 m wind speed versions of the combined transfer function was in the bias calculated from the NOR SA, which was about -4% (-0.05 mm) for the gauge height transfer functions (Table 4) and -7% (-0.09 mm) for the 10 m wind speed transfer functions (Table 7). ~~Although transfer functions based on the 10 m wind speed are included in this study, we recommend using the gauge height wind speed when~~However based on first principles, for the correction of precipitation measurements the use of a gauge height wind speed measurement or an approximation of the gauge height wind speed ~~is available~~is more defensible than the use of a 10 m height wind speed~~ or an approximation of the gauge height wind speed~~. Because differences in gauge height are common due to the necessity of mounting gauges and shields well above the highest expected snow depth, even when only 10 m wind speed measurements are available they should be adjusted to gauge height for ~~a more defensible~~the application of shielding corrections. For example, at a gauge height of 5 m, the wind speed affecting the catch efficiency of the gauge would typically be ~90% of the 10 m wind speed, while the wind speed affecting a gauge at 2 m would be ~70% of the 10 m wind speed. Using the log wind profile and/or other available wind speed profile measurements as demonstrated here, the commonly available 10 m wind speed can be used to estimate gauge height wind speeds and defensibly correct wind speed precipitation errors at all gauge heights. This approach has the advantage of being based on the arguably easier to measure and more commonly available 10 m height wind speed, but it also suffers from the disadvantage of being only an estimate of the gauge height wind speed. Due to obstacles at both the NOR and US site affecting the gauge height wind speeds, the gauge height and 10 m transfer functions were both by necessity derived from the 10 m height wind speeds, and the measurements therefore cannot easily be used to compare true gauge height wind speed and derived gauge height wind speed transfer functions. It is nevertheless interesting that a two-site 10 m wind speed transfer function performed almost as well as the estimated gauge height wind speed transfer function, especially considering the large difference between the gauge heights at the two sites. It is also worth noting that at the NOR site, where snow depth during this measurement campaign was a maximum of almost 2 m, changing snow depth did not significantly affect the relationship between the 10 m wind speed and the gauge height wind speed. Based on this it appears that despite the theoretical advantages of using gauge height wind speeds (or estimated gauge height wind speeds) for the correction of precipitation measurements, in practice the best available wind speed measurement will vary by site and by network, and in some cases the most representative wind speed measurement may not be at gauge height.~~ However in the present study, where an estimated gauge height wind speed and a 10 m wind speed were both used to create transfer functions for gauges at significantly different heights, the uncertainty (RMSE) of the combined NOR and US transfer function showed no significant change when tested on the US, NOR, or combined datasets. The only notable change caused by the use of the 10 m wind speed version of the combined transfer function was in the bias calculated from the NOR SA, which was about -4% (-0.05 mm) for the gauge height transfer function and -8% (-0.11 mm) for the 10 m *WS* transfer function.~~

~~Wind speed measurements at the gauge-height and the standard 10 m measurement height both have advantages. The 10 m height wind speed is widely available as the standard wind speed measurement height, and measurements at this height are also less likely to be affected by obstacles such as towers and precipitation gauge shields. The effects of obstacles at gauge height were apparent in both the NOR and US sites, for example. However the catch efficiency of a shielded or unshielded gauge is more closely linked to the wind speed at the gauge height. If, for example, the wind speed at gauge height is affected by the changing height of the snowpack or by vegetation or other obstacles this will affect the relationship between the 10 m and gauge height wind speed, and potentially lead to additional sources of error. It is worth noting however that at~~

~~the NOR site, where snow depth during this measurement campaign was a maximum of almost 2 m, changing snow depth did not significantly affect the relationship between the 10 m wind speed and the gauge height wind speed.~~

5~~4~~.2 A two-site~~‘Universal’~~ t~~T~~ransfer f~~F~~unction

Uncertainties and biases associated with the development and application of a single transfer function for ~~multiple~~ two separate sites within differing climate regions have been presented~~ here~~. ~~The fact that t~~The NOR site was much windier than the US site; during precipitation events the mean NOR $U_{10\,m}$ = 8.74 m s^{-1}, and the mean US $U_{10\,m}$ = 4.52 m s^{-1}. The NOR measurements were also generally noisier than the US results; for corrected SA events with similar wind speeds from both sites imposed by removing events with $U_{10\,m}$ > 10 m s^{-1}, the NOR RMSE = 0.38 mm and the US RMSE = 0.15 mm. It ~~and the NOR site is much windier than the US site makes it~~ is therefore difficult to draw site-specific conclusions based on the RMSE values as some of the gauge noise from NOR appears to be related to the specific installation rather than the climate. ~~.~~ However the bias found for the NOR SA results was more significant than for the US SA measurements. ~~This indicates that the combined transfer function was biased towards the US SA results, which outnumbered the NOR SA results by over 2:1.~~ This suggests that small ~~site~~ biases ~~do~~ between different sites and climates indeed exist, and the application of a ~~‘universal’~~multi-site transfer function based on wind speed and air temperature will be subject to such ~~site-to-site~~ variability. The soon-to-be released results of the WMO Solid Precipitation Intercomparison Experiment will provide a better opportunity for quantifying such climate~~site~~-biases, as measurements from the two sites from this study will be included along with many other sites in varying climates. Additionally, WMO-SPICE data from all sites will be processed using a common approach developed within the WMO-SPICE project (Reverdin, 2016).

Such a ~~‘universal’~~ multi-site transfer function is needed because site-specific transfer functions only exist at sites where a DFIR is already present, and therefore a more generally applicable transfer function is necessary for real-world precipitation corrections where the actual or DFIR-shielded amount of precipitation is unknown. In this study, the bias was shown to be minimized at sites in two separate climates~~sites~~ using one transfer function that was developed from combined results, and we suggest using this same approach in future studies such as the WMO Solid Precipitation Intercomparison, for which many more sites and climates will be included, allowing ~~and~~ multi-site~~‘universal’~~ transfer functions ~~can~~ to be defined more defensibly.

6~~5~~. Conclusions

Methods to address the effects of wind on precipitation measurements have been presented, and significant improvements to the measurements have been made available for unshielded gauges and other common gauge/shield combinations~~widely used wind shields~~. A new adjustment function was used to describe~~A simpler function to describe~~ catch efficiency as a function of air temperature and wind speed, and it performed as well as the more complex function suggested by Wolff et al.

(2015). In addition, the remaining uncertainty in the transfer functions used to correct or standardize precipitation measurements has also been carefully described and quantified. Significant errors persisted in the measurements even after correction for undercatch, with the RSME reduced by less than 50% for every windshield examined. Measurement error, the random spatial variability of precipitation, and variability in the type, size, density, and fall speed of hydrometeors all likely contributed to the errors that remained uncorrected. This is an active ongoing area of research that merits more attention. In addition, this study indicates that low-porosity windshields like the Belfort double Alter show great promise in reducing undercatch with a small-footprint, low-maintenance shield and merit further research.

Significant errors exist in our historical and present-day precipitation measurements. For weighing gauges that are designed to measure snowfall, these errors are affected primarily by shielding, precipitation type, crystal habits, wet vs. dry snow, and wind speed. Such errors affect the measurement of the amount of water in both seasonal and ephemeral snowpack, and therefore affect our ability to quantify the availability of water for communities and ecosystems that rely upon water from snowfall. The results and techniques presented here can be used to help create precipitation records that are traceable to a common standard, ultimately leading to a better constrained and more accurate understanding of the earth's hydrological balance.

Acknowledgements

We thank Hagop Mouradian from Environment Canada and Climate Change for contributing the mapped site locations (Fig. 1 a). We also thank Samuel Buisan from the Spanish National Meteorological Agency and Eva Mekis and Michael Earle from Environment Canada and Climate Change for carefully reviewing this manuscript. This work was greatly improved by their comments.

75. References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, Journal of Geophysical Research-Atmospheres, 108, 2003.

Alter, J. C.: Shielded storage precipitaiton gages, Monthly Weather Review, 65, 262-265, 1937.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303-309, 2005.

Blunden, J. and Arndt, D. S.: State of the climate in 2015, Bulletin of the American Meteorological Society, 97, Si-S275, 2016.

Colli, M., Rasmussen, R., Thériault, J. M., Lanza, L. G., Baker, C. B., and Kochendorfer, J.: An improved trajectory model to evaluate the collection performance of snow gauges, Journal of Applied Meteorology and Climatology, 54, 1826-1836, 2015.

Duchon, C. E. and Essenberg, G. R.: Comparative rainfall observations from pit and aboveground rain gauges with and without wind shields, Water Resources Research, 37, 3253-3263, 2001.

Folland, C. K.: Numerical models of the raingauge exposure problem, field experiments and an improved collector design, Quarterly Journal of the Royal Meteorological Society, 114, 1485-1516, 1988.

Førland, E. J. and Hanssen-Bauer, I.: Increased precipitation in the Norwegian Arctic: True or false?, Climatic Change, 46, 485-509, 2000.

Goodison, B., Louie, P., and Yang, D.: The WMO solid precipitation measurement intercomparison, World Meteorological Organization-Publications-WMO TD, 65-70, 1998.

Groisman, P. Y., Koknaeva, V. V., Belokrylova, T. A., and Karl, T. R.: Overcoming biases of precipitation measurement - a history of the USSR experience, Bulletin of the American Meteorological Society, 72, 1725-1733, 1991.

Groisman, P. Y. and Legates, D. R.: The accuracy of United-Sates precipitaiton data, Bulletin of the American Meteorological Society, 75, 215-227, 1994.

Leeper, R. D., Palecki, M. A., and Davis, E.: Methods to Calculate Precipitation from Weighing-Bucket Gauges with Redundant Depth Measurements, Journal of Atmospheric and Oceanic Technology, 32, 1179-1190, 2015.

Mekis, É. and Vincent, L. A.: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada, Atmosphere-Ocean, 49, 163-177, 2011.

Merenti-Välimäki, H.-L., Lönnqvist, J., and Laininen, P.: Present weather: comparing human observations and one type of automated sensor, Meteorological Applications, 8, 491-496, 2001.

Merenti-Valimaki, H. L., Lonnqvist, J., and Laininen, P.: Present weather: comparing human observations and one type of automated sensor, Meteorological Applications, 8, 491-496, 2001.

Nipher, F. E.: On the determination of the true rainfall in elevated gauges, American Association for the Advancement of Science, doi: doi:10.1175/1520-0493(1937)65<262:SSPG>2.0.CO;2, 1878. 103-108, 1878.

Nitu, R., Rasmussen, R., Smith, C. D., Earle, M., Roulet, Y. A., Reverdin, A., Wolff, M., Baker, C. B., and Kochendorfer, J.: WMO solid precipitation intercomparison: from experiments to results, New Orleans, LA, Jan, 2016 2016.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Theriault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, Bulletin of the American Meteorological Society, 93, 811-829, 2012.

Rasmussen, R., Dixon, M., Hage, F., Cole, J., Wade, C., Tuttle, L., McGettigan, S., Carty, T., Stevenson, L., Fellner, W., Knight, S., Karplus, E., and Rehak, N.: Weather support to deicing decision making (WSDDM): A winter weather nowcasting system, Bulletin of the American Meteorological Society, 82, 579-595, 2001.

Rasmussen, R. M., Vivekanandan, J., Cole, J., Myers, B., and Masters, C.: The estimation of snowfall rate using visibility, Journal of Applied Meteorology, 38, 1542-1563, 1999.

Reverdin, A.: Description of the Quality Control and Event Selection Procedures used within the WMO-SPICE project, Madrid, Spain, Sept, 2016 2016.

Scaff, L., Yang, D., Li, Y., and Mekis, E.: Inconsistency in precipitation measurements across the Alaska-Yukon border, Cryosphere, 9, 2417-2428, 2015.

Sheppard, B. E. and Joe, P. I.: Automated precipitation detection and typing in winter: A two-year study, Journal of Atmospheric and Oceanic Technology, 17, 1493-1507, 2000.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Springer Netherlands, 1988.

Theriault, J. M., Rasmussen, R., Ikeda, K., and Landolt, S.: Dependence of Snow Gauge Collection Efficiency on Snowflake Characteristics, Journal of Applied Meteorology and Climatology, 51, 745-762, 2012.

Thom, A. S.: Momentum, mass and heat exchange of plant communities, Academic Press, 1975.

Trenberth, K. E.: Changes in precipitation with climate change, Climate Research, 47, 123-138, 2011.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, Bulletin of the American Meteorological Society, 84, 1205, 2003.

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Fenimore, C., Gleason, K., and Arndt, D.: Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions, Journal of Applied Meteorology and Climatology, 53, 1232-1251, 2014.

Wolff, M., Brækkan, R., Isaksen, K., and Ruud, E.: A new testsite for wind correction of precipitation measurements at a mountain plateau in southern Norway, Proceedings of WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO–2010), Instruments and Observing Methods Report No. 104, WMO/TD-No. 1546, Geneva, 2010.

Wolff, M., Isaksen, K., Braekkan, R., Alfnes, E., Petersen-Overleir, A., and Ruud, E.: Measurements of wind-induced loss of solid precipitation: description of a Norwegian field study, Hydrology Research, 44, 35-43, 2013.

Wolff, M. A., Isaksen, K., Petersen-Overleir, A., Odemark, K., Reitan, T., and Braekkan, R.: Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study, Hydrology and Earth System Sciences, 19, 951-967, 2015.

Wong, K.: Performance of several present weather wensors as precipitation gauges, Brussels, Belgium2012, 25.

Yang, D.: Double Fence Intercomparison Reference (DFIR) vs. Bush Gauge for “true” snowfall measurement, Journal of Hydrology, 509, 94-100, 2014.

Yang, D. Q., Goodison, B. E., Metcalfe, J. R., Golubev, V. S., Elomaa, E., Gunther, T., Bates, R., Pangburn, T., Hanson, C. L., Emerson, D., Copaciu, V., and Miklovic, J.: Accuracy of Tretyakov precipitation gauge: Result of WMO intercomparison, Hydrological Processes, 9, 877-895, 1995.

Yang, D. Q., Goodison, B. E., Metcalfe, J. R., Louie, P., Leavesley, G., Emerson, D., Hanson, C. L., Golubev, V. S., Elomaa, E., Gunther, T., Pangburn, T., Kang, E., and Milkovic, J.: Quantification of precipitation measurement discontinuity induced by wind shields on national gauges, Water Resources Research, 35, 491-508, 1999.

Yang, D. Q., Kane, D., Zhang, Z. P., Legates, D., and Goodison, B.: Bias corrections of long-term (1973-2004) daily precipitation data over the northern regions, Geophysical Research Letters, 32, 2005.

6. Tables and Figures

Shield	Uncor RMSE	Uncor Bias
US UN	0.30 mm, 28.6%	-0.17 mm, -16.2%
All SA	0.35 mm, 34.3%	-0.16 mm, -16.1%
NOR SA	0.64 mm, 51.6%	-0.34 mm, -27.1%
US SA	0.22 mm, 23.6%	-0.11 mm, -11.7%
US DA	0.21 mm, 21.6%	-0.10 mm, -10.6%
US BDA	0.16 mm, 17.5%	-0.05 mm, -5.6%
US SDFIR	0.14 mm, 14.7%	-0.03 mm, -3.6%

Table 1. Errors and biases in the uncorrected 30-min precipitation from gauges under test, estimated using the DFIR precipitation measurements as the standard.

Shield	Sig Coef						Exp Coef			n	Max U
	τ_1	τ_2	T_τ	s_τ	θ	β	a	b	c		
US UN	0.31	0.94	-0.08	0.92	2.58	1.23	0.063	1.22	0.66	843	6 m s^{-1}
All SA	0.20	0.96	0.22	1.11	4.70	1.97	0.040	1.10	0.54	1501	9 m s^{-1}
NOR SA	0.26	1.02	0.88	0.99	3.1	1.61	0.054	0.71	0.26	352	9 m s^{-1}
US SA	0.16	0.95	-0.34	1.01	4.9	1.90	0.036	1.04	0.63	1156	6 m s^{-1}
US DA	0.00	0.92	-1.19	1.89	7.04	1.36	0.028	0.74	0.66	1392	6 m s^{-1}
US BDA	0.00	1.00	1.81	0.57	8.73	2.87	0.015	0.32	0.38	1204	6 m s^{-1}
US SDFIR	0.99	0.96	0.52	0.10	0.14	6.16	0.006	0.00	0.00	1508	6 m s^{-1}

Table 2. Transfer function coefficients for estimated gauge height wind speeds. Coefficients for the sigmoid transfer function (Sig, Eq. 3) and the exponential transfer function (Exp, Eq. 4) as well as the number of periods available (*n*) and the maximum wind speed (Max *U*) are described for the US unshielded (UN), single Alter (SA), double Altar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single Alter (SA) gauge, and the combined US and NOR SA results (All SA).

Shield	Sig RMSE	Sig Bias	Exp RMSE	Exp Bias
US UN	0.18 mm, 16.8%	0.00 mm, 0.4%	0.18 mm, 16.8%	0.00 mm, 0.4%
All SA	0.23 mm, 22.3%	-0.02 mm, -1.5%	0.25 mm, 24.0%	-0.02 mm, -1.7%
NOR SA	0.46 mm, 36.9%	-0.06 mm, -4.5%	0.47 mm, 37.7%	-0.01 mm, -1.1%
US SA	0.13 mm, 14.0%	-0.01 mm, -0.6%	0.14 mm, 14.9%	-0.00 mm, -0.5%
US DA	0.13 mm, 13.8%	0.00 mm, 0.0%	0.14 mm, 13.8%	0.00 mm, 0.1%
US BDA	0.12 mm, 13.6%	-0.01 mm, -1.7%	0.13 mm, 14.6%	0.00 mm, -0.9%
US SDFIR	0.13 mm, 14.2%	-0.01 mm, -1.6%	0.13 mm, 14.2%	-0.01 mm, -1.6%

Table 3. Transfer function results for estimated gauge height wind speeds. Sigmoid transfer function (Sig, Eq. 3) and exponential transfer function (Exp, Eq. 4) RMSE values and biases described for the US unshielded (UN), single Alter (SA), double Altar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single Alter (SA) gauge, and the combined US and NOR SA results (All SA).

Shield	Sig RMSE	Sig Bias	Exp RMSE	Exp Bias
NOR SA	0.46 mm, 36.5%	-0.06 mm, -4.7%	0.45 mm, 36.4%	-0.05 mm, -4.1%
US SA	0.14 mm, 14.6%	0.00 mm, 0.0%	0.15 mm, 15.4%	0.00 mm, -0.2%

Table 4. Transfer function results for estimated gauge height wind speeds. The same coefficients determined by the multi-site all SA transfer function (Table 2) were used to correct the results from NOR and US. Sigmoid transfer function (Sig, Eq. 3) and exponential transfer function (Exp, Eq. 4) RMSE values and biases are described for the US single Alter (SA) and the NOR single Alter (SA) gauge.

Shield	Sig Coef						Exp Coef			n	Max U
	τ_1	τ_2	T_τ	s_τ	θ	β	a	b	c		
US UN	0.31	0.94	-0.08	0.92	3.58	1.23	0.045	1.21	0.66	843	8 m s^{-1}
All SA	0.17	0.96	0.23	1.11	6.46	2.01	0.03	1.04	0.57	1501	12 m s^{-1}
NOR SA	0.25	1.03	0.95	1.06	3.99	1.61	0.05	0.66	0.23	352	12 m s^{-1}
US SA	0.12	0.95	-0.35	1.01	7.05	1.87	0.03	1.06	0.63	1156	12 m s^{-1}
US DA	0.00	0.92	-1.19	1.89	9.75	1.36	0.021	0.74	0.66	1392	8 m s^{-1}
US BDA	0.31	1.00	1.79	0.58	10.0	3.15	0.01	0.48	0.51	1204	8 m s^{-1}
US SDFIR	0.99	0.96	0.52	0.10	0.14	10.75	0.004	0.00	0.00	1508	8 m s^{-1}

Table 5. Transfer function coefficients for 10 m height wind speeds. Coefficients for the sigmoid transfer function (Sig, Eq. 3) and the exponential transfer function (Exp, Eq. 4) as well as the number of periods available (*n*) and the maximum wind speed (Max *U*) are described for the US unshielded (UN), single Alter (SA), double Altar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single Alter (SA) gauge, and the combined US and NOR SA results (All SA).

Shield	Sig RMSE	Sig Bias	Exp RMSE	Exp Bias
US UN	0.18 mm, 16.8%	0.00 mm, 0.4%	0.19 mm, 17.5%	0.00 mm, 0.6%
All SA	0.24 mm, 22.9%	-0.02 mm, -1.8%	0.25 mm, 24.4%	-0.02 mm, -2.0%
NOR SA	0.46 mm, 37.3%	-0.06 mm, -4.5%	0.46 mm, 36.7%	-0.02 mm, -1.3%
US SA	0.13 mm, 14.0%	-0.01 mm, -0.6%	0.14 mm, 15.2%	-0.00 mm, -0.5%
US DA	0.13 mm, 13.9%	0.00 mm, 0.0%	0.14 mm, 13.8%	0.00 mm, 0.1%
US BDA	0.12 mm, 13.6%	-0.01 mm, -1.7%	0.13 mm, 14.6%	-0.00 mm, -0.9%
US SDFIR	0.13 mm, 14.1%	-0.01 mm, -0.09%	0.13 mm, 14.2%	-0.01 mm, -1.6%

Table 6. Transfer function results for 10 m height wind speeds. Sigmoid transfer function (Sig, Eq. 3) and exponential transfer function (Exp, Eq. 4) RMSE values and biases described for the US unshielded (UN), single Alter (SA), double Altar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single Alter (SA) gauge, and the combined US and NOR SA results (All SA).

Shield	Sig RMSE	Sig Bias	Exp RMSE	Exp Bias
NOR SA	0.47 mm, 36.8%	-0.09 mm, -6.7%	0.44 mm, 34.6%	-0.09 mm, -7.3%
US SA	0.14 mm, 14.7%	0.00 mm, 0.5%	0.15 mm, 15.5%	0.00 mm, 1.0%

Table 7. Transfer function results for estimated 10 m height wind speeds. The same coefficients determined by the multi-site all SA transfer function (Table 3?) were used to correct the results from NOR and US. Sigmoid transfer function (Sig, Eq. 3) and exponential transfer function (Exp, Eq. 4) RMSE values and biases are described for the US single Alter (SA) and the NOR single Alter (SA) gauge.

Shield	Uncor RMSE	Uncor Bias
US UN	0.30 mm, 28.6%	-0.17 mm, -16.2%
All SA	0.35 mm, 34.3%	-0.16 mm, -16.1%
NOR SA	0.64 mm, 51.6%	-0.34 mm, -27.1%
US SA	0.22 mm, 23.6%	-0.11 mm, -11.7%
US DA	0.21 mm, 21.6%	-0.10 mm, -10.6%
US BDA	0.16 mm, 17.5%	-0.05 mm, -5.6%
US SDFIR	0.14 mm, 14.7%	-0.03 mm, -3.6%

Table 1. Errors and biases in the uncorrected 30-min precipitation from gauges under test, estimated using the DFIR precipitation measurements as the standard.

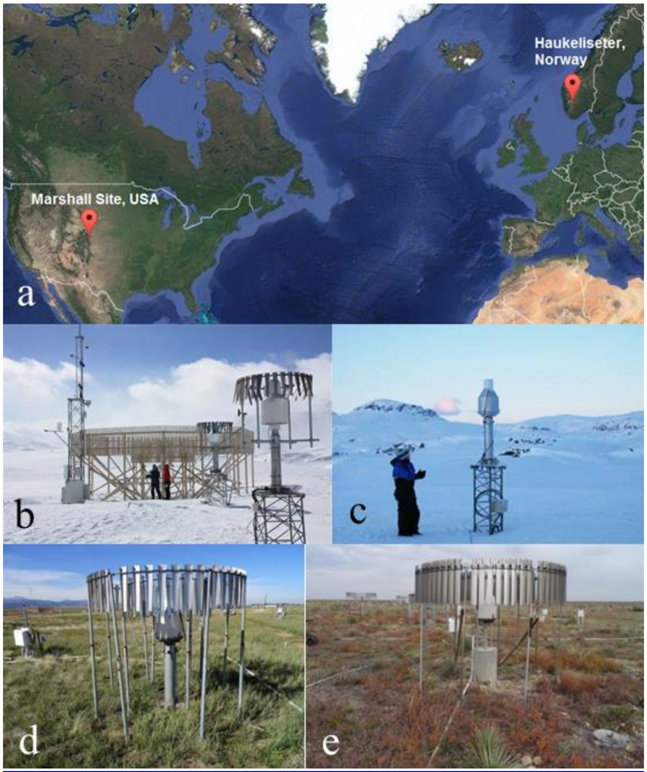
Shield	Sig Coef (τ_1, τ_2, T_τ, s_τ, θ, β)	Sig RMSE	Sig Bias	Exp Coef (a,b,c)	Exp RMSE	Exp Bias	*n*	Max *U*
US UN	0.31, 0.94, -0.08, 0.92 2.58, 1.23	0.18 mm, 16.8%	0.00 mm, 0.4%	0.063, 1.22, 0.66	0.18 mm, 16.8%	0.00 mm, 0.4%	843	6 m/s
All SA	0.20, 0.96, 0.22, 1.11, 4.70, 1.97	0.23 mm, 22.3%	-0.02 mm, -1.5%	0.04, 1.1, 0.54	0.25 mm, 24.0%	-0.02 mm, -1.7%	1501	9 m/s
NOR SA	--	0.46 mm, 36.9%	-0.06 mm, -5.2%	--	0.45 mm, 36.1%	-0.04 mm, -3.3%	352	9 m/s
US SA	--	0.14 mm, 14.5%	0.00 mm, 0.0%	--	0.15 mm, 15.4%	0.00 mm, -0.1%	1156	9 m/s
US DA	0.00, 0.92, -1.19, 1.89, 7.04, 1.36	0.13 mm, 13.8%	0.00 mm, 0.0%	0.028, 0.74, 0.66	0.14 mm, 13.8%,	0.00 mm, 0.1%	1392	6 m/s
US BDA	0.00, 1.00, 1.81, 0.57, 8.73, 2.87	0.12 mm, 13.6%	-0.01 mm, -1.7%	0.015115, 0.32, 0.38	0.13 mm, 14.6%	0.00 mm, -0.9%	1204	6 m/s
US	0.99, 0.96,	0.13 mm,	-0.01 mm,	0.006,	0.13 mm,	-0.01 mm,	1508	6 m/s

SDFIR	0.52, 0.10, 0.14, 6.16	14.2%	-1.6%	0.00, 0.00	14.2%	-1.6%		

Table 2. Transfer function results for estimated gauge height wind speeds. Coefficients for the sigmoid transfer function (Sig, Eq. 3) and the exponential transfer function (Exp, Eq. 4) and the associated RMSEs and Biases as well as the number of periods available (*n*) and the maximum wind speed (Max *U*) are described for the US unshielded (UN), single AlterAltar (SA), double Altar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single AlterAltar (SA) gauge, and the combined US and NOR SA results (All SA). The same coefficients determined by the All SA transfer function were used to separately correct the results from NOR and US.

Shield	Sig Coef (τ_1, τ_2, T_τ, s_τ, θ, β)	Sig RMSE	Sig Bias	Exp Coef (a,b,c)	Exp RMSE	Exp Bias	*n*	Max *U*
US UN	0.31, 0.94, -0.08, 0.92 3.58, 1.23	0.18 mm, 16.8%	0.00 mm, 0.4%	0.045, 1.21, 0.66	0.19 mm, 17.5%	0.00 mm, 0.6%	843	8 m/s
All SA	0.17, 0.96, 0.23, 1.11, 6.46, 2.01	0.24 mm, 22.9%	-0.02 mm, -1.8%	0.03, 1.04, 0.57	0.25 mm, 24.4%	-0.02 mm, -2.0%	1501	12 m/s
NOR SA	--	0.47 mm, 37.8	-0.10 mm, -7.8%	--	0.45 mm, 35.8%	-0.12 mm, -8.5%	352	12 m/s
US SA	--	0.14 mm, 14.7%	0.00 mm, 0.5%	--	0.15 mm, 15.5%	0.00 mm, 1.00%	1156	12 m/s
US DA	0.00, 0.92, -1.19, 1.89 9.75, 1.36	0.13 mm, 13.9%,	0.00 mm, 0.0%	0.021, 0.74, 0.66	0.14 mm, 13.8%,	0.00 mm, 0.1%	1390	8 m/s
US BDA	0.31, 1.00, 1.79, 0.58, 10.0, 3.15	0.12 mm, 13.6%	-0.01 mm, -1.7%	0.01, 0.48, 0.51	0.13 mm, 14.6%	0.00 mm, -0.9%	1204	8 m/s
US SDFIR	0.99, 0.96, 0.52, 0.10, 0.14, 10.75	0.13 mm, 14.1%	-0.01 mm, -0.09%	0.004, 0.00, 0.00	0.13 mm, 14.2%	-0.01 mm, -1.6%	1508	8 m/s

Table 3. Transfer function results for 10 m height wind speeds. Coefficients for the sigmoid transfer function (Sig, Eq. 3) and the exponential transfer function (Exp, Eq. 4) and the associated RMSEs and Biases as well as the number of periods available (*n*) and the maximum wind speed (Max U) are described for the US unshielded (UN), single AlterAltar (SA), double AlterAltar (DA), Belfort double Alter (BDA), small DIFR (SDFIR), the NOR single AlterAltar (SA) gauge, and the combined US and NOR SA results (All SA). The same coefficients determined by the All SA transfer function were used to separately correct the results from NOR and US.



~~Figure 1: Example event from the US site with accumulated precipitation measured using the Double Fence Intercomparison Reference (DFIR, dark blue), small DFIR (SDFIR, red), Belfort double Alter (BDA, yellow), standard double Alter (DA, purple), single Alter (SA, green), and unshielded (UN, light blue) weighing precipitation gauges. The 24 hr mean T_{air} was -6.6 °C, and the mean gauge height wind speed was 3.6 m s^{-1}.~~

Figure 1: Site map (a); Haukeliseter DFIR, single Alter, (b) and unshielded gauges (c); and Marshall Belfort double Alter (d) and double Alter shielded gauges (e).

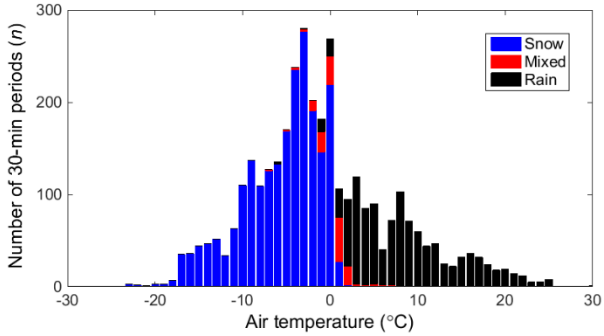


Figure 2: The temperature distribution of 30-min periods classified as snow, mixed, and rain from the US site.

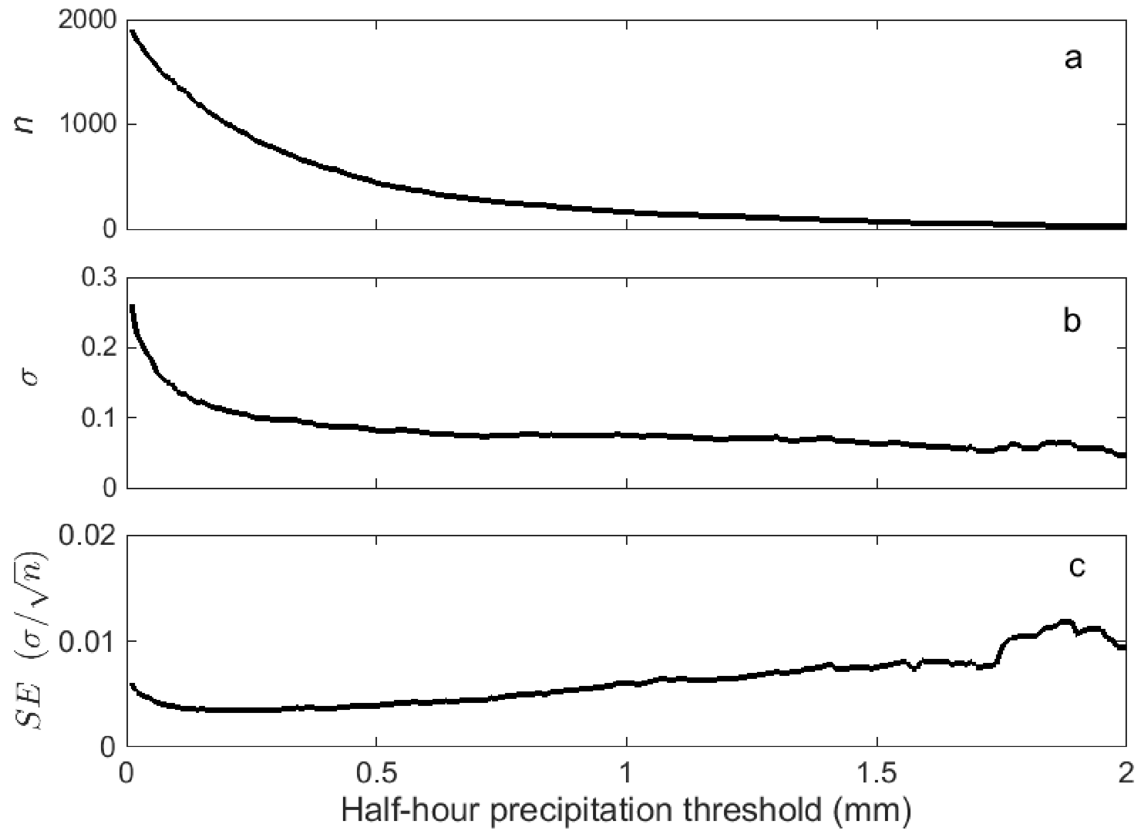


Figure 3: The effects of minimum threshold on transfer function development for 30-min periods from the US site. The number of 30-min periods (n) above the threshold (a), the standard deviation (σ) of the linear transfer function error (b), and the standard error ($SE = \sigma/\sqrt{n}$) of the transfer function (c) are shown. Only snow results are included here, with snow in this case identified by the present weather detector with more than 15 min of snow and less than 5 min of other precipitation types within each 30-min period.

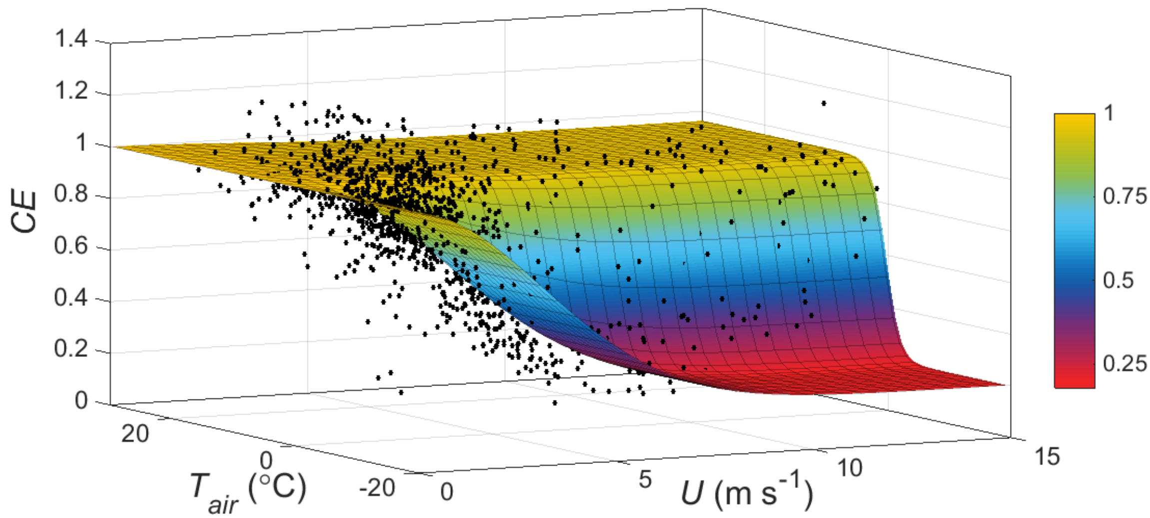


Figure 4: Example transfer function using the sigmoid function to describe the combined SA catch efficiency (*CE*) measurements from both the US and NOR sites as a function of air temperature (T_{air}) and gauge-height wind speed (*U*). Individual 30-min *CE* measurements are shown (black circles) along with the sigmoid function fit to them (coloured surface), with the colour of the surface indicating *CE* magnitude. 352 of the measurements are from the NOR site, recorded during winter periods of 2011, 2012, and 2013, and 1156 measurements are from the US site, recorded from Jan, 2009 through March, 2014.

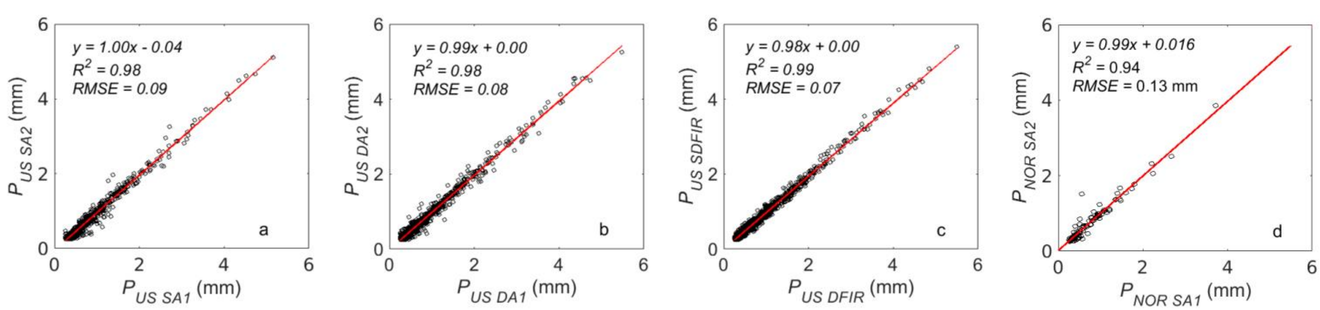


Figure 5: Comparison of 30-min precipitation (*P*) measured from ~~three~~ four pairs of replicate or near-replicate gauge-shield combinations. ~~T~~From the US site, two single Alter (SA) gauges (a), two double Alter (DA) gauges (b), and ~~a~~ double fence intercomparison reference (DFIR) and ~~a~~ small DFIR (SDFIR) gauges (c) are compared. From the NOR site, two single Alter (SA) gauges are compared (d). Only snow data are included here~~, with snow in this case identified by the Present Weather Detector~~.

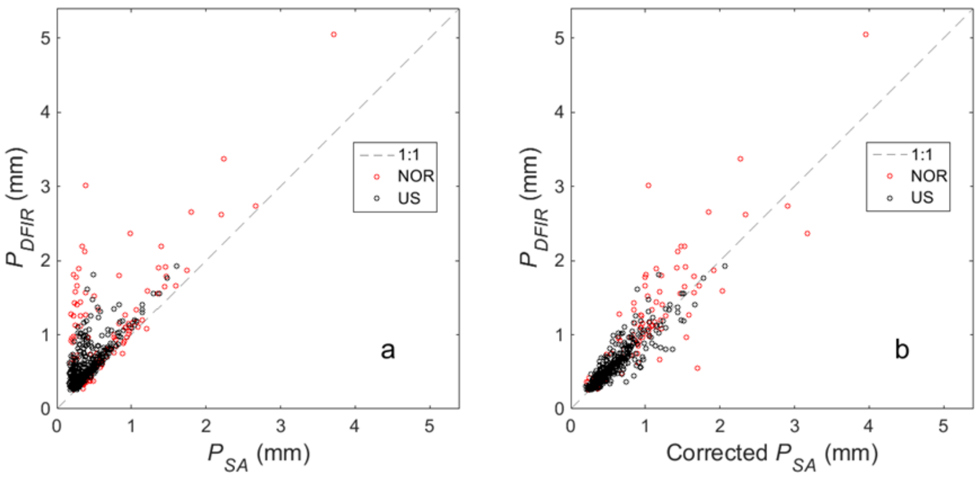


Figure 6: – Uncorrected (a) and corrected (b) SA precipitation (P_{SA}) vs. DFIR precipitation (P_{DFIR}) for snow only, where snow is here defined as T_{air} < -2.5 °C.

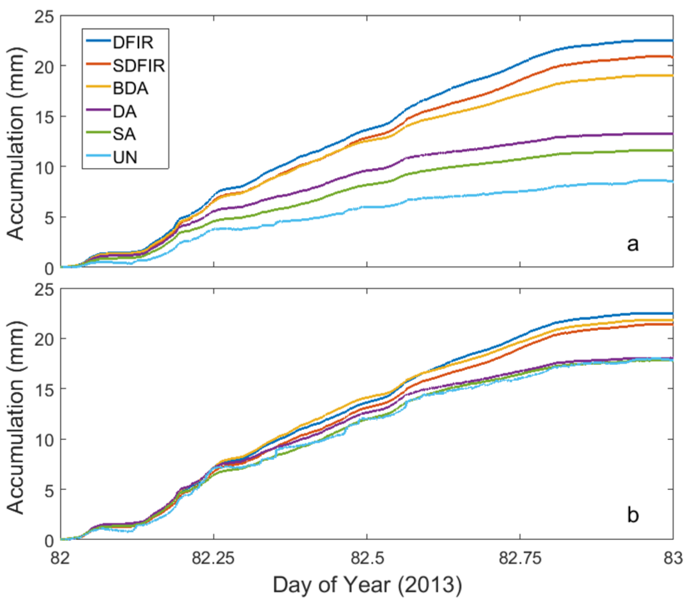


Figure 7: ~~Figure 1:~~ Example event from the US site with accumulated precipitation measured using the Double Fence Intercomparison Reference (DFIR, dark blue), small DFIR (SDFIR, red), Belfort double Alter (BDA, yellow), standard double Alter (DA, purple), single Alter (SA, green), and unshielded (UN, light blue) weighing precipitation gauges. Both the uncorrected (a) and the corrected (b) precipitation accumulations are shown, with the corrected results estimated by applying the appropriate Exp transfer function to the 1 min accumulations. The 24 hr mean T_{air} was -6.6 °C, and the mean gauge height wind speed was 3.6 m s^{-1}.

~~Example of improvement to the one day event from the US site shown in Figure 1 after applying the appropriate Exp corrections to the 1 min accumulations.~~