

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

Response to General Comments from Eva Mekis

5 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.

5 **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)
- 10 - 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
- 15 (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
- 20 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.

25 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:
- 30 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

- 35 **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.

- 5 **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.

- 10 - 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

- 15 - Line 25: Use the same notations $U_{4.5}$ vs U_{gh}

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

- 20 **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.

25

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.

30

- 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.

35

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

- 5 - 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

- 10 - 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.

15 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.

20

- 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.

- 25 **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

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

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

- 35 **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.

5 - 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).

10 **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).

15 **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.

30

- 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.

35 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.

5 - 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:

10 - 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.

20 **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.

25

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 manuscript or by changing the order of the figures.

30

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.

35

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 through 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 precipitation 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.

30

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.

35

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".

5

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,
10 '(WMO-SPICE)'.

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

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

15 **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.

20 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.

25 **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.

30

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.

35 **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.

5 **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.

10

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.

15

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.

20 **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.

25 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).

30 **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.

35

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".

5

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.

10 **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
15 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
20 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.

25

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
30 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.

35

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”.

5

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’.

- 10 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.’

- 15 **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.

20

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.)

- 25 **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!

30

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.

- 35 **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.

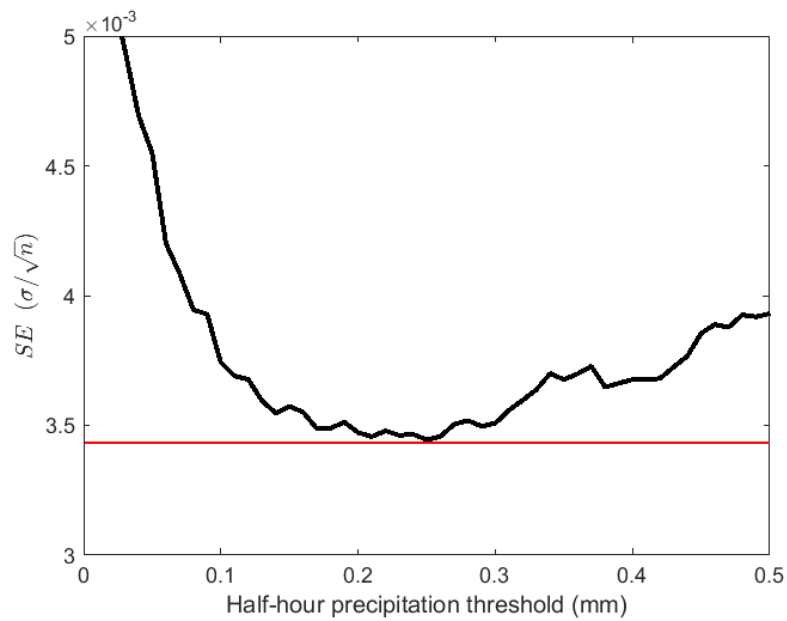


Figure 1R: The same plot shown in Figure 3 panel c in the manuscript under discussion, with the ranges of the x- and y-axes altered. The standard error ($SE = \sigma/\sqrt{n}$) of the transfer function is shown (black line) in addition to a horizontal line (red) placed to help locate the minimum in the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

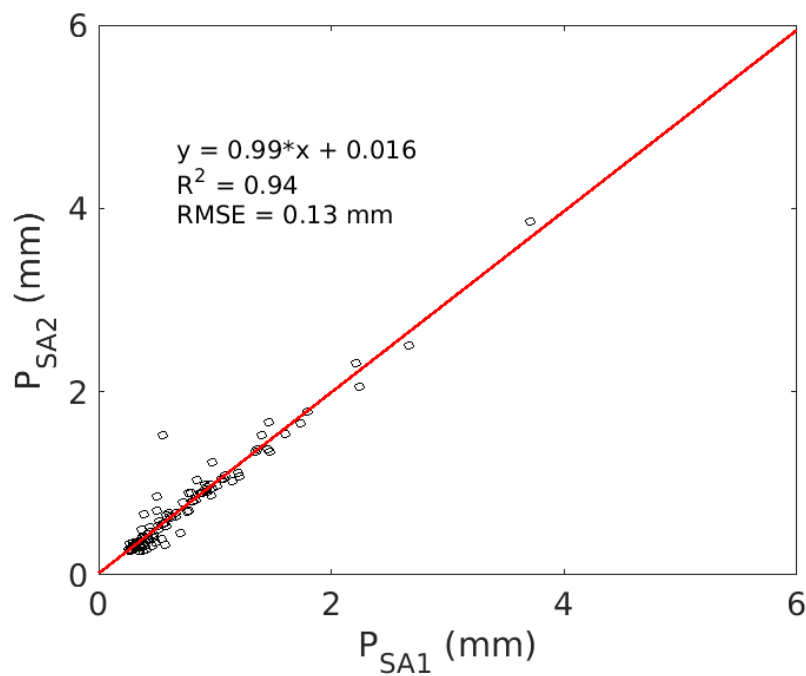


Figure 2R: The same plot shown in Figure 3 panel c in the manuscript under discussion, with the ranges of the x- and y-axes altered. The standard error ($SE = \sigma/\sqrt{n}$) of the transfer function is shown (black line) in addition to a horizontal line (red) placed to help locate the minimum in the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