

Author's Response to Anonymous Referee #2

We would like to thank Referee #2 for their thorough review of this manuscript. We appreciate the detailed comments and suggestions. Our responses to the comments and actions taken are listed below in red.

Major comments:

RC: In the abstract and section 1.1, the authors discuss using liquid, mixed phase and solid precipitation. A possible useful addition to Table 1 would be statistics on the percentage of data that was liquid or mixed phase precipitation and the percentage of the data that was solid precipitation for each site. This could potentially help explain some of the results you see in the other tables when comparing all precipitation phases to snow only.

AC: We agree that this could be useful information. We calculated the percentage of precipitation events for all phases (solid/mixed/rain) for the two seasons combined that were used in this analysis using the temperature thresholds noted in this paper and in Kochendorfer et al. (2017b). They are as follows:

Site	% Precip Events by Phase solid/mixed/rain (%)	% Total Precip by Phase solid/mixed/rain (%)	Approximate Change in RTC from All to Snow
XBK	59/16/25	43/16/40	-10%
CAR	39/32/29	28/33/39	+3%
CCR	91/8/1	85/15/1	0%
FOR	31/43/26	28/49/23	-20%
HKL	47/38/15	44/35/21	-12%
MAR	44/33/23	32/45/23	+12%
SOD	54/39/8	49/42/8	+2%
WFJ	79/19/2	78/21/1	+2%

There were really no surprises in these numbers. We believe what the referee was alluding to in their comment was that we should expect that the biggest changes in Relative Total Catch (from all precip to solid precipitation) should occur at sites with the smaller percentage of solid precip. This would support the hypothesis that the higher total catch of rain and mixed events is masking the errors associated with adjusting only solid precipitation. The largest drop in RTC does occur at FMG (~-20%) which also has the lowest percentage of solid precipitation events during the season (31%). However, the next highest drops in RTC occur at HKL (~-12%) and XBK (~-10%) and both of these sites exhibit a higher percentage of solid precipitation events during the study period (47% and 59% respectively). Alternatively, we could look at the precipitation totals for each type (second column in table above). Although the percentage value for total snow is generally lower than for event occurrence, it doesn't change the interpretation.

The relative amount of snow as compared to rain/mixed could possibly contribute to the difference in the statistics for FOR, but doesn't explain the differences at the other sites. We believe that the addition of a statement in the discussion referring to this will not improve the clarity of interpretation.

Action: Added columns in Table 1 to show the relative total precipitation by phase for each site. Added a reference to this information in Section 2.2.

RC: The authors show analysis using the wind data collected from 10m heights and gauge height and discuss many issues related to measuring the wind speeds. In the discussion starting on page 13, line 20, as well as the first part of the conclusions section, no mention is made of the possible impacts of wind variations over the 30-minute averaging period. For the windier sites used in the study, how variable were the winds over the 30 minutes and what role might that have played in some of the results? I agree with the discussions and results the authors have expressed regarding some of the other wind-related issues but this should also be addressed.

AC: This very issue was addressed in previous work during the WMO-SPICE project and was published both in Wolff et al. (2015) and Nitu et al. (2018). Wolff et al. looked at variations in both 30-minute temperature and wind speed at HKL (one of the windier SPICE sites) and concluded that removing points with large variability within the 30-minute period did not significantly change the results.

Action: The following sentences were added to the end of the discussion paragraph that formerly started on page 13, line 20: "Additional uncertainty related to wind speed may be attributed to the variability within the 30-min mean period. Although this wasn't included in the current analysis, previous work by Wolff et al. (2015) and Nitu et al. (2018) at HKL showed that the impact of high frequency variability in the wind speed over 30-min periods on transfer functions was negligible."

RC: Regarding the GEONOR gauges used in the study, the authors state they were GEONOR T-200B3 gauges. I believe this implies a three-wire GEONOR gauge. If that is the case, how were the data from the three wires used? Were they averaged? This should be discussed and/or clarified in the introduction or methods sections.

AC: The three Geonor wires were averaged before filtering.

Action: The methodology was revised to reflect this

RC: With regards to Fig 1, equation 2 shows a collection efficiency of 1.1 at a wind speed of 0, but I don't see a clear explanation for this. There is some discussion on P4, L1 that this might be indicative of rain at air temps > 2C, but equation 2 does not use temperature so I'm a bit confused as to why temperature is mentioned there and why eqn 2 would result in a catch efficiency > 1 at 0 m/s. Some clarification and additional explanation would be useful here.

AC: This is a good question and should have been addressed in the paper. Simply, the Equation 2 catch efficiency was greater than 1.0 because of the specific empirical fit to the WMO-SPICE catch efficiency data that produced the smallest error. It is due to the shape of the data that the equation was fit to, rather than a physical cause. This function was not originally published in this present manuscript, but

rather in K2017B. For this analysis, any catch efficiency that was calculated as > 1 was automatically reset to 1. This occurred relatively infrequently, since the mean wind speed during precipitation has to be less than 0.5 m s^{-1} and therefore did not significantly affect the results.

The discussion on P4, L1 does not explain why the catch efficiency for Eq. 2 is greater than 1.0. This text simply explains how liquid precipitation, classified as precipitation occurring when $T_{\text{air}} > 2 \text{ deg C}$, should be handled according to K2017b.

Action: The following has been added to the manuscript in the methodology section (2.2): “Figure 2 suggests that Eq. 2 can exceed a catch efficiency of 1 at low wind speeds. There is no obvious physical explanation for the portion of the Eq. 2 catch efficiency function (originally published by K2017b) that is > 1.0 and this is related to the empirical fit of the catch efficiency curve to the original SPICE data. For this current assessment, calculated catch efficiencies > 1 were infrequent and occurrences were automatically set to 1.”

RC: In section 3.2 and 3.3, there is mention of potential shadowing of the wind sensor at the MAR site. Was anyone at the site contacted and asked for pictures of the sensor at the site to confirm this (assuming none of the authors have visited or work at this site)? This would seem to be an easy thing to do without supposition. Could this also be a calibration issue with the wind sensor at that site?

AC: Co-author J. Kochendorfer is quite familiar with the MAR site and has done extensive work during the SPICE project on the wind measurements. This was outlined in Kochendorfer et al. 2017a. During the development of the Kochendorfer et al. (2017a) transfer functions, the ratio of the 30-min gauge-height wind speed to the 30-min 10 m height wind speed was plotted as a function of wind direction, and it was quite clear that the gauge-height anemometer was affected by obstacles in some wind sectors.

Action: P8, L31 has been changed from, “wind speed measurements may be shadowed...”, to, “wind speed measurements were shadowed...”. This statement was not based on supposition.

Minor comments:

RC: Throughout the manuscript: There is inconsistent use of the hyphen in the phrase “single Alter shielded”. In some places, there are two hyphens, in others, just one. I believe it should be “single Alter-shielded”.

Action: The wording in the text should now consistently be “single Alter-shielded”

RC: P1, L17 – Intercomparison is hyphenated here but nowhere else. The official SPICE title appears to not have the hyphen.

AC: Inter-Comparison was supposed to have been hyphenated so that the acronym SPICE made more sense, but it appears to have been dropped in the literature along the way. This should be consistent

Action: updated the text without the hyphen in Intercomparison.

RC: P1, L20 – The term “windshield” is typically used to describe the front glass on a car. In the case of precipitation measurement, I believe “wind shield” is more commonly used.

AC: agreed

Action: changed the text so that “wind shield” is consistently used.

RC: P1, L25-27, The sentence on these lines is somewhat awkwardly written.

Action: Changed to “Performance is assessed in terms of relative total catch (RTC), root mean square error (RMSE), Pearson correlation (r), and Nash-Sutcliffe Efficiency (NSE). Metrics are reported for combined precipitation types, and for snow only.”

RC: P1, L29 – What is meant by gauge configuration?

AC: This refers to the gauge model and the wind shield type (if any) that the gauge is paired with.

Action: clarified this in the text by adding “...and gauge configuration (*gauge and wind shield type*).”

RC: P5, L24 – This sentence reads a bit awkwardly and may just be missing a comma. Perhaps revise into two sentences.

Action: These sentences now read: “The algorithm removes random and systematic diurnal noise, but does not account for signal drift (an example of signal drift is a decrease in weight that occurs due to evaporation of water from the gauge bucket).”

RC: P6, paragraph 2 – How often were the temp/wind speed data missing?

AC: We can qualitatively say that the number of missing temperature and wind speed data were low, and generally the missing temperature and wind speed data coincided with missing precipitation data. Since missing data is not included in the analysis, this is largely inconsequential.

Action: none taken

RC: P6, L34 – there is an extra comma on this line

Action: fixed

RC: P6, L35 – there shouldn’t be a comma after “snow”.

Action: fixed

RC: P7, L5 – The term “both” is not necessary in this sentence.

Action: fixed

RC: P8, L6 – The term “both” is not necessary here either.

Action: fixed

RC: P11, L18 – season should be plural.

Action: fixed

RC: P11, L20 – I think you meant to say compounding instead of confounding.

Action: changed this to “complex”.

RC: P12, L16 – There is a) missing on this line.

Action: fixed

RC: Figures 5 through 8 – One suggestion the authors might want to consider is to change the colors of the bars and use warm colors (e.g. red, orange, yellow, or shades of red) for the single Alter data and cool colors (blue, green, purple or shades of blue) for the unshielded data. This would really make the differences between the two data sets much more obvious and quickly draw the reader’s attention to the points you make in the manuscript

AC: Agreed.

Action: We changed the colour of the bars so that the shielded gauges are shown in grey tones while the unshielded gauges are shown in blue tones. We also made some changes to help with discerning the bars if printed in grey scales.