General comments from Referee #1:

The manuscript "Testing and Development of transfer functions for weighing precipitation gauges in WMO-SPICE" deals with the interesting and relevant topic of how to correct undercatch especially of snowfall for different sensors. This paper is basically an extension of the Kochendorfer et al. 2017 (HESS) that again is based on the (Kochendorfer et al. 2016) study. Here the apply the previously developed correction functions to different sensors and locations and found that the previously developed, more general correction function of Kochendorfer et al. 2017 (HESS) is always recommended to use. In general the paper is mostly very written and I found no crucial error or mistake. My major concern is more related to presentation and the structure of the manuscript. Both section "Methods" and "Results (& Discussion)" are written very detailed, technically and lengthy, giving each sensor, its location, and treatment of the data its place. Figures 4 - 11 display in detail the (mostly non-existing) differences between correction functions. Partly, I found it hard to follow the overall structure/story of the manuscript. In my opinion, this structure would have been justified if the authors would have found several different best-performing equations that need to be presented alongside each sensor and in comparison with the general correction function. But, given that in most cases the general (somehow site and sensor unspecific) correction function is as well performing and recommended by the authors, a more summarizing structure would have been much more feasible and would have cause less redundancies. In its current version, the manuscript unfortunately reads quite lengthy, especially given the final outcome. Many of the interpretations named in the abstract and conclusion are a bit out of blue and should be much more elaborated. Hence, instead of the detailed description, I would recommend to show some summarizing, and comparing analyses/graphics on sensor performance.

Based on this basic concern, and the following further smaller concerns, I recommend to provide either a revised, restructured version that strongly condense the specific sensor part and put more focus on the comparison analysis, or to show the transferability of the Kochendorfer et al. 2017a Equation 3 to other sensors in a HESS-technical note. I am sorry to be that harsh, but I really do think that the level of detail and length of the manuscript does not match your findings.

Authors' response to general comments:

We think this is a fair assessment of the manuscript in some respects. We chose to separate the different types of WMO-SPICE weighing gauge measurements into different manuscripts before discovering that the 'universal' transfer functions (Kochendorfer et al., 2017a) performed well on all of the manufacturer-provided single-Alter shielded and unshielded gauges. The success of the 'universal' transfer functions on these other weighing gauges was a surprise. We had originally assumed that all of the different types of WMO-SPICE weighing gauges would require their own unique transfer function, and it was based in part on this assumption that we decided to separate the WMO-SPICE weighing gauge transfer functions into two separate manuscripts. In addition, the length and amount of detail included in the Kochendorfer et al. (2017a) manuscript would have been unwieldy if all of the weighing gauges from WMO-SPICE were added to it.

We agree that the mode of presenting the results can be simplified and improved. For example, many of the figure styles were developed for Kochendorfer et al. (2017a), and were chosen to comprehensively compare different types of transfer functions using different wind speeds. In order to achieve a more condensed view, we will change the figure style by merging gauges into the same plot and excluding some of the extraneous transfer functions. As the reviewer pointed out, the majority of the transfer functions evaluated in this manuscript were determined to be obsolete because they were no better than an appropriate 'universal' transfer function or a pre-SPICE transfer function from Kochendorfer et al. (2017b). In addition, for a given gauge the differences between any of the appropriate transfer functions evaluated were quite small. Because of this, they do not all merit such a thorough evaluation, the format

of the results can indeed be consolidated significantly in response to the reviewer's suggestion. This will help to improve the presentation of the manuscript and to focus it on the more significant results.

We maintain, however, that the results are significant enough to merit publication as a new manuscript. The manuscript provides independent validation of previously-existing transfer functions, and it also demonstrates that transfer functions derived using one type of weighing precipitation gauge can be used on another type of similar gauge. Many readers (including gauge and site developers, precipitation observers and data users) will find this significant, surprising, and also very useful. The manuscript, based on a well-executed and carefully-designed experiment, includes the evaluation of nine different types of gauges and wind shields.

Reviewer #1 should also note that in addition to the different single-Alter and unshielded gauges that were evaluated, gauges in several other types of windshields were also assessed. Some measurements recorded within these other winds shields were used to validate the results of Kochendorfer et al. (2017b), and others were shown to merit new adjustments. For example, a new set of adjustments for the shielded MRW500 precipitation gauge were derived and recommended. Arguably, such an evaluation on its own could merit publication as a full manuscript, as this is the only transfer function available for this gauge and shield so far, but due to the wealth of new measurements produced by the WMO-SPICE project, it was included with the evaluation of transfer functions for all of the other WMO-SPICE weighing precipitation gauges and wind shields.

The team of WMO-SPICE investigators who authored this manuscript has an obligation to precipitation gauge manufacturers and their customers to recommend transfer functions for each weighing gauge and wind shield tested in the WMO-SPICE project. In addition to the forthcoming WMO report, a journal article like this is the best way to disseminate the results of an international project like WMO-SPICE.

Further general concerns from Referee #1:

1. Several time differences are named significant or non-significant, but I somehow missed the section describing how a significance test was performed.

Authors' response: In response to this valid comment, T-tests have been performed to more objectively evaluate the significance of differences between the errors associated with different transfer functions and also the unadjusted measurements. The results of these T-tests will be documented in the revised manuscript, and will be used to replace the subjective determinations of significance included in the original manuscript.

2. Given that both the biases of the corrected values and the differences between the corrected functions are mostly rather small (Figure 4- 11, RMSE < 0.5 mm, biases « 0.5 mm, differences far less), I wonder how and if at one can and should interpret these differences, given the measurement accuracy of each sensor.

Authors' response: For the most part, the biases in the corrected measurements and the differences between the different corrections are indeed negligible. The restructured manuscript will state this more explicitly.

3. The paper needs to be more independent from Kochendorfer et al. 2017 (HESS). At least the essential equation 3 should also be given in this paper.

Authors' response: This is a good suggestion. The often-cited Eq. 3 will be included independently in the paper, and other changes will be made to help the manuscript stand on its own.

Further specific comments:

page 2, line 19-22: I do not understand this sentence. Please check language and try to avoid too long sentences.

Authors' response: The run-on sentence will be divided up for clarification.

page 3ff: I would recomment to use Table 1 much more often to guide the reader through the different configurations.

Authors' response: There are indeed many different gauges and configurations for the reader to familiarize himself or herself with, and the suggestion is a good one. We will try to refer to Table 1 throughout this section.

page 3, A paragraph that outlines the design of this study would help the reader to follow your following descriptions

Authors' response: We will add an introductory paragraph describing how the intercomparison was designed, with a common automated reference gauge at all sites, along with other shared types of precipitation and meteorological measurements. The overarching goal of the intercomparison was to include as many different countries, climates, and gauges as possible while still maintaining some basic standards.

page 3, line 9: typo: "a either a"

Authors' response: Thank you! This will be corrected.

page 3, line 31: CARE, acronym not yet introduced

Authors' response: CARE is "Centre for Atmospheric Research Experiments", and will be defined in the manuscript.

page 4, line 4, "have" missing in and a lower porosity

Authors' response: This will be rewritten.

page 4, line 31: 6 s and 1 min data correspond to with sensor / site. Maybe update Table 1

Authors' response: We will describe which data are available with which sensors.

page 4, line 32: "realistic" and "operational" limits, please define

Authors' response: We will clarify this. The term "realistic" was meant to signify "possible", and was used to remove periods of calibrations/validations and other impossibly high-rate precipitation and negative measurements from the record. The term, "operational" was meant to signify "within the operational specifications of the sensor", such as a weighing gauge depth that was beyond the upper limit of the gauge specifications.

page 5, Maybe I missed it, but where do you define that DFAR is the reference, and please discuss briefly its quality and its deviation from the "truth"

Authors' response: The DFAR was described and defined as the reference at the beginning of the Methods section. The "truth" is indeed difficult to determine for snowfall because pit gauge measurements, which are used as a reference for rain, are subject to blowing snow and capping, and are therefore not appropriate for snowfall. The DFIR is the reference for manual snowfall observations (Goodison et al., 1998.; Yang, 2014). In the beginning of the Methods section, we will add available references to manual DFIR and "bush" gauges, in addition to later comparisons with automated bush gauges. We will also discuss in greater detail how and why the DFAR was defined as the reference for WMO-SPICE.

page 5, line 13ff: How were temperature thresholds defined. If possible refer to citations

Authors' response: The temperature thresholds were defined based on histograms of automated precipitation type and air temperature measurements (Kochendorfer et al., 2017b; Wolff et al., 2015). An explanation and citations will be added to the manuscript.

page 5, line 18: Please provide these equation also here to be more independent from the Kochendorfer et al. (2007a) paper

Authors' response: This is a good suggestion.

page 7, line 13ff: I do not understand this interpretation. If significant site biases exists, doesn't that mean that you have to develop site-specific empiral correction functions?

Authors' response: This is indeed true, but we do not have a good way to determine such site-specific empirical corrections. These adjustments are designed for use at sites that do not have a good reference, and at such sites there is no good way to determine what the site bias is. For example, in Kochendorfer et al. (2017a) the three high-altitude sites had the largest biases, but at two of these high-altitude sites the biases were negative, and at the other high-altitude site the bias was positive.

page 10, line 31: "1500mm" Geonor ? cp. with page 11, line 1: please be consistent,
see also page 11, line 17

Authors' response: That is correct. We will correct these inconsistencies.

page 11, line 6: typo: "attributed"

Authors' response: Thank you. The spelling error will be corrected.

page 11, line 20: does the noise of the Geonor 1500 mm stem from the fact that they
have a different sensitivity than the 500 mm ?

Authors' response: Yes that is probably the case. We will suggest this in the manuscript.

page 11, line 20 ff: Are these interpretation valid, given the some deviation and the
measurement accuracy?

Authors' response: We may disagree on the definition of 'noise', but it is certainly the case that it is more difficult to measure light snowfall with a 1500 mm Geonor than a 600 mm Geonor. In general, it can be difficult to differentiate signal noise from precipitation with these weighing gauges, and this problem is augmented with the 1500 mm Geonor. This is an important issue for the measurement of

snowfall (and light rain over shorter time periods, such as 5 min) because most snowfall is associated with low precipitation rates, particularly in polar regions.

page 11, line 26, CARE and Marshall *test sites* !

Authors' response: Thank you. This will be corrected.

page 17, Table 1: please add time period of measurements considered

Authors' response: The time period will be documented in the Table 1 caption rather than the actual table. All of the measurements spanned the same time period, so it isn't necessary to describe the time period in the table for each individual site.

References

Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M. E., Reverdin, A., Wong, K., Smith, C. D., Yang, D., Roulet, Y. A., Buisan, S., Laine, T., Lee, G., Aceituno, J. L. C., Alastrué, J., Isaksen, K., Meyers, T., Brækkan, R., Landolt, S., Jachcik, A., and Poikonen, A.: Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE, Hydrol. Earth Syst. Sci., 21, 3525-3542, 2017a.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R., and Leeper, R.: The quantification and correction of wind-induced precipitation measurement errors, Hydrol. Earth Syst. Sci., 21, 1973-1989, 2017b.

WMO Solid Precipitation Measurement Intercomparison (1998, WMO/TD-No. 872, IOM No. 67, by B.E. Goodison, P.Y.T. Louie and D. Yang)

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.

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