The paper, 'The Quantification and Correction of Wind-Induced Precipitation Errors,' by Kochendorfer et al., addresses the important hydrological topic of solid precipitation underestimation by weighing gauges used in operational networks. Errors and biases are determined for several different configurations of wind shields (relative to a reference configuration), and for gauges installed at two field sites in different climate regimes. Equations with different functional forms ('transfer functions') are tested to characterize the influences of wind speed and temperature on gauge catch efficiency, with emphasis on the uncertainty of equation outputs and the height of wind speed observations used in the equations.

This paper presents a quantitative and straightforward approach to address a complex issue, and endeavours to simplify the functional forms used in transfer functions. 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.

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.

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.

Smaller things (bold text denotes recommended additions/substitutions)

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

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

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

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

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

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.

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

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.

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

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

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.

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.

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

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

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

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

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

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

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).

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.)

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.

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

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.

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.

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

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.

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?

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

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.

P12, L24: change 'efficacy' of shield to 'porosity.'

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.)

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

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.

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.

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.