Review of, “Correcting precipitation measurements of TRwS204 in the Qilian Mountains, China”.

General comments:

The manuscript describes adjustments for single-Alter shielded automated weighing gauge measurements derived from a two-year-long precipitation gauge intercomparison. Manual measurements within a DFIR shield were compared to single-Alter shielded weighing gauge measurements to estimate errors due to wind, and manual measurements within a single-Alter shield were compared to the single-Alter shielded weighing gauge to estimate ‘specific’ errors experienced by the automated gauge.

Specific comments:

1) Wind speeds were relatively low at the testbed during the measurement period (a maximum of 2 m s^{-1}, from Fig. 7), and there were only a few solid precipitation events (I count only 11 in Fig. 7c). The authors need either a longer measurement period like Chen et al. (2015), or a more thorough justification of the general applicability and usefulness of the proposed correction.

Because the purpose of the adjustments derived in the manuscript is presumably to correct gauge measurements at other sites, the adjustments must be representative of the wide range of meteorological conditions that such monitoring sites may be exposed to. Such adjustments are most significant for solid precipitation occurring in high wind speeds. For example, neglecting differences in the wind speed measurement height, Chen et al. (2015) noted at this same site, “the wind speed showed no significant effect… below 3.5 m s^{-1}”. In the manuscript under review, Fig. 2 shows that the $TRwS_{SA}$ measurements did not typically underestimate the “true” amount of precipitation, and were in fact fairly comparable to the DFIR-shielded manual gauge used as a reference. This is what one would expect given the range of conditions that the site was subject to. The problem with this is that corrections derived from such measurements will not be applicable to windy monitoring sites that are subject to solid precipitation events, where such corrections are actually most significant and necessary. The measurements presented in the manuscript could be used to test some of the transfer functions available in the WMO-SPICE Special Issue and elsewhere, but they do not comprehend a wide enough range of meteorological conditions for the derivation of valid and useful transfer functions.

2) The separation of specific and systematic/aerodynamic errors merits further examination. Based on the use of equations 1-3, two assumptions are made: 1) The single-Alter shielded CSPG is itself free from specific errors and 2) The aerodynamic error for the single-Alter shielded TRwS and CSPG are identical. While the second assumption may (or may not) be valid, the first assumption is also problematic. Comparisons of identical precipitation gauges with identical shielding show that differences between like measurements in such a field site are significant. All precipitation gauge comparisons are subject to errors due to causes such as general measurement uncertainty and the spatial variability of precipitation. Such errors are not aerodynamic, but the methodology presented in the manuscript defines all differences between the two CSPG gauges as aerodynamic; there is no specific error term in Eq. 3, which indicates that the CSPG is completely free from

specific errors. A more defensible and direct way to estimate such specific errors would be to install a DFIR-shielded TRwS and compare it to the DFIR-shielded CSPG, or to simply use low wind speed or rainfall measurements, where the effects of shielding and wind are negligible.

In general I don't understand the advantages of estimating specific and systematic errors with the indirect approach described in the manuscript. Both correctable and uncorrectable errors could be examined directly by creating a transfer function, and then quantifying the remaining uncertainty in either the transfer function or the corrected measurements.

Technical corrections:

Pg. 1 ln. 17. It would be revealing to discuss the root mean square or the mean of the absolute values of specific errors in addition to the average.

Pg. 3 ln. 20-25. Please describe these measurements in more detail. Exactly how were the 12 hr TRwS measurements estimated from the 30-min measurements? Were there different outputs available for this gauge? For example, was the change in the absolute depth calculated or were the average 30-min intensities used? Was any smoothing or averaging performed? What type of anti-freeze and oil were used, and how was the heater configured? How were the CSPG measurements taken, by weight or by measuring stick for example? Was solid precipitation melted before being measured manually? How were the meteorological measurements recorded (heights, sensors, etc.)?

Pg. 4 ln. 10. Change 'corrected' to 'subject to'.

Pg. 4 ln. 22. "The catch ratio which served as the function of environmental…" needs to be rewritten.

Pg. 4 ln. 30-31. Manual methods are also subject to specific errors.

Pg. 5 ln. 5-7. I don't know much about the TRwS, but most weighing gauges output the total depth, allowing for small changes in the total depth to be calculated over longer time periods. This issue may be due to shortcomings in the way that the gauge measurements were logged and processed, although the stated 0.001 mm resolution seems quite good. Does this resolution translate to 0.001 mm/30-min?

Pg. 5 ln. 24-25. Please clarify, "Even though the losses were less for snowfall events, their measurement ratio was minimal throughout the experiment".

Pg. 5 ln. 29. How were the corrected $CSPG_{DFIR}$ measurements "proven" to be true in rainfall? Was it using the pit gauges?

Pg. 6 ln. 2. Clarify by rewriting, "with a mean difference within this range of 0.5 mm".

Pg. 6 ln. 20-22. Clarify how these sums were estimated. Were actual totals of the specific errors used, rather than totals of the absolute values of the specific errors? If the mean specific errors were actually significant, this indicates that there is a systematic bias between the two single-Alter gauges, and that it might be more effective to derive the $TRwS_{SA}$ adjustment by comparing it directly to the $CSPG_{DFIR}$.

Pg. 6 ln. 31. Clarify how the $TRwS_{SA}$ and $CSPG_{SA}$ CR were the same. Were the CR based on systematic differences, rather than measurements?

Pg. 8 ln 8. The range is not a good estimate of average errors, especially considering that there were many more measurements available at low wind speeds. Examine this statement more carefully, and try to support it by quantifying the errors (or scatter) at different wind speeds.

Pg. 9 ln 12. Light precipitation over longer intervals may be easier to quantify using the gauge depth, rather than the precipitation intensity.