The authors have changed the content and structure of the manuscript significantly to address points raised by the referees, and they are commended for their efforts. An assessment of an automatic weighing gauge for precipitation measurements is presented relative to a manual reference configuration, and the comparison results (namely, the catch ratio for 12-h periods with precipitation) are used to test an adjustment function derived using the WMO-SPICE dataset. Errors specific to the weighing gauge under test are also investigated using a manual gauge in the same shield configuration. The results and analysis presented employ a dataset from a region not otherwise represented in WMO-SPICE, and are of particular interest to meteorological operations in China.

The manuscript flows logically from section to section, and is more conservative with respect to the attribution and quantification of errors than the previous submission. Questions related to the content and presentation of results are provided in the Comments section below, along with recommendations for changes/additions to the manuscript. A non-exhaustive list of proposed technical revisions is also provided. The intention of this Editor Comment is to complement the forthcoming Referee Comments, providing an additional perspective for the authors to consider as they prepare the next version of the manuscript.

Comments

Title

1. Propose changing title to ‘Correcting precipitation measurements from MPS TRwS204 automatic weighing gauges in the Qilian Mountains, China’

Abstract

1. The abstract should be brief, but still needs to introduce necessary background material. The gauges included in the study must be introduced (e.g. MPS TRwS204 is an automatic weighing gauge, CSPG is the manual Chinese Standard Precipitation Gauge) and the ‘existing adjustment function’ should be elaborated upon. Further, it is stated that ‘deriving adjustment algorithms has become a top priority,’ but there is no mention of what these algorithms are adjusting for (e.g. wind-induced undercatch of precipitation). The abstract should be revised to include the above points.

2. When considering the results after adjustment, the following statements are made: ‘It seems that the adjustment function is more appropriate to correct the snowfall measurements than rainfall and sleet measurements for this dataset.’ This makes sense given the results presented, but only considers the average loss relative to the reference. What about the Root Mean Square Error? What about the Bias? It is stated that ‘Overall, the results of the correction are not ideal,’ but this statement is based only on the average loss, which could be impacted by a small number of events with larger losses relative to the reference. The assessment approach should be expanded, as will be discussed further in subsequent comments.

3. It is stated that 'so many factors seem to affect the differences between measurements,' but only two factors are noted (orifice area and wind profile). Other contributing factors should be described, or this sentence should be reformulated.

4. The final sentence in the Abstract would be much stronger and more broadly applicable as 'These types of errors must be considered when correcting precipitation measurement errors for different gauge types and configurations.'

A) Introduction

1. What is meant by "false" precipitation? (P2, L12)

2. It is stated that 'the transition from manual to automatic measurements was highly encouraged' by the SPICE IOC (P2, L25-26) – can you please elaborate on this? If I recall correctly, SPICE was organized in response to the transition to automation, not to advance or recommend this transition.

3. The Introduction transitions abruptly from a discussion of errors and adjustment functions to a discussion of manual vs. automated measurements (P2, L25-32). This information is valuable, but seems out of place here. The biases in gauge measurements are assessed relative to reference measurements; historically (e.g. the first WMO Solid Precipitation Intercomparison), the reference measurements were manual measurements using the DFIR. I suggest that the authors revise the Introduction to first describe these biases (they are presently introduced without context), then describe known biases (e.g. for automated vs. manual measurements), and then get into the different errors/contributing factors and adjustment functions. In short, the Introduction should establish the context for interpreting the results that will be presented, and should flow logically from topic to topic.

4. The transition from a discussion of biases between automatic and manual measurements to the statement 'Thus, intercomparisons at different sites around the world should be conducted to the test the performance of the automatic system and correct the precipitation measurements' is confusing. Why would testing at different sites around the world be helpful? It is difficult to follow the logic of this section, as currently presented.

5. While it is true that the SPICE intercomparison sites could have their own measurement objectives, it is not necessary to state this here (P3, L1-2), as the Qilian Mountains site was not a formal intercomparison site.

6. Which existing adjustment function? (P3, L7-8)

B) Materials and Methods

1. Is the DFAR configuration (with TRwS204 gauge) at the Qilian Mountains site used in the analysis? If not – which I believe to be the case – the DFAR configuration does not need to be introduced and discussed on P3, L23-27 (i.e. these two sentences can be deleted).

2. In Figure 1, it appears that the single-Alter shield slats on the $TRwS_{SA}$ (Figure 1b) are installed differently than those on the $CSPG_{SA}$ (Figure 1c). Those on the $TRwS_{SA}$ are oriented with the flat side of the slat toward the gauge (correct), while those on the $CSPG_{SA}$ are oriented with the flat side of the slat away from the gauge (incorrect). The two shield configurations are therefore not identical. The location of the centre of mass and distribution of the slat surface area will be different in each case, impacting how the slats respond to a given wind speed. The shields being identical is an important assumption in the assessment, so this difference should be noted in the manuscript.

3. How were the $TRwS_{SA}$ measurements adjusted to match the diameter of the CSPG? You indicate that the manufacturer changed a setting (P4, L10), but additional details would be helpful.

4. In the manual quality control and filtering process, are you referring to the mean 30 min humidity? The details should be provided to guide those who may want to use a similar procedure. Also, it makes sense to remove precipitation during clear sky periods (sunshine duration = 0.5 h), but the threshold duration value is not very strict. For example, if the sunshine duration was 0.49 h, would the precipitation data be included in the analysis?

5. It is stated that $TRwS_{SA}$ precipitation 'was compared with the reference precipitation to investigate the performance of the $TRwS_{SA}$ and correct its measurements.' This makes it sound like the comparison with the reference corrects the $TRwS_{SA}$ measurements, which is not the case. Perhaps it would be clearer to state simply that the $TRwS_{SA}$ measurements were assessed relative to reference precipitation measurements from the $CSPG_{DFIR}$? I don't know if it is necessary to mention the adjusted/corrected measurements and their assessment at this point.

6. Again, no information is provided to indicate what is meant by "false" precipitation (P4, L30).

7. The reasons for not deriving transfer/adjustment functions from the experimental dataset and motivation for using the transfer function developed by Kochendorfer et al. are not clearly articulated. It is stated that 'it seems difficult to derive a valid and robust transfer function for $TRwS_{SA}$ using the dataset at this site during the experimental period.' Why is this difficult? What are the limitations of the dataset? Several important points were raised during the previous review stage and discussion, which should be reflected in the manuscript.

C) Results and Discussion

1. Numerous (20) 12-h precipitation events were noted in which the $TRwS_{SA}$ did not report precipitation, but the $CSPG_{DFIR}$ reported precipitation. Four events were rain and sleet, during which the conditions were 'nothing special'; can you please reword and elaborate on this? 16 events were snowfall events, which were evidently characterized by lower temperatures. Do you have a theory to explain why this may have been the case? What were the characteristic wind speeds? For example, if the precipitation was light and the wind speeds were higher, it would not be surprising if the single-Alter shielded gauge missed the event. Did the $CSPG_{SA}$ report precipitation during these events?

Another important concern is whether the 12-h conditions are representative of the conditions during which precipitation actually occurred during a given 12-h period. There's not necessarily a better way that you could have addressed the conditions, but the representativeness of conditions is an important point that must be noted.

2. When discussing precipitation losses (e.g. P6, L9), it is important to specify what the losses are relative to (i.e. the reference configuration).

3. Snow accumulating on the orifice and sublimating is proposed as a loss mechanism for the $TRwS_{SA}$. Were any incidents observed during the experimental period in which snow accumulated on the orifice, or is this just a theory? Any accumulated snow could also prevent incident snowfall from entering the orifice and being measured – that is, capping of the gauge may occur – which would influence the assessment.

I would assume that the CSPG is not heated, either. Did you see any snow accumulation on the CSPG? If so, is it included in the manual measurement, or is it removed prior to the measurement? Any differences in how precipitation accumulated on the gauge orifice is dealt with for the different gauge types could potentially impact the assessment.

Is there any other reason why accumulation/sublimation would be an issue for the $TRwS_{SA}$ and not the CSPG gauges? It is stated that accumulation/sublimation 'may explain most of the situations in which the $TRwS_{SA}$ recorded values of 0 for snowfall events,' but the reasons why this does not occur for the CSPG gauges are not discussed.

4. Is data loss considered to be another precipitation loss mechanism for the TRwS gauge? If the gauge reports are based on the weight of accumulated precipitation, won't the reports following any periods of data loss include any precipitation accumulated in the bucket during the loss period?

5. The title of Table 1 indicates a comparison of events on a 12-h scale; however, the notes indicate daily (24-h) mean wind speeds and temperatures are presented. Is this correct? As noted above, the conditions during 12-h periods are not necessarily representative of the conditions during which precipitation occurred. This issue will be more significant if conditions over 24-h periods are used.

6. I would consider removing the statement 'It is obvious that the $CSPG_{SA}$ performed better than the $TRwS_{SA}$ during the experiment at this site' (P6, L32). This is a subjective statement. Alternatively, it could be changed to something like, 'It is evident that the $CSPG_{SA}$ collected more precipitation relative to the $TRwS_{SA}$ during the experiment at this site.'

At the low mean wind speeds experienced at the site over the experiment (Table 1), the differences between measurements from a single-Alter shielded gauge and those from a DFIR-shielded gauge may not be significant. In this case, the $CSPG_{SA}$ and $CSPG_{DFIR}$ measurements would be expected to be very similar, and any differences relative to the $TRwS_{SA}$ could be attributed to systematic differences in the sampling, principle of operation, aerodynamic profile, etc. between the gauge types. These systematic differences are addressed in the $TRwS_{SA}$/$CSPG_{SA}$ comparison.

7. It is stated that 'the performance of the $CSPG_{SA}$ was more stable than that of the $TRwS_{SA}$' (P7, L13-14), but this comparison does not acknowledge that the CSPG is a manual gauge (how can it be unstable?), or define what is meant by 'stable.' Is this statement made in reference to the lower standard deviation for the $CSPG_{SA}$? If so, when considering measurements over such long time scales (12-h), does the standard deviation really say anything about noise or signal variability?

8. Aside from wind speed and temperature, which other 'meteorological variables have relationships with the catch ratio' (P7, L17)? Also, when noting that air temperature and wind speed are 'typically' applied in adjustment functions, it would be valuable to include supporting references.

9. The trends described for the catch ratio as a function of wind speed and temperature in Figure 3 (text on P7, L22-24) are difficult to observe in the plots. Binning the results for each precipitation type (rain, sleet, snow) and plotting as box and whisker plots would provide a much clearer representation of the data trends, and are recommended to complement the scatter plots provided in Figure 3.

10. RMSE values are computed for observed catch ratios relative to adjusted values. It would be valuable to also include RMSE values for observed $TRwS_{SA}$ accumulation values relative to the adjusted values.

11. On page 8, the performance of the adjustment functions is assessed in terms of average precipitation losses in each precipitation type. This assessment should be complemented by RMSE and bias values for the observed $TRwS_{SA}$ accumulation values relative to the adjusted values. In the work of Kochendorfer et al., it was found that the adjustments improved the bias in value relative to the reference, and had less of an impact on the RMSE; it would be interesting to see if similar trends are observed using this experimental dataset.

12. Can you please elaborate on the 'other possible errors' that may contribute to differences between $CSPG_{DFIR}$ and $TRwS_{SA}$ measurements (P9, L4-5)?

13. Do you think that the similarity of the mean absolute differences between the $CSPG_{SA}$ and $TRwS_{SA}$ in all precip types (P9, L10) may indicate a systematic difference in measurements between these gauge configurations? That is, does the combined influence of errors specific to the $TRwS_{SA}$ result in a systematic offset in measurements relative to the CSPG?

14. In Figure 5, the cumulative sums of precipitation accumulations are plotted as a function of event number – this is effectively a time series of the total accumulated precipitation for each gauge at the end of each event. It would be far more instructive to plot the individual event accumulations (accumulation at end of 12 h period minus the accumulation at the start of that 12 h period) for the $TRwS_{SA}$ vs. those for the $CSPG_{SA}$ (or vice versa) as a scatter plot. That way, each event can be compared independently, irrespective of the precipitation accumulated in previous events. A 1:1 line (indicating perfect agreement between the gauges) can be added to illustrate the accumulation trends.

15. As presented, the results in Figure 6 are difficult to interpret. I recommend generating histograms of the differences between the $CSPG_{SA}$ and $TRwS_{SA}$ measurements for each precipitation type to more

clearly illustrate the magnitude of differences between the gauges, and how those differences are distributed. Alternatively, the $TRwS_{SA}$ vs. $CSPG_{SA}$ scatter plots (see comment above) could be modified to include only points events with mean wind speeds < 1 m/s.

16. As noted above, the $TRwS_{SA}$ and $CSPG_{SA}$ technically have the same shields, but the orientation of the slats is different (Fig. 1). This potentially complicating factor should be noted in the relevant discussion in Section 3.3.

17. It is interesting that limiting the $CSPG_{SA}$/$TRwS_{SA}$ comparison to wind speeds below 1 m/s results in almost the same mean absolute differences between the measured values as observed for the full wind speed range. I wonder if this is a reflection of the full wind speed range being low (at least in terms of the mean wind speeds in Table 1), a reflection of the mean wind speeds for 12-h periods not being representative, or an indication of differences/errors not related to the aerodynamic profile, as proposed.

18. The trends indicated in Figure 7 do not appear to be conclusive for sleet and snow, given the significant scatter and small numbers of events relative to rain (I expect the R^2 correlation values would be low). This caveat should be noted when making the statement, 'it may be inferred that the amount of precipitation mainly affects the specific errors of $TRwS_{SA}$ during the experiment at this site.'

D) Conclusions

1. As noted above, I don't think that the standard deviation of losses is a representation of measurement stability; the losses are larger for the $TRwS_{SA}$, so the standard deviation of loss values will also be larger.

2. How is it 'clear' that the 'worse correction for $TRwS_{SA}$ measurements should not be attributed to the limited meteorological conditions and precipitation data during the experiment'? I don't believe that these points were addressed in the manuscript; an earlier comment requested more details in this regard.

3. I don't know if you are in a position to say that the adjustments are 'better' or 'worse' when the changes in loss values after adjustment for both gauges are within 0.2 mm for all precip types. You indicate that the measurement uncertainty is ignored (P10, L27), but what is the estimated uncertainty for each gauge type? Perhaps more important, you are using wind speeds over 12-h periods for the adjustment that do not necessarily reflect the conditions during which precipitation actually occurred, which will impart significant uncertainty on the adjusted values. I think it is OK to state the results obtained, but any broader application of these results should be treated with extreme caution.

4. The last two sentences on page 11 are very difficult to follow. Can you please rephrase these, including relevant results from the study, if possible, to demonstrate your points?

Proposed technical revisions

Abstract

P1, L10: change to 'deriving adjustment algorithms'

P1, L10: change to 'have become a top priority.'

P1, L10: remove 'mainly'

P1, L10-13: propose restructuring as 'This study analyzed precipitation measurements from single-Alter shielded MPS TRwS204 ($TRwS_{SA}$) automatic weighing gauges relative to corrected manual measurements from the Chinese Standard Precipitation Gauge in a DFIR-shield ($CSPG_{DFIR}$) in the Qilian Mountains, China. Results were compared over the period from August 2014 to April 2017, and show that precipitation collected with the TRwS204 was…'

P1, L13-14: propose restructuring as 'Applying the adjustment function reduced the average loss of the TRwS204 by 0.2 mm for snowfall events, but increased the loss by 0.2 mm for both…'

P1, L19: delete 'also'; change 'discussion' to 'investigation'

P1, L19-21: propose changing to 'Differences between precipitation measurements from the $TRwS_{SA}$ and CSPG in a single-Alter shield ($CSPG_{SA}$) were believed to result from differences in gauge orifice areas and gauge wind profiles.'

1. Introduction

P2, L1: 'Accurate and reliable precipitation datasets'

P2, L2: change 'were mainly' to 'are typically' and delete 'basic'

P2, L12: '…caused by aerodynamic effects, wetting losses, evaporation losses, trace precipitation…'

P2, L17: delete 'successively'

P2, L19: delete 'continuously'

P2, L19: change to 'With automatic field instruments used widely in national weather operations,'

P2, L20-21: delete 'which are necessary in practical work'

P2, L24: 'tested this function using precipitation measurements from Norway and the US, and proposed…'

2. Materials and Methods

P3, L16: change to 'Chinese national precipitation gauges'

P3, L23: change 'eliminate' to 'mitigate'

P3, L29-31: change to 'The precipitation dataset was obtained from two sources: the recorded values from the $TRwS_{SA}$ and the manually observed values from the $CSPG_{DFIR}$.'

P4, L5: change 'we did not enable the heating options' to 'heating was not enabled'

P4, L10-13: change to 'Before calculating the $TRwS_{SA}$ precipitation amounts, the recorded values were subject to a manual quality control and filtering process. First, 30-min precipitation data were removed if the corresponding humidity was less than 50%. Second, 30-min precipitation data were removed if the corresponding sunshine duration was equal to 0.5 h.'

P4, L20: insert 'from the $CSPG_{DFIR}$' after 'reference precipitation'

P4, L22: delete 'to some extent'

P4, L27: change 'far away to each other' to 'far away from each other'

P4, L28: delete 'like' and pluralize 'aerodynamic effects'

P4, L32: change 'effectively prevent wind effect' to 'effectively mitigate wind effects'

P5, L4: change 'For ΔP_t, it was assigned a value…' to 'ΔP_t was assigned a value…'

P5, L5: change to '$TRwS_{SA}$ measurements were considered to be subject to both specific errors and systematic errors.'

P5, L12: change 'are usually applied' to 'is usually applied'

P5, L13: change 'it seems difficult' to 'it is difficult'

P5, L14: change 'large amounts of data and various sites' to 'precipitation datasets from sites in different climate regimes'

P5, L22: change 'was aim' to 'aimed'

P5, L23: change 'means that there are no specific errors' to 'means that specific errors for the $TRwS_{SA}$ are not considered.'

P5, L24-25: change to '… Eq. (2); however, the quantification of specific errors for the $TRwS_{SA}$ is difficult. Therefore, Eqs. (1), (2) and (3) were applied to correct the $TRwS_{SA}$ measurements, ignoring specific errors.'

3. Results and Discussion

P6, L1: change 'Especially between' to 'Between'

P6, L7: delete 'obviously'

P6, L15-16: change to 'For snowfall events, wind effects will reduce the amount of snow reaching the $TRwS_{SA}$ gauge orifice.

P6, L16-17: change to 'Given that the orifice heating option was not enabled, snow accumulated on the gauge orifice may sublimate before melting and falling into the bucket.'

P7, L21-22: change 'It is obvious that these scatters disperse a lot on the plots' to 'Significant scatter is evident in both plots.'

P7, L31-32: change to 'The adjustment function in Eq. (3) is fit to the $TRwS_{SA}$ catch ratio as a function of both gauge-height wind speed and air temperature in Fig. 4.'

P8, L1-2: change to '…the corrected CRs is 0.01; this low value is not surprising, given the significant scatter observed in Fig. 4.'

P9, L14: consider changing to: 'As shown in Fig. 1, the shapes of the TRwS and CSPG are different: the TRwS is wider at the bottom, and more narrow at the top, with a 'shoulder' in between the two diameter portions; the CSPG is a cylinder of constant diameter.'

Figures and tables

P19, L4-5: change to '...and reference gauge, with 0.3 < CR < 1.7, were included.'

P21, L5-6: same as above

P21, L3: symbol for degrees Celsius looks strange

P25, L3: same as above

Table 2, 3: change 'Obse' to 'Obs'

P20 and 22, L4: change to… 'and in which 0.3 < CR < 1.7, were included.'