# An assessment of two automated snow water equivalent instruments during the WMO Solid Precipitation Intercomparison Experiment

Craig D. Smith, Anna Kontu, Richard Laffin, and John W. Pomeroy

We would like to thank the referees for taking the time to review this paper submission. Your comments and suggestions are most appreciated and we have tried to address all of them in turn, providing clarification and revisions where necessary. Our responses are highlighted in the text below. Incorporating your suggestions and addressing your comments and concerns have certainly helped to improve this manuscript.

# Response to referee comments

## Anonymous Referee #1

Although snow water equivalent (SWE) is very important information for not only disaster forecasting but also for earth science, there is still room for discussion of the methods of automatic SWE measurement. They compared different automatic SWE measurement methods (CS725, SSG1000), which are based on the different principles, with the manual SWE measurement at three sites. Then they discussed the characteristics of the comparison results at each site. Although I am very interested in their works and do not doubt that their works give the basic important information for the improvement of automatic SWE measurement methods, their work only shows few scientific new findings in the present version. From this view, I think their topic should be suitable for rather GI (or some publications which mainly treat the topics of instrument, method) publication than TC publication. In order for this manuscript to be accepted for TC, the authors should completely reconstruct the manuscript to clear the new scientific findings and scientific contribution of their works.

### Author's response

We will reconstruct the Results section and arrange the text by instrument, not by site, as suggested by Referee #2. We hope that this change will highlight our results including the new findings that came out of the SPICE intercomparison.

### Author's changes in manuscript

- Text in **3 Results and Discussion** arranged by instrument and divided into two sections, **3 Results** and **4 Discussion**.

## Anonymous Referee #2

The paper compares SWE manual measurements with automated sensors of attenuation of passive gamma radiation attenuation device in two SPICE sites and another site with rather different climatic and snow conditions; and also one scale snow in one of the SPICE sites. The topic is of high interest for a broad community as the sensors that are compared are starting to be very popular in many sites of the world, and it is necessary to discuss about their accuracy, possible sources of uncertainty, etc. Due to the limited length of snow observations, few locations analysed and some problems related with the experimental design it is not possible to give strong evidences on their accuracy and the reason of biases found between manual and automated measurements. However, I think that the paper still has enough interest for many readers, and it launches some hypothesis of interest that may serve as basis for further research.

In my opinion, the structure of the manuscript is not the best to present the results. I recommend to reorganize the presentation of them showing the equivalent figures for each site together, instead of doing subsections of the results for each site (indeed the discussion is presented in the way I suggest). Meanwhile table 1 and 2 can be combined (and added results from Fortress). In this way it would be reduced the final number of figures (the 14 current figures is excessive in my opinion). More important, it would be possible to identify common processes amongst sites and their differences, and the paper would gain consistency (in the current version some figures are made for one site but not for the others..e.g.). Thus it could be presented the validation of CS725 and SSG1000 (where available) in the three sites with a couple of multi-panel lots (one panel per site), and afterwards to show figures that allow explaining the patterns of accuracy/error shown (the figures relating air temperature and difference SWE, the soil moisture...

One concern is how to ensure that snow depth in automated sensors is the same than that were SWE is measured manually in the three presented sites. This could be another source of error not mentioned in the manuscript. In page 11 is mentioned the spatial variability in snow melting that could affect to different snow depth. Are there snow depth sensors installed above the measured areas with automated sensors. If this is the case, we could see how well the depths are similar and if there are differences, some plot could focus on estimated snow density.

Is there any evidence of a relationship between snow depth or amount of SWE with bias with the manual measurements. There are some references that CS725 may be inaccurate under very deep snowpack. - I think that Figure 11 should show the bias between manual and automated measurements to properly observe the coincidence between liquid water content and SWE differences.

It seems that the existence of water around SSG1000 may cause serious disruptions in the functioning of the device. Is it apparently due to problems in their installation or is a problem of design of the device?

Hoping my comments will result useful.

## Author's response

The referee points out the limited length of observations, few locations, and problems with experimental design. This was largely related to the choice of limiting analysis to the SPICE instrumentation over the SPICE intercomparison period and using available reference measurements. We did add CS725 data from Fortress Mountain (which is not a SPICE site) to help support our theory that CS725 offsets were due to changes in soil moisture related to sandy soil substrates. To have more data and locations, we will add one more site, Weissfluhjoch, to the SSG1000 intercomparison which adds 2 seasons of instrument intercomparisons. However, as the referee points out in regards to the existing intercomparison, the methods for the manual reference measurements differ from the other sites. This contributes to the difficulty in assessing the level of uncertainty in the reference.

The referee suggests looking at automated snow depths if they are measured in the same field of view as the SWE measurements. This, unfortunately, was not the case as the snow depth sensors were located at different locations participating in their own assessment. This deficiency is noted and we will comment that this could assist future SWE intercomparisons. We did use manual snow depth measurements at Caribou Creek and Sodankyla to estimate site variability and will add a commentary on this.

As suggested, we will rearrange the Results and Discussion section by instrument, not by site. We will also combine relevant figures and tables to reduce their number and to make comparisons between sites easier.

The problems with SSG1000 and water are due to a design feature: the cables are so short that the electronics box must be installed below the instrument on the ground. We know that Fortress Mountain site has also experienced similar problems with the instrument. After SPICE, the Sodankylä team asked the manufacturer to replace the cables with longer ones. A third measurement winter was without instrument failures, as the electronics box was now mounted about 50 cm above ground. This will be briefly mentioned in the text.

We revised Figure 11 (now Figure 6) to include the difference between the CS725 and manual SWE measurements.  This now shows the relative timing between the changes in the soil moisture/soil temperature and the changes in the instrument offset.  However, the difference in the frequency of measurements means that interpretation isn't necessarily clear.  As the manuscript states, it offers an explanation for some of the initial offset which is shown by the first intercomparison in mid-December but the large discrepancies shown in mid-February are not attributed to changes in liquid water in the soil anyways.  It does mark the coincidence between spring soil thaw and infiltration of liquid water and a corresponding increase in the sensor bias.

## Author's changes in manuscript

- Text in **3 Results and Discussion** arranged by instrument and relevant figures and tables combined. Text divided into two sections, **3 Results** and **4 Discussion**.
- Weissfluhjoch is added to the SSG1000 intercomparison.
- The SSG1000 problems will be addressed in the text.
- Revised Figure 11 (now Figure 6) to include a plot of the difference in SWE between the CS725 and the manual measurements in relation to the soil moisture/soil temperature changes
- We added further commentary in the Discussion section on how the experiment design could be improved for future SWE intercomparisons, including measuring snow depth in the same footprint as the SWE sensors.
- Added a commentary in the Discussion section on the snow depth variability at Caribou Creek and Sodankyla as an indicator of how the SWE could also vary with space and time.

## Referee #3 Charles Fierz

**General comments**

The goal of this paper, as stated by the authors, is, "to assess the use and accuracy of two instruments that were tested during the WMO - Solid Precipitation Intercomparison Experiment the Campbell Scientific CS725 and the Sommer Messtechnik SSG1000 snow scale" as well as, "to inform users of the best way to use these instruments and of any potential measurement issues that may influence their data interpretation." Unfortuately, however, I don't feel I get the promised info by reading that paper. The deficiencies of the CS725 (first 5 lines on page 4) are simply confirmed and no convincing, in depth analysis of possible source of errors are addressed for any of the two instruments. Instead, spatial variability is invoked to explain the mismatch between the continuous measurements and the manual, punctual (in time) reference measurements, the error of which are not quantified either. In view of the above and my comments below, I can hardly recommend to accept that paper for publication.

Indeed, I really doubt that the authors have enough convincing arguments and data to bring the paper in line with their goals, even after major revisions.

## Author's response

It is true that we confirm some deficiencies in the CS725, as the importance for soil moisture calibration of the CS725 is previously established (Martin et al. 2008; Wright et al. 2011). However, in previous comparisons the instrument has agreed much better with other measurements (automated or manual) than we have observed during SPICE. Our analysis shows that the CS725 assessed during SPICE did not compare as well with manual measurements as did previous intercomparisons and we propose that the larger offset is due to changes in soil moisture content after calibration and throughout the snow season. This was in turn caused by the soil type at both intercomparison sites which was predominately sand. The possible ways for soil moisture to change during winter are identified by Gray et al. (1985, 2001) and Lilbaek and Pomeroy (2008). We argue that these soil conditions did not exist for previous intercomparisons and instrument users need to be made aware of this. We also imply that this is not necessarily an instrument deficiency and some hydrological users may be able to use this measurement principle to their advantage.

We agree with the need of more in-depth analysis to support the hypotheses, and will provide this in a revised manuscript or make recommendations to guide future SWE sensor intercomparisons to fill these deficiencies.

The specific and minor comments are addressed after each comment. The resulting changes to the manuscript are also addressed after each comment.

**Specific comments**

•p. 1, lines 25-26: "These manual measurements are considered to be the reference for the intercomparison." This is one of the crux of that paper. The devices are hardly looked at while a whole lot of blame goes on these manual measurements, the error of which are hardly addressed though.

We now address some of the error in the manual measurements in the Discussion section. The literature addresses the mean bias associated with some manual samplers but the variability of the potential error in field sampling is quite large and depends highly on the skill of the user and the condition of the snow pack.

•p. 1, line 30: "throughout the intercomparison periods" is absolutely misleading and false. One full ablation period is missing and the problems of the instruments were not looked at.

One ablation period is missing for one sensor only…the SSG1000 at Sodankyla for the 2014/2015 period. However, we recognize that this limits the sample size of the intercomparison during ablation for this sensor (and for intercomparison with the other sensor). We will change the text "throughout the intercomparison periods" to "when data were available" where applicable. In addition, we will add SSG1000 intercomparison from an additional site

•p. 1, line 33: "seasonal melt" suggest replacing by "ablation period" [throughout the paper]. Furthermore, is pre-melt = accumulation period? I strongly suggest that you define these terms properly once and use them consistently throughout the paper. See for example on p. 6, line 31 for "point of maximum seasonal SWE"

We agree that consistent terminology is important. We will replace 'pre-melt" with "accumulation period" and "seasonal melt" with "ablation period".

•p. 3, line 13: "2 Instrumentation and Methods" Should try to not give interpretation in that paragraph but include it in the results section, for example as "previous intercomparison"

The commentary about previous intercomparisons has been moved to the "Introduction" section and all interpretation has been moved to "Results" or "Discussion".

•p. 3-4; lines 26-5: Is this the correct place for such comments? Should be moved to discussion part as an introduction to it.

Agreed

•p. 4, line 15: "... impact of the move are considered to be negligible." Why? Later you speak of spatial variability influencing the results.

We resolve this by including commentary on the spatial variability at this site and demonstrate how small this is over the distances of the sensor move.

•p. 4, lines 29-30: "... to stabilize the overlying snowpack and prevent ice bridging." Why does the snowpack need to be stabilized? How is ice bridging prevented? What observations do corroborate this?

This is a statement made by the manufacturer (and is stated as such) to justify their design of a larger platform with a smaller platform in the centre that does the actual weight measurement. It is out of our scope to validate their design other than to comment on the quality of the measurements.

•p. 5, line 3: "..., and the only snow scale provided ..." is incorrect. There is another SPICE site (Weissfluhjoch) equipped with a snow scale ... and a snow pillow next to it from the same provider.

The text in the manuscript is correct. The snow scale in Weissfluhjoch was provided by the site, not by manufacturer for testing during SPICE. However, based on this comment as well as some other comments about lack of snow scale intercomparison data (due to instrument failure), we have chosen to add the snow scale intercomparison for Weissfluhjoch..

•p.5, line 8: "... reliable manner ..." but not always. The simple regression does not reveal the true problems!

This statement is in reference to the actual functionality of the instrument (see the context) rather than an assessment of the measurement accuracy. The instrument functioned with a low failure rate.

•p. 5, lines 13-15: "The sensitivity..." Such a sentence belongs to the summary and conclusion section.

We agree. The sentence will be moved.

•p. 5, lines 24-25: "... has a mean measurement error less than 0.5 %." 0.5 % of what? Does this refer to the repeatability of measurements? Overall, the number looks very optimistic and the reference Farnes 1983 is hardly available. From other publications by the same author (1980 and 1982, see Kinar & Pomeroy, 2015b), this figure can hardly be reproduced. I'd strongly suggest to be more precise here.

Farnes et al. (1982) state that ESC-30 overmeasures by -0.3 % (i.e. undermeasures) of the true SWE, and that the correction factor for ESC-30 is 1.00 (no correction required). The accuracy is quoted by Goodison et al. (1987) which was added as a reference. This paper may be more available than the original Farnes et al.

paper. Of course, these errors are in ideal situations, as stated by Kinar & Pomeroy who cite Powell (1987) in reference to errors in measuring more difficult snow packs. We add a discussion on this as it relates to measurements during SPICE (especially at Caribou Creek).

•p. 5, line 25: “were taken just inside the footprint of the CS725” How do these disturbances affect the measurements?

Since the sample is 30 $cm^2$ inside an 80 $m^2$ sensor footprint, the impact is negligible but the sample area was filled in with discarded snow when possible. This was clarified in the text.

•p. 7, lines 14-15: “the instrument trends are the same as for the manual measurements” In my view your simple regression analysis fails here and does not look at problematic features. For example, how do you explain the apparent loss of mass around mid February 2014? Similar unexpected wiggles are also seen at other times on both seasons. These spurious measurements are also known to occur at Weissfluhjoch and are not to be expected from a well designed, continuously recording device.

Mid feb 2014, March 2014, March-April 2015: very cold periods (-30 C) after positive air temperatures resulted in ice bridging. The snow supports itself and the weight is not on the load cell. After adding the Weissfluhjoch data, we see it here as well (as the reviewer points out). This issue will be discussed in the revised paper in the Discussion section.

•p. 7, lines 15-21: This comparison or ‘tracker’ does not appear to work very well. Indeed, in January 2014 there is a large increase in the ‘Difference in SWE’ while air temperature plummets! Similar behaviour can be found at other times. In summary, there is another reason behind these large increases, but which?

Agreed, the correlation between temperature increase and the immediate corresponding increase in sensor bias isn’t always clear. Sources of error are now presented in the Discussion section, including the potential issues with manually sampling a snow pack that persists after a melting period as a result of ice layers, etc,

•p. 11, line 11: “systematic sampling errors” Can these be avoided?

I see the reviewer’s point. This was poorly worded. These sampling errors could be systematic but not necessarily so. The text was changed.

•p. 11, lines 24-27: “Although ...” A somehow simplistic view. In the paper you never assess any of the errors you assign the outliers to. This is definitely the biggest weakness of that paper.

The text referenced here has been changed and we now address the errors in more detail, including ice bridging, in the Discussions section.

•p. 11, lines 30-36: Poor conclusions! What does this linear relationship show? Would you calibrate the CS725 with a SSG1000? Were the deficiencies of the CS725 not already known (see your introduction)?

The linear relationship is meant to show that it’s not just the manual measurements causing the bias but rather the measurement principle of the sensor. We have tried to make this clearer in both the Discussion and the Conclusion sections. The previous literature does point out some known deficiencies in the CS725 but the behavior seen at Sodankyla, especially during melt, has never been documented. The comparison with the SSG1000 supports our conclusions that the increased bias is related to infiltration of meltwater into the sandy soils, and serves as a warning to users who may be using the instrument in similar situations.

•p. 12, line 4: “... not all increases in the bias... ” Interesting, you don’t even mention those in the discussion!

This has been corrected, see note above about sources of error

•p. 12, line 8: “... errors in the manual SWE measurements....” I agree that measurement errors can amount to a certain percentage of SWE. But you don’t even quantify these errors, even though you use them as reference. Blaming not quantified errors for the observed mismatch seems simplistic indeed.

It is difficult to estimate the measurement error in the field because it highly depends on the measurement conditions and the observer, however we do attach some loose estimates on this error in the Discussions section on Manual SWE measurements.

•p. 12, line 32: “... have a good agreement ...” Here I have really hard times to follow the logic of your conclusions. First you blame the manual measurements for mismatch and then you claim a good agreement!

The point was that there was a good agreement considering the potential issues with the manual measurements and the measurement principle of the sensor. Hopefully the revised text makes this clearer.

**Minor comments**

•Please ALWAYS put a space between numbers and units (often wrong)

We will correct these.

•p. 1; line 28: Replace “(w.e.)” by “(mm w.e.)”.

We will replace this.

•p. 1; line 29: Replace “Creek respectively” by “Creek, respectively”.

We will correct this.

•p. 2; line 7: Are these two units equivalent?

Yes they are, assuming that density of water is 1 kg/m3.

•p. 3; line 1: Replace “(SPICE; Nitu” by “SPICE (Nitu”.

Replaced with “(SPICE) (Nitu”

•p. 3; line 28: Replace “snow cores” by “snow courses”.

No, they actually compared to 8 snow samples (cores), not 8 snow courses.

•p. 3; lines 28-29: What are “snow pit densities”? Please describe. Sampler?

It means that a density profile of 5 cm high samples is measured, SWE is determined for each sample and the total SWE is calculated as a sum of the layers. This is thought to be more accurate than SWE from bulk snow sample (core). Details will be added to the revised manuscript.

•p. 4; line 30: “0.3 % of full scale” that is 3 mm w.e.! Under what conditions? Moreover, the high resolution seems useless.

This is what the manufacturer tells about the technical details of their instrument. They do not specify the conditions, or give reasons for such high resolution.

•p. 4, line 35: What is drifting? Snow? The electronics?

This was in reference to drifting snow.  This has been clarified in the text.

•p. 5; line 23: “snow tube” Is this the correct term? I’d suggest using “snow sampler” -as found elsewhere in the literature –throughout your paper.

Snow sampler can be any kind of sampler, from 5 cm high wedge sampler to a long tube. “Snow tube” defines the type of sampler used, and is a widely used term (e.g. by Kinar and Pomeroy 2015).

•p. 7; line 5: Replace “almost” by “by almost”.

We will replace this.

•p. 7; line 29: Replace “offset” by “intercept”, throughout, as used in the tables.

We will replace this.

•p. 7; line 32: What does “differential” mean?

Different melt rates at different locations at the site.

•p. 8; line 8: Try to read “2013/2014 only due to data unavailability for 2014/2015” and replace that sentence.

Clarified this sentence

p. 8; line 30: “melt and re-freeze occurred” That really questions the term “pre-melt” used elsewhere in the paper.

‘Onset of snow-melt’ and ‘snow-melt season’ are commonly used terms for the final snow melt (ablation) in the spring. They do not rule out the possibility of short melt-refreeze cycles during the accumulation period. However, we will replace ‘pre-melt’ with ‘accumulation period’ in the text for clarity.

•p. 9; lines 32-33: “gravimetric water fraction” vs “volumetric water content ”What is the relation?

Volumetric water content (theta) is the volume of water divided by the total volume. Gravimetric water content or fraction (u) is the mass of water divided by dry soil mass. They are related by theta = u SG, where SG is the soil specific gravity and depends on its density.

## Author’s changes in manuscript

Changes to the manuscript are integrated into the specific responses to the referee’s comments above.