The WMO SPICE Snow-on-Ground Intercomparison: An Overview of sensor assessment and recommendations on best practices

Craig D. Smith*, Environment and Climate Change Canada, Saskatoon, Canada, craig.smith2@canada.ca

Samuel Morin, Météo-France, Grenoble, France
Anna Kontu, FMI, Sodankylä, Finland
Samuel Buisan, AEMET, Zaragoza, Spain
Lauren Arnold, Environment and Climate Change Canada, Saskatoon, Canada
Henna-Reetta Hannula, FMI, Sodankylä, Finland
Leena Leppänen, FMI, Sodankylä, Finland
Antonella Senese, University of Milan, Milan, Italy
Javier Alastrué, AEMET, Zaragoza, Spain
José Luís Collado, AEMET, Zaragoza, Spain
Guglielmina Diolaiuti, University of Milan, Milan, Italy
Yves-Alain Roulet, MétéoSuisse, Payerne, Switzerland
Rodica Nitu, Environment and Climate Change Canada, Toronto, Canada
Audrey Reverdin, MétéoSuisse, Payerne, Switzerland
Timo Laine, FMI, Helsinki, Finland
Antti Poikonen, FMI, Helsinki, Finland
*corresponding author

Abstract: One of the objectives of the WMO Solid Precipitation Intercomparison Experiment (SPICE) was to assess the performance and capabilities of automated sensors for measuring snow on the ground (SoG), including sensors that measure snow depth and snow water equivalent (SWE). The intercomparison focused on five snow depth sensors (models SHM30, SL300, SR50A, FLS-CH 10 and USH-8) and two SWE sensors (models CS725 and SSG1000) over two winter seasons (2013/2014 and 2014/2015). A brief discussion of the measurement reference(s) and an example of the intercomparisons are included. Generally, each of the sensors under test operated according to the manufacturer's specifications and compared well with the site references, exhibiting high correlations with both the manual and automated reference measurements. The use of natural and artificial surface targets under snow depth sensors were examined in the context of providing a stable and representative surface for snow depth measurements. An assessment of sensor derived measurement quality and sensor return signal strength, where available as an output option, were analysed to help explain measurement outliers and sources of uncertainty with the goal of improving data quality and maximizing the sensor capabilities. Finally, where possible, relationships are established between the gauge measurement of solid precipitation and the measurement of snow on the ground. This paper will provide a brief summary of these results with more detail included in the WMO SPICE Final Report.

1. Introduction

The objective of the WMO Solid Precipitation Intercomparison Experiment (SPICE) is to investigate and report on the use of automated instruments for measuring snow (Nitu et al., 2012; Rasmussen et al., 2012), including the measurement of snow-on-ground (SoG). During the SPICE intercomparison period, which took place over the (northern hemisphere) winters of 2013/2014 and 2014/2015, various sensors designed to measure snow depth and snow water equivalent (SWE) were compared to a reference measurement to assess their capability to provide an accurate and reliable measurement. In addition to the assessment of the sensors, the expert team has provided many recommendations on the best practices for using various sensors and interpreting sensor data in a variety of climate regimes. Included in this summary is an assessment of several surface target types used for the measurement of snow depth, an overview of sensor diagnostic output (where available) with some suggestions on how it could be used to improve measurements, and an overview of the relationship between change in snow depth and reference gauge measurements at several SPICE sites to assess the capability of snow depth sensors for estimating precipitation.

The SPICE sites participating in the SoG assessment in varying capacities are CARE (Canada), Caribou Creek (Canada), Col de Porte (France), Formigal (Spain), Forni Glacier (Italy), Gochang (Korea), Pyramid Observatory (Nepal/Italy), Sodankylä (Finland), and Weissfluhjoch (Switzerland). Not all are discussed here. Snow depth sensors under test include the Campbell Scientific SR50A/ATH, the Dimetix FLS-CH 10, the Felix SL300, the Jenoptik/Lufft SHM30, and the Sommer USH-8. The SWE sensors under test are the Campbell Scientific CS725 and the Sommer SSG1000. A more detailed description of these instruments can be found in the Instrument Performance Reports (IPRs) that are included as an annex in the SPICE Final Report. There were other snow depth sensors provided by site hosts that are not discussed in this summary but are discussed in the SPICE Final Report. More information on the sites and the site instrumentation can be found in the site commissioning reports at http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html.

2. Intercomparison References

Intercomparisons between snow depth sensors under test (SUT) and the reference(s) were completed for the SR50A/TH (CARE, Col de Porte and Sodankylä), FLS-CH 10 (Col de Porte), SL300 (Care and Sodankylä), SHM30 (CARE, Col de Porte, Sodankylä and Weissfluhjoch), and the USH-8 (CARE and Sodankylä). Due to the nature of the sites and the project's experiment plan, the manual reference methodology differs by site. Observations of manual snow stakes are made at each site but the frequency and methods of observation differ. For example, CARE provided observations of 62 snow stakes that were visually measured by an observer on a daily basis. The intercomparison at CARE is performed using a sub-set of these snow stakes mounted at each corner of each snow depth target. The 4 snow stakes at Sodankylä, dispersed throughout the intercomparison field, were also observed daily but via photography with the depths later interpreted and extracted from the photos. The 3 snow stakes at Col de Porte were observed visually by an observer but only once per week.

An alternative for a reference measurement for snow depth is to use the mean of all automated snow depth sensors to represent a high frequency (1 minute resolution) automated reference. The theory is that by averaging measurements made by different sensors employing different measurement principles, any systematic bias experienced by an individual sensor is minimized. However, the number and type of sensors differs by site as does their configuration. For example, the sensors at Col de Porte all measure roughly the same target area while the sensors at Sodankylä are dispersed throughout the intercomparison field. At CARE, there are 3 pedestals (in relatively close proximity) each with 4 sensors measuring 3 target areas. Averaging the 4 sensor measurements at each pedestal produces an automated reference for each pedestal. Further details are provided in the Final Report and in the IPRs.

The reference for SWE sensors at both Caribou Creek and Sodankylä are manual bulk density SWE observations measured with a snow tube and obtained roughly every 2 weeks. Measurements were taken at approximately the same location each time at each of the sites. As there were typically only one or two SWE sensors at each site, an automated reference is not achievable.

3. Intercomparison Results

Only some examples of the intercomparison results are included here but all will be available in the SPICE Final report and the IPRs.

For snow depth, Sodankylä hosted 4 out of the 5 sensors under test listed above and was chosen to highlight the intercomparison here. Figure 1 shows the scatter plots of the manual reference intercomparison for the 4 sensors for the 2013/2014 and 2014/2015 combined seasons while Figure 2 shows the scatter plots for the automated reference intercomparison. Figure 3 shows an example of a time series intercomparison with the manual reference, highlighting the 2013/2014 season intercomparison for the USH-8.

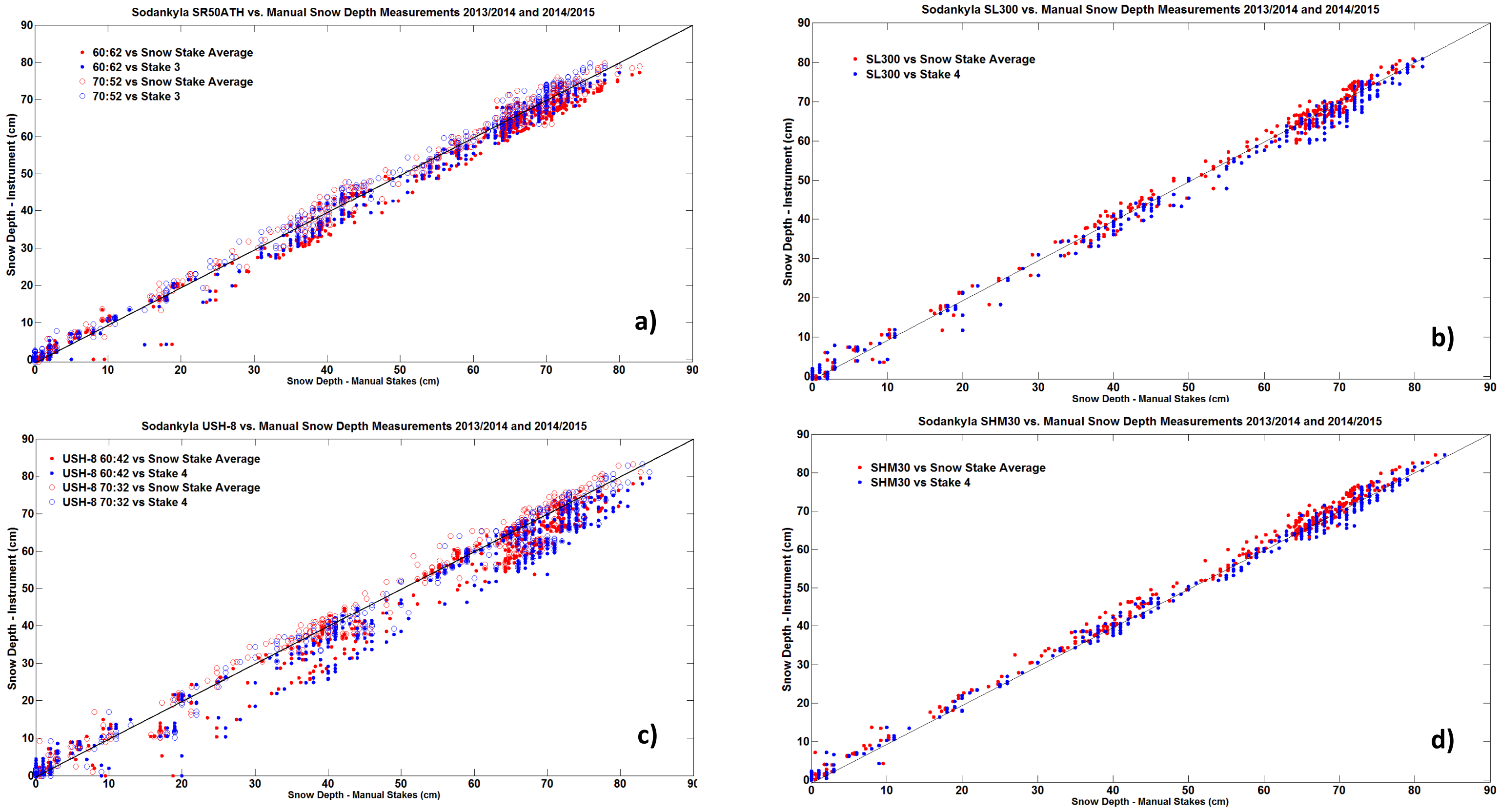

Figure 1: Intercomparison of snow depth sensors with the manual reference at Sodankylä for the a) SR50ATH, b) SL300, c) USH-8, and d) SHM30. The manual reference shown here is either an average of all four snow stakes (red) or the closest snow stake to the sensor (blue).

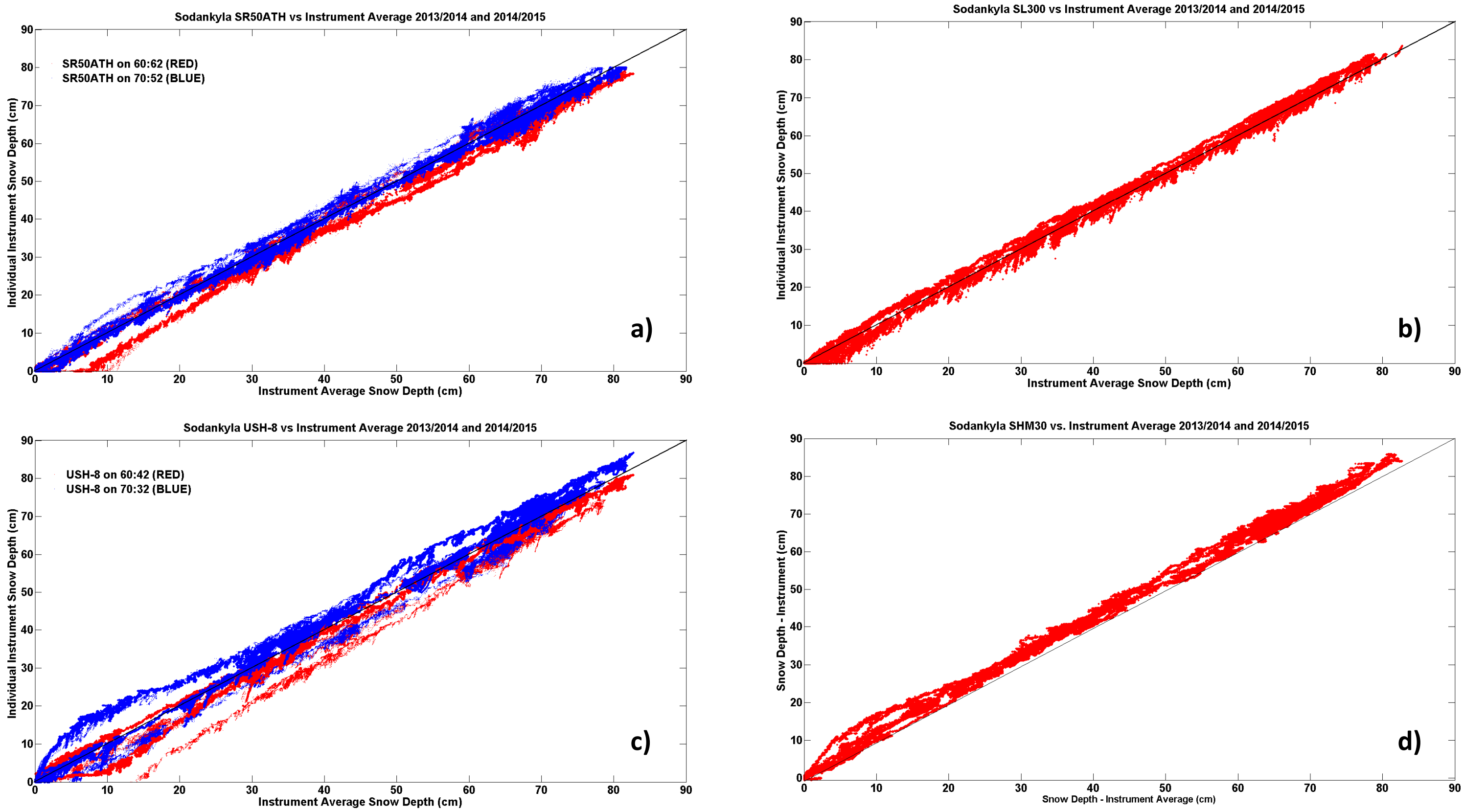

Figure 2: Intercomparison of snow depth sensors with the automated reference at Sodankylä for the a) SR50ATH, b) SL300, c) USH-8, and d) SHM30.

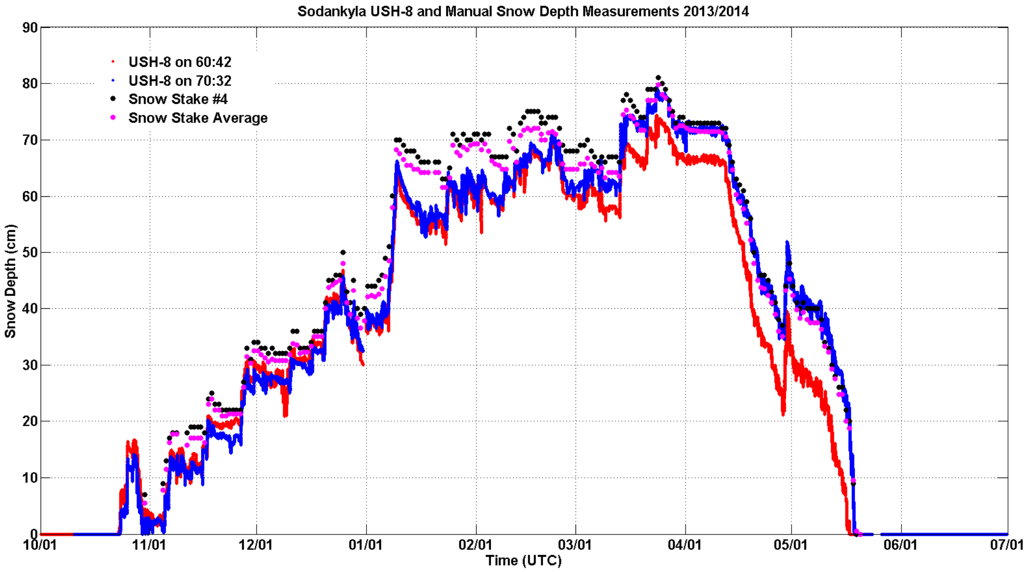

Figure 3: Time series intercomparison with the manual reference for the USH-8s at Sodankylä for the 2013/2014 season.

For SWE, the intercomparisons between the two SUT and the manual reference at Sodanlyä are shown in Figure 4 (scatterplot) and Figure 5 (time series for 2013/2014, CS725 only).

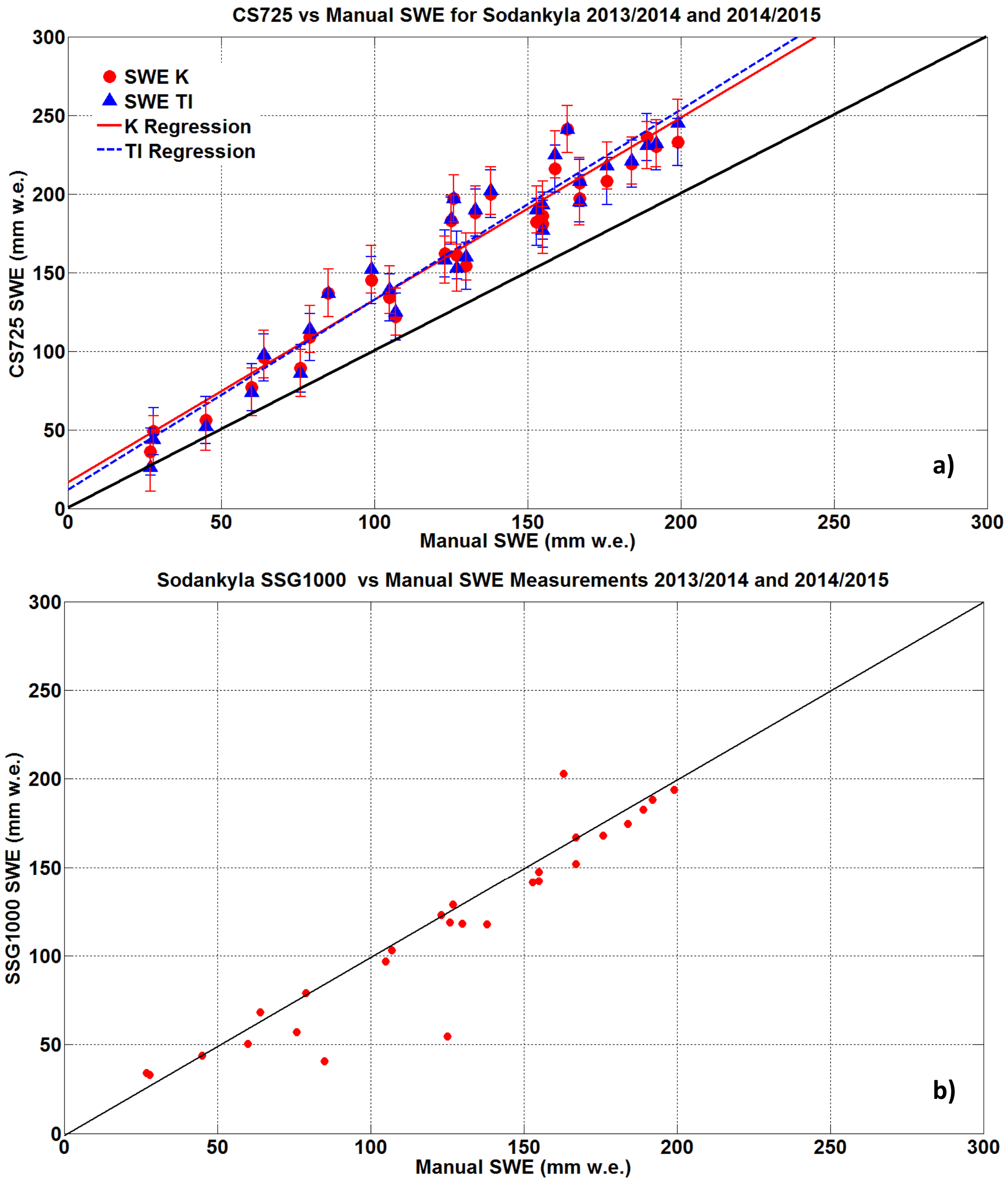

Figure 4: Intercomparison of the a) CS725 and b) SG1000 with the manual SWE reference at Sodankylä. For the CS725, the red markers represent the sensor output based on Potassium and the blue markers for Thallium. The error bars represent the manufacturer stated instrument accuracy.

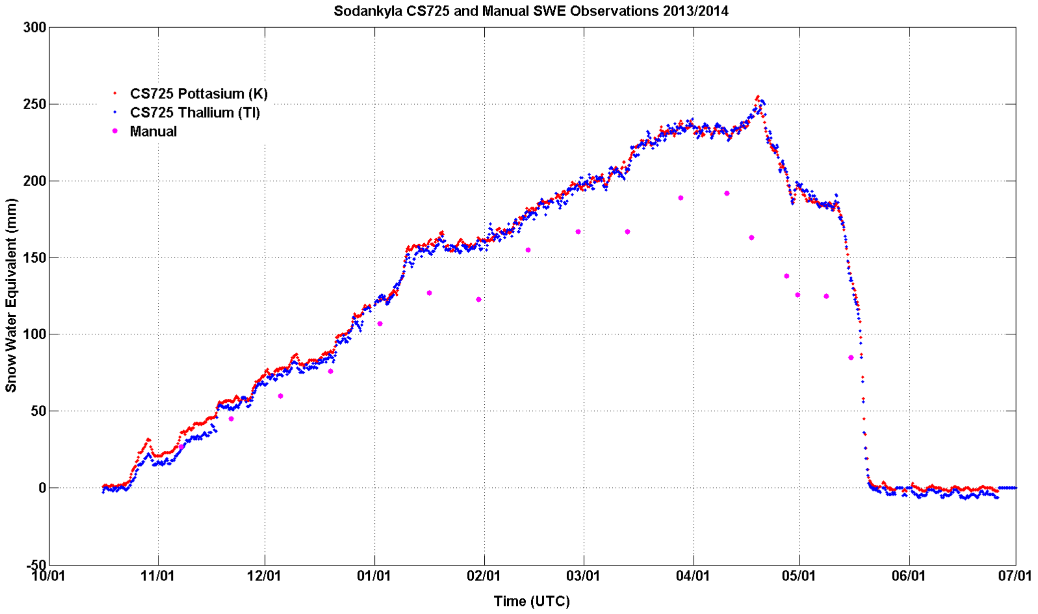

Figure 5: Intercomparison of the CS725 and manual SWE time series for the 2013/2014 season at Sodankylä.

The intercomparison results for snow depth generally show a good agreement between the SUT and the references with the data falling along with 1:1 lines in Figures 1 and 2. The r^2 values were all quite high varying from 0.86 to 1.00. For the most part, all snow depth instruments performed according to the manufacturer's specifications. In general, differences between the SUT and reference measurements are due to spatial variability in snow depth at the site rather than sensor bias. This is highlighted in Figure 3 which shows periods where there is very good agreement between the manual reference and the SUT but definite periods where the SUT is consistently measuring less snow.

For the SWE intercomparison, the CS725 (Figure 4a and Figure 5) clearly overestimates SWE as compared to the reference. The agreement between the SSG1000 measurement and the reference appears to be closer (Figure 4b) with some notable outliers that could be related to snow bridging over the measurement platform. The overestimation of the CS725 at Sodankylä (and at Caribou Creek, not shown) is believed to be related to a combination of the sensor's measurement principle and the soil conditions at the sites. This is explored further by Smith et al. (2016). The r^2 values for complete seasons ranged from 0.77 to 0.96 for the CS725 and 0.84 to 0.99 for the SSG1000.

4. Snow Depth Target Evaluation

As part of the SPICE SoG analysis, a qualitative evaluation of several snow depth sensor surface target types was completed. For SPICE, three surface targets were used. These were an artificial plastic target (CARE and Caribou Creek), an artificial turf target (Sodankylä) and natural ground (Col de Porte and Weissfluhjoch). The targets were evaluated for 1) their ability to provide a stable surface to allow the SUT to make an accurate and reliable measurement, 2) their ability to represent the surrounding landscape in capturing the first snowfall on a bare target, and 3) their ability to maintain a representative snow depth during snow melt.

Figure 6 shows the SUT measurements of snowfall events on bare targets at CARE (plastic), Sodankylä (artificial turf) and Col de Porte (natural ground). It appears that both the artificial plastic target (Figure 6a) and the artificial turf targets (Figure 6b) make a good reflective surface for sonic measurements as shown by the relatively low level of signal noise and the discernible increase in snow depth during the precipitation events. The natural target (Figure 6c) exhibits more noise than the other two targets, likely due to the presence of grass, but does not inhibit the sensor from registering a change in snow depth during the precipitation event. The target types, as evaluated for measurements made with an optical laser sensor, are mostly irrelevant suggesting that artificial targets may be more appropriate for sonic sensors and less relevant for laser sensors. However, the impact of target type (more specifically, target optical properties) on the SHM30 optical sensor signal strength output was examined and does have implications for using the signal strength output to detect light snow on bare targets (Section 5.2). More results are presented in the WMO SPICE final report.

Photographs were used to assess the target’s ability to hold snow falling onto the bare target and be representative of the surrounding landscape. Figure 7 shows photographs of the artificial plastic targets at CARE after a snowfall event. It appears that even though the targets are textured, snow tends to blow off of the surface, resulting in the surface holding less snow than the surrounding area. Figure 8 shows photographs of the artificial turf targets at Sodankylä after a snowfall event on the bare targets. The artificial turf appears to retain that first snowfall much better than the plastic targets. However, it must be noted that the wind speeds at Sodankylä are much lower than they are at CARE so caution is required when interpreting these qualitative results.

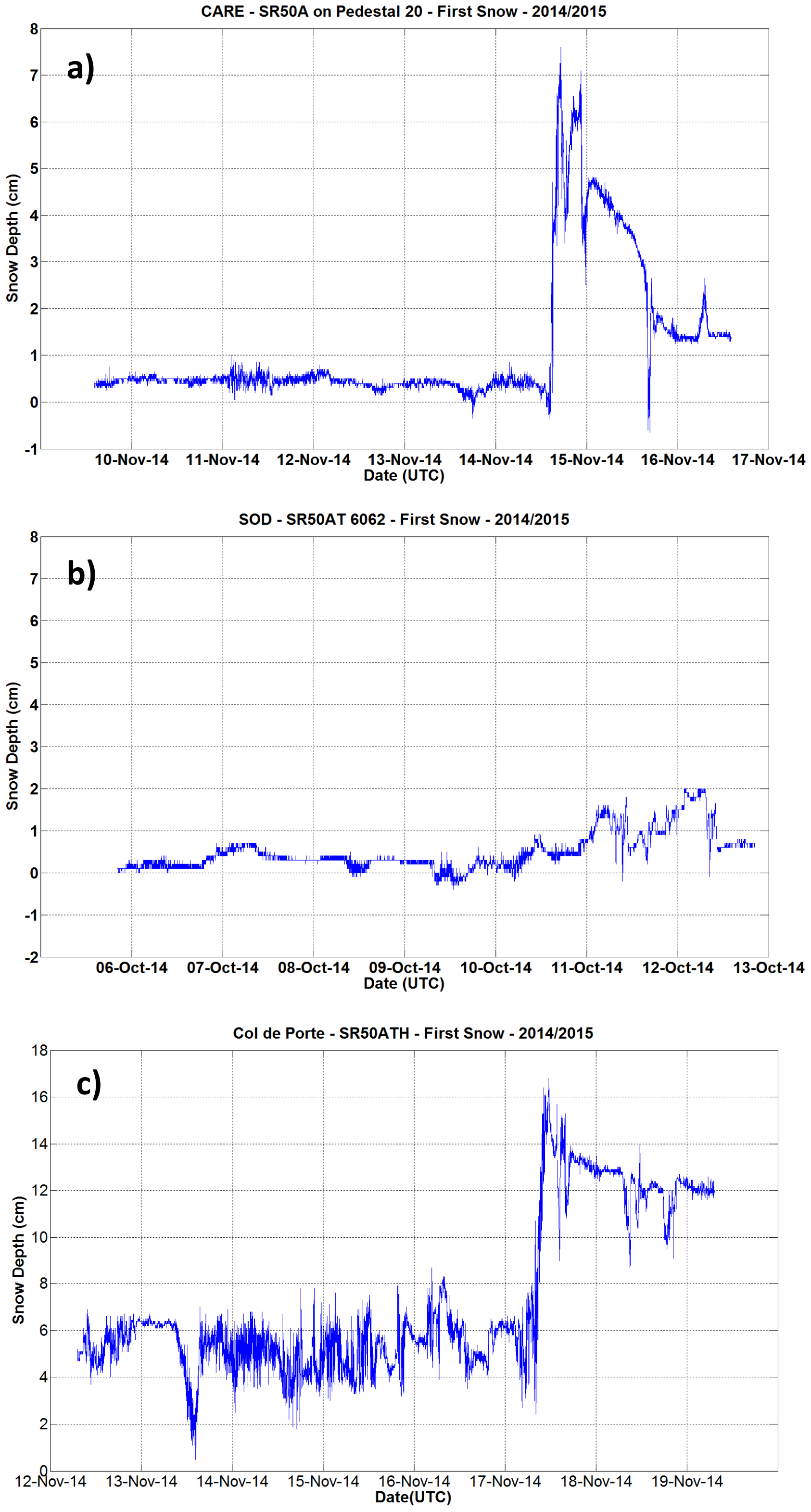

Figure 6: Measurements of bare target snow depths leading up to a snow event on the bare target as measured by the SR50A at a) CARE (plastic target), b) Sodankylä (artificial turf), and c) Col de Porte (mown grass).

Figure 7: Plastic target at CARE after snowfall.

Figure 8: Artificial turf target at Sodankylä after snowfall.

Melting of snow from artificial targets can be potentially impacted by the target's modification of the surface energy budget. If the target is darker or lighter than the surrounding area or retains more or less heat than the surrounding area, the snow on the target will melt at a different rate than the area around it potentially resulting in an unrepresentative snow depth measurement. Qualitative analysis of the plastic target at Caribou Creek (Figure 9) suggests that the melt characteristics of this target are quite similar to the surrounding landscape (grass, logging debris, and exposed sand). Snow neither melts from the target first nor last as compared to the surrounding area. The artificial turf targets at Sodankylä, however, are substantially different than the surrounding landscape (which is predominately exposed sand) and produces faster melt (Figure 10). The artificial turf targets are likely more representative when surrounded by grass rather than sand.

5. Sensor Diagnostics

Some snow depth sensors are capable of outputting a sensor diagnostic that may be used to improve (or at least assess) the measurement quality or even increase the utility of the sensor. During SPICE, we examined the potential causes of (sensor determined) reduced signal quality metrics for measurements made by the SR50A and looked at the relationship between what the sensor outputs as a reduced quality measurement and what the SPICE quality control process determines is a suspicious data value. We also look at using signal strength output from the SHM30 to increase the sensor's utility to determine the presence of snow on the target area. Since this analysis is the topic of another paper in these proceedings, it is only briefly summarized here.

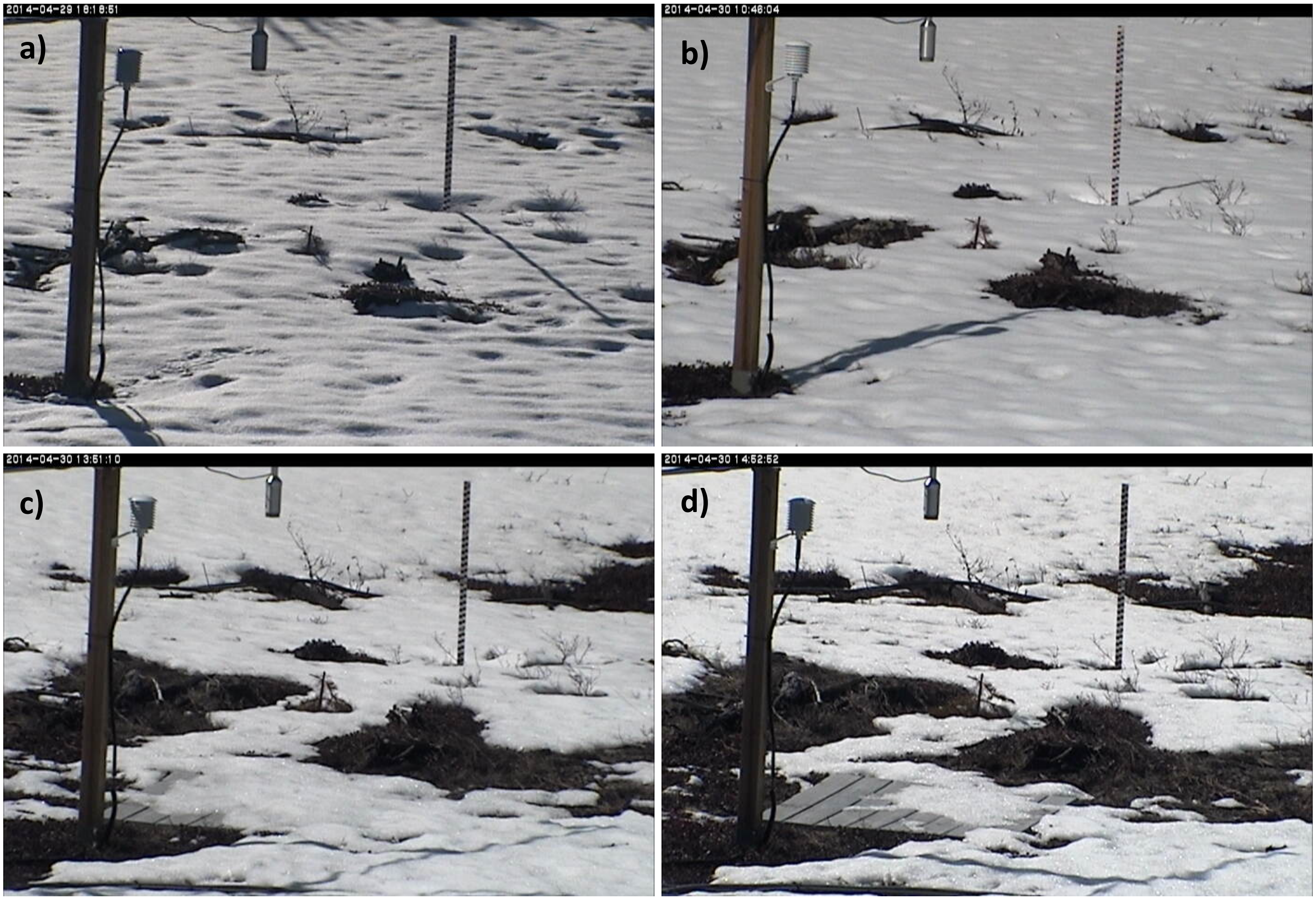

Figure 9: Time lapse photographs of the snow melt around the artificial plastic target at the Caribou Creek SPICE site a) 29 April 2014 18:18 UTC, b) 30 April 2014 10:48 UTC, c) 30 April 2014 13:51 UTC, d) 30 April 2014 14:52 UTC.

Figure 10: Bare targets in the Sodankylä intercomparison field taken on 11 October 2014 11:04 UTC, 5 hours after the photos shown in Figure 8.

5.1 SR50A Quality Numbers

Detailed analysis was carried out to cross-reference the occurrence of SR50A "Reduced Echo Signal Strength" and "High Measurement Uncertainty" quality number outputs with concurrent site conditions at Formigal, Col de Porte, CARE, and Sodankylä. Site conditions at each of these locations were quite different, including the amount of accumulated snow expected during the season. Because of expected deep snow packs at the Formigal and Col de Porte alpine sites, the sensors themselves required mounting at much greater heights than at CARE and Sodankylä, which may in turn exacerbate some of the impacts on quality numbers. In general, a decrease in the percentage of "Good" quality numbers results in an increase in the frequency of "Reduced Echo Strength" numbers rather than "High Uncertainty" or "Unable to Read" values.

At both Formigal and Col de Porte, the sensors are mounted at a height of 4m. Overall, the percentage of "Good" quality numbers at these sites was lower than that observed at the other sites with only 71% and 87% of all SR50A measurements at Formigal and Col de Porte respectively being declared "Good" by the sensor. At Formigal, measurement uncertainty was highest during the months of April and May and seems to be associated with the rapid melt of the snow pack, and perhaps the deterioration of the target surface. This did not occur at the other sites. The instruments at Formigal and Col de Porte, perhaps because of their installation heights, were sensitive to the occurrence of snowfall. The percentage of "Good" data dropped from 71% (87%) with no precipitation to 56% (75%) during precipitation at Formigal (Col de Porte). When snow depths increased by more than 2 cm in a 30 minute period, the percentage of "Good" data dropped to less than 41% (48%) at Formigal (Col de Porte).

The frequency of "Good" measurements at both CARE and Sodankylä were significantly higher than at the alpine sites, likely because the sensor installation height is only 2 m and both used highly reflective surface targets. The impact of precipitation on the quality number frequency distribution was small, even at higher precipitation rates but the impact of precipitation was higher at Sodankylä than it is for CARE. It was hypothesized that higher wind speeds at CARE may have an increased impact on the quality numbers but it appears that this is not the case.

As a follow up to the analysis summarized above, the occurrence of "Reduced Certainty" and "High Uncertainty" quality numbers was compared to the occurrence of data flagged as "Suspicious" by the SPICE quality control process. A data point flagged as "Suspicious" occurs when the automated QC process determines a data point to be out of range of reasonable snow depth values for the site, or if the point is identified manually as being "Suspicious" and flagged accordingly. It was determined that a large percentage of data (> 72%) was flagged as good by both the sensor and the QC process and only a small percentage (< 1%) was flagged as bad by both. At Formigal, 27% of the data had “Reduced Echo Strength” or “High Uncertainty” quality numbers but were not flagged as “Suspicious” by the QC process. This percentage was less than 3% at CARE. This represents data that would have been removed if the user only relied on the quality numbers for quality control. The percentage of data that were flagged as “Suspicious” by the QC process but considered of “Good” quality by the sensor was quite low at less than 1%. This is the portion of bad data that would be included in the data set if data QC was based on quality numbers alone.

When only considering data with a “Good” quality number, more than 99% of this data is also flagged as “Good” by the SPICE QC process. Most of the data (> 98%) with “Reduced Echo Strength” quality numbers are also flagged as “Good” by the SPICE QC process. Finally, over 88% of data with “High Uncertainty” quality numbers are also flagged as “Good” by the SPICE QC process. This analysis suggests that users should not base their data quality control on the sensor quality numbers alone as this would lead to the rejection of more data than necessary. Rather, users should implement their own QC process based on automated range checking and visual inspection, perhaps only using the sensor quality metric for guidance or trouble shooting.

5.2 SHM30 Signal Strength Output

The SHM30, an optical range finding instrument manufactured by Lufft/Jenoptik, outputs a signal strength value for each measurement that represents the strength of the return optical beam reflected from the target. It was illustrated by de Haij (2011) how the signal strength output from the sensor could be used to determine when snow is present in the target area, even if the depth of the snow is not great enough to be measurable by the sensor. This was explored further in SPICE over various surface target types.

The SHM30 sensors for SPICE are installed at Col de Porte (natural ground/mown grass), Weissfluhjoch (natural ground/rocks), Sodankylä (artificial turf) and CARE (grey plastic). According to the sensor user’s guide and the results reported by de Haij (2011), we expected to see the different optical properties of the targets have an impact on the signal strength return of the sensor and therefore impact the extended capabilities of the sensor to detect light snow on the target area.

It was found during SPICE that the artificial turf targets worked remarkably well, producing a very distinct change in the sensor’s signal strength output with light snow on the target. Following this, there was also a distinct change in signal strength when the target became snow free. A similar reaction was found over the natural targets, especially if the target was mown grass. However, the change in signal strength was not as distinct as it was on the artificial turf. The reaction of the sensor to snow on the grey plastic targets was quite different. The baseline (no snow) signal strength output was substantially higher than over the darker turf targets and the change with snow cover was much more erratic and less distinct.

Based on these results, if the detection of first and light snowfalls and the timing of zero snow cover are important to the sensor user, it is suggested that they use a target (whether natural or artificial) with optical properties similar to natural grass or the artificial turf that was tested during SPICE.

6. Relationships Between Change in Snow Depth and Gauge Precipitation

One of the SPICE SoG objectives was to test the relationship between total precipitation amount as measured by the SPICE reference precipitation gauge (the R2 or equivalent) and the change in snow depth as measured by an automated sensor. The intent was to test the capability of using a snow depth sensor to derive an estimate of total precipitation (snowfall) in the absence of a precipitation gauge capable of measuring the liquid water equivalent of snow. Of course, there are several challenges in converting measured snow depth to precipitation. Besides the large variation in the density of new snow, snow depth is impacted by compaction and densification, wind redistribution, sublimation and melt. These processes, which are usually site dependent, generally contest the 10:1 “rule of thumb” for converting snow depth to precipitation. Data from several SPICE sites are used to test the feasibility of using a snow depth sensor to estimate precipitation.

The relationship between R2 reference gauge measured precipitation and the change in snow depth was determined for the CARE and Sodankylä SPICE sites. Hourly gauge precipitation amount was extracted from the SPICE 60 minute SEDS (site event data set) and the corresponding change in snow depth was calculated using the SHM30 snow depth measurements. Precipitation type was derived from 1-minute disdrometer data such that rain and mixed events could be eliminated.

Figure 11 shows the scatter plot and regression line for CARE (top) and Sodankylä (bottom). The r^2 for the CARE relationship is 0.28 while the r^2 for Sodankylä is 0.58. Both relationships show significant scatter but the CARE relationship is weaker, likely because of a higher amount of wind redistribution of snow during precipitation.

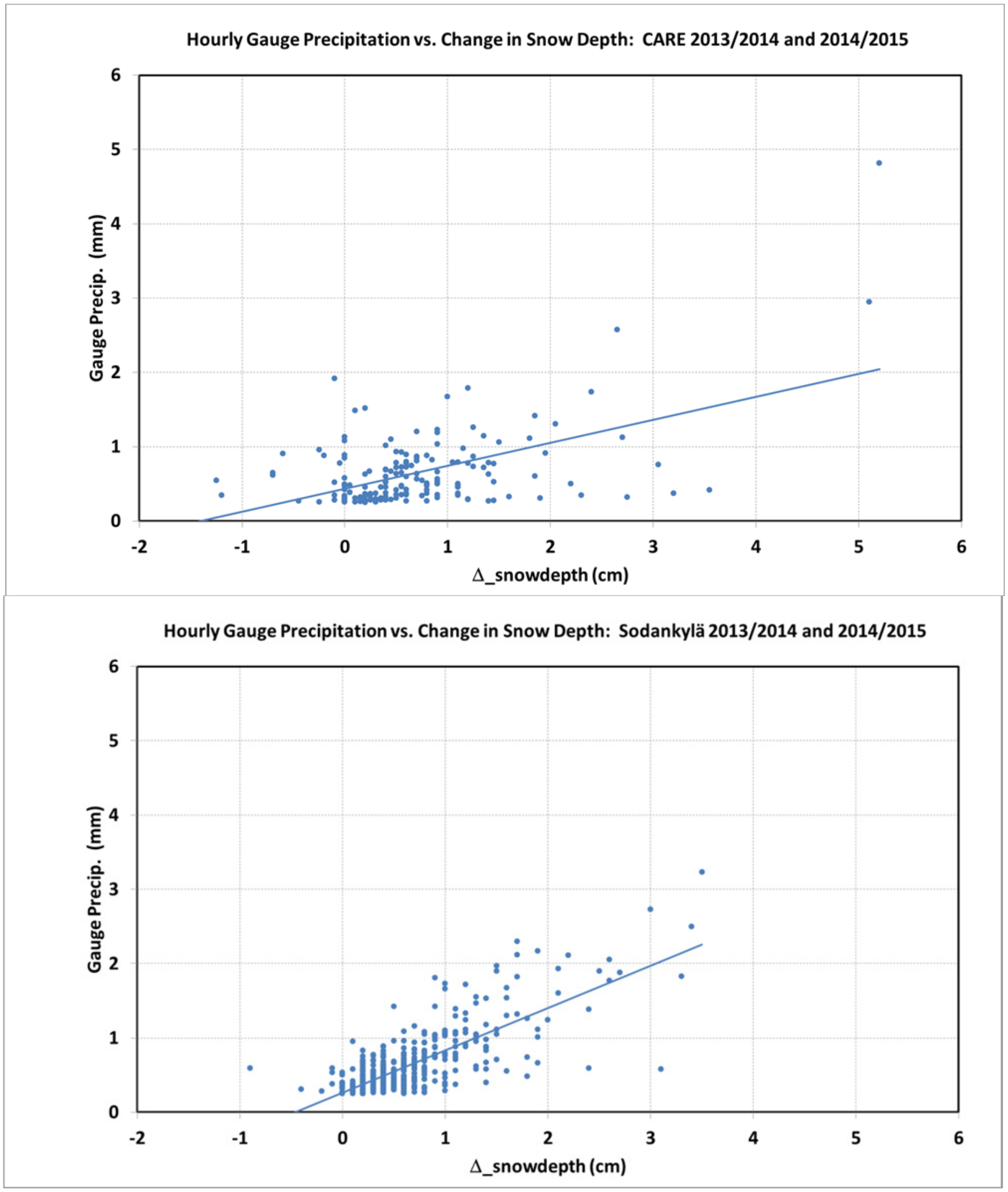

Figure 11: Relationship between hourly precipitation as measured by the reference (R2) gauge and the change in snow depth at CARE (top) and Sodankylä (bottom) for the 2013/2014 and 2014/2015 winter seasons.

The high amount of scatter in the relationships between gauge precipitation and change in snow depth at CARE and Sodankylä demonstrates that these relationships are complex and highly variable from site to site. These results suggest that is very difficult to infer precipitation amounts with any level of accuracy using the change in snow depth as determined by an automated snow depth sensor.

7. Summary and Conclusions

This paper provides a brief summary of some significant results of the WMO SPICE SoG work. The results presented here, as well as other SoG results, will be available in the SPICE final report.

Intercomparison of the various snow depth sensors with both manual and automated references generally shows that most sensors are measuring according to manufacturer specifications in an accurate and reliable manner and that most of the deviation between the sensor and the reference measurements can be attributed to spatial variability in snow depth at the sites. Differences in the measurement principle of the SWE sensors under test became apparent during the intercomparison and resulted in a diverse intercomparison, potentially highly dependent on site condtions. This is discussed in greater detail in Smith et al. (2016).

A qualitative analysis of surface target configurations for the measurement of snow depth showed advantages and disadvantages of each target type, the choice of which is largely dependent on the user's situation. The use of an artificial target appears to be more of an advantage for sonic snow depth sensors and less relevant for optical laser snow depth sensors, although the target's optical properties impacts the capability of using the laser sensor's signal strength data to identify the occurrence of light snow or when the target is snow free.

A relatively thorough review of the use of sonic sensor quality number output showed that the frequency distribution of the quality numbers varied substantially by site which is likely related to the height of the sensor above the target. Sensor quality output from sensors installed at greater heights (Formigal and Col de Porte) appear to be impacted more by the occurrence of precipitation with uncertainty (as estimated by the sensor) increasing with precipitation rate. This is less significant for sensors installed closer to the surface (CARE and Sodankylä). Cross-referencing the sensor outputted quality numbers with the SPICE quality control flags showed that there was very little overlap between the data flagged by the sensor as "Reduced Echo Strength" or "High Uncertainty" with data flagged as "Suspicious" by the QC process. This means that there is reduced utility in using the sensor quality numbers in the QC process and perhaps more guidance from the manufacturers are required for interpreting this information.

Finally, analysis was undertaken to relate changes in snow depth, as measured hourly by automated snow depth sensors, with corresponding precipitation measured by the SPICE reference gauges. As expected, relationships showed much scatter such that it would be difficult to estimate precipitation using changes in snow depth alone, especially at hourly time scales.

References

de Haij, Marijn: Field test of the Jenoptik SHM30 laser snow depth sensor, Technical report TR-325, Koninklijk Nederlands Meteorologisch Instituut, 2011.

Nitu, R., Rasmussen, R., Baker, B., Lanzinger, E., Joe, P., Yang, D., Smith, C., Roulet, Y., Goodison, B., Liang, H., Sabatini, F., Kochendorfer, J., Wolff, M., Hendrikx, J., Vuerich, E., Lanza, L., Aulamo, O., and V. Vuglinsky: WMO intercomparison of instruments and methods for the measurement of solid precipitation and snow on the ground: organization of the experiment, WMO Technical Conference on meteorological and environmental instruments and methods of observations, pages 16–18, Brussels, Belgium, 2012.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer A.P., Black, J., Thériault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed, Bull. Amer. Meteor. Soc., 93, 811–829, 2012.

Smith, C. D., Kontu, A., Laffin, R., and Pomeroy, J.W.: An assessment of two automated snow water equivalent instruments during the WMO Solid Precipitation Intercomparison Experiment, The Cryosphere Discuss., doi:10.5194/tc-2016-57, under review, 2016.