

Intercomparison of measurements of bulk snow density and water equivalent of snow cover with snow core samplers: instrumental bias and variability induced by observers

Ignacio Lopez-Moreno¹, Leena Leppänen², Bartłomiej Luks³, Ladislav Holko⁴, Ghislain Picard⁵, Alba Sanmiguel-Vallelado⁶, Esteban Alonso-González⁷, David C. Finger⁸, Ali Nadir Arslan², Katalin Gillemot⁹, Aynur Sensoy¹⁰, Arda Sorman¹⁰, Mustafa Cansaran Ertas¹⁰, Charles Fierz¹¹, Christoph Marty¹¹, and Steven Fassnacht¹²

¹Pyrenean Institute of Ecology

²Finnish Meteorological Institute

³Polska Akademia Nauk Instytut Geofizyki

⁴Slovak Academy of Sciences

⁵University of Grenoble Graduate School of Environmental and Organisational Systems Science and Engineering

⁶Instituto Pirenaico de Ecología

⁷Instituto Pirenaico de Ecología

⁸Sustainability Institute and Forum (SIF), School of Science and Engineering, Reykjavik University, Iceland

⁹University of Vienna

¹⁰Eskişehir Technical University

¹¹Federal Institute for Snow and Avalanche Research

¹²Colorado State University

May 5, 2020

Abstract

Manually collected snow data are often considered as ground truth for many applications such as climatological or hydrological studies. However, there are many sources of uncertainty that are not quantified in detail. For the determination of water equivalent of snow cover (SWE), different snow core samplers and scales are used, but they are all based on the same measurement principle. We conducted two field campaigns with 9 samplers commonly used in observational measurements and research in Europe and northern America to better quantify uncertainties when measuring depth, density and SWE with core samplers. During the first campaign, as a first approach to distinguish snow variability measured at the plot and at the point scale, repeated measurements were taken along two 20 m long snow pits. The results revealed a much higher variability of SWE at the plot scale (resulting from both natural variability and instrumental bias) compared to repeated measurements at the same spot (resulting mostly from error induced by observers or very small scale variability of snow depth). The exceptionally homogeneous snowpack found in the second campaign permitted to almost neglect the natural variability of the snowpack properties and focus on the separation between instrumental bias and error induced by observers. Under such measurement conditions, the uncertainty in bulk snow density estimation is about 5% for an individual instrument and is close to 10% among different instruments. Results confirmed that instrumental bias exceeded both the natural variability and the error induced by observers, even in the case when observers were not familiar with a given snow core sampler.

1 INTRODUCTION

Water equivalent of snow cover (SWE) is one of the most commonly measured snowpack properties (Pirazzini et al., 2018). Hydrological studies need to quantify this variable since it is mass of snow that can be released as water during the melt season. Moreover, SWE is commonly used in long-term climatological or hydrological studies (Marty, 2017; Helmert et al., 2018), and as data for calibration, evaluation, or assimilation of remote sensing and numerical models (Dong, 2018). Furthermore, long-term measurements are used for developing statistical models (e.g. Jonas et al., 2009) that can be used in combination with LiDAR measurements and other depth measurement over space (Schöber et al., 2014) or in the framework of snow data assimilation (Magnusson et al., 2014).

SWE can either be measured automatically by pressure sensors (Johnson, 2004), or retrieved by various emerging techniques like Global Navigation Satellite System antennas (Guttmann et al., 2012; Koch et al., 2014), radio frequency transmission lines (Stähli et al., 2004), passive gamma ray sensors (Choquette et al., 2013), cosmic-ray neutron sensors (Schattan et al., 2017) or ground penetrating radar (Schmid et al., 2015). Moreover, microwave remote instruments, both active and passive, proved their capability to estimate information on SWE (Takala et al., 2011), although there are clear limitations in mountain regions.

Despite the existence of a wide array of geophysical and remote sensing methods to measure bulk snow density and SWE at different spatial scales (Kinar & Pomeroy, 2015), it is still very important to perform classical manual measurements, which indeed are generally used to validate automatic sensors. Bulk snow density and SWE in a snow column can be retrieved in two ways, either by weighting snow cores extracted with a snow core sampler or by sampling snow density along the wall of a snow pit with dedicated density cutter at either regular depth intervals or layer by layer. According to a recent survey by Pirazzini et al. (2018) snow core samplers are still the most widely used devices, due to the simplicity and speed of the snow data acquisition (at least for shallow snowpacks), both in research and operational networks.

Since Church (1933) popularized the use of a snow core sampler, many surveyors have developed a variety of models that mainly differ in length and diameter, material (basically metal or plastic), weighting system (snow and sampler combined, or only the snow in a separate bag or container), and the type of the tooth-cutter applied. The specificities of each snow core sampler aim to adapt to the most common characteristics of the snowpack in a given area (mostly in terms of snow depth, hardness and presence of ice layers) or aim to facilitate its portability in the field, when relevant.

It is well known that measurements of SWE and density from snow core samplers present a number of problems that affect the accuracy of the collected data (Stuefer et al., 2013). The most common problem is the loss of part of the sample due to snowpack collapse, when the sampler encounters hard layers. Snow can also be lost when the sampler is extracted from the snowpack for weighting (Dixon & Boon, 2012). This is the reason why many surveyors only consider the samples that collect at a significant percentage of the measured snow depth to be valid. Another problem is accumulation of compacted snow at the mouth of the sampler impeding the penetration further to sample more snow. The presence of ice layers may lead to the false perception that the ground is reached, also leading to snowpack undersampling. Indeed, when density and SWE measurements from snow core samplers are compared to those from density cutters in snow pits, an underestimation of the former is generally found (Fassnacht et al., 2010; Proksch et al., 2016). Other authors have highlighted other sources of errors, including the human factor (Stuefer et al., 2013) and errors in the weighting process due to a lack in accuracy of the scales used, or the effect of wind on hanging scales (Goodison et al., 1987; Doesken & Judson, 1996). Several previous studies have attempted to quantify the error of different models of snow core samplers, considering data from snow pits or the mass of a volume of snow as ground truth, or by directly analyzing the variation between different samplers (Work et al., 1965; Peterson & Brown, 1975; Farnes et al., 1982; Dixon & Boon, 2012). Most of the results have identified a different behavior between snow core samplers under different snow conditions, and most of the reported

errors ranged between 3 % and 13% (Beaumont, 1967; Work et al., 1965; Farnes et al., 1982; Berezovskaya & Kane 2007; Dixon & Boon, 2012). These analyses are important as *in-situ* SWE measurements are often used as ground truth for the evaluation of both hydrological models (Pirazzini et al., 2018) and remote sensing products. While analyzing SWE or density data sets it is important to remember that different institutions and observers may use different snow core samplers (even changing them over time), which may introduce significant spatial and temporal inhomogeneities in the data (Beaumont & Work, 1963; Haberkorn et al., 2019).

In the framework of the COST Action ES1404 HarmoSnow (A European network for a harmonized monitoring of snow for the benefit of climate change scenarios, hydrology and numerical weather prediction, <http://www.HarmoSnow.eu/>), field campaigns were conducted to compare widely used snow core samplers under different snow conditions (depth, density). The objectives were to 1) estimate the range of differences in measurements obtained from various samplers, and 2) possibly separate the effect of the natural variability of bulk snow density and SWE from the instrumental error, and the error introduced by the different observers.

2 FIELD CAMPAIGNS, INSTRUMENTS AND METHODS

A preliminary campaign to demonstrate instruments and error sources took place in March 2016 on two plots located midway between Güzelyayla and Senyurt villages (40.195°N, 41.48°E, 2071 and 2210 m a.s.l., respectively), close to the city of Erzurum (39.9°N, 41.3°E) in Eastern Turkey. It was primarily intended to increase the awareness of local water managers and the Action participants about possible sources of errors and questions of spatial variability when performing *in-situ* snow measurements. The campaign thereby provided a first, qualitative comparison of seven snow core samplers (the results are not shown in this study) and allowed for the identification of key points for the design of the subsequent quantitative comparison campaigns.

Two such field campaigns were then carried out: the first one in March 2017 in the Bláfjöll area near Reykjavik (F1, Iceland,) and the second one in February 2018 in Sodankylä (F2, Finland). The sampling strategies followed during each campaign are presented in Figure 1. All the measurement sites had a seemingly homogeneous snowpack in terms of depth and density, small sized sampling plots with very smooth topography (slope < 5° and low roughness) and a shallow to moderate snow depth (<80 cm).

The compared snow core samplers (Figure 2) are used for operational monitoring and research in several European countries that participated in HarmoSnow (Finland, Estonia, Lithuania, Poland, Czech Republic, Slovakia, Switzerland, Italy, Turkey, Spain, Russia and other countries of former Soviet Union; Haberkorn, 2019) as well as in North America (USA and Canada). Each instrument was always operated by a person, who has a long of experience with the device. The only exception was the experiment on the potential observer bias during F2, when each instrument was operated by a so-called untrained person. These persons were the same as for the other experiments, but just switched their instrument with one from another country, which they were not familiar with. Table 1 provides a summary of the nine different instruments used in the campaigns and their main characteristics. Links to manufacturers and more detailed descriptions are provided in Table S2 in the supporting material. There is a wide variety of snow core samplers covering different diameters, lengths, materials and weighting systems (Figure 2 and Table 1). During the field campaigns a common protocol was applied by all observers. It was based on recommendations currently listed in the preliminary 2018 edition of the WMO n°8 CIMO guide (<https://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html>). Snow core samplers are inserted vertically from the top surface to the snowpack. Samplers with teeth need to be twisted, the ones with sharpened rims only need to be twisted in really dense snow. Depending on the design, samples can be excavated with (cylinder samplers) or without digging a snow pit (tube samplers). When the snowpack is deeper than the height of the cylinder, measurements need to be repeated until the soil surface is reached. To clearly separate one measurement level from the next, the application of a thin plate is favorable. However, such a plate was neither used with all cylinders nor for all measurements in the campaigns. In both cases, depending on the weighing system, either weight is measured or a direct reading of SWE is conducted. If the depth of the core in the tube was less than 80 % of the depth of the undisturbed

snow, the core had likely spilled from the tube or was not captured by the tooth-cutter. In this case, the observer had to repeat the measurement. However, snow conditions during the field campaigns were rather favorable and almost no need of repeating measurements was necessary. Snow sample weight and volume data, combined with the snow depth, result in bulk snow density and SWE values for specific measurements.

In the case of Sodankylä, the Mann-Whitney-Wilcoxon test (Wilcoxon, 1945) was used to determine if the differences in the density measurements performed by the different instruments are statistically significant ($p < 0.05$).

2.1 Field campaign in Bláifjöll, Iceland (March 2017, F1 in Fig. 1)

The first field campaign was conducted on March 1st 2017 on two plots located approximately 25 km south-east from Reykjavik in the foothills of the Hengill volcano. The two plots were in close proximity to one another but they differed in the nature of the ground surface. Plot 1 was located in an open area, approximately 200 m north of the Hellisheii Geothermal power plant (64.04°N, 21.40°W, 260 m a.s.l.). Measurements on Plot 1 were taken at a paved parking lot of the power plant. The ground was almost completely level and covered with a mix of grass and asphalt. Plot 2 was located approximately 190 m to north of Plot 1, at a lava plateau, with an irregular ground surface covered by soft moss. With such setting, uncertainty related to sampling on uneven terrain was largely reduced on Plot 1, while Plot 2 was considered as a test polygon with challenging, very high ground surface irregularity. On both plots the snowpack was cold and dry, with an average snow depth of 48 cm and 53 cm for Plot 1 and 2, respectively. Windblown surface features and ice layers resulting from rain-on-snow events were identified within the snowpack structure.

Bulk snow density or SWE measurements, depending on the type of instrument, were made with the nine different core samplers listed in Table 1. At the same time, snow depth measurements were taken at each sampling point with the snow tubes or probes (in the case of EV2 and Custom EV2). The sampling strategy was to measure along a 20 m long snow trench. Depending on the duration of a single measurement, three to six measurements were taken at each spot with one instrument, with each instrument sampling at two to three spots along the trench (Figure 1). Repeated measurements at each spot were made as close as possible to each other ([?]5-10 cm) to minimize errors resulting from natural variability in snowpack properties. Thus it was possible to associate the differences among repetitions to measurement uncertainty. This was especially true at Plot 1 where the ground surface was smooth and level. On the other hand, differences between the measurement spots (Figure 1) were due to a combination of the instrumental bias, observer errors and natural differences in snowpack characteristics along the 20 m long trench. The overall aim of the campaign was to compare the variability among repeated measurements at the same spot using one instrument to the variability observed along the trench as an attempt to separate the natural variability of the snowpack from the variability amongst the measurements.

2.2 Field campaign in Sodankylä, Finland (February 2018, F2 in Fig. 1)

The second field campaign was conducted from the 20th to 22nd of February 2018 at Sodankylä (67.37degN, 26.63degE, 175 m a.s.l.) in Lapland, Finland making use of the installation operated by the Finnish Meteorological Institute (INTERACT, 2015; Leppanen et al., 2018). The measurements were conducted during three consecutive days in a Bog plot (Bog site; Leppanen et al., 2016) and a Forest opening (Intensive Observation Area; Leppanen et al., 2016) following scheme F2A, as well as in a wide open plot (hereinafter called the Antenna plot; Sodankylä SPICE site; Nitu et al., 2018) following scheme F2 (Figure 1). As a consequence of persistent cold conditions during the whole winter, the snowpack was dry (wetness index 1; Fierz et al., 2009) and rather soft (hardness index mostly 1-3) consisting mostly of faceted crystals and depth hoar with grain size larger than 1 mm, showing very homogeneous characteristics on all three plots. All three plots are flat and the snow depth measured with probes or SWE samplers did not vary by more than 11 % ($CV < 0.11$) at any of the three plots. Average snow depths were 53.2, 71.1 and 62.7 cm at bog, forest and antenna plots respectively. The ground was frozen, facilitating the identification of the contact point between snow and soil. Low vegetation (lichen, moss, heather), height of approximately 5-15 cm on average, was present at the Forest and Antenna plots, whereas there was only isolated grass on a mostly icy ground at the Bog

plot. The aim of this campaign was to systematically distinguish the instrument based error from both the observer induced error and the natural variability of the snowpack. The sampling strategy was to divide a plot (ca. 10x20 m) into four subplots, where each of the nine snow core samplers collected five replicates. In one of the subplots at the Antenna plot, each instrument was used to measure snow density or SWE and snow depth along a transect of 10 m with a spacing of approximately 0.5-1 m between measurements where the transects were 0.5-1.5 m apart (Figure 1). In addition, at both sides of the Bog and Forest plots three measurements were taken with each instrument by observers who were not familiar with its use (Figure 1). We attempted to assess the effect of the expertise with using a specific instrument to the reliability of the measured data. Finally, stratigraphic records of the snowpack were taken in one to two snow pits at each plot (see Figures S1 to S3 and Table S1 in the supporting material). The local procedure is presented in Leppanen et al. (2016).

A major advantage of this campaign was the availability of the SnowMicroPen (SMP), which measures the penetration resistance of snow. The relative uncertainty among repeated, objective penetration resistance measurements is low and thus a good measure for the spatial variability of the snowpack is obtained (see e.g. Kronholm et al., 2004). This is further discussed in the supporting material and shown in Figures S1 to S3. Thorough analysis of such measurements can also be used to obtain snow density at the millimeter scale (see Equation 9 in Proksch et al., 2015). However, this calibration does not hold for the version of our SMP (see also Proksch et al., 2016), yielding bulk snow densities biased towards higher values. A total of 99 SMP measurements, more than 26 on each plot, were taken in undisturbed snow after all SWE measurements were completed on each subplot.

Finally, the potential role of the weighing process in the error estimation of snow density and SWE measurements was tested in two ways. Nine different weight scales were compared to check how the low air temperature may affect the accuracy of weighing. We used reference weights of 50, 200, 500, 1000 and 2000 g, performing measurements indoors (at approximately +20 oC) as well as outdoors at -26 oC after waiting for 20 minutes for the scales to cool.

3 RESULTS

3.1 Field campaign in Blafjoll

The depth, bulk density and SWE measured with six different snow core samplers (SH, VS-43, Dolfi, K-M, IG PAS and Federal) along two 20 m long snow trenches (Figure 1, scheme F1) are shown in Figure 3. In this field campaign the relative position of each instrument was annotated and the figure reproduces the order in which different devices were used. Important variability can be seen in the measurements when the different samplers are compared, but also a non-negligible variability between repetitions with each instrument at some specific spots.

For Plot 1, characterized by a very homogeneous ground surface, the coefficient of variations (CV, the ratio of the standard deviation to the mean) for snow depth and density between different spots were 0.08 and 0.11, respectively. The combination of both lead to a CV for SWE among the 14 measured spots equal to 0.22. The repeated depth measurements at each spot were almost identical (with maximum differences of 3% in some spots). In general, bulk snow density at each spot measured with the same device showed a variability of less than 5%, which was exceeded on only four out of 14 spots: twice with the Federal sampler, once with the K-M and once with the VS-43. Variability in repeated measurements of SWE was very similar to the reported variability of the bulk snow density.

Plot 2 exhibited a larger spatial variability for snow depth between spots at the plot scale, with a CV of 0.12, and also a high variability between replications at each spot (CV exceeding 0.05 in eight out of 12 spots). On this plot, however, the variability of bulk snow density between the spots was smaller compared to snow depth with a CV close to 0.1, as well as the differences among several repeated measurements at the same spot (CV never exceeded 0.1). The combination of depth and density variability leads to a CV for SWE of 0.14 among the 12 spots.

In this campaign the SH sampler provided systematically lower values for bulk snow density than most other devices. Nevertheless, the CV for bulk snow density obtained by SH at individual spots was comparable to the values obtained for the other devices. The Federal sampler provided high variability among repeated measurements in the same spot when used in Plot 1. There was no clear systematic bias induced by the instrument or the observers for the remaining samplers.

3.2 Results from the field campaign in Sodankylä

The variability of bulk snow density estimated from SMP measurements (27, 28 and 44 measurements for bog, forest and antenna, respectively) using the Proksch et al. (2015) parametrization for the three plots is shown in Figure 4. While the mean and median values are too high compared to the measurements (see Figure 5 and Section 2.2), these estimates reveal an almost negligible variability in snow density (CV not exceeding 0.01 on any of the three plots). The homogeneity of the snowpack on each plot is further supported by data of the median penetration force with a resolution of 1 mm, and in the snow pits that were dug at each of the analyzed plots (Figures S1 to S3 in the supporting material). This implies that the differences between estimates from each SWE sampler or the variability between replicates for each device are primarily due to either instrumental errors or errors induced by the observer.

Figure 5 shows the variability of all replicates for snow depth, bulk snow density and SWE measurements (Figure 1b) obtained from the nine different instruments (Table 1) at each plot. The 20 measurements (15 measurements at the Antenna plot) are made up of five close replicates on each of the subplots. In this campaign the snow-ground interface was very easy to identify during sampling, and thus the CV for snow depth measurements was very low (<0.06), being almost negligible for all observers. More significant were the observed differences between the bulk snow densities measured by the different snow core samplers. Variability between the repeated measurements was much higher than that observed for the SMP but still relatively low (CV of 0.07 to 0.10), especially compared to the field campaign in Iceland (CV ranging from 0.04 to 0.12; see section 3.1). The instruments that yield the highest variability between replicates were EV2 (in particular the customized model EV2-C), Federal, and IG PAS. The entire snowpack is sampled at once with EV2 and Federal, i.e. no digging is needed; SWE may have been underestimated due to the loss of snow from the bottom of the tube after its removal from the snowpack, due typically lacking soil plug originating from frozen soil and ice on top of the ground. This may have negatively influenced the consistency of their measurements. Measurements with SH were made after digging because its larger diameter impeded the retrieval of the snow sample directly from the top of the snow surface. On the other hand, the resolution of the scale used with IG PAS was low (50 g), which contributed substantially to the relatively high variability for the bulk snow density observations. The length of the IG PAS, which is only 50 cm, also caused problems with snow depth exceeding it. In such a case, two measurements were taken. The behavior of each sampler with regards to the plot average of all instruments is very similar on all three plots, except for the EV2-C (Figure 5). The median value for ETH, SH, and K-M lie consistently close to the plot median on all plots. In addition, SH and K-M as well as VS-43 and Dolfi provided pairwise very close values for bulk snow density on all three plots. Table 2 shows the statistically significant differences in measured bulk snow density between pairs of samplers based on the Wilcoxon rank test. Results show that K-M, SH and ETH tend to provide similar data, as do VS-43 and Dolfi, while bulk snow density measurements with the Federal were significantly different than all other samplers on all three plots. Because the CV for bulk snow density was generally low, the variability in SWE among the replicates made by the same device at the same subplot is similar to that of snow depth. Accordingly, the differences between SWE values measured with different snow core samplers were very similar to those for the bulk snow density (Figure 5).

The variability of bulk snow density obtained from close replicates by individual devices at each subplot is presented in Figure 6. Most devices show similar results for the three plots, except for EV2, EV2-C, Federal sampler and IG PAS. The patterns in the graphs are very similar for each plot, which suggests that each device has an intrinsic systematic bias. This bias was within $\pm 10\%$ except for EV2 and EV2-C in a few cases.

The snow density measurements for each snow core sampler when used by the experienced and untrained

observers are shown in Figure 7. Except for a few cases (i.e., EV2), the variability between replications was lower when the device was used by the experienced users. However, the differences were statistically significant (from the Mann-Whitney-Wilcoxon test) for only a few devices, i.e. for the SH in bog, and the Federal sampler in the forest. Inconsistency in the differences in measured average densities between the Bog and Forest plots was observed for two snow samplers. For example, the ETH snow core sampler used by untrained users provided lower average snow density at the Bog plot, but higher density at the Forest plot. This is probably due to the fact that the untrained users for each instrument were not the same at both plots. Clear differences in snow density between the two plots were also found for the IG PAS and Federal sampler.

The measured snow depth and bulk snow density at the Antenna plot following a 10 m transect sampled at 0.5-1 m intervals (Fig. 1) for each instrument are presented in Figure 8. In general, snow depth varies little along the lines with almost identical values for the 10 replicates by each observer (Figure 8b). An analysis of bulk snow density revealed that the 10 measurements performed with any instrument have a low, very similar variability, with a mean CV of 0.04 per instrument (SMP 0.02). However, notable differences were observed among the different instruments, with the mean for each instrument varying from 192 kg m⁻³ to 233 kg m⁻³ with a CV of 0.08 (Figure 8a).

4 DISCUSSION

Two field campaigns conducted in the framework of the COST Action HarmoSnow (ES1404) provided a unique opportunity to compare the snow samplers widely used in operational networks and research across Europe and North America. The results have been used to illustrate the spatial variability of snow depth, bulk snow density and SWE at small spatial scales, and to assess the extent that the use of different snow core samplers and the error involved in the measurement procedure may affect density and SWE measurements. Collected data have enabled the distinction of the three main sources of uncertainty when measuring snow density and SWE at the local scale: i) natural variability of snowpack at small spatial scales; ii) error induced in the measurement process; and iii) instrumental bias when different types of snow core samplers are used at the same time and place.

A snowpack may exhibit differences in density at very small spatial scales (Komarov et al., 2019). This fact partially explains the high variability of density (CV of 0.10 and 0.11 for plots 1 and 2 respectively) and SWE (CV of 0.14 and 0.22 for plots 1 and 2 respectively) among measurement points found in Iceland. This variability far exceeds the differences among repeated measurements at each spot that was roughly half of that between the measured spots. The CV for density still ranged between 0.03 and 0.15 for the majority of the measurements due to irregular wind crusts and ice layers. When snow is measured on a homogeneous surface like at Plot 1 (smooth parking lot and lawn) the snow depth measurement has minimal impact on uncertainty of SWE estimation. The opposite occurred at Plot 2, where snow depth was largest source of uncertainty in SWE estimation. This can be explained by uneven rocky and moss covered ground, and does confirm previous studies highlighting that snow depth measurement may be an important source of uncertainty in SWE estimation for various environments, namely when ground is covered by shrubs or unfrozen bog areas and snowpack is shallow (Sturm et al., 2010; Lopez-Moreno et al., 2013; Stuefer et al., 2013).

Homogeneous ground (frozen grass or bog) and snowpack found in Finland, provided an excellent opportunity to separate the effect of natural variability of the snow properties from the instrumental and observer induced errors. Homogeneity of snow was confirmed by 99 SMP measurements with a very low spatial variability in snow penetration resistance and density (CV of variation lower than 1% at each study plot). The low cohesion of snow during this campaign was challenging for some snow core samplers due to partially losing snow during their removal from the snowpack, preventing direct retrievals of core samples from the snowpack surface. This was the case for the SnowHydro, IG PAS, and VS-43 samplers (Table 1). This problem was avoided by digging a pit and inserting a spatula below the sampler at base of the snow-soil surface. Density measurements taken by different snow core samplers exhibited a CV between 0.02 and 0.06, while SMP revealed a CV of only 0.01. This confirms the existence of instrumental bias and error induced by an

observer that cannot be attributed to the natural variability of the snowpack. Indeed paired comparison of density data collected with different samplers exhibited a statistically significant difference according to the Mann-Whitney-Wilcoxon test (Table 2). Measurements along a 10 m transect (Figure 8) demonstrated that in a snowpack with a homogeneous density, the instrumental bias was the main source of variability while observers introduced only very low variability. It has also been detected that in most of the cases the snow densities measured by an experienced user and by the new user of the sampler did not significantly differ (Figure 7).

In Iceland the natural variability of snow along a 20 m long snow pit hindered the identification of instrumental bias, except for the SH that systematically measured a lower density compared to the average of all samplers. This is similar to the results found by Dixon & Boon (2012) who compared to the Federal sampler and the MSC (Meteorological Service of Canada) tubes. In Finland, the data also showed that the Federal sampler, EV2 and EV2-C provided the least consistent measurements (Figure 7). In general K-M, SH, Dolfi and ETH provided the most similar values and a low variability between the replicates. Three of them (K-M, Dolfi and ETH) have a relatively high diameter compared to the others, which could be beneficial for the very soft and low density snowpack we experienced over the Sodankyla campaign. The shorter snow core samplers (ETH, IG-PAS, VS-43) were in general not long enough to sample the whole snow column at once and measurements needed to be split into two steps, which increases the probability of errors. This process required more time and had a higher measurement uncertainty. Finally, it is logical that the EV2, EV2-C and Federal samplers have a higher uncertainty in Sodankyla (as was also shown for the Federal sampler in Iceland). The aforementioned difficulty to retrieve snow samples all the way to the ground under very soft and dry snow condition may explain the higher uncertainty for both devices. Indeed, the EV2 (and EV2-C) was designed to be used on glaciers and the Federal sampler was designed to sample deep and very dense, even icy snow (Marr, 1940). The reduced diameter and “no-digging” retrieval procedure may cause problems for sampling very soft snow. Dixon and Boon (2012) indicated for the Federal sampler that it performed much better when the snowpack was highly consolidated.

Even if this was not the main objective of the study, in Finland we could also confirm that scales may introduce some error, especially when they measure very light loads, with a maximum absolute error of 10% in the weighing process (see Supplementary Material, Text S2 and Table S3). Both spring scales and electronic scales yielded errors when measuring light weights, but electronic scales generally showed an increase of the error under very cold outdoor conditions. Although not analyzed in this study, it is necessary to consider that errors in electronic scales may be higher when batteries start to drain, when hanging samples are weighed in windy conditions, or if a scale is not properly leveled. In any case frequent checking and calibration of the scales is highly recommended.

Results shown here confirm the need to be cautious when assigning in situ snow measurements as ground truth, and the necessity to understand the natural variability of snow characteristics at small spatial scales and the instrumental and observer induced error. Even under relatively easy measurement conditions with a lack of ice layers and a moderately deep homogeneous snowpack, the uncertainty in snow density estimation is about 5% for an individual instrument and is close to 10% among the different instruments. Thus, for the estimation of SWE this uncertainty has to be added to the uncertainty of snow depth measurements. The homogeneous snow conditions found in Sodankyla (Finland) allowed the direct attribution of the instrumental error. Snow core samplers with larger diameter performed better than the narrower ones for those snow conditions, in line with conclusions of Farnes et al. (1982). On the other hand, snow samplers operated from the surface may miss part of the loose snow from the base of the snowpack, depending of the snow pack characteristic and the type of ground. Users should use snow core samplers best adapted to their prevailing snowpack and ground. Although careful measurement can partially eliminate some of the above uncertainties, SWE datasets composed of data from different instruments are likely to include inhomogeneities. Metadata on snow conditions and instruments used to measure the SWE together with intercomparison studies such as ours can help to estimate the accuracy of the data in such databases. Since this is not always possible to conduct the field intercomparison of the instruments, one can assume that the uncertainty of density measurements conducted by various devices in non-ideal snow conditions is approximately within 10-15%.

5 CONCLUSIONS

The results of the field campaigns provided a unique opportunity to analyze the uncertainty of measurements of bulk snow density and water equivalent of snow cover (SWE) carried out with snow core samplers that are regularly used in many European countries and beyond. To our knowledge, such a comparison in terms of number of device and environments has not been conducted before. The results showed that the devices provided slightly different uncertainties since they were designed for different snow conditions. The aim of this article was not to provide a definitive estimation of uncertainty for manual SWE measurements, but we think presented results represent a step forward in illustrating the role of the different uncertainty sources. The main conclusions can be enumerated as follows:

- In snowpack subjected to natural variability at small scales (e.g. irregular wind crusts or ice layers), as it was the case in Iceland campaign, the variability between close measurement points exceeded the one observed between repetitions at the same spot, although the latter is still high (often exceeding 5% and 10%). In case of uneven ground surface, snow depth may introduce more variability in SWE estimation compared to density. Therefore, repeated measurements in an observation field (at any time scale beyond a day) can reduce the number of SWE measurements using a fixed installed graduated snow depth stake and multiply the last measured density with the snow depth at the stake.
- In snowpack subjected to low natural variability at small scales, as it was the case in the Finland campaign, it was possible to examine instrumental bias. The largest differences were observed for snow core samplers introduced directly from the surface to the ground (no digging).
- Uncertainty induced by instrumental bias was generally less than 10% but it can reach 15%. Such differences suggest that inhomogeneities are likely to be introduced when there is a change of snow core sampler along the time, or different snow core samplers are used in a same field campaign. This is similar to measurements of other components of the water balance (e.g. precipitation) and it is unlikely that unification will be reached in near future due to varying snow and ground conditions in application areas.
- The above mentioned uncertainties refer to a homogeneous tundra-taiga snowpack (loose, mostly recrystallized snow and no wind impact) shallower than 1 meter depth. Relative errors will be smaller in deeper snowpack, with the exception of heavily wind or ice layer impacted snowpacks.
- The two measurements campaigns were conducted under very different snowpack conditions. We therefore presume that experiences made during these two campaigns are applicable to a wider range of snowpacks.
- Weighting process may introduce a considerable uncertainty especially for shallow snowpacks

ACKNOWLEDGMENTS

We thank H. Lowe and M. Jaggi for the help in analyzing the SMP data.

This research has been taken under the EU-COST Action HarmoSnow (ES1404: A European network for a harmonized monitoring of snow for the benefit of climate change scenarios, hydrology and numerical weather prediction). We also thank to project to HIDROIBERNIEVE (CGL2017-82216-R) and to the local organizers of the field campaigns who made possible this work. This work was partially supported within statutory activities No 3841/E-41/S/2019 of the Ministry of Science and Higher Education of Poland.

DATA AVAILABILITY STATEMENT

Used data is freely available in: <https://zenodo.org/record/2549310#.XErjW5VYZaQ>

REFERENCES

- Beaumont, R.T. & Work, R.A. (1963). Snow sampling results from three samplers. *Hydrological Sciences Journal*, 8(4), 74-78. doi 10.1080/02626666309493359
- Beaumont, R.T. (1967). Field Accuracy of Volumetric Snow Samplers at Mt. Hood, Oregon. *Physic of Snow and Ice Proceedings*, 1(2), 1007-1013.

- Berezovskaya, S. & Kane, D.L. (2007). Strategies for measuring snow water equivalent for hydrological applications: Part 1, accuracy of measurements. *Proceedings of 16th Northern Research Basin Symposium*, Petrozavodsk, Russia, 22–35.
- Bindon, H.H. (1964). The design of snow samplers for Canadian snow surveys. In *Proceedings of the 21st Annual Meeting of the Eastern Snow Conference*, Utica, New York, 23–28.
- Boon S., Davis R., Bladon K. & Wagner M. (2009). Comparison of Field Techniques for Measuring Snow Density at a Point. *Watershed Management Bulletin*, 12 (2), 7-12.
- Buhler, Y., Adams, M. S., Bosch, R., & Stoffel, A. (2016). Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations. *The Cryosphere*, 10, 1075-1088, doi.org/10.5194/tc-10-1075-2016.
- Church, J. E. (1933), Snow surveying: Its principles and possibilities. *Geogr. Rev.*, 23(4), 529–563, doi:10.2307/209242.
- Dixon, D. & Boon S. (2012). Comparison of the SnowHydro snow sampler with existing snow tube designs. *Hydrol. Process.*, 26, 2555-2562, 10.1002/hyp.9317
- Deems, J.S., Painter T.H. & Finnegan D.C. (2013). Lidar measurement of snow depth: a review. *Journal of Glaciology*, 59 (215), 467-479.
- Doesken, N.J. & Judson, A. (1996). A guide to the science, climatology, and measurement of snow in the United States. Ed. Colorado State University- Fort Collins (US), 87 pp.
- Dong, C. (2018). Remote sensing, hydrological modeling and in situ observations in snow cover research: A review. *Journal of Hydrology*, 561, 573-583.
- Farnes, P.F., Goodison, B.E., Peterson, N.R. & Richards, R.P. (1982). Metrication of Manual Snow Sampling Equipment. Final Report Western Snow Conference, Spokane, Washington 106 p.
- Fassnacht, S.R., Heun, C.M., Lopez-Moreno J.I. & Latron, J. (2010). Variability of Snow Density Measurements in the Rio Esera Valley, Pyrenees Mountains, Spain. *Cuadernos de Investigacion Geografica (Journal of Geographical Research)*, 36(1), 59-72.
- Fassnacht, S.R., Brown, K.S.J., Blumberg, E.J., Lopez-Moreno, J.I., Covino, T.P., Kappas, M., Huang, Y., Leone, V., & Kashipazha A.H. (2018). Distribution of snow depth variability. *Frontiers of Earth Science*, 12(4), 10 pages [doi: 10.1007/s11707-018-0714-z].
- Fernandes, R., Prevost, C., Canisius, F., Leblanc, S. G., Maloley, M., Oakes, S., Holman, K., & Knudby, A. (2018). Monitoring snow depth change across a range of landscapes with ephemeral snowpacks using structure from motion applied to lightweight unmanned aerial vehicle videos, *The Cryosphere*, 12, 3535-3550. <https://doi.org/10.5194/tc-12-3535-2018>.
- Fierz, C., Armstrong, R. L., Durand, Y., Etchevers, P., Greene, E., McClung, D. M., Nishimura, K., Satyawali, P. K. and Sokratov, S. A. (2009). The International Classification for Seasonal Snow on the Ground, UNESCO-IHP, Paris, France. [online] Available from: http://www.cryosphericosciences.org/outcomes/snowClassification/snowclass_2009-11-23-tagged-highres.pdf
- Freeman, T.G. (1965). Snow survey samplers and their accuracy. In *Proceedings of the 22nd Annual Meeting of the Eastern Snow Conference*, Hanover, New Hampshire, 1–10.
- Goodison B.E., Glynn J.E., Harvey K.D. & Slater J.E. (1987) Snow Surveying in Canada: A Perspective. *Canadian Water Resources Journal*, 12, 27-42, DOI: 10.4296/cwrj1202027.
- Haberkorn, A., Helmert, J., Leppanen, L., Lopez-Moreno, J.I., Pirazzini, R. (2019). *European Snow Booklet*, Haberkorn, A. (Ed.), 363 pp. doi:10.169904/envidat.59.

Helmert, J.; Şensoy Şorman, A.; Alvarado Montero, R.; De Michele, C.; de Rosnay, P.; Dumont, M.; Finger, D.C.; Lange, M.; Picard, G.; Potopová, V.; Pullen, S.; Vikhamar-Schuler, D.; Arslan, A.N. Review of Snow Data Assimilation Methods for Hydrological, Land Surface, Meteorological and Climate Models: Results from a COST HarmoSnow Survey. *Geosciences* 2018, 8, 489.

INTERACT Station Catalogue (2015). Eds.: Elger, K., Opel, T., Topp-Jørgensen, E., Hansen, J., Tairova, Z. and Rasch, M. DCE - Danish Centre for Environment and Energy, Aarhus University, Denmark. 305 p. (p. 72-75)

Jonas T., Marty C. & Magnusson, J. (2009). Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *J. Hydrol.* , 378, 161-167.

Kinar N. J. & Pomeroy J.W. (2015). Measurement of the physical properties of the snowpack, *Rev. Geophys.* , 53, 481-544. doi:10.1002/2015RG000481.

Komarov A.Y., Seliverstov Y.G., Grebennikov P.B. & Sokratov S.A. (2019). Spatial variability of snow water equivalent – the case study from the research site in Khibiny Mountains, Russia. *J. Hydrol. Hydromech.* , 67 (1), 110–112. doi: 10.2478/johh-2018-0016

Leppänen, L., Kontu, A. & Pulliainen, J. (2018) Automated Measurements of Snow on the Ground in Sodankylä. *Geophysica* 53(1), 43-62.

Libois, Q., Picard, G., Arnaud, L., Morin, S., & Brun, E. (2014), Modeling the impact of snow drift on the decameter-scale variability of snow properties on the Antarctic Plateau. *J. Geophys. Res. Atmos.* , 119 (11), 662–11,681, doi:10.1002/2014JD022361.

López Moreno, J.I., Fassnacht, S.R., Beguería, S. & Latron, J. (2011). Variability of snow depth at the plot scale: implications for mean depth estimation and sampling strategies. *The Cryosphere*, 5, 617-629.

López-Moreno, J.I., Fassnacht, S., Latron, J., Musselman, K., Morán-Tejeda, E. & Jonas, T. (2013). Small scale spatial variability of snow density and depth over complex alpine terrain: implications for estimating snow water equivalent. *Advances in Water Research*, 55, 40-52.

Kronholm, K., Schneebeli, M. and Schweizer, J. (2004). Spatial variability of micropenetration resistance in snow layers on a small slope. *Annals of Glaciology* , 38(1), 202–208. doi:10.3189/172756404781815257, 2004.

Marr, J. C. (1940). Snow Surveying. Soil Conservation Service, United States Department of Agriculture, Miscellaneous Publication No. 380, Washington D.C.

Marty, C. (2018). Recent Evidence of Large-Scale Receding Snow Water Equivalents in the European Alps. *Journal of Hydrometeorology*, 18, 1021-1031.

McCreight JL & Small E.E. (2014). Modeling bulk density and snow water equivalent using daily snow depth observations. *The Cryosphere*, 8, 521-536. doi.org/10.5194/tc-8-521-2014, 2014.

Nitu R., Roulet Y.-A., Wolff M., Earle M., Reverdin A., Smith C., Kochendorfer J., Morin S., Rasmussen R., Wong K., Alastrué J., Arnold L., Baker B., Buisán S., Collado J. L., Colli M., Collins B., Gaydos A., Hannula H.-R., Hoover J., Joe P., Kontu A., Laine T., Lanza L., Lanzinger E., Lee G.W., Lejeune Y., Leppänen L., Mekis E., Panel J.-M., Poikonen A., Ryu S., Sabatini F., Theriault J., Yang D., Genthon C., van den Heuvel F., Hirasawa N., Konishi H., Nishimura K., Senese A. (2018). Solid Precipitation Intercomparison Experiment 2012-2015, WMO Instruments and Observing Methods Report No. 131.

Pirazzini, R., Leppänen, L., Picard, G., Lopez-Moreno, J.I., Marty, C., Macelloni, G., Kontu, G., von Lerber, A., Melih-Tanis, C., Schneebeli, M., de Rosnay, P. & Arslan, A.N. (2018) European in-situ snow measurements: Practices and purposes. *Sensors*, 18, 7.

Peterson, N.R. & Brown, A.J. (1975). Accuracy of snow measurements. *Proceedings of the Western Snow Conference* 577-586 .

- Picard, G., Arnaud, L., Panel, J.M. & Morin, S. (2016). Design of a scanning laser meter for monitoring the spatio-temporal evolution of snow depth and its application in the Alps and in Antarctica, *The Cryosphere* , 10, 1495-1511. doi.org/10.5194/tc-10-1495-2016, 2016
- Proksch, M., Löwe, H., & Schneebeli, M. (2015). Density, specific surface area, and correlation length of snow measured by high-resolution penetrometry, *Journal of Geophysical Research: Earth Surface* , 120(2), 346–362. doi.org/10.1002/2014JF003266.
- Proksch, M., Rutter, N., Fierz, C., & Schneebeli, M. (2016). Intercomparison of snow density measurements: bias, precision, and vertical resolution, *The Cryosphere* , 10, 371-384. doi.org/10.5194/tc-10-371-2016.
- Revuelto, J., López-Moreno, J.I., Azorin-Molina, C. & Vicente-Serrano, S.M. (2014). Topographic control on snowpack distribution in a small catchment in the central Pyrenees: intra- and inter-annual persistence. *The Cryosphere*, 8 (5), 1889-2006.
- Schneebeli, M., & Johnson, J. B. (1998). A constant-speed penetrometer for high resolution snow stratigraphy. *Ann. Glaciol.* 26, 107–111.
- Stähli, M., Stacheder, M., Gustafsson, D., Schlaeger, S. & Schneebeli, M. 2004. A new in situ sensor for large-scale snow-cover monitoring. *Ann. Glaciol.* 38: 273–278.
- Stuefer S., Kane D.L. & Liston G.L. (2013). In situ snow water equivalent observations in the US Arctic. *Hydrology Research* , 44 (1), 21-34.
- Sturm M., Taras B., Liston G.E., Derksen C., Jonas T. & Lea J. (2010). Estimating snow water equivalent using snow depth data and climate classes. *J. Hydrometeorol.* , 11 (6), 1380-1394.
- Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.P., Koskinen, J., & Bojkov, B. (2011) Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sens. Environ.* , 115 (12), 3517–3529. <http://dx.doi.org/10.1016/j.rse.2011.08.014>.
- Wilcoxon F. (1945). Individual Comparisons by Ranking Methods. *Biometrics* , 1, 80-83.
- Work R.A., Stockwell H, Homer J., Freeman T.G. & Beaumont R.T. (1965). Accuracy of field snow surveys Western United States, including Alaska. *CRREL Report*, 163, 43 pp.

FIGURE CAPTIONS

Figure 1. Schematic representation of the sampling strategy applied during the two field campaigns. The scheme F1 was applied in Blafjöll (Iceland) and the scheme F2 in Sodankylä (Finland). In Sodankylä, the scheme for Bog and Forest plots was applied also in the Antenna plot, but in one subplot measurements were taken in transects (10 measurements per transect). In Bog and Forest opening sites (F2) the four subplots were measured in the same way. Letters N, O and I inform of the uncertainties contained at each measured spatial scale: Natural, induced by Observer and Instrumental bias respectively. The letter size makes reference to H high, M medium and L low relative influence of each uncertainty source at each site.

Figure 2. Instruments used in the campaigns, snow core samplers from left to right (see Table 1): Korhonen-Melander, Dolfi, VS-43, U.S. Federal, IG PAS, SnowHydro, Custom EV2, Enel-Valtecne EV2 and ETH. In addition, the SnowMicroPen is shown on the far right.

Figure 3. Boxplots showing the distribution of measured snow depth (upper panels), bulk snow density (middle panels), and SWE (lower panels) measured with different snow core samplers along two snow trenches in Iceland. Boxes stand for the 25th and 75th percentiles, vertical bars indicate the 10th and 90th percentiles and the horizontal central line is the median. Triangles at the bottom of some boxes inform about distribution skewness. Numbers above each box is the CV for repeated measurements. Dashed and dotted lines are the average and median, respectively, over all measurements on each plot.

Figure 4. Variability of the bulk snow density for the three plots as obtained from SMP measurements using the Proksch et al. (2015) parameterization (see text). The number above each box is the CV for all measurements on that plot. Boxes inform of the 25th and 75th percentiles, vertical bars indicate the 10th and 90th percentiles and the horizontal central line is the median. Dashed and dotted lines are the average and median over all measurements, respectively.

Figure 5. Variability of the 20 measurements (15 in Antenna plot) of snow depth, bulk snow density and SWE taken at the three plots (composed of 4 subplots each) conducted with different devices. The number above each box is the coefficient of variation among repeated measurements. Boxes inform of the 25th and 75th percentiles, bars indicate the 10th and 90th percentile and central line is the median. Triangles at the bottom of some boxes inform about high skewness of distribution. Dots are outliers. Dashed and dotted lines are the average and median respectively.

Figure 6. Bulk snow density difference of each sampler with respect to the total average of all samplers (in percentage) measured at each subplot for each of the three plots.

Figure 7. Bulk snow density measurements for each sampler when used by experienced observers and untrained observers on Bog and Forest plot. Six measurements were conducted with each sampler. Letters indicate the only pairs with statistically significant differences.

Figure 8. Measurements of snow depth (bottom panel) and bulk snow density (upper panel) along 10 m long transects at the Antenna plot.

Table 1. Summary of all the snow core samplers used during the campaigns and their main characteristics.

Name	Acronym	Country of origin	Length (cm)	Inner diameter (cm)	Material	Tooth-cutter	Sharpened cylinder
Korhonen-Melander sampler	K-M	Finland	70	10	Plastic	N	Y
ETH sampler	ETH	Switzerland	55	9.45	Aluminium	N	Y
SnowHydro SWE coring tube	SH	USA	165	6.13	Lexan	Y	N
Dreamly Original Light Fiberglass (Dolfi)	Dolfi	Czech Republic	100	8	Glass fiber	Y	N
Enel-Valtecne EV2	EV2	Italy	6 sections x 50	6	Stainless Steel	Y	N
Custom EV2	EV2-C	Italy	6 sections x 50	6	Aluminium	Y	N
IG-PAS snow sampler	IG PAS	Poland	50	7	Stainless Steel	N	Y
VS-43	VS-43	Russia	60	8	Aluminium	Y	N

Name	Acronym	Country of origin	Length (cm)	Inner diameter (cm)	Material	Tooth-cutter	Sharpened cylinder
Federal snow sampler	U.S. Federal	USA	7 sections x 76	3.81	Stainless steel	Y	N

Table 2. Statistically significant differences in measured bulk snow density (from the Mann-Whitney-Wilcoxon test) between pairs of snow core samplers in the Bog (B), Forest (F) and Antenna (A) plots during F2. ‘X’ means that no difference was found.

	VS-43	Dolfi	ETH	EV2	EV2-C	IG PAS	K-M	SH	Federal
VS-43									
Dolfi	X								
ETH	B,F,A	X							
EV2	B,F,A	X	F,A						
EV2-C	B,F,A	X	F,A	B,F,A					
IG PAS	B,F,A	B,F,A	X	X	X				
K-M	B,F,A	B,F,A	X	X	B,F,A	X			
SH	B,F,A	B,F,A	B	B,F,A	B,F,A	X	B,F,A		
Federal	B,F,A	B,F,A	B,F,A	B,F,A	B,F,A	B,F,A	B,F,A	B,F,A	









