

Progetto di Interesse NEXTDATA

WP1.6 Risorse criosferiche montane Task 3

Deliverable D1.6.D: Carte con risoluzione spaziale di almeno 10 km della variabilità della copertura nevosa (estensione e profondità) negli ultimi 20 anni sull'arco alpino italiano, ottenute da immagini satellitari, misure in situ e simulazioni numeriche

Preparato da:

Silvia Terzago, CNR-ISAC, Torino - s.terzago@isac.cnr.it

Climate Change Unit, Arpa Valle d'Aosta

con il contributo di:

Elisa Palazzi, Jost von Hardenberg,

CNR-ISAC, Torino

Abstract

The estimate of the past and current conditions of snow resources in the Alpine region would require reliable, kilometre-resolution, observation-based gridded data sets. Despite the growing attention in the last years and the enhancements in the observation networks, the density of station measurements is still insufficient to develop a high quality reference data set based on in-situ measurements. However, additional sources of information, mainly from remote sensing, reanalyses and climate models can be employed to get an overview of the Alpine snow resources.

The activity carried out during the project aimed at gathering the available gridded snow water equivalent data sets from remote sensing, reanalyses and climate model simulations for the greater Alpine region (GAR) in order to explore their ability to provide a coherent view of the snow water equivalent distribution and climatology in the Alpine area.

In addition to the analysis of the main satellite and reanalysis snow water equivalent products, we present here the evaluation of the simulations from the latest-generation regional and global climate models (RCMs, GCMs), participating in the Coordinated Regional Climate Downscaling Experiment over the European domain (EURO-CORDEX) and in the Fifth Coupled Model Intercomparison Project (CMIP5) respectively. We evaluate their reliability in reproducing the main drivers of snow processes – near-surface air temperature and precipitation – against the observational data set EOBS. Then we calculate for each model the snow water equivalent climatology in the last 26 years of the historical period (1980-2005) and compare it with the remote sensing and reanalysis data sets previously considered. We critically discuss the model limitations in the historical period and we explore the potential of the models in providing reliable future projections of snow water equivalent.

This analysis considering snow water equivalent climatologies in the Alpine region has been complemented with the quantification of snow water equivalent at small spatial scales (from the catchment to the local scales), in four selected areas in the Italian Alps. During the period 2015-2018, four intense field-measurement campaigns were performed, and the resulting measurements were analyzed and combined with advanced modelling tools to derive ultra-fine resolution snow water equivalent fields.

1. *Introduction*

In a context of climate change, reliable estimates of snowpack in the mountains and its changes in time are essential to develop management and adaptation strategies. Detailed studies on the impacts of global warming in snow-dominated regions are necessary for an informed management of water resources and to preserve essential ecosystem services for millions of people living in downstream areas. The density of surface stations measuring snow is currently insufficient to develop a reliable gridded snow water equivalent dataset based on in-situ measurements, thus calling for the use of alternative sources of information on snow mass, such as those derived from remote sensing observations and reanalyses (Mudryk et al., 2015).

Satellite measurements have been shown to provide a reliable picture of the global snow cover extent at few hundred meters of spatial resolution (Brown et al., 2010; Hall and Riggs, 2007), while the estimation of snow depth and snow water equivalent from satellite is typically performed at spatial scales of 25 km and is more challenging (Salzmann et al., 2014). Global reanalyses provide snow water equivalent fields at horizontal resolutions that are comparable (30 km in the zonal direction) or coarser than satellite products. Some reanalyses, such as ERA-Interim (Dee et al., 2011) and NCEP-CFSR (Saha et al., 2010), assimilate surface snow depth measurements and satellite snow cover extent while others, such as MERRA (Rienecker et al., 2011) and 20CR (Compo et al., 2011), are not constrained by measurements and thus rely on the capability of their land-surface model component to estimate snow fields.

To date, few studies have investigated the accuracy of satellite-based and reanalysis snow water equivalent (SNW) datasets against available observations, and very little is known on their performance in mountain areas.

In the frame of the NextData project, in the following we first review the available snow water equivalent (SNW) datasets and quantitatively assess the uncertainties in the estimation of the snow water equivalent in the Alpine environment. We consider global SNW gridded datasets obtained from satellite and reanalysis data and we explore how they represent the snow climatology over the Greater Alpine Region (GAR, 4–19°E, 43–49°N). Based on this analysis, we discuss the performances of state-of-the-art SNW products in this orographically complex area and we provide an estimate of the inter-dataset spread in the Alps.

Then, these results are used as a reference for evaluating the state-of-the-art climate models. We consider the models participating in the two major coordinated experiments: CMIP5, providing global simulations at spatial resolution on the order of 100 km on average, typically used as boundary and lateral conditions for regional climate simulations, and EURO-CORDEX, providing high-resolution regional simulations, on the order of 10 km, over Europe. For each of the 36 GCMs and 5 RCMs listed in Tables 1 and 2 we assess the ability to represent (i) the main drivers of snow processes, i.e., surface air temperature and precipitation, compared to the observational data set EOBS (Haylock et al., 2008), and (ii) the snow water equivalent climatology compared to the ensemble mean of the reference satellite and reanalysis data sets. We discuss the differences in the representation of snow relevant variables in the different model simulations.

Model	Institution	LSM	Res. [°lon]/Sp.Res	Reference
NSIDC-SNW	National Snow and Ice Data Center	–	25 km	Armstrong et al. (2005)
AMSR-E	National Snow and Ice Data Center	–	25 km	Tedesco et al. (2004)
CFSR	US National Centers for Environmental Prediction	Noah	0.3125	Saha et al. (2010)
MERRA	US National Aeronautics and Space Administration	Catchment LSM	0.67	Rienecker et al. (2011)
ERA-Interim/Land	European Centre for Medium-Range Weather Forecasts	HTESSEL	0.7	Balsamo et al. (2013)
20th Century Reanalysis	NOAA Earth System Research Laboratory	Noah	1.875	Compo et al. (2011)
CMCC-CM	Euro-Mediterranean Centre for Climate Change	ECHAM5	0.75/T159	Scoccimarro et al. (2011)
EC-Earth	EC-Earth Consortium	HTESSEL	1.125/T159	Hazeleger et al. (2012)
BCC-CSM1-1-M	Beijing Climate Center, China	BCC_AVIM1.0	1.125/T106	Wu et al. (2013)
MRI-CGCM3	Meteorological Research Institute, Japan	HAL	1.125/T159	Yukimoto et al. (2012)
CESM1-BGC	National Center for Atmospheric Research	CLM4	1.25	Hurrell et al. (2013)
CESM1-CAM5	National Center for Atmospheric Research	CLM4	1.25	Hurrell et al. (2013)
CESM1-FASTCHEM	National Center for Atmospheric Research	CLM4	1.25	Hurrell et al. (2013)
CCSM4	National Center for Atmospheric Research	CLM4	1.25	Gent et al. (2011)
CNRM-CM5	Centre National de Recherches Météorologiques	ISBA	1.4/T127	Voldoire et al. (2013)
ACCESS1-0	CSIRO/BOM, Australia	MOSES2	1.875/N96	Bi et al. (2013)
ACCESS1-3	CSIRO/BOM, Australia	CABLE1.0	1.875/N96	Bi et al. (2013)
CMCC-CMS	Euro-Mediterranean Centre for Climate Change	ECHAM5	1.875/T63	Scoccimarro et al. (2011)
CSIRO-Mk3-6-0	CSIRO, Australia	MOSES II	1.875/T63	Collier et al. (2011)
HadGEM2-AO	Met Office Hadley Centre	MOSES II	1.875/N96	Collins et al. (2011)
HadGEM2-CC	Met Office Hadley Centre	MOSES II	1.875/N96	Collins et al. (2011)
HadGEM2-ES	Met Office Hadley Centre	MOSES II	1.875/N96	Collins et al. (2011)
MPI-ESM-LR	Max Planck Institute for Meteorology	JSBACH	1.875/T63	Giorgetta et al. (2013)
MPI-ESM-MR	Max Planck Institute for Meteorology	JSBACH	1.875/T63	Giorgetta et al. (2013)
MPI-ESM-P	Max Planck Institute for Meteorology	JSBACH	1.875/T63	Giorgetta et al. (2013)
INM-CM4	Institute for Numerical Mathematics	INM	2.0	Volodin et al. (2010)
CESM1-WACCM	National Center for Atmospheric Research	CAM	2.5	Hurrell et al. (2013)
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory	LM3	2.5	Donner et al. (2011)
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory	LM3	2.5	Dunne et al. (2012)
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory	LM3	2.5	Dunne et al. (2012)
GFDL-CM2p1	NOAA Geophysical Fluid Dynamics Laboratory	LM2	2.5	Delworth et al. (2006)
GISS-E2-H-CC	NASA Goddard Institute for Space Studies	GISS LSM	2.5	Schmidt et al. (2006)
GISS-E2-H	NASA Goddard Institute for Space Studies	GISS LSM	2.5	Schmidt et al. (2006)
GISS-E2-R-CC	NASA Goddard Institute for Space Studies	GISS LSM	2.5	Schmidt et al. (2006)
GISS-E2-R	NASA Goddard Institute for Space Studies	GISS LSM	2.5	Schmidt et al. (2006)
NorESM1-ME	Norwegian Climate Centre	CLM4	2.5	Bentsen et al. (2013)
NorESM1-M	Norwegian Climate Centre	CLM4	2.5	Bentsen et al. (2013)
BNU-ESM	Beijing Normal University, China	BNU-CoLM3	2.8125/T42	*
CanESM2	Canadian Centre for Climate Modelling and Analysis	CLASS	2.8125/T63	Arora et al. (2011)
FGOALS-g2	LASG/CESS, China	CLM3	2.8125	Li et al. (2013)
FIO-ESM	The First Institute of Oceanography, China	CLM3.5	2.8125/T42	Qiao et al. (2013)
HadCM3	Met Office Hadley Centre	MOSES 1	3.75/N48	Johns et al. (2003)

* Reference is http://esg.bnu.edu.cn/BNU_ESM_webs/html/index.html.

Table 1: Snow water equivalent data sets, including remote sensing products, reanalyses, CMIP5 global climate models used in the study, with the corresponding land-surface model (LSM, when it applies) and the spatial/spectral horizontal resolution (from Terzago et al., 2017).

Model	Institution	LSM	Ensemble members	Reference
CCLM4-8-17	CLM Community	Terra-ML	4	Rockel et al. (2008)
ALADIN53	Centre National de Recherches Météorologiques	ISBA	–	Farda et al. (2010)
HIRHAM5	Danish Meteorological Institute	Hagemann (2002)	1	Christensen et al. (2007)
RACMO22E	Royal Netherlands Meteorological Institute	HTESSEL	2	Van Meijgaard et al. (2012)
REMO2009	Climate Service Center	Hagemann (2002)	1	Jacob and Podzun (1997)

Table 2 EURO-CORDEX regional climate models providing ERA-Interim-driven runs for the snow water equivalent variable at 0.11° spatial resolution considered in this study. For each of model we also report the land-surface model (LSM), the number of available GCM-driven runs and the reference (from Terzago et al., 2017).

The complex orography of the area, shown in Fig. 1a, and the heterogeneous pattern of steep slopes and valleys hamper the representation of climate features from both an observational and a modelling point of view. As an example, Fig. 1b points out how the topography is represented in the 1 km GLOBE digital elevation model (Hastings et al., 2000), in the CORDEX ERA-Interim-driven regional climate models and in the CMIP5 global climate models, in terms of median, 5th and 95th percentiles of the distribution of elevations. The median elevation is well reproduced by all models while the lowest and highest elevations are progressively cut out as the model spatial resolution coarsens. While global climate models, including those with the finest spatial resolution, do not properly take into account elevations above 1500 m a.s.l. in the GAR, RCMs are closer to the expected real values. The limitations of CMIP5 models have to be considered when analysing GCM

outputs over mountain areas, since the world reproduced by such coarse-scale models has a smooth orography and simplified, or highly parameterized, physical processes.

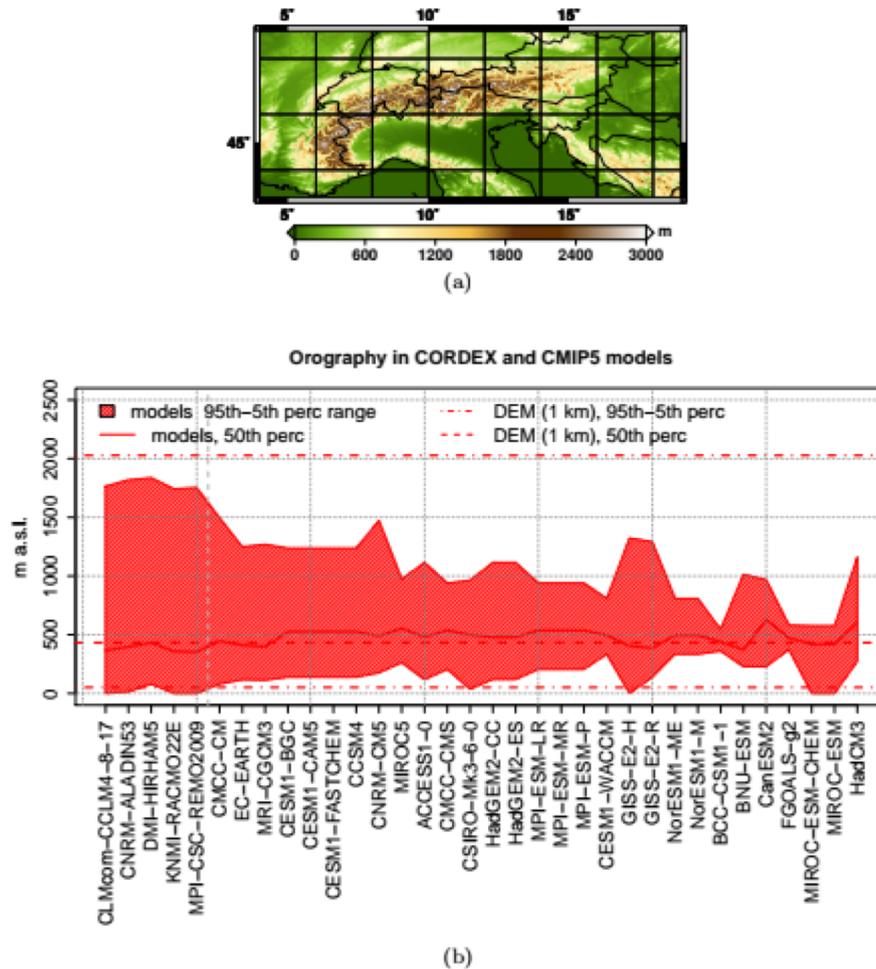


Figure 1: (a) Orography of the greater Alpine region (4–19° E; 43–49° N) as in the GLOBE 1 km digital elevation model (DEM). (b) The 95th (dash-dotted), 50th (dashed) and 5th (dash-dotted) percentiles of the elevation distribution in the DEM compared to the corresponding values obtained from the CORDEX and CMIP5 model orographies. RCM and GCM models are ordered along the x-axis from finest to coarsest spatial resolution. RCMs and GCMs are separated by a vertical dashed line (from Terzago et al., 2017).

2. Snow water equivalent in satellite products and reanalyses

We present the spatial distribution of snow water equivalent in satellite products and reanalyses, hereafter referred to as the reference data sets, and we evaluate the differences among the reanalyses in relation to possible biases in the meteorological forcing.

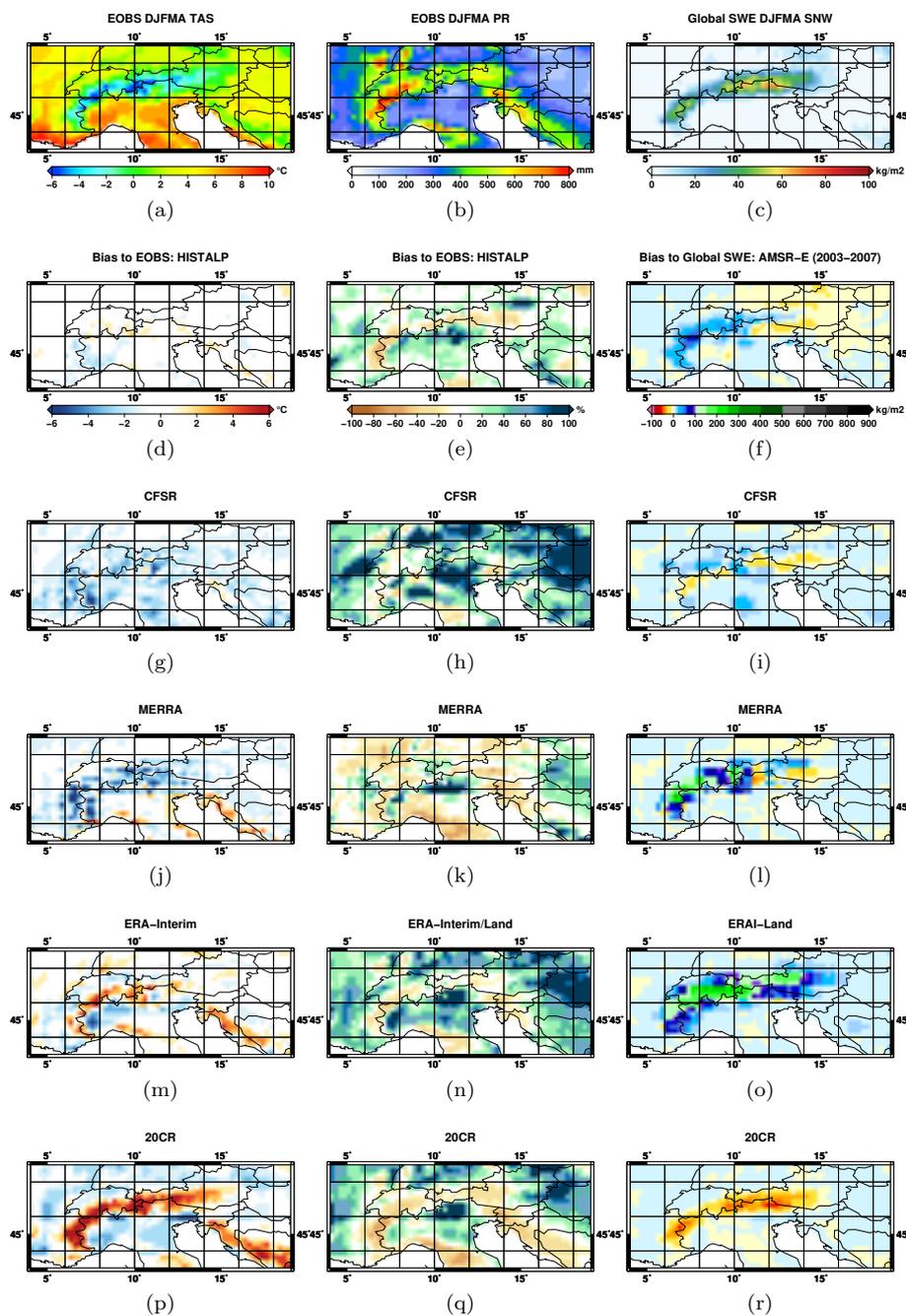


Figure 2 Multiannual mean (1980–2005) of the DJFMA average (a) air temperature, (b) total precipitation from EOBJS observational datasets and (c) snow water equivalent from NSIDC-SNW. Panels from (d) to (r) represent the bias of HISTALP, AMSR-E and reanalyses with respect to EOBJS and NSIDC-SNW data sets respectively (from Terzago et al., 2017).

Figure 2 shows the multiannual mean (1980–2005) of near-surface air temperature (TAS), precipitation (PR) and SNW averaged (or accumulated in the case of PR) over the months from December to April (DJFMA). In order to facilitate the comparison among the various data sets, we present the differences (or percent biases) with respect to a given data set, namely EOBS for TAS and PR and NSIDC-SNW for SNW. NSIDC-SNW, is in fact available for a longer period (1980–2005) than the other satellite product, AMSR-E (2003–2011). All data sets are conservatively remapped onto a regular 0.25° resolution grid. Biases are calculated over the period 1980–2005 except for AMSR-E, for which the period of overlap with the reference data set is shorter, 2003–2007. Compared to EOBS, the alternative observational, high-resolution climatology from HISTALP (Fig. 2d–e) presents a similar temperature distribution, drier conditions at high elevations and wetter conditions at low elevations. This comparison is reported to highlight the fact that uncertainties are larger in precipitation than in temperature estimates, especially in mountain areas, and also observational data sets can exhibit biases with respect to each other, which makes it even difficult to define a ground truth or reference against which to compare/validate other data sets.

Focusing on the snow water equivalent distribution, the NSIDC-SNW climatology (Fig. 2c) shows maximum values of about 50 kg m^{-2} over the western Alps and 70 kg m^{-2} over the eastern Alps. If we consider the other satellite and reanalysis products we obtain a rather heterogeneous picture.

AMSR-E (Fig. 2f) presents higher values in the western Alps and lower values in the eastern Alps compared to the NSIDC SNW. CFSR (Fig. 2g–i) shows TAS and PR patterns that are similar to EOBS over the Alpine ridge and a SNW distribution that is similar to NSIDC-SNW. The similarity in the SNW range of variability is probably due to the fact that both products integrate the Special Sensor Microwave Imager (SSM/I) data but to different extents. NSIDC-SNW is specifically derived from the Special Sensor Microwave Imager (SSM/I) data.

The CFSR snow output is mainly based on the Noah land-surface model first guess; A daily snow analysis based on several inputs, including the Special Sensor Microwave Imager (SSM/I) data, is used to constrain the model first guess (Meng et al., 2012). The CFSR snow depth/SNW is limited in the upper and lower boundaries by the snow analysis (it cannot be larger than twice and lower than half the snow analysis) but the temporal evolution of snow depth and SNW is determined by the Noah model. As a consequence, the two SNW data sets lie in similar ranges of variability, but except for this feature they can be considered independent.

The MERRA Reanalysis (Fig. 2j–l) shows a thicker snowpack with respect to NSIDC-SNW, especially over the Western Alps, as well as compared to AMSR-E. The MERRA behaviour can be explained by a cold bias over that area, partly compensated by drier conditions over the Alpine peaks.

ERA-Interim/Land (Fig. 2m–o) shows the largest SNW values, with peaks exceeding NSIDC-SNW values by more than 200 kg m^{-2} . The SNW bias is not directly explainable in terms of biases in temperature and precipitation, which indeed go towards the opposite direction (warmer and slightly drier with respect to EOBS). This result suggests that the high SNW values of ERAInterim/Land can be attributed to the snow scheme in use.

20CR (Fig. 2p–r) shows the lowest SNW values. Owing to its coarse spatial resolution, 20CR presents a warm and dry bias at high elevations and a cold and wet bias at low elevations, which in turn result in low snow accumulation and shallow snowpack over the mountain range. These simplified patterns can presumably be ascribed to an excessively smooth orography and highlight the limitations of the 20CR reanalysis in the representation of snow processes in mountain areas.

This analysis provides a quite heterogeneous picture of SNW and, despite the considerations on the biases found in the climatic drivers, at the time being it is not possible to ultimately define which product is closest to reality over the full GAR domain. For further analysis we disregard the 20CR reanalysis owing to its poor performance in this orographically complex region and the AMSR-E satellite product for its short period of availability. We thus consider as reference the mean of the other four data sets, i.e. NSIDC SNW, CFSR, MERRA and ERA-Interim/Land reanalyses. This multi-reference mean (MRM) is calculated after conservatively remapping all the data sets to the 0.7° longitude ERA-Interim/Land grid.

3. Snow water equivalent in global climate models

Global climate models clearly have too coarse spatial resolution to properly represent snowpack characteristics in mountain regions. However, it is interesting to evaluate them in the Alpine area at least in terms of temperature and precipitation biases with respect to observations owing to the fact that such models provide boundary and forcing conditions to regional climate simulations.

We analyze in detail the DJFMA TAS, PR and SNW climatologies provided by CMIP5 global climate models with spatial resolution equal to or finer than 1.25° (Fig. 3); coarser resolution GCMs are discussed further in Sect 5. CMIP5 model biases with respect to EOBS and NSIDC SNW references (Fig. 2a–c) are shown Fig. 3. The comparison period is 1980–2005. Of the four CESM-family models, namely CESM1-CAM5, CESM1-BGC, CESM1-FASTCHEM and CCSM4, three models present very similar climatologies so here we consider only one of them, CESM1-BGC, which is taken to be representative of CESM1-FASTCHEM and CCSM4.

GCMs with spatial resolution equal to or finer than 1.25° show snow amounts which are comparable to those of the reference data sets over the greater Alpine region. Compared to NSIDC-SNW, the models CMCC-CM, EC-Earth and, to a smaller extent, MRI-CGCM3 and CESM1-CAM5, show thicker snowpack at the northern slope of the Alps and in Switzerland. A common feature of all data sets is a shallower snowpack over the eastern Alps, at the border between Italy and Austria. This spatial pattern, characterized by an east–west gradient, with shallower snowpack in the eastern Alps and thicker snowpack in the western Alps, more closely resembles that provided by the AMSR-E satellite product rather than that provided by NSIDC-SNW.

BCC-CSM1-1-M and CESM1-BGC show shallower snowpacks than NSIDC-SNW, and higher temperatures with respect to the observational data sets. In these cases, the warm bias in the model can explain a less abundant snowpack. From this analysis, the precipitation bias over the Alpine ridge between the different high-resolution GCMs seems to be comparable. In fact, GCMs generally tend to a slight underestimation of winter precipitation at the ridges and to an overestimation at lower altitudes. This uniform behaviour in the precipitation pattern suggests that temperature can be the leading cause of biases in the estimation of surface snow water equivalent.

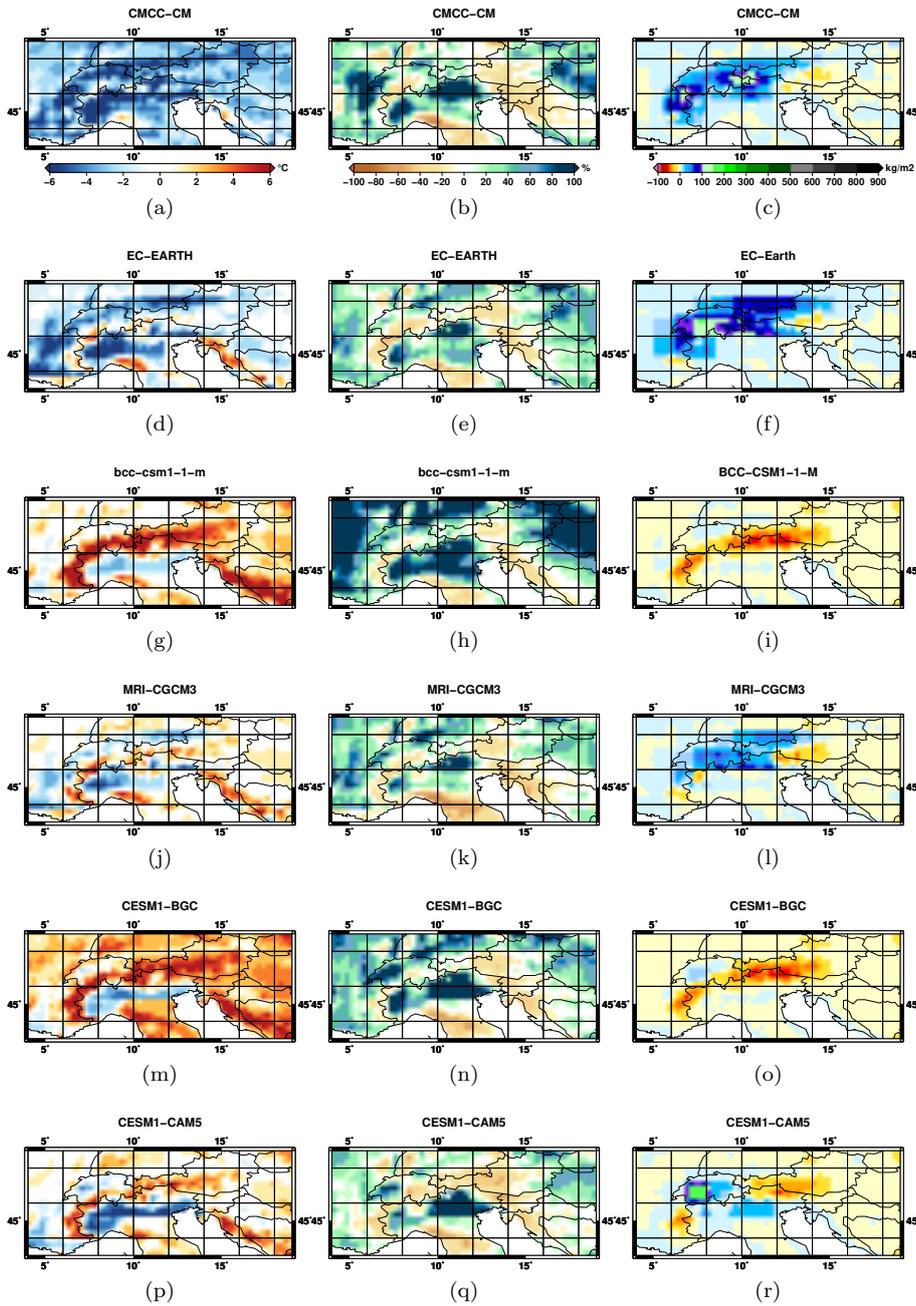


Figure 3 DJFMA (first column) air temperature, (second column) total precipitation and (third column) snow water equivalent biases of the CMIP5 global climate models with spatial resolution equal to or finer than 1.25° longitude with respect to the EOBS and NSIDC-SNWclimatologies reported in Fig. 2a, b, c (from Terzago et al., 2017)

4. SNW in regional climate models

In this section we evaluate the outputs of the EURO-CORDEX regional climate model experiments providing the highest resolution climatesimulations currently available for the European domain.

For simulating the past climate, ERA-Interim driven simulations can be used, while for projecting future evolution of snow water equivalent GCM-driven simulations are necessary. We evaluate both products in the following.

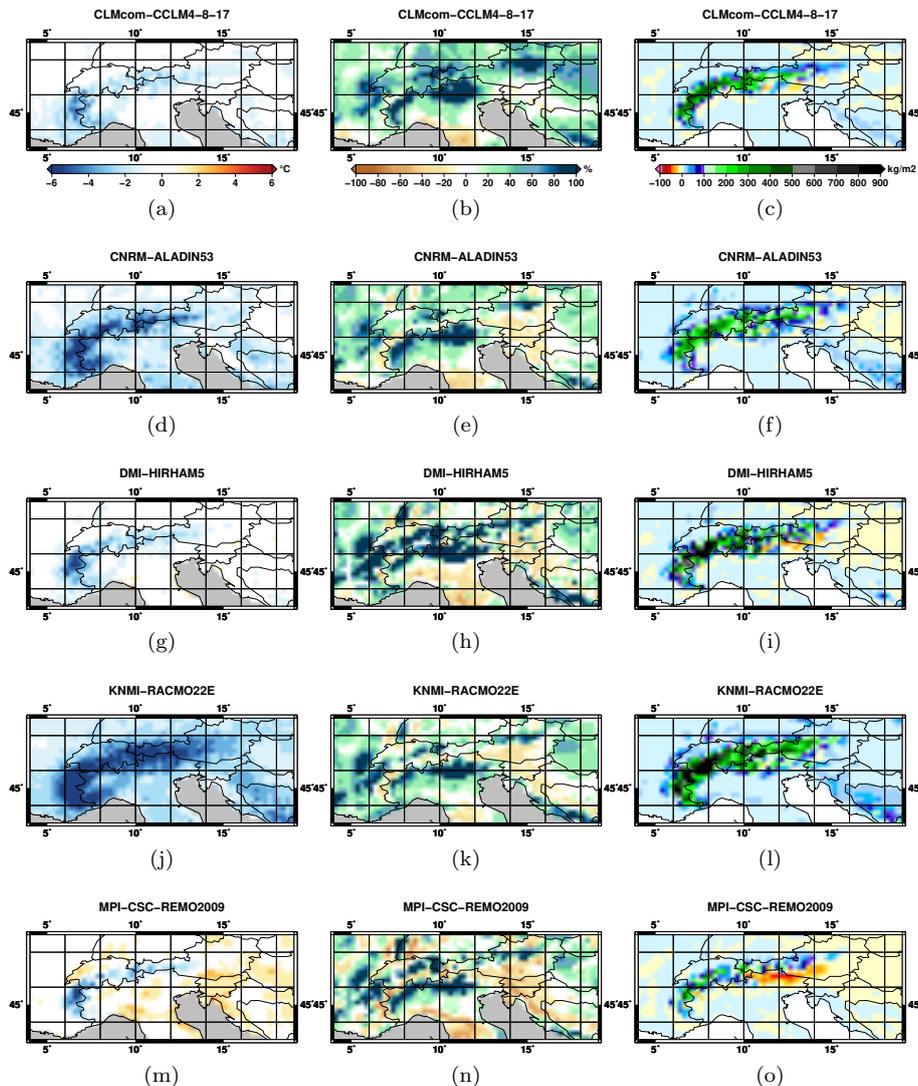


Figure 4 As in Fig. 3 but for the CORDEX ERA-Interim-driven RCM simulations, averaged over the period 1990–2005 (from Terzago et al., 2017).

Figure 4 shows the biases of ERA-Interim-driven regional climate model DJFMA TAS as well as PR and SNW climatologies with respect to the EOBS and NSIDC-SNW references, all averaged over the common period 1990–2005. All RCMs show SNW amounts several hundreds of kg m^{-2} larger than any other reference data set (Fig. 2) at the mountain ridge and lower values at low elevations. Extremely high values (shown in black) are not reliable as they correspond to areas of continuous snow accumulation and no melting, possibly areas masked as glaciers in the models. Such grid points show artificially high positive trends, which can be considered to be erroneous, and they have to be discarded from the analysis. Despite these details, RCM snow estimates are much higher than those provided by the reference data sets, and these high values can be related to the fine representation of the orography that allows, in principle, for lower temperatures in high mountain areas that are not represented in coarse-scale reanalyses, for increased solid precipitation and for longer snowpack duration. In some cases, the large SNW values in RCMs can be partly explained by cold biases (RACMO22E, ALADIN53) or wet biases (HIRHAM5) with respect to the observations. In other cases (CCLM4-8-17), despite remarkable biases in some parts of the domain, the atmospheric forcings in correspondence of the mountain ridge are in better agreement with observations and they do not show relevant deviations from the reference climatologies, so the differences have to be attributed to the snow scheme in use and/or to the finer representation of the topography. From the analysis of RCMs we can conclude that higher spatial resolution allows areas of snow accumulation to be better separated and, consequently, to reproduce higher snow maxima in correspondence of mountain peaks.

For the CCLM4-8-17 and REMO2009 models, which display no issues in the snow accumulation trends, we also investigated the GCM-driven simulations (Table 2). GCM-driven CCLM4-8-17 climatologies have a stronger negative temperature bias (CNRM-CM5, EC-Earth, HadGEM2-ES) and/or stronger positive precipitation biases (CNRM-CM5, MPI-ESM-LR) with respect to the ERA-Interim-driven runs (Fig. 5). These features result in thicker snow water equivalent. In the case of MPI-ESM-LR-driven REMO2009 the temperature bias is comparable while the precipitation bias is larger than for the ERA-Interim-driven runs. In conclusion, GCM-driven RCM simulations tend to inherit the biases already present in the driver GCM and to reflect them in SNW fields.

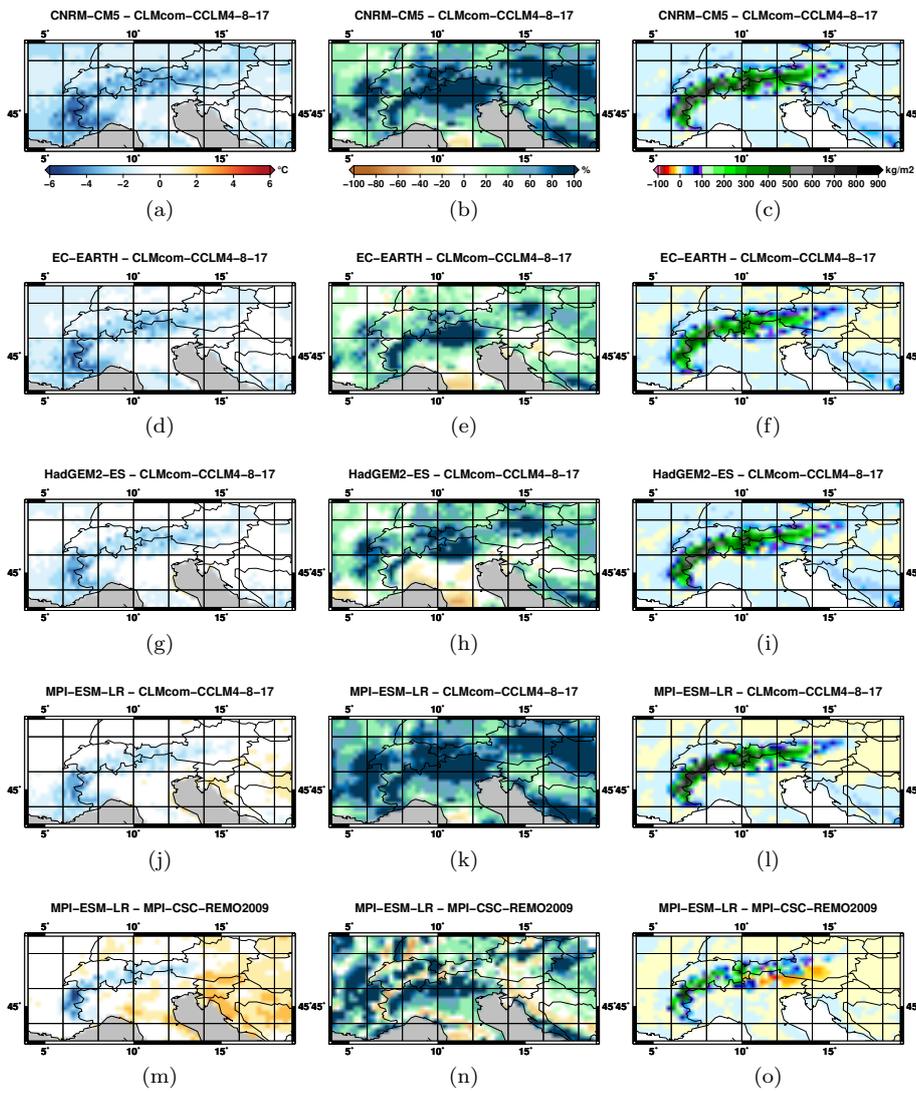


Figure 5 Biases of the GCM-driven CCLM4 and REMO2009 DJFMA surface air temperature, precipitation and snow water equivalent climatologies with respect to the reference datasets EObs and NSIDC Global SNW, over the period 1980-2005 (from Terzago et al., 2017)

5. Global view of SNW gridded products

In this section we provide a comprehensive view of all the previously considered SNW gridded data sets. The similarity of the SNW climatologies is quantified using Taylor diagrams (Taylor, 2001). Figure 6a compares the spatial distribution of the DJFMA snow water equivalent, averaged over the period 1980–2005, for the multi-reference mean (MRM, mean of the four reference data sets CFRS, MERRA, ERA-Interim/Land and NSIDC-SNW) to which all other data sets are compared; the multi-model mean (MMM), mean of all 36 CMIP5 models; the multi-model mean of the CMIP5 models with spatial resolution equal to or finer than 1.25° (MMM-HiRes, as in Terzago et al., 2014); the individual reference data sets; and the individual regional and global climate models.

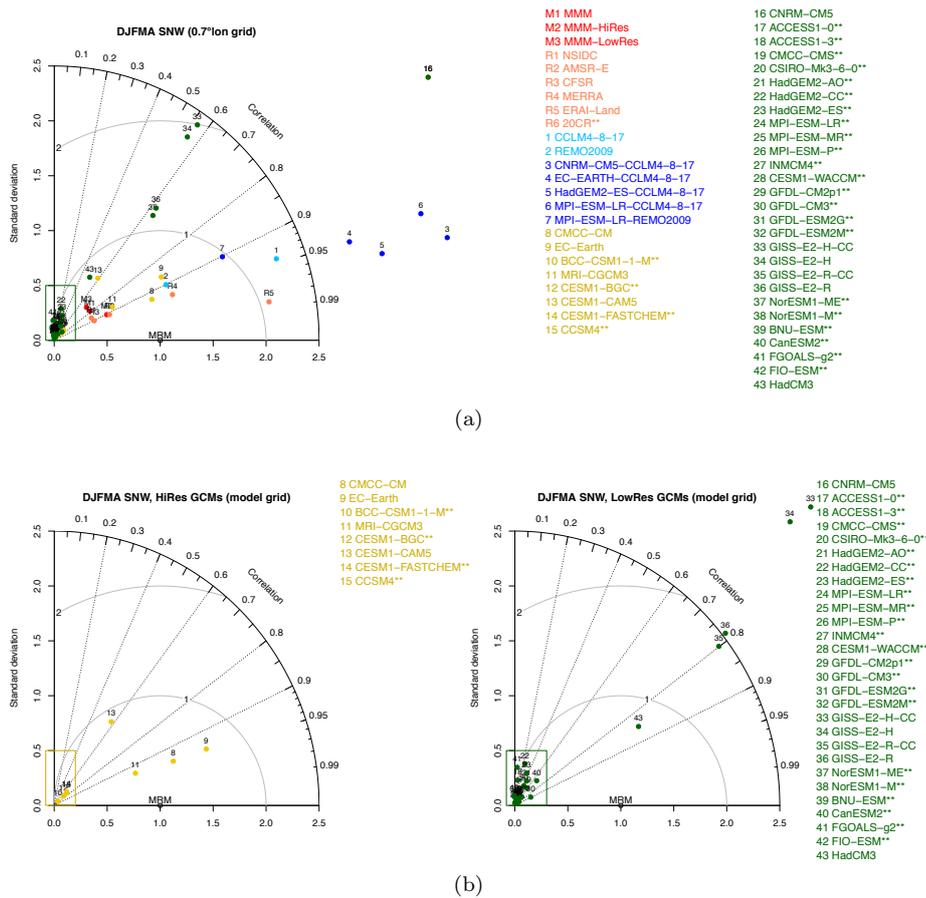


Figure 6 Taylor diagrams of the multiannual mean (1980–2005) of the DJFMA average snow water equivalent as described by climate models are kept at their original resolution and the reference data sets are remapped onto the grid of each model. Points included in the rectangles correspond to models highlighted with ** in the legend (from Terzago et al., 2017).

First we compare data sets built on different coordinate reference systems and with different spatial resolutions by reprojecting all remote sensing products, reanalyses and climate model outputs onto a common grid, specifically the ERA-Interim/Land 0.7° longitude grid. Figure 6a provides an evaluation of the individual data sets with respect to the multi-reference mean, all resampled on the same 0.7° grid.

Reference data sets are generally highly correlated with the MRM ($R > 0.85$ for all data sets except the coarsest 20CR). This feature is related to the dependence of the snow water equivalent on topography; i.e. these data sets represent larger SNW values at higher altitudes. Satellite products and the CFSR reanalysis are very close to each other, with lower variance with respect to the MRM. The MERRA reanalysis is close to the MRM, with comparable standard deviation and small RMSE. The ERA-Interim/Land and 20CR reanalyses show opposing behaviours in terms of normalized standard deviation, i.e. very high and very low respectively. ERA-Interim/Land has a wider statistical dispersion of SNW values and higher SNW peaks, clearly reflected in Fig. 2o, while 20CR has a narrow range of SNW values and a smooth SNW pattern (Fig. 2r).

Of the two RCMs considered, REMO2009 is in better agreement with the MRM in terms of RMSE and normalized standard deviation. CCLM4-8-17 has a large normalized standard deviation, which is comparable to that found in ERA-Interim/Land. All GCM-driven RCM simulations show higher variance with respect to the corresponding ERA-Interim-driven runs.

For GCMs, an important feature emerging from this analysis is that, on average, the ensemble mean of the high resolution models performs better in terms of standard deviation, root mean square difference and pattern correlation, with respect to the ensemble mean of all CMIP5 GCMs. This result highlights the importance of the horizontal resolution in simulating snowpack spatial patterns.

An alternative approach has been devised to provide a fair comparison of the GCMs. Each GCM is compared to the MRM after having conservatively remapped each reference data set onto the individual GCM grid, so that the reference is reshaped each time according to the model resolution. This approach allows for a fair evaluation of each GCM on its own grid, regardless of its resolution. For the sake of clarity, we present the results relative to this approach by separately plotting the models divided in two groups, one with resolutions finer or equal than 1.25° and one with coarser resolutions (Fig. 6b).

The clustering based on spatial resolution reveals that coarse resolution models generally have very high or very low standard deviation (please note that the CNRM-CM5 model lays outside the range of the plot). In such circumstances the ensemble mean of the models is the result of compensating extreme behaviours, and it should be considered with caution. On the contrary, individual high-resolution GCMs are generally closer to the MRM and do not exhibit extreme features, constituting a more homogeneous ensemble.

Figure 6 provides information on the similarity of SNW climatologies and, indirectly, qualitative information on the degree of interdependency of the models belonging to the same “family”. For example, among the previously mentioned four CESM-family models, namely CESM1-CAM5, CESM1-BGC, CESM1-FASTCHEM and CCSM4, three models show a high degree of similarity (Figure 6b). In the calculation of the MMM-HiRes, in order to limit the bias related to the interdependency of the models, out of these three similar models we retained only one, CESM1-BGC. We use the term “high-resolution GCMs” to indicate only the following six models: CMCC-CM, ECEarth, MRI-CGCM3, BCC-CSM1-1-M, CESM1-BGC and CESM1-CAM5. These models were further analysed in

terms of annual cycle of snow water equivalent during the historical period compared to the reference datasets and in terms of projected changes in the annual cycle at mid and end 21st century in RCP8.5 scenario in a specific study (Terzago et al., 2017).

Conclusions

The results of the analysis show that the time-averaged spatial distribution of snow water equivalent is reproduced quite differently by the different remote sensing and reanalysis datasets, which in fact exhibit a large spread around the ensemble reference mean. The information on snow water equivalent climatology provided by these datasets is affected by large uncertainties, thus it can be considered only qualitatively.

The analysis of EURO-CORDEX regional climate model simulations elucidates the extent to which horizontal resolution can affect the representation of the snow processes and climatology in mountain areas. The results from the currently available simulations at 0.11° resolution (five ERA-Interim-driven RCMs) show a much thicker average snowpack over the alpine ridge and shallower snowpack at low elevations with respect to the reference data set and global climate models. This behaviour, has been related to the finer-resolution of RCM with respect to GCMs and reanalyses.

The present work highlights how surface heterogeneity at fine scale is difficult to represent, both for remote sensing products, reanalyses and climate models. Specifically, this study contributes to the knowledge on Alpine snowpack status and changes by providing a picture of the main available snow water equivalent products, and measuring the related uncertainties in the Alpine environment. The relative assessment of the capability of satellite-based products, reanalyses, RCMs and GCMs in reproducing snowpack features provides important information to both model developers and to the community of users, allowing critical factors in the model components to be identified and raising awareness of the strengths and limitations of the available products.

Data availability. All the data sets used in this study are publicly accessible and were downloaded from the following websites: CMIP5 and CORDEX model simulations, <https://esgf-data.dkrz.de/projects/esgf-dkrz/>; NSIDC Global Snow Water Equivalent climatology and AMSR-E products, <https://nsidc.org/>; CFSR reanalysis, <https://rda.ucar.edu/>; MERRA reanalysis, <https://mirador.gsfc.nasa.gov/>; ERA-Interim/Land reanalysis, <http://apps.ecmwf.int>; EOBS, <http://www.ecad.eu>; HISTALP, <http://www.zamg.ac.at/histalp/>.

All the codes and the datasets produced for this deliverable are available upon request (email: s.terzago@isac.cnr.it).

References

- Brown, R., Derksen, C., & Wang, L., 2010: A multi-data set analysis of variability and change in Arctic spring snow cover extent, 1967–2008. *Journal of Geophysical Research: Atmospheres*, 115(D16).
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., ... & Brönnimann, S., 2011: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654), 1-28.

- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, I., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thèpaut, J.-N. and Vitart, F. 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quart. J. Roy. Meteor. Soc.*, 137(656), 553-597.
- Hall, D. K., & Riggs, G. A., 2007: Accuracy assessment of the MODIS snow products. *Hydrological Processes: An International Journal*, 21(12), 1534-1547.
- Hastings D.A., Dunbar P.K., Hittelman A.M. (2000) Assessing the global land one-km base elevation DEM. In: Schwarz KP. (eds) *Geodesy Beyond 2000*. International Association of Geodesy Symposia, vol 121. Springer, Berlin, Heidelberg
- Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M., 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.-Atmos.*, 113, <https://doi.org/doi:10.1029/2008JD010201> .
- Meng, H., Dong, J., Ferraro, R., Yan, B. H., Zhao, L. M., Kongoli, C., Wang, N.-Y., and Zavodsky, B., 2017: A 1DVAR-based snowfall rate retrieval algorithm for passive microwave radiometers, *J. Geophys. Res.*, 122, 6520-6540, doi:10.1002/2016JD026325.
- Mudryk, L. R., Derksen, C., Kushner, P. J., & Brown, R., 2015: Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. *Journal of Climate*, 28(20), 8037-8051.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... & Bloom, S., 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate*, 24(14), 3624-3648.
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., ... & Liu, H., 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015-1058.
- Salzmann, N., Huggel, C., Rohrer, M., & Stoffel, M., 2014: Data and knowledge gaps in glacier, snow and related runoff research—A climate change adaptation perspective. *Journal of Hydrology*, 518, 225-234.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, 2001.
- Terzago, S., von Hardenberg, J., Palazzi, E., and Provenzale, A.: Snowpack changes in the Hindu-Kush Karakoram Himalaya from CMIP5 Global Climate Models, *J. Hydrometeorol.*, 15, 2293–2313, <https://doi.org/10.1175/JHM-D-13-0196.1>, 2014.
- Terzago, S., von Hardenberg, J., Palazzi, E., and Provenzale, A., 2017: Snow water equivalent in the Alps as seen by gridded data sets, CMIP5 and CORDEX climate models, *The Cryosphere*, 11, 1625-1645, <https://doi.org/10.5194/tc-11-1625-2017> .



SWE Inter-comparison

Pian de L'Eyvia - Pila
March 5th, 2015

Technical report by G. Filippa, E. Cremonese, F. Diotri, U. Morra di Cella - ARPA Valle d'Aosta



In a nutshell

- 186 SWE measurements with three different methods over a 1000 m² surface, 34 snow observers involved (fig. 1).
- Two observers taking two SWE measurements on a snow pack under similar conditions and using similar methods result in an average measurement error of 3% (fig. 4).
- A group of 17 couples of observers taking SWE measurements along a 12-15 m linear transect on a homogeneous snow pack results in an average measurement error of 7% (fig. 4). The inter-operator variability (as high as 100 mm SWE, fig. 5), highlights the need for further inter-comparisons aiming at the homogenization of sampling techniques.
- The measurement of a weighted average snow density according to homogeneous strata (AINEVA method) leads to an estimation of SWE similar to that obtained by sampling density at fixed heights in the snow pack (SWE method) (figg. 2 e 6).
- SWE measurements conducted by means of a snow corer (named EV sampler) lead to an underestimation of SWE (figg. 2 e 6).



1 The data

Sampling size

Seventeen couples of observers were involved, for a total of 186 measurements. Among those, 84 measurements were taken by sampling snow density at fixed heights along the snow profile (hereafter referred to as SWE method) and 84 by sampling snow density according to homogeneous snow strata (AINEVA method). 18 measurements were taken using a snow corer named EV sampler. 3 plots were defined (A, B, and C where 76, 70 and 40 measurements were taken, respectively) (Fig.1).

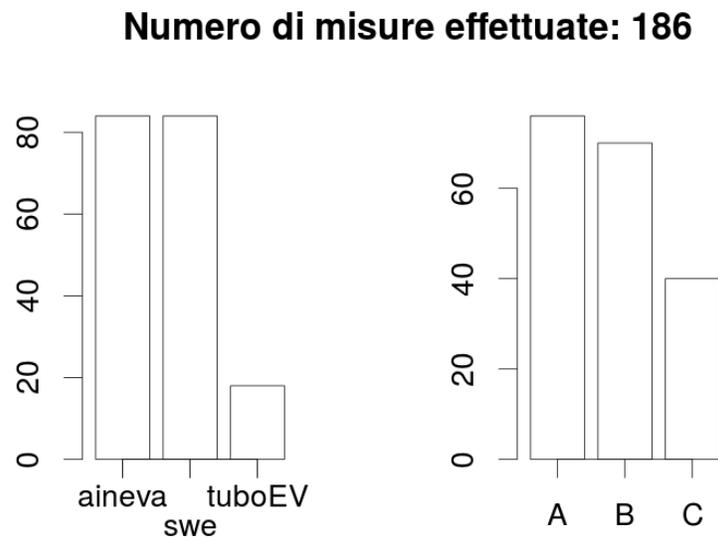


Figure 1: Number of measurements conducted with each method and on each plot



Averages across methods and plots

Figure 2 shows medians of SWE, snow depth (HS) and snow density (RHO) for the three methods. The median for each method was calculated including measurements from all couples of observers. Error bar represents the mean absolute deviation (MAD). The coefficient of variation (CV%) is also reported. The red line represents the median across all data. For SWE, we observe much lower values for SWE as measured with the EV sampler compared to other two methods. The within-method variability is similar across methods. The same is true also for snow density (RHO). Since we cannot decide *a priori* which method performs better, we assume that the global median is our ground truth.

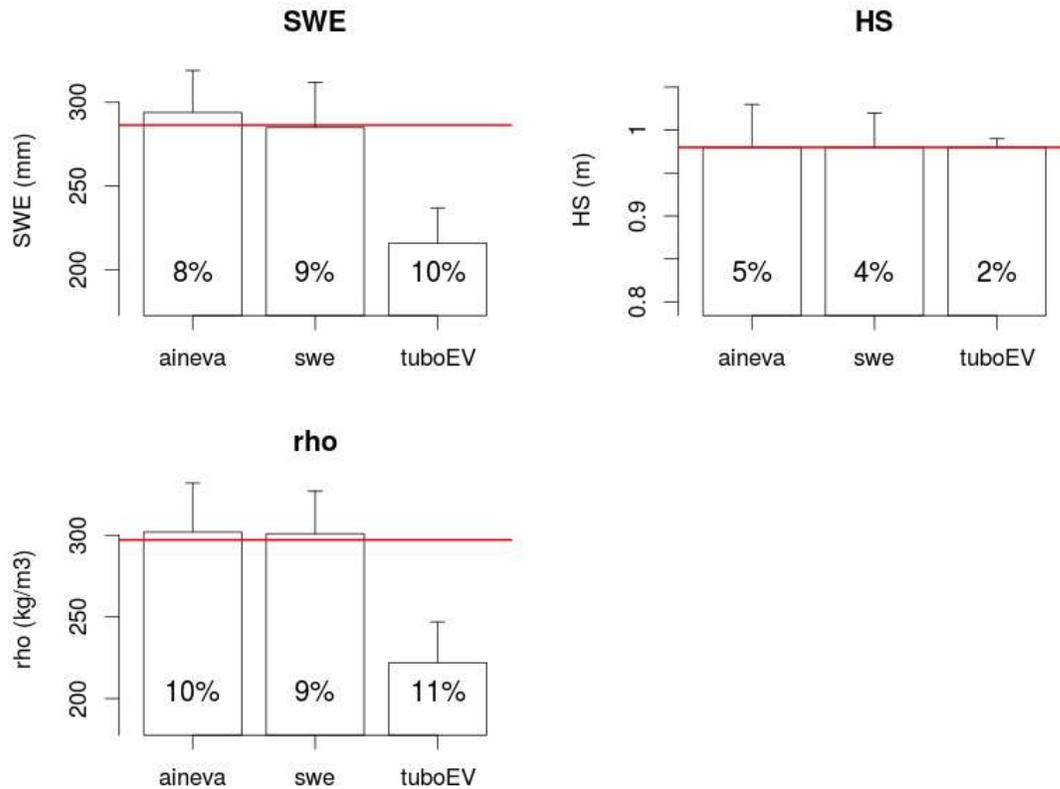


Figure 2: Medians of SWE, HS and RHO across methods. Error bar represents the MAD. CV is calculated as $MAD/median * 100$. The red line is the global median

Figure 3 is similar to fig. 2 for intra-plot variability. We observe lower snow depths in C, but with higher snow densities. The two divergent behaviors of HS and RHO compensate each other in SWE, which shows no spatial variability. However, CVs are higher between plots than between methods.

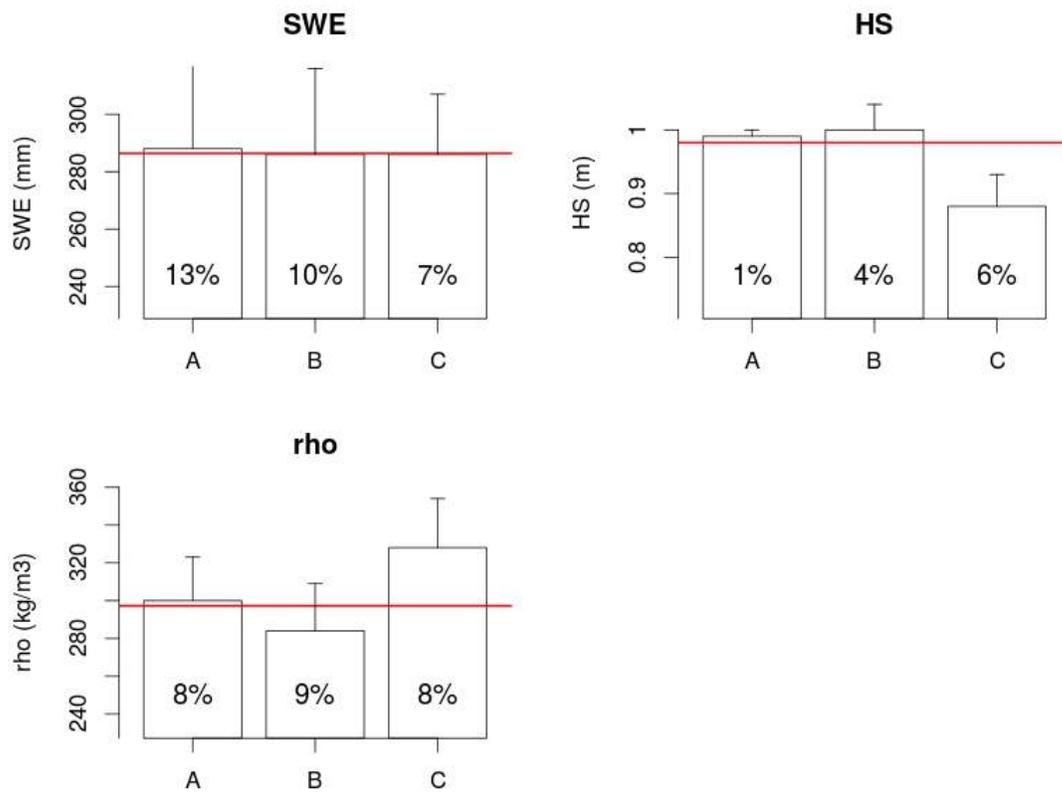


Figure 3: Medians of SWE, HS and RHO across plots. Error bar represents the MAD. CV is calculated as $MAD/median * 100$. The red line is the global median



Sources of error

The experimental design allows to disentangle SWE variability associated to four different factors: i) error with the couple of observers (observer), ii) variability between observers (inter-observer), iii) spatial variability and iv) variability across methods. Figure 4 shows the average error associated to each of the variability sources. Observer error was found to be 2.5%. This indicates that under similar snow conditions and using the same method, the same observer makes an average error of 2.5%. The probability distribution of observer error suggest that in only few cases the observer-error is higher than 5%. The inter-observer variability represents the largest source of error, around 7%. It is interpreted as the mean variability across observations conducted by different observers in homogeneous conditions. Figure 5 shows all SWE measurements split by plot and method. It is evident that inter-observer variability is the largest source of error. Note however that spatial variability within the same plot is assumed negligible along the 12-15 m transect that constitutes each plot, which may be not the case. If so, we could be overestimating inter-observer variability. Spatial variability between plots is minimal, whereas the variability due to the method is the second source of error, and will be discussed later.



Fonte	Variabilità (%)
Operatore	2.5%
Interoperatore	7.1%
Spaziale (plot)	0.1%
Metodo	3.6%

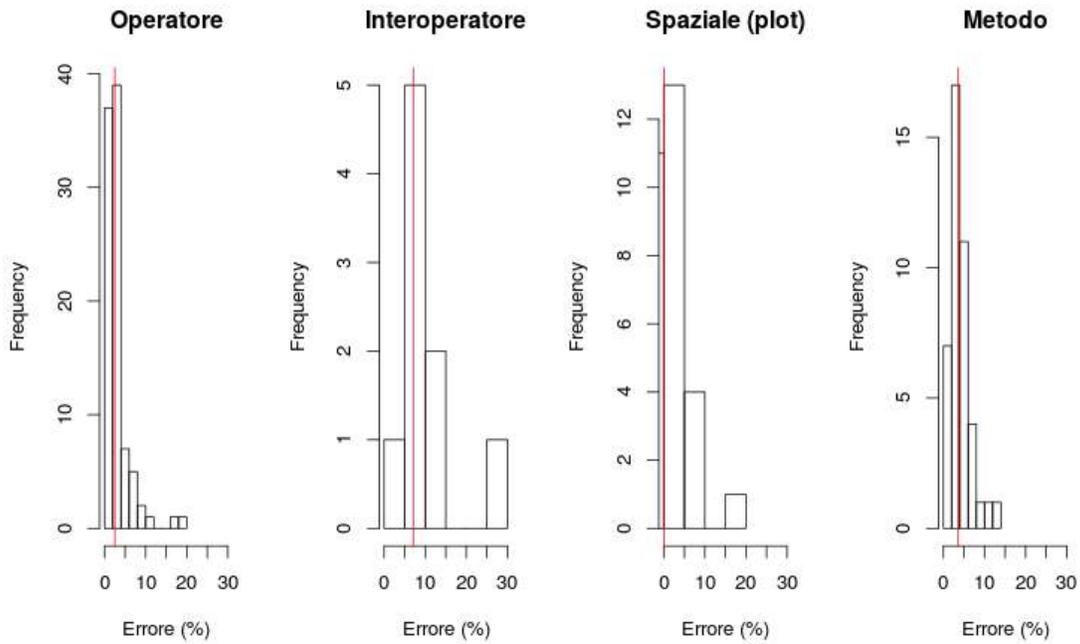


Figure 4: Mean error (%) associated to each variability source and probability density of the errors. Red line represents the median error

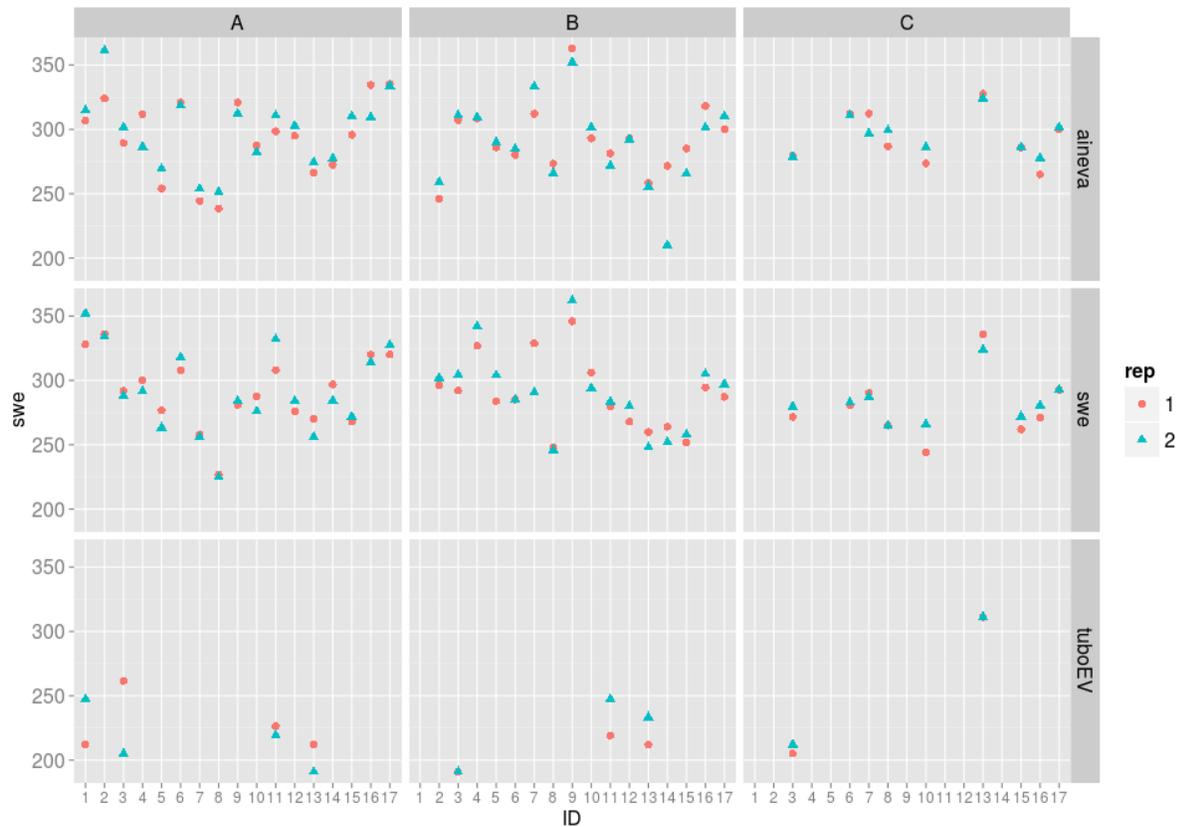


Figure 5: All SWE measurements, split by plots and methods. Different colors indicate replicates from every couple of observers.

Across-method and spatial variability

To test the hypothesis that different methods lead to different SWE estimates, and that spatial variability is not negligible, we used a mixed effect model with method and plots as fixed effects and observers as random effect. Results in fig. 6 show quantitatively what was



already apparent by simply looking at the plots, i.e. EV driller results in significantly lower SWE estimates, whereas there is no different among plots.

To further analyze the statistical model let's look at the Anova table (fig. 7) showing the effects of each factor on SWE. Method and plot effects were already discussed, but there is also an interaction between method and plots, indicating that the effect of the method is not the same across plots. Fig. 8 shows this feature. EV sampler shows lower medians in plot A and B but not in C. Note however that the sampling size for C is much lower than for A and B.

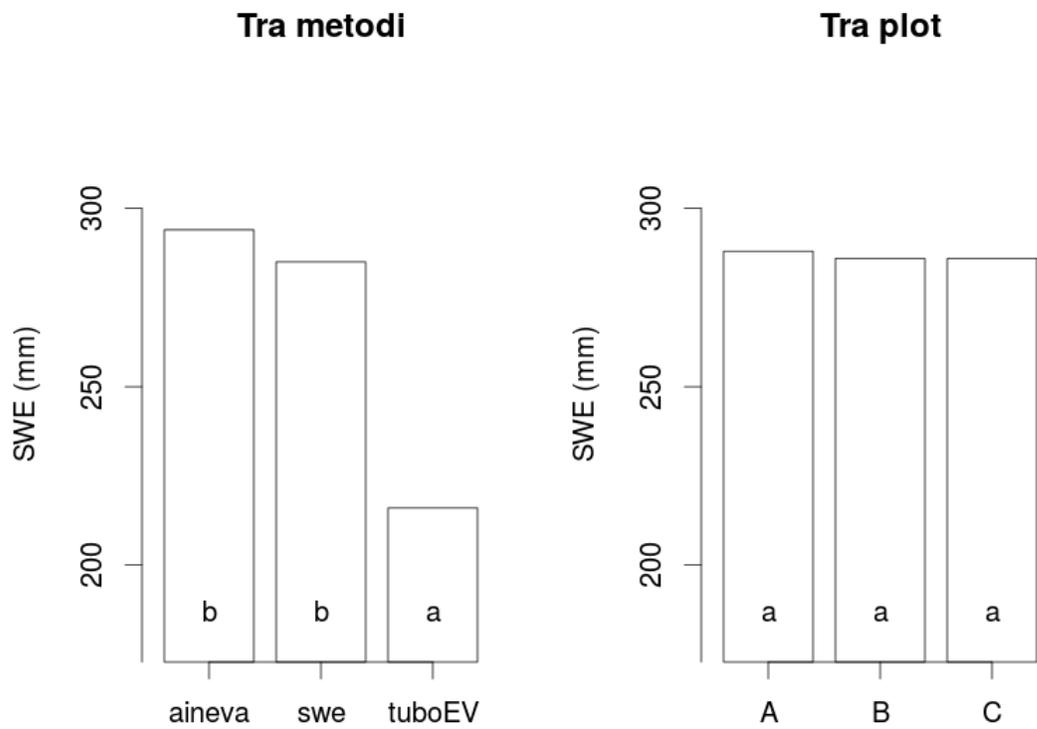


Figure 6: Results of the mixed model. Bars are the same as in fig. 2. different letters indicate significant differences between means.



Figure 7: Anova table of the model. p-values lower than 0.05 indicate significant effect of the given factor. Last line (method:plot) refers to the interaction between the two factors.

	F-value	p-value
method	37.089	0
plot	0.615	0.542
method:plot	2.7	0.033

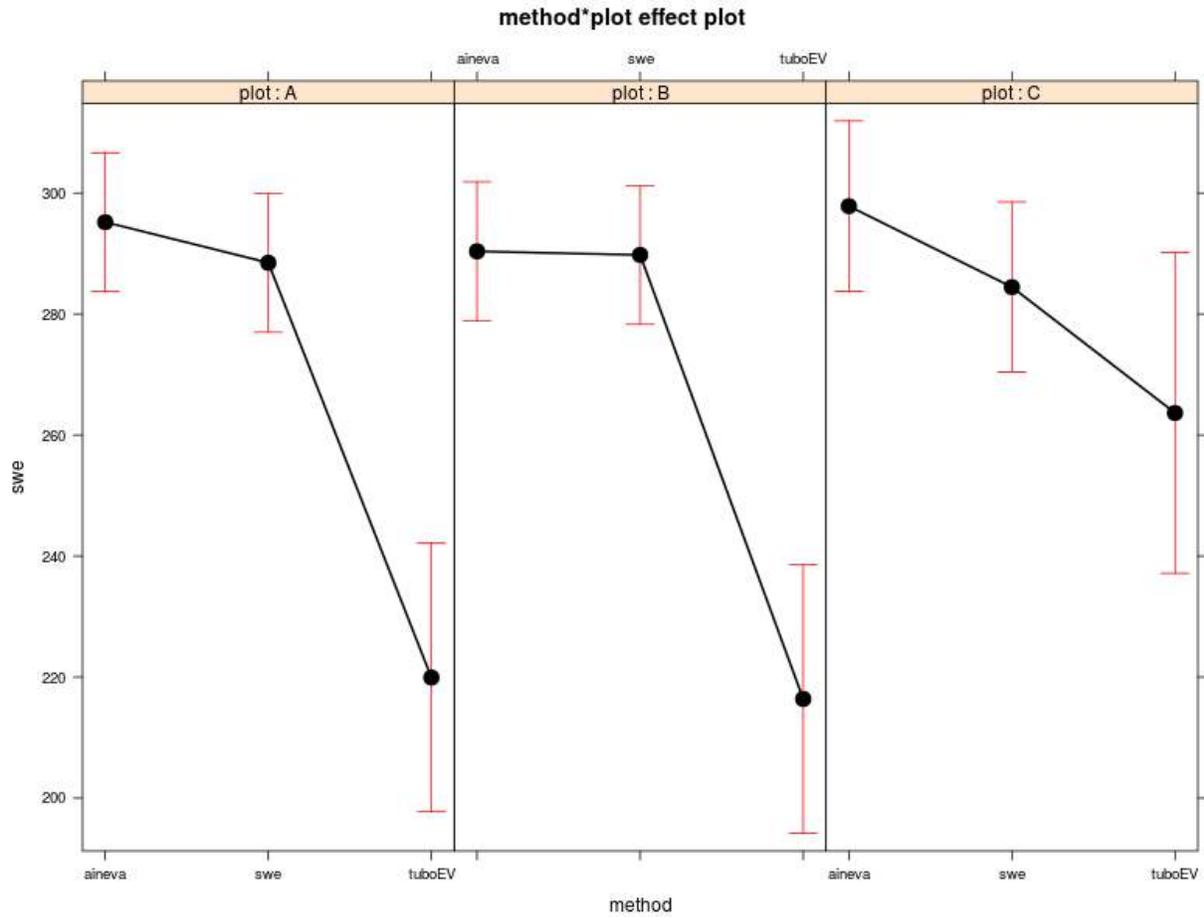


Figure 8: SWE means as a function of method and plot

We showed that the three methods lead to different SWE estimations, namely EV sampler results in a lower SWE estimate compared to the other two methods. Let's now evaluate whether any of the three methods lead to a SWE estimation statistically different from what we defined as the ground truth (i.e. the median across all observations, SWE median = 286



mm). The following table shows significance of a simple t-test. As expected, AINEVA and SWE method do not differ from the ground truth, whereas the EV sampler method results in an underestimation of the ground truth.

method	pValue
aineva	0.083
swe	0.602
tuboEV	0.001

Figure 9: p-values of the t-test used to establish whether any method is significantly different from the ground truth. p-values lower than 0.05 indicate a significant difference

General significance of results obtained

Our experiment offers interesting points of discussion, but it is worth remembering its limitations. In order to test for inter-method variability we tried to exclude excessive spatial variability by choosing a highly homogeneous site. Our results must therefore be extended with caution to markedly different snow conditions (e.g. more complex topography, wide range of snow depths and snow densities).

The unbalanced sample size between measurements conducted with AINEVA and SWE methods compared to the EV sampler may suggest that the differences we found might be an artifact. For this reason we conducted a re-sampling experiment, by randomly sampling



a number of measurements recorded with AINEVA and SWE methods equal to the number of samples taken with the EV method ($n=18$), to equal the sampling size (replicated 1000 times). Results show that in 99% of instances the conclusions would have been the same, i.e. EV sampler leads to an average SWE lower than with the other two methods.



2 Conclusion

In this filed inter-comparison, 186 SWE measurements were conducted on a relatively homogeneous, flat area of 1000 m². Measurements were taken with three different methods. We aimed at:

- i) evaluate error of a single observer
- ii) evaluate inter-observer variability
- iii) evaluate variability between methods.

Results show that observer variability lead to an error lower than 3%, whereas variability between different observers account for the largest source of error, 7%. Spatial variability was found to be the lowest, 1%. SWE and AINEVA methods lead to the same estimate fo SWE, whereas EV sampler results in an underestimate fo SWE.

We conclude that the high inter-observer variability observed points to the need for more such inter-comparisons, in order to get reliable and homogeneous measurements.

3 Supplementary

To have an overview on all measurements and not only on aggregated data, we show two plots similar to fig. 5, but showing snow depths and densities as measured by all couples of observers.

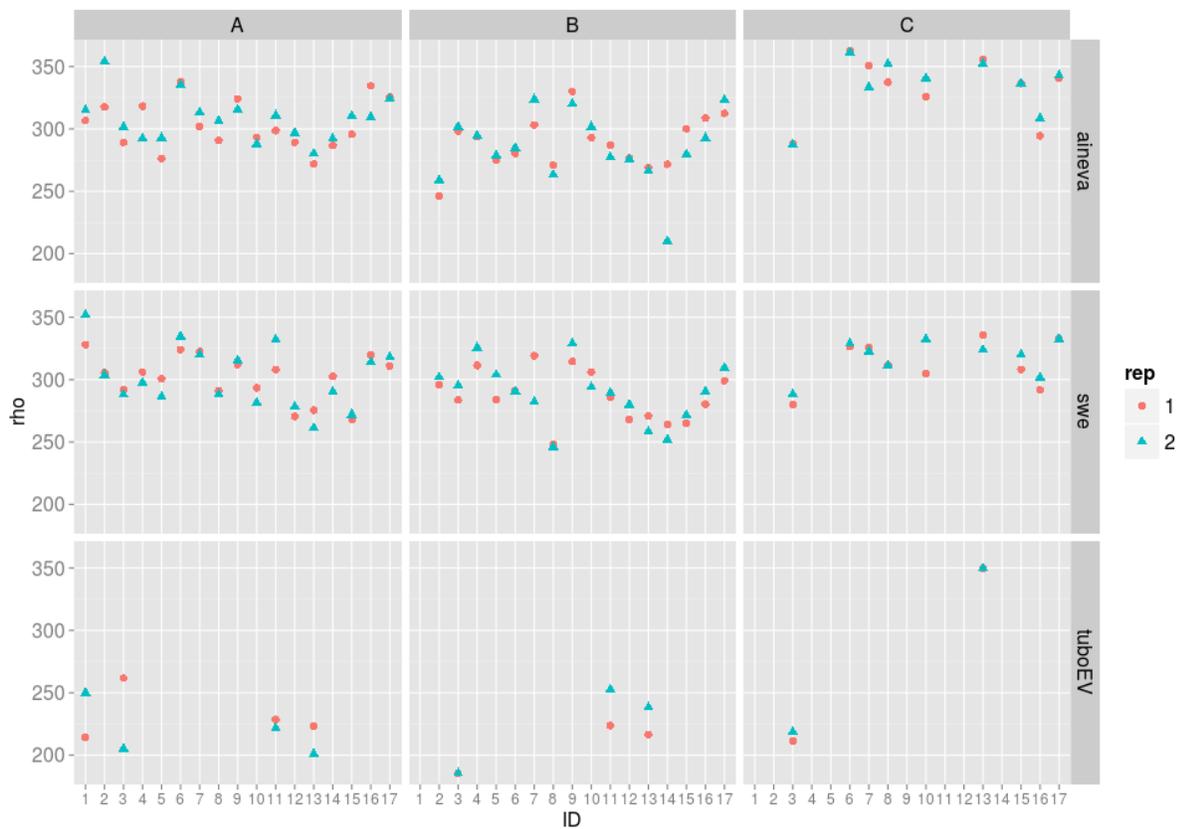


Figure 10: All RHO measurements (weighted average), split by plots and methods. Different colors indicate replicates from every couple of observers.

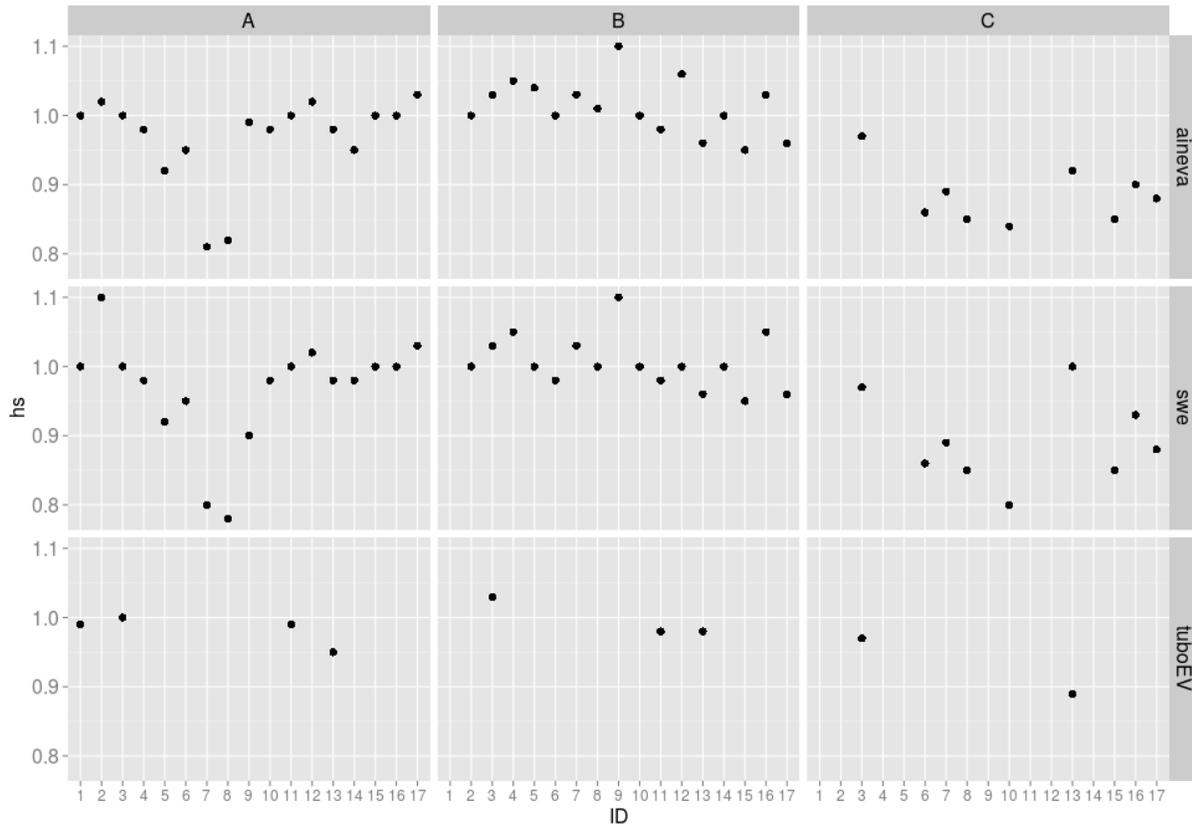


Figure 11: Same as before but for HS.



Second SWE inter-comparison

Santa Caterina Valfurva
March 3rd, 2016

Technical report by ARPA Valle d'Aosta



In a nutshell

- 184 SWE measurements with two different methods on the sides of 4 cubes (each cube, 5-m side), 24 snow observers involved (fig. 1).
- 48 SWE measurements conducted in 8 points across a heterogeneous area (itinerant campaign, fig 8).
- A couple of two observers tacking SWE measurements on a homogeneous snow pack results in an average error of 3% (fig. 4).
- Twelve couples of observers measuring 4 homogeneous cubes result in an error of 10% (fig. 4).
- The two methods (SWE method and EV sampler method) result in similar SWE estimates
- In homogeneous conditions (snow depth ranging 1.20-1.50 m, spatial variability 7%), 3 sampling points are enough to properly characterize SWE in one area, either with SWE and EV methods (fig. 7).
- In heterogeneous conditions (snow depth ranging 0.80 e 1.40 m, spatial variability 35%) 6 sampling points are needed to characterize SWE with EV method (fig.14)



1 Sampling strategy

Sampling occurred in two ways: a) with a campaign based on intensive measurements on 4 cubes of 5-meter-side, hereafter referred to as sedentary campaign; b) an itinerant campaign along a path, where 8 sampling points with markedly different snow depths were identified and sampled. In both cases, two methods for SWE determination were used: the method called SWE, where snow density was sampled at fixed heights in the snow pack with a 500 mL tube; and the EV method, with a vertical corer that allows an integrated sampling of the whole snow pack.

2 Sedentary campaign

Sample size

Sedentary campaign was conducted around 4 cubes of 5 meters side. Twelve couples of observers took part to the experiment (184 measurements in total, 92 with the EV method and 92 with the SWE method) (Fig.1). Within each couple, one observer consistently conducted measurements with SWE method and the other with the EV method. We can therefore test also for the variability of one single observer. All couples conducted measurements in the cubes A-C and only 10 of them took measurements at cube D.



Numero di misure effettuate: 184

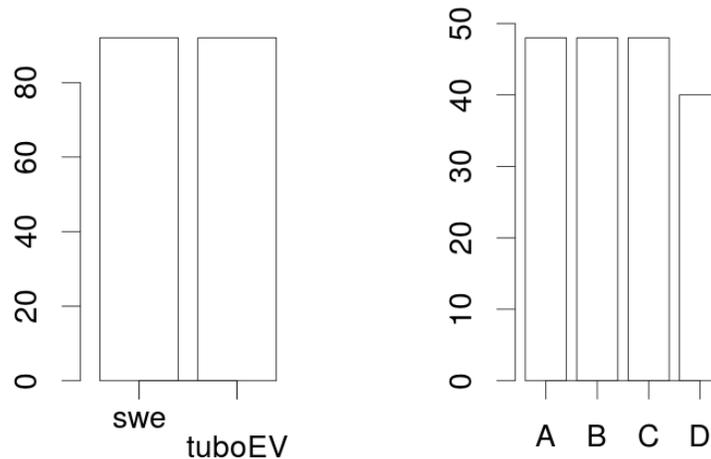


Figure 1: Number of measurements for each cube and method

Method and plot averages

Figure 2 shows medians of SWE, HS and RHO for the two methods. Error bar represents the mean absolute deviation (MAD). The coefficient of variation (CV%, calculated as $MAD/median \times 100$) is also included. The red line represents the general median. Median SWE as measured by the EV method is slightly higher than that measured by SWE method. Furthermore, EV method leads to a higher variability (13%) compared to SWE method (7%). The general median SWE is 393 mm and is considered the best estimate of the ground truth.

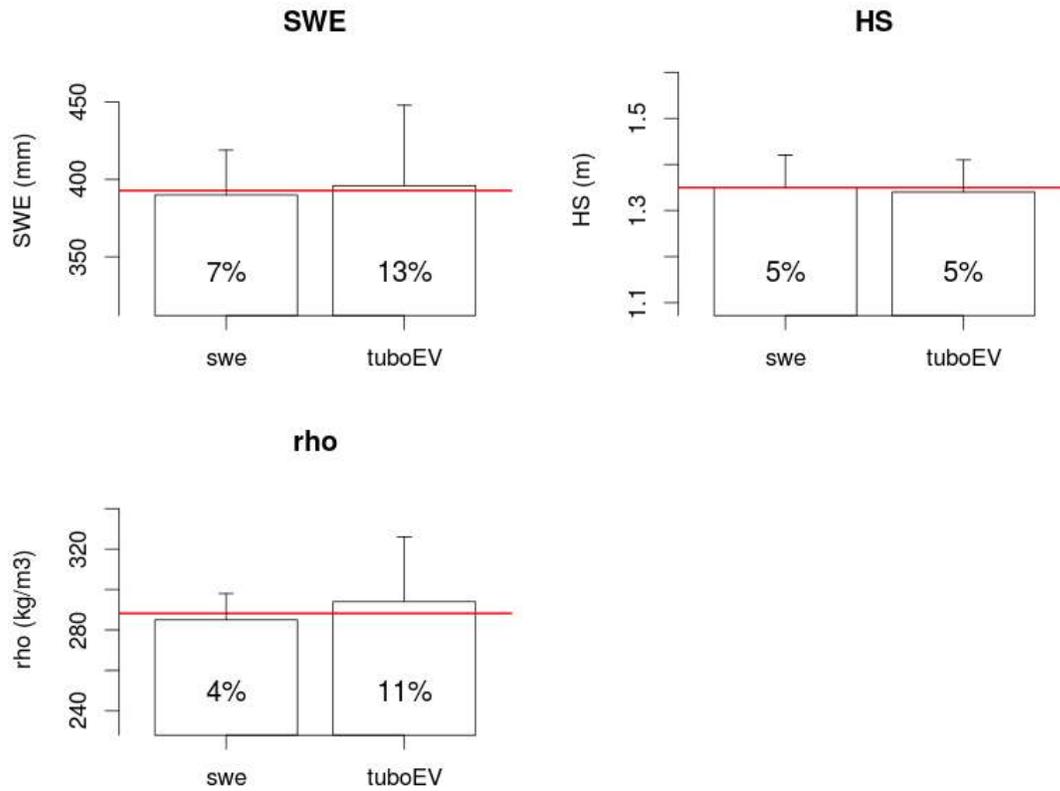


Figure 2: Medians of SWE, HS e RHO as a function of the method used. Error bar is the MAD. The red line is the general median.

Figure 3 is as fig. 2 for the between-cubes variability. A significant different SWE was found between cubes, ranging from 372 (cube B) to 424 mm (cube A).

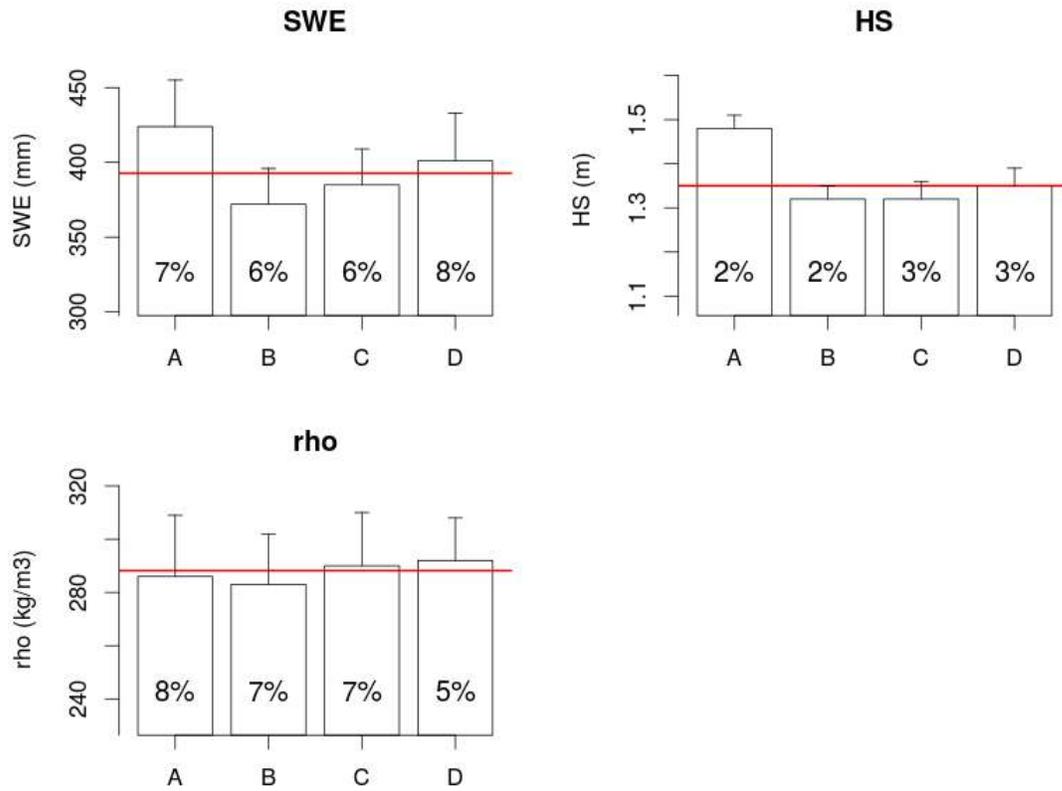


Figure 3: Medians of SWE, HS e RHO in the four cubes. Error bar represents the MAD. The red line is the general median.



Error associated to the examined factors

The experimental design allows to disentangle the SWE variability associated to 4 different factors: i) observer error, ii) inter-observer variability, iii) spatial variability, iv) variability due to method. Figure 4 shows the median error associated to each of these variability sources. The average error associated to a single observer is 2.3%, i.e. 9 mm SWE. The inter-observer variability is the largest source of error, at 10%. This variability indicates the error encountered by different observers sampling very similar snow packs.

Spatial variability between cubes is 7.4%, and variability across methods is 6.6%. Compared to results obtained during the inter-comparison 2015, where couples were distributed along a 12-15 m long transect, the conditions here are more homogeneous. Hence, the 10% inter-observer error may be considered a good estimate of the variability between different observers and we can exclude that unaccounted spatial variability is affecting the estimated inter-observer variability, as it was argued after the 2015 inter-comparison.



Fonte	mm SWE	%
Operatore	9	2.3%
Interoperatore	39	9.9%
Spaziale (plot)	29	7.4%
Metodo	26	6.6%

Figure 4: Error % associated to each variability source.



Variability between methods and spatial variability

To test the hypothesis that different methods lead to different estimations of SWE, and that spatial variability exists, we used a mixed effect model with method and plot (including interactions) as fixed effects and observers as random effect. This model allows to single out variability sources related to method, space, inter-observer and observer. Results in fig. 5 show that the small difference in SWE as measured by SWE method and EC method is statistically significant. Also spatial variability is significant, with cube A having a larger SWE compared to other plots.

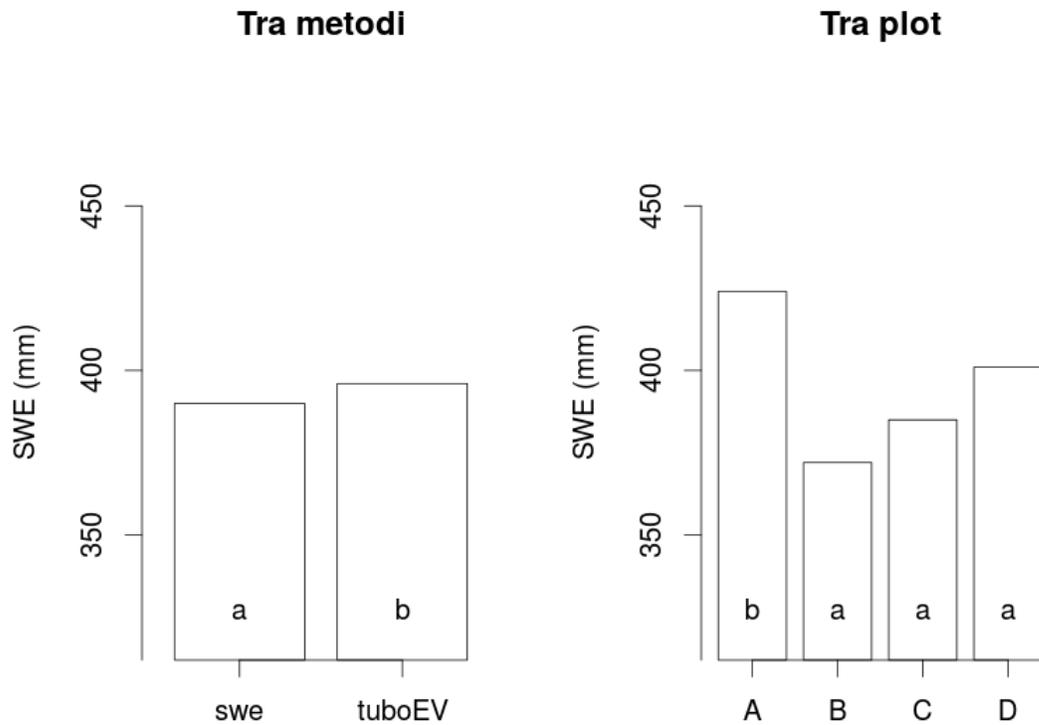


Figure 5: Results of the mixed effect model. Bars are the same shown in fig. 2. Different letters indicate significant differences between the means.

As we have shown that the two methods result in slightly different estimations, we test whether the means obtained by the methods differ from our general median, which we consider the ground truth. The following table shows the results of the t-test of the method-means against the ground truth. Both methods lead to an estimate that does not differ from the ground truth. The fact that we find statistical differences between the two methods, is



due to the high homogeneity of the data. In such conditions, minimal differences combined with the high number of observations can lead to statistically different results. Anyway, the important message is that both methods lead to a correct estimation of the ground truth.

method pValue	
swe	0.93
tuboEV	0.803

Figure 6: p-values of the t-test used to compare method-means to the ground truth. p-values lower than 0.05 indicate a significant difference between the means.

Re-sampling experiment

Previous sections demonstrate that: 1) both methods are robust and correctly estimate the ground truth and (2) observer variability is low. In order to translate the experiment into operational protocols it is worth asking: how much can we reduce the sampling effort without losing statistical power in properly determining the ground truth? We tried to answer this question with a re-sampling technique, by randomly removing some observations and run the statistical tests on randomly created subsets of the data. Results are then expressed in probabilistic form.

Briefly, the procedure is the following:



- The number of measurements to be sampled is defined (n of observations)
- We randomly sample from the complete dataset the n of observations for 1000 times
- We execute a t-test for each replication, to test whether the random subset differs from the ground truth
- We express the number of significant /not significant t-tests in probabilistic terms.

Results are shown in fig.7. The graph shows the probability (% , x-axis) to obtain an SWE estimate equal to the ground truth as a function of sample size (y-axis). Different colors represent the two methods. Two and three samples are sufficient with the SWE method and the EV method, respectively, to correctly estimate the ground truth at a 95% confidence (a typical statistical threshold). In other words, in face of a homogeneous snow field 2-3 samples would give us a 95% probability of hitting the true SWE of that area.

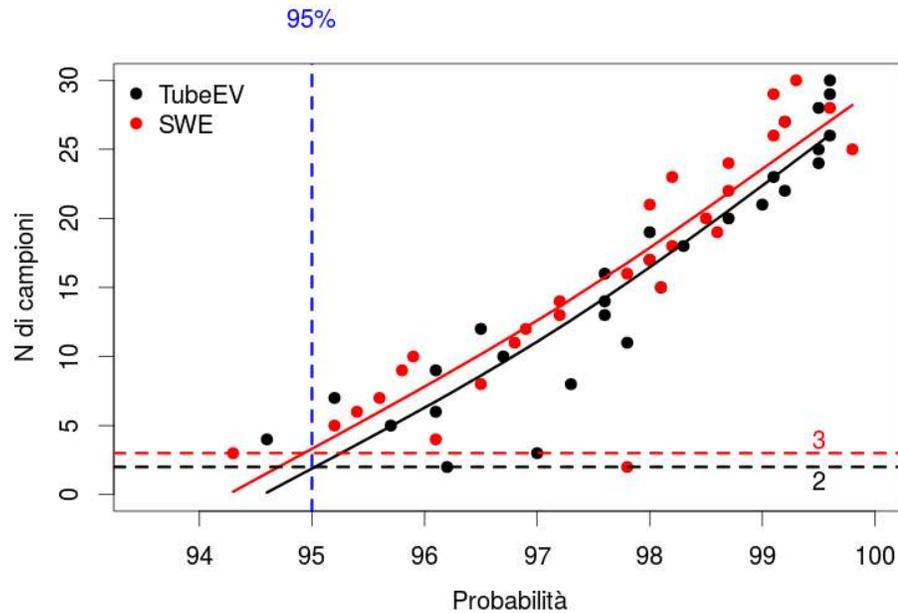


Figure 7: Results of the re-sampling experiment. Variation of the probability to obtain a correct estimate of the SWE as a function of number of sampled included.

3 Itinerant campaign

To test the two methods under less homogeneous SWE conditions compared to the flat snow field, we organized an itinerant campaign across different slopes and snow conditions. Sampling was conducted in 8 points with the two methods. For each point one observation was made with the SWE method (more time consuming) and 5 replicates with the EV



method.

SWE medians for the 8 points are shown in fig 8. As expected, compared to the sedentary campaign the spatial variability is higher. Similarly to the sedentary campaign, we calculated the variability associated to the method (10%), comparable to the one observed in the sedentary campaign, and the spatial variability (34%). Hence, we fulfilled the purpose to capture more spatial variability. A surprising result, partially in contrast to the sedentary campaign, is that EV method leads to lower SWE estimates compared to the SWE method. We can hypothesize that a snow loss at the base of the EV corer could be responsible for the discrepancy between the two methods.

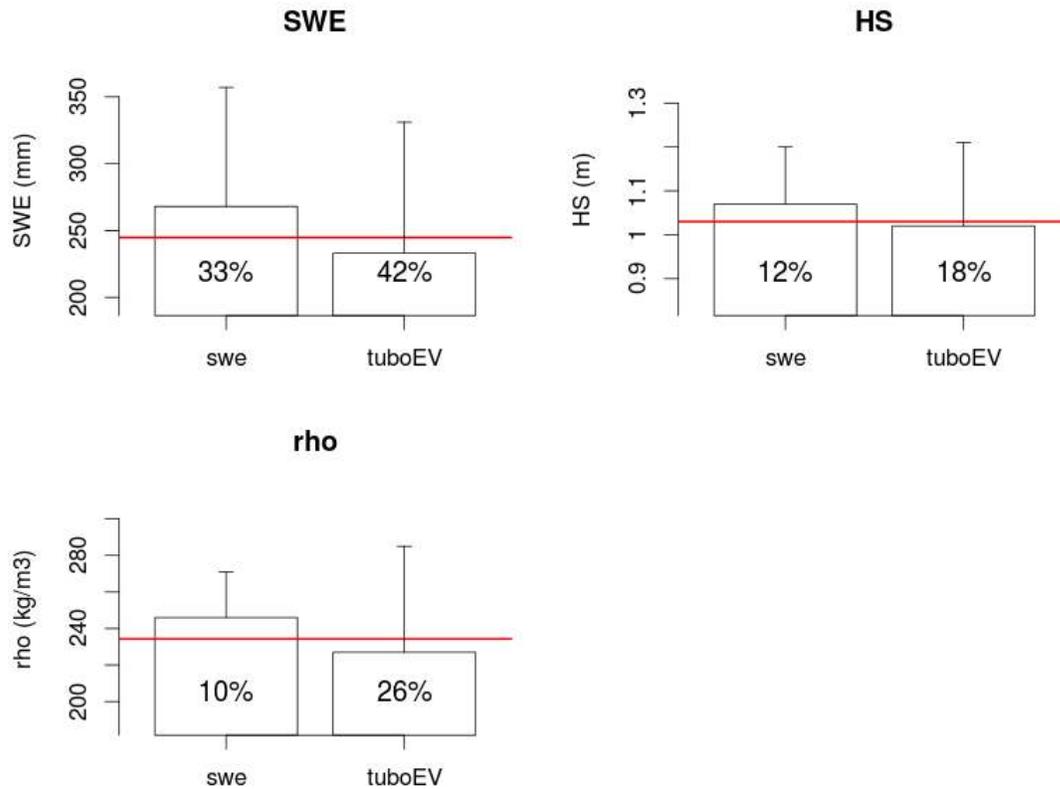


Figure 8: Median SWE, HS and RHO as a function of methods. Error bar is the MAD. The red line is the general mean.

In fig. 9 we show the median SWE observed at each of the 8 points. Along with the general mean (red line), we also show the median of the sedentary campaign as a reference (blue line). We observe that at each point the SWE median is lower compared to the flat terrain of the sedentary campaign, because the flat morphology promotes snow accumulation whereas the slopes of the itinerant campaign do not. The variability in SWE at the 8 points is



a function of different snow depths (ranging 0.86-1.33 m), but also of different snow densities (170-300 kg/m³).

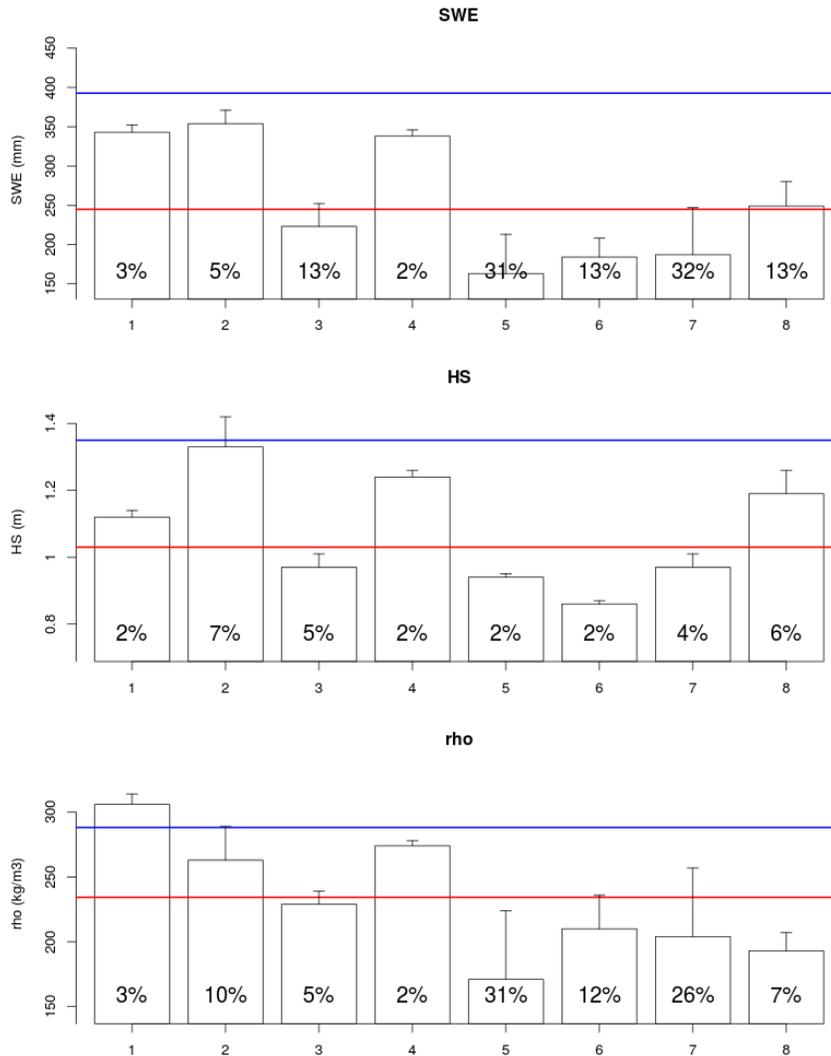


Figure 9: Medians of SWE, HS and RHO across the 8 sampling points. The red line is the general mean. The blue line is the mean of the sedentary campaign, for reference.

The mixed model on the itinerant campaign shows what was already apparent by simply



looking at the bars. We observe a statistically significant difference across methods and in space (fig. 10).

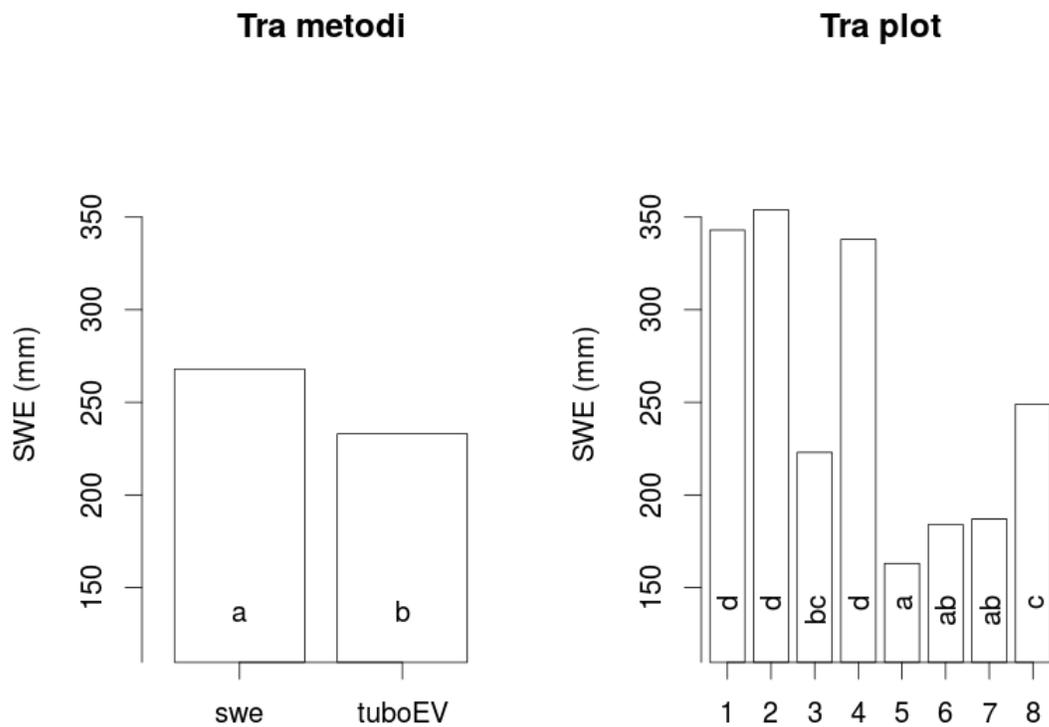


Figure 10: Results of the mixed model. Bars are the same as in fig. 8. Different letters denote significant differences between the means.



Re-sampling experiment on the itinerant campaign.

Similarly to the sedentary campaign, we tried to estimate the minimum number of sampling points required to quantify the ground truth.

In this case, however, we have to take into account that the 8 points are markedly different. We therefore formulate two different cases.

a) we have to quantify the SWE at the scale of a small slope where we have settled our 8 sampling points. The ground truth will be represented by the general mean. In such a heterogeneous landscape, how many points do we need to correctly quantify the ground truth?

With a re-sampling analysis we progressively reduce sample size and calculate the probability to correctly estimate the ground truth. Results are shown in fig. 11. If we consider a threshold of 95% confidence, 6 or more samples (EV method) are needed to correctly estimate the ground truth. This means that an observer can randomly distribute 6 sampling points with the EV method to properly characterize the SWE at this slope scale.

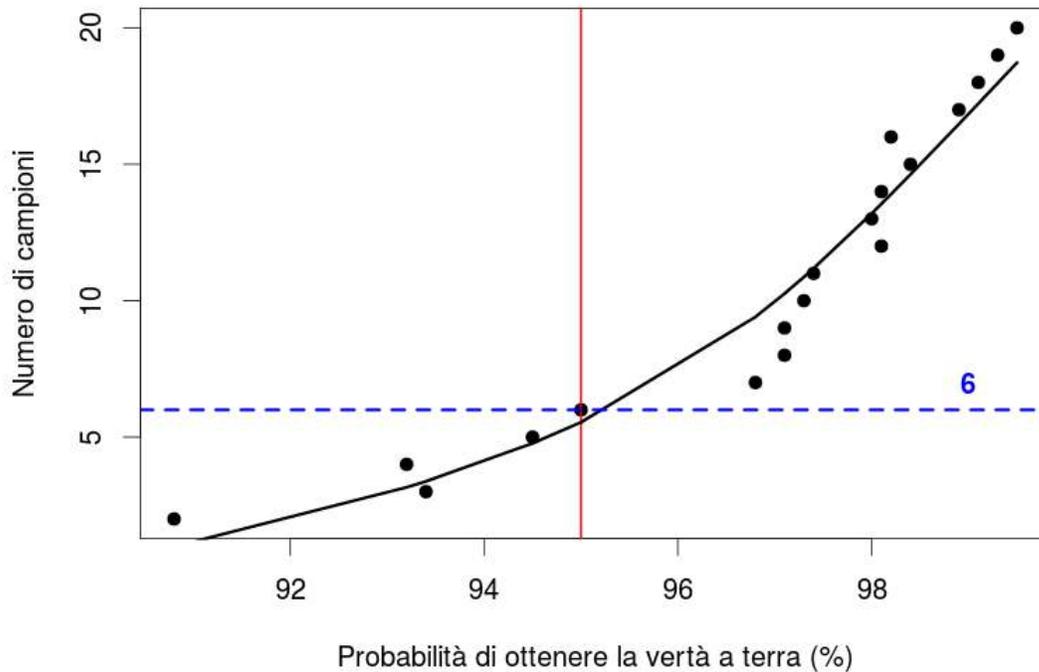


Figure 11: Re-sampling analysis: probability to correctly quantify the ground truth as a function of the number of sampling points.

Across the 8 points we identified markedly different SWE conditions. To better exploit this information, we could consider these 8 points as representative of different snow conditions. In particular, we classified the points in 3 SWE-abundance categories. In fig 13 we show that the three groups have different means.

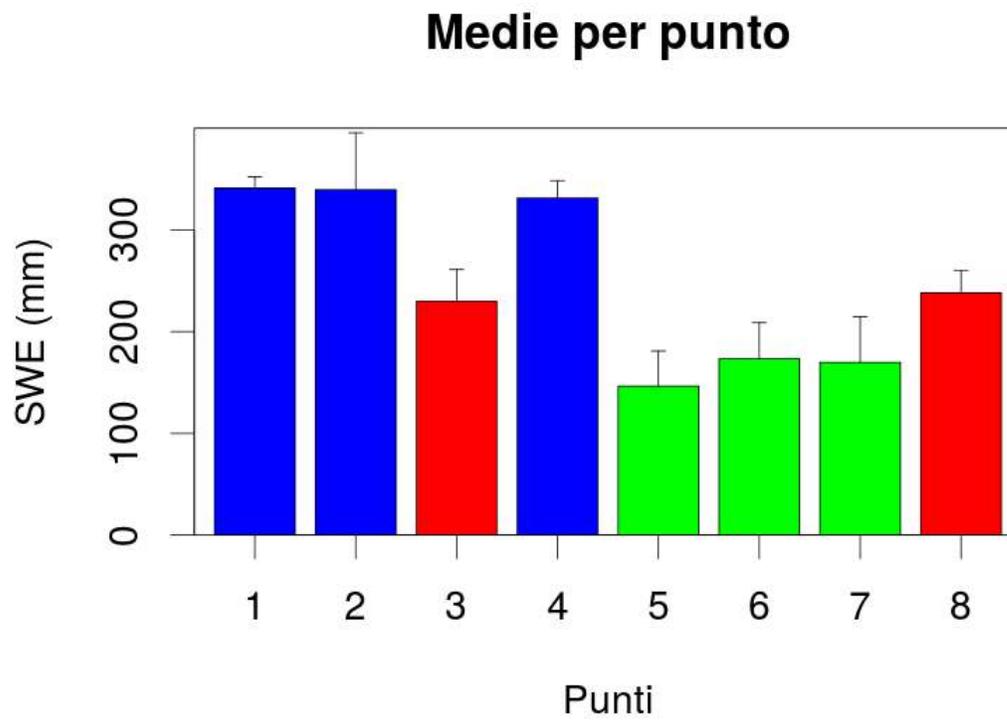


Figure 12: Average SWE in the 8 points of the itinerant campaign. Different colors show the classification according to SWE abundance, in abundant (blue), average (red) and scarce (green).

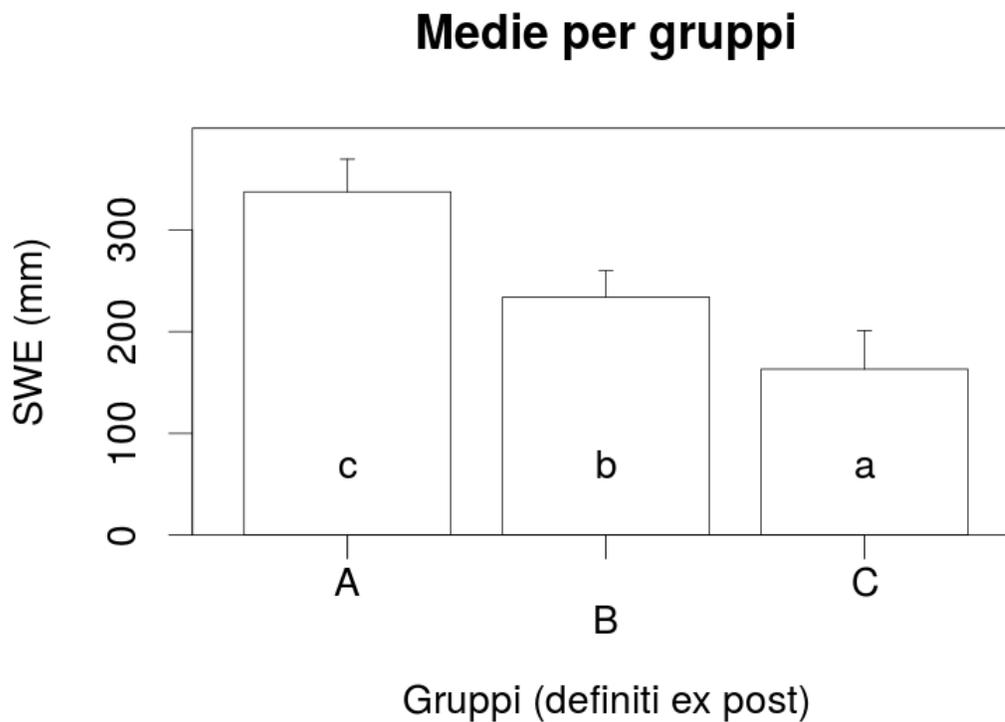


Figure 13: Average SWE for the three groups identified. Different letters indicate significant differences.

The question now is: how many samples do we need to be able to identify the differences between these three groups? Again, with a re-sampling experiment we calculate the probability to be able to maintain the reciprocal differences between groups (fig. 14). For each group we need a minimum of 6 observations to be able to preserve the variability observed with the full dataset.

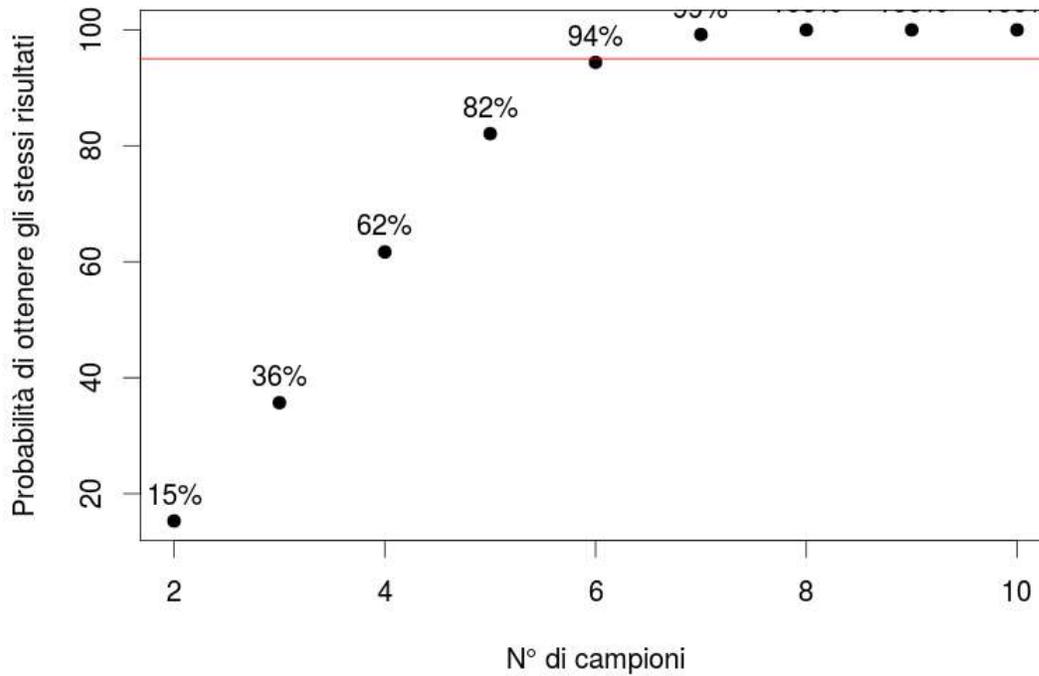


Figure 14: Variation of the probability to have the same between-group differences observed with the whole dataset, as a function of number of replicates for each of the three groups.

Results of the re-sampling suggest several operational implications on the use of the EV method under heterogeneous conditions.

- (1) To characterize a heterogeneous snow pack with snow depths ranging 80-130 cm, we need to randomly select 6 distinct points.



(2) If the objective is to characterize different situations in one area (e.g. with a clear sub-area of snow accumulation, one of snow erosion, and one with intermediate conditions), it is advisable to collect 6 samples for each of the sub-areas. In this case the observer could identify the subareas by means of fast sampling of snow depths to identify the points where the EV corer will be used.

(3) In heterogeneous conditions, we suggest that 6 is the optimal sampling size.



4 Conclusions

In this inter-comparison we conducted 184 measurements of SWE on a homogeneous surface taking advantage of the sides of 4 snow cubes (5 m per side). Moreover, 8 points were sampled to describe a typical heterogeneous situation (48 samples in total). Measurements were conducted with two different methods (SWE method and EV method)

Our objectives were:

- i) evaluate variability (and error) of a single observer.
- ii) evaluate variability (and error) between different observers.
- iii) evaluate variability (and error) between methods.
- iiii) evaluate the performance of EV method in heterogeneous conditions.

SWE and EV methods lead to similar SWE estimates.

In homogeneous conditions (hs ranging 1.20-150 m, spatial variability 7%), 3 sampling points are sufficient to characterize SWE for the area with either method. In heterogeneous conditions (hs ranging 0.80-1.40 m, spatial variability 35%), 6 points are required to characterize SWE with the EV method. With a preliminary, fast survey of snow depths, sub-areas can be identified, with different snow accumulation dynamics. For each of those area, 6 samples are required to characterize SWE.



Third SWE intercomparison

Breuil-Cervinia
April, 5-6 2017

Technical report edited by ARPA Valle d'Aosta



In a nutshell

- 2707 snow depth (HS) measurements were conducted in the Goillet Basin (Valtournenche Municipality): 6,3 km², minimum elevation 2.527 m a.s.l., maximum elevation 3.480 m (Testa Grigia).
- 36 snow observers involved, 16 measurement teams for snow depth, one team for snow density.
- 16 independent SWE simulations + one simulation with all sampling data (considered as ground truth, fig. 2).
- Total SWE in the basin was estimated at 6.3 millions m³.
- The variability across simulations for single teams varies between -30 and +20% compared to ground truth.
- Higher uncertainty in SWE estimation occurs in areas with less ground observations.
- The morphological parameters more influential on SWE distribution were those related to solar radiation.



1 Objectives 2017

After the comparison between different methods for SWE estimation under homogeneous snowpack conditions (inter-comparisons 2015 and 2016), the 2017's edition aimed at **evaluating the impact of widespread snow depth (HS) measurements on the water resource estimation for hydro-electrical power production in a medium-sized alpine basin** and in particular:

- 1) evaluate the impact on total basin SWE budget;
- 2) evaluate the impact on SWE spatial distribution.
- 3) refine current sampling protocols in the light of the results.

First objective is aimed at the quantification of the water resource directly responding to the specific need of hydro-electrical power production. The second, more specific and methodological objective consisted in the evaluation of the spatial distribution of SWE and in establishing which factors lead to larger differences in SWE distribution. The third objective, practical and operational, sought at improving the current sampling protocol.

2 Field work

Snow depth was sampled by 16 teams. Each team had two people, one directly sampling snow depth, the other registering the measurement. For data acquisition, an ad hoc smart-phone app (snowalp) was used, developed by ARPA VdA. This app allows to register snow depth associated to geographic coordinates thanks to the GPS integrated into the smart-phone. Each team was given a map with Goillet basin divided into 6 sectors (fig. 1). Each team was requested to visit as many sectors as possible, following a sampling scheme based on their own experience and skills, and sampling snow depth approx. each 100 m, with 250-300 measured points as target. Each team was asked to sample maximum variability in terms of elevation, facing, slope, morphology (erosion

vs accumulation areas), registering 0 values too in areas where snow was absent (if any).

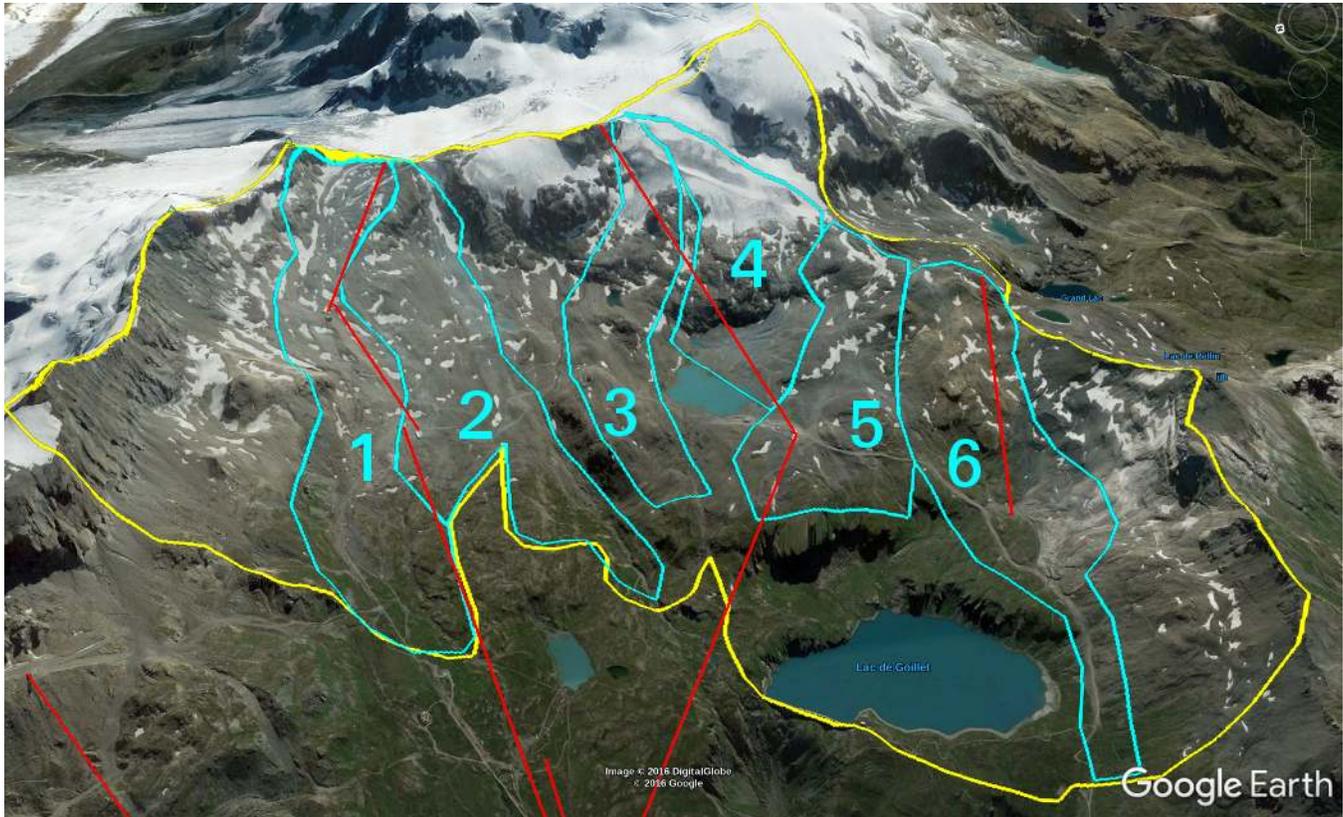


Figure 1: Goillet basin with the 6 sectors.

3 SWE simulations

The model used by ARPA for the estimation of the spatial distribution of SWE is structured in 3 steps.

- 1) Collection of various snow depth and snow density measurements;
- 2) Calculation of point SWE;
- 3) Grid distribution of SWE by means of multiple regression with dem-derived variables (elevation, slope, distance from mountain top, solar radiation, etc.) and subsequent spatialization of the resid-



uals (regression-krieking).

The final product is a SWE map for the whole basin at 500 m resolution.

Snow density

One team conducted snow density measurements coupled with a TDR measurement at 6 points. The resulting values were averaged and a basin-average snow density (280 kg m^3 , SD 39 kg m^3) was used as model input.

Snow depth

Figure 2 shows the distribution of measured HS. More than 2770 measurements were conducted across the whole Goillet basin. Some areas were not sampled or partly sampled for security reasons (glacier fields, or avalanche release area), or logistic reasons (the flat area between the two lakes was flat and therefore excluded from the study area).

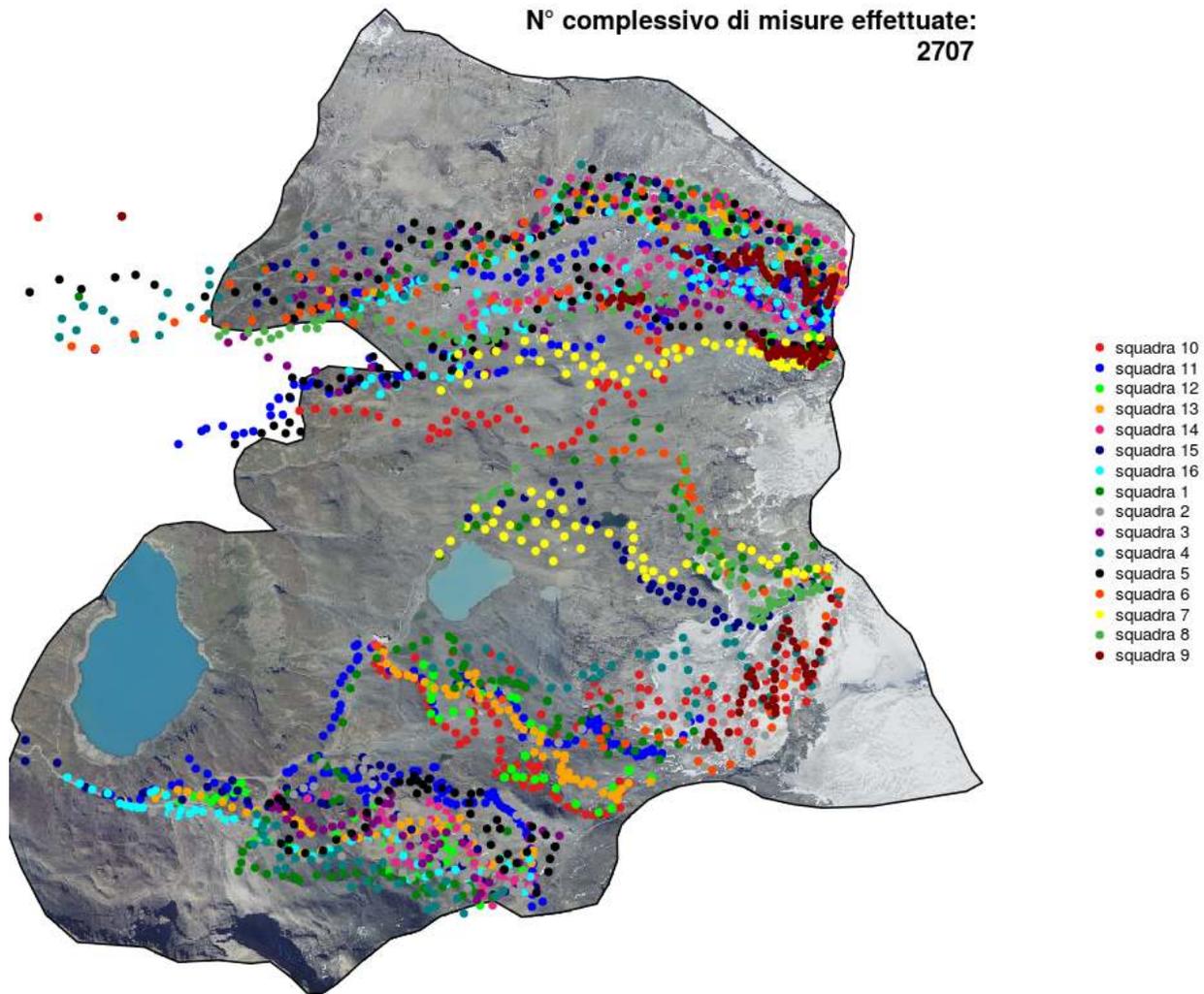


Figure 2: Spatial distribution of snow depth measurements

Figure 3 shows the same points as in fig 1, with color mapped on snow depth, in order to highlight the spatial distribution of snow depths.

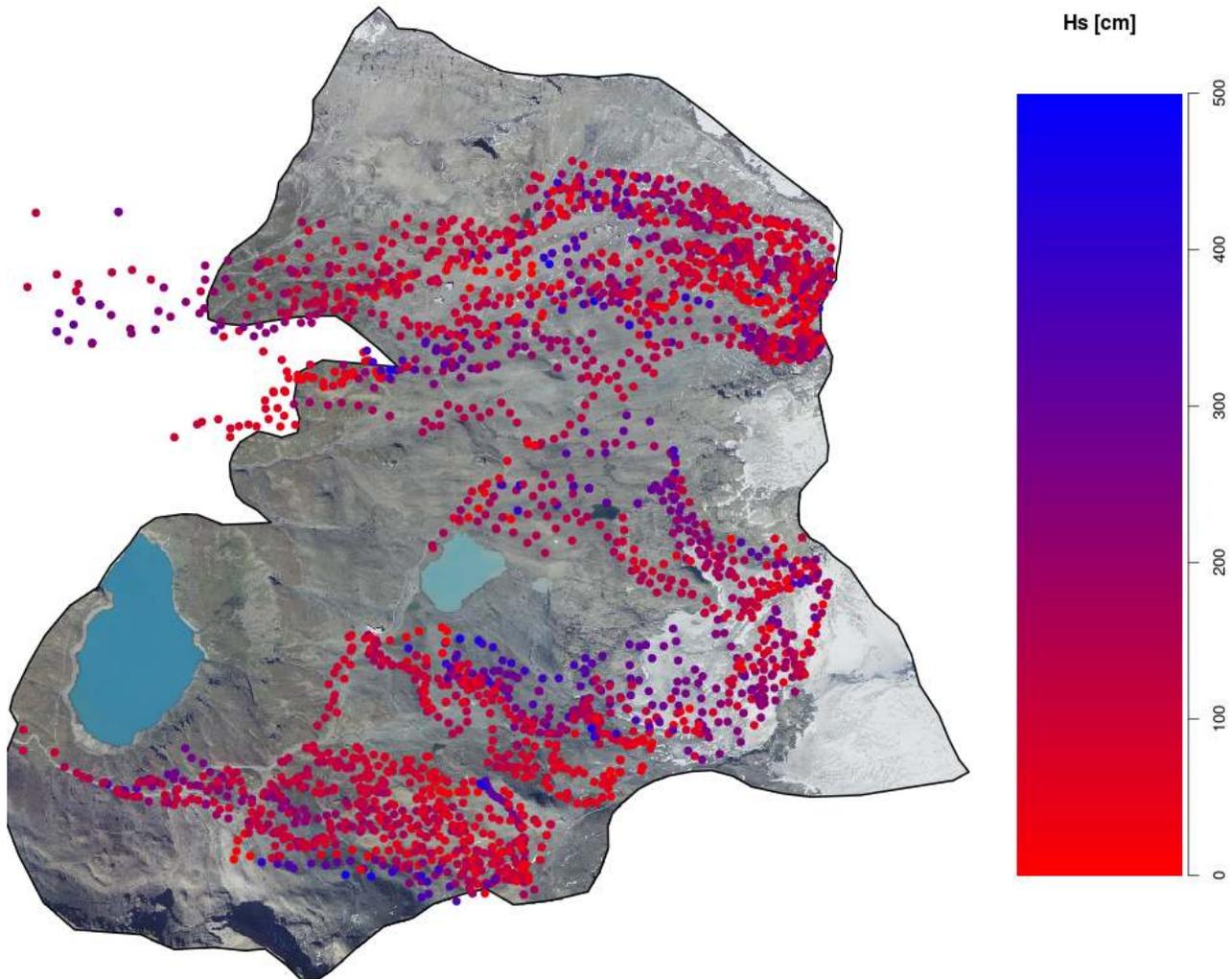


Figure 3: Spatial distribution of snow depth measurements with color mapped on snow depth values.

Figure 4 shows the variability of snow depth with a boxplot for each team. Median HS equals 158 cm, with a minimum of 0 and a maximum of 500 cm. This is not surprising, since it include the whole basin's variability. However, we also observe a significant inter-team variability by looking at the respective medians, that vary by more than 100 cm. For example, team 11 measured a median snow depth of 100 cm whereas team 10 scored a median of 200 cm, due to the different sampling

paths they have chosen.

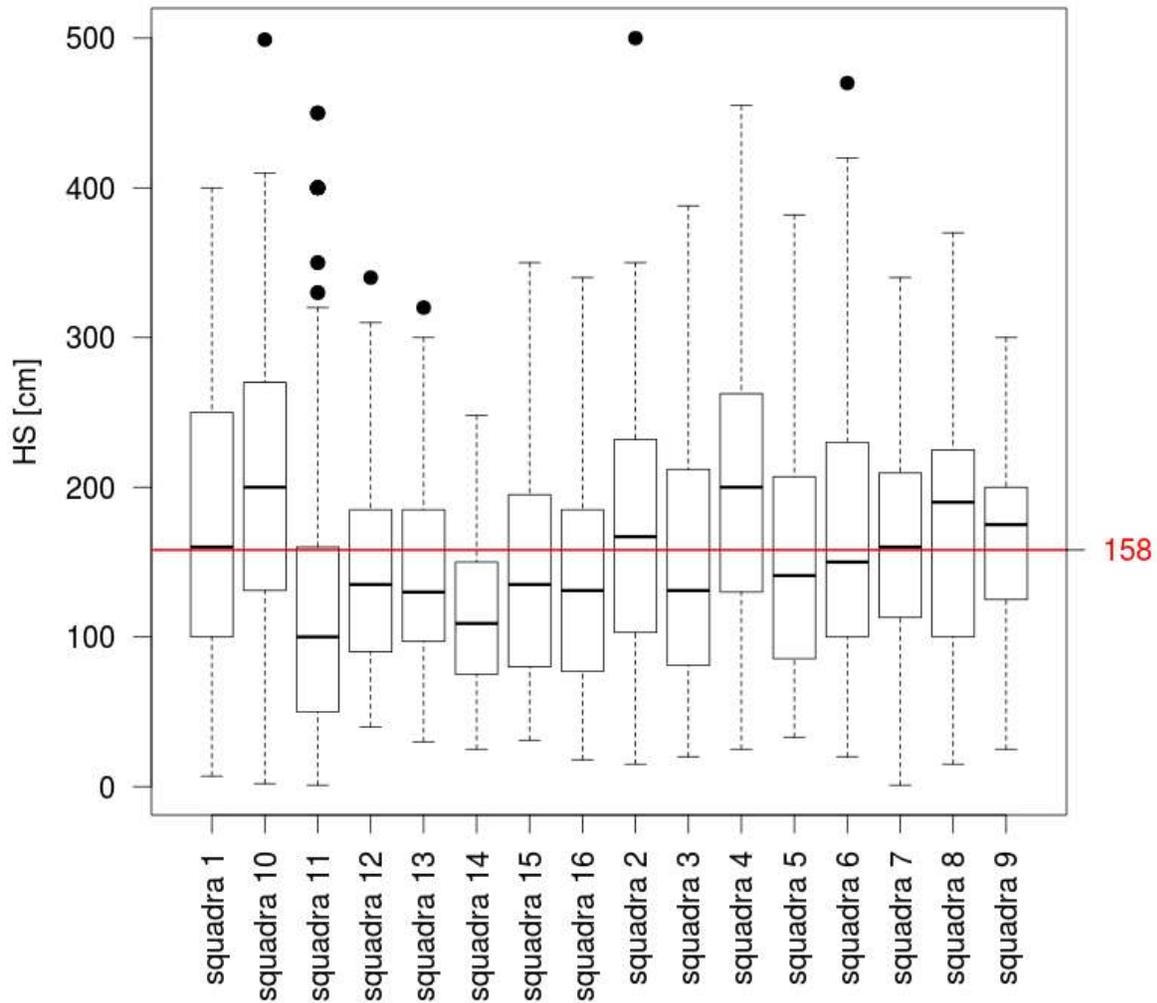


Figure 4: Boxplots of snow depth for each team. Red line is the grand median.



Spatial representativeness of sampling paths

One uncontrolled feature of the sampling scheme was the path chosen by each single team. By comparing morphological characteristics of the whole basin to those sampled during the campaign we can evaluate how representatively the morphology was sampled by the teams. This can have important consequences on the spatialisation in view of the fact that dem-derived fields are used to distribute SWE in space. Figure 5 shows the frequency distribution of some morphological variables (elevation, total radiation, concavity/convexity and slope) for the sampling points and for the whole basin. Radiation, slope, concavity/convexity on sampled points show a very similar distribution compared to the whole basin. For elevation, sampled points are distributed somewhat higher than the elevation range of the entire basin. Given that globally morphology representativeness is satisfactory, we can look at the same feature with a density distribution for each team (figs 6 and 7). For elevation, for example, it is clear that teams 9 and 14 preferentially sampled higher elevations, whereas other morphological variables were adequately sampled by all teams.

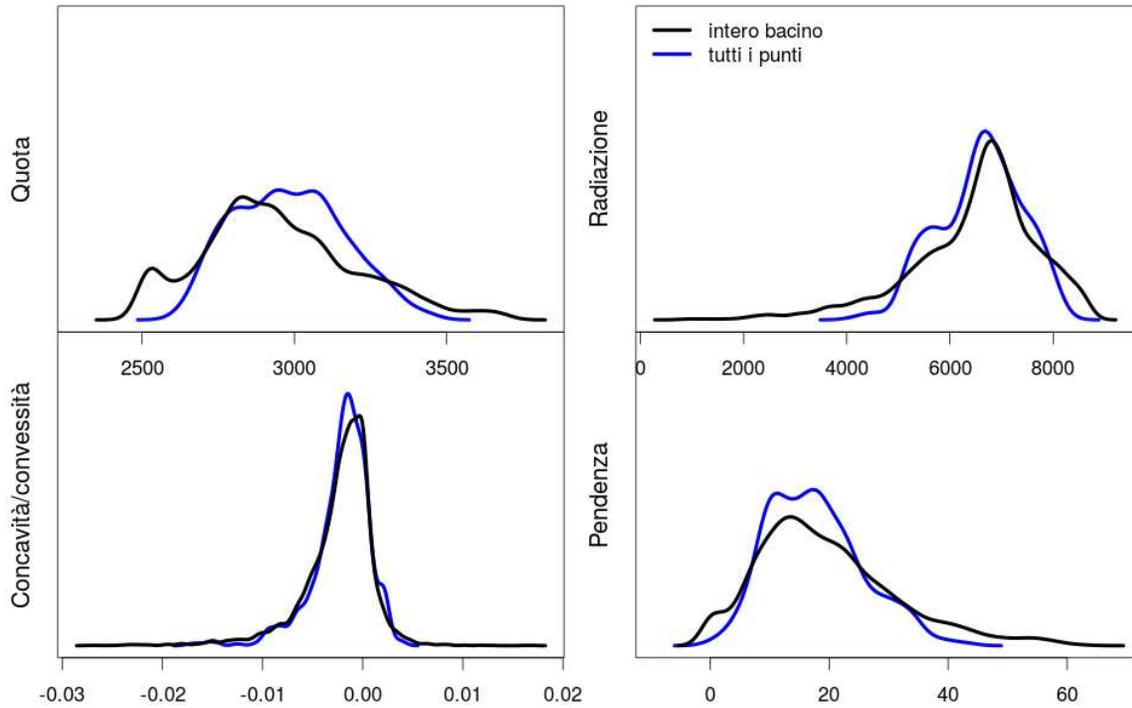


Figure 5: Frequency distribution of some morphological parameters for the sampled points compared to the whole basin (solid black line).

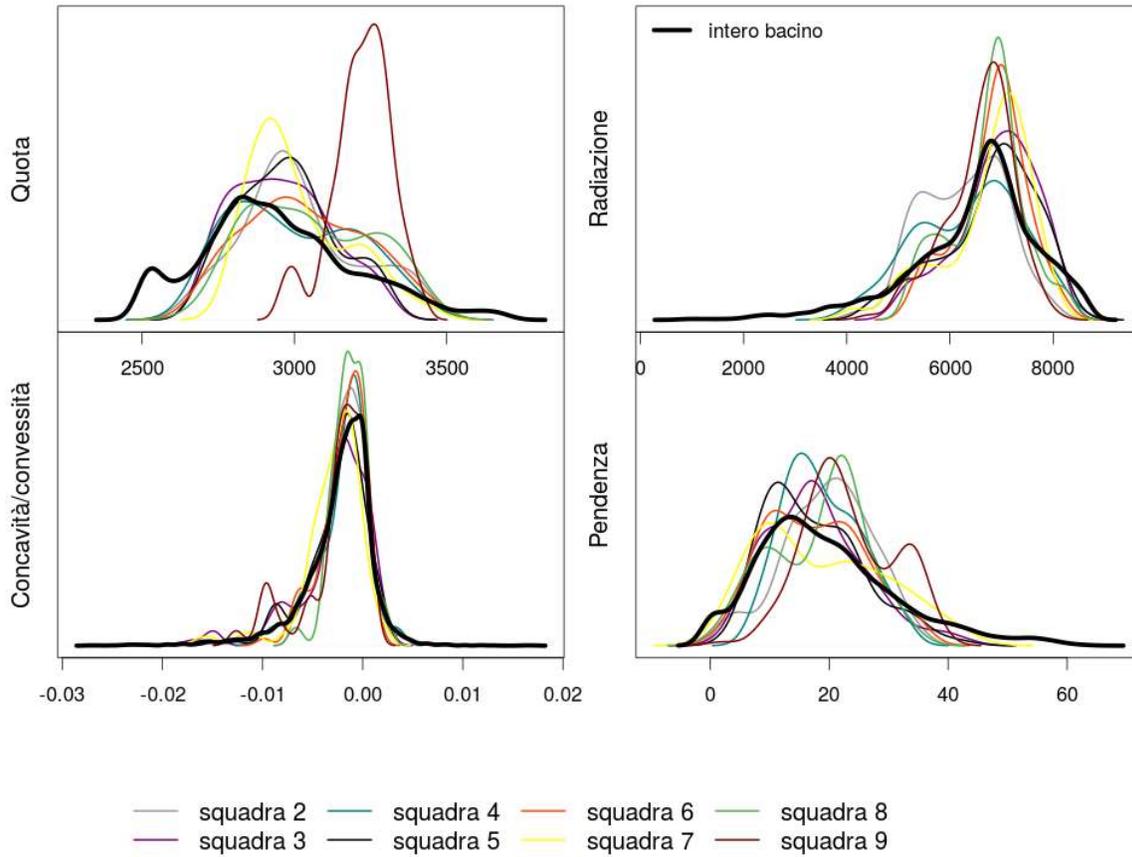


Figure 6: Frequency distribution of some morphological parameters for the sampled points compared to the whole basin (solid black line) for teams 2-9.

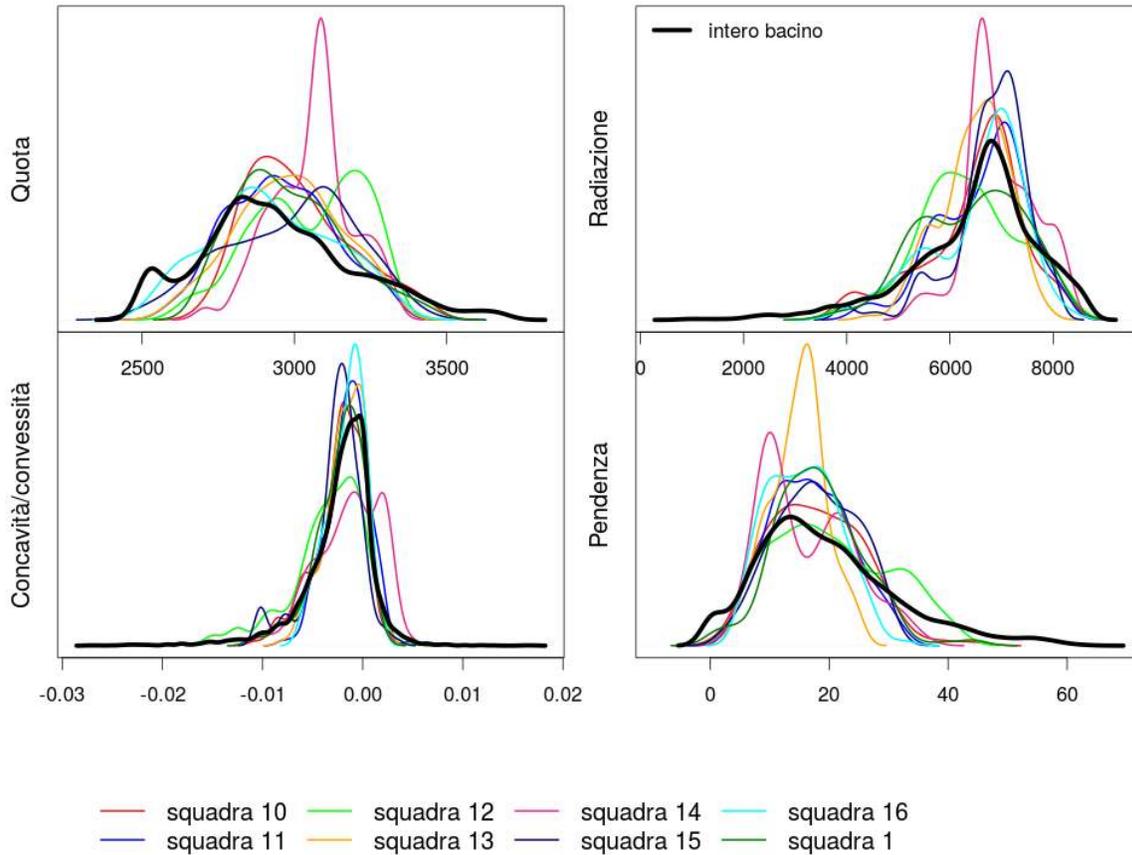


Figure 7: Frequency distribution of some morphological parameters for the sampled points compared to the whole basin (solid black line) for teams 10-16.

SWE spatialization

The set of measurements taken by each team is here considered as an independent campaign. Each of them was therefore used to produce a simulated SWE map. A global simulation was also run, by including all measurements simultaneously, and this last estimate is considered as the best approximation of the ground truth. Figure 8 shows total SWE calculated by summing the

SWE contributed by each pixel for each of the 16 simulations. We also show the average of the 16 simulations in blue and the global simulation in black. The variability among simulations is considerably high (up to 30% of the total, and on average 10%). By comparing simulations 2017 to those of the period 2012-2016, it appears that values in 2017 are considerably lower compared to the previous years' range. By looking at the totals in fig. 8 we can conclude that the average of the 16 simulations nicely matches the estimation obtained by the global model.

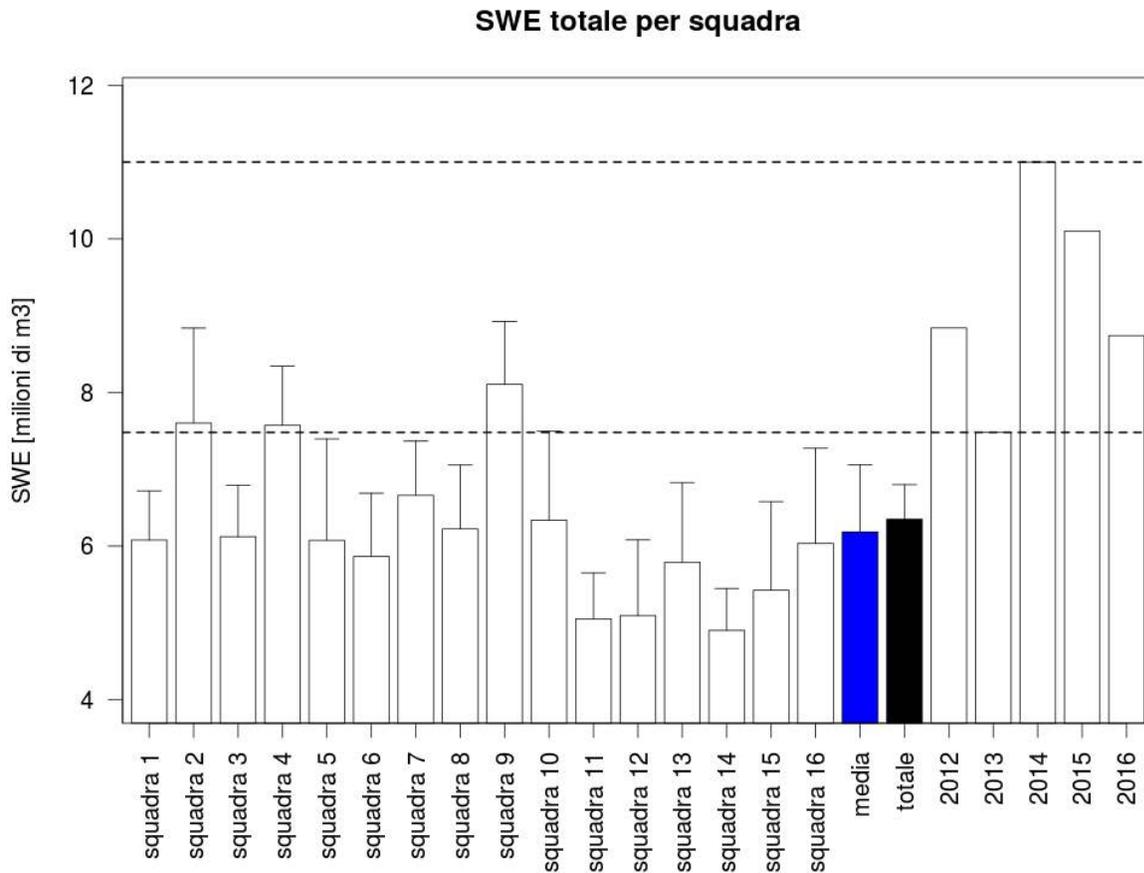


Figure 8: SWE total estimated from the 16 independent simulations, their average (in blue) and the global simulation (in black). For comparison, we show the simulation obtained in previous years.

By looking at the average SWE map (the average of 16 maps, one for each simulation) we can appreciate the distribution of SWE on the whole basin. Based on the standard deviation map, we can notice that the highest variability is found in areas with a limited number of observations, as expected. In contrast, the area between the two lakes, although with fewer observations, shows a low variability.

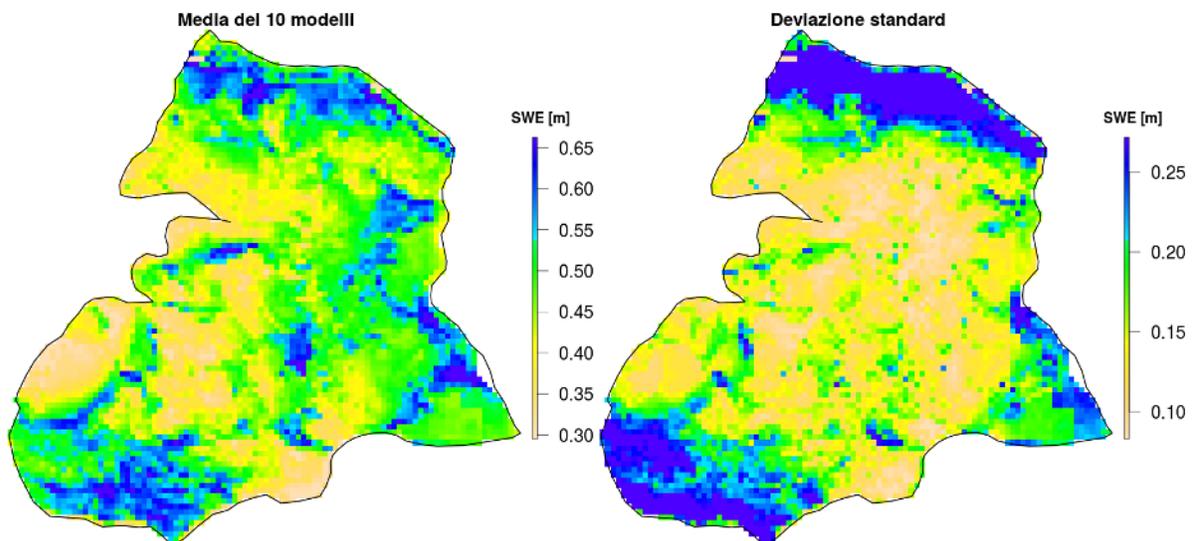


Figure 9: Map of the average and standard deviation of SWE obtained from the 16 simulations.

The map in figure 10 is obtained from the global simulation. SWE distribution shows different spatial patterns compared to the average of the 16 simulations shown in figure 9, even if it leads to a total SWE on the basin comparable (cfr fig. 8).

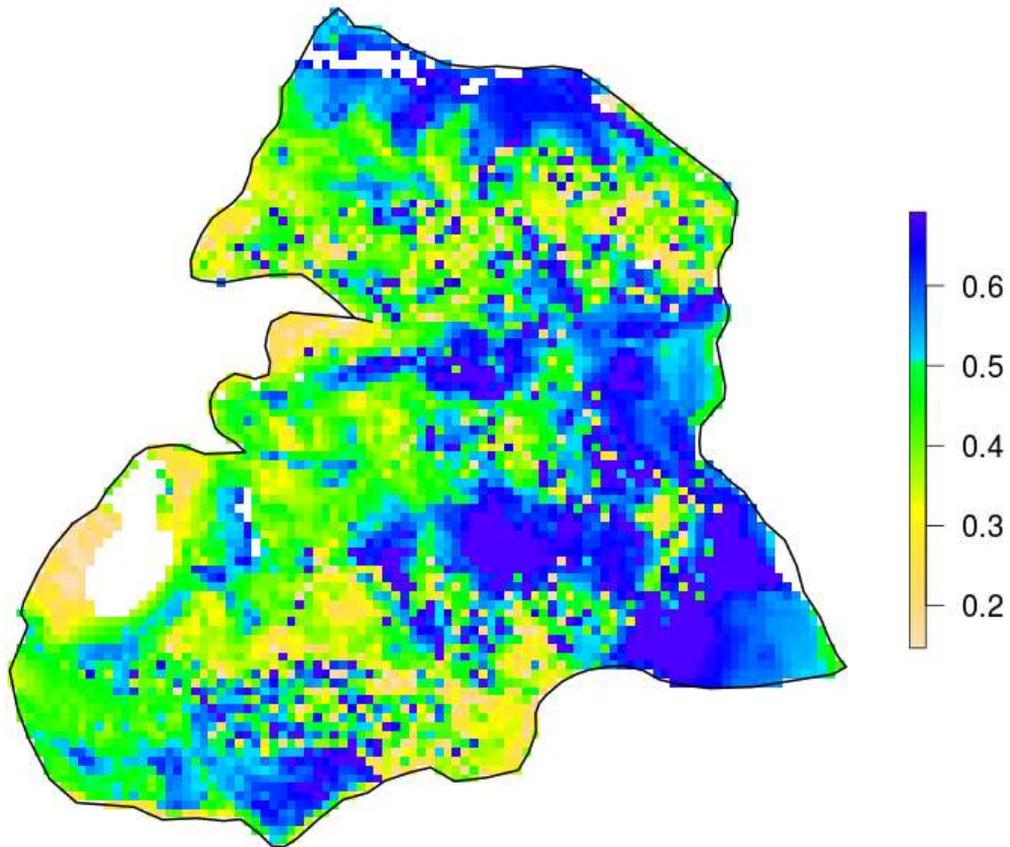


Figure 10: SWE map obtained by the global simulation (all observation feeding one single simulation).



SWE spatial distribution

To highlight the spatial differences between different simulations, we calculated the anomaly between each single observation and the global simulation map. Figure 11 shows that anomaly, with blue colors indicating SWE values higher than the global, and red colors indicating the opposite. The main differences are: 1) simulation 8 returns an area of NA values for a large area of the basin. This normally happens when the model produces negative SWE values. 2) No single teams except for simulation 14 display completely negative or positive anomalies. 3) A widespread condition occurs, where single simulations provide overestimation for certain subareas and underestimation for others.

By comparing the anomaly maps with the graph showing total basin SWE (fig. 8) we can observe that the simulations that lead to a total SWE sensibly lower than the global or the mean simulation, as for example for simulations 14, 15 and 16, these three show a negative anomaly in the south-east area of the basin. Conversely, total SWE higher than the average (as for simulations 2 and 9) are generally due to positive anomalies of the northern or south-western areas of the basin. In both cases of negative and positive anomalies, they are located in areas with lower observations thus indicating, as expected, that the lack of observations leads to higher uncertainty in SWE modeling.

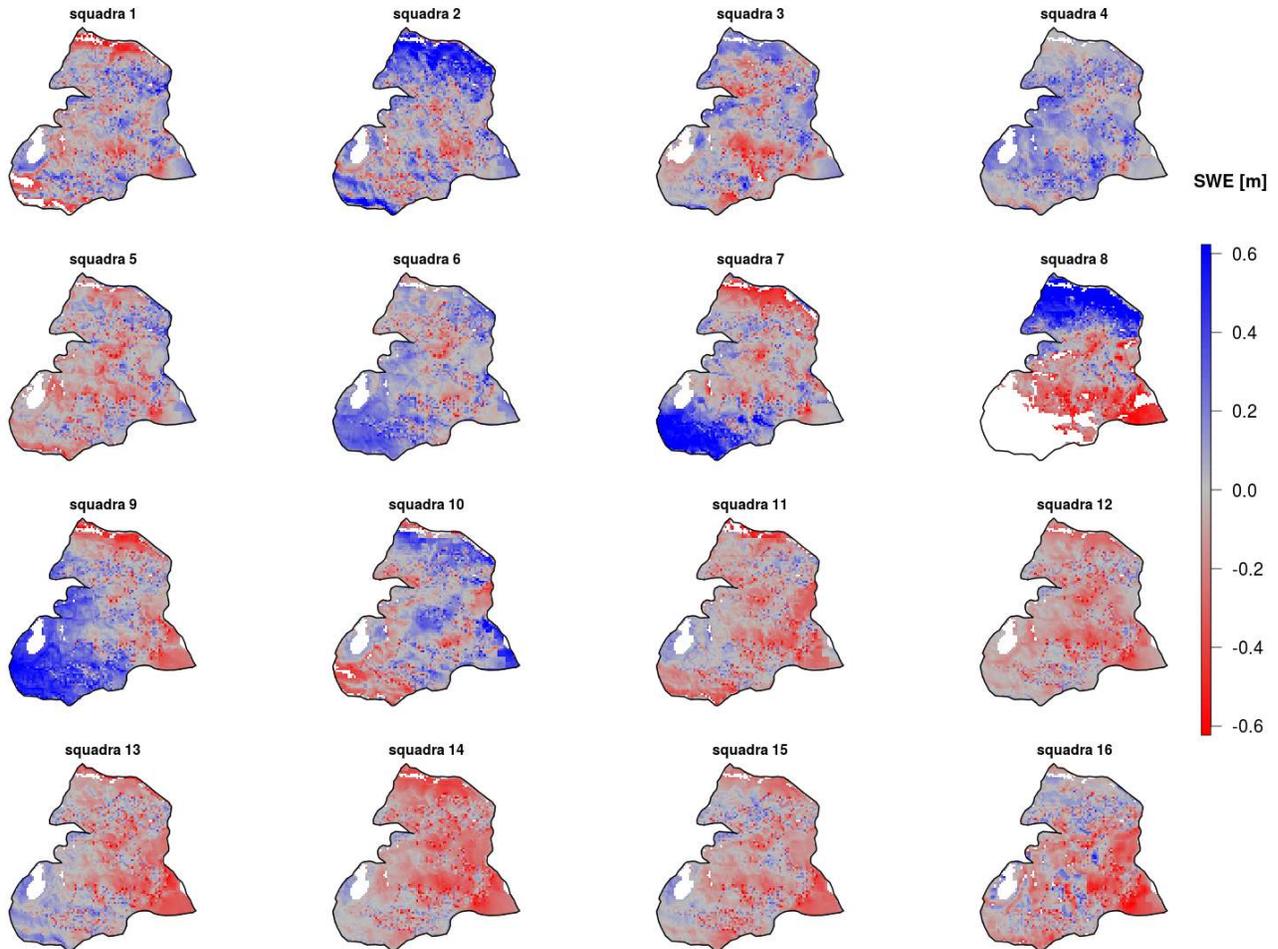


Figure 11: SWE anomaly for each simulation, obtained by subtracting to the actual simulation map the one obtained from the global model.

Factors controlling SWE distribution

The relative importance of dem-derived factors controlling SWE is illustrated in figure 12. This figure shows the variance explained by single predictors in the regression model used to spatialize SWE. In other words, the chart shows how important is the relationship between point SWE and morphological characteristics. The figure shows furthermore how often predictor category is chosen

in the model. Solar radiation is certainly the most important parameter. It appears in all simulations and explains between 20 and 60% of the variance. elevation is selected only few times (7 over 16 simulations), contrary to what we observe at larger scale, where there is a positive correlation between snow amount and elevation. In small basins such as the Goillet other morphological properties associated to meso- and micro-topography overcome elevation in determining SWE distribution. For example, the predictor category called protection, which includes the topographic attributes that make a cell more or less prone to snow erosion or accumulation, plays an important role in Goillet simulations, being chosen in all but one models and explaining 10-60% of the variance.

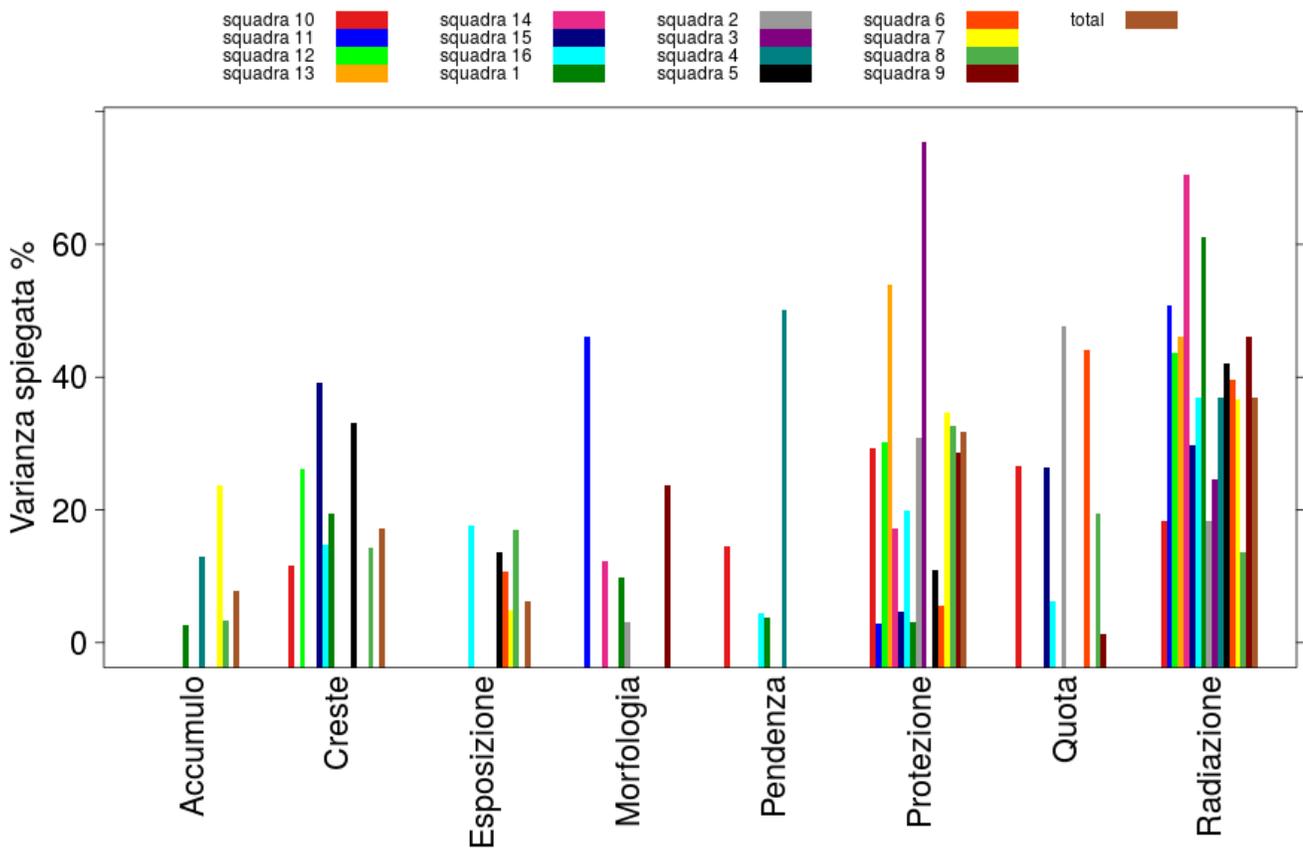


Figure 12: variance explained (%) by different categories of predictors in the multiple regression model.



4 Conclusions

We found a significant variability among simulations based on sampling strategy of the different teams. By analyzing the various paths chosen, we highlight the difficulty to sample the sectors of the basin in a proper and homogeneous manner. Differences in snow depth measurements are further amplified by the spatialization process. We stress the importance of a sound planning of the field survey.

Moreover, some areas had limited access due to i) security reasons (highest part of the slopes beneath Fürgen peak, northern slopes close to Gran Sommetta, glacial areas) or ii) logistic reasons (the flat area between the two lakes). In these areas, fewer measurements lead to larger variability between simulations: this is particularly true at the edge of the simulations (areas close to N and S mountainsides).



IV interconfronto SWE

Bardonecchia (Val di Susa, TO)
19-20 march 2018

Technical report prepared by ARPA Valle d'Aosta



In a nutshell

- 93 SWE measurements conducted in the Bardonecchia Basin by profile sampling (horizontal core) or vertical core sampling
- 65 snow observers involved, divided in 21 measurement teams
- Average basin SWE was quantified at 550 mm w. eq.
- Elevation and wind erosion explain 50% of the spatial variability in snow depth and SWE
- Solar radiation and aspect partially explain the spatial variability of snow density
- To maintain the average error in SWE estimation smaller than 20% no less than 71 and 23 measurements must be collected for snow depth and density, respectively. If our target is 10%, required measurements increase to 85 and 53.



1 Objectives 2018

Previous editions of SWE activities lead to the intercomparison between methods in homogeneous (2015 and 2016) and heterogeneous conditions (2016), the assessment of the impact of snow depth variability on the final estimation of SWE at the basin scale (2017). The current edition was organized again with the general objective to **quantify the snow water equivalent at the basin scale** and specifically:

- 1) evaluate SWE spatial variability in a specific alpine basin and at a point in time and its relationship with morphological attributes (elevation, aspect, slope)
- 2) assess the relative importance of snow depth (HS) and density (RHO) in the determination of basin-scale SWE
- 3) explore morphological attributes (elevation, aspect, slope, etc...) controlling variability in HS, RHO and SWE
- 4) estimate the minimum number of RHO and HS samples required to obtain a sufficiently accurate estimate of SWE (with an error lower than 20%).

2 The field work

The investigated area is the Bardonecchia Ski Area, alta Val di Susa (TO) (fig. 1). Sampling points were chosen to be representative of topographic conditions of the whole basin with several restrictions; access to the field was constrained by the skilifts (Colomion Spa); movements across long distance had to be avoided; the security of all involved personnel had to be guaranteed. The following classes of slope, aspect and elevation were considered (table 1).

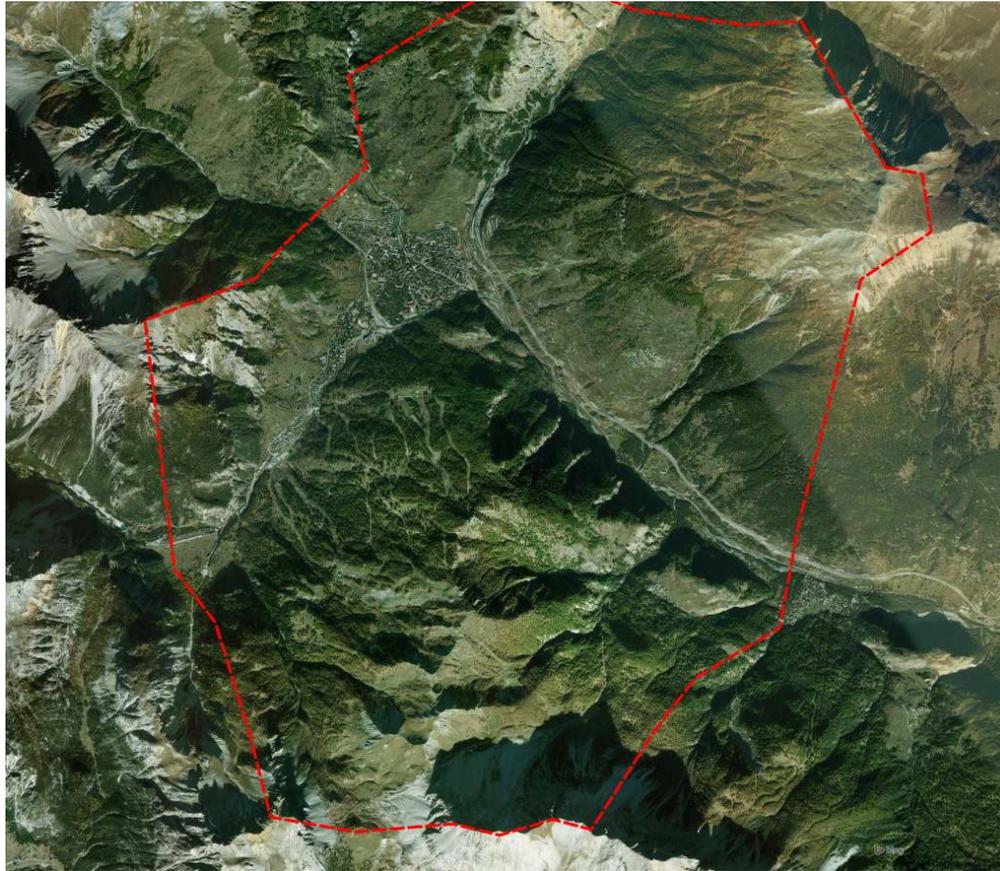


Figura 1: The study area

Tabella 1: Classes of elevation, aspect and slope considered to distribute the sampling points

Elevation	Aspect	Slope
1300-1800 m	North ($>315^\circ$, $<45^\circ$)	$<15^\circ$
1800-2300 m	East (45° - 135°)	15° - 25°
2300-2800 m	South (135° - 225°)	$>25^\circ$
	West (225° - 315°)	

By combining the various classes we obtained the distribution of topographical units of the whole area, and their respective abundances, so that the most abundant classes could be adequately repre-

sented by the sampling scheme. The final distribution of sampling points is shown in figure 2. Each of the 21 teams was assigned between 3 and 6 points (depending on their distance/accessibility). If for any reason the team could not reach a given point, we asked them to sample an extra point with the same topographic characteristics as the one assigned. Data transmission was conducted with smartphone (a picture of the filled sampling form).

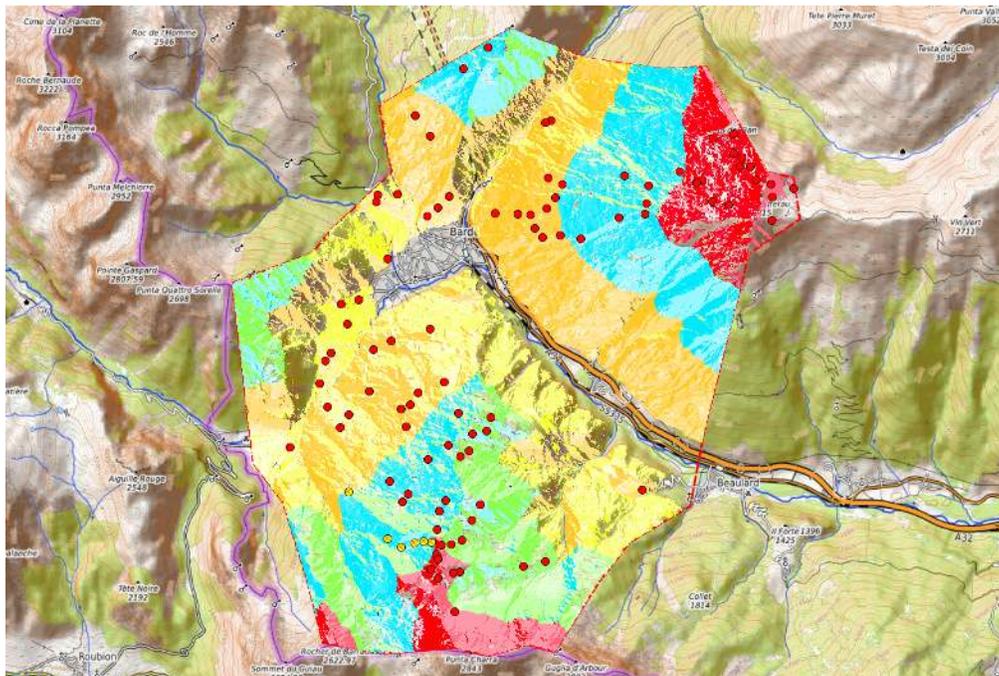


Figura 2: Distribution of the sampling points

3 Base statistics of the parameters

Ninety-three points were sampled in the sub-areas of Jafferau, Colomion, Melezet, Col des Accles, Bardonecchia, Melmise, Puy Beaulard. Points that were actually sampled roughly correspond to the ones we assigned (fig. 3).

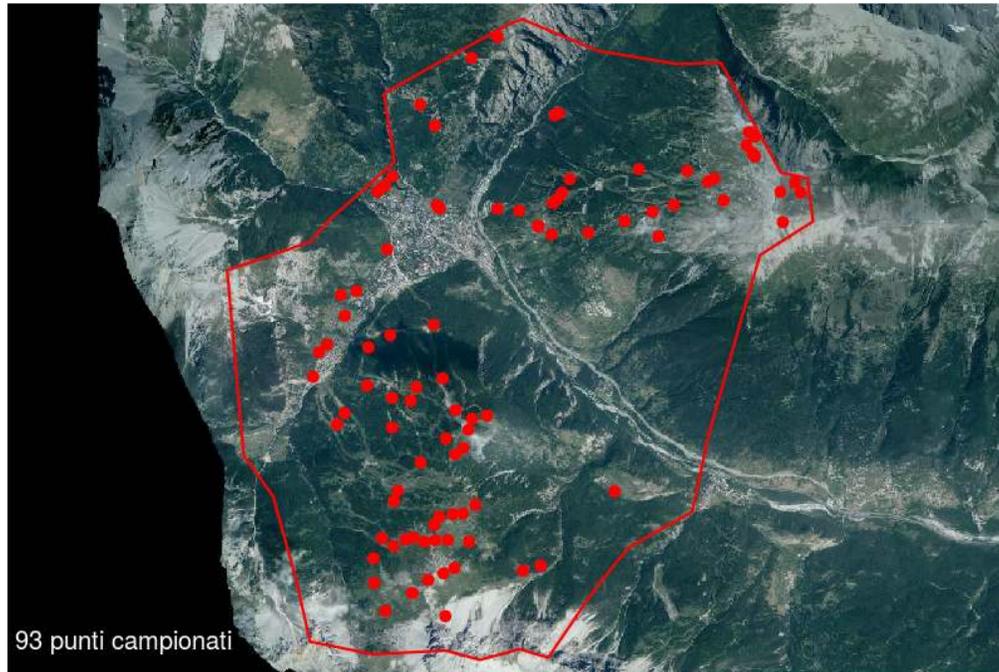


Figura 3: Distribution of the points that were actually sampled.

Average values of snow depth (HS), snow density (RHO) and SWE are reported in table 4 and fig. 5. Mean HS was 165 cm, mean RHO was 340 kg m^{-3} and mean SWE was 550 mm w. eq.. 2 points show very large values of snow depth and SWE, at 400 cm e 1500 mm, respectively. These points will be discussed further. The elevation range sampled was between 1300 m and 2770 m. Slopes varied between 0 and 43° . Prevailing facing of the points is west, which is the most frequent facing of the entire basin, with the other 3 facings with similar frequencies.



	hs	rho	swe	dem	slope	expo
Min.	:0.280	Min. :214.4	Min. :0.0980	Min. :1326	Min. :2.149	17
Median	:1.685	Median :339.3	Median :0.5502	Median :1979	Median :19.633	17
Mean	:1.653	Mean :339.5	Mean :0.5511	Mean :1942	Mean :19.539	17
Max.	:5.000	Max. :470.0	Max. :1.9840	Max. :2770	Max. :42.955	39

Figura 4: Base statistics for HS, RHO, SWE and topography. Column named expo shows the number of points sampled for each facing class, i.e. east, north, south and west, from the top.

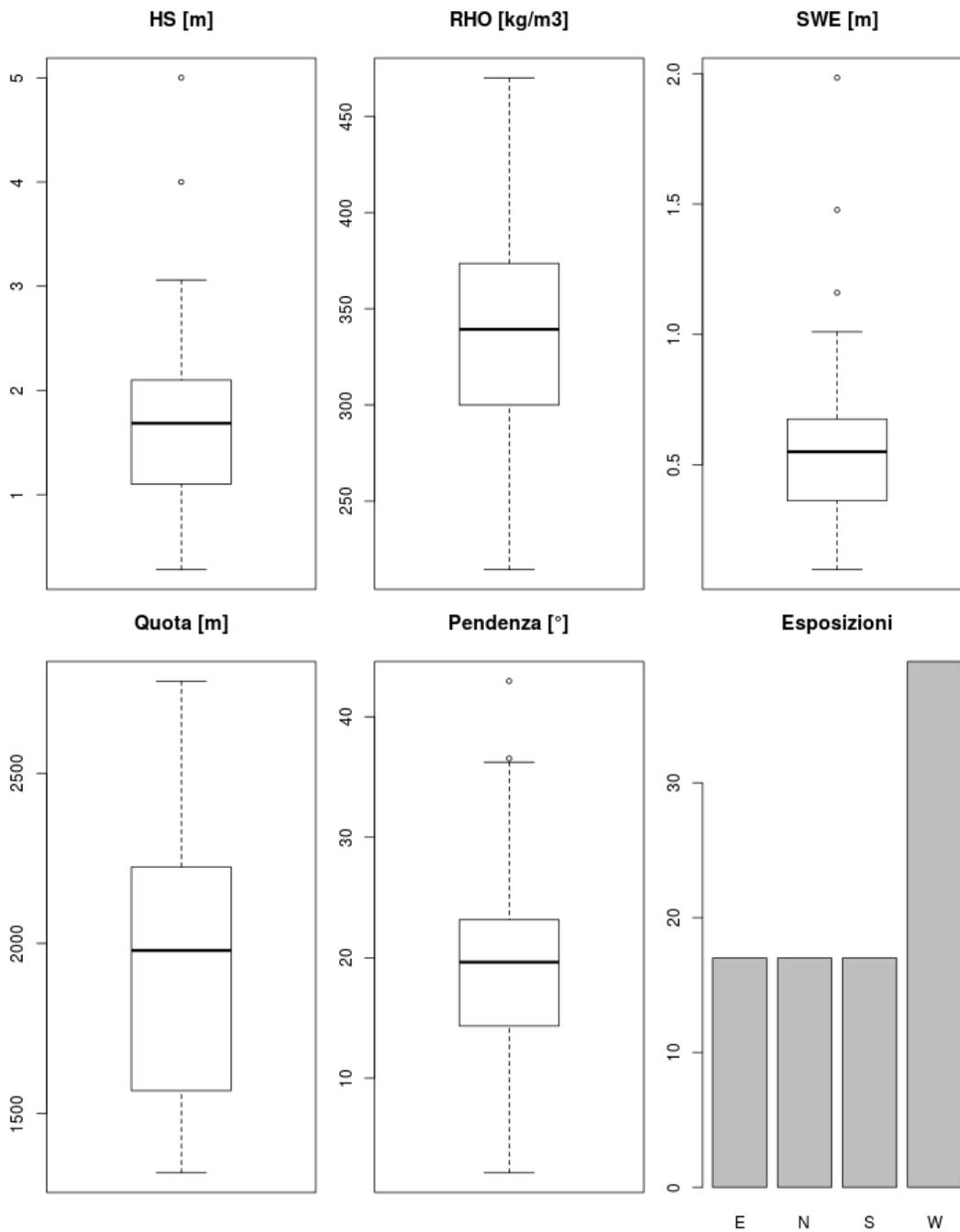


Figura 5: Boxplots for HS, RHO, SWE and topography. Black line is the median, boxes represent 25th and 75th percentiles, horizontal whiskers represent minima and maxima in the absence of outliers, represented by points



4 The relationship between topography and measured parameters.

The graph in Fig. 6 shows the relationship between measured parameters (HS, RHO and SWE) and some topographic variables. Snow depth is mainly related to elevation. Points are furthermore colored according to an erosion flag (blue = snow accumulation areas, red = snow erosion areas, black = neutral to erosion). By including erosion in the linear model the variance explained increases from 0.24 (with elevation only) to 0.5. Erosion and elevation together explain therefore 50% of the variance of HS. Snow depth is weakly related to other topographic attributes. Snow density is uncorrelated to slope and elevation but shows a correlation with facing. Warm slopes (facing south and east), have higher snow density compared to north and west. This pattern is also confirmed by the positive relationship between snow density and solar radiation. The relationship between SWE and topography reflects the one found for HS, with usually lower R^2 .

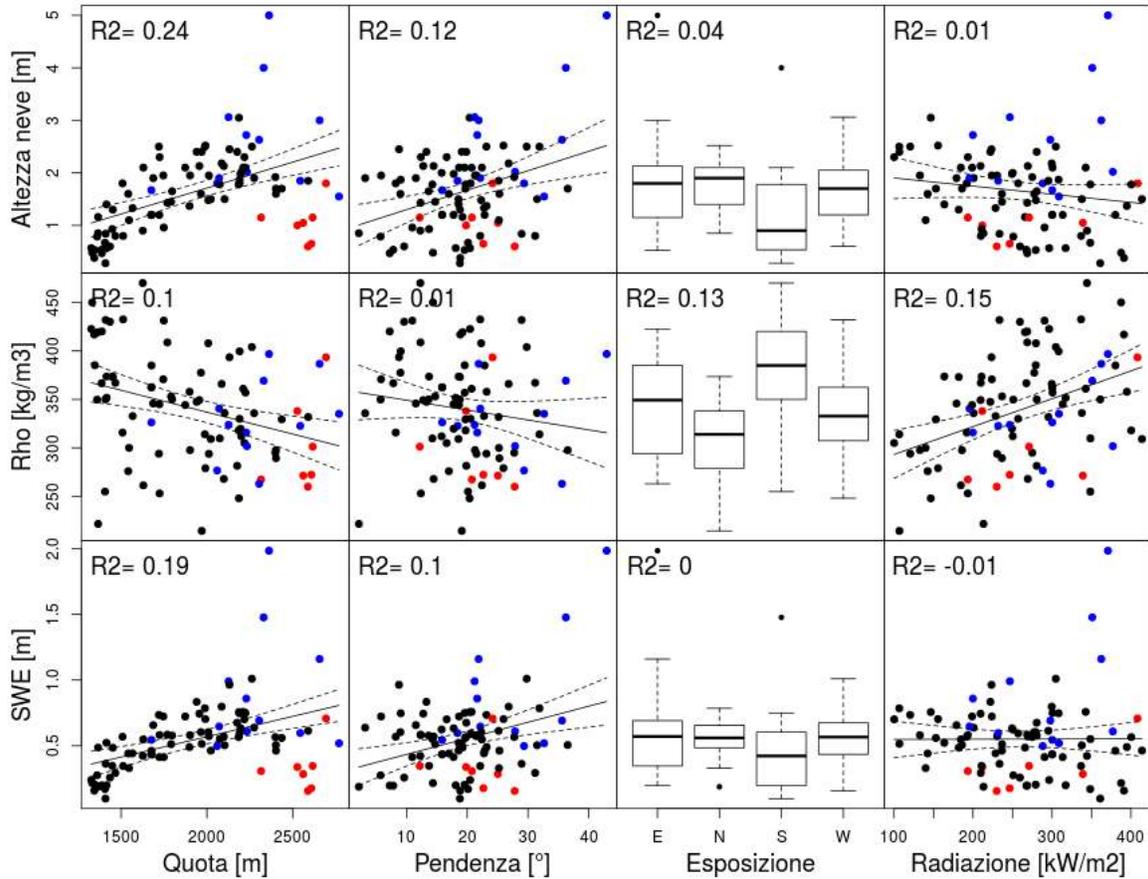


Figura 6: Relationship between measured parameters (HS, RHO, SWE) and selected topographic variables. Each panel reports the r-squared of a linear model. Blue points represent snow accumulation areas, whereas red points indicate snow erosion areas. Black points are neutral with respect to erosion. Radiation values are calculated as accumulated potential solar radiation since Jan, 1st to March, 19th, the sampling day).

In addition to the qualitative assessment of measured parameters and topography we seek a quantitative approach in order to better understand which variables are more important in the spatial distribution of SWE, RHO and HS. To this end we used the *random forest* technique, that allows to rank variable importance. This technique is robust against outliers and autocorrelated variables. Results are shown in fig. 7. Elevation explain most of the variance for HS, followed by erosion. Radiation, slope and facing are of minor importance into the model. For snow density,

the model performs worse than for HS. The most important variable is solar radiation, followed by elevation, slope, facing and erosion. Variable ranking for SWE is the same as the one described for HS.

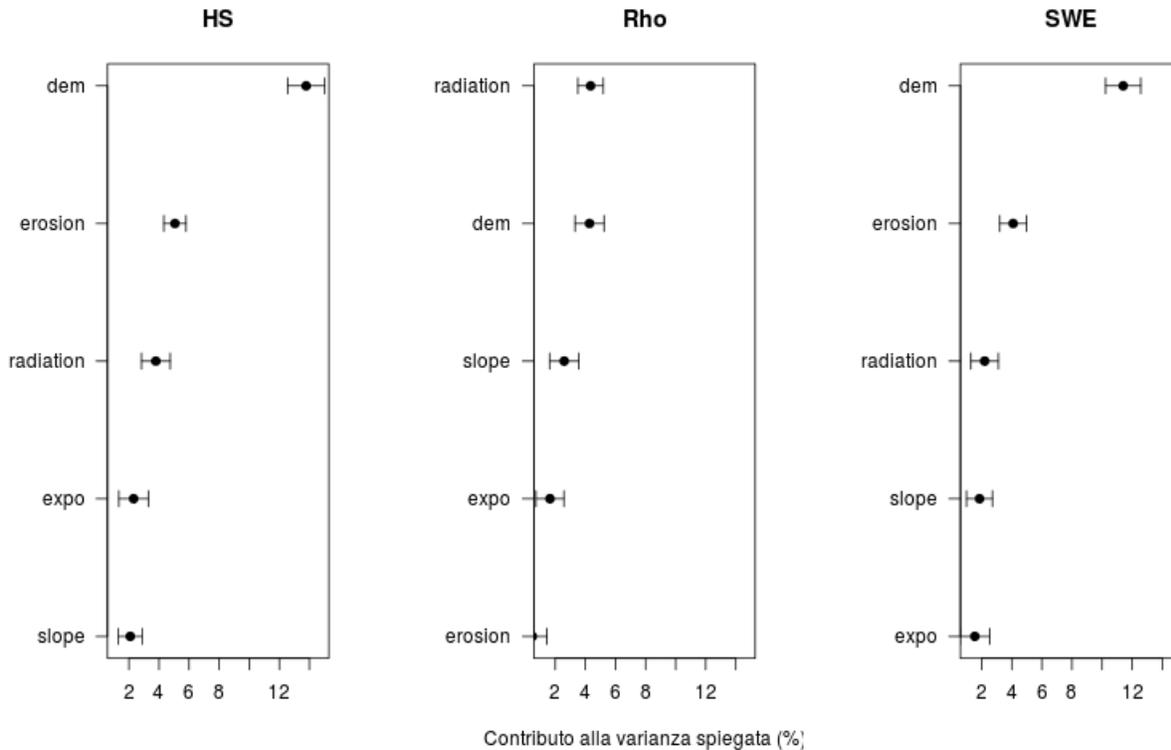


Figura 7: Random forest: importance of the different variables. X values are proportional to the explained variance by the model. The higher the contribution of a variable, the higher is its importance in the model.

Random forest analysis was conducted also on the dataset split in two elevation classes, above and below 2000 m to examine whether different drivers regulate snow distribution at different elevation belts. Results are shown in fig. 8. Variables controlling HS below 2000 m have the same ranking as for the complete data set. At high elevation, erosion becomes the most important factor in HS distribution. This is explained by the fact that erosion exerts its influence preferably at higher elevations, close to the ridges where snow redistribution by winds is an effective process. For RHO, radiation and facing are ranked at highest importance in the model, corroborating the hypotheses

from the qualitative analysis (5).

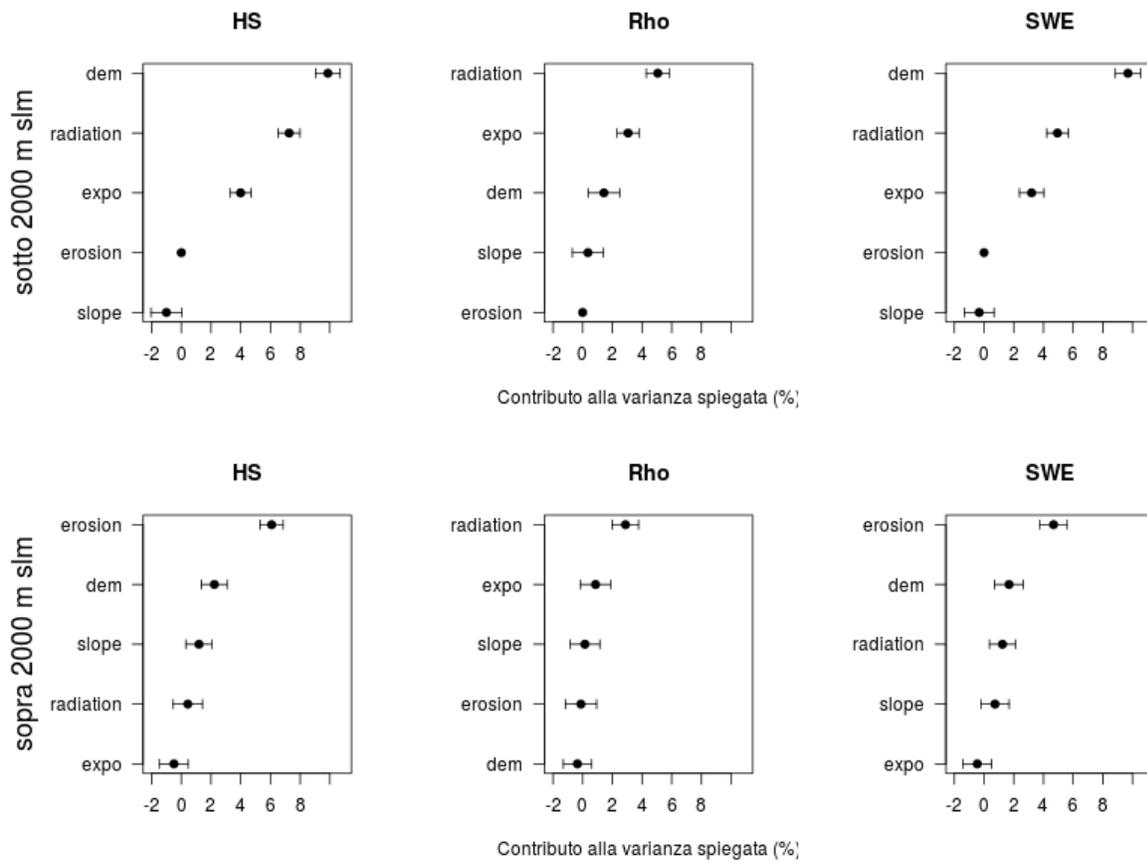


Figura 8: Same as in fig. 7 split by elevation.



5 Operational objective: the minimum number of required samples

The high number of measurements conducted in this comparison offers the possibility to run statistical analyses otherwise not possible. In this case, with 93 points we can provide a robust estimate of the mean SWE at the basin scale and consequently we can evaluate a minimum number of measurements required to achieve a given target in accuracy. This analysis aims at minimizing labor-intensive and potentially dangerous field activities.

Because the most time-consuming measurement is snow density, especially with the horizontal sampling technique, we would be particularly interested in reducing the number of RHO sampling. The following analysis simulates a reduced sampling, where actual measurements are progressively substituted by the average of the remaining samples. This exercise is run by either reducing HS and RHO points, even though the first simulation is unlikely to occur in the reality (i.e. conducting more RHO than HS measurements).

Figure 9 shows the results of the above mentioned analysis. Along the x axis from right to the left, the number of samples is progressively reduced for RHO (red line) and HS (black line). Variability (and error) increases with decreasing sampling number, but at markedly different rates for HS and RHO, due to the much higher variability of HS compared to RHO.

In this case and at that moment of the season we estimate that with 71 and 23 measurements for HS and RHO, respectively, we can estimate SWE with an average error of 20%; if we want to reduce this error below 10% we should put a larger effort and increase samples number to 85 and 53 measurements for HS and RHO, respectively.

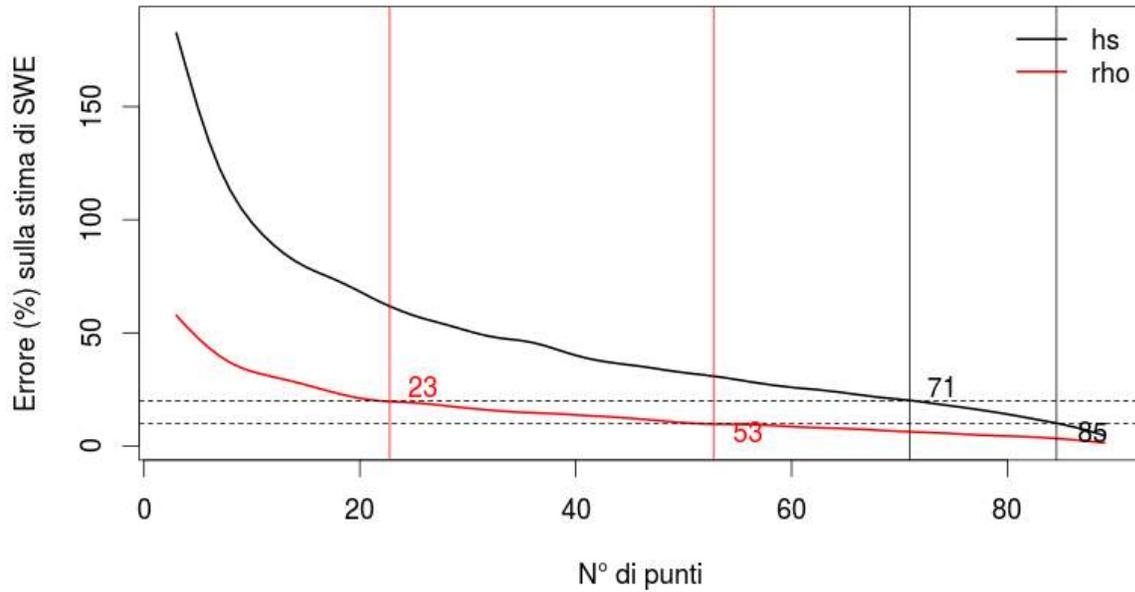


Figura 9: The effect of sample removal on the error in the estimation of SWE at the Basin scale. The red line represents the reduction of RHO measurements, keeping the complete number of HS measurements (the black line, the other way around). Dashed horizontal lines represent two uncertainty (error) targets, at 20 and 10%. The correspondent vertical lines identify the number of samples required to achieve those targets.



6 Conclusions

- The conspicuous number of measurements (93) allowed to estimate an average basin SWE on March 20th 2018 in Bardonecchia at 550 mm. This figure parallels the estimates across the Alps in this very snowy year.
- Elevation and wind redistribution contribute substantially to SWE and HS spatial distribution, whereas solar radiation (and facing) partly control the distribution of snow density.
- We estimated that to obtain an error in SWE estimation at the basin scale below an arbitrary 20% target, we need to collect a minimum of 71 and 23 samples for HS and RHO, respectively.