Answers to A. Fischer

We would like to thank Andrea Fischer for her remarks that helped to clarify and improve the manuscript. In the following you find the responses to the reviewer's questions one by one:

> **Bold: comment of the reviewer**
> Red: Answer of the authors
> *Italic: Changes to the initial manuscript*

1) The reading flux would benefit from fewer abbreviations, for example by writing CP and VP in full.

We removed many abbreviations and hope we could improve the reading flux.

2) The article would benefit from restructuring some parts. The description of the validation method in the validation section is clearly too short, and the text which would fit in this section is split across several paragraphs of the result section. This brings some confusion to the samples used for validation: in some paragraphs, d203 is used for the validation sample, in other sections d133 appears in addition to d110 and d203. In other paragraphs, 23 selected stations are mentioned, or validation results are presented which reference to 4 stations. This would be easier to follow if the presented method and samples were not spread across the results, but brought together in the sections on validation methods.

Section 3.3 includes now a summary of the validation methods. In addition Table 2 (section 2.2) gives an overview of the used grids and datasets.

3.3 Validation methods

In this study, map203 is assumed to contain more accurate gridded snow information because it is based on more snow measurements that also cover higher elevations. Therefore map203 is considered as the reference data set.

In a first step SWE estimates are compared to measured snow depth that is converted into SWE over the period 2001-2009. At each measurement station, SWE was estimated with the associated station (SWE_{orig}) and without ($SWE_{loo\text{-}cv}$). The altitude was adjusted to the station altitude by subtracting the day-specific SWE gradient from the modelled SWE values. "Orig" and "loo-cv" were compared to identify the uncertainty of the model and the impact of individual stations on the mapping results. Estimated SWE is validated taking into consideration the mean error (ME), the mean absolute error (MAE), the squared correlation coefficient (R^2) and the root mean squared error (RMSE). In order to ensure that small fluctuations in SWE are preserved, R^2 is calculated with the seasonal trend removed. Therefore the first-order differences with a lag of one year are used (Wilks, 2006; Foppa et al., 2007; Saloranta, 2012).

After implementing the quantile mapping (section 4.3), which is performed according to the results from comparing map203 with map110 (section 4.2), map.cal was validated. First the calibration was tested for the nine overlapping years with a cross-calibration approach. By dividing the data set into a training period (eight years) and a calibration period (one independent year), each year can be calibrated independently with the remaining eight years. The relation obtained between the quantiles of the two data sets was then applied to the 39 available years of map110.

For the validation of the spatial and temporal consistence "independent" stations are used (Table 2).The temporal validation outside the calibration period is done by comparing the calibrated maps with "independent" measurements from d23 (Fig. 1) during the test period (1989-2000) and the calibration period (2001-2009). Most of these 23 stations are located between 1500 and 2000 m a.s.l. and three stations are located around 2200 m a.s.l. The spatial consistence is tested with a grid that removes four stations (Stations with the black outline in Fig. 1) before estimating SWE ($map203^{-}$). These four stations are then used to validate $map203^{-}$ and the therewith calibrated grid $map.cal^{-}$.

Table 2. Grid, number of stations used to produce grid, available data period and role in the study (c: used for calibration, v: used for validation.

Grid	*Stations*	*Period*	*Role in the study*
map110	*110 (d110)*	*1971-2009*	*c*
map203	*203 (d203)*	*2001-2009*	*c/v*
map.cal	*-*	*2001-2009*	*-*
	23 (d23)	*1989-2009*	*v*
map203^{-}	*203-4*	*2001-2009*	*c/v*
map.cal^{-}		*2001-2009*	*v*

3) For potential users of the algorithms, a hint on the number and length of records needed for the calibration sample would be helpful. For example, would it make sense to calculate SWE climatologies from three 20 year long records and 20 one year records?

We are not sure if we understand the question correctly. If the question is how much data is needed to produce 'good' CDFs, we suggest a minimum of 100 data points. This can be achieved by a sufficient number of years or by expanding the time window around each day. However a minimum of five years should be available to represent of the interannual variability.

4) It would be interesting to see the altitudinal distribution of the stations, at least for the three test regions.

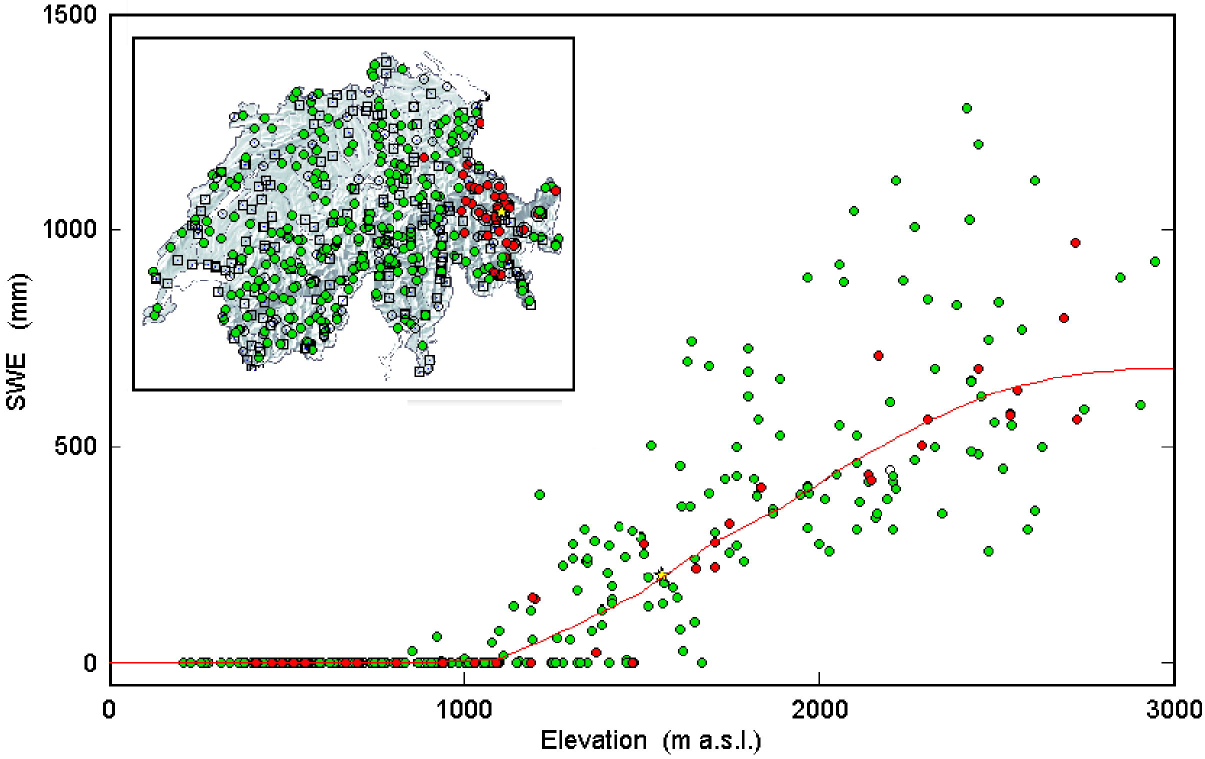
We included the distribution of the stations for the three test regions in Table 1.

Table 1. Characteristics of the three regions Alpine Rhine (AR), Valais (VS) and Thur/Töss/Glatt (TTG) and the distribution of SWE stations within these catchments.

	AR	TTG	VS
Size [km2]	6342	2586	5382
Mean elevation [m a.s.l]	1742	696	2078
Min elevation [m a.s.l]	409	345	372
Max elevation [m a.s.l]	3361	2324	4403
% above 2000	39.25	0.04	56.02
Number of SWE stations	*43*	*8*	*43*
Mean elevation of SWE stations	*1706*	*925*	*1996*
Max elevation of SWE stations	*2725*	*1610*	*2950*

5)What did the SWE/altitude curve look like, indicating also the elevation of the stations used? How did you calculate the SWE for not measured areas (elevations above 2100/2700 m)?

The Figure below is an example of the non-linear trend with elevation for an arbitrary day with the available snow depth data at the green and red points. For the 'regional' non-linear trend only the stations coloured in red are used. We hope the additional information in the text also clarifies this point. All points (red and green) would be used to determine the global (Swiss wide) trend of SWE with elevation



We make no statements for grid cells higher than 2700 m a.s.l., because we have no data for this altitude. For your information we prepared a figure for the demonstration of the regional SWE trend with elevation. The red points are the stations that are used to calculate SWE for the position of the star. The green points are all the available stations for this example. The trend continues horizontally above 2100/2700 m a.s.l. by considering the closest stations for estimating the offset from the trend.

The information of the calculation of the SWE trend with elevation is now included in section 3.1.

The longest climatology based on 110 stations only allowed for a detrending up to 2100 m a.s.l. The Thur/Töss/Glatt region is not affected, because only 2 % of the region is above 2100 m a.s.l. However a large part is above this level in the Alpine Rhine (33 %) and the Valais (52 %). Above this elevation, the SWE maps will probably underestimate the true SWE, as beyond that, the trend continues horizontally by considering the closest stations for estimating the offset from the trend. In contrast, d203 allowed for a detrending up to 2700 m a.s.l. This additional information for higher elevations reduces the fraction that is considered from the 'horizontal' trend in the Alpine Rhine (4 %) and the Valais (24 %). Consider that grid cells higher than 2700 m a.s.l. are not used for the validation.

6) How long should the calibration period be? What is the estimated accuracy of the measurements used for calculating the grids, and how accurate can the resulting grid then considered to be?

We suggest to use a calibration period of about 10 years (minimum 5 years). This guarantees to represent the inter-annual variability of the snow cover. The accuracy of the snow depth measurement is about ± 2 cm (section 2.2). However SWE measurements are assumed to have a larger uncertainty and a negative bias. Based on the bottom row in Figure 8 we assume a maximum difference of 50 mm on a pixel by pixel basis.

7) Some of the results are presented without error bars, assuming a higher accuracy than the input data. Is this plausible, and why? Or can an error bar be added to some of the presented numbers, to reflect the assumed accuracy of the results?

It is difficult to present uncertainties of spatial results. We will however try to consider in future work the uncertainties of the different components: measurement uncertainty, uncertainty of the conversion from HS to SWE, uncertainty of the spatial interpolation, natural spatial variability and uncertainty of the calibration procedure.

Detailed comments

8) Page 4246 line 10: For which period are these values calculated? Is this the mean of grid cells, or values for a mean elevation?

The values are calculated for the period 1980-2009. The indicated temperature is the mean temperature of the region averaged over all grid cells.

This is now specified in section 2.1:

Catchment average yearly precipitation sum and mean temperature are calculated from the meteorological forcing used by Zappa et al. (2012)for a hydrological simulation for all of Switzerland during the period 1980-2009.

9) Page 4247 line 10: It would be helpful to read the number of stations used in Table 1, together with the mean and maximum elevation of the measurements.

In Table 1 (see Table at Point 4) characteristics about the distribution of the stations are added and stated in the section 2.2.

The stations available in d203 in the Alpine Rhine (43 stations) and the Valais (43 stations) are well distributed. In the region of the rivers Thur, Töss and Glatt the 8 available stations (d203) are located close to its border. The mean elevation of the stations-sets equals to the mean elevation of the respective regions (Table 1). However the stations are not well representative for the highest part of the target areas.

10) Page 4249: How much of the area of the test regions is located above 2100 and above 2700 m? How did you calculate the SWE above the highest stations?

The fractions of the regions above 2100 m a.s.l. and 2700 m a.s.l. are summarised in the following table.

	Alpine Rhine	**Thur, Töss and Glatt**	**Valais**
Fraction > 2100 m a.s.l. [%]	33	2	52
Fraction > 2700 m a.s.l. [%]	4	0	24

See point 5) where this question is answered.

11) Page 4250 Line 17: Add a short reason why map 203 is assumed to be more accurate.

Map203 is assumed to be more accurate because it is based on a higher amount of stations that cover also higher elevations. This allows the snow water equivalent mapping model to use a SWE-elevation trend that is higher than for *map110.* This is now included in section 3.3 (see answer to point 2).

12) Page 4251 line 6: Wouldn't the description of the validation method fit much better in the section 3.3., Validation methods?

OK, moved to section 3.3.

13) Line 11: is this a 1 km by 1 km grid, or what does 1x1 grid mean?

Changed to: "1 km by 1 km".

14) Line 14: Blöschl, not Bloschl

Changed to "Blöschl"

15) Figure 4: It would interesting to see the number of stations in this Figure.

We added the number of station per elevation bin in Fig. 5.

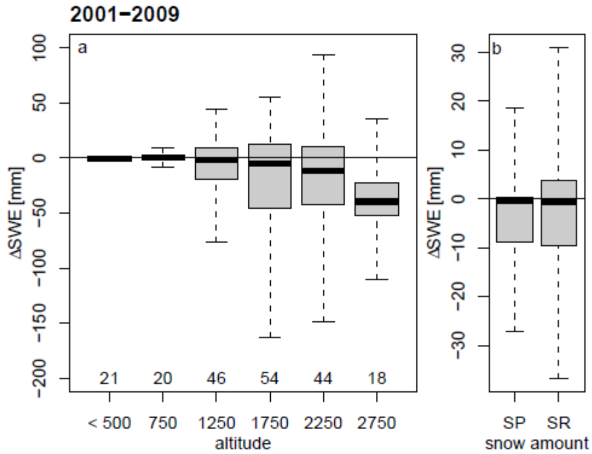


Fig. 5. The mean differences between map110 - map203 per grid cell for: (a) different altitude ranges and (b) snow-rich (SR) and snow-poor (SP) days in the whole of Switzerland during the overlapping period. The numbers at the bottom of Fig (a) are the numbers of stations per elevation bin. The boxes display the median and the interquartile range. The whiskers extend to the maximum ΔSWE, but are limited by twice of the interquartile range.

16) Page 4256 line 12: The description of this validation method would fit better into section three. How did you select the 23 stations, and how was their altitudinal distribution? What is the unit of the given differences - is it mm?

This paragraph is shifted to section 3.3 where the selection and altitudinal distribution is now described (see answer to point 2).

The unit of the differences given in section 4.2.2. is millimetres.

17) Page 4258: Please explain ME? Is 29 mm significant compared to the measurement uncertainties of the input data?

For this evaluation we changed the ME to the mean absolute error (MAE). With this score positive and negative errors do not cancel out each other. The MAE is calculated in mm water equivalent and can therefore not be compared to the measurement uncertainties of the snow depth.

In this case, the snow resources simulated with PREVAH are generally underestimated in the Alpine Rhine region (MAE: 30 mm), and partly over- and underestimated in the Valais region (MAE: 16 mm). In the region of the rivers Thur, Töss and Glatt SWE is underestimated in this example (MAE: 48 mm). These errors are in the range of the natural variability.

18) Page 4261: are 10-4Line 29: Why global (i.e. worldwide?)

The 'global' refers to the trend that is calculated if all stations are considered (here 110 or 203 depending on the dataset).

In map110 the non-linear trend of SWE over elevation is flat above 2000 m a.s.l., while this elevation trend is drawn further up to 2700 m a.s.l. in map203.

19) Page 4262 line 15: -18mm , 7*10-3 mm: are these numbers and accuracies significant compared to measurement uncertainties?

We assume that map203 and map110 have the same uncertainties. They are subject to the same measurement and model uncertainty. The mean difference between map203 and map110 is based on the lower number of stations that are available for map110. This difference could be removed with quantile mapping as can be seen in the bottom row of Figure 8.

Because map203 and map110 are based on the same measurement and model uncertainty their difference is based on the lower number of stations that are available to produce map110. This difference could be removed with the calibration method quantile mapping.

20) Table 1: please give all area ratios at the same accuracy
OK

21) Table 2: add abbreviations to the caption 0.70 instead of 0.7
OK

22) Figure 1: The application of the d133 is unclear, where are the 23 selected stations? The caption and the legend do not fit together (color descriptions). The text states that 23 stations were used for validation?
Thanks for this remark. The 23 stations are the ones that are available in a dataset of 133 from 1989-2009 but not in d110. We know do not mention the 133 stations and hope we could clarify this issue. We changed the legend in Figure 1.

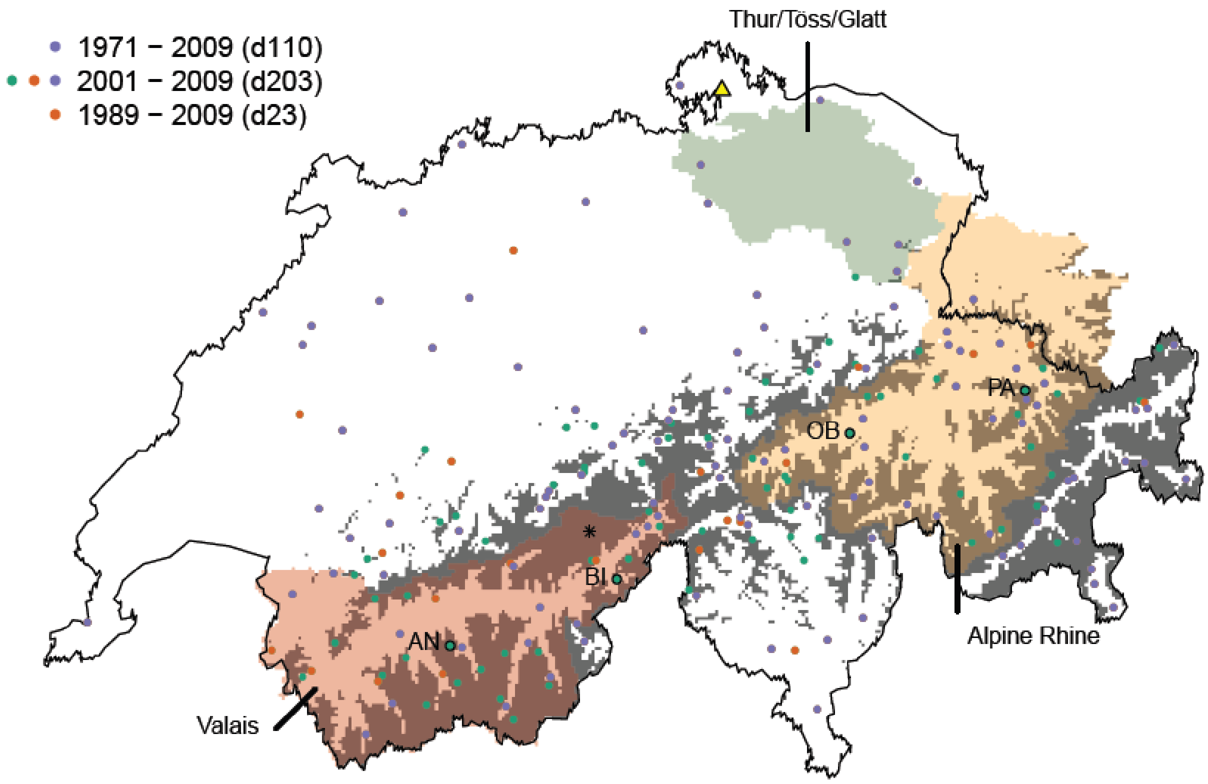


Fig. 1. Stations available during the different time periods: 110 stations from 1971-2009 (d110, purple dots), 203 stations from 2001-2009 (d203, purple, orange and green dots) and 23 stations from 1989 - 2009 (d23, orange dots). The three sub-areas, Alpine Rhine, Valais and the region including the rivers Thur, Töss and Glatt are considered in this study. Areas shaded grey are higher than 2000 m a.s.l. The four labeled stations with black outlines are used for validation. The yellow triangle is the river gauge in Neuhausen. The black star is a randomly chosen grid cell for the example in section 4.3.

23) Figure 4: label y axis
OK, y axis is labelled. We assume that the reviewer refers to the missing axis label in Figure 10.

24) Figure 13: Describe Q 1-99, Q25-75 in the caption

OK, this has been added to the caption.

25) References: Blöschl, not Bloschl
OK.

References:

Zappa, M., Bernhard, L., Fundel, F., and J\"org-Hess, S.: Forecasts and scenarios of snow and water resources in Alpine environments [in German], Forum f\"ur Wissen, 19-27, 2012.