Answers to W. Schöner

We would like to thank W. Schöner for his careful and constructive review of our paper. 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) Though the title of the paper is clear with respect to the aim of the paper I would suggest considering improvement (e.g. "Exploring the potential for SWE climatology in Alpine terrain using empirical approaches").
Thank you for motivating us to rethink about our title. We changed the title of the revised manuscript.

Homogenization of a gridded snow water equivalent climatology for Alpine terrain: methodology and applications

Detailed comments:
2) Under "2.2. Snow observations" a short statement on data quality of snow data has to be given.
The data accuracy has been specified on page 4247 line 8:"The measurement accuracy depends on the equipment and accounts for about ±2 cm."

3) Figure 2 illustrates the quantile mapping method. However, the method is well explained in the text and several references are given. Thus I suggest to skip Figure 2.
Thank you for this suggestion. We skipped Figure 2 and moved Figure 15 to section 3.2.

4) The reading of the paper could be improved by clearly keep apart the chapters on methods and results. In particular the validation method is explained under methods but again methodical concepts of validation are shown again in the results chapter.
In the revised version the methodological parts of the results are summarized in section 3.3. We hope this helps to improve the readability and clarifies some aspects.

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$^-$.

5) The authors split the sample of SWE in snow-rich and snow-poor days. Is there only statistical reason (variance) for this or is there some more motivation from physical based mechanisms?
We tested several options and got better results in dividing the dataset in snow-rich and snow-poor days. This makes sense since the spatial variability depends on the snow depth. E.g. Egli and Jonas (2009) demonstrated that the standard deviation of snow depth increases with snow depth. Therefore the slope of the CDF depends on the snow depth. As the distribution of snow in the terrain is different for different snow depth it makes sense to use different classes for different snow depth. The number of classes depends on the data.

The classification into snow rich and snow poor days is justified as the spatial distribution of snow depends on the snow depth (Egli and Jonas, 2009).

6) Chapter 4 is related to validation of quantile mapping. 4.4.1. the term "Crosscalibration" is introduced. What is meant with validation of cross-calibration during calibration period? Is it not simple validation of calibration period?
Yes, it is a validation of the calibration period. This is changed in the text and the title.

4.4.1 Validation of the calibration period
In order to validate the calibration period (...)

7) In 4.4.2 the spatial and temporal consistence is tested using 4 stations left out from the data set for calibration. However three out of the four stations appear to be spatially quite close to stations kept for calibration. Thus the validation appears not really robust to me.
We appreciate that you have drawn our attention to this problem. For this verification we tried to find stations that have not been used before 2001 and that are known to be representative for the region. In addition we wanted to look at different elevations. The distance of the stations to other stations kept for the SWE estimation is at least 5 km, which can be a large distance with respect to the snow distribution in Alpine terrain. As the interpolation of the snow model used in this study uses a Gaussian Filter (and not inverse distance weighting) we think that this procedure helps us to get a feeling for the accuracy of the SWE and helps to quantify the uncertainty.

8) Figure 14 is not clear with respect to the explaining text in the manuscript. In the text it is written that linear regression is computed for logarithmic axes of SWE and runoff. The scatter plot in Figure 14 is for the non-logarithmic data but the regression shown is for logarithmic data. This needs some clarification in the caption for Figure 14 (and would be useful to make it more clearly in the text, too). Additionally, I feel not confident with the example of Figure 14 showing the potential of SWE maps for low flow estimation. In fact the prediction band of the regression model is quite large. Is this really useful for practitioners?
We tried to clarify Figure 14 in the caption and the text. We agree that the prediction band is quite large. This is to the fact that temperature and precipitation have also an important effect on the minimal discharge. But our idea is that practitioners should be interested only in the lower line of the prediction band to get information of the worst case low flows that can be expected in the following

month. This is a first idea of how SWE data can be used to predict discharge. We hope to further develop this approach.

(…) A linear regression is carried out for the log-transformed variables, SWE and runoff. This linear regression is transformed to the non-logarithmic space (black curve). The 95 % predictive interval is calculated to specify the range where minimal runoff can be expected (Weisberg, 2005) (dotted lines). (…)
For practitioners the lower curve of the prediction interval should be of interest. This line predicts the worse case scenarios of minimum discharge for the following May, June and July given the 15 April SWE. (…)

9) The large number of abbreviations used in the study is a bit confusing and reduces the readability e.g. the different grid data sets used. A table on the grids on their role in the study would be helpful.
We removed many abbreviations and hope we could improve the reading flux. A table of the grids is now introduced in section 2.2.
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*

Minor comments:
P4245/7: delete "considerably"
OK, considerably has been deleted.

P4251/11: Ad km to 1x1 grid.
Changed to: “1 km by 1 km”.

P4257/5: SWE climatology is not a tool but a data set for hydrological model calibration/validation.
Changed to “data set”.

Figure15: units for SWE are missing.
The y axis of Figure 10 has been added.

For several Figures the axes labels are rather small.
Thanks for the comment. We increased the font size in some plots.

References

Egli, L., and Jonas, T.: Hysteretic dynamics of seasonal snow depth distribution in the Swiss Alps, Geophysical Research Letters, 36, L02501, 2009.