Answers to Anonymous Referee #3

We would like to thank the anonymous Referee #3 for the detailed and helpful comments. Based on your remarks and suggestions we revised some parts of the manuscript. In the following we will address your comments.

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

1) When analyzing the resulting maps after applying the stated procedures, you use map203 as the reference dataset. You probably did this, because there is no other independent dataset available that you could use to validate your methods of interpolation and remapping. However, this is a bit unsatisfying, as the reader simply has no clue of the quality of the resulting maps. You tried to overcome this with the leave-one-out procedure, but this is still not enough for the gridded datasets. The measurement stations used as basis for the gridded fields are usually located in flat and accessible terrain and do not represent other locations such as very steep slopes, different expositions, wind-exposed ridges,… (but these locations cover large parts of the presented terrain). The used interpolation method produces a nice elevation-dependent field with some regional differences induced by the horizontal distance weight, but cannot account for processes like e.g. wind-blown snow over crests, large northern oriented shaded faces vs. southern exposed terrain. On the one hand, the stations do not cover these areas, and on the other hand, the interpolation method does not account for additional terrain parameters such as slope or exposition. The problem might be a bit "smoothed out" with the presented 1 km resolution, but in the high alpine parts of the catchment, these processes and the corresponding variability play a role also at the presented scale. This leads me to the next remark: why or how did you confine the 1 km resolution? Have you also produced maps with a finer resolution and evaluated these results? These issues result in the following suggestions: It would be highly useful to have a reference dataset to validate the produced maps at least for some years or dates. This could be data of airborne laser scans for a limited area within the catchment and of course limited to single years. You could also use snow coverage at certain dates derived from satellite-based sensors such as Landsat ETM+ and compare these to the coverage in your maps. Even if this approach is not able to validate the amounts of SWE, this would give you insights in the performance of your interpolation methods. At least point out these issues and uncertainties and try to elaborate further validation in future work.

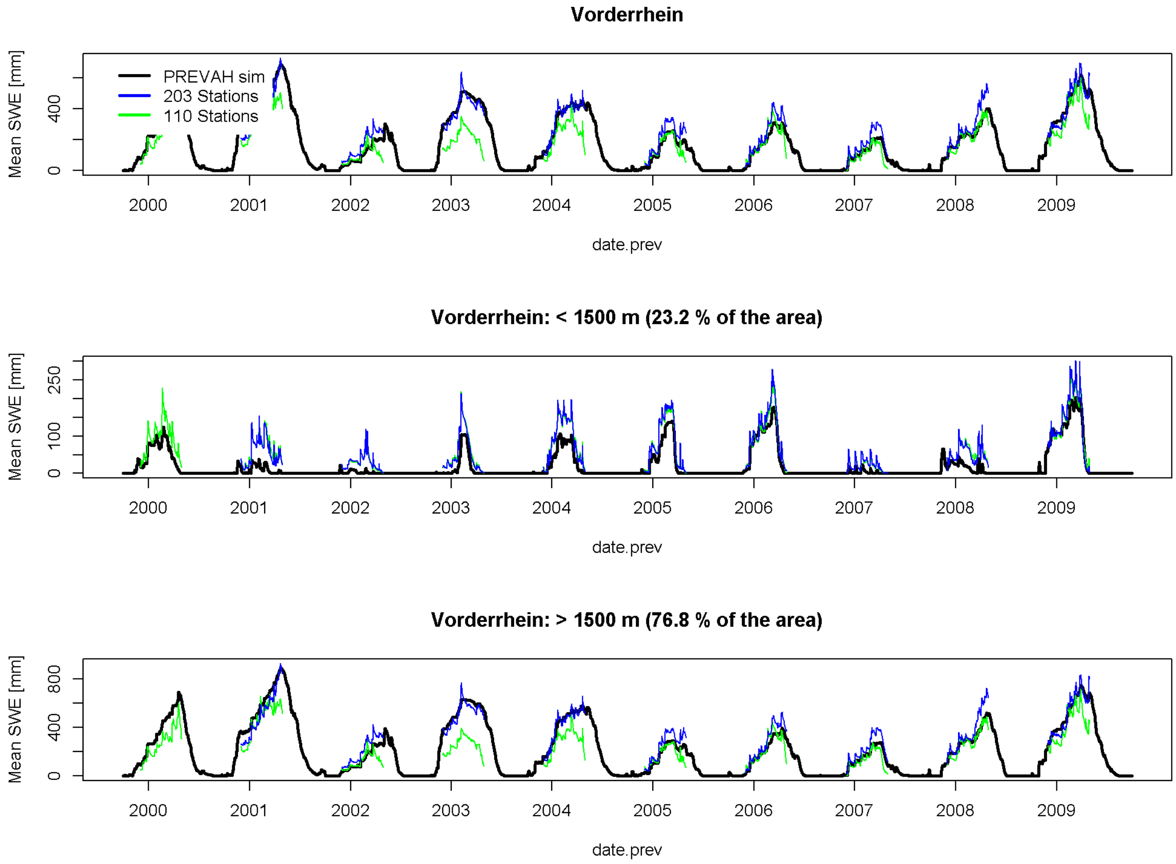
Thanks to the anonymous referee for this comment. The focus of this paper is on the methodology to homogenise gridded SWE data sets. Therefore the accuracy of the input dataset is not crucial and the validation of the input is kept to a minimum. The dataset that we are using is chosen to demonstrate the procedure because it is the best available dataset of SWE for Switzerland. We agree with the reviewer that for the application of such datasets it would be interesting to have knowledge of the accuracy of this data.

The presented input dataset to illustrate the homogenization procedure is a monitoring product that is used operationally in Switzerland. This product is not able to represent small scale variability for entire Switzerland. The resolution of 1 km that was chosen for this work is the finest resolution that can be produced with the used input data. We do not want to use a finer resolution where we have insufficient data. The procedure to homogenise gridded SWE data is independent of the resolution. The question of reasonable resolution and spatial validation of the input dataset could be addressed in a follow up paper.

Future work with gridded SWE will involve the extension of the work by Fundel et al. (2013), with probabilistic and spatial evaluation of both SWE and discharge predictions.

2) I would highly appreciate if you would add a spatial validation of the presented distributed model PREVAH in section 5.1 besides the catchment averaged SWE. This could demonstrate impressively the value of the spatially distributed SWE maps!

We are aware of the potential of this data set for spatial validation of simulated SWE. As the focus of this manuscript is not on validation of hydrological models, we would like to stop the validation in this contribution at the point of Fig. 12. The scope of an actual research is the spatial validation of SWE simulations with the hydrological model PREVAH. We would like to show you a Figure with an overview of the performance of PREVAH for the Vorderrhein below and above 1500 m a.s.l.



Specific Comments

1) P. 4243, L. 15 What "long-range model forecasts" do you mean here? Hydrological models (discharge forecasts), atmospheric models (weather or climate simulations), or others? Please specify!

We specified "long-range model forecasts" to "long-range forecasts (seasonal) of hydrological models".

2) P. 4243, L. 25 Additionally, gridded SWE maps are of high value not only for hydrological model calibration, but to validate and develop these models. Particularly, distributed hydrological models with a focus on snow that include specific processes e.g. lateral snow transport would highly benefit from these validation datasets.

Such models are presented e.g. in these publications:

- Garen, D.C., and D. Marks (2005), Spatially distributed energy balance snowmelt modelling in a mountainous river basin: Estimation of meteorological inputs and verification of model results, J. Hydrol., 315, 126– 153, doi:10.1016/j.jhydrol.2005.03.026
- Lehning, M., H. Löwe, M. Ryser, and N. Raderschall (2008), Inhomogeneous precipitation distribution and snow transport in steep terrain, Water Resour. Res., 44, W07404, doi:10.1029/2007WR006545

- Liston, G., and K. Elder (2006), A distributed snow-evolution modeling system (Snow- Model), J. Hydrometeorol., 7, 1259–1276, doi:10.1175/JHM548.1
- Strasser, U., M. Bernhardt, M.Weber, G. E. Liston, andW. Mauser (2008), Is snow sublimation important in the alpine water balance?, Cryosphere, 2, 53–66, doi:10.5194/tc-2–53-2008
- Warscher, M., U. Strasser, G. Kraller, T. Marke, H. Franz, and H. Kunstmann (2013), Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps, Water Resour. Res., 49, 2619-2637, doi:10.1002/wrcr.20219

We added some of the suggested publications to our references and added the additional value of validation and development of hydrological models.

(…), the parameter calibration, validation and development of hydrological models (Garen and Marks, 2005;Liston and Elder, 2006;Parajka and Blöschl, 2008;Warscher et al., 2013); (…)

3) P. 4244, L.16-18 "The spatial patterns of the snow depth calculated with this model agree well with the precipitation maps of the Alps contained in FOEN (2010), which are based on a different approach." Either skip this sentence or explain the methodical differences in creating these two maps. I guess that both are created by a more or less refined or extended regression with height calculation, so the matching patterns are really no surprise and don't tell anything about their quality (neither of the precipitation map, nor of the snow depth product).

We skipped this sentence.

4) P.4247, L. 18-20 Yes, I agree, but the lake also decreases the natural fluctuations! So I think this is not a reason to look at this large catchment. You could just skip this explanation because it is actually appropriate to investigate just one big catchment if this solely serves as an example for applying the SWE maps. Even if not the focus of this work, it would be very interesting to enhance the analysis to a smaller alpine headwater catchment, where anthropogenic influences are small or not present. I am quite sure you could find such a (small) subbasin within your catchment. The influence of snow on discharge could be investigated more clearly in such a catchment. However, in my point of view, for this manuscript it would be reasonable to just skip the misleading explanation (as stated above).

We skipped this sentence.

5) P. 4248, L. 22-24 How were the optimal filter widths identified?

The filter widths were identified by optimization using a leave-one-out cross-validation.

Optimized filter widths for snow mapping in Switzerland were identified by a leave one out cross-validation and are roughly around 25 km horizontally and 500 m vertically, depending on the station density and season.

6) P. 4249, L. 10 I guess you mean "correction" instead of "calibration"

Yes. Calibration has been changed to correction.

7) P. 4250, L. 20 What do you mean by "The data was stratified. . ."? Perhaps just skip this sentence!

We rephrased the sentence and hope it is clear:

To account for different distributions of SWE in different altitudes, regions and snow amounts we performed quantile mapping separately for subsets of the data predefined in section 4.2.

8) P. 4251, L. 2 I don't understand where these 330 mm RMSE (SWE, HS?) come from. Please explain!

The assumption is that a model represents the real snow depth of 2m with a mean natural variability of ± 15% with a homogeneous snow depth of 2 m. Then the RMSE of 150 randomly selected stations of the modelled and observed snow depth would amount to 33 cm. This error was calculated by using a measured variance for a normal distribution.

However, given that station data represent single point measurements with natural deviations from the mean, the resulting RMSE of the real snow depth in the model domain and the modelled snow depth would amount to 33 cm, even though the mapping model is otherwise perfect.

9) P. 4251, L. 11 "Averaged over all stations, SWEorig outperforms SWEloo-cv (Fig. 3). "I can't really detect this in Fig. 3. And: isn't this an obvious result?

Yes the result is what we expected. But the nice result is, that the difference is not really high, what gives us confidence for the modelled values between the stations. The RMSE and R^2 averaged over all stations (listed in the upper left of the Figure) state the better performance of SWE_{orig} compared to $SWE_{loo\text{-}cv}$.

10) P. 4251, L. 16 As you state here, the uncertainties increase from December to April. Could you please show this or give an error range for spring time values. This increase in uncertainty is actually very important for the following analyses, so please show it! How exactly was the SWE comparison in Fig. 3 done? At a certain point of time? Mean SWE values? Please give some additional explanation.

The comparison of SWE_{orig} and $SWE_{loo\text{-}cv}$ was done for each station separately over the period 2001-2009. In the revised version of the manuscript we include a Figure that compares the distribution of the RMSE for the winter months December-April. For the Figure in this letter we include additionally to the RMSE, the relative RMSE and R^2. From this Figure the increase in RMSE and rel. RMSE towards the end of season can be clearly seen. R^2 on the other hand decreases towards the end of season. As the same conclusion can be drawn from the three scores in the revised manuscript only the seasonal distribution of the RMSE will be shown.

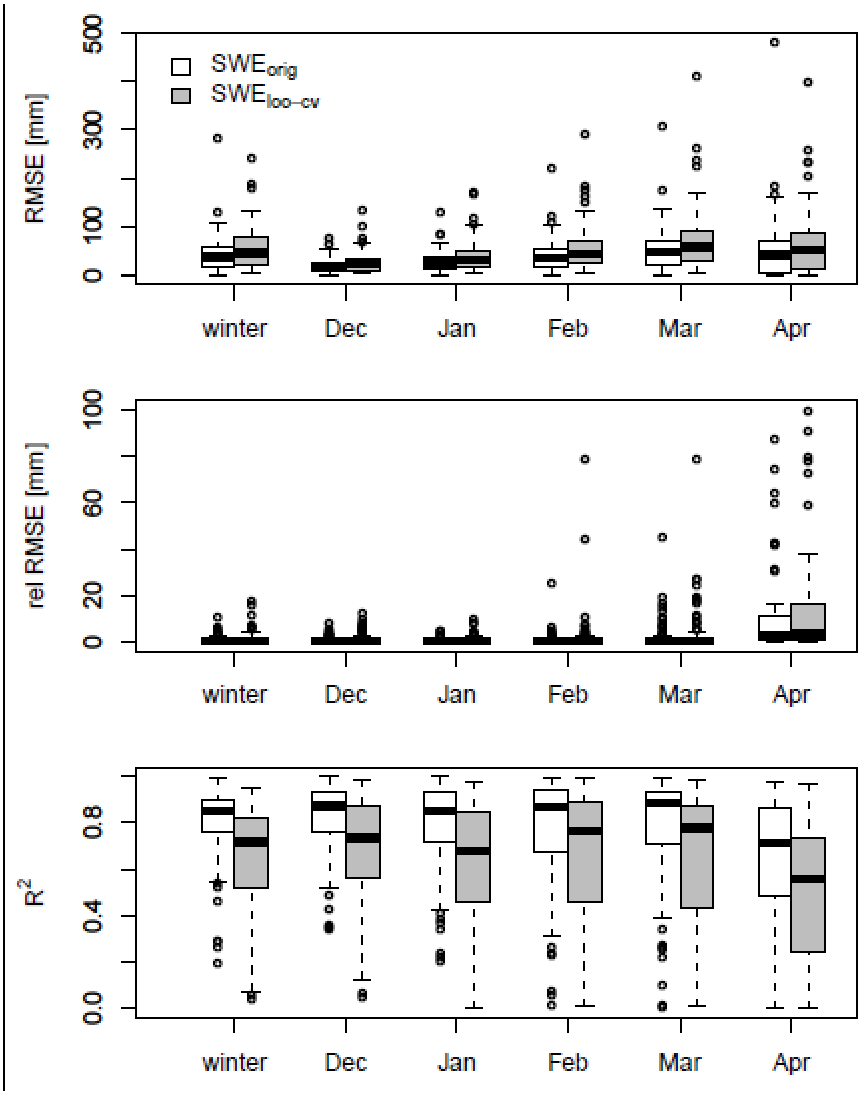


Fig. 3. The effects of season on the root mean squared error (RMSE) and the coefficient of determination (R2) over the period 2001-2009. White boxes contain the scores for all stations calculated with SWE_{orig} and the grey boxes those based on $SWE_{loo\text{-}cv}$.

(...)From December to April, the uncertainty increases at many stations (Fig. 3) because the melting process causes large SWE differences between high and low altitudes. Generally SWEorig outperforms SWEloo-cv. (…)

11) P. 4251, L. 26 See comment above, what is meant by "stratify" here? The splitting in calibration and validation period?

With "stratify" we mean finding systematic differences that make it necessary to subdivide the dataset in different subsets. To clarify the sentence now reads:

The differences between map110 and map203 (ΔSWE = map110 - map203) was analysed for all of Switzerland during the overlapping period, to identify systematic errors and to find a meaningful way to subdivide the data in different subsets for the calibration procedure. (...)

12) P. 4253, Sect. 4.3 Please extend this section, as it is a central method of this work, and it is quite hard to understand with these few sentences. You could also think about merging it with the two following sections.

We tried to clarify this section and hope it is now easier to understand.

4.3 Implementing the quantile mapping

The outcome of the comparison of map110 and map203 in the previous section led to the definition of the calibration procedure illustrated in Fig. 7. The calibration of map110 is implemented separately for each grid cell to account for spatial and altitude-dependent differences, and for each day with a moving window of ±15 days to account for seasonal effects. Because maps are only available from 1 December till 30 April reduced classes were used for dates before 15 December and after 15 April. Additionally snow-rich and snow-poor days were distinguished by means of the median SWE in map110. The ECDF of each sub-dataset were used for the calibration with quantile mapping. Finally a calibrated data set of gridded SWE maps for 39 years (map.cal) was produced.
During the nine overlapping years the calibration of a specific grid cell "x" (star in Fig. 1) on 11 March 2003 all days of this grid cell from 24 February to 26 March except those from the year 2003 are used to produce the ECDF's. Estimated SWE from map203 is used to produce $F^{h}_{11Mar,x}$ *and estimated SWE from map110 is used to produce* $F^{l}_{11Mar,x}$*.(…)*

13) P. 4255, L. 10 and Fig. 7 You can skip this result or at least Fig. 7, because it comes at no surprise, that the calibrated data fits the data you have calibrated it with. This kind of illustrates the major criticism I stated above, that no independent reference dataset is used for validation. I know, it is probably not possible because of the non-existence of such data, but then just state this and remove the figure.
This Figure has been skipped as suggested by the reviewer.

14) P. 4257, L. 6-7 ". . ., which can generally only be verified against runoff observations." Physically-based, distributed hydrological models are also verified against other variables, e.g. turbulent fluxes, soil moisture, surface temperature, . . . Regarding snow, hydrological models are often validated using remotely sensed data of snow coverage (e.g. Landsat ETM+). This is of course a limited approach, because of the missing snow mass information. Additionally, SWE or snow depth values derived from airborne laser scans are used. Your SWE maps are absolutely undoubted very valuable, so either add these validation methods or just skip the statement.
We now refer to spatial data that can be verified in hydrological models.

A lot of work is going on in verifying hydrological models against spatial data as remotely sensed data of e.g. snow (Hüsler et al., 2013;Hüsler et al., 2012;Bellinger et al., 2012;Dijk and Renzullo, 2011;Zappa, 2008;Andreadis and Lettenmaier, 2005),evapotranspiration (Kite and Droogers, 2000;Immerzeel and Droogers, 2008) and soil moisture (Bastiaanssen et al., 1998).

15) P. 4257, L. 7-10. Please rephrase the sentence, as this may be true for your application and model, but not in general.
Now reads:

SWE simulated with hydrological models is typically calculated with precipitation and temperature (Zappa et al. 2003). In this case the simulated SWE is completely independent of the calibrated SWE maps, as the SWE maps are based on measured HS and calibrated bulk density.

16) P. 4258, Sect. 5.2 If you show this example, please slightly elaborate this section and add some explanation of e.g. the consequences of the anomalies for the reservoir management.
We added some background information to elaborate this example. The Swiss government recently started a new research program called "energy turnaround". The SWE-maps will play here an important role because it is aimed to improve efficiency of hydropower systems by reliable seasonal water resources forecasts. Also in this case we can suggest the reviewer to follow our next contributions.

(...)Knowledge of the snow water resources and their anomaly as compared to long-term climatology may help operators of hydropower dams in seasonal planning of production and in the coordination of production from different reservoirs located in different regions.(...)

Technical Corrections

P. 4242, L. 1 "Gridded snow water equivalent (SWE) are valuable. . ." I think it is better to write e.g. "Gridded snow water equivalent (SWE) data are valuable. . ." or ". . . datasets are. . ." or "Gridded SWE observation products are. . .".

Thank you for this remark. Now reads:

Gridded snow water equivalent (SWE) datasets are valuable (…)

P.4242, L. 2 ". . . and verify hydrological models and other models. . ." Just write "hydrological models" or name the other models, e.g. write "verify different model systems, e.g. hydrological, land surface, or atmospheric models.

Now reads:

Gridded snow water equivalent (SWE) datasets are valuable to estimate the snow water resources and verify different model systems, e.g. hydrological, land surface, or atmospheric models.

P. 4243, L. 23 "Blöschl" instead of "Bloschl". Please check also other appearances!

Changed to "Blöschl"

P. 4245, L. 7 Typo, delete one of the two "considerably"

OK, one considerably has been deleted.

P. 4247, L. 10-12 Please merge the two sentences, e.g. like this: "Based on daily, gridded SWE maps (d110, d203, and d133), gridded SWE climatology maps (map110, map203, and map133) were produced for Switzerland with the model described in Sect. 3.1. "

Merged as follows:

Based on daily snow depth measurements (d110 and d203), gridded SWE climatology maps (map110 and map203) were produced for Switzerland with the model described in section 3.1.

P. 4249, L. 8 Delete "the"!

OK.

P. 4251, L. 12 "(1 x 1 grid)" add "km"

Changed to: "1 km by 1 km".

References

Andreadis, K. M., and Lettenmaier, D. P.: Assimilating remotely sensed snow observations into a macroscale hydrology model Adv. Water Resour., 29, 872-886, 10.1016/j.advwatres.2005.08.004, 2005.

Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M.: A remote sensing surface energy balance algorithm for land (SEBAL) - 1. Formulation, J. Hydrol., 212, 198-212, 10.1016/s0022-1694(98)00253-4, 1998.

Bellinger, J., Achleitner, S., Schöber, J., Schöberl, F., Kirnbauer, R., and Schneider, K.: The impact of different elevation steps on simulation of snow covered area and the resulting runoff variance, Adv. Geosci., 32, 69-76, 10.5194/adgeo-32-69-2012, 2012.

Dijk, A. v., and Renzullo, L.: Water resource monitoring systems and the role of satellite observations, Hydrol. Earth Syst. Sci., 15, 39-55, 2011.

Garen, D. C., and Marks, D.: Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results, J. Hydrol., 315, 126-153, 10.1016/j.jhydrol.2005.03.026, 2005.

Hüsler, F., Jonas, T., Wunderle, S., and Albrecht, S.: Validation of a modified snow cover retrieval algorithm from historical 1-km AVHRR data over the European Alps, Remote Sensing of Environment, 121, 497-515, DOI 10.1016/j.rse.2012.02.018, 2012.
Hüsler, F., Jonas, T., Riffler, M., Musial, J. P., and Wunderle, S.: A satellite-based snow cover climatology (1985–2011) for the European Alps derived from AVHRR data, The Cryosphere Discuss., 7, 3001-3042, 10.5194/tcd-7-3001-2013, 2013.
Immerzeel, W. W., and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, J. Hydrol., 349, 411-424, http://dx.doi.org/10.1016/j.jhydrol.2007.11.017, 2008.
Kite, G. W., and Droogers, P.: Comparing evapotranspiration estimates from satellites, hydrological models and field data, J. Hydrol., 229, 3-18, 2000.
Liston, G. E., and Elder, K.: A distributed snow-evolution modeling system (SnowModel), Journal of Hydrometeorology, 7, 1259-1276, 10.1175/jhm548.1, 2006.
Parajka, J., and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, J. Hydrol., 358, 240-258, 10.1016/j.jhydrol.2008.06.006, 2008.
Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., and Kunstmann, H.: Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps, Water Resour. Res., 49, 2619-2637, 10.1002/wrcr.20219, 2013.
Zappa, M.: Objective quantitative spatial verification of distributed snow cover simulations - an experiment for the whole of Switzerland, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 53, 179-191, 10.1623/hysj.53.1.179, 2008.