The Cryosphere Discuss., 8, C839–C848, 2014
www.the-cryosphere-discuss.net/8/C839/2014/
© Author(s) 2014. This work is distributed under
the Creative Commons Attribute 3.0 License.

The Cryosphere
Discussions
Open Access

Interactive comment on "Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra- and inter-annual persistence" *by* J. Revuelto et al.

Anonymous Referee #2

Received and published: 30 May 2014

1 General comments

1.1 Summary of goals, approaches and conclusions

Revuelto et al. studied with terrestrial laser scanning (TLS) data the snow distribution in a small alpine catchment in the Pyrenees. In two very different winters 12 observation days of high resolution snow depth measurements were related to topographic variables derived from digital elevation models (DEM). Their aim is to identify main topographic factors explaining the observed snow depth distribution and their evolution in time. The statistical explanatory of the topographic variables were tested with univariate and multivariate methods which are typically used to analyze snow depth

distribution. They concluded that topographic variables related to terrain curvature are the best explanatory variables, especially one variable which was rarely used in this context, the topographic position index (TPI). These variables showed persistent explanatory power in the statistical models, although the observations days covered very different, if not extreme situations. Other variables which interpret the DEM along a certain direction (e.g. wind shelter, radiation) differ between observation days.

1.2 Conceptual overview of the contribution of the paper

The main questions of the manuscript whether topographic variables are able to explain snow distribution in a catchment and whether they are variable over intra- and inter-annual time scales, are interesting and important within the scope of The Cryosphere. The knowledge of which topographic variables influence snow depth distribution has a large impact for a better understanding of runoff in snow-melt dominated catchments. This manuscript delivers and incremental advance in a series of recently published papers using high-resolution snow depth data: 12 observations in the course of two very different winter seasons allow to extend existing knowledge of the main questions to the presented interesting cases. The presentation of a very predictive variable (TPI) which is rarely used in this context may improve also similar attempts in different regions. To address the main questions, the authors used both for data acquisition as well as for data analysis appropriate methods. Some statistical issues are mentioned in the following paragraph. The authors cite relevant literature, but cited studies could be better integrated and compared. Some comparable studies are missing which used low-resolution snow depth data, but over many years of observations. The manuscript shows in general a good presentation quality with some exceptions concerning the figures. One important weakness is that no meteorological data is included, which makes the interpretation of topographical variables difficult. As a result, conclusions quantifying the persistence of topographic control are not very precise (e.g. abstract lines 23f). I also suggest to use the Sx parameter in its original way, averaged over a certain upwind window (Winstral et al., 2002). I will refer to these points in the following paragraph.

I am convinced the authors are able to address my specific comments below. Some of my comments will need a reanalysis of the data. Thus, I recommend to publish this manuscript in The Cryosphere after Major Revisions.

2. Major comments

2.1 Missing non-topographic information: wind, solar radiation and melt

The study focuses on topographic influence only. Since the distribution of snow depth is obviously an interaction of topography and wind characteristics it would be a great benefit if wind information is included in the analysis. Some topographic variables are hard to interpret without the knowledge of the wind conditions beforehand. The analysis related to Figure 4 would greatly benefit if search direction correlates best is consistent with measured wind directions. If wind was measured at the weather station mentioned in Figure 1, a closer analysis relating wind with topographic characteristics is necessary. If not, modelled results of synoptic wind conditions could fill the important missing information.

The TPI interprets the topography in all directions equally, while the Sx parameter takes sheltering effects relative to wind direction into account. Thus, the TPI is not able to explain snow drifts in leeward and less snow in windward slopes. Including wind information in this manuscript seems necessary to explain why the TPI is so important. A comparison of wind speeds to other studies could explain if the found importance of the TPI is site specific or just because other studies have never used it. My interpretation of the results could be: The conclusion that the Sx parameter (see also comments below to the usage of this parameter) explains less than the TPI hints to a rather minor role of snow redistribution in this study. I would like to see a discussion on the wind/terrain interaction in this manuscript. A strict focus on the topography seems not sufficient in this context.

Similarly, radiation was only modelled from terrain information assuming clear sky conditions. Measurements of solar radiation are needed to account for different melt rates.

Winstral and Marks (2014) used modelled melt rates. Such a model attempt would lead this manuscript to a more quantitative interpretation of different influences found in the topographic variables (abstract lines 19ff).

If the authors want to quantify the topographic control, I suggest to study this in relation to including or excluding non-topographic data and how the statistical models improve or deteriorate. If these measurements are not available, this study has a weakness in experiment design. Numerical weather models may close this gap.

In summary, an inclusion of non-topographic information (wind, radiation, melt) would help to precise the conclusion (abstract lines 23f “some similarities in snow accumulation patterns were observed”) and the conclusion referring to a persistent topographic control (Conclusions lines 25ff). When and under which conditions can we expect a persistent topographic control and when not?

2.2 Sx parameter

In relevant publications (e.g. Winstral et al. 2002, Winstral and Marks, 2002, Winstral et al. 2013) the Sx parameter was only used as an intermediate variable. Finally, only an averaged 'Sxdash' parameter was considered in their analysis. This one is averaged over a certain upwind window of 60 deg (Winstral et al., 2002). Later they used 30 deg (Winstral and Marks, 2002, Winstral et al. 2013). So I think the authors are wrong (p 1944, line 25ff: “Rather than considering the contribution from all directions at a specific location, adding all the Sx values for all directions for each cell (Winstral et al., 2002), ...”. With this wrong argument the authors appear to use Sx instead of Sxdash to account for wind directions. I think Winstral's arguments to average over a certain upwind window to create a more robust parameter seems valid, given the large fluctuation of wind speed and directions in space and time. The sheltering effect will certainly be influenced by more then only one upwind grid point. Furthermore, the dominant wind direction may not perfectly be constant in such a catchment, which may increase the catchment-wide correlation of Sxdash with snow depth compared to Sx.

Comparisons with other studies (Schirmer et al., 2011, Winstral et al., 2002), which have used Sxdash instead of Sx, are only possible if the same parameter is considered. I strongly suggest to include Sxdash for similar wind direction in this analysis.

2.3 Statistical analysis

2.3.1 Reducing the dataset

I understand the purpose of reducing the dataset for statistical tests. The number of observations is so large that statistical tests have too much power and are showing significant results for irrelevant differences. Thus, the authors reduced their data set with a Monte Carlo method, e.g. for testing if the population correlation coefficient p is significantly different from 0 given the sample.

However, I do not see a reason why the value of the sample correlation coefficient r should not be calculated for the whole dataset. To my opinion one is interested in the best estimation of the population correlation coefficient, for which in this case the largest number of observations should be best. I can see a benefit in a confidence interval around r in which p will be located. However, the mean, median and range of r values derived by subsamples does not make sense to me (Figure 4 and 5). The ranges is always quite similar in these Figures, which hint to me that there is no additional information (besides from that: for so many observations and for continuous data mean and median should be close to be the same and do not need to be mentioned both). I could not find any arguments in the cited book for this procedure (Hair et al. 1998). If the authors do, please cite these arguments with page numbers in case of citing a book. I have similar arguments for not reducing the dataset to build the linear regression function and the regression trees. For their stepwise linear regression a statistical test was used by the authors. As mentioned above, such a test would have too much power and select too many variables to be included, which leads to a overfitting

regression function. The authors worked around this problem and reduced the dataset as well (lines 16f. p. 1946). I could not find this procedure in the cited book (Hair et al, 1998). Similar to the correlation coefficient, I think the regression function is best when all data is given to the model. The overfitting problem can be solved if the argument for selecting a variable is not a statistical tests, but for example the adjusted rsquare or the AIC, which take the number of freedom into account. Another suggestion would be to separate feature selection (determined with subsets) and the determination of the final regression function (with all data).

More problematic is the overfitting problem with regression trees, since it is not just a linear function which will be trained to fit the data. But instead of reducing the dataset, I suggest to apply a method proposed by Breiman et al. (1995, p.59 – 87, p. 241ff) to determine the model parameter “tree size”. This is done with crossvalidation. In contrary, the stopping arguments used in this study are arbitrarily predetermined by the authors (lines 24ff, p. 1946) and therefore prone to overfitting. I cannot find such a procedure to determine the optimal tree size in Breiman et al., (1995). The presented tree in Figure 7b may be already at a stage of potentially overfitting to the data: TPI is in general negatively correlated to snow depth. For the subset which reaches the node TPI < -0.8 (bottom left of the tree) it is the opposite: Large values of TPI are an indication for larger snow depth. Similarly at the node TPI < 0.33 in the middle of the tree. While for me that is a sign of overfitting, the authors may have a physical argument why the general correlation may be reversed under the conditions determined by the nodes above.

In summary, I suggest to apply the whole dataset to build models, but use other methods instead of reducing the dataset to deal with overfitting problems. If the authors are interested in how much overfitting potential their models have, the quality measures rsquare and Willmott’s D, which are used in this study, can be cross-validated. Reducing the dataset for statistical tests is good practice to my opinion, as done for example to determine the threshold for significant correlations in Figure 4.

2.3.2 Interpretation of relative importance: Multicollinearity

The relative contribution of explained variance by independent variables in linear regression is influenced by correlation between them: the relative contribution of two correlated variables should have be less compared to two uncorrelated. Also rsquare of a model with correlated independent variables will be smaller. This is well described in the cited book (Hair et al., 1998). The authors stated in line 12, p. 1946 that they account to this problem with stepwise regression. I think after stepwise regression multicollinearity cannot be excluded and should be studied and presented to interpret the relative importance of variables (Table 2). A principal component analysis leading to uncorrelated input variables for the statistical methods would be helpful as well to assess how much variance can be explained by simple statistical models.

The same arguments are valid for regression trees. An important input variable may not be selected by the tree algorithm since it is masked by one or more highly correlated other variables. An analysis of multicollinearity will help to interpret Table 3.

2.4 Quantification of persistence

Above I mentioned that the persistence of topographic variables lacks a quantitative analysis. Statistical models which are trained for the whole dataset (all days and all years) can be compared to single day models. Single day models can be validated not only against data of the same day, but against all other days to quantify persistence. Models from other studies can be compared quantitatively, can be applied to this dataset. This would give more value compared to the conclusion in line 14 p. 1955 ("The scores were slightly better than reported in previous research using similar methods.."), without reporting scores nor citing studies.

2.5 Integrating studies from long-term snow observations

Recently, a long-term data set on snow depth distribution was published using snow probes (Winstral and Marks, 2014). They refer to other long-term data sets (Sturm and

Wagner, 2010; Jepsen et al., 2012). It would be a great benefit of this manuscript if the findings of these spatially limited observations are compared to this ALS dataset. Can their hypotheses be confirmed by this data set which covers a better spatial resolution, not only a different region?

2.6 Figure captions

The Figure captions are all too small.

3. Minor comments

Please mention page number in case of citing books (e.g. Hair and Zar). Please provide a formula for the TPI. Lines 18ff, p. 1943: Please provide the extent of the larger DEM, which is used for solar radiation, etc. Lines 6f, p. 1944: Please mention the window size which is used for curvature and slope calculations. Lines 18f, p. 1945: The authors excluded all search distances except of one based solely on the univariate analyis. In combination with others, important information may be lost in a multivariate model. While this may be the only practical solution, I suggest to mention univariate results between search distances, that the reader can follow this decision. Where the r values close quite similar between search distances? How easy was this decision to include only one search distance? Why are the search distances (25 m, 200 m) so different between the two variables? Line 10, p. 1946: Please mention how the coefficient were standardized line 1ff, p. 1953: How can the authors define that snow "remains longer", "melts faster", instead of accumulates more or less. I would guess it is rather the latter since the authors stated that the TPI correlation decreased during melt. And, Figure 3 is too small to identify enhanced snow depth concavities. Line 10ff, p.1953: Please check: I found in Schirmer et al. (2011) dmax = 300 m, not 200 m. In Molotch et al. (2005) I could not find a search distance specified while skimming through, searched also for "200" and found nothing. Line 15f, p. 1955: "independent" dataset for train and test. I am not convinced that a randomly selected dataset from the same day is indeed independent given the high spatial correlation of snow depth in the first tens of meters. This problem is documented for correlated time series which increases artificially the randomly cross-validated quality measures (Elsner and Schwertmann, 1994). A similar problem may occur here as well: Nearby grid points have at this day an higher explanatory power. As mentioned above I suggest to test models against other days which would guarantee independent test and train datasets, or ensure that grid point pairs closer than tens of meters are not appearing both in the testing and training dataset.

4. References

Additional references to those studies cited in the manuscript:

Breiman, L.; Friedman, J.; Olshen, R. & Stone, C. (1998): Classification and Regression Trees. Boca Raton, U.S.A.

Elsner, J. & Schmertmann, C. (1994): Assessing forecast skill through cross validation. Weather and Forecasting, 9, 619-624.

Jepsen, S. M.; Molotch, N. P.; Williams, M. W.; Rittger, K. E. & Sickman, J. O. (2012): Interannual variability of snowmelt in the Sierra Nevada and Rocky Mountains, United States: Examples from two alpine watersheds. Water Resources Research, 48, W02529.

Sturm, M. & Wagner, A. M. (2010): Using repeated patterns in snow distribution modeling: An Arctic example. Water Resources Research, 46, W12549.

Winstral, A. & Marks, D. (2002): Simulating wind fields and snow redistribution using terrrain based parameters to model snow accumulation and melt over a semi-arid mountain catchment. Hydrological Processes, 16, 3585-3603.

Winstral, A.; Marks, D. & Gurney, R. (2013): Simulating wind-affected snow accumulations at catchment to basin scales . Advances in Water Resources , 55, 64 - 79.

Winstral, A. & Marks, D. (2014): Long-term snow distribution observations in a moun-

tain catchment: Assessing variability, time stability, and the representativeness of an index site. Water Resources Research, , n/a-n/a.

Interactive comment on The Cryosphere Discuss., 8, 1937, 2014.