

Comparing methods for gap filling in historical snow depth time series

Johannes Aschauer[1], Mathias Bavay[1], Michael Begert[2], Christoph Marty[1]

[1] WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2] Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

Contact: johannes.aschauer@slf.ch

Introduction

Switzerland has a unique dataset of long-term manual daily snow depth time series ranging back more than 100 years for some stations. This makes the dataset ideal to be analyzed in a climatological sense. However, there are sometimes shorter (weeks, months) or longer (years) gaps in these manual snow depth series, which hinder a sound climatological analysis and reasonable conclusions. Furthermore, ongoing efforts towards homogenization of these manual snow depth time series (Resch et al. 2020, Buchmann et al. 2019) require continuous time series.

For historical time periods (first half of the 20th century), we are limited in the amount of available meteorological variables and lack a dense grid of snow measurement stations for spatial interpolations. In this study we focus on simple approaches to reconstruct historical snow depth time series.

Objective

Evaluate the capability of different methods to reproduce snow depth (HS) data in a single winter of missing data at different stations.

Methodology

The basic workflow:

1. Create synthetic gap of one winter (Nov-Apr).
2. Fill the gap winter with different techniques.
3. Evaluate the method’s accuracy based on the withheld data.

We use three different types of methods which are (a) Regression based methods, (b) Spatial interpolation methods and (c) Temperature index models. In case we need to train model parameters, we use data from the gap winter’s three preceding winters. Model accuracy is calculated by the two score metrics root-mean-square error (RMSE) and mean arctangent absolute percentage error (MAAPE, Kim & Kim 2016).

Data

- We use data from the period 11/1999 until 05/2018 from meteorological observation stations in Switzerland (Figure 1)
- Manually observed daily snow depth data from stations operated by both MeteoSwiss and SLF is used for spatial interpolation and regression methods.
- Daily homogenized mean temperature and precipitation data based on services from MeteoSwiss is used for driving temperature-index models.
- Evaluation is performed at 19 target stations that have a complete record of both snow depth and meteorological data.

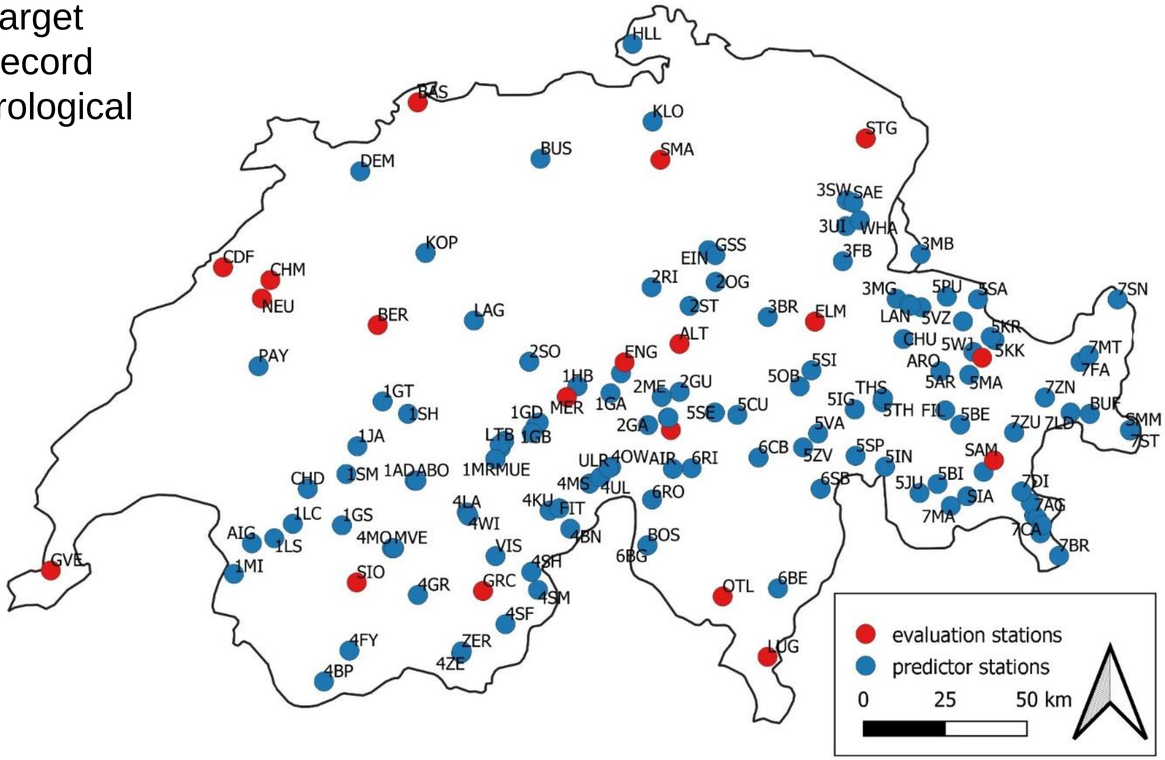

Figure 1: Spatial distribution of evaluation stations and predictor stations.

Regression based Methods

Regression models are trained with three preceding winters of the simulated gap winter (for the winters 2000-2003, the three following years are used for training). The relationship found between the predictors (i.e. HS at neighboring stations) and predictand (i.e. HS at target station) is then used to reproduce snow depth in the gap winter. The relation of neighboring stations and target station has to therefore assumed to be constant over training and gap periods.

PCA Regression

The 5 best correlated neighboring stations are used as initial predictors with standard scaling. A principal component analysis (PCA) is calculated on those predictors and two main components of the PCA are then used as predictors for a multilinear regression model.

Lasso Regression

A linear model with LASSO regularization is used (Tibshirani1996). The 10 best correlated snow depth time series of neighboring stations are given to the model as potential predictors. Predictors and predictand (i.e. target station) are standard scaled during training. The regularization parameter λ is selected during a 5-fold cross validation.

Elastic Net Regression

A linear model with elastic net regularization is used (Zou, & Hastie 2005). The 20 best correlated snow depth time series of neighboring stations are used as predictors. Predictors and predictand are standard scaled during training. Regularization parameters alpha and l1_ratio are optimized in a 5-fold cross validation.

Distance Weighting Methods

The 15 closest neighboring stations within a vertical limit of ±300 m and maximum horizontal distance of 100 km are used to spatially interpolate snow depth in the gap winter at the target station. The number of neighboring stations is reduced if less than 15 stations are satisfying these conditions.

Inverse Distance Squared

Simple inverse distance weighting is used as a benchmark without use of any lapse rate. The inverse square of the distance of a neighboring station is used to calculate its weight.

GIDS

GIDS (gradient-plus-inverse-distance-squared) is a combination of inverse distance weighting and multiple linear regression (Nalder & Wein, 1998). The method is also already used for reconstructing snow depth time series in Austria (e.g. see contribution of Resch et al. here in the same session).

Temperature Index Models

Two different temperature index models are used to calculate the snow water equivalent (SWE) of the snowpack based on daily mean temperature and precipitation.

SNOW-17

An implementation of the SNOW-17 algorithm is used to model both snow depth (HS) and SWE (Anderson 1976). SNOW-17 is a conceptual snow cover model in which the energy exchange at the snow-air interface is calculated based on air temperature only.

SLF Temperature Index Model

A temperature index model that is run by the operational snow-hydrological service at the WSL Institute for Snow and Avalanche Research SLF is used in order to model SWE at the target stations. The model is driven by daily mean air temperature and precipitation.

Density model for transferring SWE to HS

In order to transfer SWE output from the temperature-index models to snow depth, we developed a conceptual snow density model (SWE2HS). The model treats each increase of SWE during a day as a new snow layer. The density $\rho(t)$ in each snow layer at day t after deposition is calculated as

$$\rho(t) = \rho_{max} + (\rho_{new} - \rho_{max}) * e^{-t/\tau}$$

where ρ_{new} is the density of new snow, ρ_{max} is the density of settled old snow, and τ is a decay constant.

If SWE decreases during a day, snow layers are removed from the top of the snowpack in order to compensate for the loss in SWE.

The model parameters ρ_{new}, ρ_{max}, and τ are optimized in order to match measured HS series. This is done in the 3 preceding years of a simulated gap winter. The best parameters identified by the model are then used to estimate HS in the simulated gap winter from the modeled SWE series.

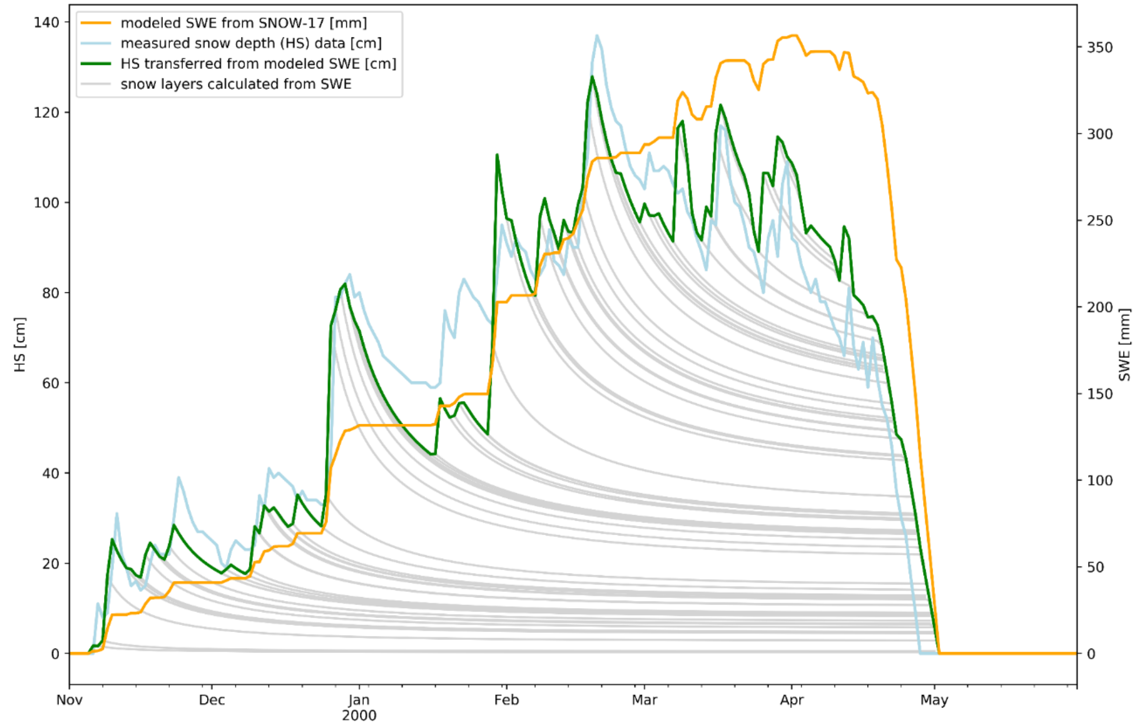

Figure 2: Density model SWE2HS employed at station DAV. SWE was calculated with SNOW-17 .

Results

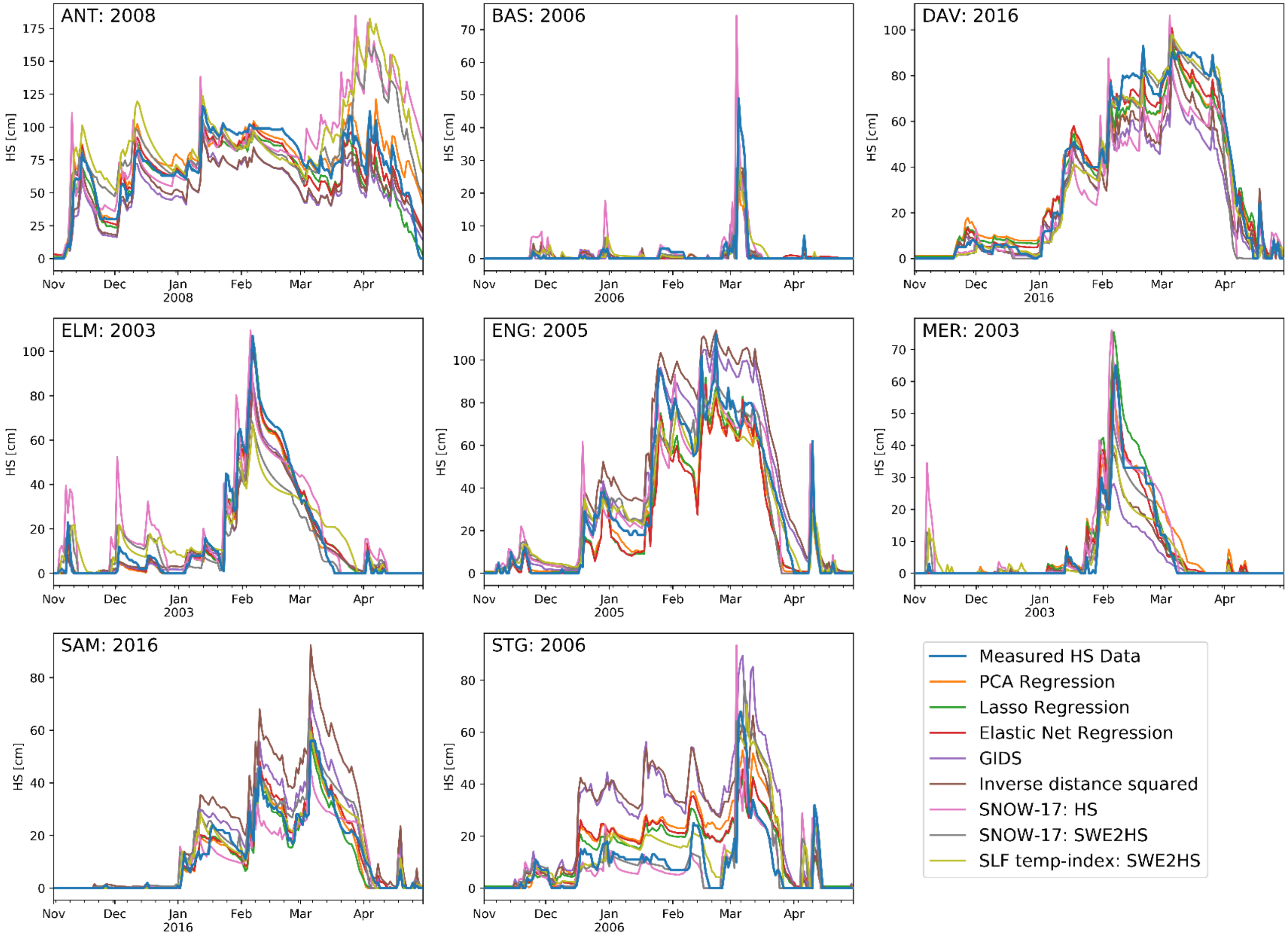

Figure 3: Examples of reproduced gaps with the different methods. Target station and hydrologic year of the gap-winter are indicated in the subplots.

Results

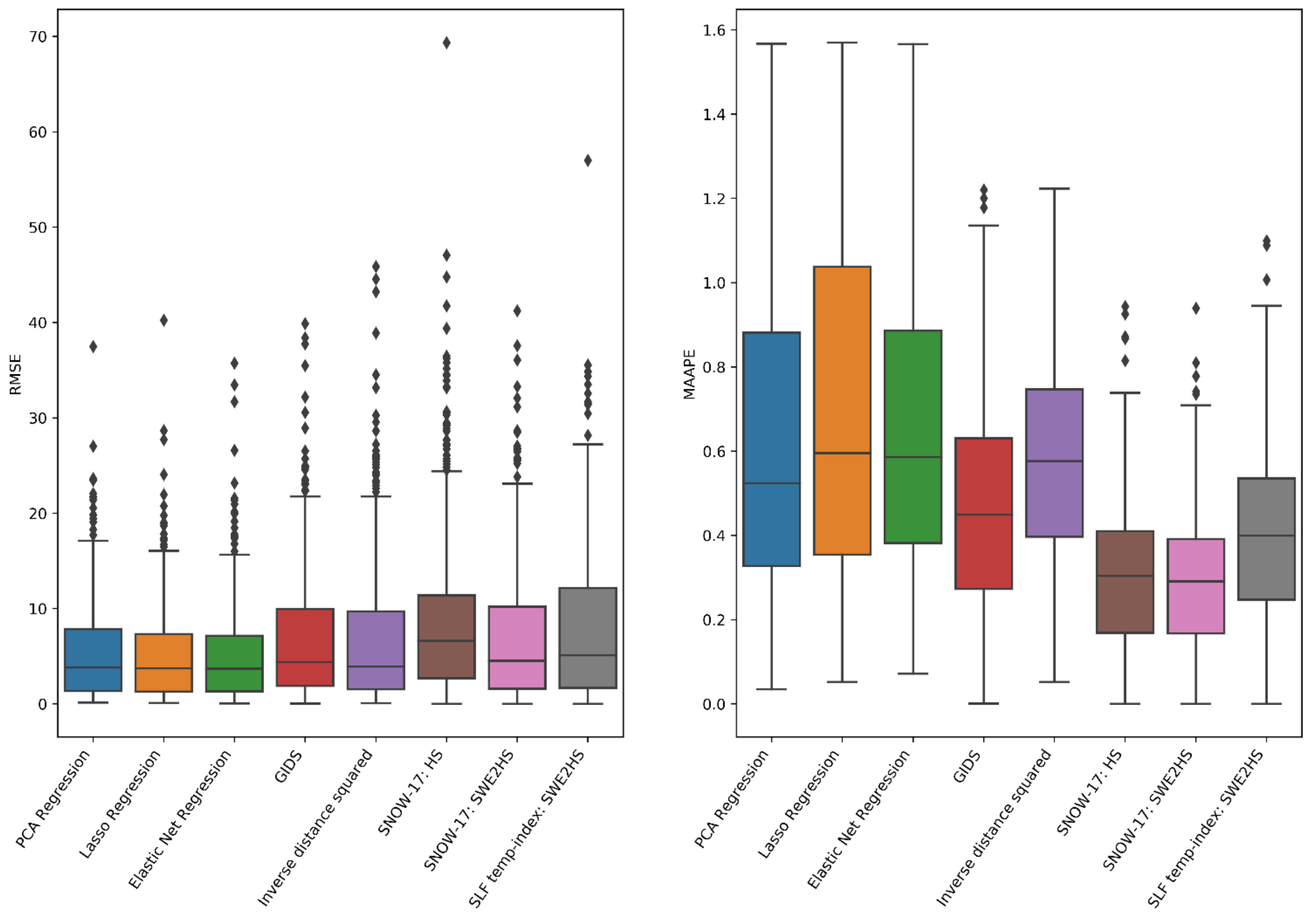

Figure 4: Boxplots of score values (RMSE left, MAAPE right) for the different methods evaluated over all target stations and gap winters.

Results

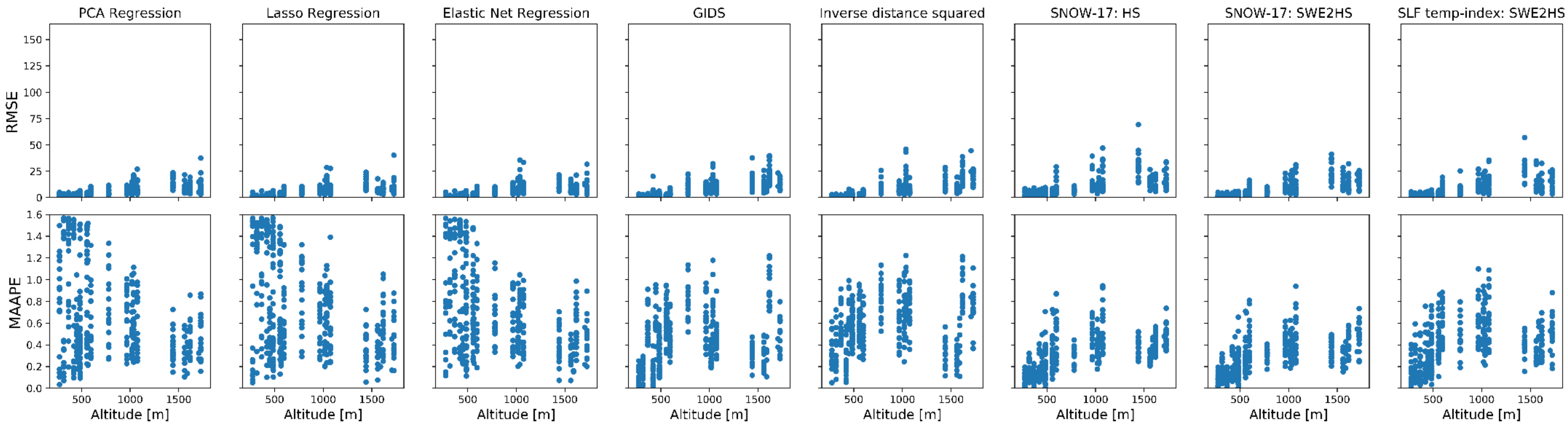

Figure 5: RMSE (top) and MAAPE (bottom) of each filled gap winter plotted against target station altitude. Each column depicts score values for a different method.

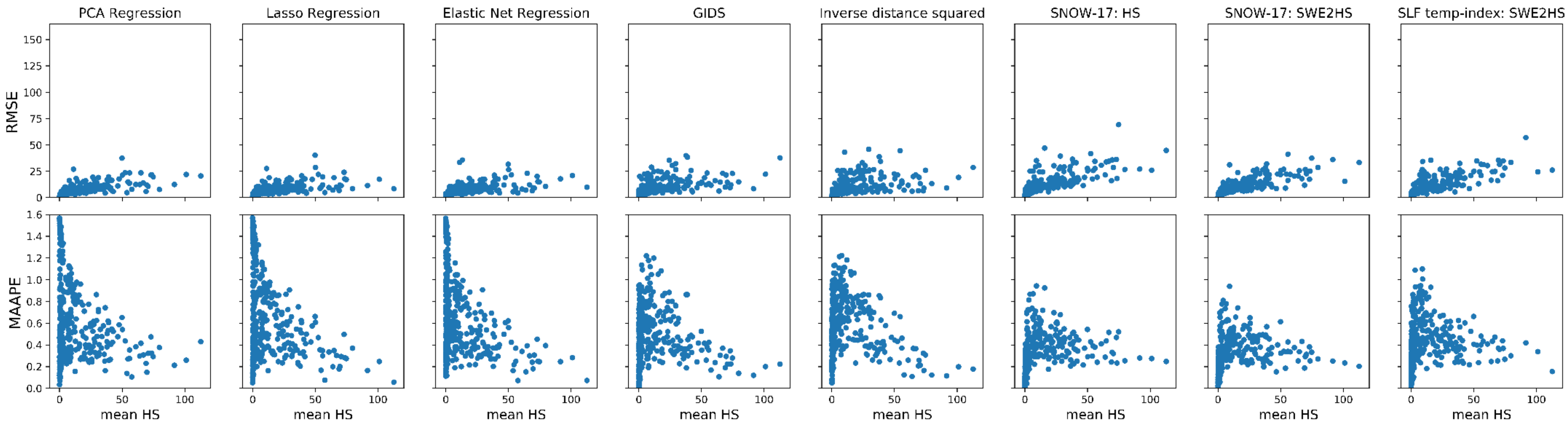

Figure 6: RMSE (top) and MAAPE (bottom) of each filled gap winter plotted against mean snow depth in the respective gap winter. Each column depicts score values for a different method.

Conclusion

- Temperature-index and regression based methods are able to reconstruct historical HS data gaps of one winter with RMSE scores below 20 cm in most example cases.
- Regression based methods do not guarantee good results for low elevation stations and require suitable predictor stations.
- Temperature-index approaches should be preferred over spatial interpolation methods whenever there is meteorological data available.
- When transferring these findings to gaps in the first half of the 20th century, it should be kept in mind that we will have a much more sparse observation grid and accuracies of the spatial interpolation methods are likely to decrease.
- Spatial interpolation methods can compete with the other approaches when introducing vertical limits for the predictor stations.

Further investigations:

- a comparison of the here presented methods with the SNOWGRID model (Olefs et. al 2013), which has been already applied for HS gap reconstruction in Austria.
- In case there is no meteorological data at a station, is it better to first interpolate meteorological variables (temperature and precipitation) and then model snow depth with the temperature index approach or should direct spatial interpolation of HS be preferred?

References

Anderson, E.A., 'A Point Energy and Mass Balance Model of a Snow Cover', NOAA Technical Report NWS 19, 150 pp, February 1976

Buchmann, M., Marty, C., & Begert, M. (2019, January). The value of parallel snow measurement time series. In Geophysical Research Abstracts (Vol. 21).

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting, 32(3), 669-679.

Olefs, M., Schöner, W., Suklitsch, M., Wittmann, C., Niedermoser, B., Neururer, A., & Wurzer, A. (2014). SNOWGRID–A new operational snow cover model in Austria. International Snow Science Workshop Proceedings 2013: 38–45.

Nalder, I. A., & Wein, R. W. (1998). Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agricultural and forest meteorology, 92(4), 211-225.

Resch, G., Chimani, B., Koch, R., Schöner, W. & Marty, C. (2020). Homogenization of long-term snow observations. In Geophysical Research Abstracts.

Tibshirani R (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society B, 58, 267–288.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.