Can artificial neural networks drive a glacier mass balance model?

Michael Binder

June 2019

Abstract

Glacier mass balance models, accurate and robust on a global scale, are of particular importance to climate scientist to analyze interactions between the earth's cryosphere and climate and ultimately predict sea level rise. Rare in-situ measurements on glaciers complicate the development of respective models that need to work globally in a variety of climate settings. Alongside most common temperature and precipitation index models this paper investigates the utilization of artificial neural networks for the prediction of glacier mass balances on different scales. The estimation of annual WGMS data by a neural network trained on monthly averaged temperature and precipitation data provides promising results and significantly outperforms corresponding index models. In contrast, more investigations are needed to verify the robustness of neural networks, when it comes to the approximation of surface height changes of glaciers based on AWS data on an hourly scale.

Introduction

Understanding the effects of climate change on the earth's cryosphere and eventually sea-level rise requires precise measurements or estimates of ice volume or mass changes all over the world. However, meteorological and glaciological measurements in mountainous terrain are scarce and spatial resolution of global climate models is too coarse to resolve specific, local weather conditions (Giesen and Oerlemans, 2012). Therefore, it is even more important to develop robust and accurate glacier mass balance models based on available data, that allow the estimation of glacier development all over the world.
Temperature (and precipitation) index models and geodetic methods currently are the state-of-the-art when it comes to the prediction of glacier mass balances without any in-situ measurements or more detailed meteorologic data. Eventually, artificial neural networks (NN) could provide superior models over T/P linear regression models in terms of stability and accuracy, when trained to a certain level. This paper tries to provide a first investigation on the application of a NN driving a glacier mass balance model.

1.1 Some theory on neural networks

Neural networks as part of the machine learning hype are becoming more and more popular for all kinds of applications, where some input variables could provide enough information to predict a certain event, value or observation and at least a small amount of training or reference data is available. Based on the number of layers within a neural network a ridiculously high number of connections between all inputs is established, which become weighted and further combined until a single or multiple outputs are generated. To train a NN the cost or error at the output layer is calculated with respect to the available reference output. The most common and also important one for the application within this paper is the Mean squared error (MSE). After its calculation at the output layer, the error is fed backwards through the network and a gradient to minimize the error at the output layer is calculated for each weight and bias within the neural network. Based on additional training parameters like learning rate, momentum, regularization and the calculated respective gradient all weights and biases are adapted to make a better prediction on the next run.

A more detailed explanation on possible configurations and the training of neural networks would go beyond the scope of this paper, but is extensively covered in the cited literature (Nielsen, 2015 and Goodfellow, Bengio, and Courville, 2016). Instead, it is probably useful to introduce the term 'overfitting', which is the excessive adaption of a NN to its training data. In the case of overfitting, the NN will perform extremely well on predicting the output for data that was within the training set, but it performs significantly worse with correct input data that was not within the training set. This scenario should always be avoided, because predicting non-training data generally is the ultimate application of a trained neural network.

Glacier mass balance modeled by a neural network

In a first attempt a NN is used to model the annual mass balance of different glaciers based on monthly temperature and precipitation data. After a successful application and a promising comparison to state-of-the-art index models, a NN has been trained to approximate the surface height change of glaciers at a local observation site.

2.1 Mass balance on an annual scale

Since air temperature generally is the most readily available data, so called temperature index (or degree-day models) use the strong relationship between snow/ice melt and air temperature to estimate the mass balance of glaciers during their ablation season (Hock, 2003). For obvious reasons accumulation is not covered within a basic degree-day model, so it makes sense to include precipitation in a model whenever respective data is available. As a result these temperature/precipitation models allow the approximation of a glacier's mass balance for a complete hydrological year on still quite restricted meteorological data. For the following comparison and evaluation of index models and neural networks, three annual glaciological mass balance records of the World Glacier Monitoring Service are put in relation to corresponding monthly temperature and precipitation data. Unlike the black box approach of neural networks, index models rely on simple linear dependencies between the input and output resulting in coefficients that leave room for further interpretation. Equation 2.1 (Marzeion, Jarosch, and Hofer, 2012) describes such a temperature/precipitation model including two factors β_1 and β_2 that establish a linear relation between the yearly sum of snowfall (precipitation at certain temperatures) and the yearly sum of positive atmospheric temperatures respectively. β_0 is the intercept of the corresponding linear regression model. Changing T_0 or T_{snow} can have a significant influence on the parameters of the linear regression and, ultimately, on the performance of the model. Some assessable testing showed that $T_0 = 0\,°\text{C}$ and $T_{snow} = 4\,°\text{C}$ yield one of the best results of the index model and are, therefore, used for comparison. Their physical interpretation and the sensitivity of the fitted parameters especially with respect to varying climate scenarios can be of great scientific interest and is further discussed in referenced literatures, but not covered within this paper (Hock, 2003 and Marzeion, Jarosch, and Hofer, 2012).

$$\dot{m}_j = \beta_0 + \beta_1 \sum_{i=1}^{12} p_{i,j}^{solid} + \beta_2 \sum_{i=1}^{12} T_{i,j}^{+} \tag{2.1}$$

$$p_{i,j}^{solid} \begin{cases} p_{i,j}, & \text{if } T_{i,j} < T_{snow} \\ 0, & \text{if } T_{i,j} \geq T_{snow} \end{cases}$$

$$T_{i,j}^{+} \begin{cases} T_{i,j}, & \text{if } T_{i,j} > T_0 \\ 0, & \text{if } T_{i,j} \leq T_0 \end{cases}$$

The index i in equation 2.1 corresponds to a monthly index from one to twelve, while the index j iterates over years. The longest of the three mass balance records has been obtained at the Claridenfirn in the Glarus Alps, Switzerland and covers 101 years from 1915 to 2015. Data for the Silvretta glacier in Austria starts four years later (97 years total) while 70 measurements are available for the Taku glacier in Alaska starting in 1946. Considering that most common data sets for training NNs hold thousands or even millions of entries, using 100 measurements for the very same task might not seem promising at first, but we are going to see that a careful network configuration and selection of training parameters can lead to more than satisfying results.

For a first and simple comparison the R^2 value (2.2) serves as a good performance measurement. It describes the variability of the depended variable (annual mass balance data) that can be described by the independent variables (temperature and precipitation)(S. Wilks, 2006).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{(y_{data} - y_{pred})^2}{(y_{data} - y_{mean})^2} \tag{2.2}$$

The performance of the T/P index model and a neural network in estimating the annual mass balance of the Claridenfirn is shown in figure 2.1. Though the index model almost predicts 50 % of the data's variability, the two-hidden-layer neural network significantly outperforms the linear regression with an R^2 of 92 %, when trained on 90 % of the available data for the Claridenfirn. When trained on only half of the available data, the NN still performs well and reaches an R^2 of 72 %, which can be an important factor for actual science cases and estimations of glacier mass balances for time frames and locations without actual measurements. Estimations of the NN for other glaciers max out at about 50 %, if trained on only one of the three glaciers, but can reach 60 % or more for two or more glaciers, if data from all glaciers is included within the training set.

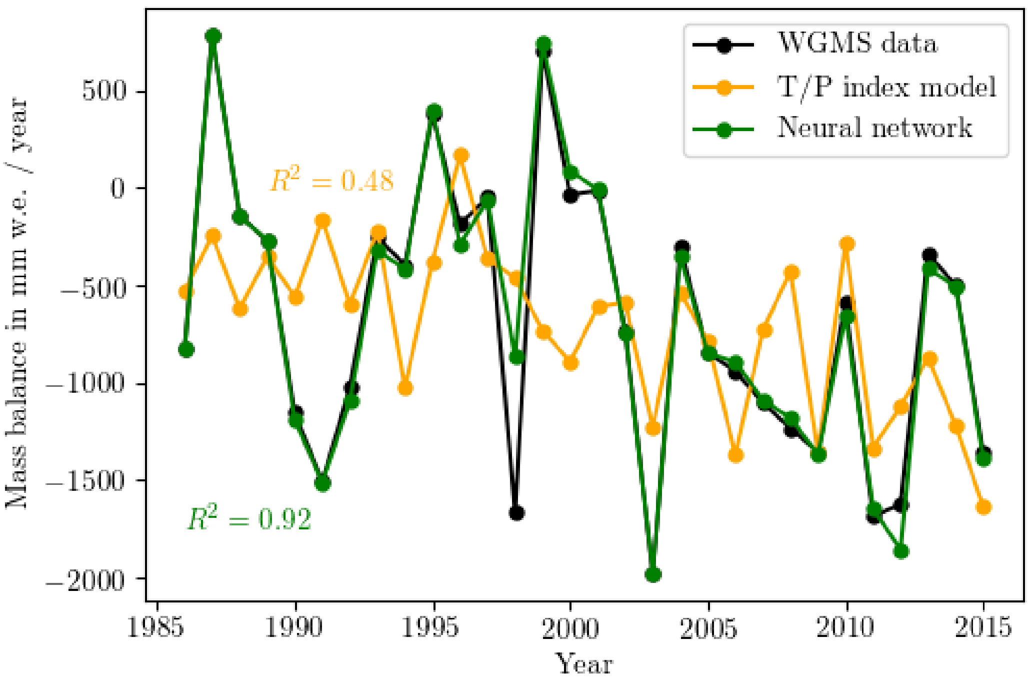

Figure 2.1: WGMS data based on the glaciological method for the Claridenfirn compared to a temperature and precipitation index model and a neural network trained on monthly temperature and precipitation data. A time frame of 30 years out of the 101-year dataset is shown for easier orientation.

The network used for this comparison ended up consisting of two fully-connected hidden layers with 24 knots (neurons) each. More complicated or deeper neural networks haven been tried, but generally performed worse than this quite simple model without any regularization or dropout layers.

2.2 Surface mass balance at an observation site

The previous section showed how effective neural networks can be used to estimate the annual mass balance of a glacier based on monthly averages of temperature and precipitation. Now the question arises if a NN can be of further use when a glacier mass balance is calculated on a more local and detailed scale based on a complete set of automatic weather station (AWS) data. Though other locations with different AWS records have been tested, too, this paper focuses on available AWS data for the Zhadang glacier in Tibet during the 2012 ablation season, because an extensive energy and mass balance model and a simple degree day model have already been available for a visual comparison. The AWS data set provides classic meteorologic parameters like temperature, pressure, relative humidity and others as well as radiation fluxes and a surface height measurement from an ultrasonic distance sensor (SR50) which provides the validation or reference data for the different models.

The data is available on an hourly basis, so every hour the neural network can use the current meteorologic conditions to estimate the surface height change until the next measurement. The ultimate results visualized in figure 2.2 are promising, but a few important observations were made during the building and training of the NN that have to be considered later in this section. The physics based energy and mass balance calculation, which is also plotted in figure 2.2, does provide a decent result, but its solution can actually

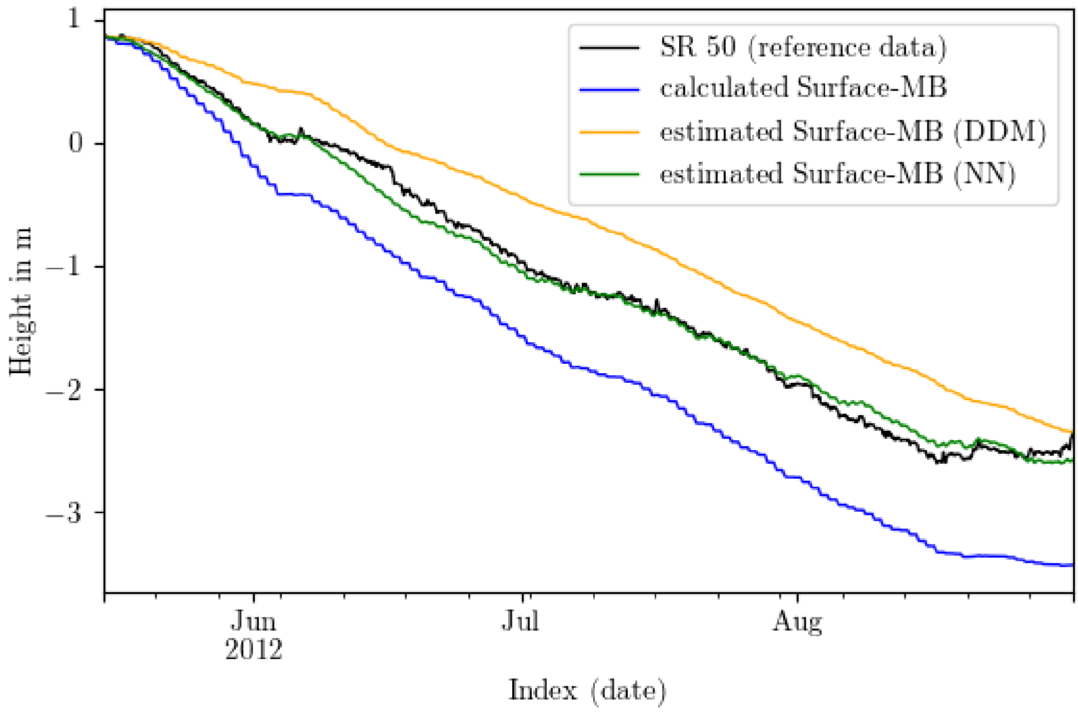

Figure 2.2: Three approaches of estimating the surface height change of the Zhadang glacier in Tibet during its ablation season 2012. SR 50 measurements serve as reference data for a physics based calculation, a Degree-Day-Model and a neural network.

be hacked or significantly improved by applying unrealistic densities for the snow and ice melting processes. For the degree day model a separation of the DDF with respect to snow and ice melting would further improve the predictions, too, but a distinction between snow and ice without the calibrated SR50 sensor would highly increase the models complexity.

Though the cumulative sum of the NN's predictions looks quite promising for the visualized ablation season, especially when compared to the other models, the R^2 for the individual changes in surface height between measurements, which are predicted by the NN, only reaches 16%. By now it is not fully understood how representative either the R^2 or the actual visualization is for the tackled problem, but further investigations of similar scenarios and the observation of other variables may provide further insights. Table 2.1 now provides a short summary on the NN and training parameters used for the estimation of the surface height change.
During the training and building of the NN a significant effect was observed by changing the loss function of the learning process from the generally used mean squared error (MSE) to a mean absolute error. The final cumulative sum of the predictions ended up being much smoother and flatter and thus suppressed most of the accumulation time slots or underestimated melting. Other important steps were the normalization of the input data and scaling of the output of the neural network. Both steps were essential and prevented the neural network from learning adequately beforehand.

Besides sticking to the MSE for the loss function, a classic exponential step decay for the learning rate without applying any momentum turned out to be the best setting for this NN. "Simplifying the network" also worked for the overall setup of the network with only one fully-connected layer with 15 neurons (ReLUs) between the input and output layer

Table 2.1: Parameters and overall settings of a NN calculating surface mass balance.

Parameter	**Value(s)**
Loss function	Mean squared error
eta (learning rate)	0.1 + exponential step decay
Momentum	-
Regularization	-
Batch size	1
Dropout	0.2
Training set	0.5 %
Layer setup [N N ...]	[15] (fully connected)
Neuron type	ReLU

performing significantly better than deeper and larger networks. Somehow unexpected was the observation that online learning (learning on each input individually instead of averaging the gradient over a certain batch size) resulted in the best training of the network. The differences to batch sizes of five or ten was observable, but not too large, so this might change again along with other modifications applied to the neural network in the future.

It was tried to avoid overfitting by applying a dropout on the input data of 0.2 and only using 50 % of the visualized data in figure 2.2 for training. Nevertheless, the application of a trained NN to other glaciers was not successful yet.

Conclusion

Estimating glacier mass balance by neural networks has the potential of highly accurate predictions at the cost of understanding relations between parameters or sensitivities of index model factors due to different climates at first. This paper shows that more research on the application of neural networks could ultimately lead to a fast, easy and reliable setup of a mass balance model for a specific input scenario.
As expected, annual mass balance cycles were easier to predict than short-term surface height changes. One-hour update rates of the AWS data may imply correlations between sequential data points which may ultimately be too high and "confuse" the NN. A recurrent neural network might provide better and more reliable results, because it also considers past data to predict a new surface height change. However, these networks are more complicated to build, so a first solution could be a downsampling of the data to a daily mean to reduce or completely avoid correlations between individual data points.
Looking more into the future, multiple investigations could be conducted on the basis of a reliable NN to evaluate atmospheric parameters on their contribution and importance for different variations and scales of glacier mass balance models.

Bibliography

Giesen, R. H. and J. Oerlemans. "Calibration of a surface mass balance model for global-scale applications". In: *The Cryosphere* 6.6 (2012), pp. 1463–1481. DOI: `10.5194/tc-6-1463-2012`. URL: `https://www.the-cryosphere.net/6/1463/2012/`.

Nielsen, Michael A. *Neural Networks and Deep Learning*. 2015. URL: `http://neuralnetworksanddeeplearnin` `com/faq.html`.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

Hock, Regine. "Temperature index melt modelling in mountain areas". In: 2003.

Marzeion, B., A. H. Jarosch, and M. Hofer. "Past and future sea-level change from the surface mass balance of glaciers". In: *The Cryosphere* 6.6 (2012), pp. 1295–1322. DOI: `10.5194/tc-6-1295-2012`. URL: `https://www.the-cryosphere.net/6/1295/2012/`.

S. Wilks, D. "Statistical Methods In The Atmospheric Sciences". In: *International Geophysics Series* 59 (Jan. 2006). DOI: `10.1016/S0074-6142(06)80036-7`.