

# The application of a non-linear back-propagation neural network to study the mass balance of Grosse Aletschgletscher, Switzerland

Daniel STEINER,<sup>1</sup> A. WALTER,<sup>2</sup> H.J. ZUMBÜHL<sup>1</sup>

<sup>1</sup>*Institute of Geography, University of Bern, CH-3012 Bern, Switzerland  
E-mail: steiner@giub.unibe.ch*

<sup>2</sup>*Deutscher Wetterdienst, Postfach 100465, D-63004 Offenbach, Germany*

**ABSTRACT.** Glacier mass changes are considered to represent key variables related to climate variability. We have reconstructed a proxy for annual mass-balance changes in Grosse Aletschgletscher, Swiss Alps, back to AD 1500 using a non-linear back-propagation neural network (BPN). The model skill of the BPN performs better than reconstructions using conventional stepwise multiple linear regression. The BPN, driven by monthly instrumental series of local temperature and precipitation, provides a proxy for 20th-century mass balance. The long-term mass-balance reconstruction back to 1500 is based on a multi-proxy approach of seasonally resolved temperature and precipitation reconstructions (mean over a specific area) as input variables. The relation between the driving factors (temperature, precipitation) used and the reconstructed mass-balance series is discussed. Mass changes in Grosse Aletschgletscher are shown to be mainly influenced by summer (June–August) temperatures, but winter (December–February) precipitation also seems to contribute. Furthermore, we found a significant non-linear part within the climate–mass-balance relation of Grosse Aletschgletscher.

## 1. INTRODUCTION

Modern climatology faces the question of whether anthropogenically induced climate change is already observable in climatic variables. Because the climate system can be regarded as non-linear (Houghton and others, 2001), traditional linear statistical models cannot describe the full complexity of its behaviour, and thus fail to answer this question. Non-linear neural network models (NNMs) provide a statistical solution to this problem.

NNMs originated in the study of the cognitive abilities of biological brain functions, i.e. how the brain processes information and how it learns (Adrian, 1926; Rosenblatt, 1958; Grossberg, 1982). A comprehensive overview of the wide range of applications of NNMs to time-series analysis is summarized in Anderson and Rosenfeld (1986) and Rumelhart and McClelland (1986).

As an analogue to biological brains, a typical NNM consists of a yet to be defined number of simple processing units. The task of a (supervised learning) NNM is to 'learn' certain features of the data, which consist of input variables and desired output responses, using a training subset. After the training process, these features are interpolated on an unknown validation subset, on which the performance of the NNM can be determined. Internal parameters of the network architecture are adjusted according to a specific learning rule so that the network ideally captures all intrinsic data features.

What distinguishes NNM from traditional statistical methods is its non-linear character, due to a non-linear mapping function. This mapping function is commonly chosen out of the class of sigmoid functions, because it has been shown experimentally by Adrian (1926) that biological neurons respond to a stimulus in a sigmoidal fashion, i.e. no output until a certain threshold is exceeded, followed by a nearly linear input–output relation and saturation from a certain input level onward.

Conventional studies of climate change are based either on physical climate models (e.g. general circulation models (GCMs) with a high computing and CPU time demand), or on statistical models such as multiple linear regression (MLR) models, which allow only a limited exploration of the state of the climate system. There is a wide variety of models for studying the behaviour of the climate system that can be ordered by increasing complexity (e.g. radiative–convective models, energy-balance models). The MLR and GCM approaches represent the extremes of this model hierarchy. Nevertheless, NNM represents a quite simple non-linear method for gaining insight into the behaviour of the climate system.

The non-linear NNM combines the advantages of physically motivated (possibly non-linear) climate models, with high complexity, and linear statistical models, with a low CPU time demand. Two recent climatological applications of NNM have been the detection and attribution of anthropogenic climate change (Walter and Schönwiese, 2002, 2003), and forecasting tropical Pacific sea surface temperatures in the El Niño region (Wu and Hsieh, 2003). For a general overview of climatological applications of neural networks, see Hsieh and Tang (1998) or Walter (2001).

In this paper, we simulate and reconstruct a proxy for annual mass balance of Grosse Aletschgletscher, Switzerland, using a back-propagation network (Rumelhart and others, 1986). This is the first time that a neural network approach has been used in this glaciological context to simulate and reconstruct glacier mass balance.

The mass balance of glaciers varies with changing climate, and many studies and measurements have been carried out to investigate this connection (e.g. Oerlemans and Reichert, 2000, and references therein). To relate annual glacier mass balance to meteorological data, a suitable combination of climate data is required. In many studies, (summer) temperature and (winter) precipitation are assumed to be the best predictors for this objective (Oerlemans and Reichert, 2000). Because mass balance is a complex

**Table 1.** Principal meteorological stations used in this study (data from the online database of MeteoSwiss)

Climate parameter	Usage in study	Station name	Altitude m.a.s.l.	Time period	Geographic coordinates
Temperature	Forcing	Sion	483	1864–2003	46°13' N, 7°20' E
Temperature	Forcing	Meiringen	595	1890–2003	46°44' N, 8°11' E
Precipitation	Forcing	Fiesch	1060	1899–1991	46°24' N, 8°8' E
Precipitation	Reduction	Ernen	1000	1967–99	46°24' N, 8°8' E
Precipitation	Reduction	Fieschertal	1095	1999–2003	46°25' N, 8°9' E
Precipitation	Reference	Brig	666	1899–1906/1959–2003	46°19' N, 7°58' E
Precipitation	Forcing	Kippel	1376	1899–1974	46°24' N, 7°46' E
Precipitation	Reduction	Ried	1500	1974–2003	46°25' N, 7°48' E
Precipitation	Reference	Leukerbad	1285	1898–2003	46°23' N, 7°37' E
Precipitation	Forcing	Lauterbrunnen	818	1899–2003	46°36' N, 7°54' E

function dependent on climate, time and other factors, it may be well suited to non-linear approaches. In fact, non-linear functions in the glacier–climate system have been found in several recent studies (e.g. Braithwaite and Raper, 2002; Lie and others, 2003). The aim of this study is to use an easy non-linear model, driven only by a combination of temperature and precipitation data.

In section 2 we provide an overview of the data used in this study and the concept behind the BPN. In section 3 we discuss how the BPN can be used to simulate a proxy for annual glacier mass balance by suitable selection of monthly instrumental data (temperature, precipitation) for the 20th century. As well as simulating glacier mass balance, this work aims to reconstruct a longer-term mass-balance series. Applying the BPN technique, we propose to use multi-proxy seasonal reconstructions of temperature (Luterbacher and others, 2004) and precipitation (Pauling and others, in press) to calculate a new annual mass-balance series back to AD 1500. In section 4 the performance of the BPN is discussed, questions are presented and further investigations are proposed.

## 2. DATA AND METHODS

### 2.1. Instrumental data of the 20th century

Figure 1 shows the topography and the locations of the meteorological stations within the greater region of Aletschgletscher, the largest stream of ice in the Alps, lying in the Bernese Alps of south-central Switzerland. Aletschgletscher is divided into main (Grosse Aletschgletscher), middle and upper parts. The main glacier (46°30' N, 8°2' E) is 24.7 km long and covers an area of 86.76 km<sup>2</sup>. Its head is at 4160 m a.s.l., the glacier terminates at 1556 m a.s.l. and its average height is about 3140 m (Haeberli and others, 1998). The exposure of Grosse Aletschgletscher is to the southeast (accumulation area) and to the south (ablation area). As shown in Figure 1a, the ice stream consists of three large firn fields: the Aletschfirn, the Jungfraufirn and the Ewigschneefeldfirn. In 1994, depth soundings made by means of radio echolot (radar) located the glacier bed at depths of 600–700 m in the Konkordia region (VAW/SANW, 1881–2002, Nos. 115/116 (1999)).

The study site which borders on the upper Valais (Rhône valley) is characterized by a generally temperate humid climate. It is more subcontinental than that of the northern

Alpine region, which is more suboceanic (Van der Knaap and Van Leeuwen, 2003). Frei and Schär (1998) point out that the dry inner Alpine conditions of the Valais contrast with the wet anomalies along the Alpine rims, especially in the north of the Aletsch region. Because the Valais has a complex topography, like the whole Alpine region, marked vertical temperature gradients are found over short distances. Furthermore, the major west–east orientation of the Valais, occasionally across the main airflow direction, can cause mesoscale meteorological phenomena like the föhn (Schüepf and others, 1978).

The fluctuations in length of Grosse Aletschgletscher have been carefully reconstructed on the basis of documents, archaeological field studies and precise dendrochronological dating. In the last 500 years, a very small advance around 1500 is dated by both dendrochronology and archaeology. Dendrochronological analysis allows an exact reconstruction of the marked advance, leading to a maximum glacier extent from 1670 to 1680. Pictures and texts document fluctuations during the 18th and 19th centuries, especially the final advance to the last Little Ice Age maximum extent around 1856. From 1892 to the present, the continuously decreasing tongue has been precisely measured every year (VAW/SANW, 1881–2002; Holzhauser and Zumbühl, 1999; Haeberli and Holzhauser, 2003).

As a continuation of the mass-balance studies of the 1950s (Kasser, 1954, 1967), Aellen and Funk (1990) estimated an average yearly net balance for the period 1931–87 from a simple hydrological model, calculating daily to annual water-storage variations in the basin of the Massa river. Based on water-flux measurements of the Massa river, a tributary of the Rhône river which issues from Grosse Aletschgletscher, and precipitation data from the surrounding area, a proxy for annual mass balances can be calculated (Paterson, 1994). Later, in order to correct systematic errors in these calculations (under- or overestimations of areal precipitation), the data were revised by a linear regression approach. Besides the uncorrected data series from Aellen and Funk (1990), information from direct glaciological mass-balance measurements on Grosse Aletschgletscher and a reference mass-balance series (average of Grosse Aletschgletscher (uncorrected), Limmeren and Plattalvagletscher) were incorporated into the linear model to determine a corrected proxy for annual mass balances (Müller-Lemans and others, 1995).

As a control of the Aletsch data, variations of total volume, surface area and mean thickness were determined for the years 1926/27, 1957, 1980 and 1999, using geodetic methods (terrestrial or aerial surveys). There is clear agreement between the Aletsch cumulative mass balance and these geodetic measurements (Vincent and others, 2004). An annual check is also provided by the average net balances computed from a stake network observed on Grosse Aletschgletscher from 1950 to 1986 (Aellen, 1995). Finally, new data made it possible to expand the mass-balance data to the 1919–99 period with the model approach presented here.

It should be noted that this proxy for annual balances of Grosse Aletschgletscher comes from a hydrological model and not directly from field measurements. So the resulting series is probably not as reliable as direct measurements, even though the data have been checked against independent geodetic measurements (Vincent and others, 2004). Nevertheless, this proxy will serve as target function in our neural network modelling approach.

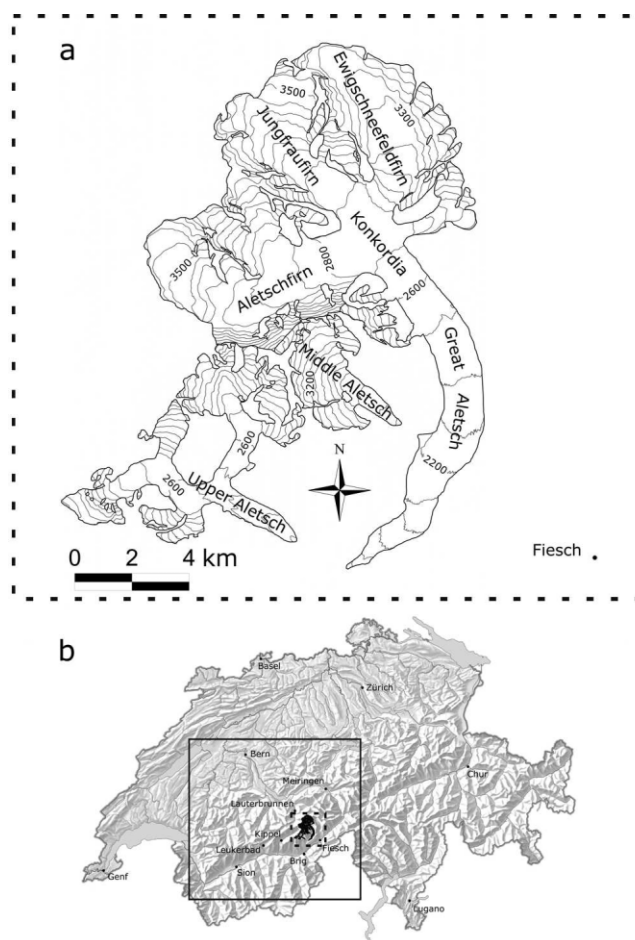
The forcing data, monthly temperature and precipitation series from the surroundings of Grosse Aletschgletscher were provided by MeteoSwiss (online database of MeteoSwiss). The data are homogenized, i.e. they are adjusted for non-climatic factors like changes in station location or changes in observation practices (Begert and others, 2003). Table 1 represents all climate series used in this study. In agreement with Schüepf and others (1978), we found a lack of long climatic time series in this region of the Swiss Alps.

As the Alpine precipitation field shows large spatio-temporal variability, there is a strong need for long-term precipitation data close to a glacier in order to relate climatic conditions to glacier activities. The precipitation series (Fiesch, Kippel) were selected according to the study by Aellen and Funk (1990). But instead of the precipitation series for Grindelwald applied there, which is missing values for the period 1903–10, we chose the highly correlated ( $r = 0.92$ ) nearby precipitation series for Lauterbrunnen. Furthermore, the Sion and Meiringen data were used as temperature datasets. With this data selection, we climatically represent both the inner Alpine dry valley of the upper Valais and the wetter northern part of the Bernese Alps. However, these data series serve as potential model inputs for the BPN.

To complete the precipitation series of Fiesch and Kippel for the missing periods 1992–2003 and 1974–2003, a method derived from the homogenization process at MeteoSwiss was used (Begert and others, 2003). Based on a highly correlated reference series, nearby precipitation series were reduced to the level of the series with missing values. For this, average ratios between all precipitation series and the reference series were calculated for the overlapping periods. Based on the proportion of these average ratios, it is possible to determine reduction factors for the nearby precipitation stations (personal communication from T. Schlegel, 2004).

The Fiesch precipitation data series ( $P_{\text{Fiesch}}$ ) was completed as follows:

1. The highly correlated ( $r = 0.92$ ) precipitation series from Brig was used as reference series.
2. The two nearby precipitation stations, Ernen ( $P_{\text{Ernen}}$ ; 1.3 km from Fiesch) and Fieschertal ( $P_{\text{Fieschertal}}$ ; 2.2 km from Fiesch), were then reduced to the level of Fiesch.



**Fig. 1.** (a) Map showing Aletschgletscher with its main branches and tributaries. (b) The Aletsch region (area with dashed outline) and the meteorological stations used as forcing or reference data in this study. Also shown is the grid range (area with solid outline) over which seasonal averages were calculated.

3. For the missing time period 1992–98,

$$(P_{\text{Fiesch}}) = q_{\text{Fiesch}} \frac{(P_{\text{Ernen}})}{q_{\text{Ernen}}} = 1.043(P_{\text{Ernen}}). \quad (1)$$

4. For the missing time period 1999–2003,

$$(P_{\text{Fiesch}}) = q_{\text{Fiesch}} \frac{(P_{\text{Fieschertal}})}{q_{\text{Fieschertal}}} = 0.764(P_{\text{Fieschertal}}), \quad (2)$$

where  $q_{\text{Fiesch}}$ ,  $q_{\text{Ernen}}$  and  $q_{\text{Fieschertal}}$  are the average ratios between the Fiesch, Ernen and Fieschertal precipitation series and the Brig reference series for the overlapping measured period.

In an analogous procedure, the Kippel data series ( $P_{\text{Kippel}}$ ) was completed using the Leukerbad reference series ( $r = 0.90$ ) and the nearby Ried precipitation series ( $P_{\text{Ried}}$ ; 3.1 km from Kippel) as a reduction series.

For the missing time period 1974–2003,

$$(P_{\text{Kippel}}) = q_{\text{Kippel}} \frac{(P_{\text{Ried}})}{q_{\text{Ried}}} = 1.044(P_{\text{Ried}}), \quad (3)$$

where  $q_{\text{Kippel}}$  and  $q_{\text{Ried}}$  are the average ratios between the Kippel and Ried precipitation series and the Leukerbad reference series for the overlapping measured period.

## 2.2. Multi-proxy reconstructions of temperature and precipitation back to 1500

In the absence of widespread instrumental data, we must rely upon indirect lines of evidence to provide information about climate variability over the past several centuries. To determine annual or higher-resolution climate variations, however, high-resolution 'proxy' climate indicators such as ice cores and tree-ring measurements, combined with the scant available documentary or instrumental evidence available in prior centuries, are required (e.g. Mann, 2002; Jones and Mann, 2004; Luterbacher and others, 2004).

In this study, two new gridded ( $0.5^\circ \times 0.5^\circ$  resolution) multi-proxy reconstructions of seasonal temperature (Luterbacher and others, 2004) and precipitation fields (Pauling and others, in press) from 1500 to 2000 for European land areas were used. The temperature reconstruction is based on a dataset that includes instrumental data series, reconstructed sea-ice and temperature indices, and proxy temperature reconstructions from ice cores and tree rings (see Luterbacher and others, 2004, for a detailed description of the method and the data used). Like the temperature reconstructions, the precipitation estimates were based on a combination of instrumental series, documentary indices (precipitation) and natural proxies (tree rings, corals, ice cores). As these reconstructions share no common predictors, they can be compared in terms of climate dynamics or temporal stability of the precipitation–temperature relationship. The temperature reconstructions end in 2004, the precipitation reconstructions in 2000. As a consequence, we studied the period mentioned above (1500–2000).

As input data for the model, a seasonal average of both temperature and precipitation sums was calculated over the area  $46\text{--}47^\circ\text{N}$ ,  $7\text{--}8.5^\circ\text{E}$ , comprising  $xy$  gridpoints (see Fig. 1b for the grid range). This grid extent covers the entire Aletsch region and clearly reflects its climatic conditions, presented in section 2.1. The seasonal averages generally show higher correlations with the observed precipitation (Fiesch, Kippel, Lauterbrunnen) and temperature measurements (Sion, Meiringen) for the overlapping 1900–2000 period than nearby single gridpoints. The correlation coefficients range from 0.87 (June–August; Meiringen) to 0.95 (March–May; Sion) for temperature, and from 0.61 (June–August; Fiesch) to 0.77 (December–February; Kippel) for precipitation. The generally lower correlation coefficients for precipitation probably reflect the greater difficulty in reconstructing and/or representing precipitation fields in mountainous regions due to their large spatio-temporal variability. As with all models, the output of a BPN depends on the quality of the input data, and possible uncertainties in the precipitation data will significantly reduce the accuracy of the trained neural networks. Related to this issue, Klein and Rossin (1999) found that errors in training data affect neural network models more severely than linear regression models.

## 2.3. Back-propagation neural networks as modelling tools

Neural networks use non-linear functions and a large number of processing units to reduce the risk of model mismatch errors. Instead of matching the architecture of the model to a problem, a model is used that can describe almost anything, and careful training of the model is used to constrain it to describe the data.

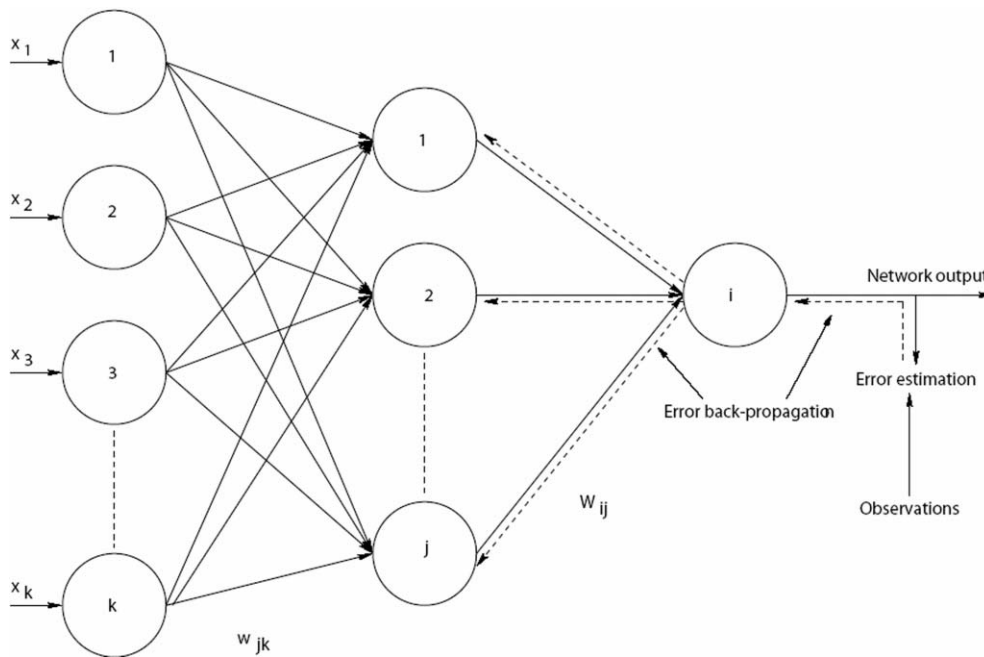
In this study, the standard NNM, the Back-Propagation Network (BPN), was applied (Rumelhart and others, 1986). This network architecture is based on a supervised learning algorithm to find the minimum of a cost function. Unlike a simulated annealing schedule (Metropolis and others, 1953), the BPN does not guarantee that the global minimum of this cost function will be reached, though it is very likely that a minimum good enough to reproduce responses in the data can be found. This approach also bears a certain risk of overfitting, so the data have to be separated into a training and a validation subset (calibration/verification in reconstruction exercises). The 'learning' process of the network is performed on the training subset only, whereas the validation subset serves as an independent reference for the simulation quality. This technique is called cross-validation (Stone, 1974; Michaelson, 1987). When applying NNM to a non-stationary time series, as in this approach, the choice of the training subset must ensure that the entire amplitude range of all forcing mechanisms considered is covered; otherwise the algorithm will fail if confronted with an extreme value during the validation process, and such an extreme value never occurred in the learning period. The training must be carried out using a representative data subset. Therefore the training data cannot be chosen continuously out of all data available but must be chosen so that all the amplitudes are covered. We used 75% of all data for training, and the remaining 25% for validation (Walter and Schönwiese, 2003).

A typical NNM consists of three layers: input, processing and output layers. Figure 2 shows a simplified neural network. Although the units of the processing layer are sometimes called 'hidden units', their weights, activation and all other internal parameters are easily accessible from the outside.

The input to an NNM is a vector of elements ( $x_k$ ), where the index  $k$  stands for the number of input units in the network. In our study,  $k = 12$  indicates the major short-term climatic influences on the mass balance of Aletschgletscher in the 20th century. For the long-term reconstructions of annual glacier mass balance since 1500, we chose  $k = 2$  (see section 3 for the selection of the input variables). These inputs are weighted with weights  $w_{jk}$ , where  $j$  represents the number of processing units, to give the inputs to the processing units

$$h_j = \sum_k w_{jk} x_k. \quad (4)$$

Using too few/many processing units can lead to underfitting/overfitting problems. That is, the simulation results are highly sensitive to the number of processing units and learning parameters chosen. The best number of processing units can be determined by repeated simulations with increasing numbers of processing units, i.e.  $j = 1, 2, \dots, n$ , and by observing the performance of these different network architectures on the validation sample. Multiple versions of the BPN must be tried to obtain robust results. Typically, the validation error decreases with increasing numbers of processing units until the network has too many degrees of freedom, i.e. too many processing units. After that, the validation error increases and the model with the optimal number of processing units, i.e. the model configuration with the lowest validation error, must be chosen for the final simulations.



**Fig. 2.** An example of a simplified three-layer  $k$ - $j$ -1 BPN architecture. Also shown is the concept behind the back-propagation training algorithm (arrows).

For the 20th century we used 12 input units as forcings. In addition, there are 6 processing units in one processing layer and 1 output unit as a target function. This neural network architecture is abbreviated as 12–6–1. For long-term simulations back to 1500, a 2–2–1 architecture was applied.

After weighting and summation of the inputs, the non-linear aspects of the BPN were encountered first, as the results were passed to non-linear activation functions in each processing unit. These functions produced the output of the processing layer

$$X_j = g(h_j) = g\left(\sum_k w_{jk}x_k\right). \quad (5)$$

To retain the original analogy between NNM and learning mechanisms in the brain (Adrian, 1926; Amit, 1989), the activation functions are commonly chosen from the class of sigmoid functions, which also will keep the model response bounded. We have chosen

$$g(h) = \tanh(h) = \frac{e^h - e^{-h}}{e^h + e^{-h}}. \quad (6)$$

The outputs of the processing units are fed to the output layer where they are again weighted

$$H_i = \sum_j W_{ij}X_j = \sum_j W_{ij}g\left(\sum_k w_{jk}x_k\right). \quad (7)$$

The use of a second activation function will finally produce the output of the network

$$Y = f(H_i) = f\left[\sum_j W_{ij}g\left(\sum_k w_{jk}x_k\right)\right], \quad (8)$$

where  $f$  is the output activation function,  $x_k$  are the chosen input variables,  $w_{jk}$  are the weights from the  $k$  input units to the  $j$  processing units, and  $W_{ij}$  are the weights from the processing units to the output unit. As an output activation

function we apply the identity function  $f(x) = x$ , so that Equation (7) is the complete BPN function.

This model is very general. However, it has been shown that ‘with one layer and an arbitrary continuous sigmoidal function, this model can approximate any continuous function, provided that no constraints are placed on the number of units or the size of the weights’ (Cybenko, 1989).

#### 2.4. The back-propagation architecture

The purpose of training an NNM is to find a set of coefficients that reduces the error between the model outputs and the given test data  $y(x_k)$ . This is usually done by adjusting the weights  $W_{ij}$  and  $w_{jk}$  to minimize the least-squares error

$$\begin{aligned} \chi^2 &= \frac{1}{2} \sum_n \sum_i [y(x_k) - Y(x_k)]^2 \\ &= \frac{1}{2} \sum_n \sum_i \left\{ y(x_k) - g\left[\sum_j W_{ij}g\left(\sum_k w_{jk}x_k\right)\right] \right\}^2, \quad (9) \end{aligned}$$

where  $n$  is the length of the time series. One way to adjust these weights, and thus to reduce  $\chi^2$ , is gradient descent. The update step in the output weights can be found by differentiating

$$\begin{aligned} \Delta W_{ij} &= -\gamma \frac{\partial \chi^2}{\partial W_{ij}} \\ &= \gamma \sum_n [y(x_k) - Y(x_k)] g'(H_i) X_j \\ &= \gamma \sum_n \Delta_i X_j, \quad (10) \end{aligned}$$

defining

$$\Delta_i = [y(x_k) - Y(x_k)] g'(H_i). \quad (11)$$

The so-called learning constant  $\gamma$ , given in Equation (10), is a scale factor which controls how big the update step is, and  $g'$  is the derivation of the sigmoidal activation function used,

here  $\tanh(x)$ . The choice of  $\gamma$  is crucial because  $\gamma$  determines the step width of the gradient descent algorithm. If  $\gamma$  is too small, the algorithm might get stuck in steep canyons of the  $\chi^2$  hypersurface. If it is too large, the algorithm might just jump over minima.

The update in the input weights can be found using the chain rule:

$$\begin{aligned}\Delta w_{jk} &= -\gamma \frac{\partial \chi^2}{\partial w_{jk}} \\ &= \gamma \sum_n \sum_i [y(x_k) - Y(x_k)] g'(H_i) W_{ij} g'(h_j) x_k \\ &= \gamma \sum_n \sum_i \Delta_i W_{ij} g'(h_j) x_k \\ &\equiv \gamma \sum_n \delta_j x_k,\end{aligned}\quad (12)$$

defining

$$\delta_j = g'(h_j) \sum_i W_{ij} \Delta_i. \quad (13)$$

The deltas for the input layer are found in terms of the deltas for the output layer by running them backwards through the network's  $W_{ij}$ 's. Thus, the errors between model output and observations are propagated backwards through the weights of the output and the processing layer, and these errors are used to adjust the weights throughout the network. Training a network by gradient descent and feeding the errors backwards through the network is called error back-propagation.

As mentioned above, this network architecture risks being stuck in local minima on the  $\chi^2$  hypersurface. To reduce this risk, an enhanced version of standard back-propagation with a momentum term  $\alpha$  was used. The momentum term introduces the old weight change as a parameter for computing the new weight change. This avoids the oscillation problems common with the standard back-propagation algorithm when the error hypersurface has a very narrow minimum area. The new weight change is computed by

$$\Delta w(t) = -\gamma \frac{\partial \chi^2}{\partial w} + \alpha \Delta w(t-1). \quad (14)$$

The term  $\alpha$ , a number between 0 and 1, is the so-called momentum parameter, and  $\Delta w(t-1)$  the weight change in the time-step  $t-1$ . The effect of these enhancements is that flat spots of the error hypersurface are traversed relatively rapidly with a few large steps, while the step size is decreased as the hypersurface gets rougher. This adaption of the step size increases learning speed significantly.

One obvious problem concerning BPN is that Equation (14) contains two parameters,  $\gamma$  and  $\alpha$ , whose optimum values will typically vary from one iteration to the next. We might therefore seek some procedure for setting these automatically as part of the training algorithm. One approach for doing this is the Polak–Ribière variant of the conjugate gradient descent. Conjugate gradient is a method of accelerating gradient descent in which the learning rate  $\gamma$  and the momentum parameter  $\alpha$  are determined in each iteration. In ordinary gradient descent, one uses the gradient to find the steepest downhill direction, then moves along that line to the minimum in that direction. With conjugate gradient, a search is made along the conjugate gradient direction to determine the step size, which minimizes the error function along that line. The momentum term  $\alpha$  is

calculated by the Polak–Ribière formula, since this controls the search direction, while the learning rate  $\gamma$  is determined by the golden section method for line minimization (Press and others, 1992; Bishop, 1995). For an excellent overview of the conjugate gradient descent method, see the unpublished draft by J.R. Shewchuk (<http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.ps>). It should be noted that for many applications the Polak–Ribière method gives slightly better results than other conjugate gradient methods or even the conventional gradient descent approach (Press and others, 1992; Bishop, 1995; Dao and Vemuri, 2002).

A second uncertainty in the BPN simulation is related to the fact that the minimum, found from the algorithm, is dependent on the initial starting point on the  $\chi^2$  hypersurface. Since the BPN represents a non-linear mapping function, slightly different initial conditions can lead to large differences in the simulation results. To eliminate this kind of uncertainty, the BPN was driven 30 times, each time only varying the initial starting point for the conjugate gradient algorithm on the  $\chi^2$  hypersurface.

### 3. RESULTS

#### 3.1. Mass balance in the 20th century

One of the most important factors in determining the success of a practical application of neural networks is the form of pre-processing applied to the data. In the simplest case, pre-processing may take the form of a linear transformation of the input and output data (see the z-transformation below). More complex pre-processing may also involve reduction of the dimensionality of the input data. The fact that such dimensionality reduction can improve performance may at first appear somewhat paradoxical, since it will decrease the information content of the input data in most cases (Bishop, 1995). Some of these techniques, such as genetic algorithms, are highly heuristic, while other methods such as principal components analysis make assumptions as to the linearity of relationships between the input data items. Finally, an exhaustive search of optimal input configurations is highly effective but computationally intensive. With the 60 input variables used in this study (five datasets with monthly resolution, each) there are  $2^{60} \approx 10^{18}$  possible subsets to consider.

In the present study, we therefore used the stepwise multiple linear regression approach to extract the potential influences on the mass balance of Grosse Aletschgletscher. Although there is an assumption of linearity behind this method, Addison and others (2004) show that the use of stepwise multiple linear regression could perform comparably to, if not better than, most other reduction methods. Stepwise multiple linear regression examines variables incorporated in the model at every stage of the regression. A variable that may have been the best choice to enter the model at an early stage may later be non-significant because of the relationships between it and other variables now in the regression. Once a variable is proven to be non-significant, it is removed from the model. This process continues until no more variables can be accepted and no more can be rejected (Von Storch and Zwiers, 1999).

We performed a stepwise multiple linear regression with the monthly averages of temperature and precipitation as independent variables and the proxy of annual mass balance

of Grosse Aletschgletscher as an output dataset (dependent). As model inputs for the BPN, we found the following subset of 12 independent variables (>99% confidence level): Meiringen (August, September), Sion (June, July), Fiesch (January, May, July, November) and Kippel (February, July, October, December).

Changes in glacier mass balance can be viewed as a direct, undelayed reaction of a glacier to climatic variations (Reichert and others, 2001; Nesje and Dahl, 2003). Furthermore, glacier mass is determined by taking into account the accumulation of snow, mostly in winter, and warm-season ablation, and thus shows an immediate response to climate variability and site meteorological conditions. Therefore, it is not surprising that 'summer' (June–September) temperature and 'winter' (October–February) precipitation turn out to be highly correlated with the annual glacier mass balance. The high correlations of May and July precipitation could indicate that the mass balance also responds to effects of spring and summer precipitation. Furthermore, it seems that there is no (linear) relationship between the mass-balance series of Grosse Aletschgletscher and the Lauterbrunnen precipitation series.

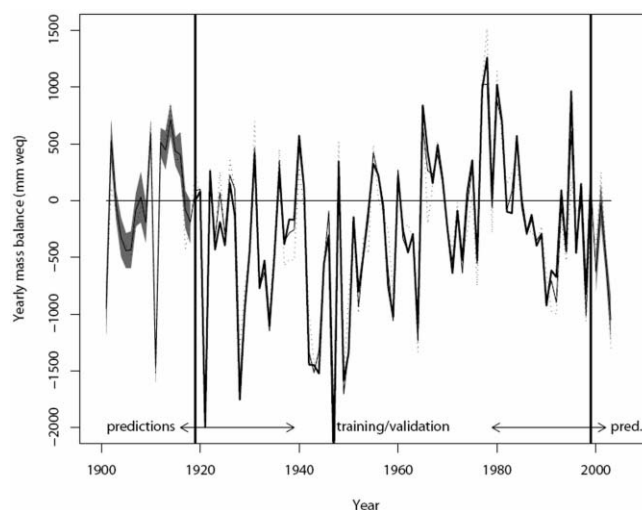
In contrast to the proxy-based reconstruction back to 1500 (see section 3.2), we use monthly data for a reconstruction of the 20th-century mass balance. In this way a choice of potential climatic driving factors is more differentiated and probably leads to better model performance.

Before feeding the BPN, the data are standardized to 1919–99 mean and standard deviation (*z*-transformation) so that temperature and precipitation are in comparable units (for robust NNM performance). After fitting the BPN with the conjugate gradient descent method, the full model was used to reconstruct a proxy of annual mass balance from the forcing data when no output data were available. In a final step, the data were rescaled by the inverse *z*-transformation.

Figure 3 is the composite made up of the annual glacier mass balance for the 1919–99 training/validation period, presented in section 2.1, and BPN predictions for those years when estimations are missing (1900–18, 2000–03). An error envelope (95% confidence interval) based on the rms errors in predictions appears around each section of BPN prediction (Mann and Jones, 2003; Reusch and Alley, 2004). The simulation and the reconstruction are the average outputs of 30 model runs to reduce the effect of falling into local minima (see section 2.4).

The simulation quality over the 1919–99 training/validation period amounts to 97.3% of explained variance, while the stepwise multiple linear regression model driven by the same input data ends with an explained variance of 86.3%. Chen and Funk (1990) applied a similar multiple linear regression approach to reconstruct the mass balance of Rhonegletscher. They used mean annual precipitation and mean summer temperature of nearby stations as model inputs. So our stepwise multiple linear regression model represents a refinement of their approach with higher-resolution data.

To ensure that the neural network has captured most relevant mechanisms, the residuals of the simulations must be tested. This is usually done by testing for Gaussian distribution of the residuals. In the case of a non-linear neural network, the residuals undergo non-linear transformations during the training of the network. Therefore the residuals do not have to be Gaussian-distributed even if the model captures all relevant mechanisms and the residuals



**Fig. 3.** Results of the BPN simulation and reconstruction (solid curve). Also shown is the proxy of annual glacier mass balance (thick curve) for the 1919–99 period and the output of the stepwise multiple linear regression (dotted curve). Confidence intervals derived from rms errors appear as grey envelopes around predictions.

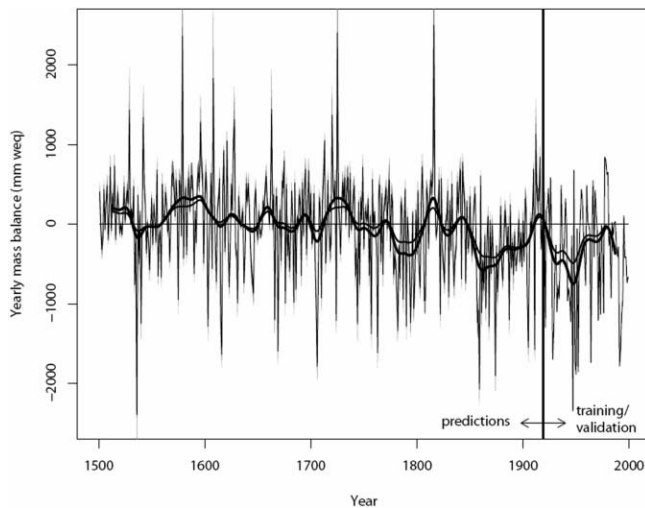
are in fact (white) noise. An alternative statistical method for testing residuals of this kind is the autocorrelation function. If the residuals are indeed white noise, i.e. uncorrelated at all lags, the autocorrelation function should have no structure (Von Storch and Zwiers, 1999; Walter and Schönwiese, 2002).

The autocorrelation coefficient of our simulations shows no significant deflection from zero (95% confidence level); therefore the BPN captured all relevant mechanisms and the remaining residuals can be treated as white noise.

The BPN used here can simulate a proxy of annual glacier mass balance for the 20th century. While the stepwise multiple linear regression approach often under- or overestimates the extremes of the mass-balance function, the trained BPN is better able to learn the intrinsic data features. A small number of NN misfits were found. Furthermore, it is possible to reconstruct missing mass-balance data. For the 2000–03 period we estimated a cumulative mass balance of about  $-2.2$  m (Fig. 3), consistent with several studies and measurements that show persistent melting of Alpine glaciers during the last two decades (BUWAL, BWG, MeteoSchweiz, 2004; Vincent and others, 2004). The beginning of the 20th century is characterized by a short period of negative mass balances (minimum around 1905) followed by a relative maximum in 1910. In 1911, a strongly negative mass balance was found, after which a period with mainly positive mass balances can be detected (maximum around 1914). Again, this is in good agreement with other mass-balance reconstructions of Alpine glaciers, such as Hintereisferner (Nicolussi, 1994; Kuhn and others, 1997).

### 3.2. Proxy-based reconstructions of mass balance back to 1500

To reconstruct annual glacier mass balance back to 1500, we used the same method described in section 3.1 but with different inputs. We focused attention on a longer-term reconstruction of mass balance. Because our data were not available in monthly resolution back to 1500, we



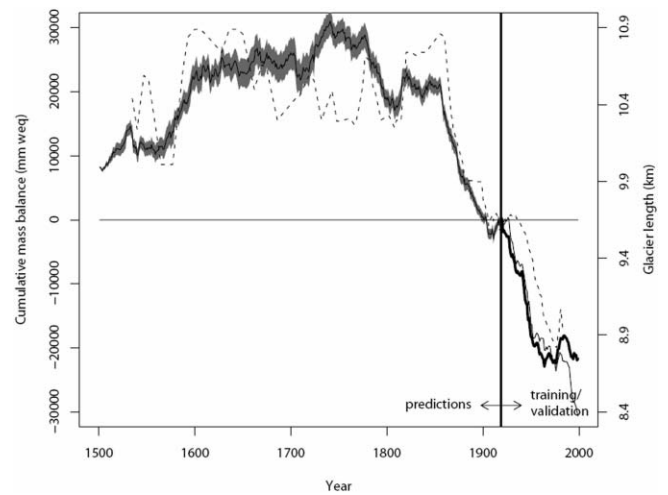
**Fig. 4.** Results of the BPN simulation and reconstruction (solid curve) of yearly mass balance. Confidence intervals derived from rms errors appear as grey envelopes around predictions. Also shown is the proxy of annual glacier mass balance for the 1919–99 period. The smoothed thick curve represents the 30 year low-pass filtered time series of the results of the BPN simulation and reconstruction. The smoothed solid curve is the 30 year low-pass filtered output of the stepwise multiple linear regression model.

used seasonal averages. After a stepwise multiple linear regression with the seasonal averages, calculated from gridded datasets, as independent variables and with annual glacier mass balance as output dataset (dependent), the potential climate influences were chosen. Significant datasets (>99% confidence level) were winter (December–February) precipitation and summer (June–August) temperature. As presented in section 3.1, the input data were standardized before training.

Figure 4 shows the simulation (1919–99) and the reconstruction (1500–1918) of the annual glacier mass balance, with an error envelope (grey shaded) around each section of BPN predictions based on its rms errors (95% confidence interval). Also given are a 30 year low-pass filtered time series for the BPN simulation and reconstruction of yearly mass balance (thick line), and the output of the stepwise multiple linear regression (solid line). Figure 5 represents the cumulative glacier mass-balance changes back to 1500, also with an error envelope (95% confidence interval). A positive slope in Figure 5 is the result of years with positive mass balance, indicating a mass gain, and the opposite for a negative slope.

Again, the BPN was trained by the conjugate gradient descent method to simulate the mass balance over the 1919–99 training/validation period. The model was then used to reconstruct annual mass changes since 1500 by averaging 30 different BPN runs that started from different points in the space of possible weights. In this case, the simulation skill amounts to 60.8% of explained variance, while the stepwise multiple linear regression approach driven by the same input data accounts for 54.8% of explained variance. Again, the autocorrelation coefficient function indicates no significant deflection from zero (95% confidence level).

A comparison of the reconstructed cumulative mass balance of Grosse Aletschgletscher and the statistics of front positions of nearby Unter Grindelwaldgletscher (e.g.



**Fig. 5.** Cumulative mass-balance changes in Grosse Aletschgletscher for the 1500–1999 period (1919 = 0). The thick curve represents the mass changes observed (1919–99); the solid curve is the result of the BPN simulation and reconstruction. Also shown is an error envelope (grey shading) around the predictions of cumulative mass balance. The dashed curve shows the length fluctuations of nearby Unter Grindelwaldgletscher, 1535–1883.

Zumbühl and others, 1983; Holzhauser and Zumbühl, 2003), provided in Figure 5, shows correlation coefficients of  $r = 0.89$  for the overlapping 1535–1883 period and  $r = 0.69$  for the 1535–1918 reconstruction period.

Figures 4 and 5 show significant trends in glacier mass-balance fluctuations over the past five centuries.

Around the mid-16th century, Grosse Aletschgletscher showed generally negative mass balances, with a consequent minimum in cumulative mass balance at this time. As shown in Figures 6 and 7 (Luterbacher and others, 2004; Pauling and others, in press), this coincides with an extended period of higher summer temperatures. This period ends when winter precipitation increases.

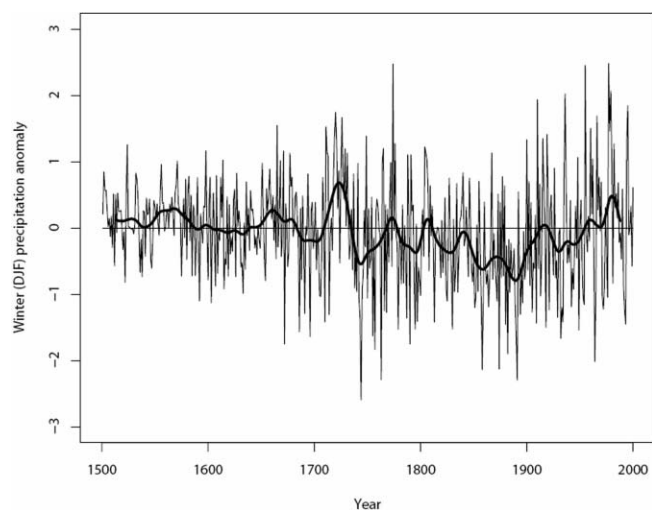
The maximum mass balances around 1600 are connected with low summer temperatures and average winter precipitation. These conditions led to low ablation rates and an extended period of increasing mass balance.

During the next century (~1610–1710), mass balance remained static and showed a slight negative trend. Summer temperature and winter precipitation are also found to stagnate. Cumulative mass balance shows a broad plateau in the 17th century, with two relative minima in the mid-17th century and at the beginning of the 18th century.

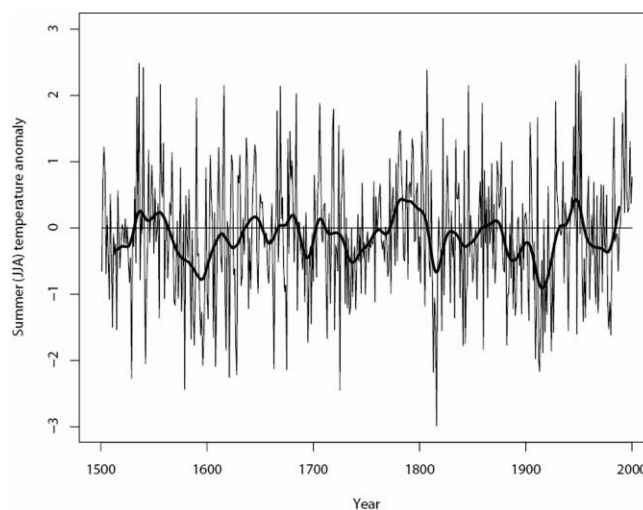
From around 1710, decreasing summer temperatures and very high winter precipitation caused a phase of uniform positive mass balance with a mass gain. The later stagnancy of glacier mass balance is represented by a nearly constant cumulative mass balance. This period of maximal cumulative mass balance coincides with a minimum in glacier length (Haeberli and Holzhauser, 2003).

From 1770 to 1810, low winter accumulation potential and high summer ablation potential overlap, producing conditions favourable for negative mass balances and a resulting continuous mass loss, with a minimum of cumulative mass balance around 1810. The 1856 glacial maximum was likely produced by 40–50 years of cool summers without anomalous positive winter precipitation, identifiable





**Fig. 6.** Winter (December–February) precipitation anomaly for the 1500–1999 period (solid curve) over the grid range used in this study (after Pauling and others, in press). Also shown is the 30 year low-pass filtered time series of the winter precipitation model input (thick curve). The time series is z-standardized relative to the 1919–99 calibration average.



**Fig. 7.** Summer (June–August) temperature anomaly for the 1500–1999 period (solid curve) over the grid range used in this study (after Luterbacher and others, 2004). Also shown is the 30 year low-pass filtered time series of the summer temperature model input (thick curve). The time series is z-standardized relative to the 1919–99 calibration average.

as two short periods of positive mass balance. After 1856, higher summer temperatures and low winter precipitation caused an extended ablation-dominated mode with an enormous mass loss until the beginning of the 20th century. From then until the 1920s, retreat rates slowed substantially, and a small mass gain was determined. Low summer temperatures coupled with increasing winter precipitation coincided with an increasing (positive) mass balance. Note that for this time a little stagnancy in length was documented for Grosse Aletschgletscher (VAW/SANW, 1881–2002; Zumbühl and Holzhauser, 1988). The mid-1950s mass loss was characterized by high summer temperatures, while the subsequent positive mass balances in the 1980s can be explained by lower summer temperatures and high winter precipitation.

The evidence presented here suggests that maxima in the mass balance of Grosse Aletschgletscher are predominantly caused by cold summers, often coupled with high winter precipitation. This is in good agreement with Oerlemans and Reichert (2000) who show that summer temperature is the important factor in drier (inner Alpine) climates. Negative mass balance is mainly driven by high summer temperature. There is probably a weaker dependence on low winter precipitation, which can be influential under strong enough forcing. This longer-term model was driven by a very small number of forcing factors, so spring and fall conditions, and the effects of summer precipitation, are not taken into account.

#### 4. CONCLUSIONS AND OUTLOOK

For the first time, an NNM has been used to reconstruct the mass balance of Grosse Aletschgletscher. Two new gridded datasets of temperature and precipitation were applied to the NNM as potential driving factors of the glacier system. This was a unique opportunity to bring these two aspects, a new method and two new datasets, together.

The results of this study show that the NNM approach is a useful tool for quantifying Grosse Aletschgletscher's mass-balance changes in a non-linear way. In fact, the reconstructed mass balance is the result of a combination of several climatic input variables (precipitation and temperature) that vary in their composition and input importance. It can be shown how the mass balance reacts to changes in local to regional temperature and precipitation. We observed maximum mass balances for Grosse Aletschgletscher around 1600, 1730, 1815/45 and 1920. Minima in glacier mass balance were found in the 1540s, 1790s, 1870s and 1950s.

Furthermore, we confirm that summer temperature is an important driving factor for variations in mass balance. Thus the results presented here suggest that neural network models capture the appropriate dependence on the relevant inputs that likely affect a glacier system. We also conclude that the climate–mass-balance relation of Grosse Aletschgletscher consists of a significant non-linear part.

In our approach we used only temperature and precipitation data as inputs for the BPN. It would be interesting to see whether other data (e.g. North Atlantic Oscillation, solar irradiance or tree rings) could improve the quality of the model. In the absence of widespread mass-balance data that can be used as training sets, the possibility of using the present approach to model glacier length variations should be further investigated.

#### ACKNOWLEDGEMENTS

This study was supported by the following institutions: Stiftung Marchese Francesco Medici del Vascello, and the Swiss National Science Foundation, through its National Centre of Competence in Research on Climate (NCCR Climate), project PALVAREX. The authors are grateful to MeteoSwiss, J. Luterbacher and A. Pauling for providing instrumental and multi-proxy reconstructions of temperature

and precipitation. Thanks also go to A. Bauder for providing data from Grosse Aletschgletscher and for helpful advice. The detailed comments and fruitful suggestions of H. Wanner, D.B. Reusch, an anonymous reviewer and the Scientific Editor, R.A. Bindschadler, helped to improve this paper.

## REFERENCES

- Addison, J.F.D., K.J. McGarry, S. Wermter and J. MacIntyre. 2004. Stepwise linear regression for dimensionality reduction in Neural Network Modelling. In Hamza, M.H., ed. *Artificial intelligence and applications*. Calgary, ACTA Press Inc., 363–368.
- Adrian, E.D. 1926. The impulses produced by sensory nerve endings. *J. Physiol.*, **61**, 49–72.
- Aellen, M. 1995. Glacier mass balance studies in the Swiss Alps. *Z. Gletscherkd. Glazialgeol.*, **31**, 159–168.
- Aellen, M. and M. Funk. 1990. Bilan hydrologique du bassin versant de la Massa et bilan de masse des glaciers d'Aletsch (Alpes bernoises, Suisse). *International Association of Hydrological Sciences Publication* 193, 89–98.
- Amit, D.J. 1989. *Modeling brain function: the world of attractor neural networks*. Cambridge, Cambridge University Press.
- Anderson, J.A. and E. Rosenfeld, eds. 1986. *Neurocomputing: foundations of research*. Cambridge, MA, MIT Press.
- Begert, M., G. Seiz, T. Schlegel, M. Musa, G. Baudraz and M. Moesch. 2003. *Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961–1990. Schlussbericht des Projekts NORM90*. Zürich, MeteoSchweiz.
- Bishop, C.M. 1995. *Neural networks of pattern recognition*. Oxford, Oxford University Press.
- Braithwaite, R.J. and S.C.B. Raper. 2002. Glaciers and their contribution to sea level change. *Phys. Chem. Earth*, **27**(32), 1445–1454.
- Bundesamt für Umwelt, Wald und Landschaft (BUWAL), Bundesamt für Wasser und Geologie (BWG), MeteoSchweiz. 2004. *Auswirkungen des Hitzesommers 2003 auf die Gewässer*. Bern, Bundesamt für Umwelt, Wald und Landschaft.
- Chen, J. and M. Funk. 1990. Mass balance of Rhonegletscher during 1882/83–1986/87. *J. Glaciol.*, **36**(123), 199–209.
- Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Contr. Signals Syst.*, **2**(4), 303–314.
- Dao, V.N.P. and V.R. Vemuri. 2002. A performance comparison of different back propagation neural networks methods in computer network intrusion detection. *Diff. Eqn. Dyn. Sys.*, **10**(1–2), 201–214.
- Frei, C. and C. Schär. 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, **18**(8), 873–900.
- Grossberg, S. 1982. How does a brain build a cognitive code? In Grossberg, S., ed. *Studies of mind and brain: neural principles of learning, perception, development, cognition, and motor control*. Boston, MA, Reidel Publishing, 1–52.
- Haerberli, W. and H. Holzhauser. 2003. Alpine glacier mass changes during the past two millennia. *PAGES News*, **11**(1), 13–15.
- Haerberli, W., M. Hoelzle, S. Suter and R. Frauenfelder, eds. 1998. *Fluctuations of glaciers 1990–1995*. Paris, International Association of Hydrological Sciences (International Commission on Snow and Ice)/United Nations Environment Programme/UNESCO.
- Holzhauser, H. and H. J. Zumbühl. 1999. Glacier fluctuations in the Western Swiss and French Alps in the 16th century. *Climatic Change*, **43**(1), 223–237.
- Holzhauser, H. and H.J. Zumbühl. 2003. Nacheiszeitliche Gletscherschwankungen. In Weingartner, R. and M. Spreafico, eds. *Hydrologischer Atlas der Schweiz. Blatt 3.8*. Bern-Wabern, Bundesamt für Landestopographie.
- Houghton, J.T. and 7 others, eds. 2001. *Climate change 2001: the scientific basis*. Cambridge, Cambridge University Press.
- Hsieh, W.W. and B.Y. Tang. 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteorol. Soc.*, **79**(9), 1855–1870.
- Jones, P.D. and M.E. Mann. 2004. Climate over past millennia. *Rev. Geophys.*, **42**(2), RG2002. (10.1029/2003RG000143.)
- Kasser, P. 1954. Sur le bilan hydrologique des bassins glaciaires avec application au Grand Glacier d'Aletsch. *International Association of Scientific Hydrology Publication* 39 (General Assembly of Rome 1954 – Snow and Ice), 331–350.
- Kasser, P. 1967. *Fluctuations of glaciers 1959–1965: a contribution to the International Hydrological Decade*. Paris, International Commission of Snow and Ice of the International Association of Hydrological Sciences/UNESCO.
- Klein, B.D. and D.F. Rossin. 1999. Data errors in neural network and linear regression models: an experimental comparison. *Data Quality*, **5**(1).
- Kuhn, M., E. Schlosser and N. Span. 1997. Eastern Alpine glacier activity and climatic records since 1860. *Ann. Glaciol.*, **24**, 164–168.
- Lie, Ø., S.O. Dahl and A. Nesje. 2003. Theoretical equilibrium-line altitudes and glacier buildup sensitivity in southern Norway based on meteorological data in a geographical information system. *Holocene*, **13**(3), 373–380.
- Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean and H. Wanner. 2004. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, **303**(5663), 1499–1503.
- Mann, M.E. 2002. The value of multiple proxies. *Science*, **297**(5586), 1481–1482.
- Mann, M.E. and P.D. Jones. 2003. Global surface temperatures over the past two millennia. *Geophys. Res. Lett.*, **30**(15), 1820. (10.1029/2003GL017814.)
- Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**(6), 1087–1092.
- Michaelson, J. 1987. Cross-validation in statistical climate forecast models. *J. Clim. Appl. Meteorol.*, **26**(11), 1589–1600.
- Müller-Lemans, H., M. Funk, M. Aellen and G. Kappenberger. 1995. Langjährige Massenbilanzreihen von Gletschern in der Schweiz. *Z. Gletscherkd. Glazialgeol.*, **30**, 141–160.
- Nesje, A. and S.O. Dahl. 2003. The 'Little Ice Age' – only temperature? *Holocene*, **13**(1), 139–145.
- Nicolussi, K. 1994. Jahrringe und Massenbilanz: Dendroklimatologische Rekonstruktion der Massenbilanzreihe des Hintereisferners bis zum Jahr 1400 mittels *Pinus cembra*-Reihen aus den Ötztaler Alpen. *Z. Gletscherkd. Glazialgeol.*, **30**, 11–52.
- Oerlemans, J. and B.K. Reichert. 2000. Relating glacier mass balance to meteorological data by using a seasonal sensitivity characteristic. *J. Glaciol.*, **46**(152), 1–6.
- Paterson, W.S.B. 1994. *The physics of glaciers. Third edition*. Oxford, etc., Elsevier.
- Pauling, A., J. Luterbacher, C. Casty and H. Wanner. In press. 500 years of gridded high-resolution precipitation reconstruction over Europe and the connection to large-scale circulation. *Climate Dyn.*
- Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery. 1992. *Numerical recipes in C: the art of scientific computing*. Cambridge, Cambridge University Press.
- Reichert, B.K., L. Bengtsson and J. Oerlemans. 2001. Midlatitude forcing mechanisms for glacier mass balance investigated using general circulation models. *J. Climate*, **14**(17), 3767–3784.
- Reusch, D.B. and R.B. Alley. 2004. A 15-year West Antarctic climatology from six automatic weather station temperature and pressure records. *J. Geophys. Res.* **109**(D4, D04103). (10.1029/2003JD004178.)
- Rosenblatt, F. 1958. The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, **65**, 386–408.

- Rumelhart, D.E. and J.L. McClelland, eds. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA, MIT Press.
- Rumelhart, D.E., G.E. Hinton and R.J. Williams. 1986. Learning internal representations by error propagation. In Rumelhart, D.E. and J.L. McClelland, eds. *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA, MIT Press, 318–362.
- Schüepf, M., M. Bouët, M. Bider and Ch. Urfer. 1978. *Regionale Klimabeschreibungen. Teil 1*. Zürich, Schweizerische Meteorologische Zentralanstalt.
- Stone, M. 1974. Cross-validation choice and the assessment of statistical predictions. *J. Roy. Stat. Soc.*, **36**, 111–147.
- Van der Knaap, W.O. and J.F.N. van Leeuwen. 2003. Climate–pollen relationships AD 1901–1996 in two small mires near the forest limit in the northern and central Swiss Alps. *Holocene*, **13**(6), 809–828.
- Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie/Schweizerischen Akademie der Naturwissenschaften (VAW/SANW). 1881–2002. *Die Gletscher der Schweizer Alpen. Nos. 1–122*. Zürich, ETH Zürich. Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie der ETH Zürich.
- Vincent, C., G. Kappenberger, F. Valla, A. Bauder, M. Funk and E. Le Meur. 2004. Ice ablation as evidence of climate change in the Alps over the 20th century. *J. Geophys. Res.* **109**(D10), D10104. (10.1029/2003JD003857.)
- Von Storch, H. and F.W. Zwiers. 1999. *Statistical analysis in climate research*. Cambridge, Cambridge University Press.
- Walter, A. 2001. *Zur Anwendung neuronaler Netze in der Klimatologie*. Offenbach, Berichte des Deutschen Wetterdienstes 218.
- Walter, A. and C.D. Schönwiese. 2002. Attribution and detection of anthropogenic climate change using a backpropagation neural network. *Meteorol. Z.*, **11**(5), 335–343.
- Walter, A. and C.D. Schönwiese. 2003. Nonlinear statistical attribution and detection of anthropogenic climate change using a simulated annealing algorithm. *Theor. Appl. Climatol.*, **76**(1–2), 1–12.
- Wu, A. and W. W. Hsieh. 2003. Nonlinear interdecadal changes of the El Niño–Southern Oscillation. *Climate Dyn.*, **21**(7–8), 719–730.
- Zumbühl, H.J. and H. Holzhauser. 1988. Alpengletscher in der Kleinen Eiszeit. *Die Alpen*, **64**(3), 129–322.
- Zumbühl, H.J., B. Messerli and C. Pfister. 1983. *Die kleine Eiszeit: Gletschergeschichte im Spiegel der Kunst*. Bern and Luzern, Swiss Alpine Museum and Gletschergarten Museum.

MS received 13 July 2004 and accepted in revised form 13 February 2005