



# Neural networks applied to estimating subglacial topography and glacier volume

G. K. C. Clarke, E. Berthier, C. G. Schoof, A. H. Jarosch

## ► To cite this version:

G. K. C. Clarke, E. Berthier, C. G. Schoof, A. H. Jarosch. Neural networks applied to estimating subglacial topography and glacier volume. *Journal of Climate*, 2009, 22 (9), pp.2146-2160. 10.1175/2008JCLI2572.1 . hal-00403093

**HAL Id: hal-00403093**

**<https://hal.science/hal-00403093>**

Submitted on 22 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Neural Networks Applied to Estimating Subglacial Topography and Glacier Volume

GARRY K. C. CLARKE

*Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada*

ETIENNE BERTHIER

*Université de Toulouse–OMP/LEGOS, and Centre National de la Recherche Scientifique–LEGOS, Toulouse, France*

CHRISTIAN G. SCHOOF AND ALEXANDER H. JAROSCH

*Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada*

(Manuscript received 9 April 2008, in final form 5 September 2008)

## ABSTRACT

To predict the rate and consequences of shrinkage of the earth's mountain glaciers and ice caps, it is necessary to have improved regional-scale models of mountain glaciation and better knowledge of the subglacial topography upon which these models must operate. The problem of estimating glacier ice thickness is addressed by developing an artificial neural network (ANN) approach that uses calculations performed on a digital elevation model (DEM) and on a mask of the present-day ice cover. Because suitable data from real glaciers are lacking, the ANN is trained by substituting the known topography of ice-denuded regions adjacent to the ice-covered regions of interest, and this known topography is hidden by imagining it to be ice-covered. For this training it is assumed that the topography is flooded to various levels by horizontal lake-like glaciers. The validity of this assumption and the estimation skill of the trained ANN is tested by predicting ice thickness for four  $50 \text{ km} \times 50 \text{ km}$  regions that are currently ice free but that have been partially glaciated using a numerical ice dynamics model. In this manner, predictions of ice thickness based on the neural network can be compared to the modeled ice thickness and the performance of the neural network can be evaluated and improved. From the results, thus far, it is found that ANN depth estimates can yield plausible subglacial topography with a representative rms elevation error of  $\pm 70 \text{ m}$  and remarkably good estimates of ice volume.

## 1. Introduction

The sea level equivalent volumes of the Greenland and Antarctic ice sheets are 7.3 and 56.6 m, respectively, whereas the combined volume of glaciers and small ice caps is far less. Yet, over the next 100 yr their 0.15–0.37-m contribution to sea level rise (Lemke et al. 2007) is expected to dominate that from shrinkage of the great ice sheets (e.g., Ohmura 2004; Meier et al. 2007). Despite such compelling reasons to be interested in the volumes of glaciers and ice caps, only a few hundred of the more than  $10^5$  glaciers and ice caps that exist today have been geophysically mapped. Estimates of glacier and ice cap

volume are therefore only weakly constrained by depth measurements, and it is extremely unlikely that this situation will change. For this reason, attention has focused on scaling relationships such as

$$V = cA^\gamma \quad (1)$$

(e.g., Chen and Ohmura 1990) between glacier area (which is readily observed) and glacier volume. The Chen and Ohmura study was based on statistical regression of data from 63 mountain glaciers and yielded  $\gamma = 1.357$  as the best-fit value for the exponent in (1). By elegant use of dimensional analysis, Bahr (1997) and Bahr et al. (1997) derived a theoretical value of  $\gamma = 11/8 = 1.375$  for the exponent, in remarkable agreement with Chen and Ohmura (1990). However, the coefficient  $c$  in (1) remains beyond the reach of dimensional analysis and must be treated as a fitting parameter that can differ from region to region.

*Corresponding author address:* Garry K. C. Clarke, Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, BC V6T 1Z4, Canada.  
E-mail: clarke@eos.ubc.ca

Several physically based approaches to estimating subglacial topography have received attention. The earliest of these (Nye 1952) is based on the recognition that ice resembles a perfectly plastic material and thus, for actively deforming glaciers, the basal shear stress  $\tau_0$  should be close to the plastic yield stress for ice. Driedger and Kennard (1986) applied this approach to estimate the volume of glaciers of known thickness and suggested that their volume estimates were accurate to  $\pm 20\%$ . A similar approach has been followed by Haeberli (1985) and Haeberli and Hoelzle (1995). Plasticity estimators require no knowledge of the mass balance forcing for the glacier but do assume, implicitly, that the glacier is healthy enough to maintain its basal stress near the yield stress. An alternative approach that requires additional assumptions is to assume that the glacier is near a steady-state configuration with respect to a known or estimated mass balance forcing. With this assumption the balance ice flux can be calculated and the balance flux can be inverted to ice thickness using Glen's flow law (Huss et al. 2008).

A shortcoming of the volume–area scaling approach is that it yields no useful information about subglacial topography—a necessary boundary condition for glacier dynamics models. In contrast, the physics-based methods allow ice thickness to be estimated but are subject to error if their underlying assumptions are not fulfilled. This motivates our interest in a fresh approach to estimating the thickness and volume of glaciers. The aim of the present contribution is to explore the potential of artificial neural networks (e.g., Bishop 1995; Reed and Marks 1999) as a tool for estimating ice thickness. In the next section we introduce artificial neural networks (ANNs), review their applications in climate science and glaciology, and describe how they are trained and used in the present study. In section 3, we describe the construction of test datasets using a numerical ice dynamics model so that the skill of trained neural networks can be objectively tested. Section 4 is concerned with testing the skill of ANNs, using this information to optimize the selection of inputs and the network architecture, and evaluating the performance of the ANN estimators for four test regions and a range of glaciation states. Section 5 summarizes the results of this study.

## 2. Artificial neural networks

Artificial neural networks are input–output systems that aim to imitate the operation of biological neural networks. These biological networks are characterized by an intricate interconnectivity among neurons and by the possibility of modifying this connectivity so that learning

can proceed. In ANNs the response  $F(u)$  of an individual neuron to an input  $u$  is approximated by either a non-linear function that smoothly or discontinuously switches between binary “on” (1) and “off” (0) states or as a linear function  $F(u) = u$ . ANNs simplify the natural situation by limiting the number of neurons and by assuming that the network connectivity has a systematic structure. The most commonly used ANNs have a multilayered architecture with a unidirectional feedforward flow of information and are termed multilayer perceptrons (e.g., Bishop 1995; Reed and Marks 1999).

In climate science, ANNs have been successfully used to reduce the dimensionality of data and to classify patterns (e.g., Cavazos 2000; Hsieh 2001; Hewitson and Crane 2002), to predict large-scale climate oscillations (e.g., Tangang et al. 1998a,b), and to approximate the behavior of complex physical systems (e.g., Monahan 2000; Tang and Hsieh 2001). In glaciology, ANNs have received less attention, but Reusch et al. (2005) used them to attempt an association of ice-core climate records from West Antarctica with synoptic climate patterns identified in a 15-yr climate reanalysis dataset. Subsequently, Reusch and Alley (2007) used self-organizing maps (a special category of ANN; Kohonen 2001) to classify patterns of the extent and concentration of Antarctic sea ice. In an intriguing series of recent papers, ANNs have been trained to emulate the fluctuations in glacier length that are associated with changes in air temperature and precipitation. This approach has allowed historical variations in glacier mass balance to be constructed (Steiner et al. 2005), past length variations to be comprehended (Steiner et al. 2008), and future changes to be predicted (Zumbühl et al. 2008)—all without reference to a physical ice dynamics model.

In this presentation we shall, for the first time, apply ANNs to the problem of estimating subglacial topography using geometric information extracted from a digital elevation model (DEM). Most field glaciologists, when standing on a glacier, would think themselves capable of guessing, albeit roughly, the thickness of ice beneath them. The underpinnings of such a guess might include a qualitative awareness of the distance between the observer and the valley walls and the steepness of the surrounding topography. The fact that such a guess is possible and has some observational basis suggests that the depth estimation process can be formalized and even automated. As illustrated in Fig. 1, our task starts with a glacierized digital elevation model, and entails the estimation of ice thickness for each of the ice-covered cells. By assembling these estimates, an ice-denuded DEM (Fig. 1c) can be constructed for the same terrain.

We assume that the process that transforms Fig. 1a to Fig. 1c can be represented by a multilayer feedforward

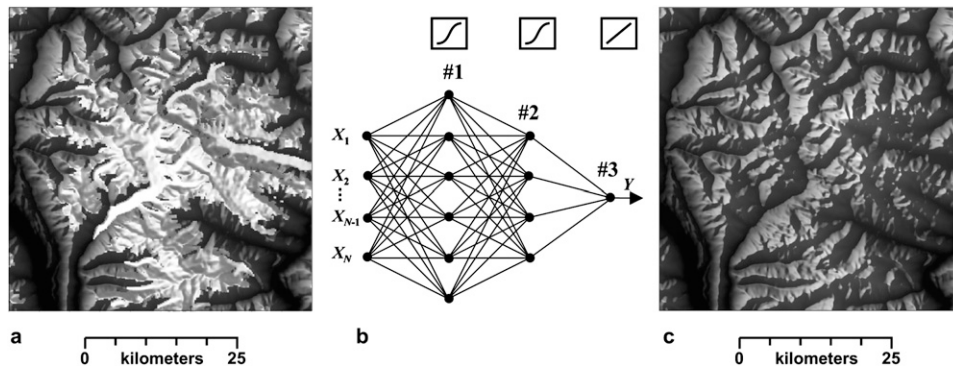


FIG. 1. Actual (ice covered) and estimated (ice denuded) DEMs for a partially ice-covered region in the vicinity of the Mount Waddington ice field, BC, Canada (map center at 51.3333°N, 125.33°W). The DEM resolution is 200 m. (a) Input DEM of surface topography with glacier-covered cells indicated by light gray shading. (b) Input–output system represented as an artificial neural network. (c) Output DEM of estimated bed surface topography using neural network operations on input DEM.

ANN (Fig. 1b) and use the MATLAB Neural Network Toolbox (Demuth et al. 2006) to train and then apply the ANN. Following the simplest options of the MATLAB Toolbox, we adopt the standard Levenberg–Marquardt back-propagation training algorithm (Levenberg 1944; Marquardt 1963) and the Widrow–Hoff least squares learning rule. Only 60% of the data in the training set are used for training the ANN; 20% are held back to allow the training result to be independently validated and part, but not all, of the remaining 20% are used to perform tests during training so that signs of overtraining can be detected and the training episode terminated.

The practical problem becomes one of optimizing the selection of inputs and the architecture of the ANN. This problem can be stated as follows: *For a DEM cell centered at some map position  $(i, j)$  with surface elevation  $S_{ij}$ , estimate the ice thickness  $H_{ij}$  or, equivalently, the bed surface elevation  $B_{ij} = S_{ij} - H_{ij}$ .* An ANN tailored to solve this problem can be applied to every ice-covered cell on a given DEM in order to generate a bed surface map  $B_{ij}$  for the region of interest; clearly,  $B_{ij} = S_{ij}$  for cells that are not ice covered.

#### a. Geomorphic premise

Our supposition that ANNs can be applied to the problem of estimating subglacial topography is based on unstated geomorphic assumptions that we shall attempt to identify: (i) Within a particular geographical region there is a sameness to the landscape that is a consequence of the sameness of the bedrock geology, geological and environmental history, and present conditions for that region. (ii) The deglaciated portions of landscapes that today are partially ice covered have geometrical similarities to portions that are glaciated;

for the most part, the areas that are now ice denuded were formerly ice covered and therefore subject to similar landscape-shaping processes that currently operate on the ice-covered landscape. (iii) Because the geological and environmental settings are spatially varying, it follows that a neural network that has been trained to estimate ice thickness in a particular geographical region may not perform well if applied to another region.

#### b. Data masks

In addition to the DEM of surface elevation  $S$ , it is necessary to define an ice mask  $\mathbf{I}$  that has the properties  $I_{ij} = 0$  for an ice-free DEM cell and  $I_{ij} = 1$  for an ice-covered one. A second mask that can prove useful is a surface slope mask  $\mathbf{G}$  that discriminates between steeply sloping and gently sloping topography. For steeply sloping topography  $G_{ij} = 1$  and for gently sloping topography  $G_{ij} = 0$ . In our work, the threshold between the two is taken as  $|\theta_0| = 25^\circ$  but is not a sensitive parameter and could be adjusted. The slope mask can be used to distinguish among gently sloping glaciers ( $\mathbf{I} \wedge \mathbf{G}'$ ), steeply sloping glaciers and ice-covered valley walls ( $\mathbf{I} \wedge \mathbf{G}$ ), and an ice-free surface of any slope ( $\mathbf{I}'$ ). We have found that for steeply sloping glaciers a physics-based estimator (discussed subsequently) gives better thickness estimates than ANN estimators, and the slope threshold determines which estimator will be enlisted. Examples of the  $\mathbf{I}$ ,  $\mathbf{G}$ ,  $\mathbf{I} \wedge \mathbf{G}'$ , and  $\mathbf{I} \wedge \mathbf{G}$  masks are presented in Fig. 2. Here,  $\wedge$  denotes the logical *and* operator, the superscript prime denotes the logical *not* operator, and  $\vee$  denotes the nonexclusive logical *or* operator.

#### c. Inputs and outputs

Referring to Fig. 1b, for each ice-covered cell there is a single output  $Y = \tilde{H}_{ij}$ , where the tilde indicates an



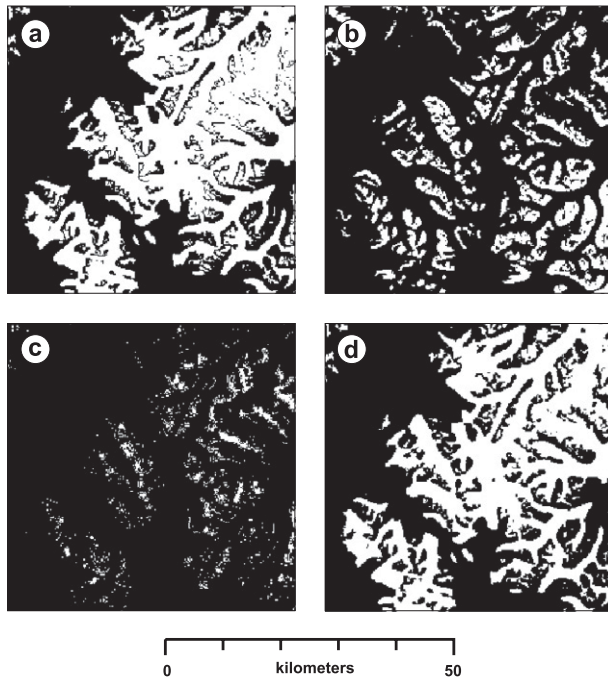


FIG. 2. Examples of ice and slope masks for site BC-3 (see Fig. 6 for its location) with roughly 50% ice cover and a deglacial phase of the climate cycle. (a) Icemark ( $I$ ). (b) Slope mask with  $G$  slope threshold  $\theta = 25^\circ$ . (c) Ice *and not* slope mask ( $I \wedge G'$ ). This mask isolates gently sloping ice-covered regions. (d) Ice *and* slope mask ( $I \wedge G$ ). This mask isolates steeply sloping ice-covered regions.

estimate of the true value  $H_{ij}$ . The situation with respect to the inputs  $X_1 \dots X_N$  is less clear-cut and starts with an enumeration of information that might provide some useful constraints on the ice thickness estimate. The possibilities that we examined are (i) the surface elevation of the ice-covered cell, (ii) the orientation of the surface slope at  $(i, j)$ , (iii) the distance between the point  $(i, j)$  and one or more points on the surrounding valley walls, and (iv) the slope at points on the surrounding valley walls. In terms of the  $X_n$  inputs in Fig. 1b, the surface elevation input would correspond to  $S_{ij}$ , and the slope orientation inputs to  $(t_x)_{ij}$  and  $(t_y)_{ij}$  where  $(t_x, t_y)$  are components of the unit vector  $\mathbf{t}$  that is aligned with the surface slope  $\nabla S$  at  $(i, j)$ .

To calculate the distance to the valley walls, we apply a sectorial stencil (Fig. 3a) obtained by dividing the compass circle into  $M$  sectors and measure the minimum horizontal distance  $R_m$  between the point  $(i, j)$  and the valley walls within that sector. This procedure yields  $M$  additional data inputs. We have experimented with two different practical definitions of “valley walls.” In early work we assumed that valley walls correspond to ice-free areas ( $I'$ ) that surrounded a glacier, but we found this definition to be too narrow and now define the valley walls as being either ice free or ice covered

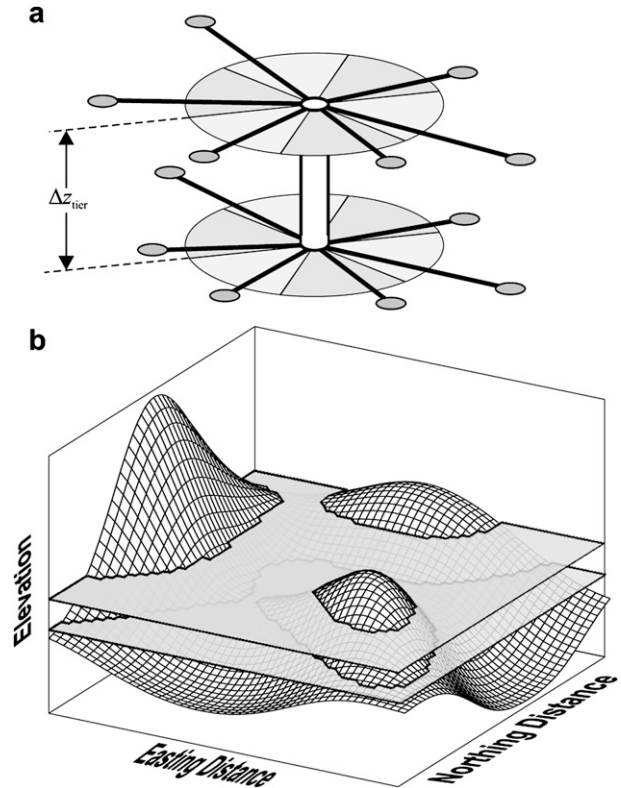


FIG. 3. (a) Example of a computational stencil for the neural network inputs. The compass circle is divided into  $M$  sectors ( $M = 6$  in the example) and within each sector the minimum range to a DEM cell satisfying a condition such as “ice free” is calculated. More than one tier is allowed with a distance  $\Delta z_{\text{tier}}$  separating the two tier levels. The logical condition can be different from tier to tier. (b) Illustration of bathtub filling of the landscape to two different levels.

but steeply sloping ( $I' \vee G$ ). To illustrate our procedure, suppose that a four-sector stencil was placed at the map point  $(i, j)$  on a glacier of unknown thickness and the stencil was aligned with the cardinal compass directions (north, east, south, west). We would first look north ( $\pm 45^\circ$ ) to find the nearest point that was ice free or steeply sloping, then east, etc., to obtain four distance inputs to the ANN. The usefulness of a two-tiered sectorial stencil (Fig. 3a), with the two elevation levels separated by some distance  $\Delta z_{\text{tier}}$ , was also tested. Valley-wall distances for the lower tier were calculated as described above and, for the upper tier, the valley-wall distances were taken as the shortest distance *at the same elevation as the upper tier*. In principle, ANN inputs from a two-tier sectorial stencil contain information on the slope as well as the range of the valley walls. An alternative approach to including valley-wall slope information is to employ a one-tier stencil but add slope magnitude information  $|\nabla S|_m$  for each of the  $R_m$  valley-wall points.

An analysis of the information value of the various ANN inputs that we tested is described in section 4. It was found that the wall-distance measurements carry essential information for the ANN and that none of other inputs delivered consistent improvements in network performance.

Throughout this study the input DEMs and masks will cover an area of  $50 \text{ km} \times 50 \text{ km}$  at a resolution of 200 m. Rather than make special assumptions at the map boundaries we avoid the boundaries entirely by setting an upper limit on range distance  $\max(R_m) = 6 \text{ km}$  and only centering the stencil on interior points that are at least 6 km from any boundary. Thus the size of the output map is  $38 \text{ km} \times 38 \text{ km}$ . It is important to choose a value for  $\max(R_m)$  that exceeds the expected distance between on-glacier points and the surrounding valley walls but is not so large that data loss at the map margins is substantial. As examples, if we had taken  $\max(R_m) = 15 \text{ km}$ , the output map size would be reduced to  $20 \text{ km} \times 20 \text{ km}$ ; whereas, if  $\max(R_m) = 0.5 \text{ km}$ , many of the range distance inputs will be set to  $\max(R_m)$  rather than the true values.

Apart from these practical considerations, decisions concerning the input map size must be guided by the understanding that neural networks yield estimates that are based on the statistical properties of the input data that were used to train the network. Thus the geomorphic premise, detailed above, would caution against the use of very large input maps (unless the ANN allowed for spatial variation). One must therefore balance conflicting requirements. The input map should be sufficiently large that data loss at the map margins is not substantial but not so large that spatial inhomogeneity of the landscape significantly undermines the geomorphic premise.

#### d. Layer architecture

In addition to selecting inputs to the ANN it is necessary to make decisions about the network architecture, including such matters as the number of active layers, the dimension  $d_k$  of each layer, and the transfer function for each layer. It is customary for the first active layer to have at least as many nodes as the number of inputs  $N$ ; as our starting assumption we take  $d_1 = 2N$  but subsequently find that  $d_1 = N$  yields comparable performance with faster training. Referring to Fig. 4 and considering one of the nodes on the first active layer, there are  $N = 4$  inputs  $x_j$  converging to any given node  $i$  and these are assigned weights  $w_{ij}$  and summed; last, a constant bias  $\beta_i$  is added to yield  $u_i = \sum_j w_{ij}x_j + \beta_i$  as the node input. The purpose of the biases is to control the position of logical thresholds within the ANN and potentially improve its performance. Next, the node input

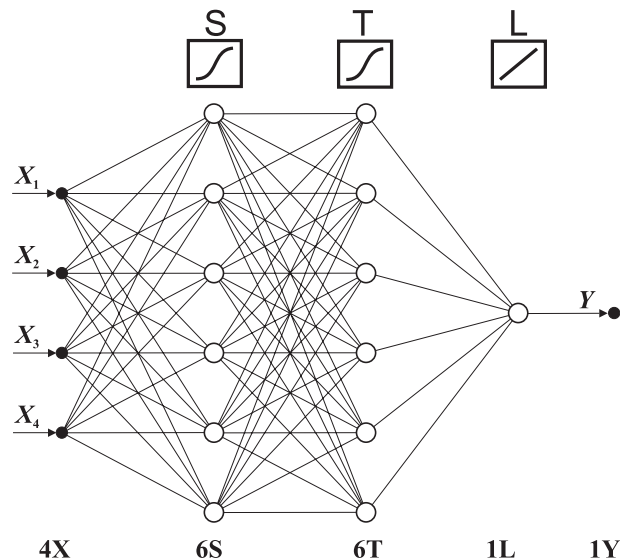


FIG. 4. Example of a multilayer feedforward artificial neural network having 4X-6S-6T-1L architecture. Solid dots represent nodes and open circles represent artificial neurons with a response functions having one of the following forms  $F(u) = 1/[1 + \exp(-u)]$  (S),  $F(u) = \tanh(u)$  (T), or  $F(u) = u$  (L). In this example the ANN has an input layer (four inputs  $X_1 \dots X_4$ ), an output layer (one output  $Y$ ), and two “hidden” layers. The three active layers, comprising the output layer and the hidden layers, are each associated with a transfer function  $F(u)$ .

is applied to a linear or nonlinear transfer function  $F(u_i)$  to produce the node output  $y_i = F(u_i)$ . Common choices for the form of the transfer function are

$$F(u_i) = \begin{cases} u_i & \text{linear} \\ 1/[1 + \exp(-u_i)] & \text{sigmoid} \\ \tanh(u_i) & \text{hyperbolic tangent} \end{cases} \quad (2)$$

(e.g., Reed and Marks 1999, p. 316). As a shorthand in subsequent discussion, we shall denote the linear transfer function by L, the sigmoid transfer function (also known as the “logistic function”) by S, and the hyperbolic tangent transfer function by T.

#### e. Training

Having discussed the inputs and architecture of the ANN, we turn attention to the problem of training the network to deliver reasonable estimates of ice thickness. In essence, training reduces to the problem of optimizing the values of the weights  $w_{ij}$  and biases  $\beta_i$  for every active node of the neural network, with the aim of minimizing the error  $\tilde{H}_{ij} - H_{ij}$  of the output estimate  $Y = \tilde{H}_{ij}$ . Here we are faced with the dilemma that the ice-covered portions of a DEM are unsuitable for training purposes because, in most cases, we have no prior information about the ice thickness. Thus the network must be trained using DEM cells that are not

ice covered. We justify this strategy by invoking the geomorphic premise: that, for the most part, glaciers exist in landscapes that are highly glaciated and the landscape signatures of glaciation are the expression of regional influences such as geology, climate, and the intensity of past glaciations. In consequence, the ice-covered parts of the contemporary landscape, if denuded of their ice cover, would resemble nearby areas that are currently deglaciated. In a subsequent section we demonstrate how these assumptions can be tested using glacier modeling but, for now, we accept them and proceed.

To generate a training set, we start with a DEM such as that of Fig. 1a that contains a mix of ice-covered and ice-denuded cells and focus attention on the ice-denuded cells. One of these cells  $(i, j)$  is selected at random and then covered by a randomly generated thickness of ice  $H_{ij}^*$  to yield an ice elevation  $Z_{ij}^*$  at that site; the entire DEM is filled to this ice level to generate a new landscape realization  $S_{ij}^* = \max(S_{ij}, Z_{ij}^*)$  as illustrated in Fig. 3b. An ice mask is generated for this filling level such that  $I_{ij}^* = 1$  when  $(I_{ij} = 1) \vee Z_{ij}^* > S_{ij}$  and  $I_{ij}^* = 0$  otherwise. (Recall that  $\vee$  represent the logical *or* operator; thus the logical expression reads: “when a cell is already ice covered or when the randomly generated ice filling level  $Z^*$  exceeds the prefilling surface elevation  $S_{ij}$ .”) The ANN inputs for the new landscape realization  $S^*$  and ice mask  $I^*$  are constructed for the point  $(i, j)$ , and the training target (i.e., the desired value of  $Y$  in Fig. 4) for this single realization is  $H_{ij}^*$ . Repeating this process (thousands of times) generates a large training set. It should be noted that there is no reward for training a neural network to perform well in irrelevant situations. We therefore edit the ANN training set so that it is concentrated on elevation bands that contain or recently contained glaciers and on ice thickness values that are plausible rather than implausible.

Clearly, real glaciers have sloping surfaces and a complex surface geometry that surely carry information concerning the subglacial geometry. It remains to be demonstrated (in section 4) that this “bathtub-filling process” leads to a sufficiently accurate representation of glaciated landscapes that ANNs trained in this manner can yield reasonable predictions of ice thickness.

By applying this “bathtub-filling process” to real glaciated DEMs such as that in Fig. 1a, we have established that well-designed neural networks, using inputs such as those described above, are highly trainable and can yield an impressive correlation between the estimated ice thickness values  $\hat{H}_{ij}$  and the target values  $H_{ij}^*$ , although with substantial scatter (e.g., Fig. 5). This result is encouraging but does not establish that the method actually performs well, because we have yet to

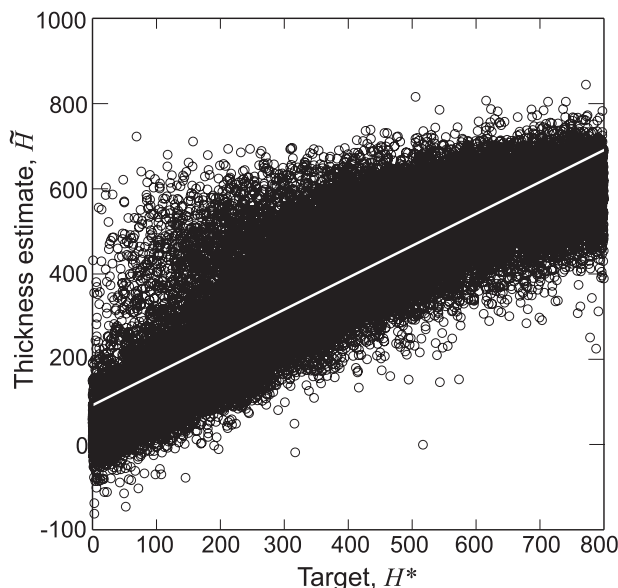


FIG. 5. A representative example of an ANN training outcome. The input DEM and ice mask are those for Fig. 1a. The correlation coefficient for the fit is  $r = 0.863$ . For this example the size of the training set was  $N_{\text{set}} = 36\,100$ .

determine whether the original hypothesis (about ice-covered topography having geometric similarities to nearby ice-denuded topography) and the bathtub-filling training strategy have shortcomings. Again, because suitable datasets are lacking, we test the success of the bathtub-filling ANN training method using synthetic data generated by a numerical ice dynamics model as described below.

### 3. Generation of test datasets

To test the skill of the ice thickness estimator, we require DEMs and ice masks for mountainous regions that have substantial ice cover and well-mapped subglacial topography. There are no suitable regions that meet these requirements. Geophysical mapping of the subglacial topography of mountain glaciers has focused on those few glaciers that have attracted scientific interest and this creates an observational bias that favors safe, accessible glaciers of modest size. Although many of Iceland's ice caps have been well mapped (e.g., Björnsson 1986), ice caps and ice sheets are unsuitable for our purposes because their geometry is only loosely controlled by subglacial topography, so the neural network method, as developed here, is unlikely to perform well. Thus we rely on glacier modeling to generate test datasets. We must necessarily assume that the modeled glaciers are acceptable surrogates for real glaciers but we do not consider this a serious shortcoming of our

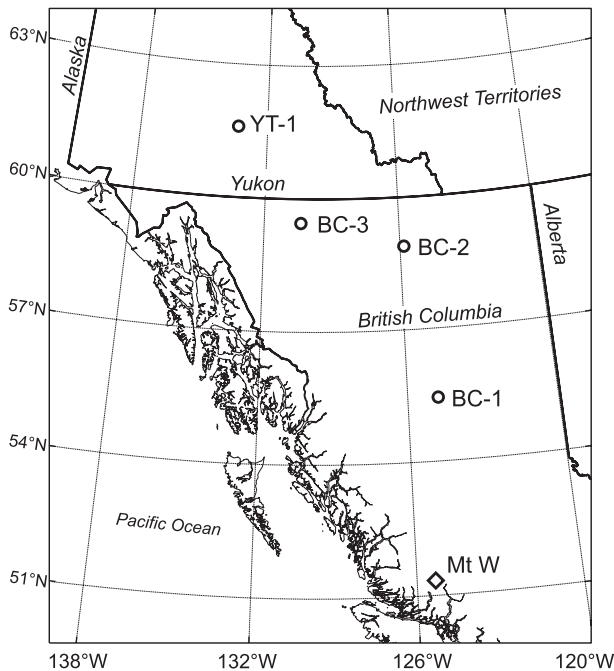


FIG. 6. Location map for the Mount Waddington area (Fig. 1) and four test sites in British Columbia and Yukon Territory, Canada.

approach. The ANNs make no use of glacier physics so the main necessity of the modeling is that the glaciers have a passing resemblance to glaciers and, in particular, that they have surfaces that slope in their direction of flow as opposed to the flat ice topography assumed for training the networks. The test is mainly to establish that the ANN training assumptions do not completely undercut the possibility of generating reasonable ice thickness estimates.

As a starting point, we select test regions in British Columbia (BC) and the Yukon Territory (YT; Fig. 6) that are mountainous and have been extensively glaciated in the past but currently lack any permanent ice cover. No effort was made to select sites that represent different bed lithologies or climatic settings, but these environmental controls can be presumed to differ owing to the large geographical separation of the test sites. The DEMs for these test sites (Fig. 7) therefore satisfy the  $H(x, y, 0) = 0$  initial condition for the ice dynamics model. To simulate the growth and shrinkage of regional ice cover, we solve the standard shallow-ice equations (e.g., Hutter 1983):

$$\frac{\partial H}{\partial t} = b - \text{div } \mathbf{Q}, \quad (3)$$

$$\mathbf{Q} = -\frac{2A(\rho g)^n |\nabla S|^{n-1}}{n+2} H^{n+2} \nabla S + \mathbf{v}_s H, \quad (4)$$

where  $H = S - B$ ;  $t$  is time;  $b$  is the ice-equivalent mass balance function ( $\text{m s}^{-1}$ );  $\mathbf{Q} = Q_x(x, y, t)\mathbf{i} + Q_y(x, y, t)\mathbf{j}$  is the ice flux ( $\text{m}^2 \text{s}^{-1}$ );  $g = 9.80 \text{ m s}^{-2}$  is the gravity acceleration;  $n = 3$  is the exponent in Glen's flow law for ice;  $A = 3 \times 10^{-24} \text{ Pa}^{-3} \text{ s}^{-1}$  is the flow law coefficient; and  $\mathbf{v}_s$  is the ice sliding rate, which we take as  $\mathbf{v}_s = 0$ . (Including sliding might add realism to the glaciation model but would also open debate about whether the sliding mechanism was correctly represented and, in any case, would be unlikely to affect the outcome of our ANN performance tests.) The mass balance forcing is assumed to be

$$b = \gamma_b(S - Z_{\text{ELA}}), \quad (5)$$

$$Z_{\text{ELA}} = Z_0 + \Delta Z \cos 2\pi t/T_0, \quad (6)$$

where  $\gamma_b = 0.001 \text{ yr}^{-1}$  is the mass balance gradient,  $Z_{\text{ELA}}$  is the equilibrium line altitude (ELA),  $Z_0 + \Delta Z$  is the ELA at  $t = 0$ ,  $\Delta Z$  is the amplitude of the ELA variations, and  $T_0 = 2500 \text{ yr}$  is the periodicity for a sinusoidally oscillating climate. We make no attempt to simulate realistic glacial histories for the test regions—our aim is simply to produce plausible ice topography for a range of glacial and deglacial conditions and use the resulting DEMs and masks to test and optimize the ANNs. The assignments of  $Z_0$  and  $\Delta Z$  differ among test sites, but for every case  $Z_0$  is chosen so that  $Z_{\text{ELA}}(0)$  lies above the highest point on the DEM and  $\Delta Z$  is chosen so that, at  $t = T_0/2$  when  $Z_{\text{ELA}}(t)$  is a minimum, the ELA is low enough to ensure an ice cover that exceeds 60% by area. Thus Eq. (6) leads to cycles of glaciation and deglaciation. Our approach to solving (3)–(6) is similar to that described by Plummer and Phillips (2003) and recently employed by Kessler et al. (2006). We approximate (3) and (4) using finite differences and develop a semi-implicit scheme to solve for  $H_{ij}$ . Intentionally, our choices of  $T_0$ ,  $Z_0$ , and  $\Delta Z$  yield modeled rates of glaciation and deglaciation that are higher than typical worldwide rates; this highlights the differences in neural network performance during phases of glaciation and deglaciation.

#### 4. Testing and optimization

In this section ANN estimators are applied to the test datasets and the results are evaluated. As already noted, four geographically separated test areas were selected and, by using the numerical ice dynamics model to glacialate and deglacialate these regions, a large suite of digital elevation models and ice masks (e.g., Fig. 8) has been generated. It is not worthwhile to consider all these model outputs in our evaluation procedure. Thus for each of the four test sites we selected six modeled ice cover maps having roughly 20%(+), 20%(-), 40%(+),



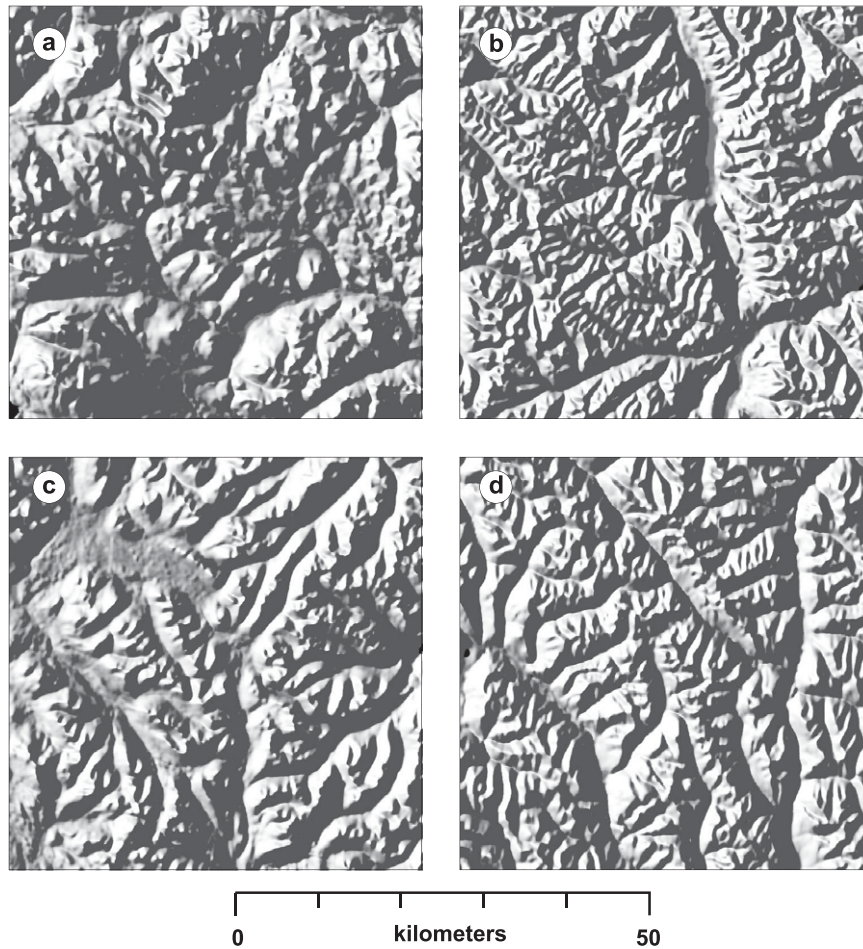


FIG. 7. DEMs for the four test sites. The geographical coordinates refer to map center positions. (a) Site BC-1 (55.4635°N, 124.8151°W). (b) Site BC-2 (58.9000°N, 125.8714°W). (c) Site BC-3 (59.4545°N, 130.3566°W). (d) Site YT-1 (61.6014°N, 133.3684°W).

40%(−), 60%(+), and 60%(−) fractional ice cover by area, which we denote by  $\alpha_I$ . This yielded a test set of 24 models. [The parenthetical signs indicate whether the ice area is increasing (+) or decreasing (−) when the model snapshots were taken. Because the glaciation/deglaciation process is hysteretic, the actual spatial distribution of ice, e.g., for the 20%(+) and 20%(−) cases, can differ markedly.] The objectives are to discover which inputs carry the most useful information and to select a network architecture that performs well under a variety of conditions. By comparing the performance of various network architectures, one can discover where computational effort is rewarded and where it is wasted and determine how many active layers are required to obtain acceptable estimates.

#### a. Special treatment for steep ice

Using ANNs to estimate the thickness of steeply sloping glaciers, such as those that hang from mountain faces

and valley walls, is unlikely to offer the best approach because the bathtub-filling assumption yields a poor representation of the geometry for such glaciers. We therefore use the slope mask to separate glaciers into two classes: gently sloping glaciers ( $\mathbf{I} \wedge \mathbf{G}'$ ), which are amenable to the neural network approach; and steeply sloping glaciers ( $\mathbf{I} \wedge \mathbf{G}$ ), which require some alternative estimation strategy.

Steeply sloping ice is approximated by an inclined slab having local thickness  $H$  and local slope  $|\nabla S| = \tan\theta$ . A long-standing rule of thumb in glaciology (e.g., Paterson 1999) is that, owing to the plasticity of ice, the shear stress at the base of slablike glaciers, taken as

$$\tau_0 = \rho g H \sin \theta, \quad (7)$$

is roughly constant (Nye 1952). If one accepts this idea, then

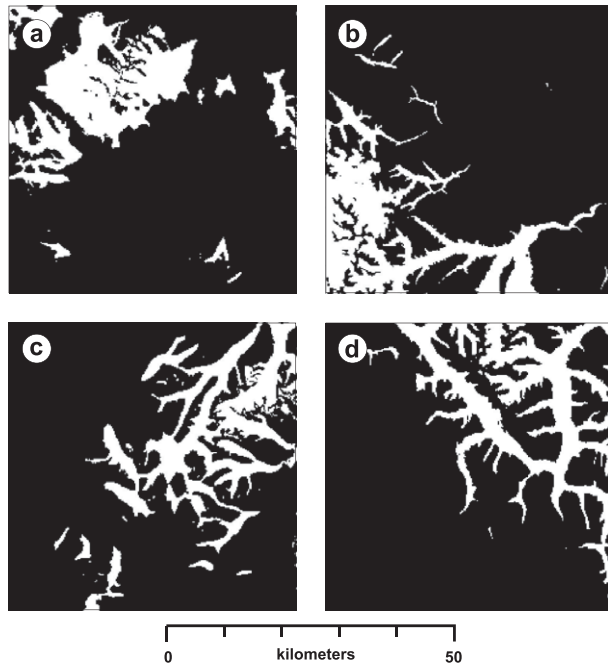


FIG. 8. Ice masks for four test sites. The ice masks represent snapshots during a cycle of glaciation and deglaciation. The area coverage of ice is roughly 20% for each map. The snapshots were taken during the deglacial phase of the cycle. The geographical coordinates refer to map center positions. (a) Site BC-1 (55.46°N, 124.82°W). (b) Site BC-2 (58.90°N, 125.87°W). (c) Site BC-3 (59.45°N, 130.36°W). (d) Site YT-1 (61.60°N, 133.37°W).

$$H(x, y) = \frac{\tau_0}{\rho g \sin \theta(x, y)} \quad (8)$$

can be used to obtain a thickness estimate  $\tilde{H}_{ij} = \tau_0 / \rho g \sin \theta_{ij}$  for the  $\mathbf{I} \wedge \mathbf{G}$  regions of the DEM. The assignment of  $\tau_0$  can be optimized to minimize the thickness estimation error for steep-ice regions of the test models but is expected to be close to  $\tau_0 = 10^5$  Pa.

### b. Optimization and performance analysis

Attention is centered on two aspects of the neural network performance: trainability and prediction error. These performance criteria are used to guide decisions about network architecture, the choice of network inputs, and the size of the training set. As a measure of trainability we take the correlation coefficient  $r$  of the ice thicknesses  $\tilde{H}$  estimated by the neural network versus the target thicknesses  $H^*$  that correspond to the random bathtub-filling levels. Using a tilde to indicate estimated quantities and angular brackets to denote ensemble averages, the bed elevation error estimates are

$$\langle \tilde{B} - B \rangle = \text{mean error in bed elevation estimates, and}$$

TABLE 1. Evaluation of ANN architecture. Preferred architectures are indicated in boldface.

Architecture	Scores		
	Training	Estimation	Combined
<b>6X-12S-1S</b>	<b>4</b>	<b>36</b>	<b>37</b>
<b>6X-12T-1T</b>	<b>9</b>	<b>34</b>	<b>37</b>
6X-12T-1S	0	26	26
6X-12S-1T	12	28	31
6X-12S-1L	1	26	26
6X-12T-1L	0	17	17
6X-12S-12S-1S	42	17	27
<b>6X-12S-12S-1L</b>	<b>65</b>	<b>15</b>	<b>32</b>
6X-12S-12S-1T	36	16	26
6X-12S-12T-1T	44	12	22
6X-12S-12T-1S	13	17	19
<b>6X-12S-12T-1L</b>	<b>56</b>	<b>21</b>	<b>35</b>
6X-12T-12T-1T	39	15	26
6X-12T-12T-1S	19	15	20
6X-12T-12T-1L	33	14	22
6X-12T-12S-1S	18	21	25
6X-12T-12S-1T	33	12	20
<b>6X-12T-12S-1L</b>	<b>56</b>	<b>18</b>	<b>32</b>

$$\langle (\tilde{B} - B)^2 \rangle^{\frac{1}{2}} = \text{root-mean-square error in bed elevation estimates.}$$

The estimated thickness of high-slope ice is denoted  $\tilde{H}_{\text{IAG}}$  and given by

$$\tilde{H}_{\text{IAG}} = \frac{\tau_0}{\rho g \sin \theta}, \quad (9)$$

and  $\tilde{H}_{\text{IAG}'}$  is the thickness of low-slope ice as estimated by the neural network.

### 1) NETWORK ARCHITECTURE

By systematically varying the number of active layers and the transfer functions between layers, we compared the performance of a variety of network architectures applied to each of the 24 models in the test dataset. To limit the possibilities, we assumed that the ANN inputs were obtained from the six range distance values  $R_m$  obtained by applying a one-tier 6-sector stencil (see Fig. 3a for an illustration of a two-tier 6-sector stencil). By using each range value twice, we applied 12 inputs to the first layer of the neural network. For brevity we employ a notational shorthand such that, for example, 6X-12S-12T-1L denotes a network with 6 data inputs (6X), a single output, and three active layers where the first active layer has 12 artificial neurons each with a sigmoidal transfer function (12S), etc. (See Fig. 4 for additional clarification of this notation.) Scoring systems were devised to quantify the trainability, estimation skill, and overall performance of the various architectures, and these results are summarized in Table 1.

TABLE 2. Evaluation of supplementary ANN inputs. Changes in training performance and depth estimation accuracy are indicated by + (improvement), 0 (negligible effect), and – (degraded).

Input	Training	Estimation
Distance to valley walls (tier 2)	+	0
Slope magnitude at valley walls	+	0
Ice surface elevation at site	+	–
Orientation of ice surface slope	+	–

Evaluation of network architecture revealed that ANNs having three active layers tended to achieve better training performance (higher  $r$  values) than two-layer networks but that this did not result in improved estimates of bed elevation. In fact the two-layer networks 6X-12S-1S, 6X-12T-1T, and 6X-12S-1T tended to match or surpass all three-layer networks. Based on our comparisons of network architecture we conclude that, for the present application, three-layer ANNs do not outperform two-layer ANNs, and, furthermore, they require substantial additional computer time for network training. The best-performing three-layer networks were 6X-12S-12T-1L and 6X-12T-12S-1S and these architectures were kept under scrutiny when additional tests were conducted. Collectively, these five networks will be referred to as the “preferred architectures” and appear in boldface in Table 1.

## 2) CHOICE OF TYPE AND NUMBER OF INPUTS

In addition to the six range distances used as input for the architecture tests, we examined the effects of including other kinds of geometric information. The results of this analysis are summarized in Table 2 and discussed below. The following additional inputs were examined: (i) range distance derived from a two-tier range stencil (as in Fig. 3a), (ii) surface slope magnitude calculated at each of the range points  $R_m$  of a one-tier range stencil, (iii) surface elevation at the map point  $P$  where ice thickness is being estimated, and (iv) the direction of local surface slope at the point  $P$ . The motivation underlying (i) and (ii) is that valley-wall slope might convey useful information about ice thickness. The motivation underlying (iii) and (iv) is that there could be systematic variations in topographic geometry associated with the elevation and aspect of the point  $P$ . (The bathtub fillings used to train the ANN assume a horizontal ice surface and therefore carry no information about slope orientation. In this situation, the orientation of the calculated bed slope  $\nabla B$  at the point  $P$  is substituted for that of  $\nabla S$ .)

Adding a second tier of range information or including slope information from the range points of a single-tier stencil improved the training performance

for all five of the preferred network architectures but the performance improvement was not substantial. The root-mean-squared error (rmse)  $\langle (\hat{B} - B)^2 \rangle^{1/2}$  actually increased for three of the five architectures tested, leading us to conclude that adding information about valley-wall slope was unjustified. Including information on surface elevation also resulted in improved training performance but considerably worsened the estimates of bed elevation, both in terms of the mean estimation error  $\hat{B} - B$  and the rmse. Finally, including information on the orientation of local surface slope had a negligible influence on the training performance but tended to result in poorer estimates of bed surface elevation.

Accepting the conclusion that range distance inputs provide the most useful geometric information for the neural networks, we turn to the question of deciding on the number of range sectors and the number of inputs to the neural network. Thus far we have assumed that the number of inputs to the first active layer of the neural network is twice the number of input data values. The merit of this assumption is that a single input can be used in more than one way by the neural network. Fixing the number of range sectors at  $M = 6$ , we examined the effect of reducing the number of inputs from 12 to 9 and 6. Interestingly, reducing the number of inputs to 6 has almost no influence on the training performance or the estimation error. Clearly, there is no argument for increasing the number of inputs beyond the number of actual data values. We also tested the effect of varying the number of range sectors from  $M = 6$  to  $M = 8, 9, 10$ , and 12 and found that increasing the number of range sectors leads to a progressive increase in training performance but only a modest improvement in estimation error. We judge that beyond  $M = 8$  the error reduction becomes small and inconsistent and therefore settle on  $M = 8$  as the practical optimum.

## 3) SIZE OF TRAINING SET

To this point we have assumed a training set of fixed size ( $N_{\text{set}} = 36\,100$ ) but we have yet to establish whether similar performance could be obtained with less training (and hence less computing effort). We tested the effect of taking  $N_{\text{set}} = 1805, 3610, 9025, 18\,050$ , and 36 100 and found that the training performance improved with increasing values of  $N_{\text{set}}$  but that the estimation error ceased to decrease significantly when  $N_{\text{set}} > 18\,050$ . Thus for subsequent applications of the neural network we take  $N_{\text{set}} = 18\,050$ .

## 4) SUMMARY

Our preferred two-layer architectures are 8X-8S-1S, 8X-8T-1T, and 8X-8S-1T, and their performance measures

TABLE 3. Averaged network performance results.

Performance parameters	Architecture		
	SS	TT	ST
Mean training correlation coefficient, $r$	0.764	0.754	0.759
Bed elevation rmse (m), $\langle(\tilde{B} - B)^2\rangle^{1/2}$	70.39	70.47	73.28
Mean of bed elevation error (m), $\langle\tilde{B} - B\rangle$	-4.06	2.08	-6.55
Number of runs used	21	24	22
Number of runs rejected	3	0	2
Total runs	24	24	24

are summarized in Table 3. Henceforth all neural networks will have  $N = 8$  inputs, so we simplify the architecture notation to 8S-1S, 8T-1T, and 8S-1T. Note that all three of the preferred two-layer architectures yield similar performance but that, by a very slight margin, the 8S-1S architecture seems to combine the best trainability with the smallest rmse if “training blunders,” which were characterized by very low values of the correlation coefficient  $r$  (as low as  $r = 0$ ), are discarded. During the course of these tests we found that the 8S-1S and 8S-1T architectures experienced training blunders that occurred when the error minimization routine became trapped in local minima and failed to converge to the true minimum error. This is a common problem of nonlinear optimization that could be remedied using standard approaches but, instead, we chose to switch to a different ANN architecture. The thickness estimation errors associated with these unsuccessfully trained networks were very large. Fortunately, training blunders are easy to identify and when this situation arises we simply discard the results obtained for the 8S-1S architecture and retrain for the 8T-1T architecture. We have not encountered a situation where all three architectures were simultaneously subject to training blunders. To ensure that well-trained networks are always used, we take  $r = 0.7$  as the acceptability threshold for training.

### c. Application to ice thickness and ice volume estimation

Having decided on the architectures and inputs of the neural networks, we now demonstrate their application to the problem of estimating ice thickness and volume for the four test sites at six stages of a glacial cycle [area fraction of ice cover  $\alpha_1 = 20\%(\pm)$ ,  $40\%(\pm)$ ,  $60\%(\pm)$ ]. Table 4 summarizes the results of this effort. Bed topography  $\tilde{B}$  and ice volumes  $\tilde{V}_I$ , etc., estimated using ANNs are compared with bed topography  $B$  and ice volumes  $V_I$  calculated from modeling output. As indicated in column three, the 8S-1S architecture was used for the ANN estimates except when a training problem occurred and the 8T-1T architecture was adopted. The

rmse and mean errors of the bed elevation estimates are given in columns 4 and 5. The middle columns give information on how the total ice volumes, estimated by the ANN and calculated from the ice dynamics model, are divided between steep-gradient ice ( $\mathbf{I} \wedge \mathbf{G}$ ) and low-gradient ice ( $\mathbf{I} \wedge \mathbf{G}'$ ) with  $\theta = 25^\circ$  taken as the threshold that distinguishes these classes. For these cases, ice volume estimates are obtained by coarse numerical integration of  $\tilde{H}$ :

$$\tilde{V}_{I \wedge G} = \int_{A_I \wedge G} \tilde{H}_{I \wedge G} dx dy, \quad (10)$$

$$\tilde{V}_{I \wedge G'} = \int_{A_I \wedge G'} \tilde{H}_{I \wedge G'} dx dy, \quad (11)$$

with the total ice volume estimate being  $\tilde{V}_I = \tilde{V}_{I \wedge G} + \tilde{V}_{I \wedge G'}$ . The three rightmost columns give the estimated and actual ice volumes and the fractional error in these estimates, expressed as a percentage.

It is immediately apparent that the network performance is related to the amount of ice cover and to whether the ice cover is increasing or decreasing with time. The rmse in  $\tilde{B}$  increases with increasing  $\alpha_1$  but there is no clear relationship between mean error and  $\alpha_1$ . We attribute the tendency for rmse to increase as  $\alpha_1$  increases to a partial breakdown of the geomorphic premise. As more of the landscape becomes hidden by ice cover, the average geometrical properties of the exposed landscape and those of the buried landscape progressively diverge, leading to reduced skill for the ANN. The fractional error in ice volume (rightmost column), which has relevance to sea level predictions, is very strongly linked to whether ice cover is increasing or decreasing with time. The fractional error tends to be large for the case of increasing ice cover and surprisingly small (45% is the worst case) when ice cover is decreasing. Noting, in the  $V_I$  column of Table 4, that for a given fractional coverage of ice  $\alpha_1$  the ice volume during deglaciation (20–, 40–, 60–) is systematically much larger than the ice volume during glaciation (20+, 40+, 60+) for each of the test sites, we conclude that during glaciation the ice configuration resembles a snowfield, whereas during deglaciation the ice configuration is more glacierlike. We speculate that the geomorphic premise is better matched to the situation of glacier recession than to glacier advance; the geometric form of snowfields is not closely related to glacial processes that shape glaciated landscapes so their configuration carries less information about subglacial topography.

Globally, glaciers are in retreat and thus performance measures for the 20–, 40–, and 60– cases are more relevant to the present situation than those for 20+, 40+, and 60+. Generally, the high-slope volume estimates



TABLE 4. Neural network performance for test sites. Values based on ANN estimates (indicated by a tilde, e.g.,  $\tilde{B}$ ) are compared with those calculated from numerical ice dynamics simulations. The default architecture for the ANNs was 8S-1S but the 8T-1T architecture was used in cases when an 8S-1S network experienced a training blunder. Here  $\alpha_I$  is the area fraction of ice cover,  $\langle(\tilde{B} - B)^2\rangle^{1/2}$  is the rmse of the bed elevation estimate,  $\langle\tilde{B} - B\rangle$  is the mean error of the bed elevation estimate,  $\tilde{V}_{IAG}$  is the estimated volume of steeply sloping ice,  $V_{IAG}$  is the modeled volume of steeply sloping ice,  $\tilde{V}_{IAG'}$  is the estimated volume of gently sloping ice,  $V_{IAG'}$  is the modeled volume of gently sloping ice,  $\tilde{V}_I$  is the estimated total ice volume,  $V_I$  is the modeled total ice volume, and  $(\tilde{V}_I - V_I)/V_I$  is the relative error of the estimated ice volume.

$\alpha_I$ (%)	Site	Net	$\langle(\tilde{B} - B)^2\rangle^{1/2}$ (m)	$\langle\tilde{B} - B\rangle$ (m)	$\tilde{V}_{IAG}$ (km <sup>3</sup> )	$V_{IAG}$ (km <sup>3</sup> )	$\tilde{V}_{IAG'}$ (km <sup>3</sup> )	$V_{IAG'}$ (km <sup>3</sup> )	$\tilde{V}_I$ (km <sup>3</sup> )	$V_I$ (km <sup>3</sup> )	$(\tilde{V}_I - V_I)/V_I$ (%)
60−	BC-1	8S-1S	74.6	7.6	0.23	0.18	114.70	121.09	114.94	121.27	−5%
60−	BC-2	8S-1S	112.5	28.6	1.29	0.87	130.55	151.07	131.84	151.94	13%
60−	BC-3	8T-1T	118.0	68.1	3.24	2.10	63.35	119.80	66.59	121.91	−45%
60−	YT-1	8S-1S	107.6	37.9	1.10	0.79	134.37	164.02	135.47	164.81	−18%
40−	BC-1	8T-1T	62.8	11.7	0.17	0.09	58.80	65.49	58.97	65.58	−1%
40−	BC-2	8S-1S	85.1	21.1	0.66	0.42	73.79	83.39	74.44	83.81	11%
40−	BC-3	8T-1T	88.6	43.6	1.53	0.84	49.06	75.49	50.58	76.33	−34%
40−	YT-1	8S-1S	80.9	32.5	0.62	0.52	86.29	103.46	86.92	103.98	−16%
20−	BC-1	8S-1S	51.1	1.1	0.09	0.04	27.42	27.79	27.51	27.84	−1%
20−	BC-2	8S-1S	77.4	−3.7	0.32	0.19	26.76	26.11	27.08	26.30	3%
20−	BC-3	8S-1S	59.8	3.8	0.48	0.24	30.11	31.33	30.59	31.57	−3%
20−	YT-1	8S-1S	66.5	−7.2	0.27	0.21	43.22	41.32	43.49	41.53	5%
20+	BC-1	8S-1S	40.7	−24.6	1.54	1.26	10.10	3.46	11.64	4.72	147%
20+	BC-2	8T-1T	32.6	−13.1	7.37	3.75	0.75	0.33	8.11	4.08	99%
20+	BC-3	8T-1T	39.5	−14.6	7.19	4.19	2.24	0.94	9.43	5.13	84%
20+	YT-1	8S-1S	45.2	−33.3	5.81	2.58	6.72	1.32	12.53	3.89	222%
40+	BC-1	8S-1S	69.1	−44.4	1.34	1.32	40.96	14.48	42.30	15.80	168%
40+	BC-2	8S-1S	58.6	−15.6	13.14	10.80	7.07	4.89	20.20	15.69	29%
40+	BC-3	8T-1T	70.5	−5.5	10.76	11.27	12.97	11.36	23.73	22.62	5%
40+	YT-1	8S-1S	55.5	−26.4	9.88	7.70	14.64	7.13	24.52	14.83	65%
60+	BC-1	8S-1S	85.7	−44.5	1.04	1.18	90.60	51.94	91.64	53.17	72%
60+	BC-2	8S-1S	62.1	4.3	15.28	15.67	24.65	26.16	39.93	41.85	−5%
60+	BC-3	8S-1S	72.2	5.9	10.32	13.33	40.80	44.44	51.13	57.77	−11%
60+	YT-1	8S-1S	62.3	−17.0	10.98	12.20	40.64	31.90	51.61	44.10	17%

[based on Eq. (8)] are superior to the low-slope estimates using neural networks, but this does not necessarily imply that Eq. (8) offers a useful approach to estimating the low-slope thicknesses and volumes.

Figure 9 shows histograms of the elevation estimation error  $\varepsilon_B = \tilde{B} - B$  for all four test sites and a deglaciating ice cover with  $\alpha_I = 20\%$ ,  $40\%$ , and  $60\%$ . The error distributions tend to be roughly symmetric about  $\varepsilon_B = 0$ , which accounts for the surprisingly good estimates of ice volume, and, as noted above, the rmse increases with increasing ice cover.

Figure 10 shows maps of the estimated ice thickness (left panels) and thickness estimation error  $\tilde{H} - H$  (right panels) for the four test sites and  $60\%$  (−) ice cover. (Note that the colored regions are ice covered and the white regions are ice denuded.) The ice thickness map for site BC-1 (Fig. 10a) manifests an interesting pathology not shared by the other examples—a fanlike patterning of the predicted ice thickness. This is related to the fact that the areal distribution of ice for site BC-1 closely resembles a continuous ice field rather than an arterial system of flowing glaciers (as for the remaining sites). Within this ice mass, many points lie at

distances that exceed the assumed 6-km range limit of the computational stencil. Thus many of the inputs to the neural network will be set to  $R_m = 6$  km and the ANN will lose its effectiveness. For the three other sites the predicted ice thicknesses seem highly plausible: ice thickness lies in the 0–400-m range, typical for glaciers of modest size, and thickness varies smoothly with distance, tending to be maximum along the central axes of valley channels. Close examination of the thickness estimation errors (right panels), draws attention to the fact that although the estimated thickness patterns are plausible they are not necessarily correct. As also suggested by the histograms (Fig. 9), there can be large errors in the ice thickness estimates.

## 5. Concluding remarks

This study is thought to represent the first attempt to apply neural networks to the problem of ice thickness estimation. As a first attempt, the results are encouraging and suggest that additional development is warranted, ideally in tandem with estimation strategies that are rooted in glacier physics. Of particular interest

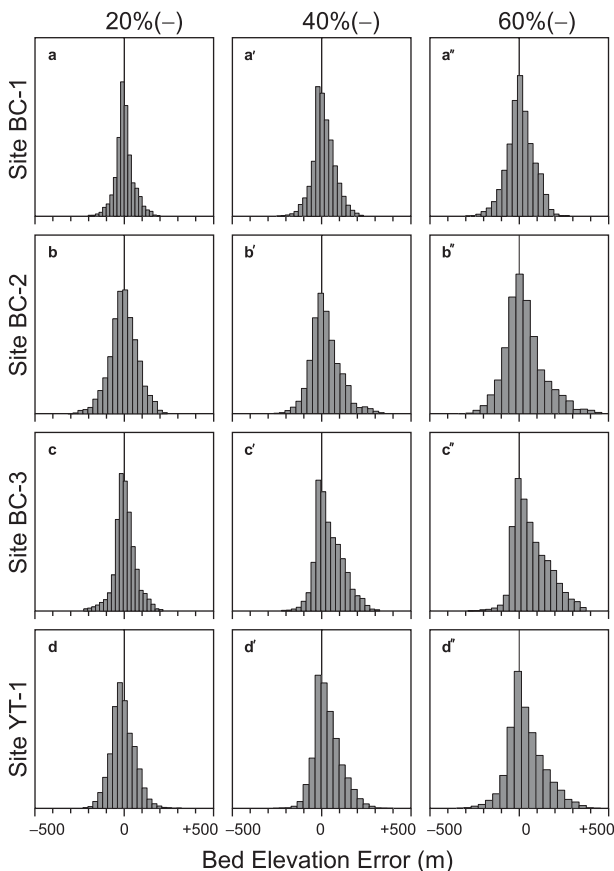


FIG. 9. Histograms of the estimation error of bed elevation for the numerically modeled glaciation of four test sites. The column headers 20%(-), 40%(-), 60%(-) indicate the approximate areal percentage of ice cover; the negative sign indicates that the ice cover was decreasing at the time of the model snapshot. (a) Site BC-1. (b) Site BC-2. (c) Site BC-3. (d) Site YT-1.

is the indication that, despite unavoidable errors in the ice thickness estimates, the resulting errors in estimated ice volume are surprisingly small. Thus neural networks can be used to estimate the ice volume of earth's mountain glaciers and yield estimates that are completely independent of those obtained by conventional volume–area scaling analysis. Our use of ice masks underscores the importance of collecting accurate information on glacier outlines, one of the objectives of the Glacier Land Ice Measurements from Space (GLIMS) program (additional information is available online at <http://www.glims.org/>).

The neural network estimators perform best during deglacial phases (like the present one) when the fractional area of ice coverage is not large and when the ice masses are characterized by distinct glacierlike elements rather than aggregated into extensive ice fields. One of our aims was to predict subglacial topography in order to provide a realistic boundary condition for nu-

merical ice dynamics models. In this we can claim a qualified success: the estimates of subglacial topography yield plausible, though not necessarily accurate, estimates of the bed surface. As yet, there are no good ways for estimating subglacial topography that do not involve drilling or geophysical measurements, but neural networks can at least provide better results than simply ignoring the existence of the ice cover (i.e.,  $\tilde{H} = 0$ ). As an approach to estimating glacier ice volume during deglacial phases, the neural networks seem to yield very good estimates of ice volume that are possibly superior to those obtained using conventional volume–area scaling approaches [as in Eq. (1)]. Shortcomings of the ANN approach are that it is computationally intensive and that to apply the method over very large regions, for example, Alaska–Yukon or the Himalayas, requires repeated application of the method over a patchwork of overlapping approximately 50 km  $\times$  50 km subregions.

Worthwhile directions for future study would be to investigate the potential of physics-based methods and to compare ANN and volume–area scaling estimates of ice volume for glaciers of known thickness as well as for numerically simulated glaciers with appropriate mass balance forcings and proper model physics. Our numerical experiments with increasing and decreasing degrees of ice coverage could have ominous implications for the volume–area scaling approach because we find that glaciers having identical areas can have greatly differing volumes, depending on whether glacier area is increasing or decreasing with time. Thus the volume–area scaling is likely to work best when glaciers are near a steady state, an explicit assumption of Bahr's scaling analysis [Bahr et al. 1997, their Eq. (3a)].

Estimating the thickness of ice caps and ice sheets appears to be a distinct problem that will require different approaches from the ones considered here. These ice masses have been favored targets for geophysical study and much is already known of their subglacial topography, so the situation is not hopeless.

**Acknowledgments.** We thank two anonymous reviewers for extremely helpful suggestions and Erik K. Schieffer and Gabriel J. Wolken for assistance with the DEMs and ice masks. Financial support was provided by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) and the Natural Sciences and Engineering Research Council of Canada. EB acknowledges postdoctoral support from the Marie Curie Outgoing International Fellowship program of the European Union. This paper is a contribution to the Polar Climate Stability Network and to the Western Canadian Cryospheric Network, both of which are funded by CFCAS and consortia of Canadian universities.

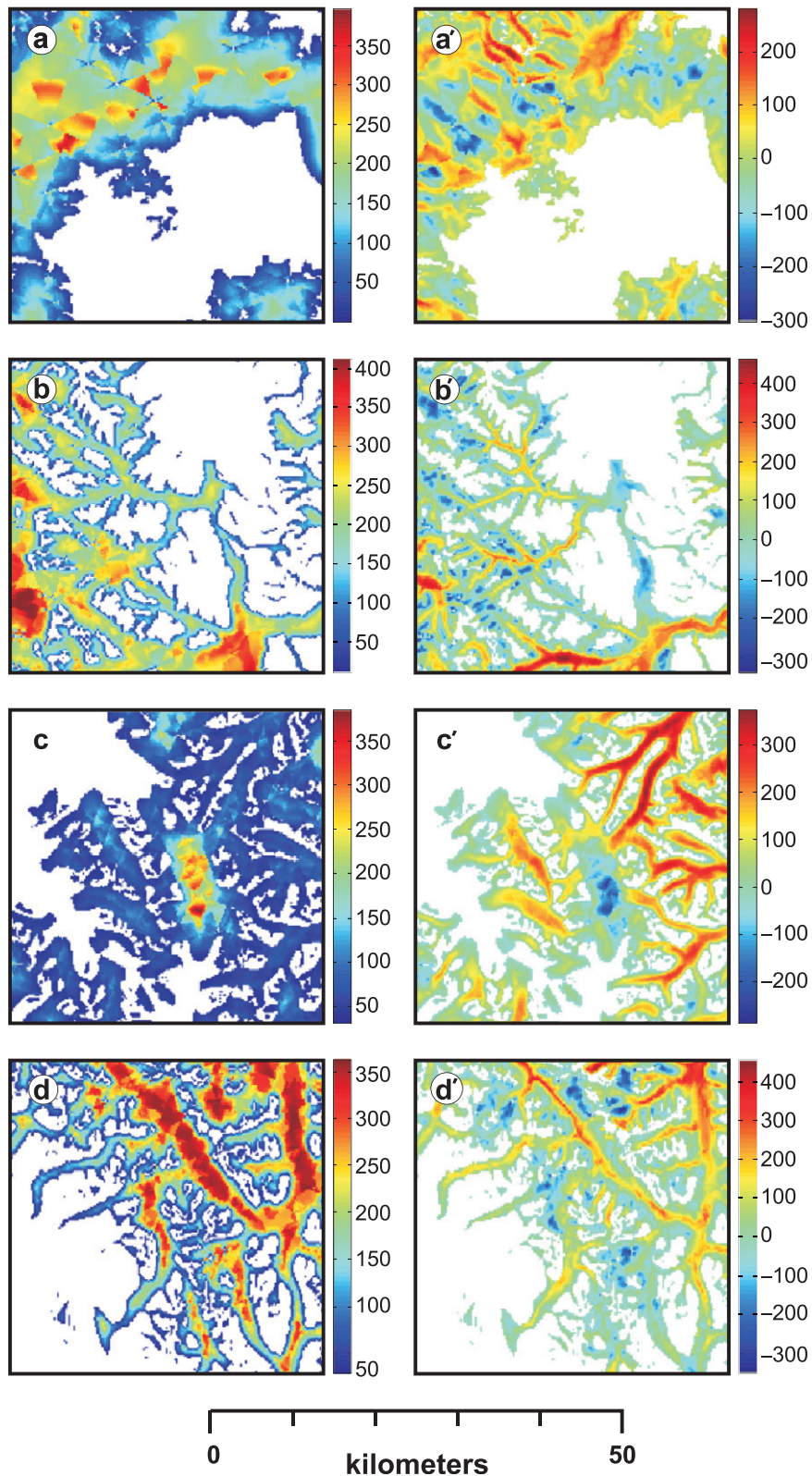


FIG. 10. Maps of (left) estimated ice thickness and (right) ice thickness estimation error for test sites. Scales for the color bars indicate thickness in meters. For these maps the areal coverage of ice was roughly 60% with ice cover decreasing with time. These maps match to the rightmost histograms in Fig. 9. (a),(a') Site BC-1. (b),(b') Site BC-2. (c),(c') Site BC-3. (d),(d') Site YT-1.

## REFERENCES

- Bahr, D. B., 1997: Global distributions of glacier properties: A stochastic scaling paradigm. *Water Resour. Res.*, **33**, 1669–1679.
- , M. F. Meier, and S. D. Peckham, 1997: The physical basis of glacier volume-area scaling. *J. Geophys. Res.*, **102**, 20 355–20 362.
- Bishop, C. M., 1995: *Neural Networks for Pattern Recognition*. Oxford University Press, 482 pp.
- Björnsson, H., 1986: Surface and bedrock topography of ice caps in Iceland mapped by radio echo soundings. *Ann. Glaciol.*, **8**, 11–18.
- Cavazos, T., 2000: Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. *J. Climate*, **13**, 1718–1732.
- Chen, J., and A. Ohmura, 1990: Estimation of Alpine glacier water resources and their change since 1870s. Hydrology in Mountainous Regions I, IAHS Publication 193, 127–135.
- Demuth, H., M. Beale, and M. Hagan, 2006: Neural network toolbox user's guide. The MathWorks, Inc., 452 pp.
- Driedger, C., and P. Kennard, 1986: Glacier volume estimation on Cascade volcanoes—An analysis and comparison with other methods. *Ann. Glaciol.*, **8**, 59–64.
- Haeberli, W., 1985: Global land-ice monitoring: Present state and future perspectives. *Glaciers, Ice Sheets, and Sea Level: Effect of a CO<sub>2</sub>-Induced Climatic Change*, National Academy Press, 216–231.
- , and M. Hoelzle, 1995: Application of inventory data for estimating characteristics of and regional climate-change effects on mountain glaciers: A pilot study with the European Alps. *Ann. Glaciol.*, **21**, 206–212.
- Hewitson, B. C., and R. G. Crane, 2002: Self-organizing maps: Application to synoptic climatology. *Climate Res.*, **22**, 13–26.
- Hsieh, W. W., 2001: Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *J. Climate*, **14**, 2528–2539.
- Huss, M., D. Farinotti, A. Bauder, and M. Funk, 2008: Modelling runoff from highly glacierized alpine drainage basins in a changing climate. *Hydrol. Proc.*, **22**, 3888–3902.
- Hutter, K., 1983: *Theoretical Glaciology*. D. Reidel Publishing Company, 510 pp.
- Kessler, M. A., R. S. Anderson, and G. M. Stock, 2006: Modeling topographic and climatic control of east-west asymmetry in Sierra Nevada glacier length during the Last Glacial Maximum. *J. Geophys. Res.*, **111**, F02002, doi:10.1029/2005JF000365.
- Kohonen, T., 2001: *Self-Organizing Maps*. 3rd ed. Springer-Verlag, 501 pp.
- Lemke, P., and Coauthors, 2007: Observations: Changes in snow, ice and frozen ground. *Climate Change 2007: The Physical Basis*, S. Solomon et al., Eds., Cambridge University Press, 337–383.
- Levenberg, K., 1944: A method for the solution of certain nonlinear problems in least squares. *Quart Appl. Math.*, **2**, 164–168.
- Marquardt, D. W., 1963: An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.
- Meier, M. F., M. B. Dyurgerov, U. K. Rick, S. O'Neel, W. T. Pfeffer, R. S. Anderson, S. P. Anderson, and A. F. Glazovsky, 2007: Glaciers dominate eustatic sea-level rise in the 21st century. *Science*, **317**, 1064–1067.
- Monahan, A. H., 2000: Nonlinear principal component analysis of neural networks: Theory and application to the Lorenz system. *J. Climate*, **13**, 821–835.
- Nye, J. F., 1952: A comparison between the theoretical and measured long profile of the Unteraar Glacier. *J. Glaciol.*, **2**, 103–107.
- Ohmura, A., 2004: Cryosphere during the twentieth century. *The State of the Planet: Frontiers and Challenges in Geophysics, Geophys. Monogr.*, Vol. 150, Amer. Geophys. Union, 239–257.
- Paterson, W. S. B., 1999: *The Physics of Glaciers*. 3rd ed. Butterworth-Heinemann, 496 pp.
- Plummer, M. A., and F. M. Phillips, 2003: A 2-D numerical model of snow/ice energy balance and ice flow for paleoclimatic interpretation of glacial geomorphic features. *Quat. Sci. Rev.*, **22**, 1389–1406.
- Reed, R. D., and R. J. Marks II, 1999: *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. MIT Press, 358 pp.
- Reusch, D. B., and R. B. Alley, 2007: Antarctic sea ice: A self-organizing map-based perspective. *Ann. Glaciol.*, **46**, 391–396.
- , B. C. Hewitson, and R. B. Alley, 2005: Towards ice-core-based synoptic reconstructions of West Antarctic climate with artificial neural networks. *Int. J. Climatol.*, **25**, 581–610.
- Steiner, D., A. Walter, and H. J. Zumbühl, 2005: The application of a non-linear back-propagation neural network to study the mass balance of Grosse Aletschgletscher, Switzerland. *J. Glaciol.*, **51**, 313–323.
- , A. Pauling, S. U. Nussbaumer, A. Nesje, J. Luterbacher, H. Wanner, and H. J. Zumbühl, 2008: Sensitivity of European glaciers to precipitation and temperature—Two case studies. *Climatic Change*, **90**, 413–441, doi:10.1007/s10584-008-9393-1.
- Tang, Y., and W. W. Hsieh, 2001: Coupling neural networks to incomplete dynamical systems via variational data assimilation. *Mon. Wea. Rev.*, **129**, 818–834.
- Tangang, F. T., W. W. Hsieh, and B. Tang, 1998a: Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J. Geophys. Res.*, **103**, 7511–7522.
- , B. Tang, A. H. Monahan, and W. W. Hsieh, 1998b: Forecasting ENSO events: A neural network-extended EOF approach. *J. Climate*, **11**, 29–41.
- Zumbühl, H. J., D. Steiner, and S. U. Nussbaumer, 2008: 19th century glacier representations and fluctuations in the central and western European Alps: An interdisciplinary approach. *Global Planet. Change*, **60**, 42–57.