We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Using Artificial Neural Networks to Produce High-Resolution Soil Property Maps

Zhengyong Zhao, Fan-Rui Meng, Qi Yang and
Hangyong Zhu

Additional information is available at the end of the chapter

Abstract

High-resolution maps of soil property are considered as the most important inputs for decision support and policy-making in agriculture, forestry, flood control, and environmental protection. Commonly, soil properties are mainly obtained from field surveys. Field soil surveys are generally time-consuming and expensive, with a limitation of application throughout a large area. As such, high-resolution soil property maps are only available for small areas, very often, being obtained for research purposes. In the chapter, artificial neural network (ANN) models were introduced to produce high-resolution maps of soil property. It was found that ANNs can be used to predict high-resolution soil texture, soil drainage classes, and soil organic content across landscape with reasonable accuracy and low cost. Expanding applications of the ANNs were also presented.

Keywords: ANN, soil drainage, soil texture, soil organic carbon, DEM, topography, hydrological index, vertical slope position

1. Introduction

1.1. Soil properties

Difference of soils in physical and chemical determined what type of plants grows in a soil or what particular crops grow in a region. Jack pine (*Pinus banksiana*), for example, occurs on coarse sands, poor drainage, and shallow soils, and sugar maple (*Acer saccharum*) grows best on deep, fertile, moist, well-drained soils in Ref. [1]. The most important soil properties included soil texture, soil drainage, and soil organic carbon (SOC).

Soil texture is defined as relative proportions (percentages) of clay, sand, and silt contents. These percentages are used to confirm soil textural classes in a soil texture triangle (**Figure 1**).
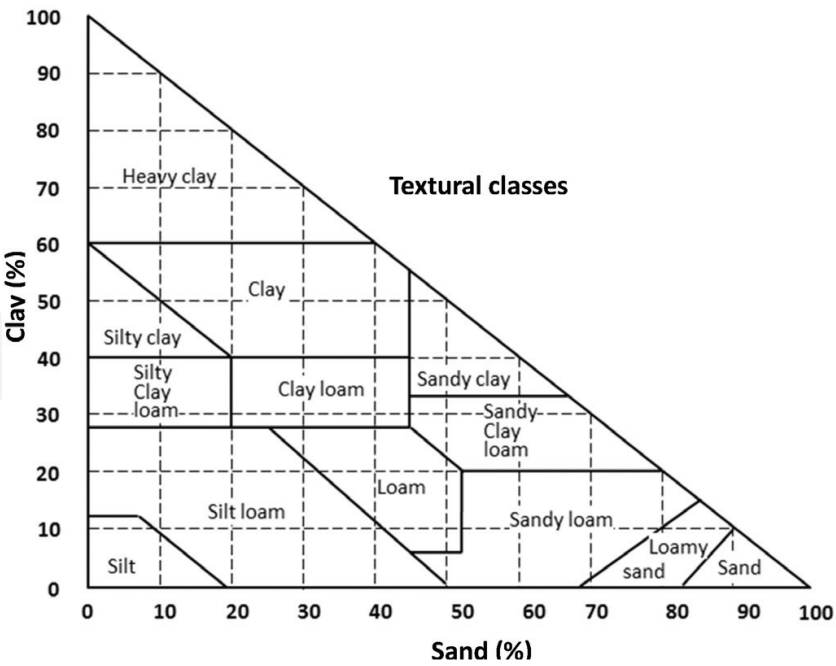
Figure 1. Canadian soil texture triangle in Ref. [2].

Soil texture not only directly affects the porosity of soil, but also determines water-holding and nutrient-holding capacity, flow characteristics, and long-term soil nutrient regime. For example, soils with heavy clay in general have higher percentage of smaller pores, higher water-holding capacity at lower water potentials and are often associated with poorly drained conditions with limited aeration for plant growth. As a contrast, soils with heavy sand normally have relatively higher percentage of larger pores with lower water-holding capacity under relatively dry conditions. Soil texture also affects the risk of soil erosion and soil erodibility.

Soil drainage was defined as the frequency and duration of periods of water saturation or partial saturation, and soil drainage classes reflect average soil moisture conditions in Ref. [3]. Soil drainage is associated with water-holding and nutrient-holding capacities, flow characteristics, and solute transport. Soil drainage is also directly related to plant growth. For example, plants grown on soil with poor drainage often suffer from reduced growth, leaf dieback as a result of root suffocation, and root disease in Ref. [4]. Plants experiencing root decline from excess water are also more susceptible to attack by secondary diseases and insects in Ref. [5]. Under natural conditions, soil drainage characteristic is one important factor that determines which types of plants grow on a particular landscape site. For precision forestry and precision agriculture, high-resolution soil maps are especially important in Ref. [6]. Soil drainage classes are closely related soil texture and slope position (**Figure 2**).

Soil organic carbon refers to the carbon (C) occurring in soil organic matter of the soil. SOC can help to improve soil physical properties by increasing water-holding capacity, stabilizing soil structure in Ref. [8], soil chemical properties, and nutrients holding capacity in Ref. [9]. From a land management perspective, SOC plays important roles in reducing soil erosion and improving crop productivity. For this reason, SOC content has been used as a required input variable
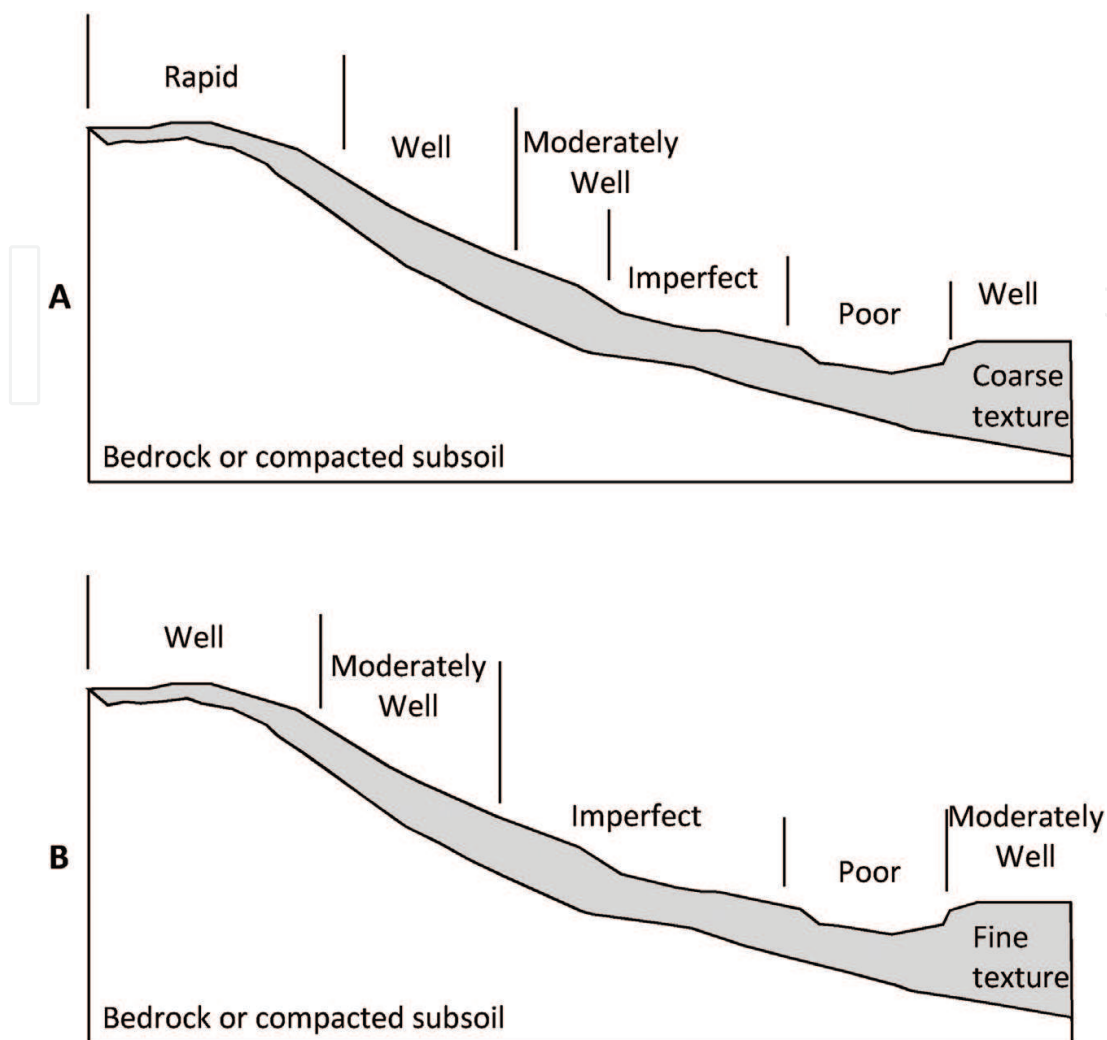
Figure 2. Generalized soil drainage patterns and drainage classes for soils with coarse texture soil (A) and fine texture soil (B) as influenced by slope position in Ref. [7].

for a number of hydrological simulation models in Ref. [10] and many landscape level models for estimating soil water retention, cation exchange capacity, and soil bulk density in Ref. [11].

1.2. Mapping soil properties

Field soil surveys have been the primary method for determination of soil properties, including soil texture, soil drainage, and SOC. For mapping purposes, soil surveys are normally conducted with point samples, either systematically or randomly over a given area, and then the point data are usually interpolated to produce soil maps. Various interpolation methods have been used to produce soil maps, especially the kriging method in Ref. [12]. There is a major limitation about interpolation method, i.e., the assumption that the spatial distributions and changes of the interpolated properties are continuous. Therefore, large amount of data are often required to produce accurate high-resolution soil maps. With the purpose of improving

the interpolation accuracy with sparsely distributed sample points, various improved kriging methods have been developed in Ref. [13]. However, the methods still require substantial amounts of field samples to define the spatial autocorrelation and the precision of the resultant maps will still depend upon the density and distribution of original data points in Ref. [14]. Due to high spatial variability of soil characteristics, large numbers of sampling points are required to generate an accurate high-resolution soil map. Although the accuracy of a soil map may be increased with increasing data points, intensive field surveys are expensive and time-consuming. Furthermore, the accuracy is affected by the quality of the data, which, to a great extent, depends on the field experience of the soil surveyors in Ref. [15]. As an alternative, various models have been developed to produce soil property maps.

Statistical models with predictive powers could potentially overcome the problem of interpolation methods in Ref. [16]. Bell et al., for example, related soil drainage class to parent material, terrain, and surface drainage with the help of discriminant function analysis in Pennsylvania, USA in Ref. [17]. According to this method, soil drainage probability maps were predicted well when compared with published soil drainage maps. Campling et al. applied a logistic model to successfully predict the probability of drainage classes in a tropical area using terrain properties (elevation, slope, distance-to-the-river channel) and vegetation indices from a Landsat TM image in Ref. [18]. By applying discriminant function analysis and a co-kriging method, Kravchenko et al. created soil drainage maps using topographical data, i.e., slope, curvature, and flow accumulation, and soil electrical conductivity data in central Illinois, USA in Ref. [19]. But empirical models derived with traditional statistical methods may hinder the real relationships between soil properties and independent data because the relationships are rarely linear in nature.

1.3. Artificial neural networks

In recent years, artificial neural network (ANNs) have been increasingly used to overcome non-linear problems. The ANN is a form of artificial intelligence that was inspired by the studies of the human neuron and has been used to analyze biophysical data in Ref. [20]. ANNs have the ability to auto-analyze the relationships between multi-source inputs (including combinations of qualitative and quantitative data) by self-learning, and produce results without hypothesis. Some ANNs have been successfully used to map soil properties in Ref. [21]. For example, in Licznar and Nearing's study, soil loss was predicted quantitatively from natural runoff plots with the ANN method in Ref. [22]. The results showed that correlation coefficients (predicted soil loss versus measured values) were in the range of 0.7–0.9. Ramadan et al. applied two different multivariate calibration methods (PCA and back-propagation ANN) to predict soil properties (sand, silt, clay, etc.) with the help of DNA data from microbial community in Ref. [23].

1.4. Objectives

In the chapter, we focused on describing a general approach for using ANNs to produce high-resolution soil properties, from preparing data, building ANN structure, training ANNs, optimizing networks, to simulating ANNs.

2. Data preparation for modeling soil properties

Preparing data, including input data and target data, is an important and indeed a critical step before building ANN for soil properties.

2.1. Input data

Input data were composed of potential variables that describe or determine the predicting soil properties, including DEM-generated topo-hydrological variables, such as slope steepness, soil terrain factor (STF), sediment delivery ratio (SDR), vertical slope position (VSP), topographic witness index (TWI), and potential solar radiation (PSR) (**Figure 3**), and existing coarse resolution soil map, such as soil property map, geology map, surficial parent material map, and hydrologic map, because (1) at local levels, soil properties are assumed to have been modified by hydrological processes that are associated with topography and they can be modeled with a
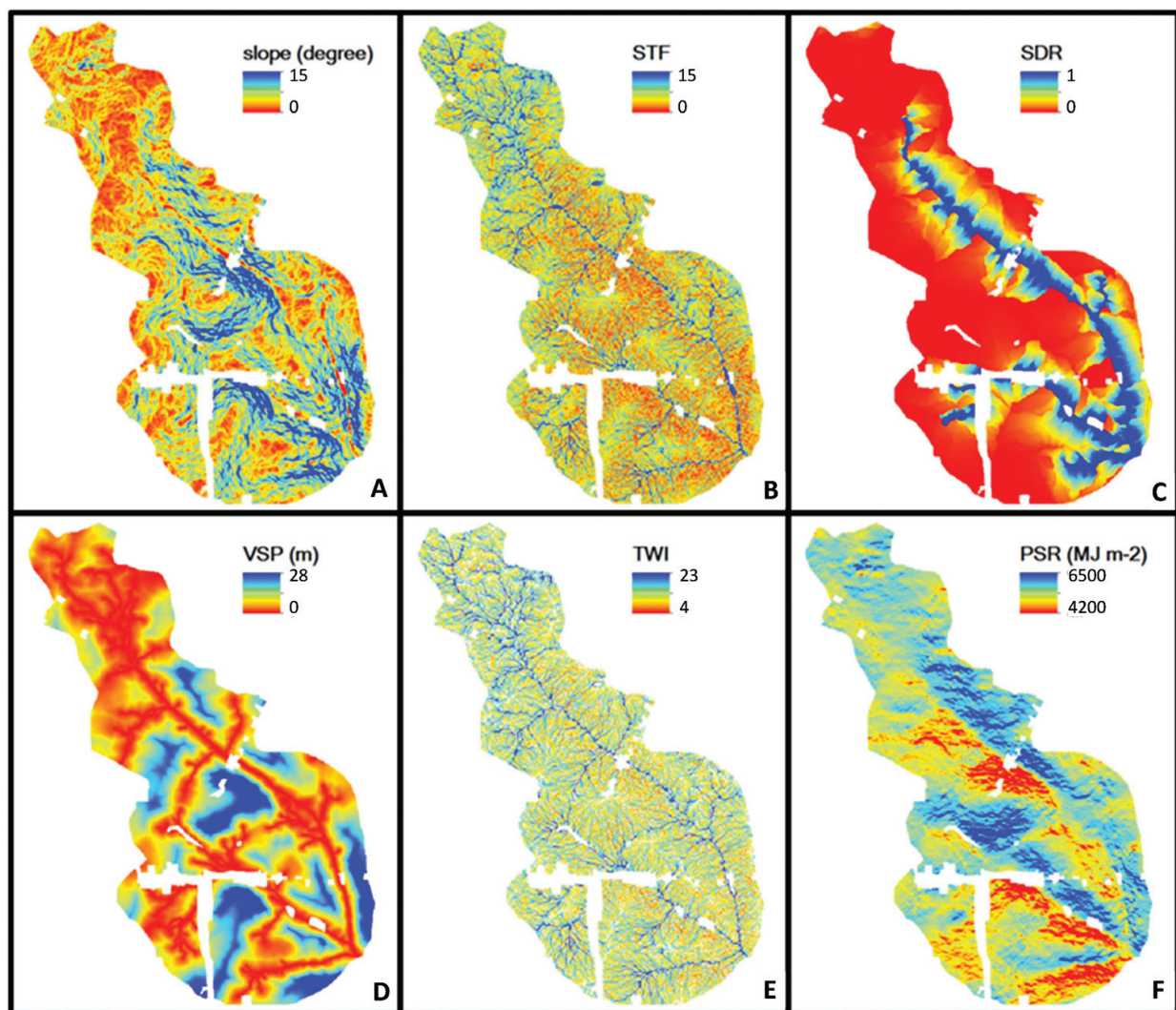
Figure 3. The images of slope (A), soil terrain factor (B), sediment delivery ratio (C), vertical slope position (D), topographic witness index (E), and potential solar radiation (F) in the black brook watershed, New Brunswick, Canada.

DEM in Ref. [24]; (2) at landscape levels, average soil properties were related to geological formations and soil parent materials. These landscape features are assumed to have been captured by existing coarse resolution soil maps.

Soil terrain factor is a modified version of the hydrological similarity index in Ref. [25]. It considers total drainage area and slope as well as the clay content in rooting zone. The STF was calculated using Eq. (1):

$$STF = \ln \frac{(A+1)P_{clay}}{(s+k)^2} \tag{1}$$

where A is the flow accumulation (m^2); P_{clay} is the clay content (wt. %) from the coarse resolution soil data; k is a parameter (=1); and s is the slope steepness (m m^{-1}).

Sediment delivery ratio is the percent of sediment delivered to surface waters from the total amount of soil eroded in a watershed. The ratio, calculated by Eq. (2), indicates the efficiency of sediment transport in the watershed and is largely influenced by topography and the flow distance to streams in Ref. [26].

$$SDR_i = \exp\left(-\beta t_i\right) \tag{2}$$

where t_i is the travel time from cell i to the nearest channel (s); and β is a watershed-specific constant.

Traveling time, t_i, is defined by Eq. (3):

$$t_i = \sum_{j=1}^{N_p} \frac{l_j}{v_j} \tag{3}$$

where N_p is the total number of cells from cell j to the nearest channel, along the flow path (m); l_j is the length segment cell j along the flow path (m); and v is flow velocity (m s^{-1}).

Flow velocity, v, is got based on Eq. (4) in Ref. [27].

$$v = ds^{1/2} \tag{4}$$

where s is slope steepness (m m^{-1}) and d is a coefficient dependent on surface roughness characteristics (m s^{-1}) for cell i.

By using HYDRO-tools extension in ArcView, the flow length, t_i, was calculated in order to acquire travel time, with an inverse velocity grid used as a weighting factor in Ref. [28].

The watershed parameter, β was estimated by numerically solving Eq. (5):

$$SDR_w = \frac{\sum\limits_{i=1}^{N} \exp\left(-\beta t_i\right) l_i^{0.5} s_i^2 a_i}{\sum\limits_{i=1}^{N} l_i^{0.5} s_i^2 a_i} \tag{5}$$

where SDR_w is the watershed average SDR, which was calculated with an empirical formula similar to $SDR_w = pA_T^c$ in Ref. [29]. Parameters p and c were confirmed as 0.42 and -0.125 because they represent a good general approximation between SDR_w and SDR in Ref. [30].

N is total number of cells over the watershed, a_i is area of the cell (m^2), l_i is the length of cell i along the flow path (m), A_T is the area of the watershed (km^2).

Vertical slope position (m) is defined as the elevation differences between the land and the nearest water surface and calculated by integrating the elevation difference for each cell alone the path to the nearest water body using the following Eq. (6) (**Figure 4**):

$$VSP = \min \sum (ds) \tag{6}$$

where d is the distance between two adjacent cells (m); s is slope steepness (m m^{-1}).

Topographic wetness index is a steady-state wetness index that reflects soil moisture and drainage conditions, defined as a function of the natural logarithm of the ratio of local upslope contribution area and slope angle in Ref. [32].

$$TWI = \ln\left(\frac{A}{s}\right) \tag{7}$$

where A is the flow accumulation (m^2) and s is the slope steepness (m m^{-1}).

Potential solar radiation (MJ m^{-2}) is the total of annual potential solar radiation. PSR reflected the potential light distribution along with the change of topography. The higher the value, the stronger the light radiation. Potential solar radiation takes into account the central Latitude, days of 1 year from 1 to 365 and hours of 1 day from 1 to 24 by an ArcView Extension in Ref. [33].

Coarse resolution soil maps are widely available. These maps usually reflected average soil properties over a large area (**Figure 5**). Researches indicated that coarse resolution soil data had a significant influence on the distribution of high-resolution soil property maps, especially around the boundary in Ref. [34].
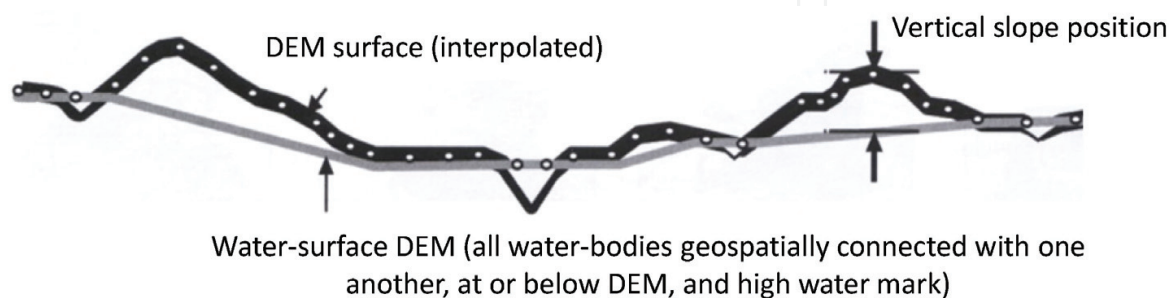
Figure 4. Vertical slope position of a slope profile in ref. [31].
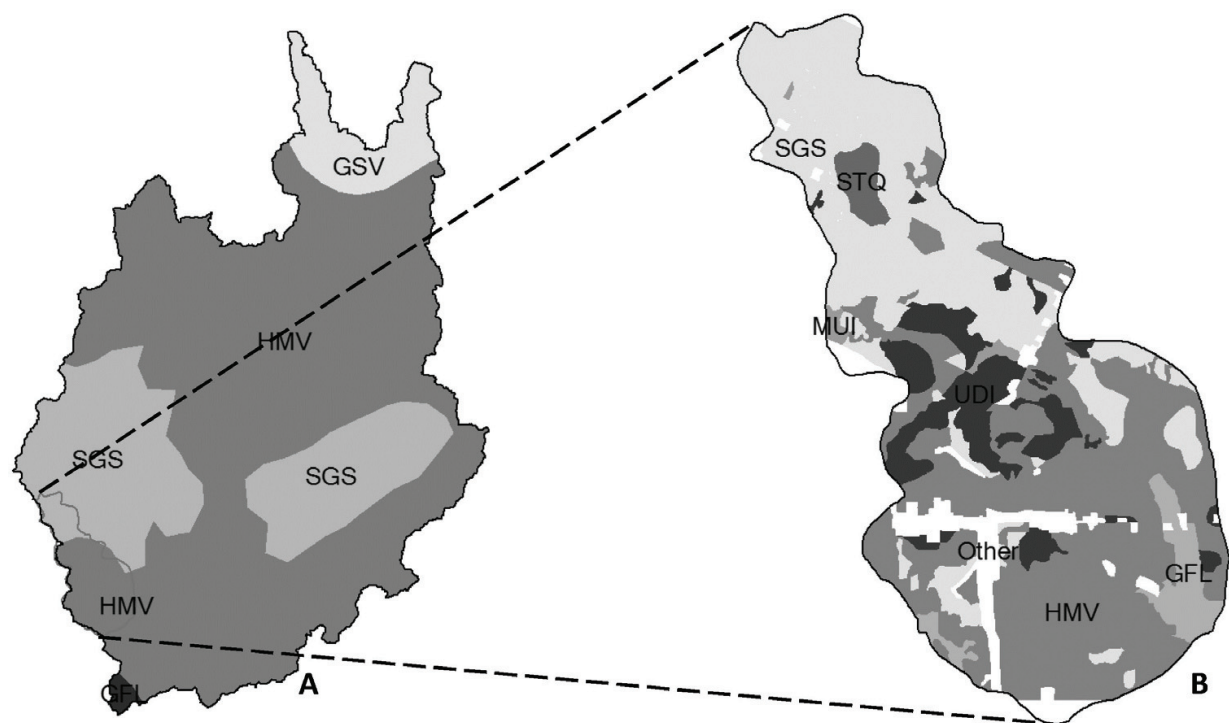
Figure 5. Comparison of coarse resolution soil map (A) and high-resolution soil map (B).

2.2. Target data

Target data, used as reference data in training ANNs, were composed of collecting field soil samples with soil property data (**Figure 6**). Representativeness and density of target data will directly affect the performance of ANNs.

3. Building ANNs for soil properties

A full process of modeling soil properties with ANNs was composed of building ANN structure, training ANNs, and network optimization.

3.1. Building ANN structure

The most popular ANN in modeling soil properties is back-propagation (BP) ANN because this kind of ANNs can map non-linearity when limited discontinuous points exist between input and output data in Ref. [35]. Common BP ANN has three layers: the input layer contains the independent variables used to make model predictions; the output layer contents variables to be predicted; hidden layer connects the input layer and output layer. Each node in one layer is linked with all nodes of the adjacent layer. The number of nodes in the hidden layer determined the complexity of the model. The input weight matrix consisted of all links between the input layer and the hidden layer and the output weight matrix consisted of all links between the hidden layer and the output layer. Weight (w), which affects the propagation value (x) and the output value (o) from each node, was fine-tuned using the value from the preceding layer based on Eq. (8).
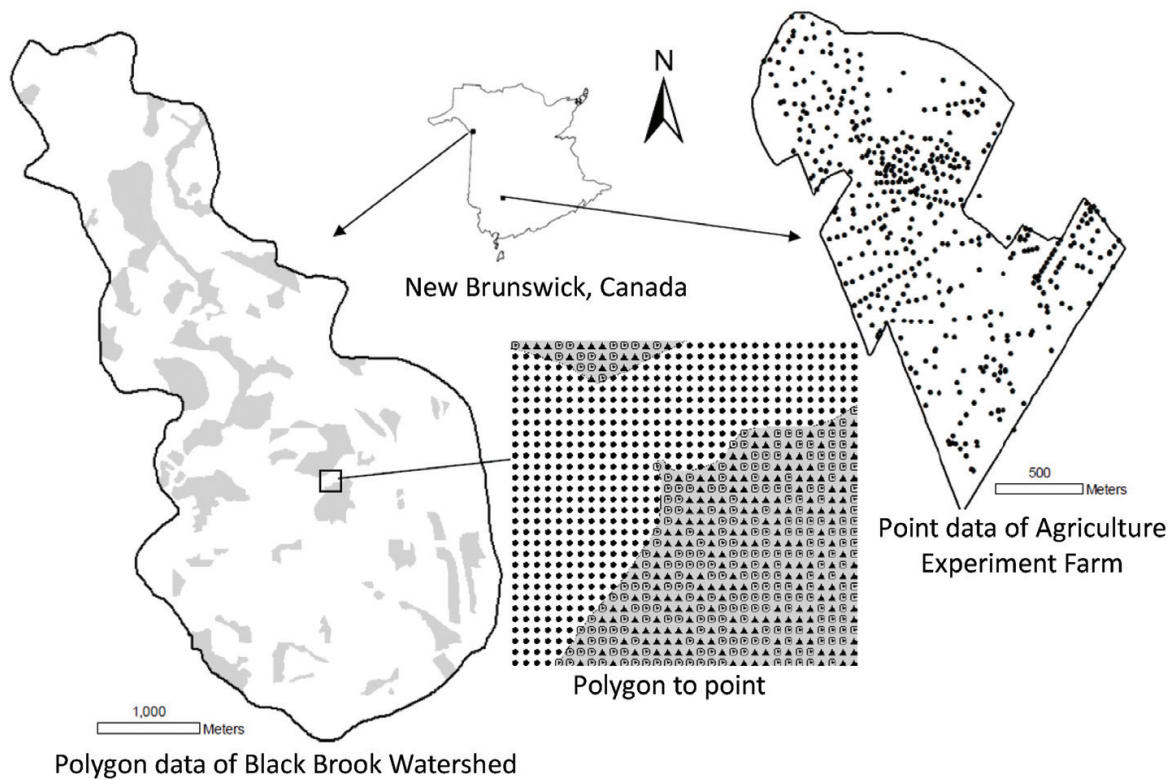
Figure 6. A sample of target data referring to polygons and points.

$$o = f\left(-T + \sum w_i x_i\right) \tag{8}$$

where T was a specific threshold (bias) value for each node; f was a non-linear sigmoid function, which increased monotonically.

When building ANNs for soil properties, the combinations of coarse resolution soil data (i.e., average soil drainage, sand, clay, silt contents) and DEM-derived topo-hydrological data (i.e., slope, STF, SDR, VSP) composed the input layer nodes. Predicted soil properties were the nodes in output layer.

3.2. Training ANNs

The aim of training ANN is confirming coefficients according to different rules or algorithms. BP ANN is trained by self-adjusting weight and bias values of each neuron along a negative gradient descent to minimize the mean squared error (MSE) in Ref. [36]. The MSE between the network outputs (o) and targeted values (t) was calculated through each training cycle (i) by Eq. (9). Training was stopped when the MSE could not be reduced by a set threshold. Frequently-used algorithm included the Levenberg-Marquardt (LM) algorithm and the resilient (RP) algorithm. The LM algorithm was based on Levenberg-Marquardt optimization theory in Ref. [37]. The RP was a kind of rebound back-propagation algorithm in Ref. [38].

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (t_i - o_i)^2 \tag{9}$$

An early stopping method was used to avoid "over-fitting", which has the effect of decreasing prediction accuracy outside of the training data, and improving ANN generalization in Ref. [39, 40]. Through this method, in order to compute the gradient, update the network weights and estimate biases, a training set was used. Another data set, that is, the validation set, was applied to monitor the training process with the purpose of preventing "over-fitting". If training MSE decreased but the validating MSE increased, the training of the ANN model was stopped.

3.3. ANN optimization

The purpose of ANN optimization is adjusting networks structure and improving prediction accuracy of ANNs. It included two parts: (1) selecting the best combination of inputs. The schemes of combining inputs should follow one-variable, two-variable, three-variable, etc. (2) selecting the fittest number of hidden layer's nodes. When the number of hidden layer nodes was too small, prediction accuracy of the ANN was low. When the number of hidden layer nodes was too large, there was a potential over-fitting.

4. Built ANNs for soil properties

4.1. ANNs for soil texture

A BP ANN was developed to estimate soil texture with three-layer structure in **Figure 7** in Ref. [41]. The input layer had six nodes, including average clay and sand contents from coarse
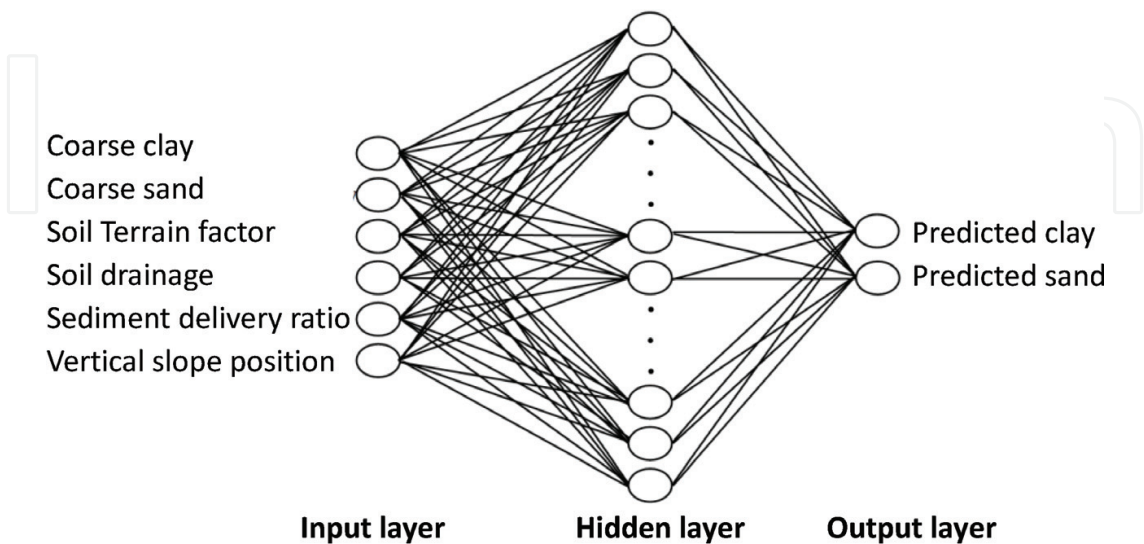
Figure 7. ANN structure for predicting high-resolution clay content and sand content.

resolution soil data, and four DEM-generated topo-hydrologic variables. The output layer contained two nodes: predicted high-resolution clay and sand contents.

The predictive capability of the ANN trained with LM and RP methods was assessed when the hidden layer nodes changed from 5 to 40, and training cycles changed from 25 to 250.

Accuracy of ANN models with the LM and RP training methods when 100 training cycles to various net structures is reported in **Table 1**. Results showed that the ANN models trained with the LM methods had much higher ROA ±5% and lower MSE than the models trained with the RP methods when holding the same number of hidden layer nodes. The LM trained ANN models had better prediction capability. With increasing the number of hidden layer nodes, the MSE of ANNs trained by the LM method was decreasing, but the ROA ±5% got the highest value with 25 hidden layer nodes. According to the results, the best ANN model of predicting clay and sand was a 6-25-2 ANN. Results also directed that when the number of hidden layer nodes was less than 25, the hidden layer scale was too small and the accuracy of model prediction was low. However, over-fitting happened when the number of hidden layer nodes exceeding 25. When the ANN model has been over-fitted, the training accuracy (MSE)

Training algorithm	Net structure	MSE (%)	ROA ±5% (%)[*]	
			Clay	Sand
Levenberg-Marquardt back-propagation (LM)	6-5-2	29	81	76
	6-10-2	26	86	76
	6-15-2	25	85	80
	6-20-2	24	86	80
	6-25-2	24	88	81
	6-30-2	24	85	80
	6-35-2	24	86	81
	6-40-2	23	84	74
Resilient back-propagation (RP)	6-5-2	61	34	33
	6-10-2	39	75	70
	6-15-2	38	79	70
	6-20-2	38	74	68
	6-25-2	35	76	71
	6-30-2	33	80	76
	6-35-2	31	75	72
	6-40-2	28	74	72

[*]Relative overall accuracy (ROA) ±5%, a parameter of assessing the relative accuracy of model predictions, was calculated by counting all predictions within a 5% range of the referenced clay and sand content.

Table 1. Prediction accuracy of ANNs trained with LM and RP algorithms with 100 epochs and nodes of hidden layer changing from 5 to 40 in ref. [41].

increased but the prediction accuracy decreased. In another word, over-fitted ANN models would have poor "generalization" and could lead to inaccurate prediction when using to other input data than the original training set. The same results were presented for the nets trained by the RP method, but the RP method had the highest value of prediction accuracy with 30 hidden layer nodes and the best net structure was 6-30-2.

Prediction accuracies of the 6-25-2 network using the LM training method with training cycles of 25–250 are showed in **Table 2**. As presented, the values of ROA ±5% had the maximum value after 100 epochs. The results indicated that when the epochs of training was more than 100, the ANNs could be over-trained, which is another form of over-fitting.

It can be concluded that net structure, training algorithms, and training cycles would have significant impacts on performance of an ANN.

4.2. ANNs for soil organic carbon

A set of ANNs were developed to predict SOC distribution across the landscape in Ref. [42]. The ANNs used widely available coarse resolution soil map data, high-resolution DEM-generated topo-hydrologic variables, and detailed land use data as inputs. In order to select the best combination of inputs, the various schemes of combining inputs were designed and showed in **Table 3**.

Results from the two-input-node ANN (Level 1) are shown in **Figure 8**. The STF was the poorest predictor of SOC with a MSE of 84 and ROA ±1% (a parameter of assessing model predictions, calculated by counting all predictions within a 1% range of the referenced SOC value) of 66%. The VSP stood out as the best predictor of SOC, with MSE of 29 and ROA ±1% of 70.6%. These results indicated that VSP was the best predictor of SOC distribution across the landscapes.

For Level 2, VSP combined with SDR was the best three-input-node ANN SOC prediction model with MSE of 22. The model of VSP combined with PSR also exhibited a slightly higher

Training cycles	MSE (%)	ROA $\pm$ 5% (%)[*]	
		Clay	Sand
25	27	86	76
50	25	83	72
100	24	88	81
150	24	87	80
200	23	83	80
250	23	84	81

*Relative overall accuracy (ROA) ±5%, a parameter of assessing the relative accuracy of model predictions, was calculated by counting all predictions within a 5% range of the referenced clay and sand content.

Table 2. Predicted soil clay and sand content based on 6-25-2 ANN model using the LM method when the epoch was 25, 50, 100, 150, 200 and 250 times in ref. [41].

Scheme	Level 1	Level 2	Level 3	Level 4	Level 5[*]
1	CSOC, STF	CSOC, VSP, STF	CSOC, VSP, SDR, slope	CSOC, VSP, slope, PSR, sand	CSOC, VSP, slope, PSR, land use
2	CSOC, slope	CSOC, VSP, slope	CSOC, VSP, SDR, PSR	CSOC, VSP, slope, PSR, silt	CSOC, VSP, slope, PSR, land use, drainage
3	CSOC, PSR	CSOC, VSP, PSR	CSOC, VSP, slope, PSR	CSOC, VSP, slope, PSR, clay	—
4	CSOC, SDR	CSOC, VSP, SDR	CSOC, VSP, SDR, slope, PSR	CSOC, VSP, slope, PSR, drainage	—
5	CSOC, VSP	—	—	—	—

*CSOC: coarse resolution SOC data; sand, silt, clay, drainage: high-resolution sand, silt, clay, and drainage data; land use: detailed land use map with 13 classes.

Table 3. Schemes of combining inputs with different levels.
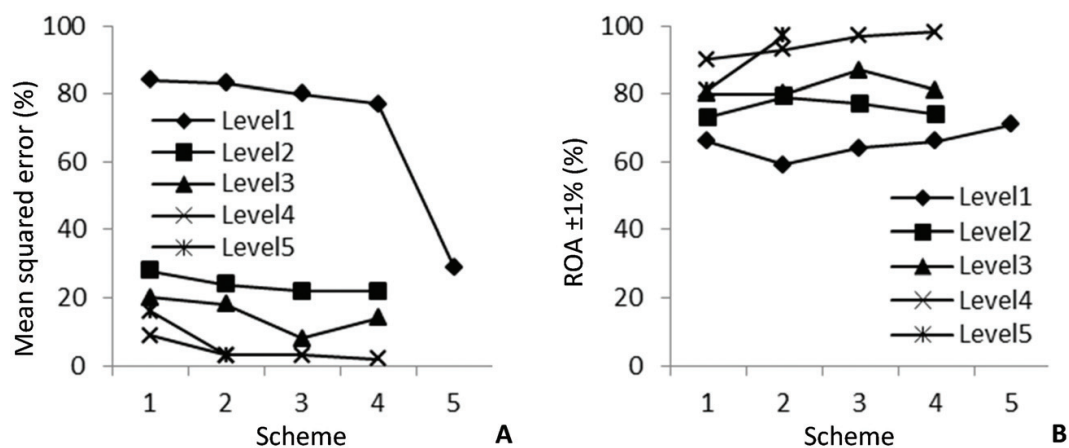
Figure 8. Mean squared error of ANNs (A) and prediction accuracy referring to relative overall accuracy ±1% (B) under different schemes of combining inputs.

MSE (23). However, in terms of MSE, the difference between the two models was considered to be insignificant. Furthermore, the CSOC-VSP-PSR ANN had better performance when measured with ROA ±1% (77 vs. 74%) than the CSOC-VSP-SDR ANN. The model of VSP combined with slope showed the highest values of ROA ±1% (79%).

Within the four input node ANN models (Level 3), the CSOC-VSP-slope-PSR ANN had the best performance, while the CSOC-VSP-SDR-slope ANN had the poorest accuracy of prediction. A further increase of input nodes by adding other DEM-generated topo-hydrological variables could not improve the accuracy of model prediction. As shown in **Figure 8**, the method of adding SDR as a new input node into the CSOC-VSP-slope-PSR ANN could cause a decrease in the accuracy of model prediction.

Input data extracted from high-resolution soil maps significantly improved model prediction accuracy (Level 4). For example, the addition of one soil parameter reduced MSE from a range of 8–20 (level II) to 2–9. Based on the results, soil parameters that were extracted from high-

resolution soil maps could significantly improve the accuracy of model prediction. Both of the content of silt and clay and soil drainage classes were better predictors than the sand content. With MSE decreased to 2 and ROA ±1% increased to 98%, soil drainage was the best additional parameter for modeling SOC.

When land use was introduced as an input layer node in addition to the best four-input-node ANN, CSOC-VSP-slope-PSR, the MSE increased from 2 to 3 but the ROA ±1% decreased from 98 to 97% (Level 5).

4.3. ANNs for soil drainage

An ANN was developed and trained to predict high-resolution soil drainage class maps following the flowchart in **Figure 9**. The research indicated that the best ANN for mapping soil drainage had five input nodes (two from coarse resolution soil maps: average soil drainage class, sand content; three from DEM-generated topo-hydrological variables: slope, SDR, and VSP) and 20 hidden nodes in Ref. [34]. After training, the calibration correlation coefficient of the ANN was 0.69, which was slightly higher than the prediction correlation coefficient (0.65), with MSE of 0.758.

The trained ANN was used to produce a high-resolution soil drainage map for a little watershed (**Figure 10**). An error matrix was constructed using soil drainage records (measured soil
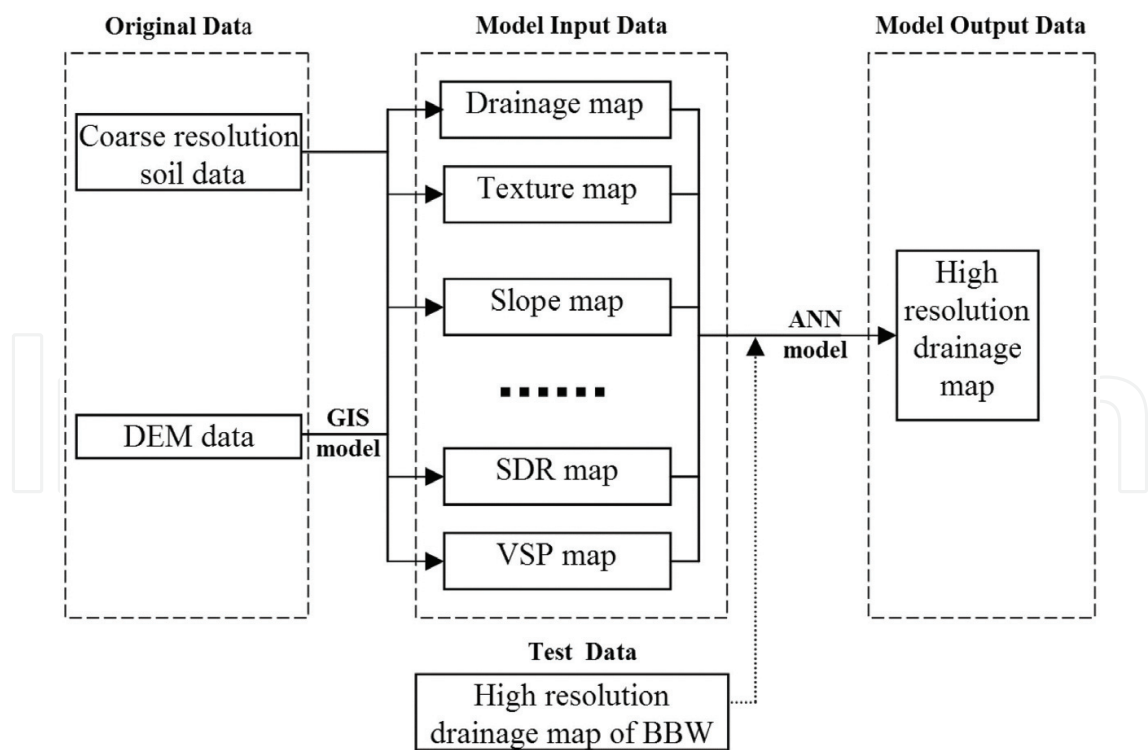
Figure 9. Schematic diagram showing structure and flow of the artificial neural network for predicting soil drainage in ref. [34].
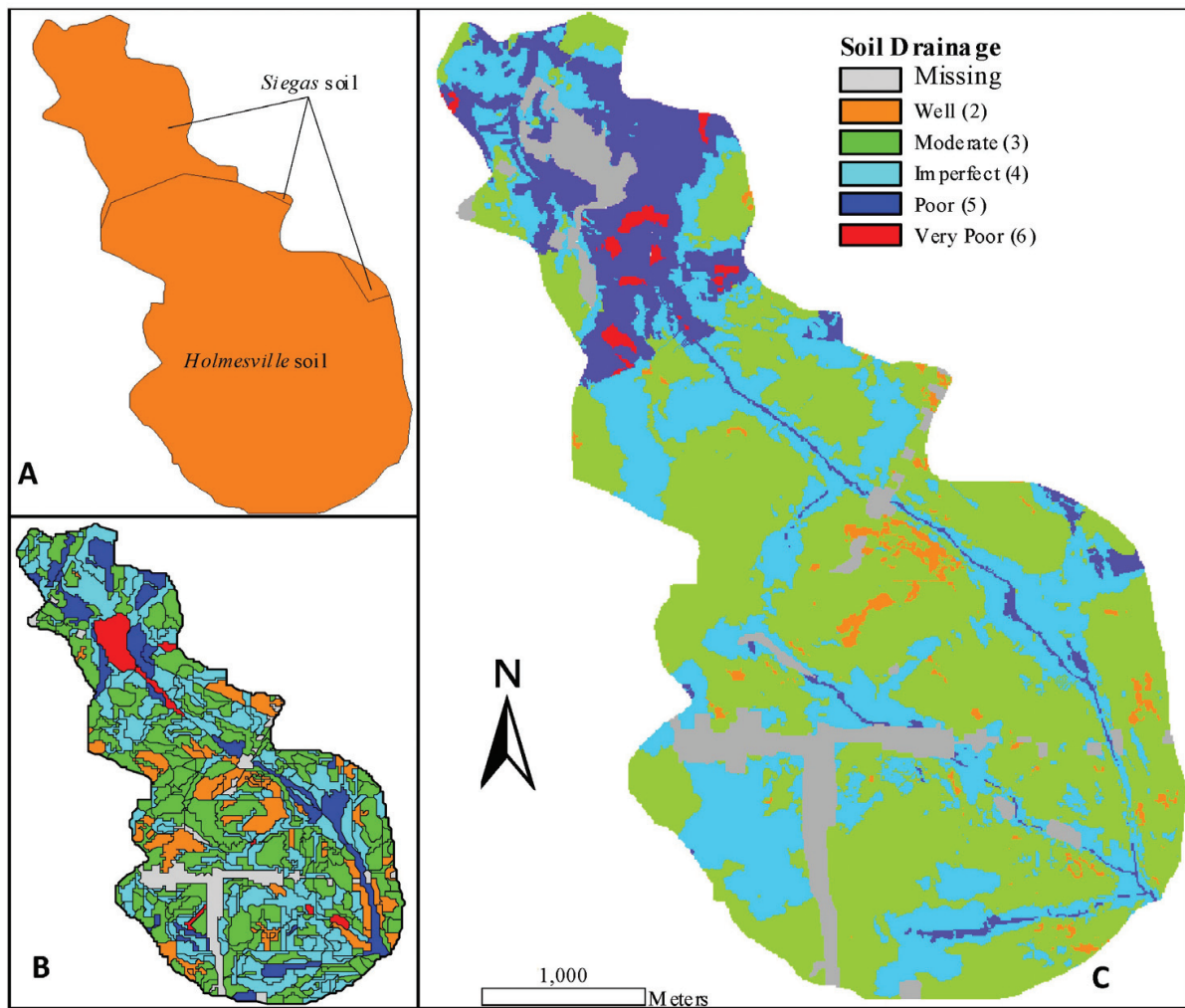
Figure 10. Low-resolution soil drainage map (A), high-resolution soil drainage map (B) and predicted soil drainage map based on artificial neuron network model (C) in ref. [34].

drainage classes) from 1:10,000 soil maps as reference data (**Figure 10B**) and predicted soil drainage classes using the ANN (**Figure 10C**). Results indicated that 52% of model-predicted drainage classes were exactly the same as the field assessment. About 94% of model-predicted drainage classes were within ±1 class compared to the field assessment.

The comparing of coarse resolution soil drainage map (**Figure 10A**) and predicted soil drainage map using ANN model (**Figure 10C**) showed that the predicted soil drainage maps have more detailed soil drainage condition information than the coarse resolution soil drainage map. As shown in **Figure 10C**, the original soil polygon boundaries of coarse resolution soil map are still visible in the high-resolution soil map, which indicated that coarse resolution soil data had a significant influence on the distribution of soil drainage in high-resolution soil drainage map produced. This implied that the accuracy of the coarse resolution soil sand content data, especially around the boundary, will affect the accuracy of predicted high-resolution soil drainage maps.

5. Expanding applications of ANNs

5.1. Deducing general rules from an ANN-analysis approach

It is well documented that soil properties, especially those associated with soil drainage, can be describe in terms of DEM-generated topo-hydrologic variables. However, relationships between soil drainage and these variables are usually difficult to define with conventional statistical methods because of their intense non-linearity. ANNs provide a useful tool to address the non-linear mapping. However, ANNs are "black boxes" with little or no possibility to understand their internal behaviors in Ref. [43] and as a result, relationships between soil drainage class and independent variables are not transparent to users. Furthermore, ANN-prediction accuracy is heavily dependent on the data used to calibrate the model. ANNs also potentially can over-fit the calibration data, which has the effect of decreasing prediction accuracy outside of the calibration data in Ref. [34]. These problems inherently limit the use of ANNs outside areas where the model was originally developed. ANNs could, however, be used to analyze relationships between soil drainage class and topo-hydrologic variables that were quantified by network-parameter.
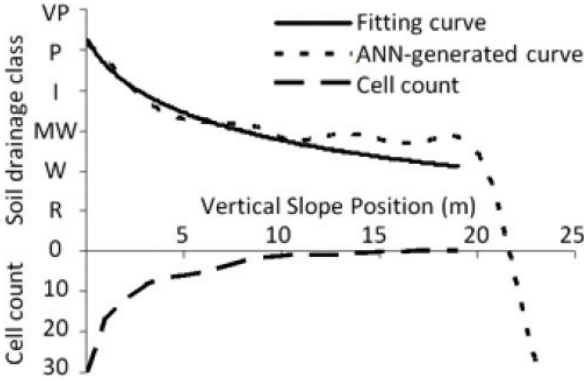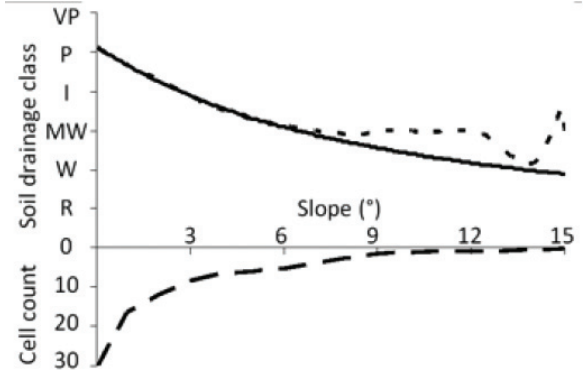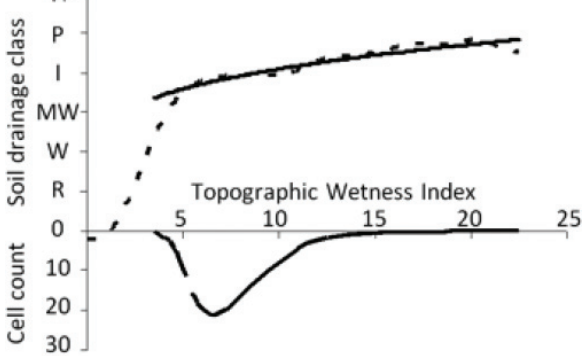
Once the ANNs were trained and tested, they were used to generate the relationships (curves) between ANN-predicted soil drainage classes and topo-hydrologic variables (**Table 4**). Within ANNs with one topo-hydrologic variables, ANN-predicted soil drainage classes (dependent variable) were plotted against independent single variables, with coarse resolution soil drainage data (CSD) being set as constants. Within ANNs with two topo-hydrologic variables, ANN-predicted soil drainage classes were plotted as three dimension surfaces against the two variables, with CSD being set as constants.

The ANN-generated soil drainage-variable relationships (curves) were subsequently formulated as simple mathematical equations using non-linear regression method. Parameters of soil drainage equations were estimated with the Curve Fitting Tool of MATLAB. The used weighted least-squares regression that minimizes the error estimate was used to avoid biases in Ref. [39], included an additional scale factor (the weight factor; the cell count (%) of topo-hydrologic variables) based on Eq. (10):
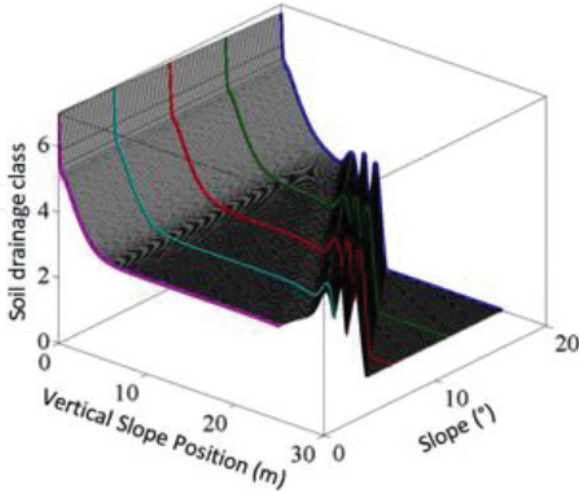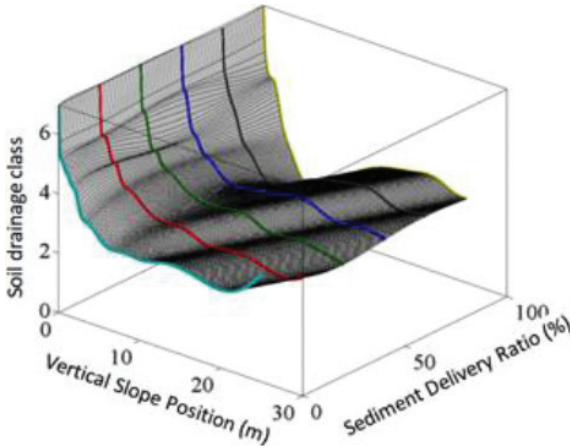
$$S = \sum_{i=1}^{n} w_i \left(y_i - \widehat{y}_i \right)^2 \tag{10}$$

where w_i are the weights, n is the number of data points included in the fit, S is summed square of residuals, y_i is the observed response value, $\widehat{y}_i$ is the fitted response.

Soil drainage equations with single topo-hydrologic variables are summarized in **Table 4**. Most of the soil drainage equation curves (fitting curves) compared well to the corresponding ANN-generated curves, it indicated that prediction performance of soil drainage equations agreed with ANNs in most cases. The maps predicted by the best soil drainage-single variable equation (soil drainage-VSP equation) had accuracies of 44%. Compared to the corresponding ANNs, reductions of accuracy were 2% for the equations.

Inputs of ANNs[*]	Soil drainage curves	Soil drainage equation[**]
Coarse resolution soil drainage (CSD) VSP		$$SD = \dfrac{319.6}{5.921 + VSP^{0.7543}}$$
CSD, slope		$$SD = \dfrac{504.2}{9.845 + slope^{1.044}}$$
CSD, SDR		$$SD = a + bSDR^3 + cSDR^2 + dSDR$$ a = 43.77 b = 0.00005794 c = −0.004975 d = −0.01193
CSD, TWI		$$SD = 12.84 + 14.42 TWI^{0.2878}$$

Inputs of ANNs[*]	Soil drainage curves	Soil drainage equation[**]
CSD = well, VSP, slope		—
CSD = well, VSP, SDR		—

[*]ANN structure: input layer's nodes: (inputs) hidden layer's nodes (20) output layer's nodes (1).
[**]Digital soil drainage classes: rapidly drained (VR)-0, rapidly drained (R)-1, well drained (W)-2, moderately well drained (MW)-3, imperfectly drained (I)-4, poorly drained (P)-5, very poorly drained (VP)-6.

Table 4. ANNs, ANN-generated curves with fitting curves, and equations for soil drainage.

Some disagreements also were observed between soil drainage equation curves (fitting curves) and ANN-generated curves. It implied an advantage of soil drainage equations. These disagreed sections are most likely to occur when there are no or few data points in calibration or validation data sets. In these cases, ANN model predictions appeared unrealistic. For example, when VSP was >18.5 m, CSD-VSP ANN predictions demonstrated a sudden change, which could not be explained and was highly unrealistic. In contrast, the corresponding soil drainage equations curve (fitting curve) logically extended its curvilinear trend, which could avoid the unrealistic predictions made by ANNs in value range where there are insufficient calibration data. Thus, the obtained soil drainage equations could overcome the poor generalization problem of ANN models.

In addition, no requirement for special software support when performing predictions was another advantage of soil drainage equation, compared to ANNs using MATLAB software in Ref. [34] or soil landscape models using ARC/INFO software in Ref. [17].

For ANNs with two topo-hydrologic variables, we intended to produce three-dimensional surfaces (**Table 4**). However, the results were not able to produce meaningful mathematical equations because of the complexity of the data and the uneven distributions of data points across the range of independent variables. For example, the soil drainage surface (CSD = well) from the CSD-VSP-slope ANN model has a contour surface that was too difficult to formulize because of lack of general patterns.

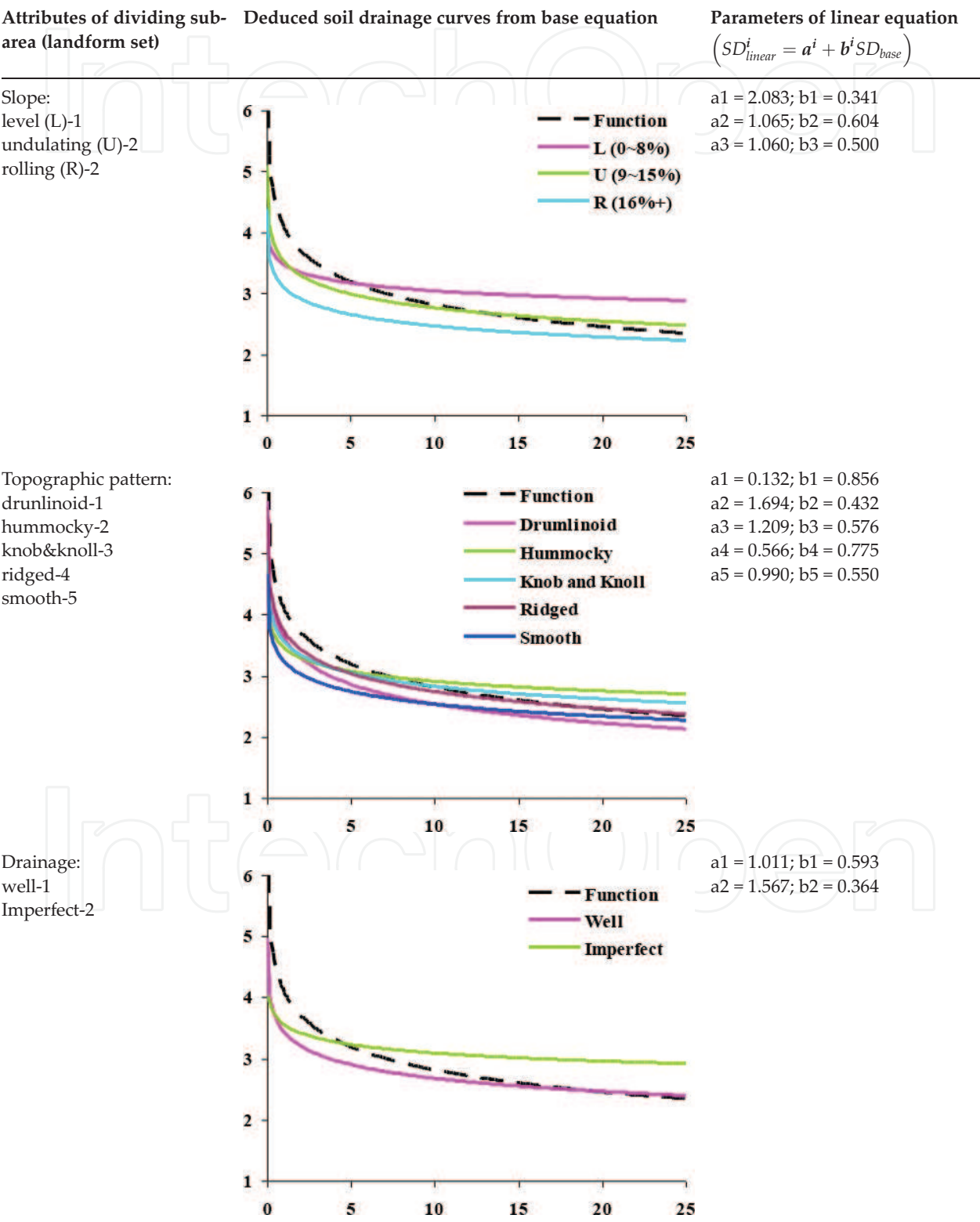## 5.2. Mapping soil property maps over a very large area

Various models, including ANNs, have been developed to predict soil properties. However, it is difficult to use these existing models to produce high-resolution soil property maps over a very large area (>1000 km^2). This is because these models are either interpolation models or statistics models that were built based on the relationships between local environment variables and observed soil property conditions in the field. When applied over a large area, these models may perform well in areas with similar landforms where field samples were collected, but have trouble in areas with significantly different landforms. It is also difficult to build a new model that can produce soil property maps over a large area because it is very difficult to collect sufficient field samples for either interpolation or model calibration. In order to produce soil drainage map over a very large area with limited number of field samples, a two-stage approach was used to produce soil drainage map over a large area (e.g. the province of Nova Scotia) in Ref. [44]. In the first stage, soil drainage-VSP equation, generated from a soil drainage ANN in BBW, was used as the base model because it can capture the general trend of soil drainage distribution rules along topographic gradient. The base equation was directly used to predict soil drainage maps in the province of Nova Scotia. In the second stage, after dividing the entire provincial area into sub-area (landform) based on different division methods, corresponding linear transformation models were subsequently developed to adapt soil drainage classes produced by the base model to fit field samples. Each linear transformation model is composed of a set of linear equations and each linear equation responded to a special landform. Each linear equation was designed as Eq. (11).

$$SD^i_{linear} = a^i + b^i SD_{base} \tag{11}$$

where SD_{base} is the initial drainage classes produced by base model.

a^i, b^i and SD^i_{linear} responded to a special landform (i) of Nova Scotia. a^i is the shifting parameter, which described average difference of soil drainage conditions between the BBW and a special landform of Nova Scotia. b^i is the stretching parameter, which described the change rate of soil drainage conditions between two the BBW and a special landform of Nova Scotia. SD^i_{linear} is the adapted soil drainage classes. Attributes of coarse soil maps were used as the criteria to divide the entire area of Nova Scotia into sub-area (landforms), including slope, topographic pattern, drainage, and texture. Each dividing criteria responded to a set of landforms (**Table 5**).

For each landform of each linear transformation model, using all of field samples within the landform (sub-area) as calibration data, parameters a^i and b^i of the landform (i) were estimated with the regression analysis tool. Only linear equations that passed $P < 0.05$ based on F and t test for the significance of the correlation coefficient were kept. In order to reduce the number
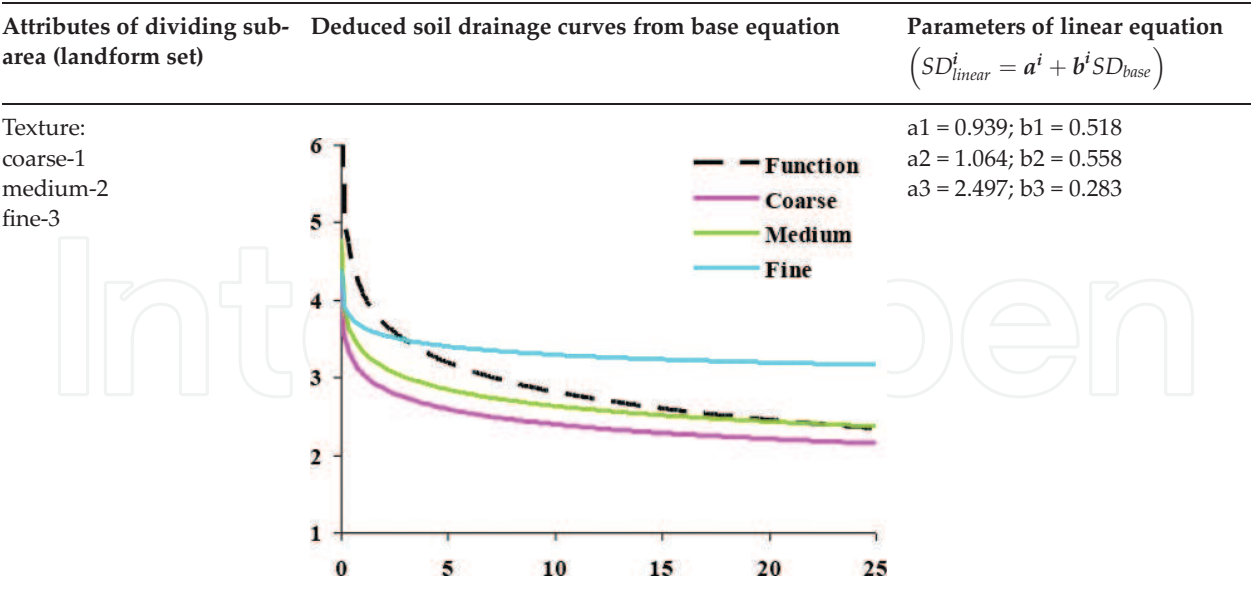
Attributes of dividing sub-area (landform set)	Deduced soil drainage curves from base equation	Parameters of linear equation $\left(SD^i_{linear} = a^i + b^i SD_{base}\right)$
Slope: level (L)-1 undulating (U)-2 rolling (R)-2		a1 = 2.083; b1 = 0.341 a2 = 1.065; b2 = 0.604 a3 = 1.060; b3 = 0.500
Topographic pattern: drunlinoid-1 hummocky-2 knob&knoll-3 ridged-4 smooth-5		a1 = 0.132; b1 = 0.856 a2 = 1.694; b2 = 0.432 a3 = 1.209; b3 = 0.576 a4 = 0.566; b4 = 0.775 a5 = 0.990; b5 = 0.550
Drainage: well-1 Imperfect-2		a1 = 1.011; b1 = 0.593 a2 = 1.567; b2 = 0.364

Attributes of dividing sub-area (landform set)	Deduced soil drainage curves from base equation	Parameters of linear equation $\left(SD^i_{linear} = a^i + b^i SD_{base} \right)$
Texture: coarse-1 medium-2 fine-3		a1 = 0.939; b1 = 0.518 a2 = 1.064; b2 = 0.558 a3 = 2.497; b3 = 0.283

Table 5. Linear transformation models with different landform sets.

of linear equations, field samples that come from different landforms were combined when no significant differences were detected ($P > 0.05$).

As showed in **Figure 11**, prediction accuracies of linear transformation models under different landform sets (**Table 5**) were always better than prediction accuracy of base equation. It indicated that the two-stage methods provide a viable way to extend base equation to generate soil drainage maps over a large area with limited number of field samples.
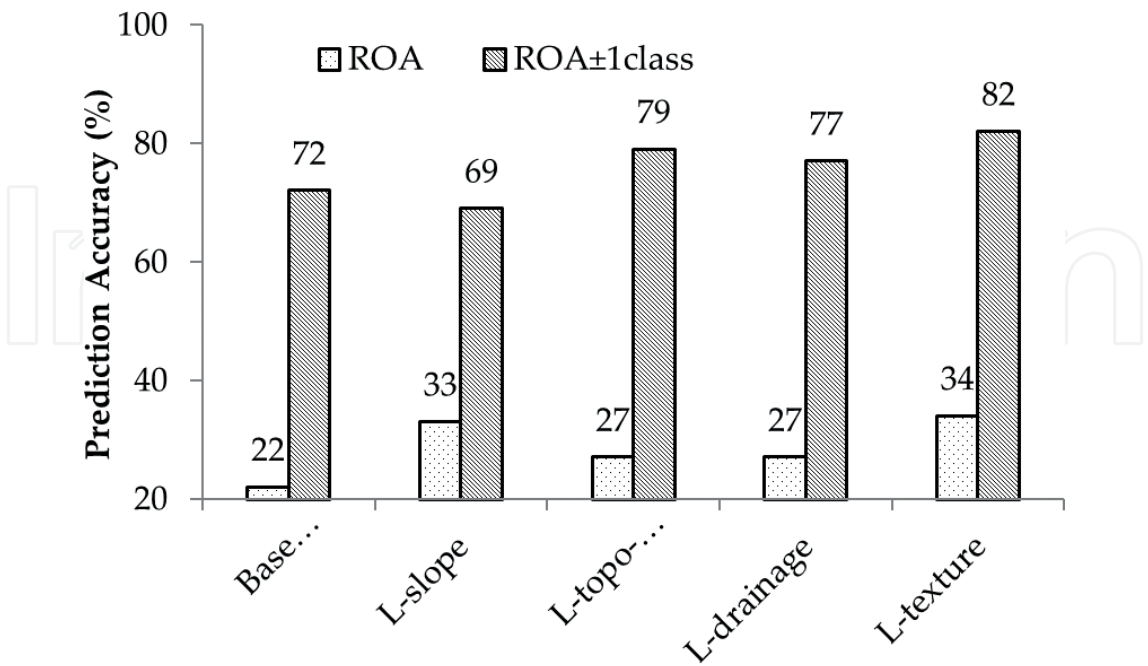

Figure 11. Accuracy comparison of base equation and linear transformation models with different landform sets.

6. Summary

This chapter presented a general approach in using ANNs to produce high-resolution soil properties. It started from preparing input and target data, following by building ANN structure, training ANNs, and optimizing networks. Three successful ANNs for soil texture, SOC, and soil drainage described how to select the fittest hidden layer's nodes, how to select the best combination of inputs, and how to produce high-resolution maps. Two extending applications of the ANNs gave advices in using the obtained ANNs outside the area of ANN calibration.

Acknowledgements

Author details

Zhengyong Zhao[1]*, Fan-Rui Meng[2], Qi Yang[1] and Hangyong Zhu[3]

*Address all correspondence to: z.zhao@unb.ca

1 Guangxi Key Laboratory of Forest Ecology and Conservation, Forestry College, Guangxi University, Nanning, China

2 Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, New Brunswick, Canada

3 Harbin City Forestry Academy, Harbin, China

References

[1] Farrar JL. Trees in Canada. Ottawa, Canadian Forest Service: Markham: Fitzhenry & Whiteside Limited; 2010

[2] Shields JA, Tarnocai C, Valentine KWG, MacDonald KB. Soil Landscapes of Canada: Procedures Manual and User's Handbook. LRRC Contribution Number 88–29. Land resource Research Centre: Ottawa; 1991

[3] Brady NC, Weil RR. The Nature and Properties of Soils. 14th ed. Upper Saddle River: Pearson Education, Inc.; 2008

[4] Allmaras RR, Fritz VA, Pfleger FL, Copeland SM. Impaired internal drainage and Aphanomyces euteiches root rot of pea caused by soil compaction in a fine-textured soil. Soil & Tillage Research. 2003;**70**:41-52

[5] Colombo SJ, Cherry ML, Graham C, Greifenhagen S, McAlpine RS, Papadopol CS, Parker WC, Scarr T, Ter-Mikaelian MT, Flannigan MD. The impacts of climate change on Ontario's forests. In: Colombo SJ, Buse LJ, editors. Forest Research Information Paper No. 143. Forest Research Institute, Marie, ON, Canada. 1998

[6] Oberthür T, Dobermann A, Neue H. How good is a reconnaissance soil map for agronomic purposes? Soil Use and Management. 1996;**12**:33-43

[7] Arp PA. Soils for Plant Growth: Field and Laboratory Manual. Fredericton: University of New Brunswick; 2005

[8] Leeper GW, Uren NC. Soil Science: An Introduction. 5th ed. Melbourne: Melbourne University Press; 1993

[9] Chivenge PP, Murwira HK, Giller KE, Mapfumo P, Six J. Long-term impact of reduced tillage and residue management on soil carbon stabilization: Implications for conservation agriculture on contrasting soils. Soil & Tillage Research. 2007;**94**:328-337

[10] Somaratne S, Seneviratn G, Coomaraswamy U. Prediction of soil organic carbon across different land-use patterns: A neural network approach. Soil Science Society of America Journal. 2005;**69**:1580-1589

[11] Tranter G, Minasny B, Mcbratney AB, Murphy B, Mckenzie NJ, Grundy M, Brough D. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management. 2007;**23**:437-443

[12] Voltz M, Webster R. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. Journal of Soil Science. 1990;**41**:473-490

[13] McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM. An overview of pedometric techniques for use in soil survey. Geoderma. 2000;**97**:293-327

[14] Thattai D, Islam S. Spatial analysis of remotely sensed soil moisture data. Journal of Hydrologic Engineering. 2000;**5**:386-392

[15] Webster R. Fundamental objections to the 7th approximation. Journal of Soil Science. 1968;**19**:354-365

[16] McBratney AB, Santos MLM, Minasny B. On digital soil mapping. Geoderma. 2003;**117**:3-52

[17] Bell JC, Cunningham RL, Havens MW. Soil drainage class probability mapping using a soil-landscape model. Soil Science Society of America Journal. 1994;**58**:464-470

[18] Campling P, Gobin A, Feyen J. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Science Society of America Journal. 2002;**66**:1390-1401

[19] Kravchenko AN, Bollero GA, Omonode RA, Bullock DG. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. Soil Science Society of America Journal. 2002;**66**:235-243

[20] Wasserman PD. Neural Computing: Theory and Practice. New York: Van Norstrand Reinhold Co.; 1989

[21] Levine ER, Kimes DS, Sigillito VG. Classifying soil structure using neural networks. Ecological Modelling. 1996;**92**:101-108

[22] Licznar P, Nearing MA. Artificial neural networks of soil erosion and runoff prediction at the plot scale. Catena. 2003;**51**:89-114

[23] Ramadan Z, Hopke PK, Johnson MJ, Scow KM. Application of PLS and back-propagation neural networks for the estimation of soil properties. Chemometrics and Intelligent Laboratory Systems. 2005;**75**:23-30

[24] Ren LL, Liu XR. Hydrological processes modeling based on digital elevation model. Geographical Research. 2000;**19**:369-376

[25] Ambroise B, Beven K, Freer J. Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity. Water Resources Research. 1996;**32**:2135-2146

[26] Fernandez C, JQ W, McCool DK, Stockle CO. Estimating water erosion and sediment yield with GIS, RUSLE, and SEDD. Journal of Soil and Water Conservation. 2003;**58**:128-136

[27] Haan CT, Barfield BJ, Hayes JC. Design Hydrology and Sedimentology for Small Catchments. San Diego: Academic Press; 1994

[28] Schauble EA. Applying stable isotope fractionation theory to new systems. Reviews in Mineralogy and Geochemistry. 2004;**55**:65-111

[29] Vanoni VA. Sedimentation Engineering. ASCE Manuals and Reports on Engineering Practice No. 54. Reston: American Society of Civil Engineers; 2006

[30] Ferro V, Porto P, Tusa G. Testing a distributed approach for modeling sediment delivery. Hydrological Sciences Journal. 1998;**43**:425-442

[31] Murphy PNC, Ogilvie J, Castonguay M, Zhang C-F, Meng F-R, Arp PA. Improving forest operations planning through high-resolution flow-channel and wet-areas mapping. Forestry Chronicle. 2008;**84**:568-574

[32] Beven KJ, Kirkby MJA. Physically based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin. 1979;**24**:43-69

[33] Meng F-R, Castonguay M, Ogilvie J, Murphy PNC, Arp PA. Developing a GIS-based flow-channel and wet areas mapping framework for precision forestry planning. Proceeding for IUFRO precision forestry symposium 2006, 5-10 March, 2006, Stellenbosch, South Africa; 2006.

[34] Zhao Z, Chow TL, Yang Q, Rees HW, Benoy GA, Xing Z, Meng F-R. Model prediction of soil drainage classes based on DEM parameters and soil attributes from coarse resolution soil maps. Canadian Journal of Soil Science. 2008;**88**:787-799

[35] Li ZY. Supervised classification of multispectral remote sensing image using B-P neural network. Journal of Infrared and Millimeter Waves. 1998;**17**:153-156

[36] Sigillito VG, Hutton LV. Case study II: Radar signal processing. In: Eberhart RC, Dobbins RW, editors. Neural Network PC Tools. San Diego: Academic Press Professional, Inc.; 1990

[37] Fun MH. Training modular networks with the Marquardt-Levenberg Algorithm. Master's Thesis. Stillwater, OK: Oklahoma State University; 1996

[38] Riedmiller M, Braun H. A direct adaptive method for faster back propagation learning: The RPROP algorithm. Proceedings of the IEEE international conference on neural networks, 28 March-1 April, 1993, San Francisco, CA; 1993.

[39] The MathWorks Inc. The Help Document. Natick, MA: The MathWorks, Inc.; Copyright 1984–2007

[40] Xu D, Wu Z. System Analysis and Design Based on MATLAB 6.X-Neural Network. Xian: Xidian University Press; 2002

[41] Zhao Z, Chow TL, Rees HW, Yang Q, Xing Z, Meng F-R. Predict soil texture distributions using an artificial neural network model. Computers and Electronics in Agriculture. 2009;**65**:36-48

[42] Zhao Z, Yang Q, Benoy GA, Chow TL, Xing Z, Rees HW, Meng F-R. Using artificial neural network models to produce soil organic carbon content distribution maps across landscapes. Canadian Journal of Soil Science. 2010;**90**:75-87

[43] Lopez C. Looking inside the ANN "black box": classifying individual neurons as outlier detectors. International Joint Conference on Neural Networks, July, 1999, Washington, DC; 1999.

[44] Zhao Z, Ashraf MI, Meng F-R. Model prediction of soil drainage classes over a large area using limited number of field samples: A case study in the province of Nova Scotia. Canada. Canadian Journal of Soil Science. 2013;**93**:73-83