CGU HS Committee on River Ice Processes and the Environment
14th Workshop on the Hydraulics of Ice Covered Rivers
Quebec City, June 19 - 22, 2007

Estimation of River Ice Thickness Using Artificial Neural Networks

Karem Chokmani[1], Bahaa M. Khalil[1], Taha B.M.J. Ouarda[1], Raymond Bourdages[2]

1 : Institut national de la recherche scientifique - Eau, Terre & Environnement, 490 rue de la Couronne, Québec, Qc, Canada G1V 2X1

2 : Environment Canada, Water Survey of Canada, 373, promenade Sussex, Ottawa, Ontario, Canada, K1A 0H3

Abstract

The purpose of this study is to assess the ability of the artificial neural network (ANN) models in estimating river ice thickness using easy available climate data. A site specific ANN models were developed for two hydrometric stations at two rivers in Alberta (Canada). The ANN models were found to adequately estimate ice thickness. Ways to improve the performances of the ANN models are proposed.

1. Introduction

River ice has an important socio-economic impacts on Nordic Countries through its effects on flooding, hydropower, navigation, ecology and the environment (Morse and Hicks, 2005; Shen, 2003). In this context, ice thickness plays an important role in cold regions engineering since it is required for the design of structures in ice-affected rivers as well as for the numerical modelling of streamflow under ice, ice jam formation and release or for flood insurance studies (Hicks, 2002; White, 2004).

In many cases ice thickness measurements are unavailable or incomplete (Andres and Van der Vinne, 2001). Given the lack of observation data, there is an urgent need to develop effective methods of estimating river ice thickness. Generally, there are two basic methods of ice thickness numerical modelling: stochastic techniques which are based on past records and deterministic approaches which are based on physical ice growth and decay principles.

All deterministic models use more or less simplified version of energy budget and they differ from each other by the level of details used to compute the ice energy budget (Andres and Van der Vinne, 2001; Hicks *et al.*, 1997; Ma and Fukushima, 2002). This kind of models require a large amounts of input data such as wind speed, vapour pressure, water temperature, and cloud cover that are not always available at the site of interest (Andres and Van der Vinne, 2001). Moreover, many assumptions or simplifications for these models directly lead to low accuracy of the models and limitations in their practical applications (Chen and Ji, 2005). Consequently, simplified version of the energy balance models are developed such as

the widely used Stefan's equation, which relies only on accumulating freezing degree-days and correction factors accounting for snow accumulation and wind exposure (White, 2004).

On the other hand, stochastic models require only a set of historical ice thickness observations and their corresponding easily measurable and commonly available meteorological and hydrometric parameters, which are thought to influence ice thickness. The relationship captured in the past records is used for ice thickness prediction. These models are generally simple to operate, while still maintaining an acceptable level of accuracy. The most widely used approach is the multiple regression (Dornan, 2005; Williams and Stefan, 2006). Artificial neural networks represent a promising stochastic technique for ice thickness estimation: Chen and Ji (2005) and Wang *et al.* (2005) have both applied ANN for forecasting the ice condition (dates of ice run, freeze-up and break-up; water temperature; ice flow density; etc.) of the Yellow River (China). More recently, Seidou *et al.*, (2006) have applied ANN for ice growth modeling on a number of Canadian lakes and they showed that ANN are more effective than Stefan's-based models.

The major advantages of the neural networks algorithm over traditional modeling tools are their non-parametric nature and its ability to adapt to different types of data formats and configurations. However, in spite of that ANN have been successfully applied to hydrological modeling, and have shown a great potential in the prediction of a wide range of hydrological parameters (Dawson and Wilby, 2001; Maier and Dandy, 2000), their application for the estimation of river ice thickness remains very limited.

The objective of this study is to develop and test site specific ice thickness models based on hydrometric and meteorological data, using ANN technique.

2. Methods

2.1. Data

The ANN models were developed using ice thickness direct measurements from winter gaugings conducted by Water Survey of Canada (WSC) and climate data from Environment Canada's National Climate Data and Information Archive. Models were applied on ice thickness direct measurements (dependent variable) from two hydrometric stations in Alberta (Canada): Athabasca River at Fort McMurray (07DA001) and Clearwater River at Draper (07CD001). Both stations were associated to the closest meteorological station: Fort McMurry A (Figure 1). Meteorological observations at this station were used to calculate the explanatory variables assumed representative to climate conditions at the two hydrometric stations.

Data are available from winter season 1978-1979 up to winter season 2004-2005. All meteorological records were measured on daily basis with some missed values. Data available for ice thickness were only few measures for each winter season (ranges between 1 and 7 measures per season). The total number of ice thickness observations during the period from 1978 to 2005 found to be only 89 records for station 07DA001 and 90 for station 07CD001. Table 1 summarises the characteristics of available data.

For each station, 8 records (about 10%) were randomly selected to be used only for the ANN model validation. In order to ensure the generalization capacity of the model, records corresponding to the minimum and the maximum values of the ice thickness were included into the model calibration data set.

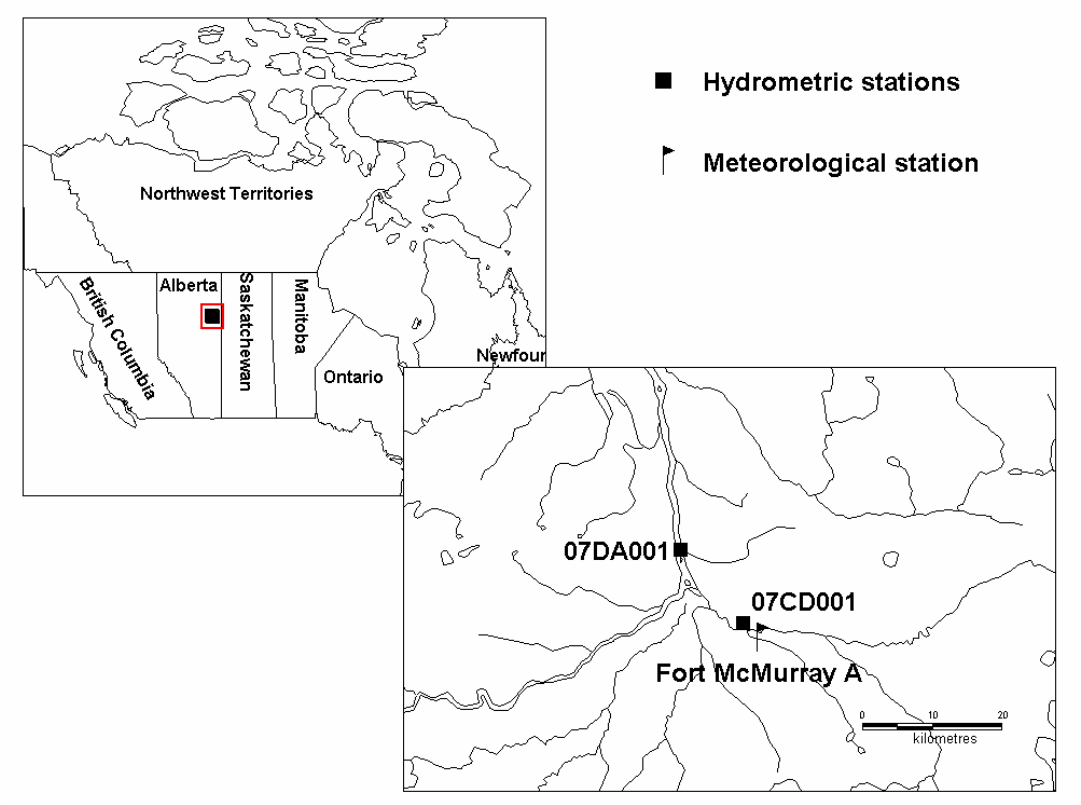

Figure 1: Meteorological and hydrometric stations location

Table 1: Meteorological and hydrometric observations characteristics

Hydrometric station				Meteorological station (observation period)
Name	**ID**	**Winter Gaugings†**	**Distance to meteo. station (km)**	
Athabasca at Fort McMurray	07DA001	1979-2005 **89** [2; 6; 3.5]	18	**Fort McMurray A** (1978-2005)
Clearwater at Draper	07CD001	1979-2005 **90** [1; 7; 3.4]	4.5	

†: First line: observation period, bold: total number of winter gaugings and, between hooks, respectively: minimal number of gaugings per winter, maximum number of gaugings per winter and average number of gaugings per winter.

The following explanatory variables were used as ANN inputs: Accumulated Freezing Degree Days (AFDD), Cumulative Snow (CUMS) and Accumulated Solar Radiation (ASR). Since river ice thickness is a function of changes in meteorological parameters over time (Dornan, 2005) then cumulative climate parameters were used. These parameters were accumulated starting from the ice onset date for each season which was identified using the 'B' flag from the published streamflow WSC database.

2.1.1. Accumulated Freezing Degree Days

Calculation of the AFDD was mainly based on the average temperature during ice presence. AFDD represents the sum of the mean daily temperatures below zero throughout the ice season. If a daily mean temperature is above the freezing point, its value is subtracted from

the AFDD. AFDD starts accumulating from the ice onset date corresponding to the first 'B' flag of the season.

2.1.2. Cumulative Snow

Cumulative Snow (CUMS) represents the sum of daily product of actual snow depth of the snow cover on the ice and the average snow density. This parameter was used to account for the insulating effects of the snow cover on the ice as recommended by Dornan (2005). Surface snow cover, especially newly fallen light density snow, can significantly retard the growth of static ice by slowing the heat loss to the atmosphere. It is defined as follows:

$$CUMS = \sum H * \rho_S \tag{1}$$

Where, ρ_S represents the average density of the snow cover, and H represents the actual depth of the snow cover on the ice.

The average density of the snow cover was estimated from the depth of the snow cover on the ice (H) and the cumulative solid precipitation which has occurred since the ice onset date ($\sum P$). It is computed as the depth of the measured snow water cover equivalent divided by its depth:

$$\rho_s = \frac{\sum P}{H} \tag{2}$$

The depth of the snow cover on the ice (H) is calculated using the snow depth on the ground available at the meteorological station. It is set to zero on the first day of river ice formation. Subsequent changes to the snow cover on the ground were added to the snow cover on the ice.

2.1.3. Solar Radiation

Accumulated Solar Radiation (ASR) was computed as the sum of total daily solar radiation at the top of the atmosphere, which is the raw incoming direct radiant energy from the sun, in Mega Joule per square meter (MJ/m^2). Due to the lack of daily information on atmospheric content and nebulosity, corrections for atmospheric attenuation or for clouds reflections were not applied. Total daily solar radiation was estimated as the sum of hourly solar radiation from the sunrise until the sunset, assuming the solar radiation to be constant within each hour. Hourly solar radiation, sunrise hour and sunset for each hydrometric station were calculated using the station geographic coordinates (for computing details, one can refer to: http://solardat.uoregon.edu/SolarRadiationBasics.html).

2.2. ANN model description

A multi-layer neural network consists of a number of interconnected nodes arranged into layers (Figure 2) where each node (neuron) operates as a simple processing element (Figure 3). Each node in the network is interconnected with all nodes in the preceding and following layers. There are no interconnections within nodes of the same layer. The number of layers and the number of nodes by layer represent the network architecture. The input layer serves as an entry for the vector of input data presented to the network where each node corresponds to one explanatory variable (input). The output layer represents the output data (target). All layers between the input and output layers are referred to hidden layers.

As illustrated in Figure 3, the input (I_j) to a node (j) is the weighted sum (using w_{ij}) of the outputs (O_i) from the nodes of the preceding layer (i). This sum is then passed through an activation function (f) to produce the node's output (O_j) within the range of the activation function. The activation function is usually a sigmoid or hyperbolic tangent, which is a non-linear function with an asymptotic behavior. In our case, we used the hyperbolic tangent sigmoid function.

The interconnecting weight values are adjusted and updated during the training phase to achieve minimal overall training error between the desired and calculated output vectors. All data should be rescaled to ensure they receive equal attention during the training process. In this study, data were rescaled to the interval [-1, 1]. Further details concerning the internal computation between nodes can be found in Rumelhart *et al.* (1986).

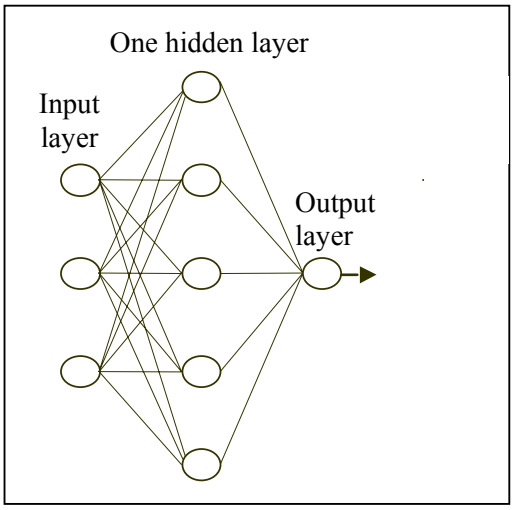

Figure 2: Three layers neural network.

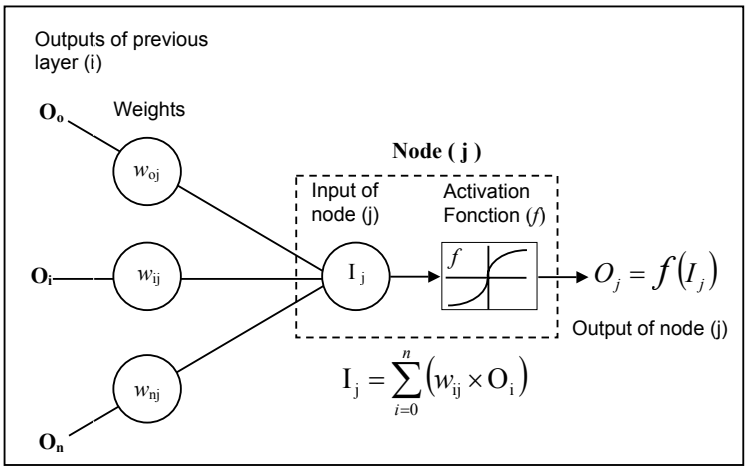

Figure 3: The basic element of a neural network: node computation.

2.2.1. Model architecture

The best neural network architecture can only be determined experimentally for each particular application. According to Hornik *et al.* (1989) and de Villiers and Barnard (1993), ANN models with one hidden layer can approximate any continuous function. Thus, in this study, model architecture of one hidden layer was used.

The number of nodes in the hidden layer(s) should be large enough to ensure a sufficient number of degrees of freedom for the network function and small enough to minimize the risk of loss of the network's generalization ability. Furthermore, it's important to note that a useless increase of the neural network size will lead to a significant increase in training and running time. As recommended by Maier and Dandy (2001), the node number upper limit in the hidden layer(s) was fixed as follows:

$$N^H = \min(2N^I + 1; \frac{N^{TR}}{N^I + 1}) \tag{3}$$

Where, N^H is the number of nodes in the hidden layer(s), N^I number of input nodes and N^{TR} the number of training sample.

2.2.2. Performance assessment

To assess the ANN model performance, one should not relay on individual error measure. Consequently, in this study the following error measure were employed: correlation coefficient (*r*), coefficient of determination (r^2), root mean square error (*RMSE*), relative root mean square error (*RMSEr*), mean absolute error (*MAE*), mean absolute relative error (*MARE*), *Nash* criterion (*Nash*), and index of agreement measure (*d)*. They are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

$$RMSEr = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{y_i - \hat{y}_i}{y_i})^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{7}$$

$$Nash = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

$$d = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (9)$$

where y, $\bar{y}$ and $\hat{y}$ are, respectively, original, mean value and estimated values of the dependent variable and n is the sample size.

It should be noted that the *Nash* criterion compares the model performances to the hypothetic case where the mean value of the dependent variable is used as an estimator. Given a perfect fit, the *Nash* criterion is equal to 1, whereas if the model is worse than the mean of the dependent variable, the *Nash* statistic will be negative. In general, one can expect very satisfactory results from a model with a *Nash* criterion higher than 0.8. However, it was shown that the *Nash* criterion is very sensitive to extreme values, and the index of agreement (d) has been proposed as a possible alternative since it is less sensitive to extreme values (Maier and Dandy, 2001).

2.2.3. Training strategy

The back-propagation algorithm presented by Rumelhart *et al.* (1986) has been used in this application. This training algorithm is now the most commonly used to train neural networks in a wide range of applications.

In order to achieve the best training results and ensure a good generalization capacity for the ANN model, the proper selection of training data is a crucial step, especially in data scarcity situation, as it is the case in this study. One of the major concerns in the neural network training process is the overtraining (over fitting). When overtraining occurs, the neural network's generalization ability will be compromised and the prediction performance becomes narrowly defined around the training datasets. To avoid overtraining, the available training data has been divided into three subsets. The first one is the training data set; which is used for computing and updating the network weights. The second subset, test data set, is used to avoid the overtraining of the neural network by monitoring the error during the training process. Normally, as it is for the learning set error, the error computed on the test set decreases during the initial phase of training. However, when the network begins to over-fit the training data, the error on the test set will begin to increase slowly for the next iterations. At this time, the training process will be stopped, and the neural network weights corresponding to the minimum test error will be maintained for the next steps. The third subset, validation data set, is not used during the training process, and is only used to benchmark the neural network and to compare different models. Indeed, because of the limited number of observations, and to compensate for the small sample size, several models were trained and the best model according to validation group was retained.

3. Results

Preliminary data analysis of ice thickness records showed the existence of an extreme value at each location. Box plots in Figure 4 show that an extreme record of 2.84 m was recognised at location (07DA001) and a record of 1.84 m labelled as extreme value at location (07CD001). Therefore, Grubb's test was applied to verify if these values could be considered as outliers. The test showed that the two records are significantly outliers. The p value, which is the probability of accepting the null hypothesis that the record under consideration is not an

outlier, was 0.0006 and 0.016 for 07DA001 and 07CD001, respectively. Thus, both extreme records were considered as outliers and consequently they were excluded from the data sets prior any further analysis.

The next step was to select randomly the validation group for each studied locations. Table 2 summarises descriptive statistics for dependent and explanatory variables for the calibration data sets from both studied locations. Also, a correlation matrix for each location is presented in Table 2.

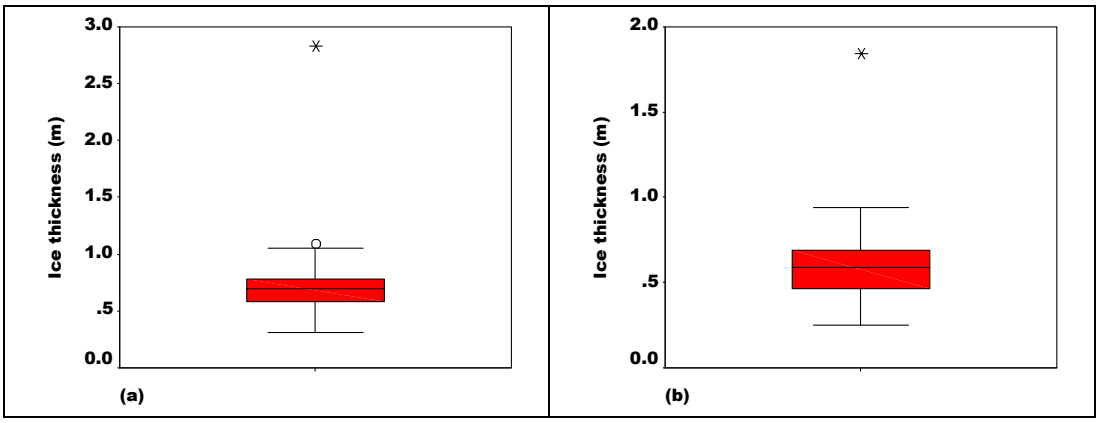

Figure 4: Boxplot on ice thickness values: a) 07DA001, b) 07CD001

From Table 2, one can see that the maximum ice thickness measured at 07DA001 was 1.09 m recorded in April 1979 associated with the maximum AFDD record of 2641.2 degree day, while the minimum record was 0.31 m measured in December 1981. The minimum AFDD record was in November 1979 associated with an ice thickness measure of 0.59 m. For location 07CD001, the ice thickness measures ranges between maximum value of 0.94 m and minimum value of 0.25 m. While, AFDD measures ranges between 2635.8 degrees day associated with ice thickness of 0.69 m measured during April 1979 and 29.5 degrees day associated with ice thickness of 0.65 m measured during November 1979. The mean and the median values for ice thickness were so close indicating symmetry of the data under consideration.

Correlation pattern for both locations indicated that the ice thickness is more correlated to the AFDD than to the ASR. Also, high correlation detected between the AFDD, the ASR and the CUMS as well. Correlation between predictors or explanatory variables should be considered to avoid multicollinearity when using regression models, while it is not the case when using ANNs.

Table 2: Descriptive statistics of dependent and explanatory variables calculated on the calibration set

07DA001	**Statistics**						
		N	Mean	Median	Max	Min	Std-Dev
	Ice thickness	80	0.69	0.7	1.09	0.31	0.17
	AFDD	80	1300.11	1333.25	2641.2	129.1	635.4
	CUMS	64	1.55	1.19	4.72	0.03	1.31
	ASR	80	727.86	750.3	1839.07	129.7	482.83
	Correlation						
			Ice thickness	AFDD	CUMS	ASR	
	Ice thickness		1	0.57	0.41	0.51	
	AFDD		0.57	1	0.73	0.79	
	CUMS		0.41	0.73	1	0.66	
	ASR		0.51	0.79	0.66	1	
07CD001	**Statistics**						
		N	Mean	Median	Max	Min	Std-Dev
	Ice thickness	81	0.57	0.58	0.94	0.25	0.15
	AFDD	81	1299.76	1320.3	2635.8	29.5	647.93
	CUMS	68	1.53	1.08	6.83	0.01	1.46
	ASR	81	724.58	554.21	2003.76	47.85	478.73
	Correlation						
			Ice Thickness	AFDD	CUMS	ASR	
	Ice Thickness		1	0.59	0.45	0.51	
	AFDD		0.59	1	0.75	0.84	
	CUMS		0.45	0.75	1	0.77	
	ASR		0.51	0.84	0.77	1	

Comparison between estimated ice thickness using ANN models versus measured values are illustrated in Figure 5 for the calibration data sets. At both stations, ANN model seems to fit quite well with calibration data, except for low and high ice thickness values (below 0.5 m and above 0.9 for 07DA001 and below 0.3 m and above 0.8 m for 07CD001). These values correspond to the beginning and the ending periods of the ice thickness growth cycle, which has "S" variation shape with time: At the beginning of the winter season, river ice grows slowly due to the system's inertia. By the end of the season, the ice reaches its maximum growth rate due to the decrease of the system response to the atmospheric forcing. This results in an underestimation of high ice thickness values and an overestimation of low values (Figure 5). The failure of the ANN model on fitting low and high ice thickness values could be attributed to the choice of the activation function, the data scaling method and/or the number of the hidden layers. The same results were obtained for the validation data set. As shown in Figure 5, the ANN model of 07DA001 seems to suffer from underestimation of high values (> 0.9 m) and overestimation of low values (<0.5 m), while mid-range ice thickness values were more precisely estimated. As for 07CD001, the estimated values were relatively more precise since the validation data values were in the well fitted range of the ANN model (0.3 m to 0.8 m).

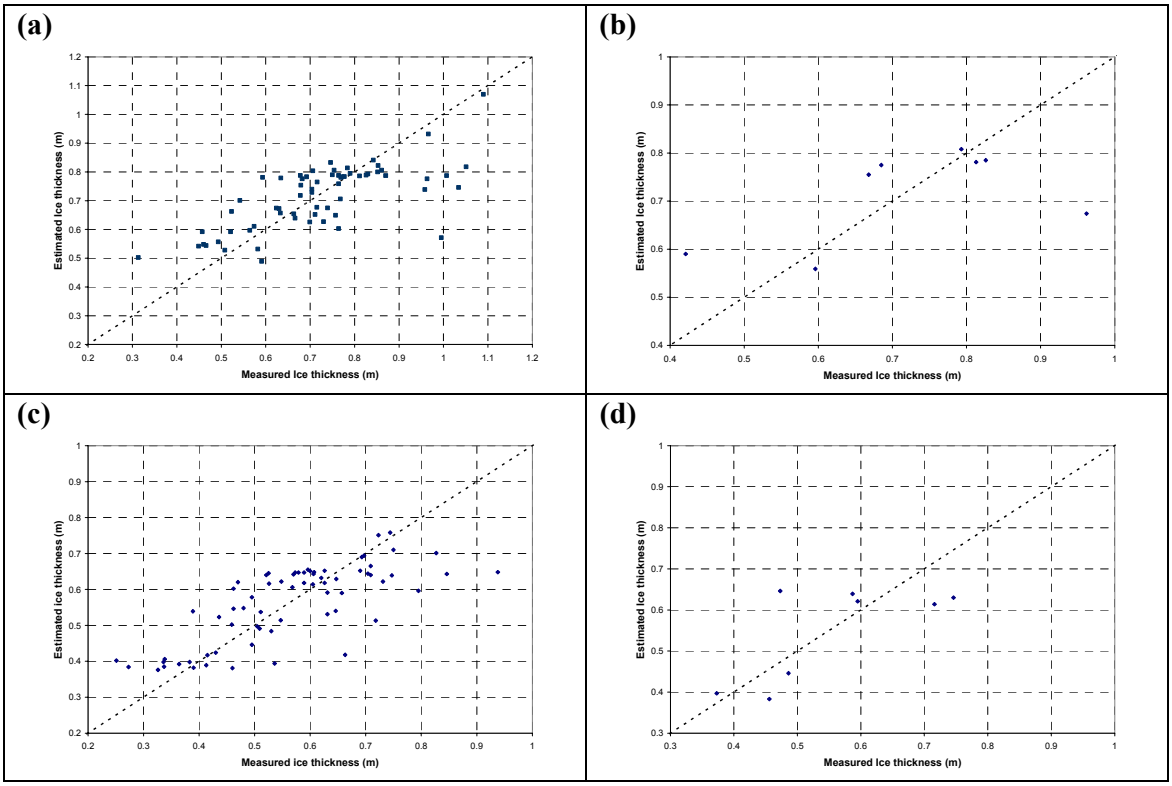

Figure 5: 07DA001 ANN results a) calibration: b) validation; 07CD001 ANN results c) calibration: d) validation

Table 3 presents the ANN model calibration and validation error measures for both studied locations. As expected, calibration results were better than validation. In general, the ANN model performances were not very satisfactory, especially for 07DA001. This could be explained by the proximity of the station 07CD001 to the Fort McMurry meteorological station (Table 1, Figure 1) where observations are more representative of climate conditions at 07CD001 hydrometric station. This was supported by the correlation analysis where climate explanatory variables were more correlated to ice thickness at 07CD001 (Table 2).

According to validation results, the ANN model estimated river ice thickness with a *RMSE* of 9 cm at the station 07CD001 and 13 cm at the 07DA001. For both locations, the *Nash* criterion was smaller than 0.50, far away from the required value of 0.80. The r^2 was less than 0.60 and the *RMSEr* and the *MARE* were both higher than psychological limit of 10%. However, the *d* criterion, which is less conservative than the *Nash,* was in the acceptance zone (>0.80), especially for the 07CD001. If the two ice thickness values out of the best fitting zone of the ANN model for 07DA001 are removed (Figure 5b), the validation results will improve dramatically: the *Nash* criterion reaches 0.55, the r^2 0.61, and the *RMSEr* and the *MARE* drop to 8%and 7%, respectively.

Table 3: ANN model performances

	07DA001		07CD001	
	Calibration	Validation	Calibration	Validation
r	0.71	0.56	0.75	0.70
r^2	0.51	0.32	0.57	0.50
RMSE (m)	0.11	0.13	0.09	0.09
RMSEr(%)	16.2	19.1	17.5	16.7
MAE (m)	0.08	0.09	0.07	0.08
MARE (%)	11.96	14.1	13.3	13. 8
NASH	0.50	0.32	0.57	0.46
d	0.81	0.70	0.85	0.84

4. Conclusions and recommendations

The purpose of this study was to develop and apply site specific ANN models for estimating river ice thickness using easy available meteorological information. The model was tested on two hydrometric stations (Athabasca River at Fort McMurray and Clearwater River at Draper - Alberta, Canada). The results showed that the ANN models represent viable approach for river ice thickness estimation and capable to estimate it with a *RMSE* ranging from 9 cm to 13 cm. These results are comparable to results published in similar studies that have been using stochastic approaches for ice thickness estimation (Dornan, 2005; Seidou *et al.*, 2006).

However, the developed ANN models were not optimum since the models failed to correctly estimate low and high ice thickness values. Therefore, it is believed that the developed ANN models could be improved by: First, rescaling the data in an interval of [0.1, 0.9] may accommodate extreme values (low and high) occurring outside the best fitted data range. Second, using more than one hidden layer may improve the ANN ability to describe the more complex process underlying the ice growth at the beginning and the ending of winter season.

5. Acknowledgments

Financial support and data provided by Water Survey of Canada (Environment Canada) are gratefully acknowledged

6. References

Andres, D.D. and Van der Vinne, P.G., 2001. Calibration of ice growth models for bare and snow covered conditions: a summary of experimental data from a small Prairie pond, 11th workshop on river ice: River ice processes within a changing environment. CGU HS Committe on River Ice Proccesses and the Environment (CRIPE).

Chen, S. and Ji, H., 2005. Fuzzy optimization neural network approach for ice forecast in the Inner Mongolia reach of the Yellow River. Hydrological Sciences Journal, 50(2): pp. 319-329.

Dawson, C.W. and Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography, 25(1): pp. 80-108.

de Villiers, J. and Barnard, E., 1993. Backpropagation neural nets with one and two hidden layers. IEEE Transactions on Neural Networks, 4(1): pp. 136-141.

Dornan, L., 2005. Development of site specific ice growth models for hydrometric purposes, 13th Workshop on the Hydraulics of Ice Covered Rivers. CGU HS Committe on River Ice Proccesses and the Environment (CRIPE), Hanover, NH, USA, pp. 101-131.

Hicks, F., 2002. Continuous monitoring of streamflow under ice conditions. KW605-1-2622, Environnement Canada, Water Survey of Canada, Ottawa, Canada.

Hicks, F., Cui, W. and Andres, D.D., 1997. Modelling thermal breakup on the Mackenzie River at the outlet of Great Slave Lake, N.W.T. Canadian Journal of Civil Engineering, 24(4): pp. 570-585.

Hornik, K., Stinchcombe, M. and White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5): pp. 359-366.

Ma, X. and Fukushima, Y., 2002. A numerical model of the river freezing process and its application to the Lena River. Hydrological Processes, 16(11): pp. 2131-2140.

Maier, H.R. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software, 15: pp. 101-124.

Maier, H.R. and Dandy, G.C., 2001. Neural network based modelling of environmental variables: a systematic approach. Mathematical and Computer Modelling, 33: pp. 669-682.

Morse, B. and Hicks, F., 2005. Advances in river ice hydrology 1999-2003. Hydrological Processes, 19(1): pp. 247-263.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning internal representation by error propagation. In: Parallel distributed processing: Exploration in the microstructure of cognition. D.E. Rumelhart and J.L. McClelland (Editors). MIT Cambridge Press, Cambridge, MA (USA), pp. 318-364.

Seidou, O., Ouarda, T.B.M.J., Bilodeau, L., Hessami, M., St-Hilaire, A. and Bruneau, P., 2006. Modelling ice growth on Canadian lakes using artificial neural networks. Water Ressources Research, 42(W11407, doi:1029/2005WR004622).

Shen, H.T., 2003. Research on river ice processes: progress and missing links. Journal of Cold Regions Engineering, 17(4): pp. 135-141.

Wang, T., Yang, K.L., Guo, Y.X. and Huo, S.Q., 2005. Application of artificial neural networks to forecasting of river ice condition. Journal of Hydraulic Engineering, 36(10): pp. 1204-1208.

White, K.D., 2004. Method to estimate river ice thickness based on meteorological data. : ERDC/CRREL Technical Note TN-04-3, U.S. Army Engineer Research and Development Center, Hanover, New Hampshire.

Williams, S.G. and Stefan, H.G., 2006. Modeling of lake ice characteristics in North America using climate, geography, and lake bathymetry. Journal of Cold Regions Engineering, 20(4): pp. 140-167.