Inductive transfer applied to
modeling river discharge in Nova Scotia

Daniel L. Silver[1], Ian S. Spooner[2*], and Lisa Gaudette[1]

1. Jodrey School of Computer Science, Acadia University, Wolfville, Nova Scotia B4P 2R6, Canada
2. Department of Earth and Environmental Science, Acadia University, Wolfville, Nova Scotia B4P 2R6, Canada
* Corresponding author: <ian.spooner@acadiau.ca>

ABSTRACT

Effective watershed management requires accurate modeling of river discharge. Many years of data collection are often required to capture variations in seasonal trends and produce accurate predictive and descriptive models. In this study artificial neural networks that employ inductive transfer are used to develop models that predict the discharge (flow rate) of streams in Nova Scotia from weather data. The models use two days of weather data to predict the discharge for the following day. The objective is to show that transfer of knowledge from previously learned models for streams can be used to reduce the time and cost associated with collecting large amounts of data for modeling the discharge of a nearby river.

Results show that models developed using only 180 days of training data with transfer from related streams perform as well on independent test data as models constructed using five years of training data and no transfer. The results also show that a considerable variance in stream discharge and stream morphology can be accommodated and that the induced models may be acceptable for management of the resource when little data are available. There is scope for improving the method of transfer by taking into consideration the degree of relatedness between the streams, watersheds, and their associated climate conditions.

RÉSUMÉ

Une bonne gestion des bassins versants exige une modélisation exacte du débit fluvial. Il faut souvent de nombreuses années de collecte de données pour rendre compte des variations saisonnières et produire des modèles de prévision et des modèles descriptifs exacts. Dans le cadre de cette étude, des réseaux neuronaux qui font appel au transfert induit de données météorologiques sont utilisés pour prédire le débit de cours d'eau en Nouvelle-Écosse. Les modèles retenus reposent sur deux journées de données météorologiques pour la formulation de prévisions du débit pour le lendemain. Il s'agit d'illustrer qu'il est possible d'utiliser des connaissances tirées d'autres modèles de cours d'eau déjà établis et de réduire ainsi le temps et les coûts associés à la collecte de données de modélisation du débit d'une rivière proche.

Les résultats indiquent que les modèles établis qui se fondent uniquement sur 180 jours de données de formation et tirées d'autres cours d'eau apparentés donnent d'aussi bons résultats à l'aide des données d'essai indépendant que par les données de modèles établis à partir de cinq années de données de formation, sans transfert. Les résultats établissent par ailleurs qu'il est possible de rendre compte d'un écart considérable dans le débit et la morphologie des cours d'eau et que les modèles produits à l'aide de données préalables sont acceptables pour la gestion de la ressource s'il y a peu de données. Il serait possible d'améliorer ce transfert de connaissances en tenant compte du degré de similitude entre les cours d'eau, les bassins versants et les conditions climatiques connexes.

[Traduit par la redaction]

INTRODUCTION

Stream discharge data are an important component in the assessment and management of ground water and surface water resources and can be applied across a broad range of scales, including engineering design, flood forecasting, reservoir operations, water supply, recreation, and environmental management. Growing populations and competing priorities for water are spurring the demand for more accurate, timely, and accessible water data. However, the field collection of data for modeling stream flow is time consuming and expensive. Time can be the most significant problem because five or more years worth of weather and stream data are commonly required to capture variation in seasonal trends and to produce accurate predictive and descriptive models using standard statistical methods (Cheng *et al.* 2002; Costa *et al.* 2000; Stiff 2000). As well,

doi:10.4138/atlgeol.2009.009

stream flow gauging station data are typically discontinuous as many stream gauging stations have not been maintained. If accurate models of stream discharge could be constructed from a temporally limited data set, such models would have considerable value.

To reduce the time and cost of collecting data, we propose the application of an advanced machine learning technique called inductive transfer to model stream discharge using artificial neural networks. Previously learned knowledge of related streams is used as a source of inductive bias in the development of a model for a new target stream.

This work should be distinguished from prior studies that employed neural networks to watershed management and focused on rainfall-runoff modeling (Bhattacharya and Solomatine 2003; Dawson and Wilby 2001; Dibike and Solomatine 2000; Jain and Chalisgaonkar 2000; Kompare *et al*. 1997), prediction of discharge (Muttiah *et al*. 1997), and modeling of chemical characteristics (Bastarache *et al*. 1997). These studies were done on large transboundary rivers such as the Brahmaputra (Sharma *et al*. 2005) and none employed inductive transfer from related streams. In this study we focus on 2nd to 4th order rivers, which are becoming an increasingly important surface water resource, and on the use of inductive transfer to supplement small sets of training data for these streams.

The streams investigated in this study are all located in Nova Scotia, Canada. The models make short-term predictions employing two consecutive days of weather data as input, with stream discharge predicted for the following day. This study is not a trivial undertaking, as the relationship between weather data and stream flow rate in Nova Scotia is complex (e.g., Stiff 2007).

BACKGROUND

The stream gauging station data used in this study were gathered from three, 2nd order to 4th order streams located in drainage basins in the Annapolis Valley in western Nova Scotia
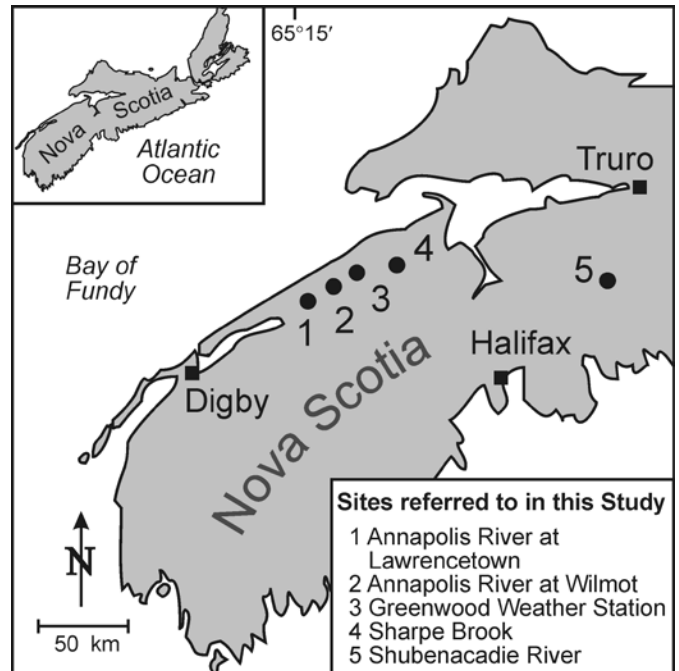

Fig. 1. Locations of the four streams (Annapolis River at Lawrencetown, Annapolis River at Wilmot, Sharpe Brook, and Shubenacadie River) and the source of the weather data (Greenwood, Nova Scotia).

and in the Shubenacadie Valley in Central Nova Scotia (Fig. 1). The Annapolis Valley is about 100 km long and 10 to 15 km wide. The topography varies from the steep slopes and scarp of the North Mountain, to the low relief valley floor, to the slopes and the raised peneplain of the South Mountain (Neily *et al*. 2003; Hamblin 2004; Rivard *et al*. 2007; Trescott 1967, 1968).

The three gauging stations in the Annapolis Valley are located within 60 km of each other (see Fig. 1 and Table 1). They were chosen because robust and easily accessed data sets are available for all three sites and a weather station (Greenwood, Fig. 1) for which a high quality data set exists is located nearby.

Table 1. Summary of stream statistics. Source: Rivard *et al* . (2007).

Name	Distance to Greenwood	Drainage Area	Average Discharge	Standard Deviation	Maximum Discharge
	km	km^2	m^3/s		m^3/s
Annapolis River at Lawrencetown	22	1020	23.68	25.21	276
Annapolis River at Wilmot	9	546	12.26	13.83	140
Sharpe Brook	23	8.81	0.22	0.3	4.79
Shubenacadie River	109	389	12.4	11.21	75.3

Two gauging station sites from which data were gathered for this study are located on the Annapolis River. One station is located near the town of Lawrencetown, Nova Scotia (Fig. 1). Henceforth this site will be referred to as the Annapolis River at Lawrencetown (ARL). The second gauging station site is also along the Annapolis River and is located near the town of Wilmot (Fig. 1) This site is referred to as the Annapolis River at Wilmot (ARW). The third gauging station site is located near the outlet of Sharpe Brook, a 2nd order tributary of the Cornwallis River (Fig. 1). All three sites are located within the Annapolis Lowland physiographic region and Annapolis Valley climate region with headwaters in the Western Nova Scotia climate region. These climate zones are characterized by relatively low rainfall (800–1200 mm annually) and warmer temperatures than are typical of eastern Nova Scotia (Davis and Browne 1996).

Sharpe Brooke is a typical 2nd order valley side stream; it is ungraded along much of its length and has a fairly steep gradient averaging 27 m/km. The watershed for Sharpe Brook above the gauging station covers approximately 19 km². Sharpe Brook typifies streams in Nova Scotia for which discharge data may be available as these streams are typically used for irrigation or are part of a municipal water supply system. The Annapolis River above the ARW gauging station is a 3rd order, primarily valley bottom stream that is fed by streams very similar to Sharpe Brook. Most of the 190 km² watershed for this gauging station is located on the south side of the Annapolis Valley. The ARL gauging station is located about 17 km downstream of the ARW gauging station; the watershed for this site encompasses 780 km². The Annapolis River (and the Shubenacadie River, see below) is typical of the valley bottom streams in Atlantic Canada that are a significant component of the groundwater system and serve as sources of water for irrigation and power generation. These large rivers can also be prone to flooding and surface runoff contamination (Stiff 2007). Gauging stations are rare on these systems and data quality and quality are commonly poor, hence the need to explore methods of discharge modelling. The ARL is the target site for this study.

Weather data were gathered for the town of Greenwood on the floor of the Annapolis Valley from the National Climate Data and Information Archive (Fig. 1). The valley floor is primarily drained by the Annapolis River, which flows southwest to the Annapolis Basin, and the Cornwallis River, which flows northeast to the Minas Basin (Rivard et al. 2004, 2007). The population in the Annapolis Valley is rural, with towns and villages located primarily along the rivers. The most densely populated region in the Annapolis Valley is the Coldbrook-Wolfville Urban Corridor to the east of Greenwood which supports 40% of the population of Kings County (Fig. 1). Agriculture dominates the Annapolis Valley; in 1996 970 farms were recorded in Annapolis and Kings Counties with gross farm receipts totaling $152 M (Government of Nova Scotia 2002).

The Annapolis Valley is underlain mainly by Triassic sandstone and conglomerate and is flanked to the south by metamorphosed Paleozoic and granitic rocks of South Mountain and to the north by the Jurassic basalt highland of North Mountain (Rivard et al. 2007). Wisconsinan glaciation initiated at approximately 75 ka in Nova Scotia, and the Annapolis Valley was ice free by about 12 ka (Stea et al. 1992). In southwestern Nova Scotia, four regional ice flow stages have been recognized during this time (Stea 1987). The resulting surficial geology is complex. Till, glaciofluvial, and fluvial deposits underlie most of the study area. Till is the most common glacial deposit throughout the Annapolis Valley, mantling much of the study area, and varies in thickness from 1 m to possibly more than 20 m (Rivard et al. 2007; Trescott 1968). It rarely overlies a significant thickness of stratified drift, and may occur with inclusions of stratified drift. Till may also underlie bedded silt and clay estuarine deposits.

Local climatic conditions in the study area are heavily influenced by the topography of the Annapolis Valley. For instance, the North Mountain and South Mountain uplands effectively funnel westerly winds through the Annapolis Valley, yet also serve as buffers to provide some protection from weather systems travelling over the Bay of Fundy and the Atlantic Ocean (Neily et al. 2003; Rivard et al. 2007). No part of the study area is more than 60 km from a large body of water (Atlantic Ocean or Bay of Fundy), which can lead to significant variation in both temperature and precipitation. The valley floor is partially protected from direct coastal influences from the Bay of Fundy by North Mountain. As a result, some of the highest summer temperatures in Nova Scotia have been recorded in the Annapolis Valley, particularly in the eastern part. Winter temperatures normally average about -4.5 °C. Total annual precipitation generally ranges between 1100 to 1300 mm in the Annapolis Valley (Neily et al. 2003; Rivard et al. 2007).

The fourth stream gauging station is on the Shubenacadie River, located in the central Nova Scotia Uplands about 100 km east of the Sharpe Brook site (Fig. 1). The Shubenacadie River is primarily a 3rd and 4th order, low relief, valley bottom drainage system similar to the Annapolis River, with a watershed above the gauging station that encompasses 900 km². The Shubenacadie River site is located in the Eastern Nova Scotia climate region, a diverse geographic area with high average rainfall (1000–1400 mm annually) and generally cool temperatures (Davis and Browne 1996). As in the Annapolis Valley, the local bedrock and surficial geology is complex and the thickness of overburden is highly variable. Land use in the region is primarily agricultural. This site was chosen to determine how regional climatological differences might affect transfer of knowledge from models of streams located in different physiographic regions.

Prior to the 1960s little hydrologic information was available for the Annapolis Valley other than weather records (Trescott 1968). Since the 1960s more data have become available, including water quality data for the Annapolis and Cornwallis Rivers (Clean Annapolis River Project and Friends of the Cornwallis River Society, respectively), discontinuous mean daily discharge recorded at eleven gauge stations from 1915 to 2002 throughout the Annapolis Valley, and weather data published to the internet that had been recorded at seven climate stations throughout the province by Environment Canada. A

number of hydrological studies have been conducted in the Annapolis Valley. Trescott (1968) investigated groundwater resources and hydrogeology of the Annapolis-Cornwallis Valley. Hennigar (1992) studied the hydrogeological and groundwater conditions in the Annapolis-Cornwallis Valley, and Myers (1997) looked at the fluvial dynamics and river restoration strategies in Mill Brook, a tributary of the Cornwallis River. Fenton (1998) studied surface water - ground water interaction and stream discharge in Elderkin Brook, a moderate gradient, 1st order tributary of the Cornwallis River. Levy (1998) investigated connectivity between input and stream discharge for Fishwick Brook, a 1st order, low gradient tributary of the Cornwallis River. Cook (2000) studied fluvial dynamics and river restoration strategies in Elderkin Brook and the South Annapolis River, a moderate gradient tributary of the Annapolis River. Blackmore (2007) completed a groundwater vulnerability assessment for the Annapolis-Cornwallis Valley and associated watersheds as part of a Geological Survey of Canada-funded investigation of the hydrogeology of the Annapolis-Cornwallis Valley region (Rivard *et al.* 2007).

The study by Rivard *et al.* (2007) and associated studies by Blackmore (2007), Trépanier (2008), and Gauthier (2008) represent a comprehensive evaluation of the groundwater resources in the Annapolis Valley and have contributed significantly to understanding groundwater system dynamics in valley-ridge settings.

METHODOLOGY

Site Selection

In this study three gauging station sites are used as sources for knowledge transfer (ARW, Sharpe Brook, and Shubenacadie River) to train the target site (ARL). Data quality and quantity from all sites are good (see Table 3). The ARW (4th order stream, valley bottom site) and Sharpe Brook (2nd order stream, valley side and upland site) represent the spectrum of sites from which gauging station data is commonly available in Atlantic Canada. As well, high quality weather station data

Table 2. Details of weather and discharge variables. Source: Government of Canada (2005a)

Parameter Name	Unit	Min.	Max.	Description
Month	12-Jan	1	12	The month of the year
Max Temp	°C	-17.4	35.6	Maximum daily temperature
Min Temp	°C	-35.5	22.6	Minimum daily temperature
Mean Temp	°C	-23.5	26.7	Mean daily temperature
Total Rain	mm	0	106.7	Total daily rainfall
Total Snow	cm	0	53.6	Total daily snowfall
Total Precipitation	mm	0	106.7	Total daily precipitation
Snow on Ground	cm	0	112	Depth measured once per day
Dew Point	°C	-28.16	22.096	Hourly, averaged over the day
Relative Humidity	%	38	99.583	Hourly, averaged over the day
Mean Wind Speed	km/h	0	61.208	Hourly, averaged over the day
Visibility	km	0.7083	25.45	Hourly, averaged over the day
Atmospheric Pressure	kPa	96.611	104.54	Hourly, averaged over the day
Discharge	m3/s	varies	varies	Hourly, averaged over the day

Table 3. Years used for training, validation, and test datasets.

Stream	1986	1987	1988	1989	1990	1991	1992/3	1994	1995
Annapolis River at Lawrencetown									
Annapolis River at Wilmot									
Sharpe Brook									
Shubenacadie River									
	Training					Val.	Test		

are available from a site (Greenwood) that is central to ARL, Sharpe Brook and ARW. Training the ARL using secondary task data from the Shubenacadie River gauging station provides a test for the regional transportability of this technique.

Machine Learning and Artificial Neural Networks

Machine learning is the study of computing systems that improve automatically from experience they acquire from their environments. Machine-learning algorithms use experience, in the form of training examples, such as daily weather and stream discharge data, to develop or *induce* models to predict future events such as how rivers will react as a function of weather conditions.

Artificial neural networks use a method of machine learning based on computational models of biological neurons and networks of neurons as found in the central nervous system of humans. Neural network modeling systems take advantage of massive numbers of parallel processing nodes which work cooperatively to solve a problem. A range of neural networks have been developed; however, the basic structure of an artificial neuron remains the same - to integrate its inputs and to generate an output value as a function of this input. Learning is achieved by modifying the *weight* of the individual connections between the neurons. In an artificial neural network (simulated by computer software) the effectiveness of an input, x_i, from some other neuron is determined by the weight, w_i, of the connection from that neuron. Each neuron has an additional input which is referred to as a *bias*, x_b, (not to be confused with inductive bias mentioned earlier and discussed later in this section). The input value for the bias is fixed to 1, however its weight, w_b, is modified during the process of learning. Input integration is accomplished by an input function which for a unit, j, is most often a simple summation, $I_j = \sum_i x_i w_{ij}$. The output of the neuron, j, is produced by pushing the value of I_j through an *activation* function. The most commonly used activation is the *sigmoid* function given by $y_j = 1/(1 + e^{-Ij})$ where y_j is the output of unit j. This function maps its input to the interval (0,1) becoming asymptotic as the absolute value of summation I_j increases. The behavior of an artificial neural network, depends upon three fundamental aspects; (1) the input and activation functions of the unit (neuron structure), (2) the input connectivity from other neurons (network architecture), and (3) the weight on each of the input connections. Given that the first two aspects are fixed, the behavior of the ANN is defined by the learned values of the weights.

The paper presents the results of using multi-layer feed-forward neural networks as shown in Figure 2. The networks are simulated as part of a general machine learning environment developed at Acadia University called the Research and Application Sequential Life-long Learning system, or RASL3 (ML3 2009). Each network consists of an input layer, a hidden layer, and an output layer of neurons connected in a strictly feed-forward fashion. The network accepts inputs and generates outputs that are continuous and in the range (0,1). To classify an example, the set of attribute values are presented to the input nodes. Each input node forwards the value on to all nodes in the hidden layer. The hidden nodes compute their activations and forward them on to all nodes in the output layer. The activation value(s) produced by the output node(s) indicate(s) the predicted class.

To learn a task using neural networks of the type shown in Figure 2, the weights of the connections must be adjusted to produce the hypothesis with greatest generalization. The most widely used learning algorithm for this type of network is the back-propagation of error algorithm (Mitchell 1997). The sum of squared errors between the output(s) of the network and the target output(s) as provided by the training examples is backward propagated from the output layer down through each of the hidden layers. The change in each weight is expressed as the derivative of the error with respect to weight. At each node, each incoming connection weight is adjusted to minimize the error contributed by that weight to the global error. Thus, the process of learning is one of iteratively presenting the training examples and making small weight changes to reduce the error. The algorithm stops when the error across all examples reaches a minimum. This learning process can be described as gradient descent through a space of all weights in the network in search of a set of weight values that minimizes the error for all training examples.

The performance of a model, and therefore the machine learning algorithm that produced it, is normally based on how accurately it predicts output values for a previously unseen set of test examples. This is referred to as the model's generalization performance. Sufficient generalization performance is necessary to provide confidence in the model's ability to make future predictions. To ensure the development of a neural network model that will have good generalization (predictive accuracy), a randomly chosen *validation* or *tuning* set of examples is used to monitor when the model starts to overfit the training data. In summary, when developing a network model, one typically creates training, validation, and test sets from the available data. The training and validation sets are used to train the model, and the independent test set is used to judge the generalization performance of the model.

Inductive Transfer and MTL Neural Networks

Every machine learning method has a space of models or hypotheses (*e.g.,* linear equations, logical expressions, graphical structures, or probability tables). The development of a predictive model can be considered a search over this space for the hypothesis that best matches the training examples. Anything that constrains the search within this hypothesis space, beyond the training examples, is called *inductive bias* (Mitchell 1997). Inductive bias is essential for the development of a hypothesis with good generalization from a practical number of examples. Without inductive bias accurate learning cannot occur because the training examples are insufficient for selecting the best model. Ideally, a learning system can select its inductive

bias to tailor the preference for hypotheses according to the task being learned (Thrun 1997). One type of inductive bias is prior knowledge of the domain of tasks being learned. A domain of tasks is defined by some sense of relatedness between the tasks, minimally they all share the same inputs and, more specifically, they share similar invariances or features over the input space.

The retention and use of task domain knowledge (DK) as a source of inductive bias has become known as *inductive transfer* or *transfer learning,* and remains an open problem in machine learning (Caruana 1997; Thrun 1997, Silver and Bennett 2008). The goal of inductive transfer research is to find ways of using prior knowledge to develop more accurate hypothesis (models) with fewer training examples as quickly and efficiently as possible (Silver and Mercer 2002).

Knowledge Transfer in MTL Networks.

Multiple task learning (MTL) neural networks are one of the better documented methods of inductive transfer (Caruana 1997). An MTL network is a feed-forward multi-layer network with an output for each task that is to be learned. The standard back-propagation of error learning algorithm is used to train all tasks in parallel (Mitchell 1997). Consequently, MTL training examples are composed of a set of input attributes and a set of target outputs, one for each task. Figure 2 shows an MTL network containing a hidden layer of nodes that are common to all tasks. The sharing of internal representation is the method by which inductive bias occurs within an MTL network (Baxter 1996). The more tasks are related the more they will share representation and create beneficial inductive bias.

Sequential Learning through Task Rehearsal.

The task rehearsal method was introduced by Silver and Mercer (2002) as a machine life-long learning system that is able to retain and recall task knowledge. After a task, T_k has been successfully learned, its hypothesis representation is saved in a domain knowledge store. This representation acts as a surrogate for the space of input-output examples that defines task T_k. Virtual examples of the input-output space for T_k can be produced by passing inputs to the domain knowledge representation for T_k and recording the outputs. When learning a new task, T_0, the domain knowledge representa-
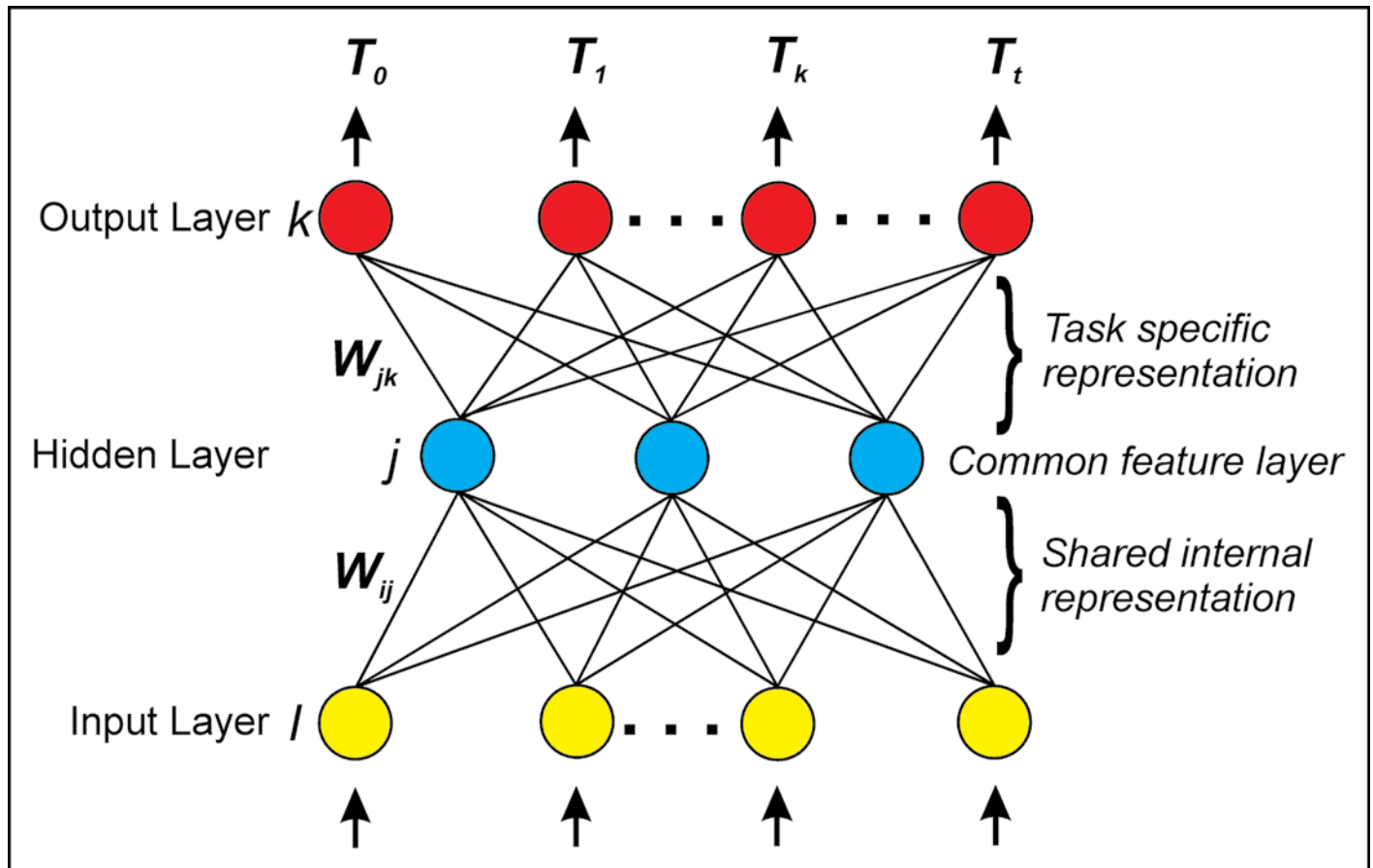

Fig. 2. An example of a multiple task learning (MTL) network with an output node for each task being learned in parallel. Inductive transfer between tasks occurs as a function of sharing internal representation (connector weights Wij) below the common feature layer.

tions for tasks $T_1...T_k...T_t$ are used to generate corresponding virtual output values from the set of T_0 training examples. The resulting set of virtual examples is used to relearn, or rehearse, the domain knowledge tasks in parallel with the learning of T_0 in an MTL network. MTL training can be started from either random initial weights or from the prior domain knowledge weights (O'Quinn *et al*. 2005; Poirier and Silver 2004). It is through the sharing of internal representation and the rehearsal of previously learned tasks that prior knowledge is transferred to the new task.

Theory and Approach to Model Development

The objective of this research was to determine the value of inductive transfer as applied to modeling of discharge for a target stream using the data from one or more gauging stations as a source of transfer. Our hypothesis was that a previously developed model of discharge for a specific stream can be used as a source of inductive transfer when developing a model for a distinct but geospatially and geomorphologically related stream. Through the use of prior model knowledge and inductive transfer, fewer years of training data are required to construct accurate models. The more physically and hydrologically similar, or *related,* the streams are, the greater the expected benefit from the transfer.

In this study, the sources of knowledge transfer are stream discharge datasets that exist for ARW, Sharpe Brook, and the Shubenacadie River near Enfield (Fig. 1). The primary target task is predicting discharge at the ARL gauging station (Fig. 1).

To demonstrate the value of inductive transfer, the following approach is taken. First, standard single task learning (STL) models are created using standard back-propagating neural networks for the Annapolis River at Lawrencetown for data sets ranging from 180 days to five years. Next, domain knowledge (DK) models based on 5 years of training data are constructed for (1) the ARW, located upstream of the ARL and having approximately half the discharge, (2) Sharpe Brook, located nearby but in a different drainage basin, (3) the previous two streams together in an MTL back-propagating network, and (4) the Shubenacadie River at Enfield, located in a geomorphologically distinct drainage basin located about 120 km from the other sites with a unique, local climate. Finally, the four DK models are each used as a source of transfer for learning the primary task (the ARL) in MTL networks task rehearsal as described in Section 2.2. To ensure a fair comparison, the models developed with inductive transfer are trained with the same primary task data used to develop the STL models.

The models were supplied with two days of weather data from Greenwood, Nova Scotia, to predict the discharge for the following day as this time frame best captures the duration of precipitation events in the study region. The STL and MTL models were tested against an independent test set and compared. The Mean Absolute Error (MAE) is the average of the absolute value of the error of each example and is reported in cubic meters per second (m³/s). The correlation measures

the covariance of the actual and predicted discharge. Graphs of the actual vs. predicted discharge over a period of time are also used to analyze where the models are performing well or poorly. Paired, two-tailed T-tests of the difference of the MAE are used to measure if the difference between the STL and inductive transfer models are significant.

Data Collection and Preparation

The weather and stream data used in this study were obtained from two distinct sources within Environment Canada for the years 1986–1995. The weather data come from the on-line National Climate Data and Information Archive (Government of Canada 2005a). It was received in hourly, daily, and monthly formats depending upon the variable (temperature, precipitation, pressure, etc.). For the purposes of this study all data was converted to daily values (total, mean, maximum or minimum). Table 2 provides descriptions and statistics of the various weather parameters used in the study. Table 3 shows which years of data were used for training, validation, and testing.

The weather data were reasonably complete. Where necessary, missing values were input using the average of the previous and next day values. A large amount of weather data are missing for the period from December 1992 to November 1993. Hence, the decision was made not to use data from this period. Consequently, the year labeled "1992/93" in Table 3 is composed of data from January to November, 1992, and data from December, 1993.

Discharge data for several streams were obtained through the on-line Water Survey of Canada "HYDAT" database (Government of Canada 2005b). Stream discharge data are currently being recorded at nearly 3000 sites across Canada, and are available in near-real time from just under half of those sites. The ARW is the only site used in this research which is currently monitored in real time. Historic data for Sharpe Brook and the Shubenacadie River are available through to April, 1995, whereas the ARL has data available through to December 2000 and for the entire year of 2003.

The weather and stream data required a significant amount of preparation prior to modeling. The data were combined into a single tab-delimited file using a series of Perl scripts that removed header information, calculated daily statistics, and combined data from multiple files into a single file ready for use by the neural network modeling software. Each row of the final example sets contains 25 input variables (the current month, 12 weather variables for yesterday, 12 weather variables for today), and one target output variable (the discharge value for tomorrow).

The studies consider the effect of secondary task transfer with varying amounts of training data for the primary task, which is stream discharge for the ARL. For the primary task, training sets containing five years, three years, one year, and 180 days of data were used, along with one year of validation data and two years of test data. The validation or tuning set of

data was used to prevent over fitting the neural network model to the training data. For the DK models constructed as sources of transfer, five years of training data were used, with one year of validation data and two years of test data. The data used for ARL was shifted forward a year so as to realistically challenge the task rehearsal method to create novel virtual examples for rehearsal of the secondary tasks. Table 3 shows the years of data used for each stream. For the ARL, the five-year period consists of January 1, 1987, to December 31, 1991 (1825 examples); the three-year period is from 1987 to 1989 (1095 examples), the one-year period is 1987 (365 examples), and the 180 days were selected at random from the 1987 data ensuring 10 to 15 days per month. The data from 1992/93 (365 examples) was used as validation data to prevent overfit to the training set. The independent test data are from the years 1994 and 1995 (730).

Model Development and Analysis

This section covers the development and comparison of the various models as outlined in the previous section on theory and approach to model develpment.

Neural Network Architecture and Learning Parameters

Three layer networks were used for all models (Fig. 2). Twenty hidden nodes were chosen for the STL and MTL networks. This provided sufficient representation for multiple tasks within the MTL networks when using one secondary DK task. Thirty hidden nodes were chosen for MTL networks with two secondary DK tasks. Provided a validation set is used to prevent over-fitting, additional hidden nodes do not hinder the development of accurate models. A learning rate of 0.0025 was used to produce faster training when transfer was not used, without loss of accuracy, and 0.001 was used for the MTL networks when transfer is occurring from the secondary tasks. The momentum term remained constant at 0.9 and random initial weights were chosen in the range (-0.1 to 0.1).

STL Models for the Annapolis River at Lawrencetown (ARL)

The purpose of this experiment was to develop models using single task learning (STL) neural networks with varying amounts of training data that predict the following day's discharge for the ARL. The performance of the models on the test set were compared to each other and subsequently to models constructed with inductive transfer using prior knowledge. This experiment used the four different training sets described earlier and shown in Table 3. The models were developed using a 25-20-1 (input-hidden-output) network. Five repetitions were performed for each set of training data with different random initial weights. Up to 1.5 million training iterations were allowed for each repetition.

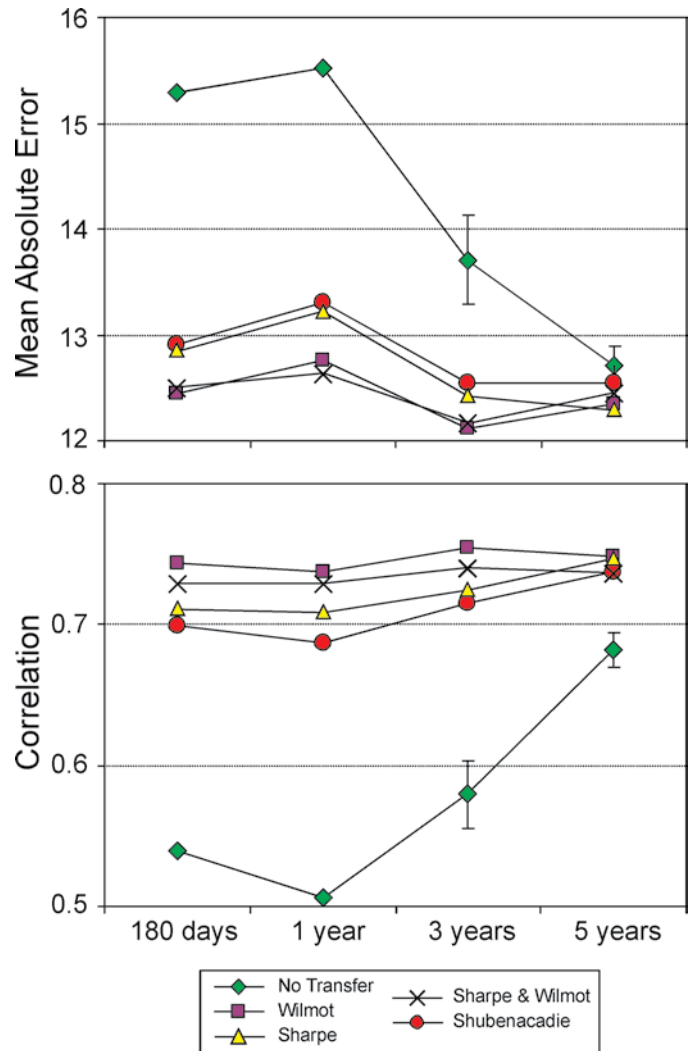Figure 3 presents the performance (and 99% confidence in-

Fig. 3. Performance of Annapolis River at Lawrencetown models on the test set with and without transfer as a function of the amount of training data.

tervals) of the various STL models on the test set as a function of the number of training examples. The graphs show that the models steadily improve as more training data were used. The exception is the set of models developed with only 180 days of data. These models have slightly better performance than those developed with one year's worth of data. We speculate that this is because, by chance, the random sample of data chosen for the 180-day training set is closer to that of the test set than the one year training set. That is to say, the one-year dataset contains some additional noise.

Figures 4 and 5 compare the test set graphs of the actual vs. predicted stream discharge for the 180 day and five-year STL models. The five-year STL model can be seen to better predict the discharge peaks and generally follow the trends of the actual data. This reflects the improvement in MAE and correlation performance over the 180-day model.
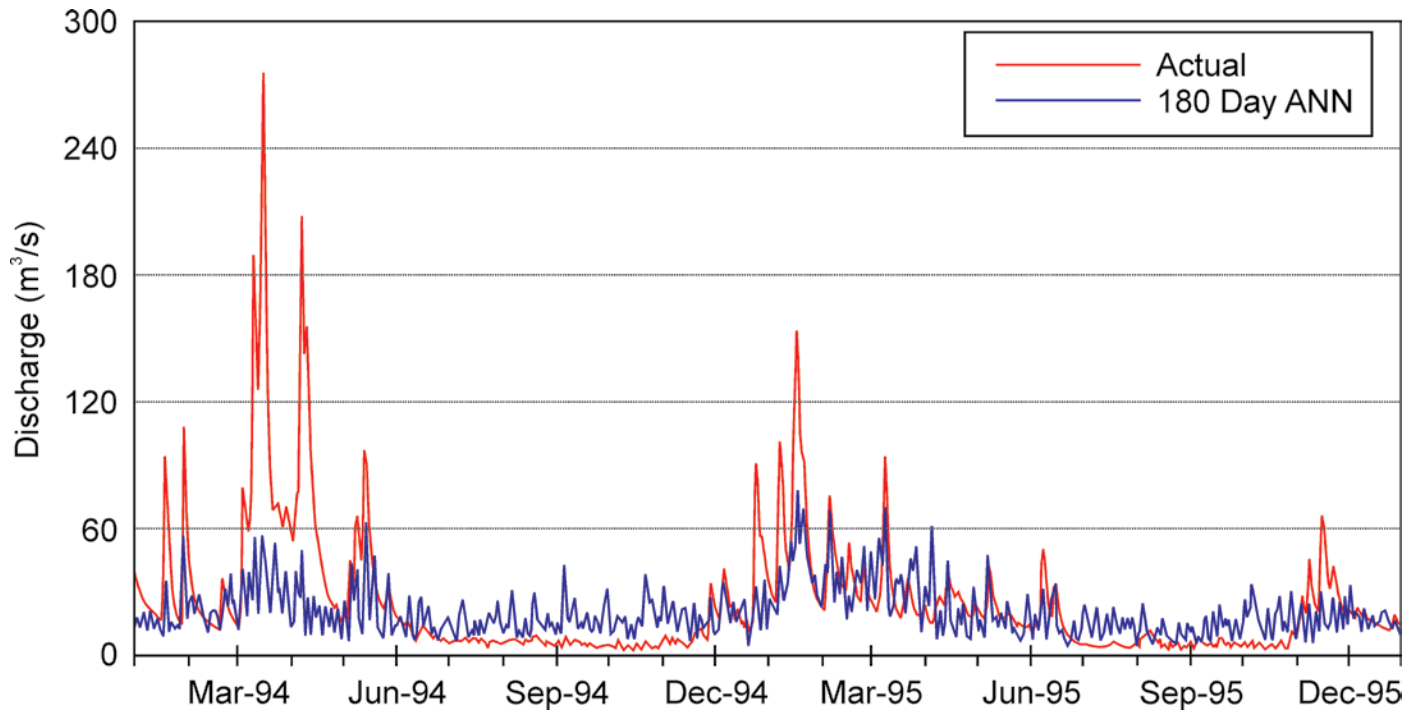
Fig. 4. Annapolis River at Lawrencetown actual discharge vs. predicted discharge for a 180 day single task learning (STL) model without transfer.
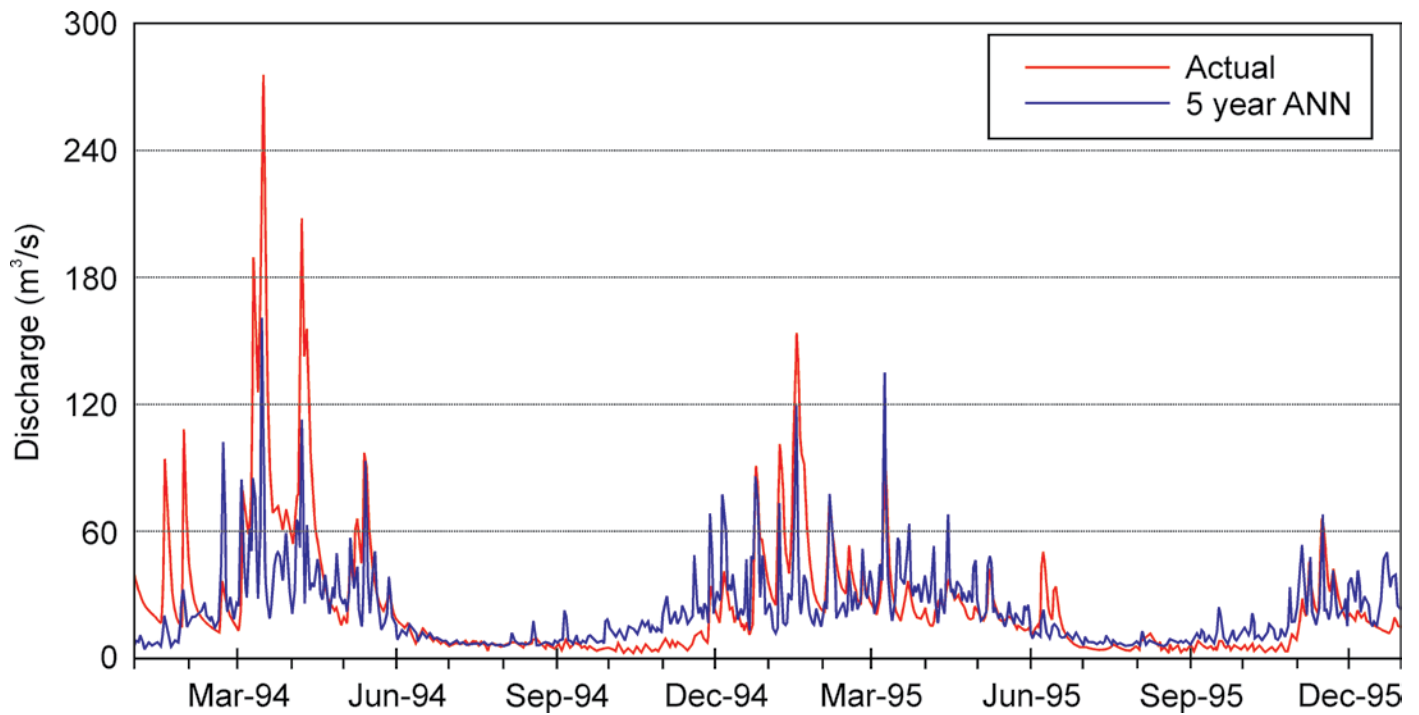

Fig. 5. Annapolis River at Lawrencetown actual discharge vs. predicted discharge for a 5-year STL model without transfer.

Domain Knowledge Models for the Secondary Streams

The objective of this experiment was to develop STL and MTL DK models for the secondary streams to be saved for transfer to the primary task. A secondary objective was to compare the accuracy of models for these streams. The model for the Shubenacadie River is of particular interest given that it is the most distant from the source of weather data. All models used five years of data for training (1986–1990), one year for validation (1991), and two years for testing (1992/93 and 1994), as shown in Table 3. STL models were constructed for Sharpe Brook, the ARW, and the Shubenacadie River. An MTL model was also constructed for Sharpe Brook and the ARW to observe the effect of simultaneous transfer from two prior models. The STL models used 20 hidden nodes, while the MTL models used 30 hidden nodes. The four models were saved as separate DK for later use in inductive transfer.

Table 4 shows that the correlation performance for all models is at the same level (~ 0.7) and near that of the five-year STL models for the ARL. The variance in the mean absolute discharge error (m³/sec) between the streams is due to variability in stream catchments (Sharpe Brook watershed is much smaller than the others considered in this study). The MTL models developed for Sharpe Brook and the ARW are not significantly better or worse than the individual STL models. The Shubenacadie River models have the lowest correlation of all models at 0.671. Graphs of actual versus predicted discharge (not shown) reveal that the Shubenacadie River is somewhat different from the other streams, and that the associated models suffer from a number of over-predictions made in the summer and fall. This result is likely because the weather data used was from Greenwood which is a significant distance from the Shubenacadie River site.

MTL Inductive Transfer Models for the Annapolis River at Lawrencetown (ARL) Gauging Site

The purpose of this experiment was to compare models of the ARL using transfer from each of the four DK models of the secondary streams developed in the previous section with that of the STL models presented in the last section. The same training, validation and test data used during STL model development were used for this experiment. All modeling was done using MTL networks which were initialized with the representation of a previously learned DK model. An additional output is added to the network for the target ARL task. The models developed from STL DK used 20 hidden nodes and two outputs, for a 25-20-2 network, while the models developed from the MTL DK used 30 hidden nodes, with a total of three outputs, for a 25-30-3 network.

The performance results are presented in Figure 5. All models developed under MTL with inductive transfer using 180 days, one year, and three years of training data performed significantly better than the associated STL models and no inductive transfer. MTL models developed with five years of data

Table 4. Performance of domain knowledge models on their respective test sets.

Stream	MAE	Correlation
Sharpe Brook STL	0.131	0.702
Annapolis River at Wilmot STL	6.754	0.709
Sharpe Brook MTL	0.140	0.707
Annapolis River at Wilmot MTL	6.721	0.706
Shubenacadie River STL	7.149	0.671

performed as well as or better than the associated STL models. All sources of transfer are beneficial, with the ARW as the best source and the Shubenacadie River as the worst.

DISCUSSION

The objective of this research is to determine the value of inductive transfer as applied to the modeling of discharge. The target stream (ARL) is typical of larger, 4th order and higher valley bottom streams in Atlantic Canada for which accurate discharge models are required. Our hypothesis was that previously developed models for discharge of a distinct gauging site (ARW) or stream (Sharpe Brook) can be used as a source of inductive transfer for a distinct but spatially and geomorphologically related stream (ARL). We also tested whether an unrelated stream (Shubenacadie River) in a distinct physiographic region could be used as a source of inductive transfer. In all cases, our results indicated that fairly accurate models can be developed.

It is important to compare the consistent performance of the models developed under transfer to those without transfer. The performance of MTL models varies only slightly as the amount of training data changes, particularly when the source of transfer is the ARW (see Figure 5). Clearly this site is the best source of prior knowledge for developing models for the ARL. This result makes sense as the two locations are closest to each other and are on the same river system.

The models for the ARL developed using the Sharpe Brook and Shubenacadie River as sources of inductive transfer do not perform quite as well as models developed with transfer from the ARW (Fig. 5). Both Sharpe Brook and the Shubenacadie River are in different drainage basins than the ARL. Sharpe Brook is a significantly smaller stream than the ARL and the Shubenacadie River is the more distant stream. This result supports the theory that more related prior knowledge leads to more beneficial inductive transfer. Combining Sharpe Brook with the ARL as a source of transfer creates models that perform better than models with transfer from only Sharpe Brook, but worse than models with transfer from only the ARW.

Although the ARL model developed with 180 days of training data and transfer from ARW (Fig. 6) does statistically as well as the STL model developed with 5 years of training data (Fig. 4), the model with transfer tends to be more conservative in its predictions. The MTL transfer model does not predict the
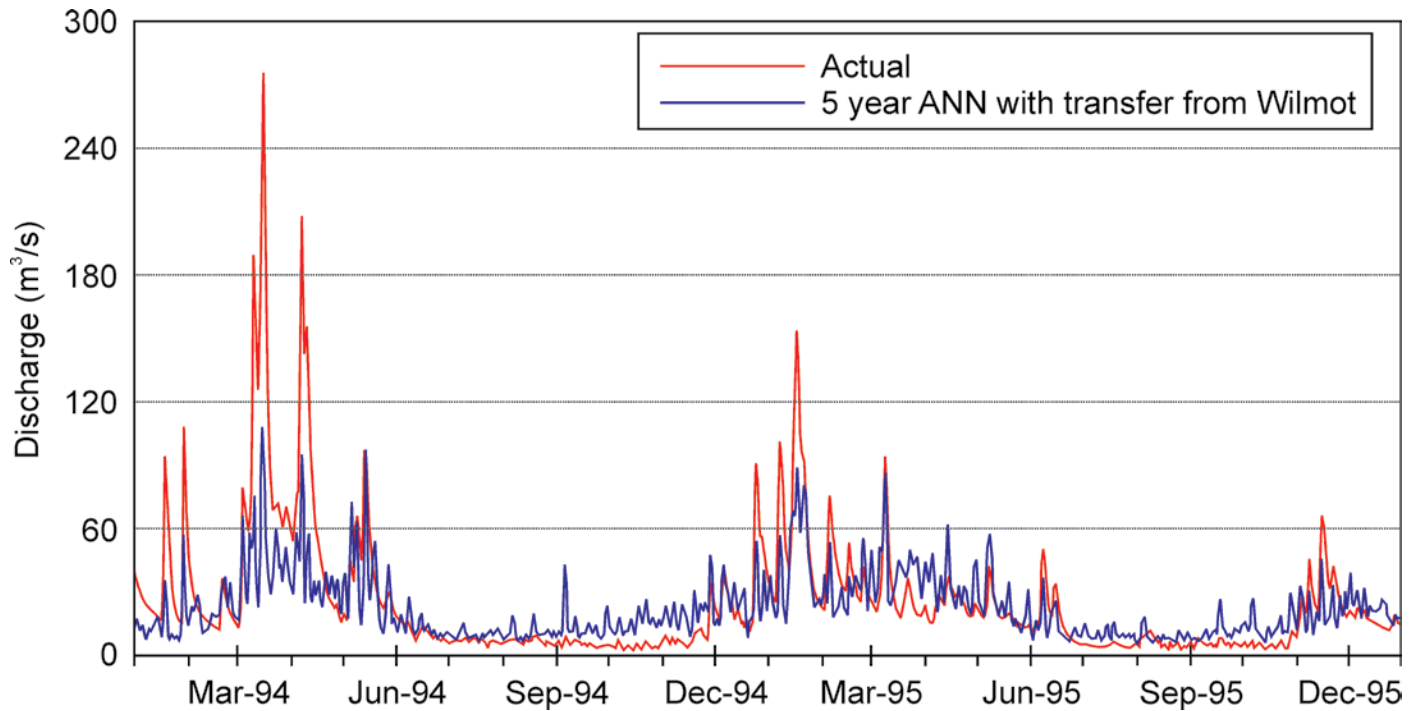
Fig. 6. Actual versus predicted discharge for a model of Annapolis River at Lawrencetown developed with 180 days of training data and with transfer from the Annapolis River at Wilmot.

highest discharge value (March 1994) as accurately as the STL model based on five years of training data. Two reasons for this are (1) portions of the chosen test set are atypical as they contain some very high discharge values, and (2) the models used only two days of weather data as input. The peak discharge recorded in March, 1994 (276 m³/s) for the ARL is the highest value recorded for all streams in the study. By comparison, the highest value in the 180 day training set is 168 m³/s, while the highest value in all other training sets is 205 m³/s recorded in 1987 (Government of Canada 2005b). Because none of the models are trained with target outputs of this magnitude the models perform poorly on the 1994 portion of the test set. This explains why models developed from the 180 day training set tend to under-predict peak values as compared to models developed with 5 year training sets. If the performance of the best models are examined using only the 1995 portion of the test set, the MAE decreases from 12.6 m³/s to less than 10 m³/s. This occurs because the highest value recorded in 1995 was a more typical 154 m³/s.

The second reason why the models err on extreme discharge values is that only two days of weather data are used and knowledge of recent discharge levels is not provided. This can lead to under-prediction when the stream level is at bankfull stage or higher (typically in the spring and late fall) and over-prediction when the stream level is at or below baseflow stage (typically in the summer). An efficient method of adding more days of weather data and knowledge of recent stream discharge would increase the performance of the models.

Based on the above, one could argue that the models devel-

oped with inductive transfer do better at predicting when an extreme in stream discharge will occur than they do in predicting the magnitudes of those extremes. Despite doing well in the normal range, the graphs show that the MTL models under-predict a number of the peak discharge values. Nonetheless, the models provide a significant performance advantage over the less accurate alternatives; models developed without transfer using standard STL. Therefore, we propose that the MTL models provide a valuable starting point when little data are available. As time passes, new data collected for the primary stream can be combined with the existing data so as to continually develop more accurate models.

CONCLUSIONS

The results of the research supports the hypothesis that inductive transfer of previously learned knowledge can reduce the number of years of training examples required to construct accurate models of stream discharge. Inductive transfer via MTL and task rehearsal with as little as 180 days of training data for the primary task produces predictive models that are statistically equivalent to models developed from five years of training data and no transfer. The experiments also demonstrated that the more related prior models of stream discharge (the Annapolis River at Wilmot) generated the best models for the target task (discharge at the Annapolis River at Lawrencetown). We conclude that inductive transfer methods should be considered when modeling river discharge and that

value exists in the systematic retention and reuse of models from any environmental problem domain.

Several interesting directions for future work have been identified. More accurate models could be developed by increasing the amount of weather data used as input beyond two days. Extending this window of weather data should increase the accuracy of the models. Keeping in mind that additional inputs increase the sample complexity and training times, we are currently investigating the use of recurrent neural networks that can maintain a sense of context while only requiring a single day's weather as input (Mitchell 1997). A second approach to improving model performance is to use the known discharge of one stream as an input to the model for another stream. For example, the discharge of the ARW, which is constantly monitored, could be used as an input to a model for the ARL. This approach would allow all streams in a drainage basin to benefit from monitoring one stream.

We have also identified several questions regarding the limits of inductive transfer applied to stream discharge and more generally to environmental modeling. Can we use data from large streams that have been monitored for some time to better model smaller secondary streams for which we have little data? Our intention is to explore this in future work. Can prior knowledge of a model in one region, using weather data from that region, be used to transfer knowledge to a stream in a different region, using weather data from that second region? There is no reason why this cannot be the case. An issue related to both of these questions is determining a measure of relatedness between streams and their associated watersheds and climatic conditions to ensure the approach is beneficial to the development of accurate models.

ACKNOWLEDGMENTS

REFERENCES

Bastarache, D., El-Jabi, N., and Clair, T.A. 1997. Predicting conductivity and acidity for small streams using neural networks. Canadian Journal of Civil Engineering, 24, pp. 1030–1039. doi:10.1139/cjce-24-6-1030

Baxter, J. 1996. Learning model bias. *In* Advances in neural information processing systems. *Edited by* M. Mozer, D. Touretzky, and M. Perrone. The MIT Press, Cambridge, Massachusetts, USA. pp. 169–175.

Bhattacharya, B., and Solomatine, D.P. 2003. Neural networks and M5 model trees in modeling water level-discharge relationship for an Indian river. *In* European Symposium on Artificial Neural Networks 2003 Proceedings. Bruges, Belgium, pp. 407–412.

Blackmore, A. 2007. An evaluation of the basic geochemistry of ground and surface waters in Kingston, Nova Scotia and its interaction with the surficial geology and land use practices of the surrounding watershed. Unpublished M.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 249 p.

Caruana, R. A. 1997. Multitask learning. Machine Learning, 28, pp. 41–75. doi:10.1023/A:1007379606734

Cheng, R.T., Costa, J.E., Haeni, F.P., Melcher, N.B., and Thurman, E.M. 2002. In search of technologies for monitoring river discharge. *In* Advances in water monitoring research. *Edited by* T. Younos. Water Resources Publications LLC, pp. 203–219.

Cook, S.B. 2000. An assessment of in-stream restoration structures on tertiary streams in Annapolis Valley, Nova Scotia. Unpublished B.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 77 p.

Costa, J.E., Spicer, K.R., Cheng, R.T., Haeni, FP, Melcher, N.B., Thurman, E.M., Plant, W.J., and Keller, W.C. 2000. Measuring stream discharge by non-contact methods—a proof-of-concept experiment. Geophysical Research Letters, 27, pp. 553–556. doi:10.1029/1999GL006087

Davis, D., and Browne, S. 1996. Natural history of Nova Scotia volume II. Nova Scotia Museum of Natural History and Nimbus Publishing, Halifax, Nova Scotia, 312 p.

Dawson, C.W., and Wilby, R.L. 2001. Hydrological modeling using artificial neural networks. Progress in Physical Geography, 25, pp. 80–108.

Dibike Y.B., and Solomatine D.P. 2001. River flow forecasting using ANNs. Journal of Physics and Chemistry of the Earth, Part B: Hydrology. Oceans and Atmosphere, 26, pp. 1–8.

Fenton, H. 1998. A model of storage potential and input transferral patterns of the Elderkin Brook Watershed in Kings County, Nova Scotia. Unpublished B.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 53 p.

Gauthier, M.-J. 2008. Étude des interactions entre l'eau de surface et l'eau souterraine aux échelles locale et regional dans la vallée d'Annapolis, Nouvelle-Écosse. Unpublished M.Sc. thesis. Université du Québec, National Institute of Scientific Research, Water, Earth and Environment Centre (INRS-ETE), 163 p.

Government of Canada. 2005a. Environment Canada Weather Office. National climate data and information archive. URL <http://www.climate.weatheroffice.ec.gc.ca> July 2005.

Government of Canada. 2005b. Water survey of Canada. URL <http://www.wsc.ec.gc.ca> July 2005.

Government of Nova Scotia. 2002. Nova Scotia statistical review 2002. Nova Scotia Department of Finance, Statistics Division. URL <http://www.gov.ns.ca/finance/publish/statsrev/sr2002.pdf> July 2005.

Hamblin, A.P. 2004. Regional geology and hydrogeology potential of the Triassic/Jurassic Fundy Group, Annapolis-Cornwallis Valley, Nova Scotia. Geological Survey of Canada Open File 4678, 30 p.

Hennigar, T. W., Gibb, J., Cruickshanks, F., and Rothfischer,

J. 1992. Hydrogeological and Groundwater Interests in the Annapolis-Cornwallis Valley. Geological Association of Canada - Mineralogical Association of Canada, Joint Annual Meeting, Wolfville, '92. Field Trip C-4, Guidebook. 29 p.

Jain, S.K., and Chalisgaonkar, D. 2000. Setting up stage-discharge relations using Artificial Neural Networks. Journal of Hydrologic Engineering, 5, pp. 428–433. doi:10.1061/(ASCE)1084-0699(2000)5:4(428)

Kompare, B., Steinman, F., Cerar, U., and Dzeroski, S. 1997. Prediction of rainfall runoff from catchment by intelligent data analysis with machine learning tools within the artificial intelligence tools. Acta hydrotechnica (in Slovene language) 16/17, pp. 79–94.

Levy, A.L. 1998. A short term watershed analysis of Fishwick Brook in Berwick, Nova Scotia during a drought. Unpublished B.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 67 p.

Mitchell, T. M. 1997. Machine learning. McGraw Hill, New York, New York, USA, 414 p.

ML3. 2009. Machine life-long learning website. URL <http://ml3.acadiau.ca/>, July 2009, (contains a link to the RASL3 software which can be freely downloaded).

Muttiah, R.S., Srinivasan, R., and Allen P.M. 1997. Prediction of two-year peak stream discharges using neural networks. Journal of the American Water Resources Association, 33, pp. 625–630. doi:10.1111/j.1752-1688.1997.tb03537.x

Myers, M.S. 1997. A detailed investigation of the morphological impact of in-stream remediation structures on Mill Brook, Kings County, Nova Scotia. Unpublished B.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 69 p.

Neily, P.D., Quigley, E., Benjamin, L., Stewart, B., and Duke, T. 2003. Ecological land classification for Nova Scotia, Volume 1 – mapping Nova Scotia's terrestrial ecosystems. Report DNR 2003-2. Nova Scotia Department of Natural Resources, Renewable Resources Branch, Stewiacke, Nova Scotia, 83 p.

O'Quinn, R., Silver, D., and Poirier, R. 2005. Continued practice and consolidation of a learning task. *In* Proceedings of the meta-learning workshop, international conference on machine learning (ICML 2005), Bonn, Germany, pp. 60–67.

Poirier, R., and Silver, D. L. 2004. Sequential consolidation of learned task knowledge. *In* Advances in artificial intelligence, 17th conference of the Canadian Society for Computational Studies of Intelligence. *Edited by* A. Tawfik and S. Goodwin. Springer-Verlag, Berlin Heidelberg, Germany, pp. 217–232.

Rivard, C., Ross, M., Michaud, Y., Hamblin, T., Drage, J., Blackmore, A., Webster, T., Paniconi, C., and Deblonde, C. 2004. Preliminary hydrogeological characterization of the Annapolis Valley, Nova Scotia. URL <http://cgcq.rncan.gc.ca/Annapolis/projet/documents/articles/AIH_2004_93585.pdf> July 2009.

Rivard, C., Deblonde, C., Boivin, R., Bolduc, A., Paradis, S.J., Paradis, D., Lioa, S., Gauthier, M.- J., Blackmore, A., Trepanier, S., Castonguay, S., Drage, J., and Michaud, Y. 2007. Canadian groundwater inventory: hydrogeological atlas of the Annapolis Valley, Nova Scotia. Geological Survey of Canada Open File 5541, 128 p.

Sharma, N., Sankhua, R.N., and Pandey, A.P. 2005. Spatio-temporal modeling of hydrological variability for the river Brahmaputra using artificial neural networks. Role of Water Sciences in Transboundary River Basin Management, Thailand, 2005 pp. 25–31.

Silver, D. L., and Mercer, R. E. 2002. The task rehearsal method of life-long learning: overcoming impoverished data. *In* Advances in artificial intelligence, 15th conference of the Canadian Society for Computational Studies of Intelligence. *Edited by* R. Cohen and B. Spencer. Springer-Verlag, Berlin Heidelberg, Germany, pp. 90–101.

Silver, D., and Bennett, K. 2008. Guest editorial: inductive transfer learning. Machine Learning, 73, pp. 215–220. doi:10.1007/s10994-008-5087-1

Stea, R. R. 1987. Quaternary glaciations, geomorphology, and sea-level changes: Bay of Fundy region. NATO ASI-IGCP Project 200: Later Quaternary Sea Level Correlations and Applications, Field Trip Guidebook. Halifax, Nova Scotia, pp. 1–34.

Stea, R. R., Mott, R.J., Belknap, D.F., and Radtke, U. 1992. The pre-Late Wisconsin chronology of Nova Scotia, Canada. *In* The last interglaciation/glaciation transition in North America. *Edited by* P.U. Clark and P.D. Lea. Geological Society of America, Special Paper 270, pp. 185–205.

Stiff, D. 2007. Investigating flood risk in an ungauged watershed using Lidar, GIS, and HEC tools. Unpublished M.Sc. Thesis, Acadia University, Wolfville, Nova Scotia, Canada, 168 p.

Thrun, S. 1997. Lifelong learning algorithms. Learning to learn, Kluwer Academic Publisher, Norwell, Massachusetts, USA, pp. 181–209.

Trépanier, S. 2008. Caractérisation et modélisation de la contamination de l'eau souterraine dans un sous-bassin de la vallée d'Annapolis (Nouvelle-Écosse), Unpublished M.Sc. Thesis, Université du Québec à Montréal, Montréal, Québec, Canada, 123 p.

Trescott, P. C. 1967. An investigation of the groundwater resources of the Annapolis-Cornwallis Valley, Nova Scotia. Unpublished Ph.D. Thesis, Graduate College, University of Illinois, Urbana, Illinois, USA, 269 p.

Trescott, P. C. 1968. Groundwater resources and hydrogeology of the Annapolis-Cornwallis Valley. Memoir 6, Nova Scotia, Province of Nova Scotia Department of Mines, Halifax, Nova Scotia, 159 p.

Editorial responsibility: Sandra M. Barr