Inductive Transfer Applied to Stream Discharge Modeling

Daniel L. Silver, Lisa Gaudette and Ian Spooner

Abstract—**Artificial neural networks and inductive transfer are used to develop models that predict the discharge (flow rate) of fresh water streams in Nova Scotia from weather data. The objective is to show that transfer can be used to reduce the time and cost associated with collecting large amounts of the data for environmental modeling. The models use two days of weather data to predict the discharge for the following day. The models can be applied to land use, water management and flood predictions for sections of streams where continuous monitoring is not feasible. 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.**

I. Introduction

The collection of data for modeling environmental phenomenon typically requires a lot of time and money. Data gathering equipment can be expensive and the on-going labor costs can be high. From a modeling perspective, time can be the most significant problem because several years' worth of data, that captures variations in seasonal trends, is often required to produce good predictive and descriptive models. For example, five or more years worth of weather and stream data are required to produce sufficiently accurate models of stream flow rate, or discharge, under standard statistical and machine learning methods. There would be considerable value in constructing models of stream discharge from only a years worth of data.

To reduce the time and cost of collecting years of data, this research proposes the use of previously learned knowledge of related streams as a source of inductive bias when developing a model for a new target stream. The work involves the first use of our artificial neural network based inductive transfer system on a real-valued problem (as opposed to a classification problem) [8]. Multiple-task learning (MTL) and task rehearsal is used to transfer knowledge from previously learned discharge models of one or more streams to the model for a new stream [10].

The streams used in this research are all located in Nova Scotia, Canada. The models make short-term predictions - two consecutive days of weather data are used as input, and stream discharge is predicted for the following day. This is not a trivial undertaking as the mapping between weather data and stream flow rate is surprisingly complex. And such a model provides substantial lead time for local authorities to plan emergency water management measures or to prepare for a potential flood situation.

Manuscript received March 16, 2007. This work was supported in part by Acadia Univeristy and the National Science and Engineering Research Council of Canada.

Danny Silver and Lisa Gaudette are with the Jodrey School of Computre Science, Acadia University, Wolfville, NS, Canada, B4P 2R6 (phone: 902-585-1105; fax: 902-585-1067; email: danny.silver@acadiau.ca).

Ian Spooner is with the Department of Geology, Acadia University, Wolfville, NS, Canada, B4P 2R6 (email: ian.spooner@acadiau.ca).

II. Background

A. Streams and Weather Systems

While only 0.016% of the Earth's water is in streams and lakes at any given time they are important components of the earth's hydrological cycle [6]. The water in streams has a significant impact on the Earth's surface and on the plants, animals and humans that live near them. Streams are used for many purposes, such as transportation, recreation, irrigation, and groundwater recharge. Consequently, streams can cause problems when their levels rise or fall abnormally.

Formally, a stream is "any body of flowing water confined within a channel, regardless of size" [6]. The water in a stream comes from an area known as its drainage basin. The discharge of a stream is the volume of water flowing through a given cross section in a specified length of time, generally measured in cubic meters per second (m3/s). Stream levels, and therefore discharge, can vary considerably by season and by year. In the area studied, the general pattern is that the highest levels occur in the spring, with the lowest levels occurring in the summer months. Flooding occurs, but is rare. Typically, a small flood may occur as often as every 2-3 years, with more severe floods being less common.

It is important to emphasize that water enters streams in ways other than direct precipitation. For a few days after a precipitation event, water could still be entering the stream after percolating through the ground or through surface runoff, and the effect of snow melt is significant during the spring and warmer times during the winter. These factors throughout the drainage basin influence the stream. Streams will respond to input events in different time frames, based on factors such as the size of the drainage basin, geology, soil types, and the dampness of the soil [6]. A mountain stream with a small drainage basin over rocky terrain or frozen soil may be prone to sudden flash flooding within hours of an input event, while a river in a flatter area with more absorbent soil may take days to respond.

B. Inductive Transfer and MTL Neural Networks

The constraint on a learning system's hypothesis space, beyond the criterion of consistency with the training examples, is called inductive bias [5]. Inductive bias is essential for the development of a hypothesis with good generalization from a practical number of examples. Ideally, a life-long learning system can select its inductive bias to tailor the preference for hypotheses according to the task being learned [11]. One type of inductive bias is prior knowledge of the task domain.

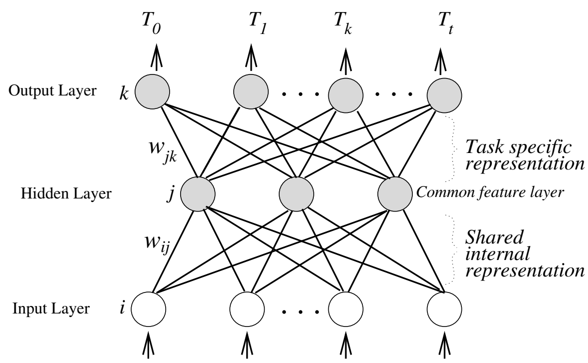

Fig. 1. A multiple task learning (MTL) network with an output node for each task being learned in parallel

The retention and use of task domain knowledge (DK) as a source of inductive bias has become known as inductive transfer and remains an open problem in machine learning [4,11]. The goal of research in inductive bias is to find ways of using prior knowledge to develop more accurate hypothesis (models) with fewer training examples as quickly and efficiently as possible [10].

1) Knowledge Transfer in MTL Networks: Multiple task learning (MTL) neural networks are one of the better documented methods of inductive transfer [4]. 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. Consequently, MTL training examples are composed of a set of input attributes and a target output for each task. Figure 1 shows a 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 [1]. The more tasks are related the more they will share representation and create positive inductive bias.

2) Sequential Learning through Task Rehearsal: In [10] the task rehearsal method, which builds on the theory of pseudo-rehearsal [9], was introduced as a 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 representations 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 [7,8]. It is through the sharing of internal representation and the rehearsal of previously learned tasks that knowledge is transferred to the new task.

III. Theory and Approach to Model Development

The objective of this research is to examine the value of inductive transfer applied to an environmental modeling problem. Our hypothesis is that a previously developed model of discharge for stream A can be used as a source of transfer when developing a model for related stream B. Through the use of prior model knowledge and inductive transfer, fewer years of training data will be required to construct accurate models. The more related the prior model is to the target task, the greater the expected benefit from the transfer. More specifically, the sources of knowledge transfer are stream discharge models of the Annapolis River at Wilmot and Sharpe Brook both near Greenwood, Nova Scotia, and the Shubenacadie River near Enfield, Nova Scotia. The primary target task is the Annapolis River at Lawrencetown. Details of these streams are presented in the next section.

To demonstrate the value of inductive transfer, the following approach is taken. First, standard single task learning (STL) models are created for the Annapolis River at Lawrencetown for different amounts of data from 180 days to five years. Next, four domain knowledge (DK) models are constructed: (1) for the Annapolis River at Wilmot, located upstream of the Annapolis River at Lawrencetown and having half the discharge, (2) for Sharpe Brook, located nearby but in a different drainage basin, (3) for the previous two streams together in an MTL network, and (4) for the Shubenacadie River at Enfield, located in a more distant drainage basin with somewhat different climate. Finally, the four DK models are each used as a source of transfer for learning the primary task in MTL networks. To ensure a fair comparison, the models developed with inductive transfer are trained with the same data used to develop the STL models.

The models are given two days of weather data from Greenwood, NS, to predict the discharge for the following day. The STL and MTL models are 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 reports 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.

IV. Data Collection and Preparation

Figure 2 shows the locations of the four streams (Lawrencetown, Wilmot, Sharpe Brook and Shubenacadie) and the source of the weather data (Greenwood). Table 1 and

2 presents details of the weather data and the statistics of streams, respectfully. The weather and stream data was obtained from two distinct sources within Environment Canada for the years 1986-1995. The weather data comes from the on-line National Climate Data and Information Archive [2]. It was received in hourly, daily and monthly formats depending upon the variable. For the purposes of this study all data was converted to daily values (total, mean, maximum or minimum). Table 1 provides a listing of the various weather parameters used in the study.

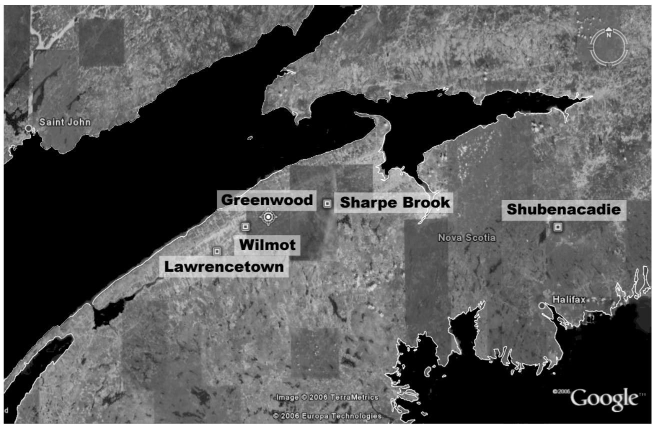

Fig. 2. Locations of streams and the Greenwood weather station (Source: Google Earth).

The weather data was reasonably complete. Where necessary, missing values were imputed using the average of the previous and next day values. The period from December 1992 to November 1993 had a large amount of missing weather data. The decision was taken to not use data from this period. Consequently, the year labeled "1992/93" in Table 2 is composed of data from January to November, 1992, and data from December, 1993.

Discharge data for several streams was obtained through the on-line Water Survey of Canada "HYDAT" database [3]. Stream discharge data is currently being recorded at nearly 3000 sites across Canada, and is available in near-real time from just under half of those sites. The Annapolis at Wilmot 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, while the Annapolis at Lawrencetown 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 was 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 contain 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, stream discharge for the Annapolis River at Lawrencetown. For the primary task, training sets containing five years, three years, one year, and 180 days of data are used, along with one year of validation data and two years of test data. For the DK models constructed as sources of transfer, five years of training data are used, with one year of validation data and two years of test data. The data used for Lawrencetown is shifted forward a year so as to challenge the method to create novel virtual examples for rehearsal of the secondary tasks. Table 3 shows the years of data used for each stream. For Lawrencetown, the five year period consists of January 1st 1987 to December 31st 1991; the three year period is from 1987 to 1989, the one year period is 1987, and the 180 days are selected at random from 1987.

V. Model Development and Analysis

This section covers the development and comparison of the various models as outlined in the approach described in section 3.

A. Neural Network Architecture and Learning Parameters

Three layer networks were used for all models. 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 0.0-0.1 range.

B. STL Models for the Annapolis River at Lawrencetown

1) Objective: The purpose of this experiment is to develop models using single task learning (STL) neural networks with varying amounts of data for the primary stream, the Annapolis River at Lawrencetown. These models will be compared to each other and later to models constructed with inductive transfer using prior knowledge.

2) Method: This experiment uses the four different training sets described in section 4 and shown in Table 3. The models are developed using a 25-20-1 network. Five repetitions are performed for each set of training data with different random initial weights. Up to 1.5 million training iterations were allowed for each repetition.

3) Results: Figure 6 presents the performance (and 99% confidence intervals) 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 is 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.

Figure 4 and 5 compare the graphs of the actual vs. predicted 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.

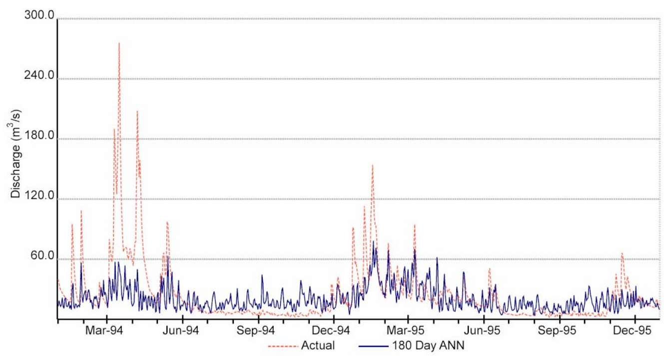

Fig. 4. Lawrencetown actual vs. predicted Discharge for a 180 day STL model.

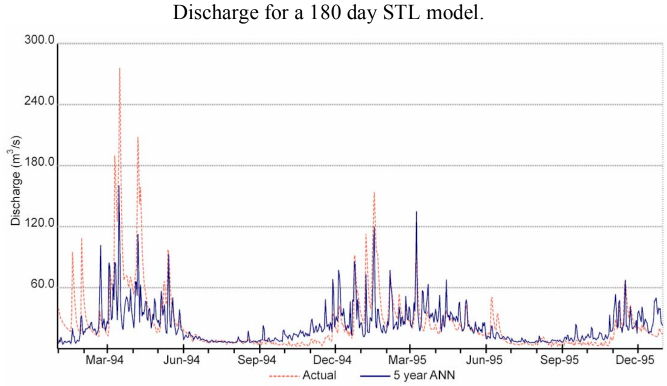

Fig. 5. Lawrencetown actual vs. predicted discharge for a 5 year STL model.

C. Domain Knowledge Models for the Secondary Streams

1) Objective: The objective of this experiment is to develop STL and MTL DK models for the secondary streams to be saved for transfer to the primary task. A secondary objective is 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.

2) Method: All models use 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 2. STL models were constructed for Sharpe Brook, the Annapolis River at Wilmot, and the Shubenacadie River. An MTL model was also constructed for Sharpe Brook and the Annapolis River at Wilmot. The STL models use 20 hidden nodes, while the MTL models use 30 hidden nodes. The four models were saved as separate DK for later use in inductive transfer.

3) Results: 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 Annapolis River at Lawrence. The MTL models developed for Sharpe Brook and Wilmot are not significantly better or worse than the individual STL models. The Shubenacadie 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.

D. MTL Inductive Transfer Models for the Annapolis River at Lawrencetown

1) Objective: The purpose of this experiment is to compare models of the Annapolis River at Lawrencetown using transfer from each of the four DK models of the secondary streams developed in the previous section.

2) Method: The same training, validation, and test data used during STL model development described in section 5.2 are used for this experiment. All modeling is done using MTL networks which are initialized with the representation of a previously learned DK model. An additional output is added to the network for the target Shubenacadie task. The models developed from STL DKs use 20 hidden nodes and two outputs, for a 25-20-2 network, while the models developed from the MTL DK uses 30 hidden nodes, with a total of three outputs, for a 25-30-3 network.

3) Results: The performance results are presented in Figure 6. All models developed under MTL with inductive transfer using 180 days, one year, and three years of training data perform significantly better than the associated STL models and no transfer. MTL models developed with five years of data performed as well as or better than the associated STL models. All sources of transfer are beneficial, with Wilmot as the best source and Shubenacadie as the worst.

Figures 7 shows a graph of the actual versus predicted discharge for a model of Annapolis River at Lawrencetown developed with 180 days of training data and with transfer from Wilmot.

4) Discussion: It is important to compare the consistent performance of the models developed under transfer relative to those without transfer. There is only a small variation in the performance of models as the amount of training data changes, particularly when the source of transfer is the Annapolis River at Wilmot.

The models for Annapolis at Lawrencetown developed using the Sharpe Brook and Shubenacadie River as sources of transfer do not perform quite as well as models developed with transfer from Annapolis at Wilmot (see Figure 6). Both Sharpe Brook and Shubenacadie River are in different drainage basins than Lawrencetown. Sharpe Brook is a significantly smaller stream than Lawrencetown and Shubenacadie is the more distant stream. This supports the theory that more related prior knowledge results in more beneficial inductive transfer. Combining Sharpe Brook with Wilmot 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 Wilmot.

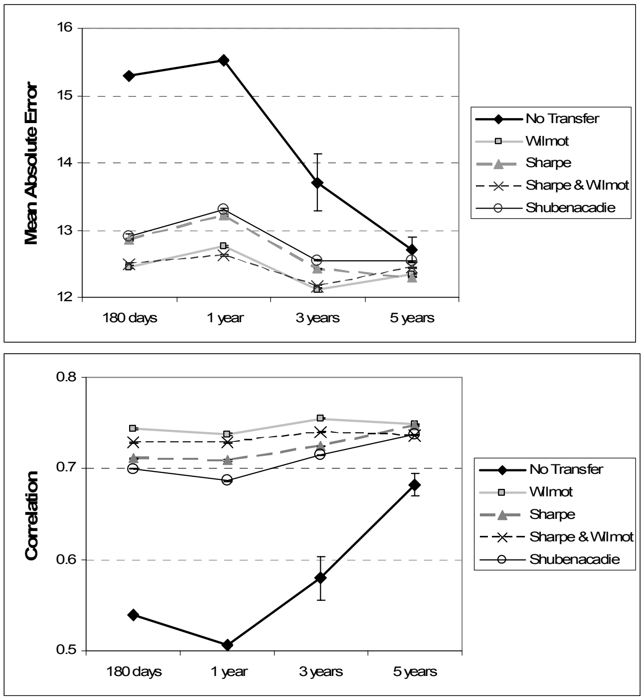

Fig. 6. Performance of Annapolis at Lawrencetown models on the test set with and without transfer as a function of the amount of training data.

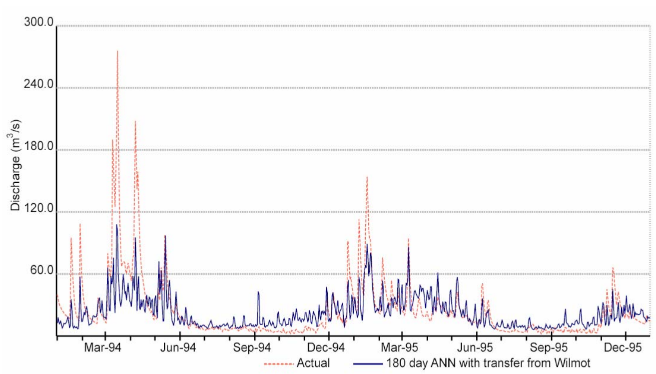

Fig. 7. Actual vs. predicted discharge; 180 day training set with transfer from Wilmot.

Although the model developed with 180 days of training data and transfer from Wilmot (see Figure 7) does statistically as well as the STL model (see Figure 5), the model with transfer tends to be more conservative in its predictions. It does not predict the highest value, in March 1994, as well as the STL model based on five years of training data. There are at least two reasons for this: (1) portions of the chosen test set are particularly difficult as they contain some very high discharge values, and (2) the models use only two days of weather data as input. The peak discharge recorded in 1994 (276 m3/s) for Lawrencetown 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. 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 is used and knowledge of recent discharge levels is not provided. This can lead to under-prediction when the stream level is abnormally high (typically in the spring) and over-prediction when the stream level is abnormally low (typically in the fall and winter). An efficient method of adding more days of weather data and knowledge of recent stream discharge would increase the performance of the models.

VI. Conclusions

The results of the experiments support our 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 produces 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 (Wilmot) generated the best models for the target task (Lawrencetown). We conclude that inductive transfer methods should be considered when modeling environmental problems and that there is value in the systematic retention and reuse of models from an 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 [5]. 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 Annapolis River at Wilmot, which is constantly monitored, could be used as an input to a model for the Annapolis River at Lawrencetown. 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 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? A related question is determining how similar the streams and climatic conditions must be in order for this approach to be beneficial.

REFERENCES

[1] Baxter, J. (1996). Learning model bias. Advances in Neural Information Processing Systems, (pp. 169–175). The MIT Press.

[2] Government of Canada. Environment Canada Weather Office. National Climate Data and Information Archive. Retrieved July 2005 from http://www.climate.weatheroffice.ec.gc.ca

[3] Government of Canada. Water Survey of Canada. Retrieved July 2005 from http://www.wsc.ec.gc.ca.

[4] Caruana, R. A. (1997). Multitask learning. Machine Learning, 28, 41–75.

[5] Mitchell, T. M. (1997). Machine learning. New York, NY: McGraw Hill.

[6] Montgomery, C.W. Environmental Geology (6th Edition). McGraw Hill. 2002.

[7] O'Quinn, R. and Silver, D. and Poirier, R. 2005. Continued Practice and Consolidation of a Learning Task. Proceedings of the Meta-Learning Workshop, International Conference on Machine Learning (ICML 2005), Bonn, Germany, 7-11 Aug, 2005, p. 60-67.

[8] Poirier, R., & Silver, D. L. (2004). Sequential consolidation of learned task knowledge. Lecture Notes in Artificial Intelligence, 17th Conference of the Canadian Society for Computational Studies of Intelligence (AI'2004), 217–232.

[9] R. Robins, A. V. (1995). Catastrophic forgetting, rehearsal, and pseudo-rehearsal. Connection Science, 7, 123–146.

[10] Silver, D. L., & Mercer, R. E. (2002). The task rehearsal method of life-long learning: Overcoming impoverished data. Advances in Artificial Intelligence, 15th Conference of the Canadian Society for Computational Studies of Intelligence (AI'2002), 90–101.

[11] T. Thrun, S. (1997). Lifelong learning algorithms. Learning to Learn, 181–209.1 Kluwer Academic Publisher.

TABLE I
DETAILS OF WEATHER AND DISCHARGE VARIABLES.

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 II
SUMMARY OF STREAM STATISTICS.

Name	Distance to Greenwood km	Drainage Area Km^2	Average Discharge m^3/s	Standard Deviation	Maximum Discharge m^3/s
Lawrencetown	22	1020.00	23.68	25.21	276
Wilmot	9	546.00	12.26	13.83	140
Sharpe Brook	23	8.81	0.22	0.30	4.79
Shubenacadie	109	389.00	12.40	11.21	75.3

TABLE III
YEARS USED FOR TRAINING, VALIDATION, AND TEST DATASETS.

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

TABLE IV
PERFORMANCE OF DOMAIN KNOWLEDGE MODELS ON THE TEST SET.

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