Understanding the Range of Applicability of Artificial Neural Network Forecasting Models

G.J. Bowden, H.R. Maier and G.C. Dandy

Centre for Applied Modelling in Water Engineering, Department of Civil and Environmental Engineering, Adelaide University, Adelaide SA, Australia (gbowden@civeng.adelaide.edu.au)

Abstract: When an operational artificial neural network (ANN) model is deployed, new input patterns are collected in order to make real-time forecasts. However, ANNs (like other empirical and statistical methods) are unable to reliably extrapolate beyond the calibration range. Consequently, there is a need to determine if these new input patterns are similar to the input data used in training the model. In order to address this problem, a novel hybrid forecasting model consisting of a Self-Organizing Map (SOM) and a backpropagation ANN (BPN) is presented. The SOM combines each new input pattern with the training data and determines if the new pattern corresponds to patterns within the training set and the BPN is used to obtain the forecast. In this way, it is possible to define the range of applicability of the model and there is likely to be greater confidence in a forecast resulting from an input pattern that is similar to those in the training data. A case study is presented in which an ANN model is developed to forecast salinity in the River Murray at Murray Bridge (South Australia) 14 days in advance. The model is developed using 6 years of daily salinity, flow and river level data. Once the model has been developed, it is then combined with the front-end SOM classifier and used as an operational model. The operational model is then trialled on a further 3-year period to determine whether the SOM is capable of indicating when the model will fail to generalise. The results indicate that this approach is very successful and it is clear that the method has widespread potential for developing confidence in real-time forecasting/prediction applications.

Keywords: Artificial Neural Networks; Self-Organizing Map; Forecasting; Water Quality; Salinity Modelling

1. INTRODUCTION

In the past decade, artificial neural networks (ANNs) have risen to prominence as a viable alternative to many traditional water resources models, particularly in the field of forecasting hydrologic variables. Some of the salient features that have contributed to their popularity include their ease of implementation, their ability to learn from examples without explicit knowledge of the underlying physics and their powerful generalisation abilities.

ANNs are well established in research circles and have been applied to a wide variety of water resources problems including rainfall-runoff modelling, precipitation forecasting, streamflow forecasting, groundwater and water quality modelling [ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Maier and Dandy, 2000]. However, one limitation of ANNs is that (like other empirical methods) they are unable to reliably extrapolate beyond the range of the data used for training [Flood and Kartam, 1994]. Accordingly, if the data used to train the ANN model are limited, it is very difficult to determine when the model will fail to generalise and to understand the range of applicability of the ANN model.

It has been acknowledged in the past that an ANN is susceptible to becoming "...a prisoner of its training data" [Minns and Hall, 1996]. Once the ANN model has been deployed in an operational sense, it is likely to perform poorly if faced with inputs that are far removed from the examples that it was presented with during training. This led the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology [2000] to pose the following question, "Very often we may have no alternative but to proceed with limited data. Under these circumstances can we say when generalization will fail so that we understand the range of applicability of the ANN"? Once the model has been calibrated and deployed, this equates to knowing when the model is likely to fail and when the model needs to be recalibrated to incorporate new, uncharacteristic patterns that it has not been trained on.

To improve generalisation ability beyond the calibration range, Imrie et al. [2000] added a guidance system to the original cascade correlation learning architecture used in their study and analysed the effect of using different output activation functions. The guidance system involved adjusting the cascade correlation algorithm to include cross-validation learning. It was found that the guidance system improved the results on a validation set and increased the maximum flow prediction. The use of a cubic polynomial as the output activation function was found to further enhance the capability of the ANN models to extrapolate beyond the calibration range. However, no system was developed to detect uncharacteristic data, apart from comparing the maximum and minimum values in the training and validation sets.

Other real-time hydrological forecasting experiments have been investigated in the literature [see Coulibaly et al., 2000a; Coulibaly et al., 2000b; Thirumalaiah and Deo, 2000], however, to date no system has been developed to determine when the deployed model needs to be recalibrated. In general there are three options available to the modeller: (1) no recalibration, (2) recalibration at some arbitrary time interval, and (3) recalibration given some knowledge of when a pattern that is outside of the training domain is encountered.

Recently, Bowden et al. [2001b] proposed a method for diagnosing uncharacteristic data patterns using a self-organizing map (SOM). It was found that by combining the new data with the training data and clustering these data using the SOM, regions of poor performance could be identified by examining the resulting clusters. If the new data formed a cluster that did not contain training data, then these data were diagnosed as uncharacteristic. It was found that the ANN model performed poorly on these uncharacteristic data since it had not considered these events during training. This is because ANNs are exceptionally good at interpolation, but since the activation function (usually sigmoid or tanh) saturates, extrapolation is unreliable. To determine when the ANN is extrapolating rather than interpolating, it is necessary to know what the distribution of the training data is, however, this can be rather difficult to determine with a large number of inputs. One way to address this problem is by plotting histograms of the inputs in the training set, to see which values are most common and which values are rare or absent from the training set. But this is somewhat subjective and becomes increasingly difficult as the number of inputs increases. In the present study, a SOM is used to compare multivariate distributions and diagnose when a new m-dimensional data pattern differs from all m-dimensional patterns in the training set, where m is the number of inputs.

In this paper, a hybrid model has been employed consisting of two important components. The first is a SOM that combines each new input pattern with the data used for training and determines if the new pattern clusters within the training domain. The second is a backpropagation network (BPN) which is used to perform the forecast. When a new input pattern is found to be uncharacteristic, there is a large degree of uncertainty associated with the corresponding forecast and consequently, a warning is issued. This pattern is then placed in the training set. Once the corresponding output has been collected, the BPN is retrained with this pattern included. In this way, the ANN model is able to adapt to new information as it is encountered.

To assess the efficacy of this approach, a case study is considered in which the hybrid model is used to forecast salinity in the River Murray at Murray Bridge (South Australia) 14 days in advance. The model is developed using 6 years of daily salinity, flow and river level data. Once developed, it is used as an operational model and is trialled on a further 3-year real-time forecasting test period consisting of independent data. In this paper, three recalibration scenarios are investigated, including: (1) no recalibration, (2) recalibration after each new data sample is collected, and (3) recalibration when an uncharacteristic pattern is detected by the SOM procedure.

2. CASE STUDY

The real-world case study used to demonstrate the effect of different retraining regimes is that of forecasting salinity in the River Murray at Murray Bridge, South Australia, 14 days in advance. ANN models have previously been developed for this case study by Maier and Dandy [1996] who used daily salinity, flow and river level data at various locations in the river as input variables for the period 01-December-1986 to 30-June-1992. Data from this period and at the same locations were also used in the present study. Maier and Dandy [1996] found that the ANN models trained on the input set shown in Table 1 performed best for this case study. Consequently, these 51 inputs were used to develop the BPN model. A description of how these inputs were determined is given in Maier and Dandy [1996].

Table 1. Summary of Model Inputs.

Variable	Location	Lags (days)	Total No.
Salinity	Murray Bridge	1, 3, ..., 11	6
Salinity	Mannum	1, 3, ..., 15	8
Salinity	Morgan	1, 3, ..., 15	8
Salinity	Waikerie	1, 2, ..., 5	5
Salinity	Loxton	1, 2, ..., 5	5
Flow	Overland Corner	-19, -17, ..., 7	14
Level	Lock 1 Lower	-3, -1, ..., 5	5
Total Number of Inputs			51

An additional data set consisting of daily salinity, flow and river level data at various locations in the river for the period 01-July-1992 to 30-June-1995 were used to simulate the real-time forecasting period.

3. METHODS

3.1 Self-Organizing Map (SOM)

The Self-Organizing Map (SOM) was developed by Kohonen [1982] and arose from attempts to model the topographically organized maps found in the cortices of mammal brains. The underlying basis behind the development of the SOM was that topologically correct maps can be formed in an n-dimensional array of processing elements (PEs) that did not have this initial ordering to begin with. In this way, input stimuli, which may have many dimensions, can come to be represented by a one- or two-dimensional vector which preserves the order of the higher dimensional data and provides a non-parametric estimation of the underlying distribution.

The SOM employs a type of learning commonly referred to as competitive, unsupervised or self-organizing, in which adjacent cells within the network are able to interact and develop adaptively into detectors of a specific input pattern [Kohonen, 1990].

Sorting items into categories of similar objects is a challenging, yet frequent task. The SOM achieves this task by nonlinearly projecting the data onto a lower dimensional display and by clustering these data. However, the SOM has only been used in a limited number of water resources applications. Applications have included the estimation of rainfall rates from infrared satellite and ground surface data [Hsu et al., 1997], the identification of flow regimes in horizontal air-water flow in an experimental pipeline [Cai et al., 1994] and the classification of flood data into classes defined by Representative Regional Catchments (RRCs) [Hall and Minns, 1999]. Details of the SOM algorithm are given in Bowden et al. [2001b]

The SOM implemented in this research consisted of a 20 by 20 Kohonen layer grid. There is no theoretical principle for determining the optimum size of the Kohonen layer grid [Cai et al., 1994], hence, the grid size was optimized by trial-and-error. The learning rate used in the SOM decreased linearly from an initial value of 0.7 down to 0.01 using

$$\alpha(i) = \max\left[\left(\alpha(0)(1-\frac{i}{rlen})\right), 0.01\right] \qquad (1)$$

where $\alpha(i)$ is the learning rate at iteration i, $\alpha(0)$ is the initial learning rate and $rlen$ is the running length of the training i.e. number of samples fed to the network. The neighbourhood size (N) was also a function of the training time and its size decayed linearly as training progressed, in accordance with

$$N = \max\left[\mathrm{int}\left(D(1-\frac{i}{rlen})\right), 1\right] \qquad (2)$$

where D is the maximum dimension of the Kohonen layer's columns and rows i.e. D=max(*column dimension*, *row dimension*). The SOM was trained for a total of 300 epochs.

3.2 Backpropagation Network (BPN)

A BPN model was developed using the available data from the period 01-December-1986 to 30-June-1992. The data were divided into training, testing and validation sets by using a genetic algorithm (GA) so as to minimize the statistical difference (as measured by the mean and standard deviation) between training, testing and validation data sets [Bowden et al., 2001b].

Maier and Dandy [1998] conducted empirical trials on the salinity data set and determined that 30 hidden layer nodes provided optimal performance. Consequently, a network with 51 nodes in the input layer, 30 hidden layer nodes and 1 node in the output layer was used for the BPN component of the hybrid model developed in this study. To ensure that overtraining did not occur (i.e. when the network performs well on the training data, but poorly on independent test data), cross-validation was used as the stopping criterion. In this approach, a test set is used to determine the BPN's generalisation ability. The test set root mean square error (RMSE) is calculated every 1000 iterations and the network with the best test results is saved during the run. After 100 iterations with no further improvement in the test set results, training is stopped and the network that performed best on the test set is used as the final model.

3.3 Hybrid SOM-BPN Model

A hybrid model (Figure 1) was developed that utilises a SOM to determine when an uncharacteristic input pattern is encountered and a BPN to perform the forecasting. After the SOM clusters the data, the proposed hybrid model diagnoses each new pattern as either characteristic or uncharacteristic depending on the presence or absence of training data in the new sample's cluster.

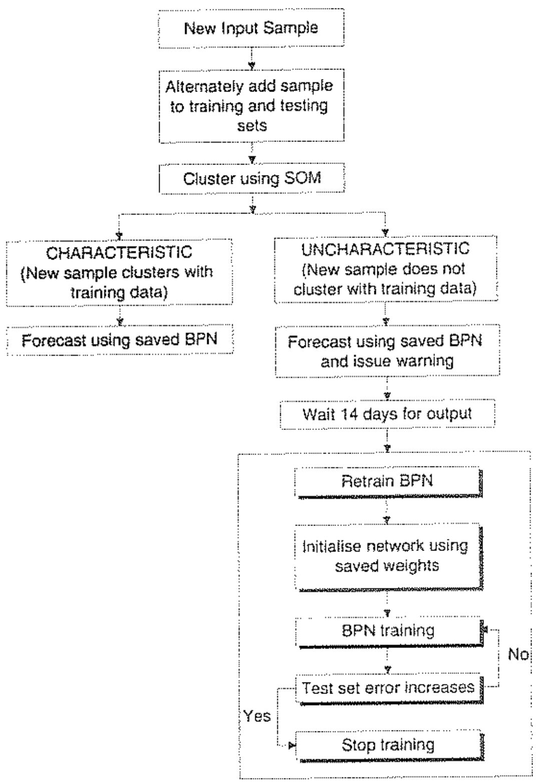

Figure 1. Hybrid SOM-BPN model.

To help ensure that the training and test sets remain statistically representative of the same population, new samples are alternately placed in each set. When a new sample is added to the training set, it is then clustered using the SOM component of the hybrid model. If the new sample is found to be uncharacteristic, it is necessary to wait 14 days (the forecasting interval) for the corresponding output before retraining the BPN. However, if the new sample is found to be characteristic data, the currently saved BPN is used to obtain the 14-day forecast.

When the BPN is retrained, the weights are initialised by using the saved weights of the previous model. This was found to significantly decrease the time needed to retrain the BPN.

4. RESULTS AND DISCUSSION

The first scenario investigated was the effect of not retraining the BPN model. The BPN model was developed using the data from the period 01-December-1986 to 30-June-1992 and the train, test and validation set RMSEs were 31.4, 30.3 and 31.2 EC units, respectively. Figure 2 shows the forecasts obtained without retraining the BPN model for the real-time forecasting test period.

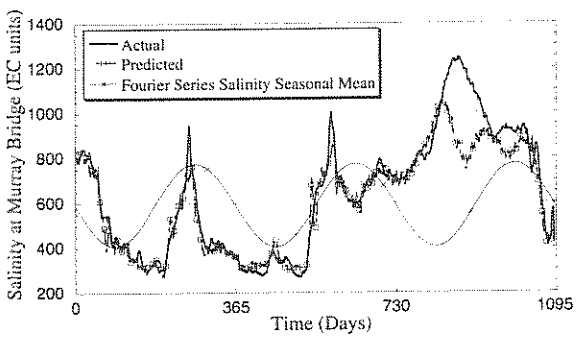

Figure 2. Results of not retraining the BPN model for the real-time forecasting test period (July 1992 - June 1995). The Fourier Series Seasonal Mean is also shown for the salinity at Murray Bridge.

In Figure 2 it can be seen that the model performs poorly on all peaks that exceed 900 EC units. In particular, the model was unable to predict the magnitude and duration of the major salinity peak that occurred in the third year at around day 900. This is also shown by the high RMSE obtained by this model for this test period (98.5 EC units). It is interesting to note that a previous study conducted by Bowden et al. [2001a], identified that this major salinity peak corresponds with an unseasonal low flow event. A Fourier series was fitted to the mean monthly salinity at Murray

Bridge for the model development period (i.e. 01-December-1986 to 30-June-1992) and this is shown in Figure 2 for the real-time forecasting test period 01-July-1992 to 30-June-1995. It is evident that the high salinity/low flow event that occurred in the third year is unseasonal and unlike any of the data used in calibrating the BPN model. This has been shown by Bowden et al. [2001b], who used a SOM to diagnose that these data are outside of the training domain. Consequently, the model was unable to match the large peak.

The second scenario investigated the effect of retraining the BPN network after each sample is obtained. It must be noted that to simulate an operational model, 14 days must elapse until the relevant output can be obtained before retraining of the model can commence. The results of this retraining scenario for the real-time forecasting test period are shown in Figure 3.

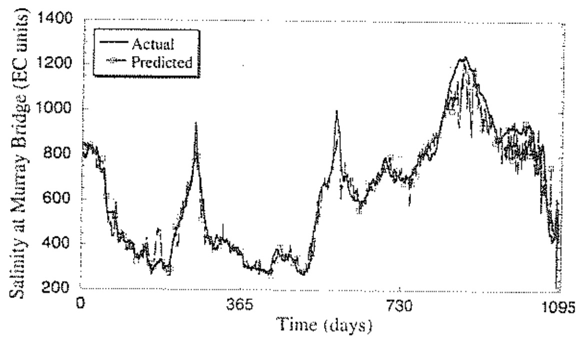

Figure 3. Results of retraining the BPN model after each sample for the real-time forecasting test period (July 1992 - June 1995).

As expected, this retraining scenario provided better predictions for the peak salinity values and this is reflected by a lower RMSE (63.8 EC units) in comparison with the scenario where no retraining was performed. Of particular importance is the improved ability of the model to predict the major salinity peak. However, a concern is the large degree of noise in the forecast that manifests itself in the last year of the test period. It was found that this noise could be ameliorated by resplitting the data after two years into training and test sets and retraining the model.

The results of resplitting the data are shown in Figure 4. Whilst still underpredicting the peak value, the forecasting noise was reduced and the RMSE decreased from 63.8 EC units to 61.1 EC units. Therefore, it is hypothesised that noise in the forecast was due to the training and test sets becoming statistically unrepresentative of the full data set.

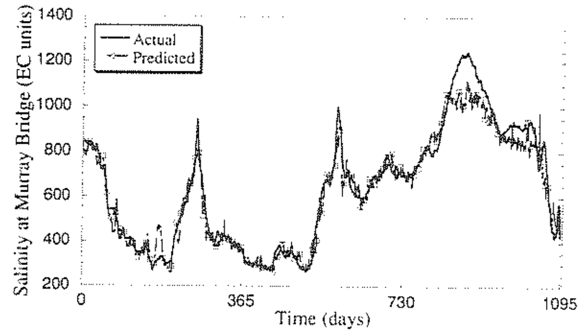

Figure 4. Results of retraining the BPN model after each sample (with data resplit into training and test sets after 2 years) for the real-time forecasting test period (July 1992 - June 1995).

The third scenario investigated involved the use of the hybrid SOM-BPN model to selectively retrain based on the identification of uncharacteristic samples. As with the previous scenario, in order to reduce the forecasting noise it was better to resplit the data into training and test sets after two years. The SOM-BPN model's results for the real-time forecasting test period are shown in Figure 5. The hybrid model performed particularly well in forecasting the peak in the second year. It underestimated the major peak in the third year, but for this peak it was able to equal the performance of the model that was retrained after every sample (Figure 4).

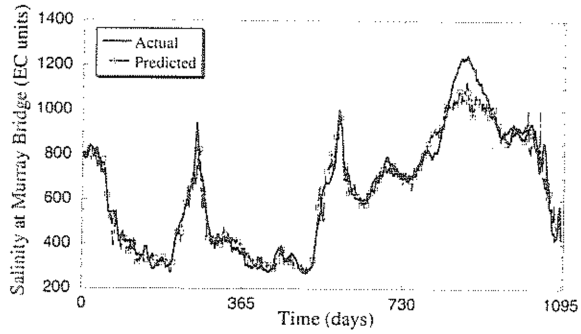

Figure 5. Results of the hybrid SOM-BPN model (with data resplit into training and test sets after 2 years) for the real-time forecasting test period (July 1992 - June 1995).

There were 178 warnings issued in the 3-year period, which indicated 16.3% of the patterns in this test period were diagnosed as uncharacteristic. Despite only retraining 16.3% of the time, the SOM-BPN model was able to obtain a RMSE of 58.1 EC units. This represents an improvement in the RMSE of 4.9% when compared with the model that retrained 100% of the time. In addition, the results from the SOM-BPN represent a 41.0%

improvement in the RMSE when compared with the no retraining scenario.

5. CONCLUSIONS

This paper presents a new retraining methodology for real-time forecasting models based on a hybrid SOM-BPN model. In addition, the performance of the SOM-BPN model was compared with two alternative retraining methods involving no retraining and continuous retraining, respectively. All three retraining methods were trialled on a 3-year real-time forecasting period using real data from the River Murray. The results indicate that the SOM-BPN model provides an effective means of identifying when retraining is required. The front-end SOM component of the hybrid model was able to diagnose the data that were outside of the training domain and retrained 16.3% of the time during the 3-year real-time forecasting test period. This resulted in a significant improvement over the scenario in which the ANN model was not retrained. In addition, selectively retraining based on the presence of uncharacteristic data was found to produce a lower RMSE when compared to continuously retraining after each new sample.

6. REFERENCES

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, Artificial neural networks in hydrology. II: Hydrologic applications, *Journal of Hydrologic Engineering, ASCE*, *5*(2), 124-137, 2000.

Bowden, G. J., G. C. Dandy, and H. R. Maier, Data transformation for neural network models in water resources applications, *Water Resources Research (submitted)*, 2001a.

Bowden, G. J., H. R. Maier, and G. C. Dandy, Optimal division of data for neural network models in water resources applications, *Water Resources Research (in press)*, 2001b.

Cai, S., H. Toral, J. Qiu, and J. S. Archer, Neural network based objective flow regime identification in air-water two phase flow, *The Canadian Journal of Chemical Engineering*, *72*, 440-445, 1994.

Coulibaly, P., F. Anctil, and B. Bobee, Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, *Journal of Hydrology*, *230*, 244-257, 2000a.

Coulibaly, P., F. Anctil, and B. Bobee, Neural network-based long term hydropower forecasting scheme, *Computer Aided Civil and Infrastructure Engineering*, *15*, 335-364, 2000b.

Flood, I., and N. Kartam, Neural networks in civil engineering. I: Principles and understanding, *Journal of Computing in Civil Engineering*, *8*(2), 131-148, 1994.

Hall, M. J., and A. W. Minns, Classification of hydrologically homogeneous regions, *Hydrological Sciences Journal*, *44*(5), 693-704, 1999.

Hsu, K. L., X. G. Gao, S. Sorooshian, and H. V. Gupta, Precipitation estimation from remotely sensed information using artificial neural networks, *Journal of Applied Meteorology*, *36*(9), 1176-1190, 1997.

Imrie, C. E., S. Durucan, and A. Korre, River flow prediction using artificial neural networks, *Journal of Hydrology*, *233*, 138-153, 2000.

Kohonen, T., Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, *43*, 59-69, 1982.

Kohonen, T., The Self-Organizing Map, *Proc. IEEE*, *78*(9), 1464-1480, 1990.

Maier, H. R., and G. C. Dandy, The use of artificial neural networks for the prediction of water quality parameters, *Water Resources Research*, *32*(4), 1013-1022, 1996.

Maier, H. R., and G. C. Dandy, The effect of internal parameters and geometry on the performance of back-propagation neural networks: an empirical study, *Environmental Modelling and Software*, *13*, 193-209, 1998.

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

Minns, A. W., and M. J. Hall, Artificial neural networks as rainfall-runoff models, *Hydrological Sciences Journal*, *41*(3), 399-417, 1996.

Thirumalaiah, K., and M. C. Deo, Hydrological forecasting using neural networks, *Journal of Hydrologic Engineering*, *5*(2), 180-189, 2000.