The Application of Artificial Neural Network for Forecasting Dam Spillage Events.

[1,2]Anita Talib, [3]Yahya Abu Hasan
[1]School of Distance Education, Universiti Sains Malaysia, USM, 11800, Penang, Malaysia.
[2]Center for Marine and Coastal Studies CEMACS, Sains Malaysia, USM, 11800, Penang, Malaysia.
[3]School of Mathematical Sciences, Universiti Sains Malaysia, USM, 11800, Penang, Malaysia.
anita@usm.my

Abstract

Artificial Neural Network (ANN) is applied for forecasting of dam spillage using a 10 year hydrological dataset of Ahning Dam in the northern Malaysia at the Pedu-Muda area. During the rainy season, the Ahning Dam will overflow due to heavy rainfall. Increased siltation due to logging could contribute to worsen the overflow during this time. This will lead to increasingly destructive floods downstream. Rainfall in the Pedu-Muda area is not constant. It is highly influenced by the monsoon seasons. During the dry season, the dams may dry up. Siltation will further hamper the dams' capability to store water for irrigation.

This study has shown that a simple ANN, based on 3 input variables; dam inflow, rainfall and dam release can forecast the complex non-linear hydrological processes of dam spillage events.

1.Introduction

The Ahning River Dam project was completed in 1989 and is owned by the Kedah State Government, Malaysia. Apart from Ahning Dam, there are two other large dams in Kedah, known as Muda and Pedu dams, that were built in 1969 under the Muda Irrigation Project, aimed to enable double cropping of rice cultivation in the Muda area. Though their main function is to provide irrigation, dam releases are usually carried out for domestic supply purposes during the early January to March. (Othman and Ali, 2002). To better coordinate the Ahning and Pedu dam releases to meet the irrigation and domestic demands, the operation and maintenance of both dams is managed by Muda Agricultural Development Authority (MADA) (Othman and Ali, 2002).

One of the problems faced is that rainfall in the Pedu-Muda area is not constant. It has two distinct seasons, a wet season and dry season. During the rainy seasons (wettest in October), the Dam overflows due to heavy rainfall. Increased siltation due to logging, also worsen the situation. This leads to destructive floods downstream. And during dry seasons, the dams may dry up (Berga,1998).

The aim of this study is to model the hydrological dynamics of Ahning Dam. Different modelling approaches that have been applied to modelling water resources include differential equation, statistical and computational methods (Chapra, 2008). The use of various approaches of finding patterns and to apply them for forecasting and prediction on is also part of data mining. Ecological data mining refers specifically to search and discover patterns and forecast processes in ecological datasets.

In this paper we consider the use of artificial neural network. Artificial neural network (ANN) is an abstraction of the human neuron, a mathematical model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

Basically, the advantages of neural networks are that they are able to represent both linear and non linear relationships and are able to learn these relationships directly from data. Many researchers compared statistical methods to ANNs (Lek *et al*, 1996; Maier and Dandy, 1996). They concluded that ANNs perform better compared to traditional Multiple Regression and other classes of statistical modelling. ANNs are more flexible, hence more suitable for prediction, more accurate and the results could be replicated. ANNs have successfully been applied in temporal studies of salinity in rivers (Maier and Dandy, 1996) benthic communities(Chon *et al*, 2000) in streams, eutrophication and algal blooms in lakes (Karul *et al*, 1999; Talib, 2008). A distinct advantage of ANNs compared to statistical and differential equation approaches is that they can predict the timing and magnitude of species succession (Recknagel and Wilson, 2000). Study on various water bodies has shown that ANNs could predict algal blooms with abundance and succession patterns of blue-green algae species (Recknagel *et al*, 1997; Talib, 2006)

In this case study, we use the chosen techniques for modelling of Ahning Dam. In the following sections, we present brief descriptions of the approaches that we considered, followed by an analysis of the hydrological events related to the data.

2. Study Area and Dataset

Ahning River is a tributary of Padang Terap River in Kedah (see Figure 1). Water released from Ahning and Pedu dams, flows into Padang Terap River. It then flows into the Bukit Pinang and Pelubang water treatment plants.

This study was conducted using a sample of Ahning Dam monthly datasets from 1992-2000.The datasets were preprocessed as follows:

i. scaling

ii. removing missing variables and outliers

iii. analysing data with descriptive statistical analysis

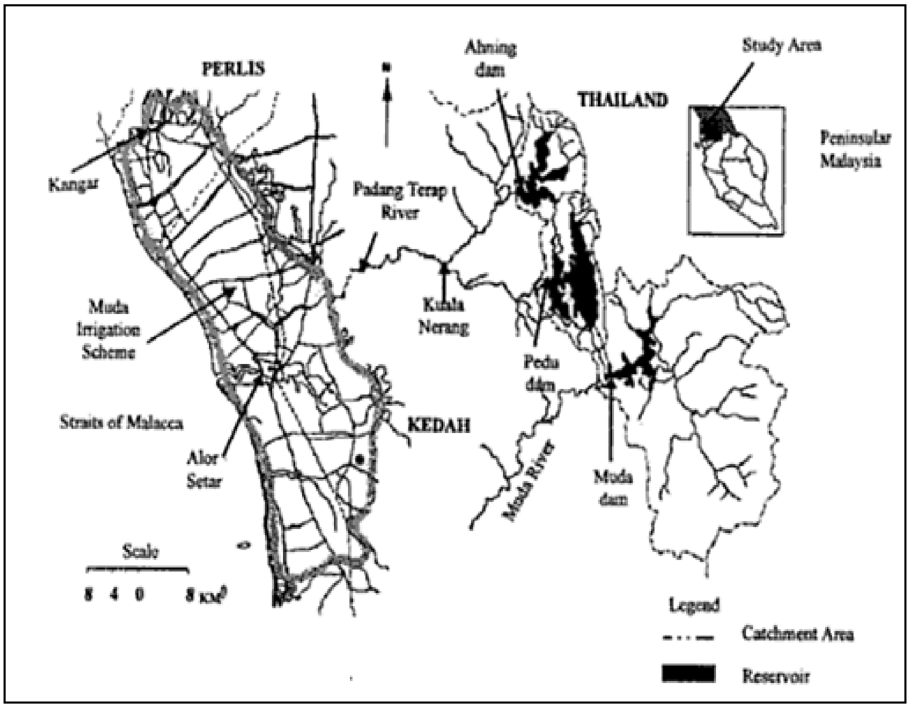
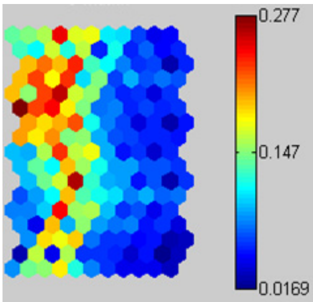


Figure 1 Location of Ahning Dam and other dams in the Padang Terap catchment area.

3. Artificial Neural Networks ANN

Non Supervised ANN were applied to the ordination, clustering and mapping of relationships between physical, chemical and biological time-series data of the two adjacent lakes. The ANN models were developed by using the MatLab 5.3 SOM Toolbox functions (Vesanto, 1999) developed at the Laboratory of Computer and Information Science at Helsinki University of Technology. The ANN are also known as Self-Organizing Maps (SOM) or Kohonen networks referring to Kohonen (1995) who invented principles of unsupervised training of ANN. Clustering was carried out by the K-means method (MacQueen, 1967; Vesanto, 2000).
The SOM algorithm using the k-means method has separated data sets into three clusters corresponding with the following water quality ranges:
Cluster 1: January to April
Cluster 2: May to August
Cluster 3: September to December
The U-matrix identified overlapping areas at the borders of the clusters, and is shown in Figure 2a together with the K-means clusters (Figure 2b). The plane quality was very good with a low quantization error of 0.070, based on the results of the training.
For the supervised ANN, the ANN architecture was selected manually, i.e. 3-2-1 architecture. The hyperbolic tangent was selected for both input and output activation function, with sum of squares as output error function. Three variables were selected for inputs, including monthly rainfall, monthly inflow and dam release with an output of 1 month ahead dam spillage. Various lag parameters were tested as inputs with 1 month lag as the optimal test result.



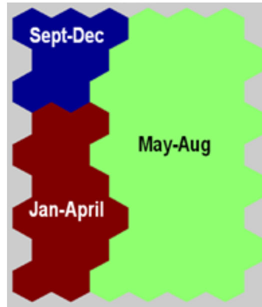


Figure 2a Distance matrix map (U-matrix) **Figure 2b Partitioned map (K-means) of seasons**

Figure 2 The U-matrix and K-means clusters

4 Results and Discussion

This study involves training and testing the supervised ANN model by trial and error approach. ANN model with the 3-2-1 architecture, with conjugate gradient descent as the training algorithm is chosen and applied (see Table 1). ANNs are computationally intensive and many parameters have to be determined with few guidelines and no standard procedure to define the architecture (Lek and Guegan, 1999). There is no global method for when to stop training and how to determine overtraining. Training was stopped based on the number of iterations, i.e 1000 and as the error on the validation set increases.

Forecasting of ecological time series data sets using ANN is a complex task. Basically, the four main steps taken in this forecasting study is to initially choose a suitable model representation based on expert-knowledge. The next step is to train the model and Validate is, that is by testing the model on unseen data sets to determine its validity.

To interpret involve finding the insights, causal explanations based on the hypothesis tested.

Although ANNs are increasingly being used for the prediction and forecasting of water resources variables, the problem of assessing the optimality of the results exists. Apart from the importance of preprocessing, specific mapping of ANN depends on the architecture of the network, training techniques and modeling parameters.

ANNs are sensitive to the composition of the training data set and to the initial network parameters (Bishop 1995). For this dataset (see Table 1), the split-set validation is chosen for testing the models, with sequential data partitioning i.e. training (68%), validation (16%) and test (16%). Another validation technique, the leave-one-out bootstrapping and leave-k-out cross-validation have been attempted by Wilson (2004) whereby no user decisions are required regarding division of data into training and testing sets. Another problem with ANN, is that it is difficult to predict an unknown event that has not occurred in the training data. The values of training data should therefore cover as wide a range as possible (Aoki et al., 1999).

Table 1 ANN Parameters for training and validation

Parameters		
	Training	Validation
Absolute Error	677.99	1613.64
Network Error	0.0029	0
Error Improvement	0.000042	
Iteration	134	
Training Speed, iter/sec	134	
Architecture	[3-2-1]	
Training algorithm	Conjugate Gradient Descent	
Training stop reason	No error improvement	

Heavy rainfall usually triggers spill, when almost top water level conditions are achieved. For flood controls, release of water in specific quantities is practiced when conditions permit. Water released at designated times also coincides with paddy planting seasons and in August, coincides with inter monsoon rains (Berga, 1998).
The results of the 1 month ahead dam spillage forecast for Ahning Dam is as shown in Figure 3 with R^2 values (where R is a measure of correlation) of 0.50. The model forecasted the right timing and magnitude for dam spillage events during 1995 and 1997 and in the first spillage event of 2001. For the second spillage event of 2001 and 2002 spillage, the model overestimated the spillage. In 1997, heavy storms caused serious flooding in this region, the effect of Zita, a tropical cyclone that originated in Luzon, 20 August 1997, whereby 2800 people were evacuated.

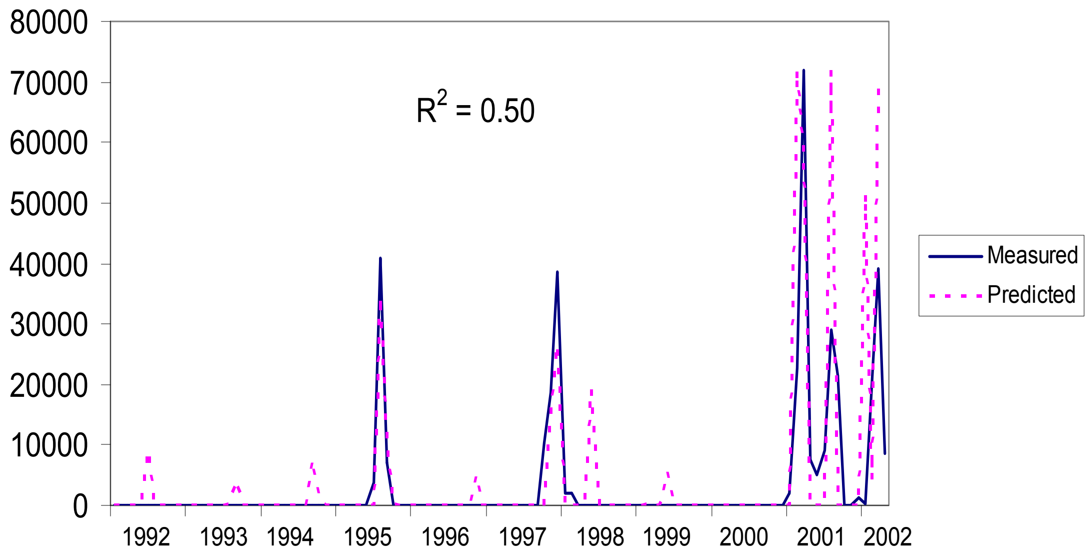


Figure 3 Results of 1 month ahead forecast of Dam Spillage using dataset from Ahning Dam.

Table 2 Input importance (%) based on the training of ANN

Input Column Name	Importance, %
Monthly Rainfall	18.4
Monthly Inflow	68.0
Dam Release	13.6

Our 1 month forecast results correlate with the Non Supervised ANN model. The results of the clustering using SOM are as shown in Fig 4.
According to Berga (1998), there are two seasons, a wet season (August to October) coinciding with the south-west monsoon and dry season (between December to March).The moderate period is between April to July.
The results of this study (see Figure 4) on ordination and classification of dam spill, dam release, monthly inflow and monthly rainfall data of Ahning Dam using SOM have proven that the dynamics over the long term period can be patternised into 3 seasonal clusters. Between September to December, there is high inflow, also related to high rainfall in this period. This is also correlated to the previous high rainfall seasonal period of May to August. Subsequently dam spillage occur in this period of September to December. Dam release is highest during January to April, as a counter measure of the dam spill events that occurred in the September to December period.

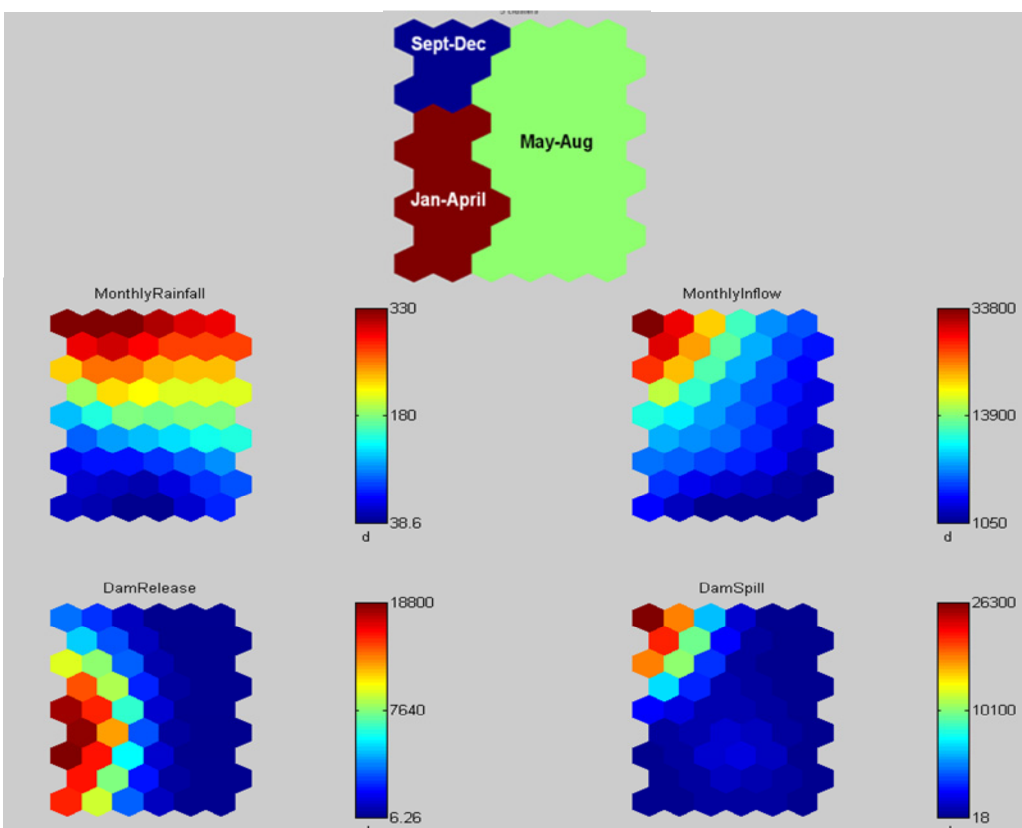


Figure 4 Results SOM clustering patterns correlating rainfall, inflow, dam release and dam spillage events.

Both methods of ANN highlighted the important key variables related to dam spillage events in Ahning Dam. The forecasting of dam spillage correlates with the seasonal patterns of hydrological dynamics and this knowledge can be utilized by water resource development and management, for future flood mitigation plans.

5.Conclusion

Various computational techniques can be applied to data mining ecological datasets. The results of the study show that it is possible to forecast 1 month ahead dam spillage for Ahning Dam using a simple ANN with 3-2-1 architecture that could reasonably forecast in terms of timing and magnitude of dam spill events. Apart from the importance of ANN architecture and the training algorithm for forecasting, good ordination and clustering using ANN approach is also important as it allows us to visualize the patterns related to the dynamics in the datasets, over a certain time period, or seasons. The results from this study can be utilized by managers and decision makers as part of the dam management measures.

References

Aoki, I., Komatsu, T., and Hwang, K. (1999) Prediction of response of zooplankton biomass to climatic and oceanic changes. Ecol. Model 120: 261-270.

Berga L (1998). Proceedings of International Symposium on New. Trends and Guidelines on Dam Safety, Barcelona, vol 2, pp. Balkema, Rotterdam.

Bishop, C.M. (1995) Neural Networks for Pattern Recognition: Oxford Uni. Press.

Chapra S. (2008) *Surface Water Quality Modeling*. Waveland Pr Inc

Chon, T.-S., Park, Y.-S., and Park, J.H. (2000) Determining temporal pattern of community dynamics by using unsupervised learning algorithms. Ecological Modelling 132: 151-166.

Karul, C., Soyupak, S., and Yurteri, C. (1999) Neural network models as a management tool in lakes. Hydrobiologia 408/409: 139-144.

Kohonen, T. (1989) Self-Organization and Associative memory. Berlin: Springer-Verlag.

Kohonen, T. (ed) (1995) Self-organizing maps. Heidelberg: Springer-Verlag.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., and Aulagnier, S., Application of neural networks to modelling nonlinear relationships in ecology, *Ecological Modelling*, 90,1996, pp 39-52.

Lek S, Guegan JF Artificial neural networks as a tool in ecological modelling, an introduction, *Ecological Modelling* 120, 1999, pp 65-73.

Maier, H.R., and Dandy, G.C. The use of Artificial Neural Network for the prediction of water quality parameters, *Water Resources Research* 32, 1996, pp1013-1022.
MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley,: University of California Press, pp. 281-297.

Othman, A.S. and Ali, A. 2002. eds Biodiversity resources of Ahning Lake Universiti Sains Malaysia. Malaysia

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K.-I., Artificial neural network approach for modelling and prediction of algal blooms, *Ecological Modelling*, 96,1997, pp1-28.

Recknagel, F., and Wilson, H., Elucidation and prediction of aquatic ecosystems by artificial neuronal networks, In *Artificial Neuronal Networks: Application to Ecology and Evolution*, Springer Verlag, 2000, pp. 143-156.

Talib, A., Abu Hasan, Y, and Varis O. Application of machine learning techniques in data mining of ecological datasets. *Proceedings of International Conference on Environmental Research and Technology ICERT 2008*, 2008, pp 677-681

Talib A. (2006), Comparative Ecological Study of Two Dutch Lakes Using Computational Modelling PhD Thesis. University of Adelaide, Australia.

Othman, A.S. and Ali, A. 2002. eds Biodiversity resources of Ahning Lake Universiti Sains Malaysia. Malaysia

Vesanto, J. (1999) SOM-Based Data Visualization Methods. *Intelligent Data Analysis* 3: 111-126.