

# Combined Sewer Overflow forecasting with Feed-forward Back-propagation Artificial Neural Network

Achela K. Fernando, Xiujuan Zhang, and Peter F. Kinley

**Abstract**—A feed-forward, back-propagation Artificial Neural Network (ANN) model has been used to forecast the occurrences of wastewater overflows in a combined sewerage reticulation system. This approach was tested to evaluate its applicability as a method alternative to the common practice of developing a complete conceptual, mathematical hydrological-hydraulic model for the sewerage system to enable such forecasts. The ANN approach obviates the need for *a-priori* understanding and representation of the underlying hydrological hydraulic phenomena in mathematical terms but enables learning the characteristics of a sewer overflow from the historical data.

The performance of the standard feed-forward, back-propagation of error algorithm was enhanced by a modified data normalizing technique that enabled the ANN model to extrapolate into the territory that was *unseen* by the training data. The algorithm and the data normalizing method are presented along with the ANN model output results that indicate a good accuracy in the forecasted sewer overflow rates. However, it was revealed that the accurate forecasting of the overflow rates are heavily dependent on the availability of a real-time flow monitoring at the overflow structure to provide antecedent flow rate data. The ability of the ANN to forecast the overflow rates without the antecedent flow rates (as is the case with traditional conceptual reticulation models) was found to be quite poor.

**Keywords**—Artificial Neural Networks, Back-propagation learning, Combined sewer overflows, Forecasting.

## I. INTRODUCTION

COMBINED sewers are a common feature in urban drainage systems in towns and cities with well-established old sewerage systems. The overflows from these systems are a major source of pollution in the urban waterways. In coastal cities, an association between storm events, urban runoff, and coastal water quality has been established [1]. In separated wastewater systems too, overflows can get triggered by excessive inflow and infiltration during rain events.

Achela K. Fernando is a lecturer in the School of the Built Environment of the Unitec New Zealand, PO Box 92025, Mt Albert, Auckland (phone: 64-9-8154321-X7036; e-mail: afernando@unitec.ac.nz).

Xiujuan zhang is a student of the School of the Built Environment of Unitec New Zealand (e-mail: xiujuanzhang968@yahoo.com).

Peter Kinley is the modeling team leader of Metrowater, Auckland (e-mail: peter.kinley@meterowater.co.nz).

Auckland City has a population of approximately 370,000. Auckland City Council (ACC) owns the stormwater and combined sewer systems in Auckland City. ACC is the sole shareholder in Metro Water Ltd (Metrowater), the Council Controlled Organisation that owns, operates and manages the separated wastewater system in Auckland City. Metrowater also operates and manages the combined sewer system on behalf of ACC.

Under the Resource Management Act (1991) and Auckland Regional Council's proposed Air, Land and Water Plan, Metrowater and Auckland City are required to obtain Resource Consents to operate their drainage networks. A requirement to obtain the Resource Consents is that an understanding of each of the overflow structures is developed and the performance of the overflow structures is described, in terms of discharge volumes and frequencies and spill rates. A \$23 million modelling study that studied the overflows was completed in December 2005.

The standard method to predict the operation of constructed overflows is to develop a calibrated hydrological/hydraulic model that is then fed in with the real or design rainfall. This involves the estimation of the amount of RDII (Rainfall dependant inflow and infiltration) in the model which is a difficult and complex task given the unpredictability associated with seasonal variations when a long term prediction is required. Various investigations are being carried out to come up with solutions to this challenge [2]. The process of hydraulic model development, calibration and simulation is, however, quite time consuming. Many of these deterministic and conceptual models require a large quantity of good quality data, sophisticated approaches for calibration, and a detailed understanding of the underlying physical processes. Black-box models such as ANNs attempt to develop a relationship among input and output variables without considering the physical processes involved [3, 4].

The existing calibrated hydrological and hydraulic model of the trunk wastewater system serving Auckland City has had 100 years of stochastic rainfall representing the existing climate, with data at five minute intervals run through it to gain an understanding of the performance of the overflows, in terms of volumes, frequencies and spill rates. This process took approximately 18 months. The outputs have been used to help select options for managing the overflows and to

prioritise overflow structures for remedial work through long term works programmes.

This study is intended to investigate the possibility of replacing the traditional hydrological/hydraulic model with an ANN type model given the ability of the ANN to learn from the available data. It is envisaged that if the ANN model succeeds, then the process of overflow frequency estimation can be accomplished with less effort and time than has been hitherto possible.

## II. THEORY

Various forms of ANNs have found applications in predicting and forecasting [5, 6, 7] The ANN used in this application is a Multi Layer Perception (MLP) type network. The learning algorithm follows the back propagation technique that was first established by Rumelhart et. al. [8,9]. The method adopted in this study is described below. A conceptual representation of the used network is shown in Figure 1. The generic case of an N-dimensional input vector  $X_p = (X_{p1}, X_{p2}, X_{p3}, \dots, X_{pN})$  that is to be mapped to the desired M dimensional output vector  $Z_p = (Z_{p1}, Z_{p2}, Z_{p3}, \dots, Z_{pM})$  through a multi-layer perceptron L nodes in the hidden layer is considered for the establishment of the algorithm [10]. Let the output from the network be  $O_p = (O_{p1}, O_{p2}, O_{p3}, \dots, O_{pM})$  which is different from the desired output. The output layer weights are super-scripted "o" and the hidden layer weights are super-scripted "h". A typical node in the input, hidden and output layers respectively are i ( $1 \leq i \leq N$ ), j ( $1 \leq j \leq L$ ) and k ( $1 \leq k \leq M$ ). Only the p<sup>th</sup> pattern or sample in the training set is considered here.

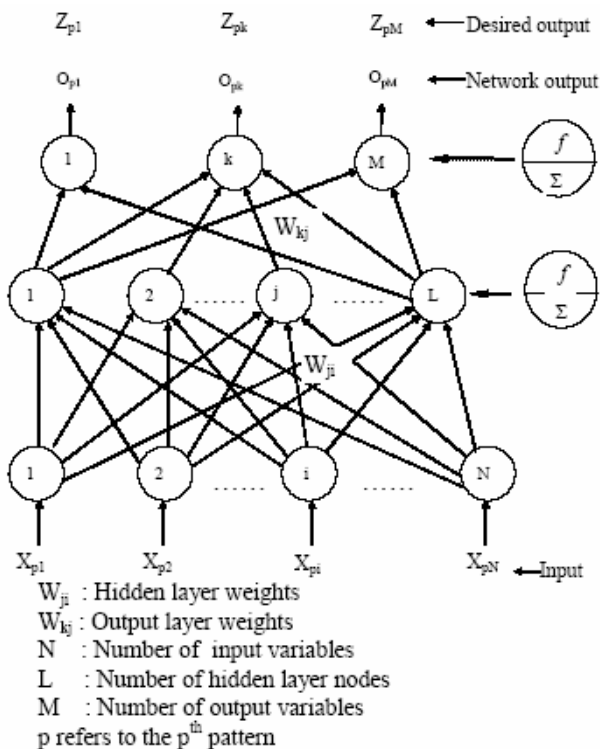


Fig. 1 Generic representation of the ANN of MLP

### A. Hidden Layer Input

The net input into the j<sup>th</sup> node in the hidden layer is

$$\text{net}_{pj}^h = \sum_{i=1}^N w_{ji}^h X_{pi} + \theta_j^h$$

where the last term is the bias term which was set to zero in this application.

### B. Hidden Layer Output

The outgoing signal from the j<sup>th</sup> hidden node is,

$$i_{pj} = f_j^h(\text{net}_{pj}^h)$$

where  $f$  is the transfer function. A Hyperbolic tangent function was used in this application.

### C. Output Layer Input

Input into k<sup>th</sup> node in the output layer is the weighted sum of all the incoming signals, i.e.,

$$\text{net}_{pk}^o = \sum_{j=1}^L w_{kj}^o i_{pj} + \theta_k^o$$

where again the last term is the bias which was set to zero in this case.

### D. Output Layer Output

The output from the k<sup>th</sup> output node is,

$$O_{pk} = f_k^o(\text{net}_{pk}^o)$$

### E. Weight Update to Output Layer Connections

The desired weight change for the synaptic weight connecting the k<sup>th</sup> output layer node with the j<sup>th</sup> hidden layer node when the p<sup>th</sup> pattern is introduced to the network is

$$\Delta_p w_{kj}^o = \eta (Z_{pk} - O_{pk}) f_k^o(\text{net}_{pk}^o) i_{pj}$$

where the value  $\eta$  is the learning rate which was set at 0.05. The weights on the output layer were updated according to,

$$w_{kj}^o(t+1) = w_{kj}^o(t) + \Delta_p w_{kj}^o(t)$$

where t and t+1 respectively refer to the previous and current time steps in the iterative learning procedure.

### F. Weight Update to Hidden Layer Connections

The desired weight change for the synaptic weight connecting the j<sup>th</sup> hidden layer node with the i<sup>th</sup> input layer node when the p<sup>th</sup> pattern is introduced to the network is

$$\Delta_p w_{ij}^h = \eta f_j^h(\text{net}_{pj}^h) X_{pi} \sum_{k=1}^M (Z_{pk} - O_{pk}) f_k^o(\text{net}_{pk}^o) w_{kj}^o$$

Defining a hidden layer error term as

$$\delta_{pj}^h = f_j^h(\text{net}_{pj}^h) \sum_{k=1}^M \delta_{pk}^o w_{kj}^o$$

the updated hidden layer weight can be expressed as,

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \eta \delta_{pj}^h X_{pi}$$

where t and t+1 respectively refer to the previous and current time steps in the iterative learning procedure.

A computer program in FORTRAN was developed to represent this training algorithm and extended to freeze the connection weights once the training has been completed and compute the final outputs for any given set of input data.

#### G. Normalization of the Raw Data

The raw data associated with this application (overflow rate, rainfall depth) can vary widely and it is general practice to normalize (or standardize) the raw data to lie between a specific range [11]. It is common to use a linear normalization technique for this purpose. However, this normalization limits the future use of the trained ANN to forecast for ranges outside the data ranges used during training. This application uses a non-linear technique which has previously shown to enable some extrapolation outside the data ranges used for learning [12]. Other methods to enable extrapolation have also been reported [e.g. 13].

### III. DATA FOR NETWORK TRAINING

#### A. Estimation of Overflows using Traditional Models

The traditional model was simulated with synthetically generated rainfall for the period of 2011 – 2020. The results of the model simulation included the overflow rates at the constructed overflow structures.

#### B. Data used

For the purpose of this study, only one specific overflow structure was chosen. The data used included the traditional model-predicted overflow rates for the overflow structure and the synthetically generated rainfall data for the rain-gauge that is in the closest proximity. The overflow rates were in m<sup>3</sup>/s while the rainfall depths were in mm; The time resolution of all the data was 15 minutes.

### IV. METHOD

#### A. Choice of Overflow

The chosen overflow structure (numbered 5EAA220w1) is situated in Avondale, a suburb of Auckland city approximately 10km away from the city centre. This structure is known to overflow a few times in an average year and was deemed to be a good test case. Another attribute of this overflow is that a permanent raingauge is operational in the proximity, data from which have been used to synthetically generate rainfall for a 10 year period. As mentioned previously, the overflow rates from the traditional model and the synthetic rainfall data for the Avondale rain-gauge were chosen for this study.

#### B. Choice of Input Data

Although it is qualitatively presumed that the overflow rate at the future time step  $[Q(t)]$  is bound to be a function of antecedent rainfall  $[R(t-1), R(t-2), \dots, R(t-i)]$  and antecedent overflow rates  $[Q(t-1), Q(t-2), \dots, Q(t-j)]$ , it is impossible to gauge qualitatively how many time steps into the past would allow the best predictability, i.e., it is not known *a priori* what

the values of  $i$  and  $j$  are.

From the ANN point of view, determining the appropriate architecture is important since the network topology directly affects the computational complexity and the generalization capability. Thus it is crucial to choose the “just right” number of inputs to, on the one hand, reduce redundancies, and on the other, enable proper learning of the input-output mapping function by the ANN [14, 15]. While too small a network may fail to capture the underlying function, too big a network may cause inefficiency in the form of redundancies in the connection weights in the ANN.

To enable the right choice of appropriate input variables, i.e., the choice of most suitable values for  $i$  and  $j$ , the serial- and cross-correlation between the overflow and rainfall data were investigated. This method, previously successfully used for similar studies [12] provides useful information to determine the size of the ANN in order to capture the underlying function efficiently.

The cross-correlation between the overflow rates and rainfall data and serial correlation amongst the overflow rates were determined for the first three years of data and plotted as shown in Fig.2.

Similarly, the serial correlation values for the overflow rates were plotted for the first three years' of data. The serial correlation plot is shown in Fig. 2.

From Fig. 2, it can be concluded that the cross correlation values increase with increasing lag time, peak around a time lag of approximately 9 units and then decreases with increasing lag time. In general, a high correlation can be observed between approximate lag time units of 5 and 16. Thus, the appropriate rainfall input to forecast  $Q(t)$  were  $R(t-5), R(t-6), R(t-7), R(t-8), R(t-9), R(t-10), R(t-11), R(t-12), R(t-13), R(t-14), R(t-15), R(t-16)$ .

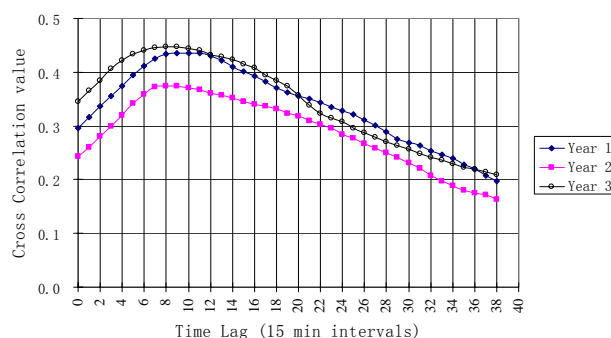


Fig. 2 Cross correlation between overflow rate (Q) and rainfall (R)

For serial-correlation (Fig. 3), on the other hand, the correlation values decreases gradually, as expected, with increasing lag time. The overflow rate input values chosen were therefore  $Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6)$ .

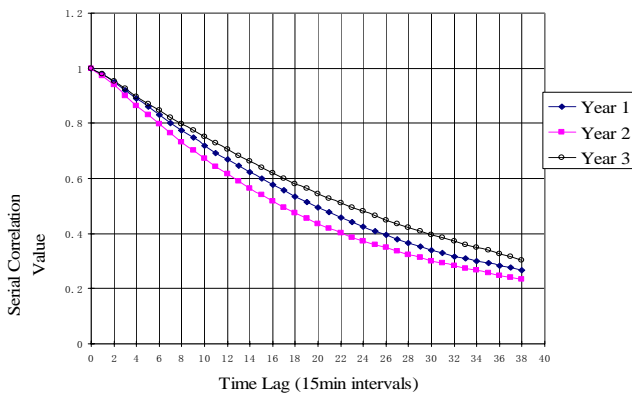


Fig. 3 Serial correlation amongst overflow rates (Q)

The ANN model (ANN-I) for forecasting the current value of overflow rate  $Q(t)$  thus consisted of 18 input nodes (12 antecedent overflow rates and 6 antecedent rainfall data). A hidden layer nodes in the MLP was set to 9.

For comparison purposes with the traditional methods, another model (ANN-II) which uses only the antecedent rainfall data was also developed. ANN-II model used 12 input nodes with the antecedent rainfall  $R(t-5)$  to  $R(t-16)$ .

It is now accepted in practice, after Amari et. al. [16], that no over-fitting to the data occurs if the ratio between the training sample size to the number of weights is larger than 30 [17]. This condition is satisfied in this application.

### C. Standardization of Training Data

In order to enable forecasting into values unseen during training, the commonly used linear normalization of input data was replaced with a modified normalization [9]. The raw values ( $v$ ) were first normalized to obtain the linearly standardized value  $V$  to span between 0 and 1 as follows:

$$V = \frac{v - v_{\min}}{v_{\max} - v_{\min}}$$

The linearly standardized value  $V$  was then further normalized using a hyperbolic tangent function to obtain the standardized value  $SV$  as

$$SV = \frac{1 - e^{-V}}{1 + e^{-V}}$$

This transformation is illustrated in Fig. 4.

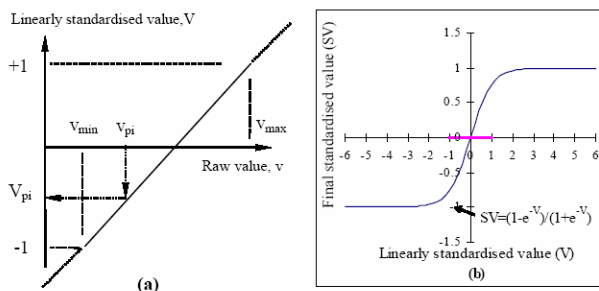


Fig. 4 Standardization of the raw data

### D. Learning

The first three years' data were used for training the ANN while the remaining seven were used for testing. Of the various stopping criteria [18], termination following a fixed number of training iterations was adopted for simplicity. The ANN program developed by the principal author begins by assigning a random set of weights for the network connections. Several runs were completed and the one with the least total mean squared error (MSE) for the training set was used as the model. Fig. 5 shows an example of one such training run that has been stopped at 30 error-reducing iterations (66 total iterations) with a final MSE of 0.430.

Both the ANNs were trained several times (20) starting with random connection weights and stopped when the total mean squared error reduced 30 times. The connection weights of the trained ANN that ended up with the least total MSE were frozen and the resulting ANN models were used for forecasting.

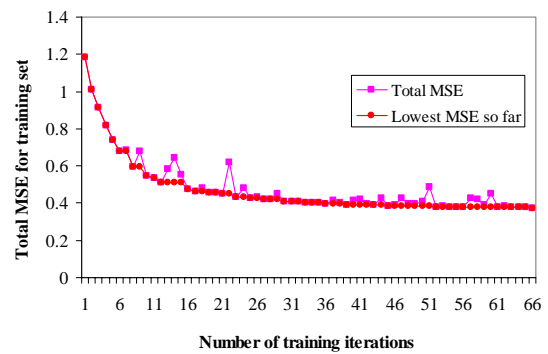


Fig. 5 Variation of error during learning for ANN-I

Table I below summarizes the characteristics of the ANN models developed to forecast overflow rates.

TABLE I BASIC CHARACTERISTICS OF THE ANN MODELS		
ANN Model	ANN - I	ANN - II
Input layer nodes	18	12
Hidden layer nodes	9	6
Input flow rate variables	$Q(t-1)$ to $Q(t-6)$	
Input rainfall variables	$R(t-5)$ to $R(t-16)$	$R(t-5)$ to $R(t-16)$
Training set	Year 1 to Year 3	Year 1 to Year 3
Testing set	Year 4 to Year 10	Year 4 to Year 10

## V. RESULTS AND DISCUSSION

### A. ANN - I Model Forecasts

The results for the training set show that the forecast

overflow rates closely follow the actual ones. Fig. 6 shows the actual flow rates plotted against the forecast overflow rates while Fig. 7 shows the specific overflow events (leaving out the zero valued no overflow periods) concatenated for the training set.

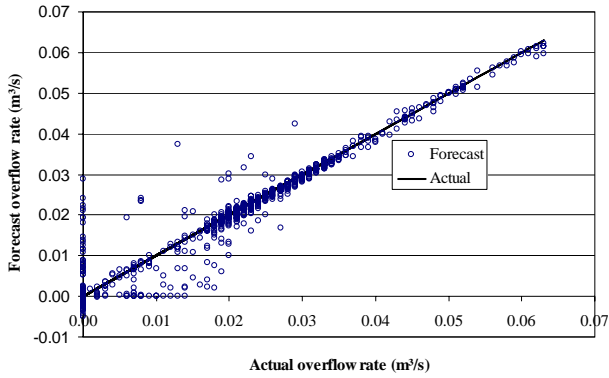


Fig. 6 Performance of the ANN-I model for training set

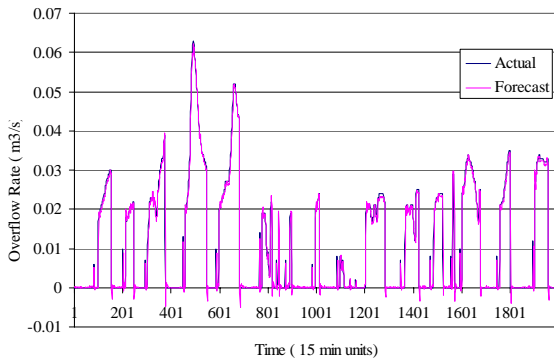


Fig. 7 ANN-I forecasts and actual values for training set

With a correlation coefficient of 0.98, the agreement between the actual and forecast values for the training set is very satisfactory for the ANN-I model.

Figs 9 through 14 show the correlation between the forecast and actual over flow rates for each year in the testing set. The agreement of forecasts at higher overflow rates are very good as indicated by the close fit to the straight line at the upper end. However, there is considerable scatter at the lower end of forecasts, particularly those below 0.03 m<sup>3</sup>/s.

Table II gives a summary of the statistics of the forecasts. The root mean squared error (RMSE) expressed as a percentage of the actual mean value of the overflow rate gives a measure of the disagreement of the forecasts with the actual values. This varies from 12.7% to 18.3% for the years in the training set. Overflow rate in Year 6 (0.064m<sup>3</sup>/s) goes marginally outside that in the training set (0.062m<sup>3</sup>/s) and the forecast of this high value is almost perfect which seems to indicate that the modified normalization technique has been effective.

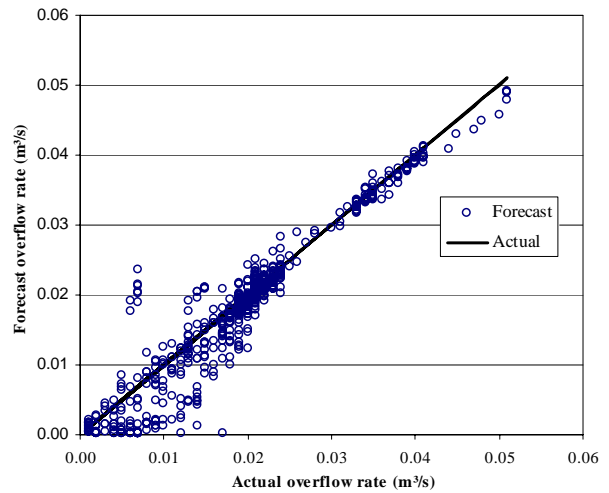


Fig. 8 Forecast and actual overflow rates for Year 4

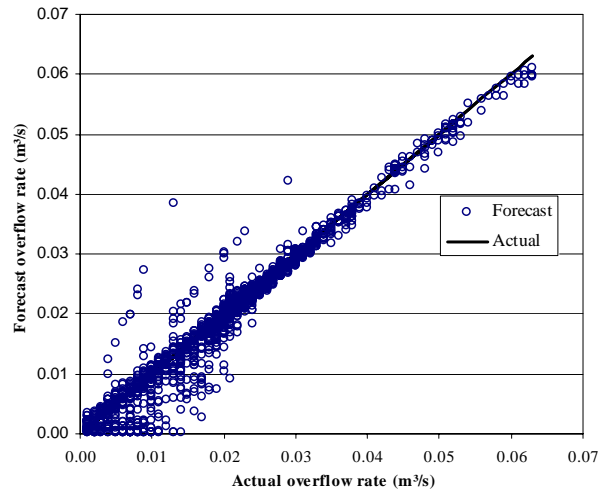


Fig. 9 Forecast and actual overflow rates for Year 5

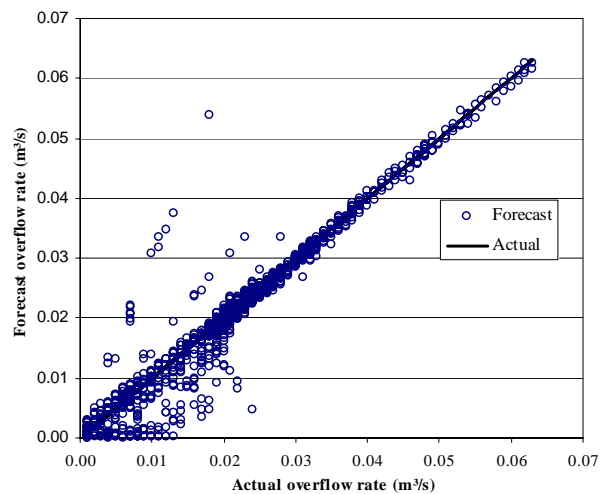


Fig. 10 Forecast and actual overflow rates for Year 6

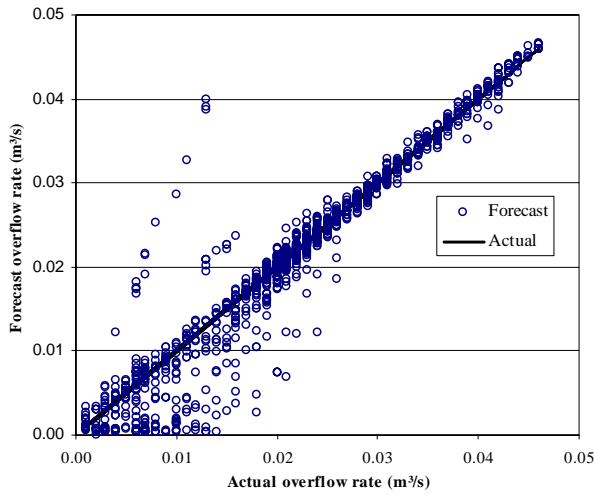


Fig. 11 Forecast and actual overflow rates for Year 7

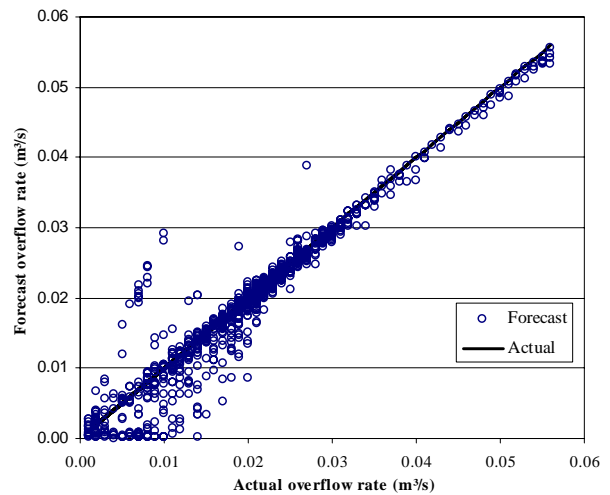


Fig. 14 Forecast and actual overflow rates for Year 10

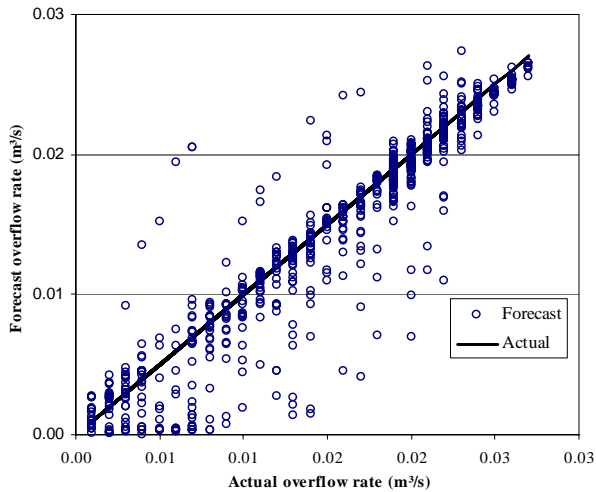


Fig. 12 Forecast and actual overflow rates for Year 8

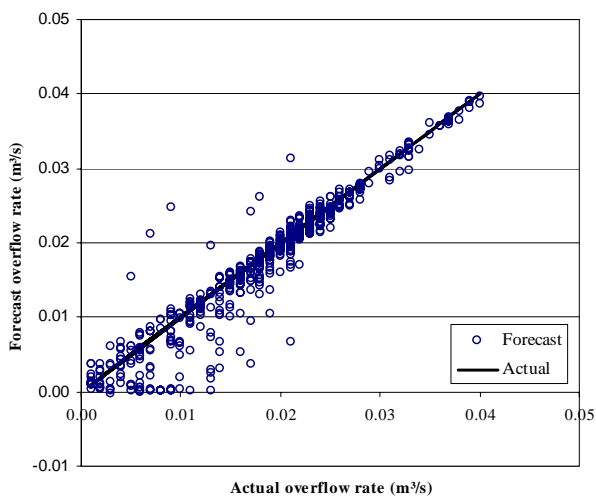


Fig. 13 Forecast and actual overflow rates for Year 9

TABLE II  
 THE STATISTICAL PROPERTIES OF THE FORECASTS FROM ANN-I MODEL

Year	Actual Mean (m <sup>3</sup> /s)	RMSE expressed as a % of Actual Mean
Training set (Years 1-3)	0.010	25.1
Year 4	0.019	16.6
Year 5	0.021	12.7
Year 6	0.022	13.7
Year 7	0.022	13.8
Year 8	0.015	18.3
Year 9	0.020	16.0
Year 10	0.020	13.9

*B. ANN – II Model Forecasts*

The ANN-II has difficulty learning a pattern relating the antecedent rainfall to future overflow rates. As can be seen in the Fig. 15 below, the agreement between the forecast and actual overflow rate is extremely poor for the training set revealing a complete contrast to Fig. 7. The performance on the testing sets, although not presented here, proved to be even poorer. This leads to the conclusion that the ANN-II model, based on the rainfall input alone, cannot effectively forecast the overflow rates. The ANN-I model, however, is effective and gives good agreement between the forecast and actual overflow rates.

It can be concluded that the ANN-I model can be used for overflow forecasting provided that this predictive tool can be connected an on-line to a monitor installed at the overflow structure. Installing telemetry to critical or important overflow structures for this purpose can be done easily.



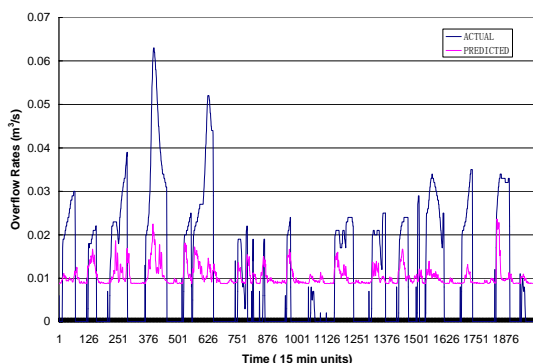


Fig. 15 ANN-II forecasts and actual values for training set

#### ACKNOWLEDGMENT

The financial support from Unitec New Zealand to present this paper at the International Conference on Computer Science (ICCS2006) in Vienna, Austria is thankfully acknowledged.

#### REFERENCES

- [1] R. H. Dwight, J. C. Semenza, D. B. Baker and B. H. Olson, "Association of urban runoff with coastal water quality in Orange County, California" *Water Environment Research*, Water Environment Federation, pp. 82-90, Jan/Feb 2002
- [2] Water Environment Research Foundation, "WERF Project Investigates Sanitary Sewer Overflow Prediction Technologies", Available <http://www.werf.org/press/fall99/page6.cfm>.
- [3] S. Birikundavvi, R. Labib, T. Trung and J. Rousselle, "Performance of neural networks in daily streamflow forecasting", *Journal of Hydrological Engineering*, Vol. 7, No. 5, pp. 392-398, 2002
- [4] A. Jain and S. Srinivasulu, "Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural networks techniques", *Water Resources Research*, Vol. 40, doi:10.1029/2003 WR002355, (2004) Americal Geophysical Union, 2004.
- [5] D. A. K. Fernando and A.W. Jayawardena, "Runoff forecasting using RBF networks with OLS Algorithm", *Journal of Hydrologic Engineering*, ASCE Water resources Engineering division, Vol. 3, No. 3, pp. 203- 209, July 1998.
- [6] M. Tomasino, D. Zunchettin and P. Traverso, "Long term forecasts of River Po discharges based on predictable solar activity and a fuzzy neural network model", *Hydrological Sciences Journal*, Vol. 49, No. 4, pp. 673-684, August 2004.
- [7] Y. Lin, Y. Wang, Y. Li, B. Zhang and G. Wu, "Earthquake prediction by RBF neural network ensemble", *Advances in Neural Networks*, Proc. of the International Symposium on Neural networks, China, Ed. F. Yin, J. Wang, and C. Guo, pp. 962-969, Springer, August 2004.
- [8] D. E. Rumelhart, R. Durbin, R. Golden, and Chauvin, "Backpropagation: The Basic Theory, In *Mathematical Perspectives on Neural Networks*", Eds. P. Smolensky, M. C.Mozer, and D. E. Rumelhart, Lawrence Erlbaum Associates, Mahwah, NJ., pp. 533-566, 1996.
- [9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning internal representations by error propagation*, In *Parallel Distributed Processing*, Vol. 1, Eds., D. E. Rumelhart, and J. L. McClelland, MIT Press, Cambridge, MA. pp. 318-362, 1986.
- [10] Wasserman, P.D., *Advanced methods in neural computing*, Van Nostrand Reinhold, New York, 1993.
- [11] M. A. Rao, J. Srinivas, *Neural networks: algorithms and applications*, Pangbourne, Eng.; Alpha Science International, 2003.
- [12] Fernando, A.K. and Kerr, T., "Runoff forecasting with Artificial Neural Network Model", The 3<sup>rd</sup> Pacific Conference on stormwater and aquatic resource protection, Auckland, New Zealand Water and Wastes Association publication of conference proceedings in CD-ROM, 2003.

- [13] H. K. Cigizoglu "Estimation, forecasting and extrapolation of river flows by artificial neural networks" *Hydrological Science, Journal*, Vol. 48, No. 3, pp. 349-361, June 2003.
- [14] M. Campolo, A. Soldati, P. Acdreussi, "Artificial Neural Network approach to flood forecasting in the River Arno" *Hydrological Sciences Journal*, 48(3), pp. 381-398, June 2003
- [15] H. R. Maier and G.C. Dandy, "Determining inputs for neural network models of multivariate time series", *Microcomputers in Civil Engineering*, Vol. 12, pp. 353-368, 1997.
- [16] S. Amari, N. Murata, K. R. Muller, M. Finke and H.H. Yang, "Asymptotic statistical theory of overtraining and cross-validation" *IEEE Transactions on Neural Networks*, Vol. 8, No. 5, pp. 985-996, 1997.
- [17] W. Wang, P.H.A.J.M. Van Gelder and J.K. Vrijling, "Some issues about the generalization of neural networks for time series prediction" in *Advances in Neural Networks*, Proc. of the International Symposium on Neural networks, August 2004, China, Ed. F. Yin, J. Wang, and C. Guo, Springer, pp. 559-564, 2004
- [18] L. Prechelt, 1998, Early stopping – but when? In G. B. Orr, K-R Mueller, Eds. *Neural networks: Tricks of the trade*, Berlin, Springer, pp. 55-99, 1998..

**Achela K. Fernando** received her B.Sc. (Eng) in Civil Engineering from the University of Peradeniya, Sri Lanka in 1987, M.Eng. in Environmental and Sanitary Engineering from Kyoto University, Japan in 1992 and Ph.D. from the University of Hong Kong in 1997.

She has been working as a lecturer in the School of Built Environment of Unitec New Zealand since July 2003 and prior to that has worked for civil and environmental engineering consultancy practices in Sri Lanka, Hong Kong and New Zealand.

Dr. Fernando is a Member of the Institution of Professional Engineers in New Zealand and a Chartered Professional Engineer.

**Xiujuan Zhang** completed her Masters postgraduate in Municipal Engineering at Taiyuan University of Technology in Shanxi Province, China in 1995. Currently she is a student in the Environmental Engineering discipline in the School of Built Environment of Unitec New Zealand.

**Peter F. Kinley** received his BE (Civil) from the University of Auckland in 1996.

He has been working as Modelling Team Leader at Metrowater in Auckland, New Zealand, since June 2002 and has previously worked for a range of local authorities, consultants and contractors around New Zealand.

Mr Kinley is a Member of The Institution of Professional Engineers New Zealand and is a Chartered Professional Engineer in New Zealand.