

# Infilling streamflow data using feed-forward back-propagation (BP) artificial neural networks: Application of standard BP and pseudo Mac Laurin power series BP techniques

M Ilunga<sup>1\*</sup> and D Stephenson<sup>2</sup>

<sup>1</sup> Civil Engineering, Wits University, P Bag X03, WITS 2050, South Africa

<sup>2</sup> Department of Civil Engineering, University of Botswana

## Abstract

Hydrological data (e.g. rainfall, river flow data) are used in water resource planning and management. Sometimes hydrological time series have gaps or are incomplete, or are not of good quality or are not of sufficient length. This problem seems to be more prevalent in developing countries than in developed countries. In this paper, feed-forward artificial neural networks (ANNs) techniques are used for streamflow data infilling. The standard back-propagation (BP) technique with a sigmoid activation function is used. Besides this technique, the BP technique with an approximation of the sigmoid function by pseudo Mac Laurin power series Order 1 and Order 2 derivatives, as introduced in this paper, is also used. Empirical comparisons of the predictive accuracy, in terms of root mean square error of predictions (RMSEp), are then made. A preliminary case study in South Africa (i.e. using the Diepkloof (control) gauge on the Wonderboomspruit River and the Molteno (target) gauge on Stormbergspruit River in the River summer rainfall catchment) was then done. Generally, this demonstrated that the standard BP technique performed just slightly better than the pseudo BP Mac Laurin Orders 1 and 2 techniques when using mean values of seasonal data. However, the pseudo Mac Laurin approximation power series of the sigmoid function did not show any substantial impact on the accuracy of the estimated missing values at the Molteno gauge. Thus, all three the standard BP and pseudo BP Mac Laurin orders 1 and 2 techniques could be used to fill in the missing values at the Molteno gauge. It was also observed that a linear regression could describe a strong relationship between the gap size (0 to 30 %) and the expected RMSEp (thus accuracy) for the three techniques used here. Recommendations for further work on these techniques include their application to other flow regimes (e.g. 4-month seasons, mean annual extreme, etc) and to streamflow series of a winter rainfall region.

**Keywords:** Infilling, feed-forward backpropagation, artificial neural network, pseudo Mac Laurin power series

## Introduction

For planning, management and effective control of water resource systems, a considerable amount of data on hydrological variables such as rainfall, streamflow, etc. are required. Very often in some developing countries, hydrological data sequences at a given network have gaps or are incomplete, or are not of good quality or are not of sufficient length (This problem is more prevalent in developing countries than in developed countries). This can severely affect the reliability of the design of, e.g. a hydropower plant, the construction of dams, etc. Generally, in those countries, the overwhelming majority of gaps are caused by temporary absence of observers, the cessation of measurement or absence of observations prior to the commencement of measurement (Makhuvha et al., 1997) or by limited financial resources (Balek, 1992).

Several streamflow hydrological data infilling techniques have been used, e.g. artificial neural networks (ANNs), regression methods, etc. Despite the criticisms formulated against

ANNs techniques, these techniques were found to be powerful tools when compared to multivariate regression based models for infilling the missing data (Panu et al., 2000). ANNs techniques can be used to express a non-linear mapping between variables with no prior assumptions on the variables (linear or non-linear as in regression methods) and these techniques can cope with missing data (French et al., 1992). In the past decade, ANNs have been used intensively in hydrology and water related fields. However, apart from a few papers published (Panu et al., 2000; Khalil et al., 2001; Elshorbagy et al., 2000), its application for infilling streamflow data remains sparse.

In this paper, feed-forward ANNs techniques are used for streamflow data infilling. On one hand the standard back-propagation (BP) with a sigmoid function (Freeman and Skapura, 1991) is used and on the other hand the BP technique with an approximation by pseudo Mac Laurin power series (Order 1 and Order 2 derivatives) to the sigmoid function, as introduced in this paper, is also used. Empirical comparisons of the predictive accuracy, in terms of root mean square error of predictions (RMSEp) are then made. A preliminary case study is made to demonstrate the performance of these three techniques. In what follows, the terms algorithm and technique can be used interchangeably. McL1BP and McL2BP will mean pseudo Mac Laurin Order 1 derivative and Order 2 derivative respectively (refer to Figs. 2, 3, 4, 6 and 7).

\* To whom all correspondence should be addressed.

☎ +2711 771 7152; fax: +2711 339-1762;

e-mail: [masengoi@civil.wits.ac.za](mailto:masengoi@civil.wits.ac.za)

Received 9 March 2004; accepted in revised form 17 December 2005.

## Streamflow data infilling techniques

### Artificial Neural Networks (ANNs) overview

ANNs are networks of interconnected simple units (nodes) that are based on a greatly simplified model of the brain. There are two main types of ANNs, i.e. feed-forward networks (where the signal is propagated only from the input nodes to the output nodes) and recurrent networks (where the signal is propagated in both directions). The advantage of the ANNs, even if the “exact” relationship between sets of inputs and outputs data is unknown but is acknowledged to exist, is that the network can be trained to learn that relationship, requiring no prior underlying assumptions (non-linear vs. linear) as in conventional methods, and they are regarded as ultimate black-box models (Minns and Hall, 1996). ANNs seek to learn patterns, but not to replicate the physical processes in transforming input to output (Minns and Hall, 1996). As opposed to conventional methods, the ANNs are thought to have the *ability to cope with the missing data* and, perhaps most importantly, are able to generalise a relationship from the small subsets of data whilst remaining relatively robust in the presence of *noisy or missing inputs*. Thus, ANNs can learn in response to a changing environment (Wilby and Dawson, 1998). Since the early nineties, ANNs have been successfully used in the area of water resource engineering related to rainfall/runoff forecasting (Minns and Hall, 1996; French et al., 1992; Agarwal and Singh, 2001) and infilling streamflow data (Panu et al., 2000; Khalil et al., 2001; Elshorbagy et al., 2000, etc.). However, apart from the above-mentioned applications of ANN techniques, the application of ANNs in infilling hydrological data remains sparse. For infilling streamflow data, these authors developed ANN techniques for cases where data exist before and after missing gaps (e.g. consecutive missing values). Three-layered ANNs have been intensively used in that respect.

### Standard back-propagation (BP) technique

The standard BP technique is only outlined in this section and for more details, the reader is referred, for example, to Freeman and Skapura (1991). Given a three-layered ANN as depicted in Fig. 1, in standard BP the adjustment of the interconnecting weights during training employs a method known as *error back-propagation* in which the weight associated with each connection is adjusted by an amount proportional to the strength of the signal in the connection and the total measure of the error. The total error at the output layer is then reduced by redistributing this error value backwards through the hidden layers until the input layer is reached. This process is repeated until the total error for all data sets is sufficiently small. The weights change equations on the output layer and hidden layers are respectively:

$$w_{kj}^0(t+1) = w_{kj}^0(t) + \eta \delta_{pk}^0 i_{pj}, \text{ and} \quad (1)$$

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \eta \delta_{pj}^h x_i \quad (2)$$

where:

- i = unit node in the input layer
- j = unit node in the hidden layer
- p = pattern and k is related to the output layer
- $\eta$  = learning rate
- $\delta_{pk}^0$  and  $\delta_{pj}^h$  = error terms (which encompass a derivative part) for output units and hidden units respectively
- t = t<sup>th</sup> iteration.

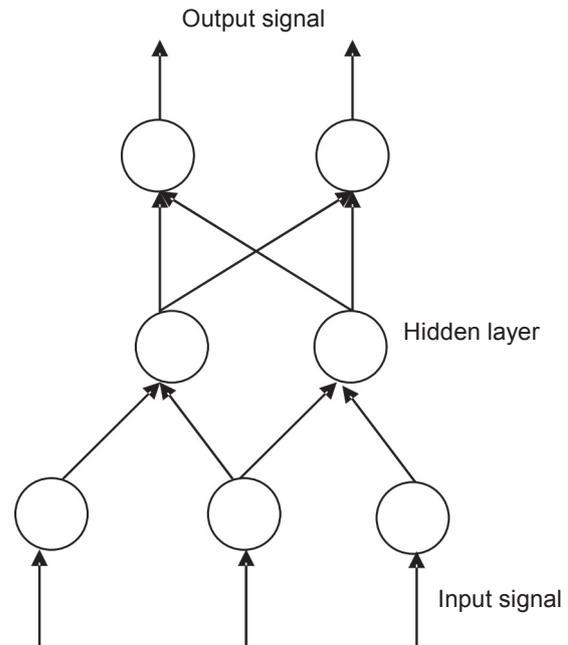


Figure 1  
A three-layer feed-forward ANN

For practical considerations, it is sometimes suggested to remove the bias terms altogether: their use is optional (Freeman and Skapura, 1991).

In the standard BP, basically the learning process is done through sequential mode and batch mode. In the former mode of learning the process of learning is governed by the error of each data set one by one while in the latter mode weights at each iteration are adjusted only after all data sets have been processed.

An activation function is used to express the non-linear relationship process between the input and output data. This function can be any threshold function or any continuous function. It is normally a monotonic non-decreasing function and differentiable everywhere for x values. The activation function most commonly used is a sigmoid, non-linear continuous function between 0 and 1 and is represented as:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Freeman proposed that a range of x values from 0.1 to 0.9 should be used for practical purposes. This range will be adopted in this paper. Thus, the input data and output data will be scaled (during training of ANNs) to fall under the above-mentioned range. A linear scaling was used here as in Hines (1997). Scaling input data and output data has the advantage on the speed of convergence of the system and it gives each input equal importance and prevents premature saturation of the activation function (Hines, 1997). Therefore, the formulas used in this paper should not contain any unit as they apply to scaled numbers during training of ANNs.

The first derivative of the sigmoid activation function, which is used in the updated Eqs. (1) and (2), is given by:

$$f'(x) = f(x)(1 - f(x)) \quad (4)$$

The standard BP (which is a gradient descent method) has been criticised because convergence to an optimal solution is not al-

ways guaranteed (Agarwal and Singh, 2001). Thus, several variants of the BP such as Newton's method, Adaptive step-size and the Levenberg-Marquardt algorithm were proposed. Despite these criticisms, it appears in practice that the BP leads to solutions in almost every case and that standard multilayer feed-forward networks are capable of approximating any measurable function to any desired degree of accuracy, as stated by Minns and Hall (1996).

In the following section a modification to the standard BP, by approximating the sigmoid function by "pseudo" Mac Laurin power series Order 1 and 2 derivatives, is introduced.

### Standard BP technique with sigmoid function approximated by pseudo Mac Laurin power series

This technique is the same as the one outlined in the previous section but with the only difference that a Mac Laurin power series approximation was applied to the sigmoid activation as follows:

$$f(x) = \frac{1}{1+e^{-x}} = \frac{1}{2-x+\frac{x^2}{2}-\frac{x^3}{6}+\frac{x^4}{24}-\dots} \quad (5)$$

The Mac Laurin power series (which is actually a particular case of a Taylor power series) approximates the function  $f(x)$  when  $x$  approaches zero. In other words, for small values of  $x$  such that  $0 < x \ll 1$ , a good approximation of  $f(x)$  can be achieved by a Mac Laurin power series. The Mac Laurin first order derivative approximation of Eq. (3) is given by:

$$f(x) \approx \frac{1}{2-x} \quad (6)$$

The derivative of Eq. (6) is given by:

$$f'(x) \approx \frac{1}{(2-x)^2} \approx (f(x))^2 \quad (7)$$

Similar to Eq. (4), Eq. (7) can be used in the weights update equations of the neural network.

The Mac Laurin second order derivative approximation of Eq. (3) is given by:

$$f(x) \approx \frac{1}{2-x+\frac{x^2}{2}} \quad (8)$$

The derivative of Eq. (8) is given by:

$$f'(x) \approx \frac{1-x}{(2-x+\frac{x^2}{2})^2} \approx (1-x)(f(x))^2 \quad (9)$$

Similar to Eq. (4), Eq. (9) can also be used in the weights update equations of the neural network. Like the sigmoid function, Eqs. (7) and (9) are also continuous, monotonic non-decreasing functions and differentiable under the interval (0; 0.9).

For this paper, no strict limitation on the range of values of  $x$  (e.g.  $x$  is greater than 0 but approaching 0) was set for the application of the Mac Laurin power series. However, the Mac Laurin power series approximation is just applied to an interval such that  $0 < x < 1$ , e.g. (0.1; 0.9), for scaled input and output data. That is why the prefix "pseudo" is introduced. The Mac Laurin (Order 1 and Order 2) approximation is done purposely for this interval just to evaluate the impact on the accuracy of the estimated missing values.

### Data availability

A preliminary test was done with mean values of seasonal naturalised streamflow data of two rivers belonging to the Orange River drainage system (D) of South Africa, specifically in the secondary drainage region D1 of the Eastern Cape (Midgley et al., 1994). The geographical location of these rivers, located in the summer rainfall region, is given below (refer to Table 1). The mean monthly flows and the mean annual runoff for the selected rivers are given in Table 2 and Table 3 respectively. Two seasons of a 6-month period each were assumed (wet-October to March, and dry-April to September). This was considered just to test preliminarily the approach as presented in this paper. Generally speaking, four seasons should have been considered for South Africa. This has been suggested in the conclusion. Recall that Pegram (1997) found that the months of October and September could fall into earlier summer (e.g. wet) and dry seasons respectively. The D1H004 gauge (Molteno) was taken as the target gauge and D1H001 (Diepkloof) as the control gauge. The hydrological year starts in October and ends in September.

### Results and discussion

The selected streamflow data set was complete and thus exhibited no gaps. However, for testing of the different infilling techniques (i.e. the standard BP and the pseudo Mac Laurin Order 1 and Order derivatives BP techniques), some consecutive gaps (e.g. 6.7 %, 13.3 %, 20 %, 30 % of missing data, and starting in 1934) were created on the target streamflow gauge data set, e.g. D1H004. It was noticed that starting gaps earlier (e.g. 1928) on the record of the target gauge D1H004 did not sensitively have any impact on the accuracy of the estimated values for the different techniques. The three techniques were applied to mean monthly flows. The ANNs were trained in a sequential

**TABLE 1**  
**Geographical location of rivers**

Gauge	Name	River	Latitude	Longitude	Area (km <sup>2</sup> )	Period of records used	% Missing
D1H001	Diepkloof	Wonderboomspruit	31000'03''	260 21'11''	2 397	1924-53	0
D1H004	Molteno	Stormbergspruit	310 24'00''	260 22'17''	348	1924-53	0

**TABLE 2**  
**Mean monthly flows (million m<sup>3</sup>/month) for gauges D1H001 and D1H004**

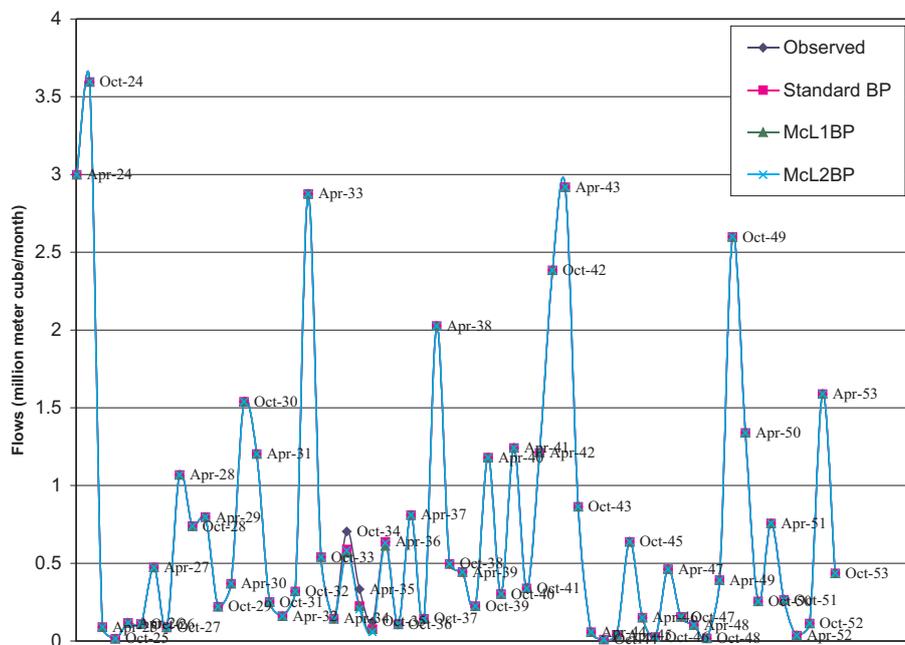
Gauge	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.
D1H001	1.57	3.02	2.99	4.22	6.51	11.75	4.97	3.69	1.04	0.51	2.92	2.30
D1H004	0.24	0.77	0.71	0.74	0.89	1.45	0.86	0.66	0.12	0.19	0.05	0.21

Gauge	MAR (million m <sup>3</sup> )
D1H001	44.271
D1H004	7.858

Algorithm	RMSEp (million m <sup>3</sup> /month)			
	6.7%	13.3%	20%	30%
Standard BP	0.082	0.180	0.304	0.505
Mac Laurin order 1 BP	0.093	0.20	0.292	0.524
Mac Laurin order 2 BP	0.095	0.185	0.307	0.520

mode on the concurrent parts of observed data and the weights obtained were then used to estimate the missing values. A single input-output ANN with three nodes in the hidden layer was used and the bias terms were assumed to be zero as their use is optional (Freeman and Skapura, 1991). The learning rate was set to 0.35 throughout for quite reasonable results, although a wide range of values (e.g. between 0.01 and 0.9) for the learning rate was tried. Input and output values were scaled to fall within the range 0.1 to 0.9 as mentioned earlier.

Table 4 summarises the results from the three techniques, i.e. the standard BP and the pseudo Mac Laurin (Order 1 and Order 2 derivatives) BP techniques.



**Figure 2**  
Mean monthly seasonal flows at D1H004 (6.7 % missing data from 1934)

From Table 4, generally, it follows that the RMSEp increases with increase in the proportion of missing values (gap size) for all three techniques. Thus, the accuracy decreases as the gap size increases. Generally, the standard BP performs just slightly better than the pseudo Mac Laurin (Orders 1 and 2) BP. This could be due to the fact that the error terms in the updated Eqs. (1) and (2) which encompass a derivative part, are slightly bigger for pseudo Mac Laurin (Orders 1 and 2 derivatives) BP techniques than for the standard BP technique (e.g. the slopes of functions (6) and (8) are steeper than the one of function (3) within the range 0.1 to 0.9). However, the pseudo Mac Laurin approximation did not show any substantial negative impact on the accuracy of the estimated missing values. The graphical plots (refer to Figs. 2, 3 and 4) confirm these results, where the differences in estimated missing values at gauge D1H004 are generally small, except for Fig. 4 (20 % missing data) where the flows are exaggeratedly overestimated for the year 1939. Figures 5, 6 and 7 show the root mean square errors of predictions (RMSEp), thus the accuracy vs. the gap size (% of missing values), at gauge D1H004 for the standard BP, pseudo Mac Laurin (Orders 1 and 2 derivatives) BP algorithms respectively. From Figs. 5, 6 and 7, it is seen that for all algorithms, the bigger the gap size, the bigger the RMSEp, thus the accuracy becomes increasingly less. However, it is observed from these figures that a linear regression can strongly describe the relationship between the gap size and the expected RMSEp (thus accuracy) for the three techniques.

The coefficients of determination (which are very close) were found to be 0.972, 0.969, and 0.974 for standard BP, pseudo Mac Laurin Order 1 and Order 2 BP algorithms respectively (refer to R<sup>2</sup> values in Figs. 5 to 7). This correlates with the observation that the differences in estimated values were small for the respective techniques at different gap sizes (0 to 30%). It was noticed that increasing the number of data points (e.g. up to seven gap sizes: 6.7%, 10%, 13.3%, 15%, 20%, 25% and 30%) did not affect substantially the relationship between the gap size and the expected RMSEp for the three techniques.

From the results obtained here, it can be said that all three the standard BP and pseudo Mac Laurin Orders 1 and 2 BP algorithms are acceptable to fill in the missing values for gauge D1H004. This can be done within the range 0 to 20 % without any significant violation of either the accuracy of estimated values or the statistical properties (i.e. the mean and the variance of the incomplete and infilled series).

## Conclusion and suggestions

Besides the standard BP algorithm, two other techniques, viz. the pseudo Mac Laurin (Order 1 and Order 2 derivatives) BP have been introduced for scaled input and output data in the interval (0.1; 0.9). These preliminary

results showed that the pseudo Mac Laurin approximation does not affect substantially the accuracy of the estimated values at gauge D1H004, when compared to the standard BP. Thus, both techniques were acceptable to fill in the missing values. However, it was observed that a linear regression could describe a strong relationship between the gap size and the expected RMSEP for the three algorithms under investigation. It is suggested that the impact of the Mac Laurin power series of order relatively higher (e.g. 3, 4 etc.) on the estimated values also be investigated. The batch-training mode has to be tried and other activation functions (e.g. hyperbolic tangent) as well. The techniques herein evoked should also be tested on other data sets. Recall that these techniques have been applied to mean values of seasonal streamflow data. Other flow regimes should also be tried (4-month seasons, mean annual, extreme, etc.). The three techniques should also be applied to streamflow series of a winter rainfall region.

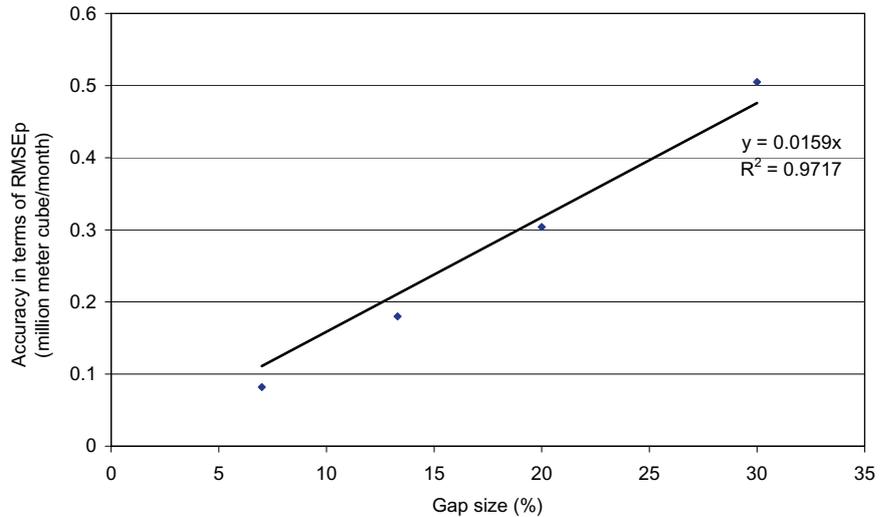
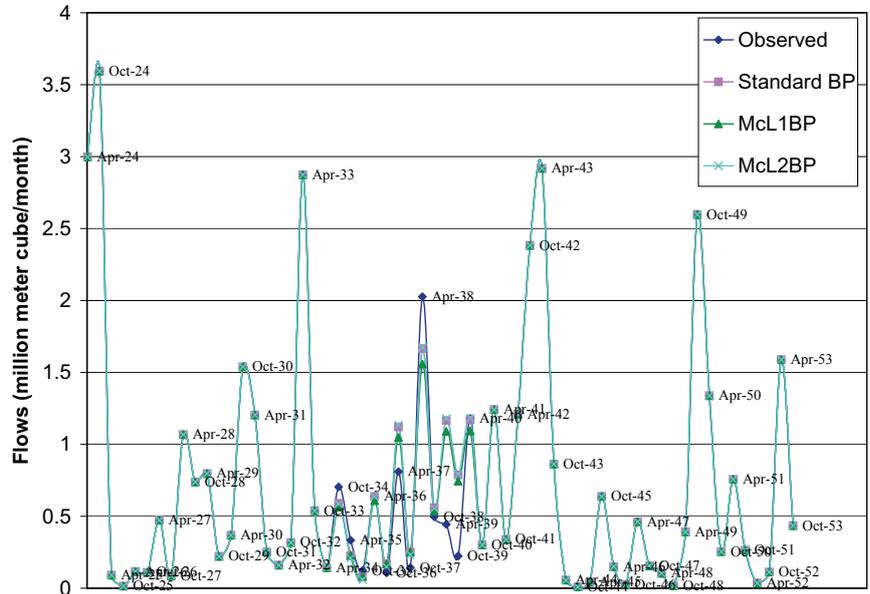
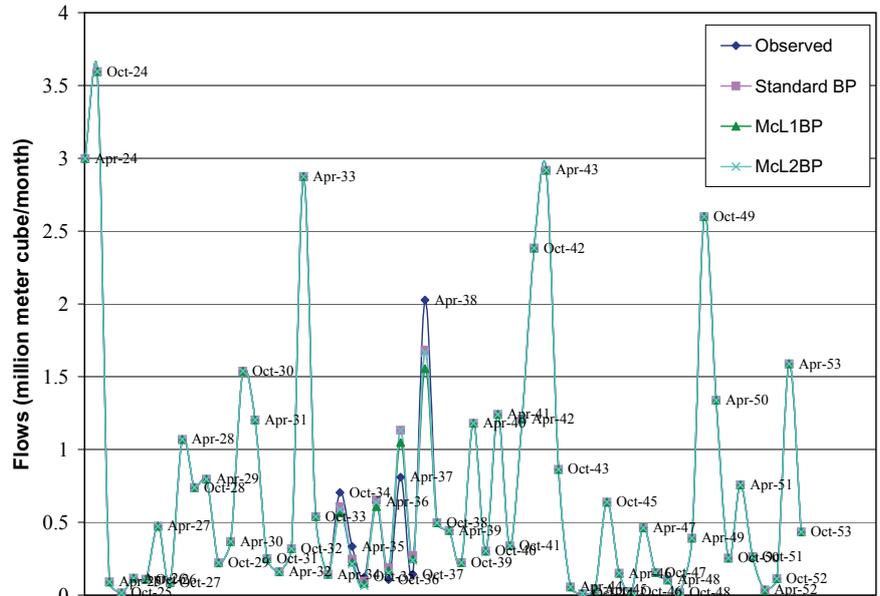
**References**

AGARWAL A and SINGH JK (2001) Pattern and batch learning ANN process in rainfall-runoff modeling. *Indian Association of Hydrologists. J. Hydrol.* **24** (1) 1-14.

BALEK J (1992) An application of the inadequate hydrological data of the African tropical regions in engineering design. *Proc. 2<sup>nd</sup> Int. Hydrol. Symp.*, 29-37. Date and place?

ELSHORBAGY AA, PANU US and SIMINOVIC SP (2000) Group-based estimation of missing hydrological data: I. Approach and general methodology. *J. Hydrol. Sci.* **45** (6) 849-866.

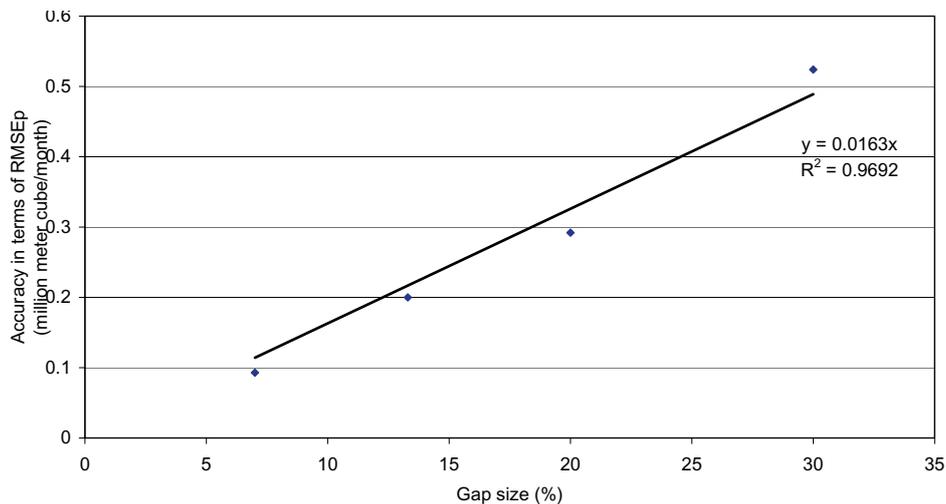
FREEMAN AJ and SKAPURA, MD (1991) *Neural Networks: Algorithms, Application and Programming Techniques*. Addition Wesley Publications Company, Inc.



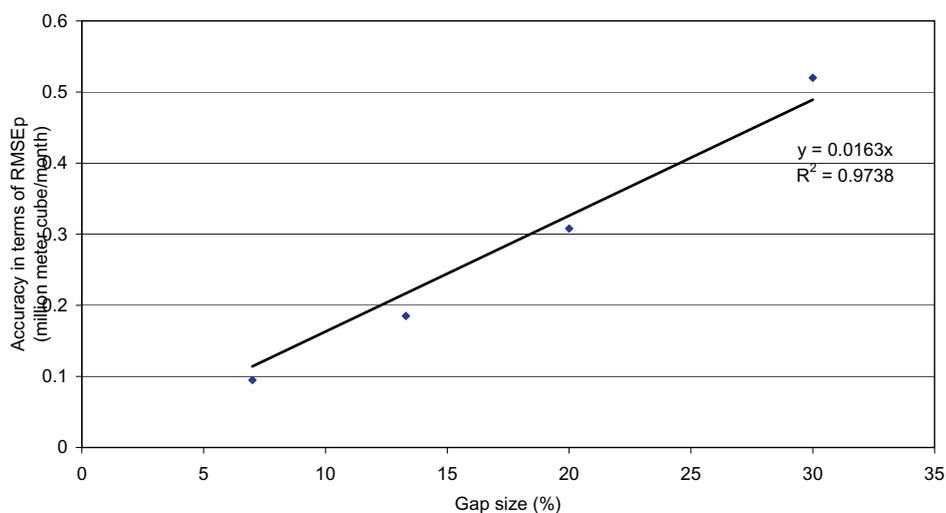
**Figure 3 (Top right)**  
Mean monthly seasonal flows at D1H004 (13.3 % missing data from 1934)

**Figure 4 (Middle right)**  
Mean monthly seasonal flows at D1H004 (20 % missing data from 1934)

**Figure 5 (Bottom right)**  
Accuracy vs. gap size at D1H004 (Standard1BP)



**Figure 6**  
Accuracy vs. gap size at D1H004  
(McL1BP)



**Figure 7**  
Accuracy vs. gap size at D1H004  
(McL2BP)

FRENCH N, KRAJEWSKY F and CUYKENDALL R (1992) Rainfall forecasting in space and time using neural network. *J. Hydrol.* **137** 1-31.

HINES W (1997) MATHLAB supplement to fuzzy and neural approaches in engineering. *Adaptive and Learning Systems for Signals Processing, Communication and Control*. Johnson Wiley & Sons, Inc.

KHALIL M, PANU US and LENNOX WC (2001) Group and neural networks based streamflow data infilling procedures. *J. Hydrol.* **241** 153-176.

MAKHUVHA T, PEGRAM G, SPARKS R and ZUCCHINI W (1997) Patching rainfall data regression methods. Best subset selection, EM and pseudo-EM methods: theory. *J. Hydrol.* **198** 289-307.

MIDGLEY DC, PITMAN WV and MIDDLETON BJ (1994) Surface Water Resources of South Africa, 1990. WRC Report No. 298/3.1/94 (1<sup>st</sup> edn.).

MINNS AW and HALL MJ (1996) Artificial neural network as rainfall-runoff models. *J. Hydrol. Sci.* **41** (3) 399-417.

PANU US, KHALIL M and ELSHORBAGY A (2000) Streamflow data Infilling Techniques Based on Concepts of Groups and Neural Networks. *Artificial Neural Networks in Hydrology*. Kluwer Academic Publishers. Printed in the Netherlands. 235-258.

PEGRAM G (1997) Patching rainfall data using regression methods, 3. Grouping, patching and outliers detection, *J. Hydrol.* **198** 319-334.

WILBY R and DAWSON CW (1998) An artificial neural network approach to rainfall-runoff modeling. *J. Hydrol. Sci.* **43** (1) 47-66.