

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^{1*} and D Stephenson²

¹ Civil Engineering, Wits University, P Bag X03, WITS 2050, South Africa

² 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

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