Interpolating Gaps in River Suspended Sediment Records using Artificial Neural Networks

Michael Gorman, Rick Millane
Dept. of Electrical and Computer Engineering
University of Canterbury
Christchurch, New Zealand
rick.millane@canterbury.ac.nz

Murray Hicks
NIWA
Christchurch, New Zealand
murray.hicks@niwa.co.nz

***Abstract*—Sediment transport is important in the management of rivers, catchments and floodplains. Sediment is monitored by measuring the suspended sediment concentration in rivers, but these measurements can be interrupted by various sensor malfunctions, leading to gaps in the record. An artificial neural network is developed for predicting suspended sediment concentration in these gaps, and is trained using high sampling rate contiguous records of quickflow and sediment concentration. The approach is evaluated by application to records from the Motueka River in the South Island of New Zealand.**

Keywords—rivers; suspended sediment; artificial neural networks; ANN

I. Introduction

Suspended sediment transport in rivers is important in many areas of environmental and resource management, and in engineering projects. These include effects on fish habitat, river aesthetics, reservoir filling and capacity, widening of flood plains, water quality and filtration, erosion around structures, channel navigation, hydro equipment damage and dam planning, to name a few [1-2]. Prediction or measurement of suspended sediment concentration (SSC) is therefore of importance in many areas of science and engineering. The suspended sediment in a river is made up of granules of soil and other materials, and in most rivers the diameter of these granules is usually smaller than ten microns [3]. The sediment concentration is the mass of the sediment per volume of fluid, usually measured in mg/L. Sediment yield is the result of a complex interaction of geomorphological processes, and depends on basin and river characteristics such as topography, land cover, land use, and climate. The high variability of these parameters makes modelling of the sediment process cumbersome and prediction of the sediment loading difficult.

Sediment transport in rivers is traditionally estimated using so-called sediment rating curves (SRCs) which are an empirical relationship between SSC and flow rate, Q, of the form

$$SSC = aQ^b \tag{1}$$

where a and b are constants that are determined by a regression analysis of data from a particular river under particular conditions. However, this relationship is over-simplified and inaccurate [4], is poor at predicting sediment load peaks, and can be in error by up to an order of magnitude [5,6].

The flow rate Q of a river, usually in m^3s^{-1}, can be measured using various methods such as with a pressure sensor coupled with river cross-section information, or by measuring water level coupled with appropriate calibration measurements, to provide a continuous record. Sediment concentration is commonly measured with two different methods. The first is an auto sampler which consists of a magazine of single use containers which collect a water sample which is then analysed for SSC in a laboratory. However, due to the limited magazine capacity, sampling of the SSC using an auto sampler is quite limited. The second method uses a turbidity sensor which is based on measurements of light scattering from the suspended particles. Continuous records of river SSC can be obtained using a turbidity sensor calibrated with water samples collected during flood runoff. Often, however, the turbidity record has gaps due to episodes of sensor lens bio-fouling, sensor over-ranging (i.e., the in-river turbidity exceeds the range of the sensor), power failure, flood damage to the sensor or its mounting, or other sensor malfunction. Typically, record is lost during flood events, which is when the data are most valuable if the purpose of monitoring is to determine the sediment load. Data from a concurrently operating auto sampler is generally too sparse to fill these gaps. The lost record may be replaced with SSC estimated from an SRC if the flow rate is concurrently measured but, as noted above, simple SRC relationships can vary considerably within and between high-flow events, on a seasonal basis, and over years due to long-term changes in sediment supply within the catchment, such as following a large storm event, earthquake, or land use conversion. As a result, SSC estimates from a simple, time-stationary function can be in error by up to an order of magnitude.

An alternative approach to estimate SSC is to use an adaptive time-varying estimator based on flow rate and SSC measurements in the vicinity of the missing record. In view of the complex nonlinear nature of the SSC time series, here we investigate the use of an artificial neural network (ANN) to patch gaps in the SSC record. The network is trained using data from a Q/SSC dataset collected from the Motueka River in the South Island of New Zealand and tested by interpolating artificially introduced gaps in the SSC record.

There have been other investigations of the use of ANNs to model SSC in rivers. Three such studies are briefly reviewed here. In all cases, predictions were generally superior to those obtained using simple SRCs or regression analysis. Kisi [5] applied ANNs to a two-year dataset of flow rate and SSC from USGS data from two locations in Puerto Rico. Data were sampled at one-day intervals and one year of data was used for training and one year for testing. For both locations, the best ANN inputs to estimate the current SSC were found to be the current and one previous flow rate. Addition of the previous SSC as an input added little to the accuracy of the predicted SSC. Alp and Cigizoglu [7] used 6.5 years of one-day sampled flow rate and rainfall data from three locations in the catchment, from the Juniata River in Pennsylvania. Five years of data were used for training and one year for testing. Rainfall data alone were found to be a poor predictor. For flow data alone, the current and two previous samples gave the best SSC prediction. Incorporating rainfall data, as a weighted average over the three rainfall sites, the current and previous flow rate and rainfall gave a slightly better prediction than flow rate alone. Kumar et al. [8] used 14 years of daily flow rate, SSC and rainfall data from a number of locations in the catchment of the Sutley River in Northern India. They used 10 years of data for training and 4 years for validation. They concluded that a 6-input ANN gave the best predictions, with inputs consisting of the current and previous flow rate, the previous SSC and the current rainfall rate at three locations.

In this study we consider a somewhat different dataset to the above studies in that we have available the flow rate and SSC at 15 minute intervals. We also consider input flow rates to the ANN that can consist of either total flow and/or quickflow (which is the component of storm runoff that drains quickly from a catchment – described further below). Furthermore, the objective is different in that we aim to fill in gaps in the SSC record rather than to predict the SSC.

II. ANN Implementation

We used a three-layer feed forward ANN consisting of an input layer, a single hidden layer and an output layer. The single output node is the estimate of the SSC at the current time index. The number of input nodes was variable as we investigated various combinations of possible inputs. The number of hidden layer nodes was chosen to be equal to 4. Every input node is connected to every hidden layer node though a weight, which for simplicity is denoted w_i where i indexes all the weighted interconnections in the network. Each hidden layer node consists of a summer and a tangent sigmoid transfer function. The network is trained by updating the weights using a backpropagation algorithm. During the training, the output associated with a particular input is compared with the known output and the weights updated by the equation

$$w_{i(\text{new})} = w_{i(\text{old})} + n(y_d - y)x_i \tag{2}$$

where $w_{i(\text{new})}$ is the updated weight, $w_{i(\text{old})}$ is the current weight, n is the "learning rate" of the network, y_d is the desired output, y is the predicted output, and x_i is the input associated with weight w_i.

The ANN was implemented using the FANN library [9]. The application was written in Python to give a portable executable and is multi-threaded.

III. Results

The dataset consisted of flow rates and SSCs collected in the Motueka Catchment in the northern South Island of New Zealand (Fig. 1). The data used in this study were collected at the mainstem Motueka at the Woodman's Bend site which captures the discharge of sediment to the coast. The data were collected over a six year period in 2003-2008 at 15-minute intervals, giving a total of about 210,000 data samples [10]. The SSC was measured using a turbidity sensor which was calibrated to SSC with samples collected by an automatic sampler throughout this period. Short gaps in the turbidity record were generally patched using SSC data derived directly from an auto-sampler. Gaps in the turbidity data were flagged in the data record. Flow rate was measured in the standard way, using a water level record and water-level/discharge rating curves.

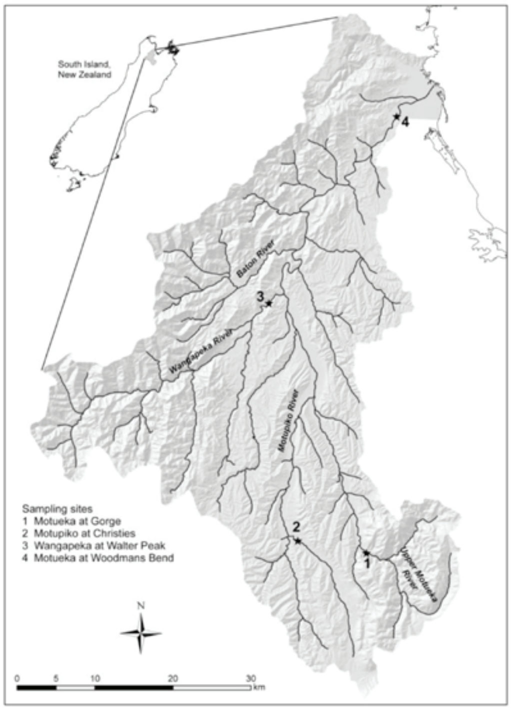

Fig. 1 The Motueka Catchment showing the Motueka at Woodman's Bend measurement site (top) and an image showing discharge of suspended sediment into Tasman Bay (bottom).

Peaks in the SSC occur during floods where there is a transient increase in the flow rate above the base flow of the river. Therefore, the increased flow above the base flow, the so-called quickflow, was considered as a potentially useful variable for predicting the SSC. The flow data were separated into base flow and quick flow components. The base flow $B(t)$ was calculated using a simple hydrograph slope separation algorithm as

$$B(t) = \min(Q(t)\, ,\, B(t-1) + Ct) \qquad (3)$$

where t is time and the constant C is the maximum rate of change of the base flow. The quickflow is $F(t) = Q(t) - B(t)$. For this dataset, the base flow and quickflow were calculated with $C = 0.0062$ Ls^{-2}. Examples of the base flow and total flow, and of the flows and SSC time series, are shown in Fig. 2.

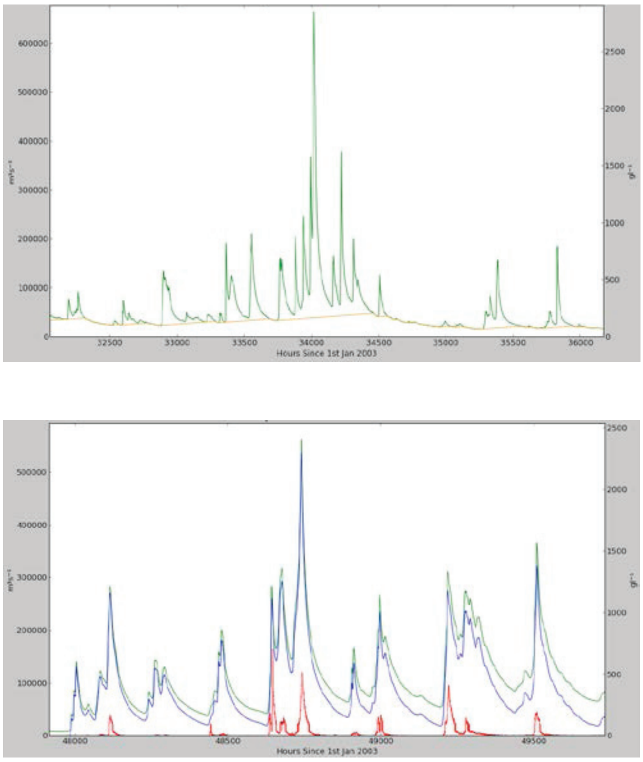

Fig. 2 Examples of base flow (brown) and total flow (green) (top), and a typical time series showing total flow (green), quickflow (blue) and SSC (red) (bottom).

An ANN learning rate of 0.7 was used. Initial experiments with training and testing showed that the quickflow data are more effective than the total flow in predicting the SSC, so only those data were used as inputs. Quickflow data over the previous 400 samples were considered as potential inputs. The best previous quickflow samples to use as inputs were determined by a greedy forward search feature selection algorithm. The ANN is first trained with a single input using each of the quickflow samples individually as the input, and the prediction error calculated. The input with the lowest error is retained. A second input is then added and the error calculated over the remaining options for the second input. The second input with the lowest error is retained. This process is continued in this fashion, adding an additional input at each stage until the error starts to increase. Using this algorithm, 7 quickflow inputs were found to be optimum, consisting of data with delays of 0, 1, 48, 112, 189, 211, 214 samples. The error versus the number of inputs is shown in Fig. 3. A simple incorporation of previous SSC data as an input was not successful and is discussed further in the Conclusions section.

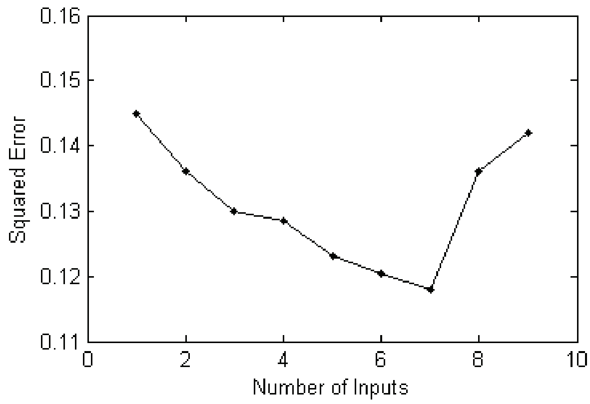

Fig. 3 The squared error versus the number of inputs to the ANN.

Tests showed that the ANN predicted the SSC reasonably well in magnitude and also predicted the correct phasing of the SSC peak relative to the discharge peak during events. This is important because SSC often tends to peak earlier than water discharge.

An example of the predicted and actual SSC for a 70 day period is shown in Fig. 4, and the approximately correct magnitude and phasing is evident. The mean absolute error between the predicted and actual SSC is 0.3%. For this simulation, a 7000-sample contiguous segment was removed from the dataset and the remaining data used for training. The removed segment did not have any gaps in the SSC record. The SSC in the removed segment was then predicted using the trained ANN. For this dataset, the gaps in the SSC record varied in length between 1 and 28 days. Therefore, the ability to predict a 70-day period is sufficient to patch these gaps.

A second simulation used the SSC calculated from the ANN to predict the accumulated sediment load for the 6-year period, which was compared with the actual load. The accumulated sediment load $L(T)$ is given by

$$L(T) = \int_0^T S(t)Q(t)dt \qquad (4)$$

and is a quantity of importance since it shows the total amount of material transported by the river over time period T. In this case, the data were divided into two sets, consisting of 2/3 and 1/3 of the data, which were used for training and testing,

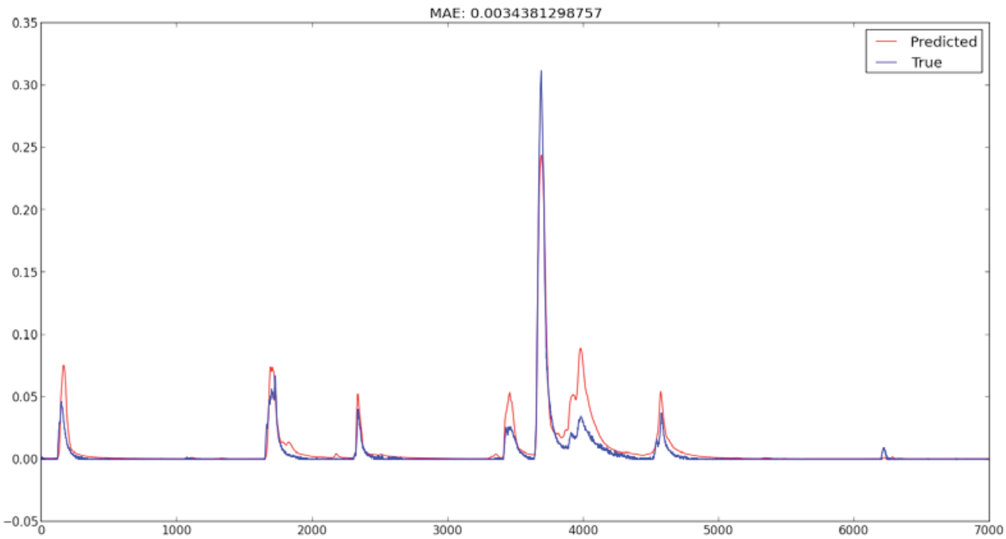

Fig. 4 Predicted (red) and actual (blue) SSC over a 70 day period. The x-axis is time in samples (15-minutes) and the y-axis is a scaled SSC.

respectively. The testing set consisted of every third quickflow datum, and the training set consisted of the remainder of the samples. The ANN was trained using this training set. The SSC in the testing set were then predicted using the trained ANN. The actual and predicted accumulated sediment load was then calculated using (4) using the actual and predicted SSC from the testing set. The predicted SSC in the actual SSC gaps were not used in calculation of the predicted sediment load so as to provide a valid comparison with the actual sediment load calculation. The actual and predicted loads calculated in this way are shown in Fig. 5. The agreement is seen to be reasonably good. Note that the load estimated in this way is expected to be very close to one-third of the actual load over the 6 year period since one-third of the data were used and were evenly spread over the full period. The relative error in the final accumulated sediment load after 6 years is only 0.36%. However, the average accumulated error, i.e. the absolute difference between the predicted and actual loads

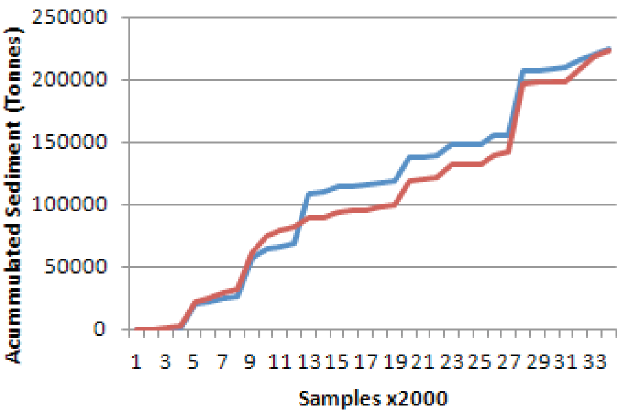

Fig. 5 Predicted (red) and actual (blue) accumulated sediment load as described in the text.

integrated over the period, was 18%. We consider this to be indicative of the method's accuracy for patching short-term data gaps, and this is expected to be superior to the accuracy obtained from attempting to fill such gaps using sediment rating curves.

IV. Discussion

The results presented here demonstrate the potential utility of using ANNs to interpolate gaps in SSC records. Quickflow data appear to be sensitive in predicting the SSC, which is perhaps not surprising since this increased flow during flood events is correlated with peaks in the SSC. The results are preliminary however, and indicate a number of useful avenues to pursue. The high temporal resolution of this dataset may be too fine-grained, and low-pass filtering of the data before input to the ANN might provide a more effective input for prediction. Incorporation of previous SSC as an input to the ANN is potentially useful, but is complicated by the presence of gaps in the SSC record. However, it is likely that a simple scheme for incorporating this data can be devised. The results presented here indicate that, at least for this data set, the gap-free data can be used to reliably predict the SSC over time periods that exceed the lengths of the gaps in the record, and are thus sufficient to patch these gaps. It is possible that characteristics of the gaps themselves, such as information on the distribution of gap lengths and gap spacings for example, could be used to advantage when interpolating in the gaps. Information on any correlation between the location of the gaps and the timing of flood events might also be relevant.

Acknowledgment

The data collection was funded by the Foundation for Research, Science and Technology under contract C09X0305 for Landcare Research's Integrated Catchment Management Programme.

REFERENCES

[1] P. McCully, Silenced Rivers: The Ecology and Politics of large Dams. London: Zed Books, 2001.

[2] J.K. Edzwald, Water Quality and Treatment, 6th ed. New York: McGraw-Hill, 2011.

[3] D.E. Walling and P.W. Moorehead, "The particle size characteristics of fluvial suspended sediment: an overview," Hydrobiologia, vol. 176/177, pp. 125-149, 1989.

[4] E.A. McBean and S. Al-Nassri, "Uncertainty in suspended sediment transport curves," J. Hydrologic Eng. ASCE, vol. 114, pp. 63-74, 1988.

[5] O. Kisi, "Suspended sedimant estimation usiong neuro-fuzzy and neural network approaches," Hydrolog. Sci. J., vol. 50, pp. 683-696, 2005.

[6] M. Demirci and A. Baltaci, "Prediction of suspended sediment in river using fuzzy logic and multilinear regression approaches," Neural Comput. Applic., vol. 23 (Suppl. 1), pp. S145-S151, 2013.

[7] M. Alp and H.K. Cigizoglu, "Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data," Environ. Model. Software, vol. 22, pp. 2-13, 2007.

[8] A.R.S. Kumar, C.S.P. Ojha, M.K. Goyal, R.D. Singh and P.K. Swamee, "Modeling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic, and decision tree algorithms," J. Hydrologic Eng. ASCE, vol. 17, pp. 394-404, 2012.

[9] http://leenissen.dk/fann/wp/

[10] L.R. Basher, D.M. Hicks, B. Clapp and T. Hewitt (2011), "Sediment yield response to large storm events and forest harvesting, Motueka River, New Zealand," N.Z. J. Marine and Freshwater Res., vol. 45, pp. 333-356, 2011.