Journal of Hydrology (2008) **355**, 192–201

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Structural optimisation and input selection of an artificial neural network for river level prediction

Paul Leahy *, Ger Kiely, Gearóid Corcoran

Centre for Hydrology, Micrometeorology and Climate Change, Department of Civil and Environmental Engineering, University College, Cork, Ireland

Received 27 August 2007; received in revised form 25 January 2008; accepted 18 March 2008

KEYWORDS
Rainfall-runoff modelling;
Artificial neural networks;
Structural optimisation;
Simulated annealing;
Flood prediction

Summary Accurate river level prediction, necessary for reliable flood forecasting, is a difficult task due to the complexity and inherent nonlinearity of the catchment hydrological system. Although artificial neural networks (ANNs) offer advantages over mechanistic or conceptual hydrological models for river level prediction, their applicability is limited by the fact that each ANN has to be specifically optimised and trained for a particular prediction problem and suitable input vectors selected. A recently developed novel optimisation algorithm combining properties of simulated annealing and tabu search is used to arrive at an optimal ANN for the prediction of river levels 5 h in advance. The algorithm seeks to minimise the value of a cost function based on the complexity and performance of the ANN. This is done by removing inter-neuron connections and adjusting the weights of the remaining connections. The candidate inputs presented to the algorithm were: current values of river levels at the flood location and two upstream locations; the change in level over the previous 4 h at the flood point, mean sea level pressure (SLP) and the change in SLP over the previous 24 h. The optimisation removed 79% of the network connections and three of the candidate inputs, leaving the current levels at the two upstream locations and at the flood point as the only inputs. The optimised ANN was then trained using the standard backpropagation algorithm. This methodology produces an ANN of greatly reduced complexity albeit with a reduced performance compared to an unoptimised ANN trained with backpropagation only. However, it has the advantage of being generally applicable and represents an improvement over trial and error as a method of ANN structural optimisation and input selection. For this prediction problem, current levels at two upstream locations and at the flood point are the best predictors of the level at the flood point.

* Corresponding author. Tel.: +353 21 4903816; fax: +353 21 4276648.
E-mail addresses: paul.leahy@ucc.ie (P. Leahy), g.kiely@ucc.ie (G. Kiely), g.corcoran@ucc.ie (G. Corcoran).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2008.03.017

Introduction

A successful flood management strategy requires accurate forecasting of river flows (White, 2001). Flows are dependent on diverse factors such as the spatial and temporal distribution of precipitation, topography, soil type, vegetation, land use, catchment hydrology and the built environment, making flood forecasting a difficult exercise. Approaches to flow prediction can be broadly divided into three categories: mechanistic modelling, statistical or 'black box' modelling and conceptual modelling. A typical mechanistic approach combines precipitation observations or forecasts with detailed physical models of the river catchment (Ivanov et al., 2004). The statistical approach is based on the properties of observed data (such as time series of river stages or precipitation) rather than on the physical properties of the catchment system itself. Conceptual modelling lies between these approaches, relying on a simplified representation of the physical system, which can be calibrated using past data.

Mechanistic modelling has the disadvantage of being data-intensive and therefore expensive, requiring spatially and temporally resolved meteorological input data and detailed physical descriptions of the characteristics of the catchment (Elshorbagy et al., 2000). The underlying physical system is complex and several of the driving factors may interact with each other in unpredictable ways (Jain and Prasad Indurthy, 2003). Conceptual models, while not requiring such detailed knowledge of the physical properties of the system, generally require some physical knowledge to formulate the model and some data in order to calibrate it. No knowledge of the underlying hydrological system is required for statistical modelling, but it is necessary to have previous information on the system's behaviour in order to derive such a model.

One purely statistical method, the artificial neural network (ANN), is a computational tool with the ability to represent a complex nonlinear system without any *a priori* knowledge of the underlying physical system itself. ANNs combine nonlinear functions of variables presented as inputs in order to model a prescribed output. The combination of functions is optimised via a process known as training the network in order to best match the output of the network with the desired or target value (Haykin, 1994).

This approach of using a network of interconnected, simple neurons to perform computations was introduced by McCulloch and Pitts (1943) as a representation of synaptic processes in the brain. Applications of ANNs in machine learning became widespread in the early 1980s with the advent of affordable microprocessors and by the late 1990s ANNs had been applied to many topics in water resource management and hydrology (see review by Govindaraju, 2000a), such as the modelling of flood events and the prediction of bacterial concentrations in seawater (Campolo et al., 1999; Kashefipour et al., 2005). Among the characteristics of ANNs that make them suited to such applications are: noise rejection; tolerance of errors and gaps in the input data; and the lack of any requirement for exogenous inputs (Nayak et al., 2005). A key feature of ANNs is their ability to generalise, i.e. to generate an output from a previously unseen specific combination of inputs (Haykin, 1994).

Kisi (2004) applied various ANN types and auto-regressive moving average (ARMA) models to the problem of forecasting monthly river flow based on past flows and concluded that ANNs provided better results than ARMA models. At shorter time scales, ANNs have also been found to perform better than both ARMA and conceptual models for daily streamflow forecasting and to outperform ARMA models in real-time forecasting of hourly river stage based on inputs of previous stages (Birikundayvi et al., 2002; Thirumalaiah and Deo, 2000). Toth et al. (2000) presented a hybrid approach, using various time series models (ARMA, ANNs and non-parametric nearest-neighbours) to generate short-term precipitation forecasts for a catchment, and then fed these predictions into a conceptual rainfall-runoff model in order to predict river runoff. Of the time series models investigated, ANNs provided the best prediction accuracies, especially at longer prediction intervals. Dawson et al. (2006) reported on a genetic algorithm to evolve neural network runoff prediction models for a catchment of 3315 km^2 in northern England, based on inputs of upstream stages and rainfall and found good performance at 6 h prediction intervals.

Supervised training methods such as backpropagation (Werbos, 1974; Rumelhart et al., 1986) are capable of optimising the connection weights of a given ANN, but a suitable ANN structure for the application must be chosen before training can take place. Trial and error is a widely used method of ANN structural optimisation (Hsu et al., 1997; Jain and Chalisgaonkar, 2000) but this approach is not rigorous and offers no guarantee of arriving at a truly optimal structure. Furthermore, trial and error becomes impractical for ANNs with large numbers of neurons and weights and is therefore of little use as a general approach to structural optimisation. Network growing techniques such as cascade-correlation learning and network pruning techniques such as weight decay have also been successfully used as means of structural optimisation (Fahlman and Lebiere, 1990; Weigend et al., 1991; Thirumalaiah and Deo, 2000). More recently, genetic algorithms have been employed in order to remove surplus weights and derive optimal ANN structures (Sexton et al., 2004; Dawson et al., 2006; Zanchettin and Ludermir, 2007). Constructive, rather than destructive, algorithms, where networks are built up from a minimal size instead of pruned from a maximal size, have also shown promise for structural optimisation of ANNs (Sexton et al., 2004).

Selection of suitable input data remains a problem in ANN applications. Bowden et al. (2005) presented two methodologies for ANN input selection: one based on partial mutual information (PMI) and one based on self-organising maps (SOM), generalised regression neural networks (GRNN) and genetic algorithms (GA). However, the methodology to be described here is more general in that it simultaneously addresses both the problem of input selection and the related problem of structural optimisation of the ANN itself.

Ludermir et al. (2006) proposed a two-stage hybrid global optimisation methodology for simultaneously optimising the structure and weights of an ANN using elements of both simulated annealing and tabu search (reproduced in Table 1). In this approach, network weights and structure are optimised together in the first stage, prior to a second stage of fine

Table 1 Optimisation algorithm (after Ludermir et al., 2006)

1. $s_0 := [C_0, W_0]$ (initial solution)
2. $T_0 := 1$ (initial temperature)
3. $s_{BSF} := s_0$ (initialisation of best solution so far)
4. for $i = 0$ to $I_{max} - 1$
5. if $(i + 1) = m\, I_T$ $[m = 1, 2, 3, \ldots, I_{max}/I_T]$
6. $T_{i+1} := rT_i$ (cool to new SA temperature)
7. Else
8. $T_{i+1} := T_i$
9. if (stopping criterion satisfied)
10. Break
11. Generate set of K new solutions from s_i
12. Choose best solution s' from set
13. if $f(s') < f(s_i)$
14. $s_{i+1} := s'$
15. else
16. $s_{i+1} := s'$ with probability $\exp(-[f(s') - f(s_i)]/T_{i+1})$
17. if $f(s_{i+1}) < f(s_{BSF})$
18. $s_{BSF} := s_{i+1}$
19. end for
20. Use backpropagation to train the network defined by s_{BSF}

tuning of the optimised network by conventional backpropagation training. A maximal complexity network is initially defined. The optimisation algorithm is iterated and allowed to randomly prune inter-neuron connections and adjust connection weights until the value of a cost function (based on the network prediction error and the number of remaining connections in the network) is minimised. In common with tabu search methods, several new candidate solutions are randomly generated during each iteration and the best (i.e. lowest cost) of these is selected and evaluated. If the lowest cost candidate solution has a lower cost than the current solution, then it replaces it. In common with simulated annealing, even if the lowest cost candidate solution has a higher-cost than the current solution, it may still replace it. A random variable determines whether such a transition to a less favourable, higher-cost solution is allowed. With each iteration of the algorithm the maximum allowed value of the transition probability is reduced by decreasing a controlling parameter known as the temperature. These unfavourable cost transitions allow the optimisation process to escape from local minima. A record of the lowest cost, or best so far (BSF), solution is retained throughout the optimisation process.

Unlike the weight elimination complexity regularisation method (Haykin, 1994), this method optimises the network structure *prior* to training by backpropagation. After the ANN structure has been optimised using the algorithm, conventional backpropagation is used to arrive at the final weights. This methodology also has the capability to eliminate redundant inputs by disconnecting them from the output.

The relationship between river flow and stage is straightforward in most river channels, and the problem of flood prediction can be reduced to the prediction of the river stage at the flood location corresponding to a flood event (Henderson, 1969). The objective of this study was to optimise and test ANN models for the prediction of river levels at Mallow, a town on the Munster Blackwater river in southwestern Ireland which is prone to regular flooding (Corcoran, 2004). We investigate if an optimised ANN with six or fewer inputs can provide accurate forecasts over a prediction interval of 5 h.

Catchment description

The Munster Blackwater catchment (Fig. 1) covers an area of 3324 km^2 in the southwest of Ireland. 1186 km^2 of the catchment is upstream of the town of Mallow, 45 m above sea level, and 90% of the vegetation cover in the catchment is grassland. The catchment drains from west to east, the flow being influenced by year round rainfall, transported by south-westerly prevailing winds. Significant evaporation losses only occur during the summer months of May to September. The annual average rainfall varies from 1465 mm to 980 mm (west to east) across the catchment with approximately 400 mm of annual evapotranspiration. January and July mean daily temperatures at Donoughmore (17.5 km SW of Mallow, 180 m a.s.l.) are 5.4 and 14.3 °C, respectively. The Blackwater extends 75 km upstream of Mallow (45 m a.s.l.) and the gradient (S_{1095}) of this section is 2.3 m km^{-1}.

Methods

Instrumentation and data

Three river level measurement stations were used in this study. Stations S1 and S2 are located at Duarrigle and Dromcummer, 38 km and 19 km upstream of the flood location at Mallow town (S3, Fig. 1 and Table 2). River stages were measured using both pulley-based shaft encoders (Thalimedes, OTT-Hydrometry, Germany) and radar sensors (Kalesto, OTT-Hydrometry, Germany) to provide redundancy in the event of a sensor failure and all levels were logged by multi-channel data loggers (Hydrosens, OTT-Hydrometry, Germany) every 15 min and transferred via GSM modem to a central archive. Sea level pressure was measured at the synoptic weather station at Cork Airport, approximately 50 km south of Mallow. Routines were developed in the MATLAB environment (Mathworks, USA) to create, optimise, train and evaluate the neural network models. Data gathered during the period 2001–2002 were used in the study. Approximately 36,000 records were suitable after some data were removed due to failing quality control tests.

ANN description

The multi-level perceptron (MLP; Rumelhart et al., 1986) is a widely used ANN configuration and has been frequently applied in the field of hydrological modelling (Govindaraju, 2000a). In an assessment of network types the MLP has been found to perform well, with superior generalization properties to the radial basis function network. However the MLP model is more difficult to optimise (Senthil Kumar et al., 2004). The basic processing units (the neurons) of a MLP are arranged in layers, each layer containing several neurons (Fig. 2). A MLP consists of an input layer, one or more

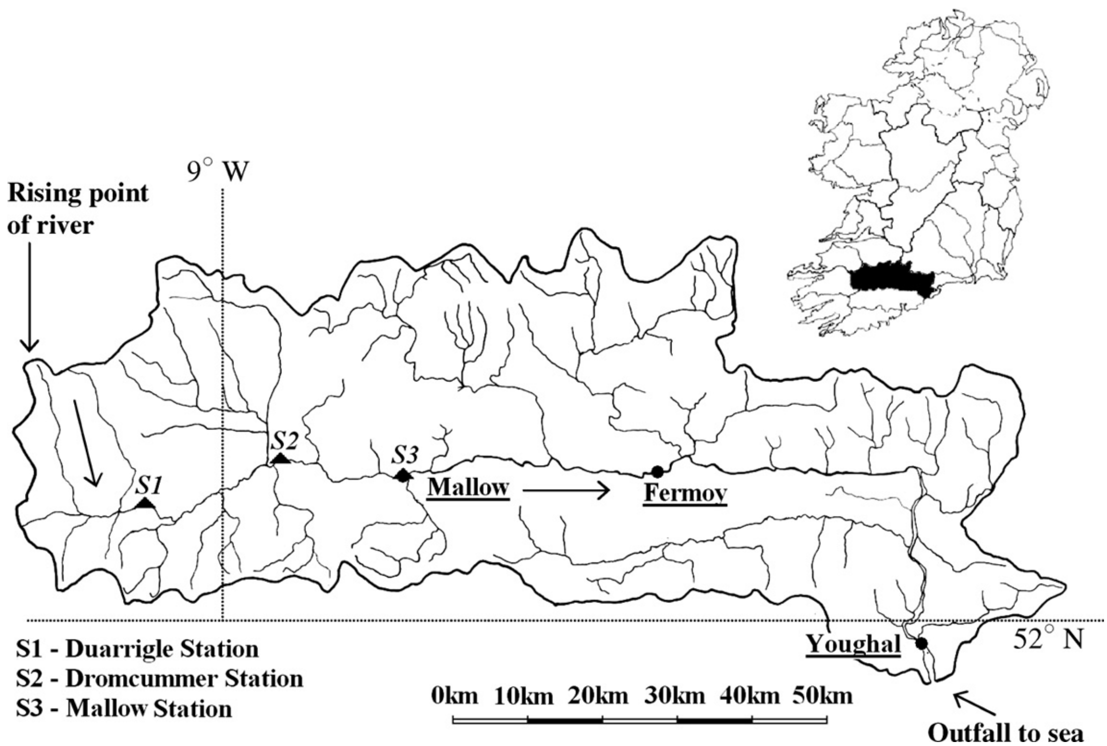

Figure 1 Map of the Munster Blackwater catchment showing river level monitoring stations. Arrows indicate direction of flow. Inset: map of Ireland showing location of catchment.

Table 2 Locations and physical details of river water level monitoring stations

Station ID	Location	Distance upstream of flood point (km)	% of subcatchment upstream of station	Elevation (m a.s.l.)
S1	Duarrigle	38	24	100
S2	Dromcummer	19	83	65
S3	Mallow	0	100	45

hidden layers and an output layer. Input data are presented to the input layer and then passed to each hidden layer in sequence, and finally to the output layer. The one-directional flow of data through the network results in this being known as a 'feed-forward' network.

The number of neurons in the input layer is equal to the number of inputs. The input layer performs no computations on the data, but merely distributes the values it receives to the first hidden layer. Normally, all the neurons of the input layer are connected to all the neurons of the first hidden layer (Fig. 2a). All connections between neurons have an associated weight. The neurons of the hidden layers calculate the weighted sum of their inputs. A nonlinear transformation known as an activation function is then applied to this weighted sum. The activation function serves two purposes: it scales the output value of the neuron to a normalised interval such as [0,1] and it introduces the nonlinearity necessary for the ANN to be able to represent nonlinear systems. A sigmoid function such as $(1 + \exp(-t))^{-1}$, where t is the untransformed neuron input, is a common choice for the activation function (Jain and Prasad Indurthy, 2003).

The optimisation algorithm of Ludermir et al. (2006) allows for inter-neuron connections to be randomly removed and the effects of these removals on network prediction performance to be evaluated. A maximal complexity network with six inputs and a single hidden layer containing eight neurons was defined as the starting point for the algorithm. Neuron biases were not used in this exercise.

Input selection

The selection of appropriate inputs is crucial to the success of any ANN model. Too many inputs will result in a parameter space that is too large to be efficiently optimised during training. Redundancy must be avoided in choice of inputs in order to reduce the risk of local minima being returned as optimisation solutions (Govindaraju, 2000b). The choice of inputs can be guided by knowledge of the physical system to be modelled. For example: upstream river levels are expected to be a good predictor of future river levels downstream in the same river system (Govindaraju, 2000a). Furthermore, there are practical reasons (such as the cost

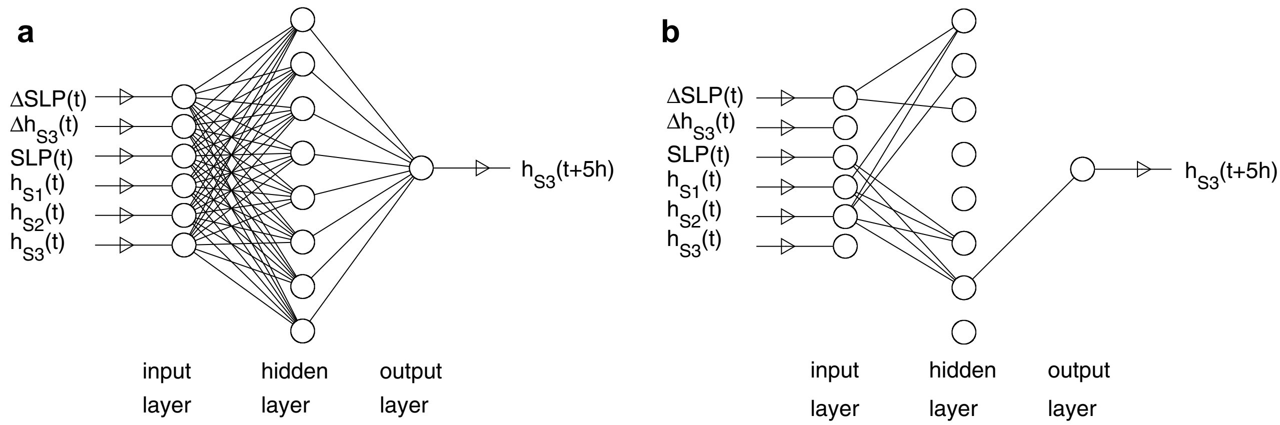

Figure 2 (a) Maximally connected six input ANN topology with a single hidden layer of 8 neurons, illustrating intra-neuron connections; (b) Optimised ANN topology with three inputs and reduced connectivity.

of purchasing and maintaining instrumentation) to keep the number of inputs to a minimum.

In order to determine suitable lag times for inputs, cross-correlations between each of the upstream (S1 & S2) measured stages and the downstream stage at Mallow (S3) were performed. The peak in the cross-correlation spectrum of S2 and S3 was at 135 min and the peak of the S1 and S3 spectrum was at 225 min. The correlation peaks are useful as guidelines for input selection, but they are not definitive, as the river system is incompletely gauged. Increased flow at the downstream location could emanate as lateral inflow from areas of the basin that are not upstream of either S1 or S2 (see Table 2 for the percentages of the catchment which are upstream of each monitoring point). It should also be noted that as the catchment drains predominantly from west to east and that the prevailing winds (carrying moisture from the Eastern Atlantic) come from the southwest, bands of precipitation may often follow the course of the streamflow, resulting in an increased proportion of the observed flow at S3 which has not passed S1 or S2. This makes the task of river stage prediction more difficult and may also have the effect of shifting the cross-correlation peak back in time (if compared to the ideal scenario of a point discharge upstream of S1). Hence, the correlation peaks should only be considered as lower bounds for the selection of lag times for inputs.

Modelling of systems that exhibit hysteretic behaviour can be achieved by incorporating previous values of input variables. This extra information should be sufficient to allow the ANN to distinguish the two limbs of the hydrograph curve. This approach has been used for modelling loop-rating curves (Jain and Chalisgaonkar, 2000) and soil water retention curves (Jain et al., 2004). We added an additional candidate input variable: the recent change in stage ($\Delta h/\Delta t$) at the downstream flood location at S3. Using such a measure rather than lagged values of actual stage should allow the network to distinguish the rising limbs of flood hydrographs ($\Delta h/\Delta t$ positive) from falling limbs ($\Delta h/\Delta t$ negative) more easily, reducing the prediction error due to the hysteresis of the system. The change in stage was calculated over 2, 4 and 6 h intervals and 4 h was found to provide slightly better results than 2 h, with the worst results from a 6 h interval (data not shown).

Sea level pressure (SLP) has been used with success to condition stochastic models of daily precipitation (Kiely et al., 1998) therefore current SLP and the change in SLP over the previous 24 h were incorporated as additional ANN inputs.

Thus, the following parameters were used as network inputs:

- Current river stage at all stations (S1, S2, and S3).
- Sea level pressure.
- Change in S3 river stage over the previous 4 h.
- Change in sea level pressure over the previous 24 h.

All input data were linearly mapped to the range [0,1].

Network structure optimisation

In addition to selecting the ANN inputs, the properties of the ANN's structure must be correctly chosen. Haykin (1994) recommends that the ratio of the number of training pairs to the number of neurons to be trained should be greater than 30. Given the large size of the training set, this is not a concern in this case, where we limit the total number of neurons in all layers to 15 (6 + 8 + 1). With a large number of potential inputs and ANN properties available to configure, it is not possible to test every single possible configuration. However, the structure may be optimised by defining an initial maximal structure and applying the methodology of Ludermir et al. (2006). Network complexity is represented by a parameter, γ, the percentage of network connections used. A connection, i, may be enabled or disabled, therefore each connection is represented by a connectivity bit, c_i, which has a value of one if the connection is used and zero if the connection is unused. After the network optimisation, some of the hidden layer neurons may be redundant if all their input connections have been removed (Fig. 2b). These neurons were removed prior to the subsequent step of backpropagation training.

If N_1 is the number of inputs, N_2 the number of neurons in the hidden layer and N_3 the number of outputs, the maximum possible number of connections is given by

$$N_{max} = N_1 N_2 + N_2 N_3 \quad (1)$$

and γ is then defined by

$$\gamma = \frac{100}{N_{\max}} \sum_{i=1}^{N_{\max}} c_i \tag{2}$$

The cost function is then defined as

$$f(s) = \frac{1}{2}(kE + (1-k)\gamma) \tag{3}$$

where k is a bias factor to allow adjustment of the relative importance of the number of network connections and the prediction error; E is a network performance indicator, such as the mean squared error (MSE; Eq. (4)) expressed as a percentage of the maximum value

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} (o_i - p_i)^2 \tag{4}$$

In Eq. (4), N is the number of test points and o_i (p_i) is the ith observation (simulation) of river stage at the flood point.

The parameters used for the optimisation were:

$I_{\max} = 200$ number of iterations of optimisation algorithm;
$N_{\text{epochs, max}} = 100$ maximum number of epochs for backpropagation training;
$N_1 = 6$ number of input layer nodes;
$N_2 = 8$ maximum number of hidden layer nodes;
$N_3 = 1$ number of output layer nodes;
$T_0 = 1$ initial simulated annealing temperature;
$r = 0.9$ geometric cooling factor;
$I_T = 10$ number of iterations between coolings;
$p = 0.2$ probability of inverting connectivity bits;
$n_K = 20$ number of new solutions to generate per iteration;
$k = 0.75$ weighting factor of prediction error in cost function.

The maximal network thus had 56 ($= N_1N_2 + N_2N_3$) internal connections. A total of five different network variants were applied to the prediction problem. Firstly a maximally connected network was created with six inputs, one hidden layer with eight neurons and a single output. A further four networks were derived from the maximally connected network, allowing the different parts of the optimisation algorithm to be compared. The maximal, unoptimised network was trained using classical backpropagation. The optimisation algorithm was applied to both the weights and connections of the untrained maximal network. In a further variant, only the network weights were optimised. Finally the two optimised variants were subsequently trained using backpropagation, giving the following complete list of networks:

[MaxC|NoOptW|BP]: initial, maximal network, no optimisation, trained with backpropagation;
[OptC|OptW|noBP]: weights and connections optimised and no backpropagation training;
[MaxC|OptW|noBP]: weights only optimised, fully connected, no backpropagation training;
[OptC|OptW|BP]: network with optimised weights and connections subsequently trained with backpropagation;
[MaxC|OptW|BP]: fully connected network with weights only optimised subsequently trained with back propagation.

Training

For each input element of the training set, there is a corresponding target value. Together, these are known as training pairs. During training, each training pair was fed into the ANNs and the weights of the model were iteratively adjusted using Levenberg–Marquardt backpropagation until the prediction error on the validation set failed to improve for five successive epochs or a limit of 100 epochs was reached. The use of the intermediate validation stage prevents overtraining of the ANN, a phenomenon where an ANN begins to model the noise in its training data as well as the underlying trends (Govindaraju, 2000b), resulting in poorer performance on unseen data. Finally, the prediction error was calculated on the previously unseen test set.

Performance measures

Legates and McCabe (1999) recommend the use of a combination of relative and absolute measures of goodness-of-fit when assessing model performance. The mean squared errors (MSE, Eq. (4)) is the absolute error value used in the network training phase. The MSE is computed for each epoch and the network parameters are adjusted in order to minimise the MSE. The MSE or sum of squared errors (SSE) from the test phase provides a good overall assessment of the accuracy of the model predictions. The coefficient of efficiency E_2 (Nash and Sutcliffe, 1970), and correlation coefficient, R, are relative measures of goodness-of-fit widely used in hydrological modelling. E_2 compares the model performance against using the mean as a predictor, with $E_2 = 1$ representing a perfect model. E'_n, a generalised modification of the coefficient of efficiency (Eq. (5)) uses the base flow rather than the mean flow as the reference point, as the baseline is more likely to be a good predictor than the mean for the majority of samples.

$$E'_n = 1 - \frac{\sum_{i=1}^{N} |o_i - p_i|^n}{\sum_{i=1}^{N} |o_i - o'|^n} \tag{5}$$

In Eq. (5); o' is the baseline flow; n is a positive nonzero integer (usually 2) and N is the total number of samples.

A test for comparing different learning algorithms has been proposed by Mitchell (1997) in which the available data is partitioned into k disjoint subsets of equal size, where a suggested value of k is at least 30. The networks to be compared are trained and tested k times. Each of the k subsets is used in turn as the test set, with the remaining $k-1$ subsets used to train the networks. The means of the errors over the k trials can be used to compare the learning algorithms and obtain a confidence interval.

Results

Fig. 3 and Animation 1 show an example of the optimisation algorithm in progress. Rapid decreases in both the value of the cost function and the network connectivity (γ) occur in the first 20 iterations of the optimisation. Updates of the best so far (BSF) network configuration

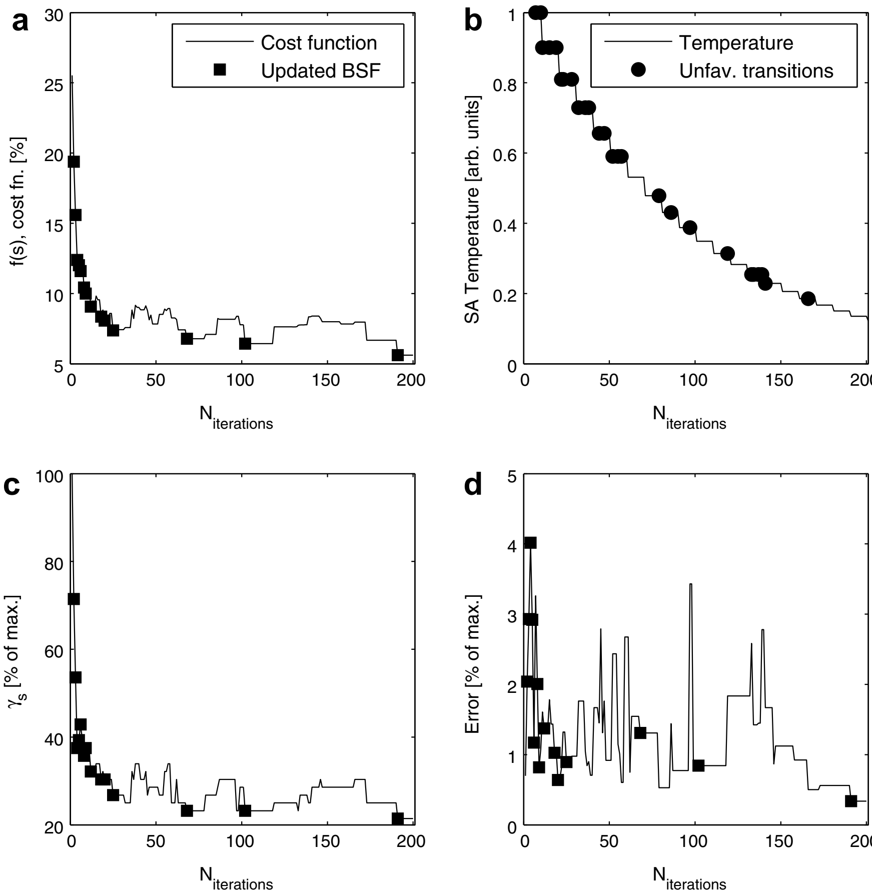

Figure 3 Evolution of: (a) the cost function, $f(s)$; (b) the simulated annealing temperature, with randomly generated unfavourable cost transitions indicated by filled circles; (c), the network connectivity parameter, γ, and (d), the simulation error, E over 200 iterations of the optimisation scheme, prior to network training. Filled squares in (a), (c) and (d) indicate best so far (BSF) optimisations.

are always associated with a decrease in the value of the cost function (Fig. 3a), which may be attributed to either a reduction in γ (Fig. 3c) or in the prediction error (Fig. 3d) or both. A comparison of Figs. 3b and 3c shows that several transitions to higher-cost solutions were permitted by the randomly generated transition probability within the first 50 iterations. After 50 iterations these transitions are less frequent as the temperature is cooled and they become less likely to be allowed. By iteration 67 most of the internal ANN connections have been removed and by iteration 102 only a single input remains connected to the output. However, this configuration represents a local minimum and is not a BSF solution and by iteration 119, one input has been reconnected (Animation 1).

All 30 trials of the fully optimised network [OptC|OptW|noBP] arrived at the same selection of three inputs from the six candidate inputs: namely the current river levels at the two upstream stations S1 and S2, and at the flood point, S3. All trials also converged to the same number of internal network connections (Table 3). The coefficient of variation in the performance over 30 trials was 22% for [OptC|OptW|noBP] and 10% for [OptC|OptW|BP], indicating that optimisation of the weights did depend to some extent on network initialisation and partitioning of data into training, validation and test sets. The coefficient of variation of performance over the 30 trials of the maximal network [MaxC|NoOptW|BP] was 23%.

Observed and simulated hydrographs for a single set of trials on a test dataset are shown in Fig. 4. Fig. 4a shows the fully optimised network before [OptC|OptW|NoBP] and after [OptC|OptW|BP] backpropagation training. The backpropagation-trained variant exhibits closer estimation of the baseflow, however both variants tend to overestimate smaller peaks. The fact that the test dataset is a random set of 1/30th of the available data means that some peaks, e.g. that of 9th March 2002 are not represented in the test set. In Fig. 4b hydrographs simulated by the two weights only optimised network variants are shown. After backpropagation training, the variant [MaxC|OptW|BP] provides better estimates of baseflows than the fully optimised variants shown in Fig. 4a, and peak flow magnitudes are well simulated. Finally, Fig. 4c shows the performance of the maximal network after backpropagation [MaxC|NoOptW|BP] which provides a simulated hydrograph with good representation of base and peak flows, similar to that of [MaxC|OptW|BP] in Fig. 4b.

The complexity and performance measures of each ANN are presented in Table 3. The ANNs with optimised connections retain only 21% of the connections of the maximal case, i.e. only 12 of the 56 initial connections remain in

Table 3 Averages over 30 trials of: complexity (γ), number of epochs before training was stopped, and performance on test set (as baseline-adjusted coefficient of efficiency, E_2', correlation coefficient, R, and SSE as percentage of maximum SSE) of: unoptimised, maximally connected ANN trained using conventional backpropagation (MaxC|NoOptW|BP); ANN with optimised connectivity and weights (OptC|OptW|noBP); maximally connected ANN with optimised weights (MaxC|OptW|noBP); ANN with optimised connectivity and weights and further training by backpropagation (OptC|OptW|BP); and maximally connected ANN with optimised weights and further training by backpropagation (MaxC|OptW|BP)

	MaxC|NoOptW|BP	OptC|OptW|noBP	MaxC|OptW|noBP	OptC|OptW|BP	MaxC|OptW|BP
γ	100	21	100	21	100
N_{inputs}	6.00	3.00	6.00	3.00	6.00
$N_{\text{epochs (BP)}}$	100.00	–	–	9.00	64.30
E_2'	0.99	0.83	0.85	0.90	0.98
R	0.99	0.89	0.88	0.91	0.98
SSE	0.17	1.95	1.72	1.16	0.26

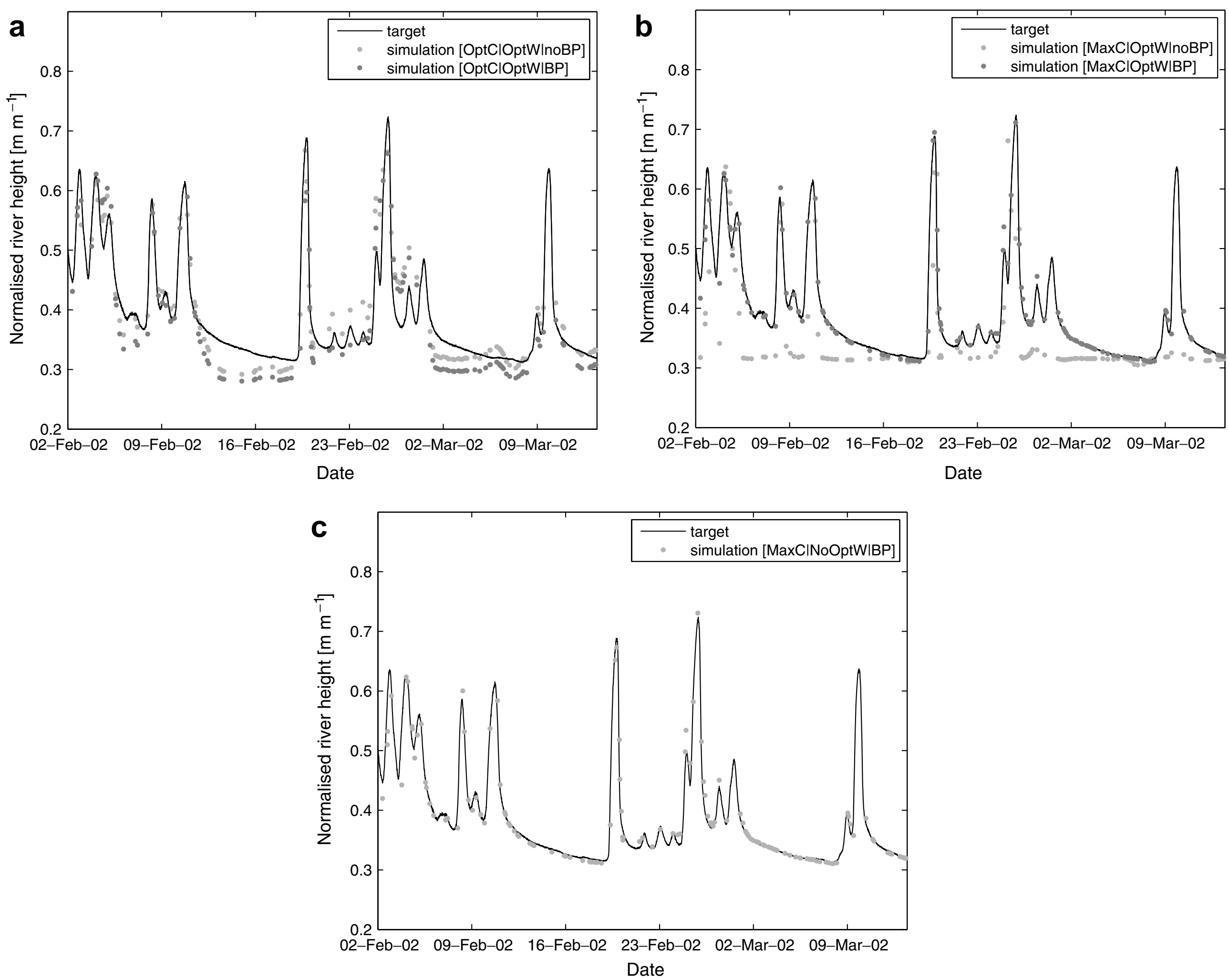

Figure 4 Target and normalised 5 h predicted river levels from trial number 30: (a) from fully optimised networks before [OptC|OptW|noBP] and after [OptC|OptW|BP] backpropagation; (b) from weights only optimised networks before [MaxC|OptW|noBP] and after [MaxC|OptW|BP] backpropagation; (c) from unoptimised, maximal network trained by backpropagation [MaxC|NoOptW|BP].

the fully optimised network [OptC|OptW|BP]. Furthermore, of the remaining connections, all but four were redundant (as they did not influence the ANN output) and were removed prior to backpropagation training. Of the six candidate input vectors presented to the optimisation algorithm, only three remain after full optimisation in all 30 trials. This implies that the current levels at the upstream stations and the flood location are better predictors of levels 5 h ahead, for this catchment than SLP, its recent change in value, or the recent change in level at the flood point.

Over the 30 trials, the fully optimised, trained networks, [OptC|OptW|BP] performed worse on test datasets than the trained maximal network [MaxC|NoOptW|BP] by a SSE margin of 0.99 ± 0.04 (where the 95% confidence interval was derived using the approach of Mitchell (1997)). The performance of the weight-only optimised networks [MaxC|OptW|BP] was close to the maximal networks, with a difference in SSE of 0.09 ± 0.03 (Table 3). The E_2' goodness-of-fit measure also shows the loss of network performance of [OptC|OptW|BP] (0.90) compared with [MaxC|NoOptW|BP] (0.99). The E_2' of [MaxC|OptW|BP] (0.98) was close to that of the maximal ANNs. A similar trend can be observed by comparing the correlation coefficients (R) of each variant.

Discussion

Current river levels were found to have more predictive power for this particular problem than sea level pressure, change in sea level pressure, or recent change in river level. The fact that all 30 trials of the optimisation [OptC|OptW|noBP] selected the same set of inputs from random network initial states and randomly partitioned training/validation/test data shows that the methodology has the ability to find a global solution regardless of initial conditions. In the case of the fully optimised ANN [OptC|OptW|BP], the optimisation process removed all but three inputs and twelve internal network connections, leaving a greatly reduced parameter space for the backpropagation algorithm.

In training of the fully optimised networks [OptC|OptW|BP] only nine epochs of backpropagation were required before the early stopping criterion was satisfied. During backpropagation of the weight-only optimised network variants [MaxC|OptW|BP] an average of 64.3 epochs was required. The early stopping criterion was not satisfied during the maximum allowed 100 epochs of backpropagation training of the maximal network [MaxC|NoOptW|BP]. Therefore, the fully optimised networks can be trained much faster than unoptimised networks with conventional backpropagation, which is largely due to their greatly simplified internal complexity. However, the fact that even the partially (weights only) optimised variants [MaxC|OptW|BP] satisfied the early stopping criterion before the maximum allowed number of epochs shows that this methodology also reduces the required number of backpropagation epochs through weight optimisation. However, these reductions in backpropagation training epochs are achieved at a performance penalty, which is small but statistically significant, in the case of [MaxC|OptW|BP] and larger in the case of [OptC|OptW|BP].

Algorithms which simultaneously optimise ANN structures and weights such as this one and the constructive scheme reported by Sexton et al. (2004) are generally slow in execution compared to classical backpropagation as they require evaluation of large numbers of candidate solutions. In this case, processing time on a desktop computer for 200 iterations of the optimisation algorithm was approximately twelve times longer than for 100 epochs of backpropagation training of the maximal network. However, such algorithms remain valuable for their ability to reduce the number of input vectors, particularly if the processing time available for optimisation is not a constraint.

Conclusions

This study has demonstrated that a global optimisation methodology for ANN architecture and weights can be employed successfully to a river level prediction problem, but with a performance penalty relative to a fixed architecture trained using conventional backpropagation only. Optimisation of weights alone increases the ANN performance to a value close to that of the fixed architecture. The benefits of the combined optimisation of weights and connections are: a large reduction in the network complexity; a consequent reduction in the number of epochs of backpropagation training required; and the identification of the most parsimonious set of network inputs.

The methodology is more robust than trial and error, has the demonstrated ability to escape local minima, can eliminate unnecessary input vectors, and also offers the advantage that it is not application-specific, i.e. that it can be applied to any similar problem as a general approach. However, in common with other approaches to structural optimisation and input selection, there remain several parameters controlling the scheme which have to be prescribed prior to the optimisation.

In this study, current river levels at the flood location and the two available upstream monitoring stations were found to be the most useful predictor of levels 5 h ahead at the flood point. Using the prescribed cost function, other input vectors such as the recent change in river level at the flood location, sea level pressure or the change in sea level pressure did not improve prediction accuracy to the point where the extra network complexity added by their introduction was justified by an associated improvement in prediction performance.

The methodology offers further potential for automating the ANN design and optimisation process as it is also possible to incorporate neuron input delays and neuron biases within the parameter space given to the optimisation scheme prior to backpropagation training. The use of more sophisticated cost functions in the optimisation procedure may offer further improvements in predictive performance and greater control over the performance versus complexity trade-off. The methodology has also recently been modified to incorporate a genetic algorithm showing promising results on test problems (Zanchettin and Ludermir, 2007).

Acknowledgement

Funding for this work was provided by the Office of Public Works, Ireland and OTT Hydrometry (Ireland) Ltd. The authors acknowledge the assistance of Dr. Gordon Lightbody of UCC, Jack O'Keeffe and Charles McCarthy of Cork County Council, Peter Newport of the OPW and Micheál Mac Carthaigh of the Irish EPA/Met Éireann for providing sea level pressure data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhydrol.2008.03.017.

References

Birikundayvi, S., Labib, R., Trung, H.T., Rousselle, J., 2002. Performance of neural networks in daily streamflow forecasting. J. Hydrol. Eng. 7 (5), 392–398.

Bowden, G.J., Dandy, G.C., Maier, H.R., 2005. Input determination for neural network models in water resources applications. Part 1 – background and methodology. J. Hydrol. 301, 75–92.

Campolo, M., Andreussi, P., Soldati, A., 1999. River flood forecasting with a neural network model. Water Resour. Res. 35 (4), 1191–1197.

Corcoran, G.F., 2004. Development and examinations of flood warning systems for the Munster Blackwater at Mallow. University College Cork, M. Eng. Sci. Thesis, 146 pp.

Dawson, C.W., See, L.M., Abrahart, R.J., Heppenstall, A.J., 2006. Symbiotic adaptive neuro-evolution applied to rainfall–runoff modelling in northern England. Neural Networks 19, 236–247.

Elshorbagy, A., Simonovic, S.P., Panu, U.S., 2000. Performance evaluation of artificial neural networks for runoff prediction. J. Hydrol. Eng. 5 (4), 424–427.

Fahlman, S.E., Lebiere, C., 1990. The cascade-correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

Govindaraju, R., 2000a. ASCE task committee on application of artificial neural networks in hydrology, 2000a. Artificial neural networks in hydrology. II: hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.

Govindaraju, R., 2000b. ASCE task committee on application of artificial neural networks in hydrology, 2000b. Artificial neural networks in hydrology. I: preliminary concepts. J. Hydrol. Eng. 5 (2), 115–123.

Haykin, S., 1994. Neural networks, a comprehensive foundation. first ed.. Macmillan College Publishing Company, New York.

Henderson, F.M., 1969. Open channel flow, first ed. Macmillan Series in Civil Engineering Macmillan Publishing Company, New York, USA.

Hsu, K., Gupta, H.V., Sorooshian, S., 1997. Application of a recurrent neural network to rainfall-runoff modeling. In: Proceedings of the Conference on Aesthetics in the Constructed Environment. ASCE, New York, USA, pp. 68–73.

Ivanov, V.Y., Vivoni, E.R., Bras, R.L., Entekhabi, D., 2004. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. J. Hydrol. 298, 80–111. doi:10.1016/j.jhydrol.2004.03.041.

Jain, A., Prasad Indurthy, S.K.V., 2003. Comparative analysis of event-based rainfall-runoff modeling techniques – deterministic, statistical, and artificial neural networks. J. Hydrol. Eng. 8 (2), 93–98.

Jain, S.K., Chalisgaonkar, D., 2000. Setting up stage-discharge relations using ANN. J. Hydrol. Eng. 5 (4), 428–433.

Jain, S.K., Singh, V.P., Van Genuchten, M.T., 2004. Analysis of soil water retention data using artificial neural networks. J. Hydrol. Eng. 9 (5), 415–420.

Kashefipour, S.M., Lin, B., Falconer, R.A., 2005. Neural networks for predicting seawater bacterial levels. Proceedings of the Institution of Civil Engineers – Water Management 158 (3), 111–118.

Kiely, G., Albertson, J.D., Parlange, M.B., Katz, R.W., 1998. Conditioning stochastic properties of daily precipitation on indices of atmospheric circulation. Meteorol. Appl. 5, 75–87. doi:10.1017/S1350482798000656.

Kisi, Ö., 2004. River flow modeling using artificial neural networks. J. Hydrol. Eng. 9 (1), 60–63.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of ''goodness-of-fit'' measures in hydrologic and hydroclimatic model validation. Wat. Resour. Res. 35 (1), 233–241.

Ludermir, T.B., Yamazaki, A., Zanchettin, C., 2006. An optimization methodology for neural network weights and architectures. IEEE Transactions On Neural Networks 17 (6), 1452–1459.

McCulloch, W., Pitts, W.H., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.

Mitchell, T.M., 1997. Machine LearningMcGraw-Hill Series in Computer Science. McGraw-Hill Higher Education, New York, NY, USA.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models I, A discussion of principles. J. Hydrol. 10, 282–290.

Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2005. Short-term flood forecasting with a neurofuzzy model. Wat. Resour. Res. 41, W04004. doi:10.1029/2004WR003562.

Rumelhart, D., Hinton, D., Williams, G., 1986. Learning internal representations by error propagation. In: Rumelhart, D., McClelland, F. (Eds.), Parallel Distributed Processing 7. M.I.T. Press, Cambridge, MA, USA.

Senthil Kumar, A.R., Sudheer, K.P., Jain, S.K., Agarwal, P.K., 2004. Rainfall-runoff modelling using artificial neural networks: comparison of network types. Hydrol. Process. 19 (6), 1277–1291. doi:10.1002/hyp.5581.

Sexton, R.S., Dorsey, R.E., Sikander, N.A., 2004. Simultaneous optimization of neural network function and architecture algorithm. Decision Support Systems 36, 283–296.

Thirumalaiah, K., Deo, M.C., 2000. Hydrological forecasting using neural networks. J. Hydrol. Eng. 5 (2), 180–189.

Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. J. Hydrol. 239 (1–4), 132–147.

Weigend, A.S., Rumelhart, D.E., Huberman, B.A., 1991. Generalization by weight elimination with application to forecasting. In: Lippmann, R.P., Moody, J.E., Touretzky, D.S. (Eds.), Advances in Neural Information Processing Systems III. Morgan Kaufmann, San Mateo, USA, pp. 875–882.

Werbos, P., 1974. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. Thesis, Harvard University, Cambridge, MA, USA.

White, W.R., 2001. Water in rivers: flooding. Proceedings of the Institution of Civil Engineers – Water and Maritime Engineering 2 (2), 107–118.

Zanchettin, C., Ludermir, T., 2007. Comparison of the effectiveness of different cost functions in global optimization techniques. In: Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12–17.