Artificial neural networks and high and low flows in various climate regimes

DIONYSIA PANAGOULIA
Department of Water Resources & Environmental Engineering, School of Civil Engineering, National Technical University of Athens, 5 Heroon Polytechniou, 15780 Zographou, Athens, Greece
dpanag@hydro.civil.ntua.gr

Abstract An algorithm coupling linear least squares and simplex optimization (LLSSIM) is used to examine the ability of a three-layer feedforward artificial neural network (ANN) to simulate the high and low flows in various climate regimes over a mountainous catchment (the Mesochora catchment in central Greece). The plot of the long-term annual catchment pseudo-precipitation (rain plus snowmelt) simulated by the snow accumulation and ablation model (SAA) of the US National Weather Service (US NWS) showed trends of three climatically distinct periods, described by clearly descending, rising and moderately descending segments in pseudo-precipitation. The ANN model was calibrated for each of the three climate types and each was validated against the others. A set of statistical measures and graphs adapted for high and low flows showed the robustness of the ANN model under various climates and transient conditions. The ANN model proved capable of simulating well the daily high and low flows when it is calibrated for increasing pseudo-precipitation and validated for moderately decreasing pseudo-precipitation. For the entire period, the ANN model provided a better simulation of high and low flows than the conceptual soil moisture accounting (SMA) model of the US NWS, which was also employed in this study. Because the ANN is not a physically-based model, it is by no means a substitute for the SMA model. However, it is concluded that the ANN approach is an effective alternative for daily high- and low-flow simulation and forecasting in climatically varied regimes, particularly in cases where the internal dynamics of the catchment do not require an explicit representation.

Key words artificial neural network; conceptual modelling; high flows; linear least squares; low flows; simplex optimization

Réseaux de neurones artificiels et crues et étiages en régimes climatiques variés

Résumé Un algorithme couplé de moindres carrés linéaires et d'optimisation simplex (LLSSIM) a été utilisé afin d'examiner l'aptitude d'un réseau de neurones artificiel (RNA) sans rétroaction à trois niveaux à simuler les crues et les étiages selon les régimes climatiques variés d'un bassin versant montagneux (le bassin versant de Mesochora en Grèce centrale). Les graphiques de la pseudo-précipitation (pluie plus fonte nivale) à long terme, simulée par le modèle d'accumulation et d'ablation de la neige (SAA) du service national météorologique des Etats Unis (US NWS), ont révélé les tendances de trois périodes climatiques distinctes, correspondant à des segments de décroissance forte, croissance et décroissance modérée. Le modèle RNA a été calé pour chacun des trois types climatiques et validé par rapport aux deux autres. Un ensemble d'indices statistiques et de graphiques adaptés aux crues et aux étiages a montré la robustesse du modèle RNA selon différents climats ainsi qu'en conditions transitoires. Il est apparu que le modèle RNA est apte à bien simuler les crues et les étiages journaliers lorsqu'il est calé en période de pseudo-précipitation croissante et validé en période de pseudo-précipitation modérément décroissante. Pour la période entière, le modèle RNA a donné de meilleures simulations de crues et d'étiages que le modèle conceptuel de prise en compte de l'humidité du sol (SMA) de l'US NWS qui a également été utilisé dans cette étude. Le modèle RNA ne peut pas être substitué au SMA puisqu'il n'est pas à bases physiques. Il apparaît cependant que l'approche RNA est une alternative efficace pour la simulation et la prévision des crues et des étiages journaliers en régimes climatiques variés, en particulier lorsqu'aucune représentation explicite des dynamiques internes au bassin versant n'est requise.

Mots clefs réseau de neurones artificiel; modélisation conceptuelle; crues; moindres carrés linéaires; étiages; optimisation simplex

INTRODUCTION

The natural climatic variations including droughts, low river flows, and floods are of great relevance to water resources system design and operation. For some regions such extreme events have been responding to large-scale climatic forces (Trenberth *et al.*,

Open for discussion until 1 February 2007

1988; Nichols, 1989; Richey *et al.*, 1989; Eltahir, 1996), thus providing a physical explanation for the Hurst phenomenon (Hurst, 1951; Mandelbrot & Wallis, 1968; Klemes, 1974). Numerous other studies have established links between natural climate variability and hydrological variables on monthly and annual time scales (e.g. Lins, 1985; Redmond & Koch, 1991; Kahya & Dracup, 1993; Dettinger & Cayan, 1995; Lall & Mann, 1995). Other investigations have considered the issue of non-randomness in hydrological extremes (e.g. Wall & Englot, 1985; Booy & Lye, 1989; Lisi & Villi, 1997).

Since the assumption of randomness may be clouded by natural climate variability, for specific problems (e.g. predicting the impacts of climate change on hydrological extremes), the relationship between hydrological extremes and climate must be investigated. In this respect, two major questions have arisen: the first concerns the detection of climate variability in the existing records of extreme hydrological events and the possible record decomposition into high and low flows, and the second concerns the ability of models to capture such extreme values.

Although a serious obstacle to answering the first question is the shortness of the hydrological records, an effective means for testing whether there have been changes in hydrological extremes is to use the regional frequency analysis, or simply to search for trends in meteorological records. For example, Bradley (1998) used a regional frequency analysis approach to indicate the non-randomness (due to climatic variability) in annual maximum precipitation for 43 stations over 42-year period (1949–1990) in the Southern Plains of the USA. Bradley applied statistical tests, such as Kendall's correlation coefficient (Hirsch *et al.*, 1991, 1993), the *S* statistical test and the moving average test, to identify trends and non-randomness in annual quantile series.

In other related investigations, Mimikou *et al.* (1991) and Panagoulia (1992b) explored the plot of the long-term annual precipitation and pseudo-precipitation (rain plus melt) over the medium-sized mountainous catchment of Mesochora, in central Greece, for a 15-year period (1972–1986). The exploration of the long-term annual pseudo-precipitation by Panagoulia (1992b) revealed three segments with trends of varied climate conditions: the first (1972–1977) with a clearly decreasing pseudo-precipitation, average annual value of 1738 mm and lowest of 1252 mm (in the year 1977); the second (1978–1983) with an increasing pseudo-precipitation, average annual value of 2067 mm and highest value of 2460 mm (in the year 1979); and the third (1984–1986) with a moderately decreasing pseudo-precipitation, average annual value of 1912 mm, highest value of 2021 mm (in the year 1985), and lowest value of 1890 mm (in the year 1986). To gain more insight into these results, this study makes an intercomparison of the high and low flows of these climatically varied sub-periods for observed and modelled flows.

Considerable reported studies exist on the comparison of the two major approaches for modelling rainfall–runoff (R–R) processes, i.e. the conceptual (physical) modelling and the system-theoretic (or black-box) modelling (e.g. Gupta, 1984; Young & Wallis; 1985; Singh, 1988). The conceptual rainfall–runoff (CR–R) models usually incorporate simplified schemes of physical laws and are generally nonlinear, time-invariant and deterministic, with parameters that are representative of watershed characteristics. The most well known CR–R models, such as the Stanford IV model (Crawford & Linsley, 1980), the Sacramento soil moisture accounting (SMA) model of the US National Weather Service (Burnash *et al.*, 1973), the Tank model (Sugawara, 1974), the NAM (Nielsen & Hansen, 1973) and the ARNO model (Todini, 1988), have been compared to verify their ability to reproduce the measured flow rates, the easiness of calibration and estimation, the performance of several sub-processes (e.g. interflow, infiltration, evapotranspiration), as well as their physical interpretation (Franchini & Pacciani, 1991; Tingsanchali & Gautam, 2000). Although these models are valid in capturing the important features of watershed responses, such as the beginning of the rising limb of the hydrograph and the flow volume, they present some discrepancies related to their ability to reproduce extreme flows (especially high flows). In addition, the CR–R models are interpolative, i.e. when they are calibrated to a given set of hydrological signals (time series), there is no guarantee that the conceptual models can predict accurately when they are used to extrapolate beyond the range of calibration or verification experience (Gan & Burges, 1990; Panagoulia, 1992b).

In the system-theoretic approach, the models connect inputs and outputs without detailed consideration of the internal structure of the physical processes. The most common linear time series models are the ARMAX (auto-regressive moving average with exogenous inputs) developed by Box & Jenkins (1976). Although these models have provided satisfactory predictions in many fields (e.g. Bras & Rodriguez-Iturbe, 1985; Salas *et al*., 1980), their lack of nonlinearity in the transformation of rainfall to runoff often hampers their performance. This is why the coupling of deterministic models, such as the TANK and NAM, with the ARMA stochastic model (Box & Jenkins, 1976) applied, for example, to a river basin in Thailand (Tingsanchali & Gautam, 2000), improved significantly the flood forecasting capabilities of the models, especially the time-to-peak and rising limb of the hydrograph.

More recently, the use of artificial neural networks (ANNs) for nonlinear theoretical modelling has shown great potential. An ANN is capable of representing arbitrarily complex, nonlinear processes that relate inputs to outputs in a wide variety of fields (Vemuri & Rogers, 1994). The ANNs have proved to be an effective and efficient means to model R–R processes in the case where explicit knowledge of the internal hydrological processes is not required. The ANN modelling is widely reported in hydrological literature (Raman & Sunilkumar, 1995; Maier & Dandy, 1996; Loke *et al.*, 1997; Zhang & Stanley, 1997; Brion & Lingireddy, 1999; Abrahart & See, 2000; Tingsanchali & Gautam, 2000; Kim & Barros; 2001, Hu *et al.*, 2001; Wilby *et al.*, 2003; Cigizoglu, 2003; Campolo *et al.*, 2003; Tomasino *et al.*, 2004; Hu *et al.*, 2005; Giustolisi & Laucelli, 2005).

Although ANNs have already been shown to produce river flow predictions well compared to conventional models (Crespo & Mora, 1993; Karunanithi *et al.*, 1994; Hsu *et al.*, 1995; Abrahart & Kneale, 1997; Dawson & Wilby, 1998; Abrahart & See, 2000; Tingsanchali & Gautam, 2000), their ability to capture high and low flows is restricted to the research environment (Minns & Hall, 1996), and they often overestimate or underestimate high and low flows (Dawson & Wilby, 1998; Campolo *et al.*, 1999; Karunanithi *et al.*, 1994).

Reasons for this inability have been discussed in the literature. For example, Dawson & Wilby (1998), and Campolo *et al.* (1999) suggested that the under-estimation of peak flows could be attributed to a lack of information provided to the network, such as the antecedent rainfall. Karunanithi *et al.* (1994) suggested that the problem could be alleviated by including more high-flow patterns in the training data sets, while Hsu *et al.* (1995) proposed log-transformations of flow values to reduce the gap between the high- and low-flow conditions. Minns & Hall (1996, 1997) emphasized the need to exercise care when scaling the calibration data prior to ANN training. Scaling the calibration data within the range [0.1–0.9] (Hsu *et al.*, 1995; Campolo *et al.*, 1999) or [0.2–0.8] showed that the standardization using these reduced ranges bore little improvement (Minns & Hall, 1997).

Model generalization in itself is another important consideration in predicting high and low flows. A standard procedure is cross-validation, whereby another set of data is used to monitor the generality of the model during training. The study by Imrie *et al.* (2000), revealed that the generalization beyond the calibration range depends on the use of specific functions with respect to inherent hydrological properties of the catchment and calibration data. However, the aforementioned issues concern the modelling of the absolute extreme flows (high or low), rather than the identification of extreme events in climatically different rainfall–runoff (R–R) processes.

In this study, an attempt is made to explore the validity of ANNs to simulate high and low flows considered as above-mean and below-mean flows (Hsu *et al.*, 1995) in a climatically-varied environment beyond their calibration experience. The adopted approach is based on the philosophy of Klemes (1982, 1985) concerning the suitability of simulation models to predict the effects of climatic variability or change by using the dynamics of a differential split-sample test. Here, the various climates have been revealed, as reported previously, in the three segments of the long-term annual "rain plus melt" over the Mesochora catchment in central Greece. The "rain plus melt" has been produced from the implementation over the catchment of the snow accumulation and ablation model (SAA) of the US National Weather Service (US NWS).

The algorithm of three-layer feedforward ANNs coupled with the linear least squares and the nonlinear simplex (LLSSIM) optimization (Hsu *et al.*, 1995) was adopted here. The ANN model was calibrated (trained) for each of the three sub-periods and validated on the other two. The physically-based conceptual soil moisture accounting (SMA) model of the US NWS was also employed for comparison purposes, since the SMA has been tested extensively in the examined catchment (Panagoulia, 1992b; Panagoulia & Dimou, 1997a). The SMA was also used for impact assessment of climate variability and change on river flows at daily and sub-daily time steps in the study catchment and elsewhere (e.g. Nemec & Schaake, 1982; Lettenmaier & Gan, 1990; Panagoulia, 1991, 1992a; Panagoulia & Dimou, 1997a).

The choice of the specific ANN model used here is supported by the literature (e.g. Campolo *et al.*, 1999; Imrie *et al.*, 2000), since it is more effective and efficient than the widely used back-propagation algorithms. Furthermore, the model has been shown to capture the global or near-global solutions of a problem with fewer function evaluations. Moreover, the LLSSIM algorithm incorporates an automatic procedure for expansion of the network size, which can result in a rapid model development without user intervention.

In the remainder of the paper, a brief overview is given of the architecture of the three-layer feedforward ANN modelling and the LLSSIM algorithm for training the ANN model weights. Next, the ANN, SAA and SMA model identification is briefly presented. Subsequently, the ANN-LLSSIM methodology is examined to simulate the high and low flows under various climate conditions over the specified catchment. Adapted global statistical measures and graphical displays are used to quantify model performance and differences over the calibrated/validated sub-periods. The SMA model results are used for comparison purposes for the entire study period. Finally, the analysis of the results and the conclusions are presented.

NEURAL NETWORKS AND LLSSIM TRAINING ALGORITHM

Several ANN structures, e.g. feedforward networks, self-organizing feature maps, radial basis networks etc., have been proposed in the literature and their properties extensively studied. In this paper, a feedforward ANN is used (Rumelhart *et al.*, 1986) for reasons of improved performance. A typical three-layer feedforward ANN is shown in Fig. 1 (Hsu *et al.*, 1995). Each layer consists of nodes (processing elements) that are connected to other nodes of neighbouring layers. Thus, the first layer (input layer) is connected directly to the input variables $x_i(p)$ and to hidden layer nodes; the second layer (hidden layer) is connected to both the input and output layer nodes; while the third layer (output layer) is connected to hidden layer nodes and to the output variables $z_k(p)$.

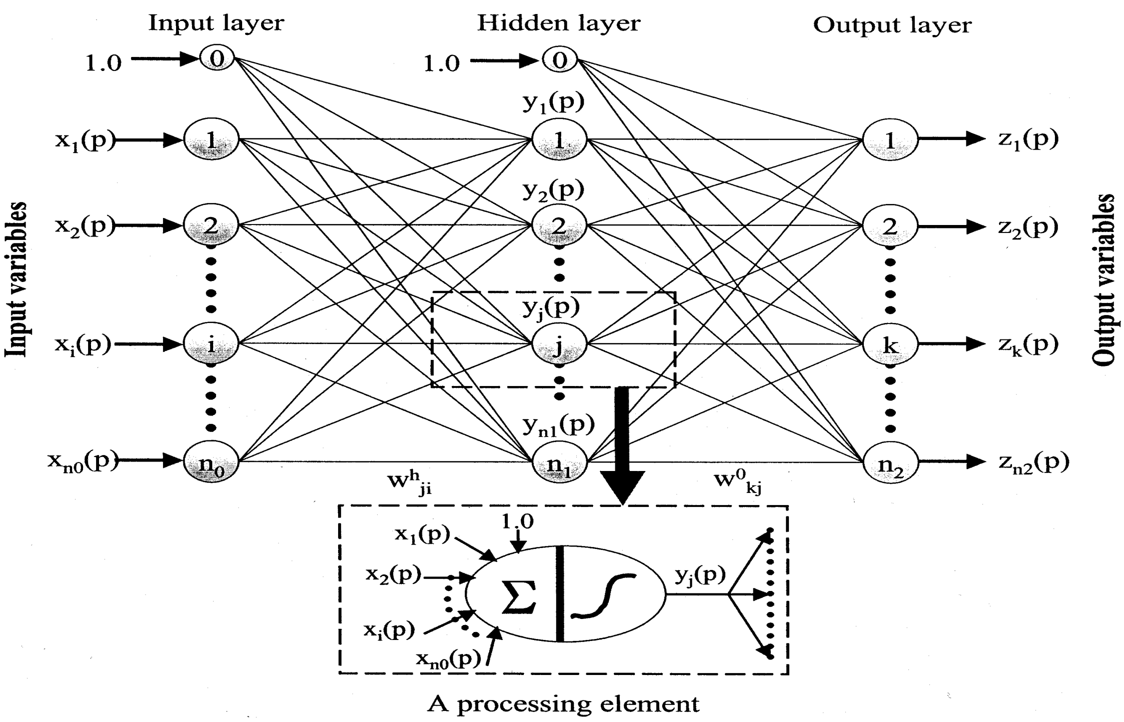


Fig. 1 Typical three-layer feedforward neural network.

A typical node j, is shown in Fig. 1. It receives incoming signals x_i from each node of the previous layer. Within node j, a linear combination of the incoming signals x_i is formed, with weights w_{ji}, to produce the effective incoming signal s_j:

$$s_j = \sum_{i=0}^{n_o} w_{ji} x_i \tag{1}$$

Still within node j, the signal s_j is passed through a nonlinear activation function (sigmoid function) to produce the output y_j of the node. The sigmoid function is a smooth, monotonically increasing and bounded function. In this study, the logistic function:

$$y_j = f(s_j) = \frac{1}{1+\exp(-s_j)} \tag{2}$$

is used as a sigmoid function. Thus the outputs $y_j(p)$ and $z_k(p)$ of hidden layer node j and output node k, respectively, can be expressed as follows:

$$y_j(p) = f\left[\sum_{i=0}^{n_0} w_{ji}^h x_i(p)\right] \tag{3}$$

$$z_k(p) = f\left[\sum_{j=0}^{n_h} w_{kj}^o y_i(p)\right] = f\left[\sum_{j=0}^{n_h} w_{kj}^o f\left(\sum_{i=0}^{n_0} w_{ji}^h x_i(p)\right)\right] \tag{4}$$

where $x_i(p)$ is the input of input node i, $w^h{}_{ji}$ is the connection weight from input node i to hidden node j, $w^o{}_{ki}$ is the connection weight from hidden node j to output node k, n_0 is the number of input nodes and n_h the number of hidden layer nodes.

ANNs of this type can be trained to approximate a given input–output relationship. The objective of network training is to choose optimal connection weights $w^h{}_{ji}$ and $w^o{}_{ki}$ so that network predicted outputs $z_k(p)$ are a best fit to the measured output values, say $t_k(p)$. In the present case, where the rainfall–runoff relationship is to be approximated, a single output value $t(p)$ (runoff) is measured; hence, the approximating ANNs must have a single output node with output value $z(p)$ given by equation (4) with the index k dropped. In order to train the network, rainfall–runoff measurements are organized in m sets of input–output patterns (indexed by p), each having n_0 inputs $[x_1(p), x_2(p), \ldots, x_{n0}(p)]^T$ and a single target value (runoff), $t(p)$. Training the ANN is equivalent to minimizing the following error function:

$$\begin{aligned} F(w^h, w^o) &= \frac{1}{2}\sum_{p=1}^{m}(t(p)-z(p))^2 = \frac{1}{2}\sum_{p=1}^{m}\left[t(p) - f(\sum_{j=0}^{n_h} w_j^o y_i(p))\right]^2 \\ &= \frac{1}{2}\sum_{p=1}^{m}\left[t(p) - f\left(\sum_{j=0}^{n_h} w_j^o f(\sum_{i=0}^{n_0} w_{ji}^h x_i(p))\right)\right]^2 \end{aligned} \tag{5}$$

with respect to the weights w_{ji}^h and w_j^o.

As regards the LLSSIM network algorithm (Hsu *et al.*, 1995) that was adopted here, it partitions the weight space into two groups, namely input-to-hidden layer weights $w^h{}_{ji}$ and hidden-to-output layer weights $w^o{}_j$, employing a different training strategy for each group. Given a set of values for the input-to-hidden layer weights $w^h{}_{ji}$, the optimal values for hidden-to-output weights are determined explicitly by using linear least squares (Scalero & Tepedelenlioglu, 1992). With respect to the input-to-hidden layer weights $w^h{}_{ji}$, the optimization is performed using the nonlinear simplex algorithm of Nelder & Mead (1965). Thus the nonlinear part of the search is performed in a reduced dimensional space (input-to-hidden layer weights alone), resulting in acceleration of the training process. Improved global search characteristics are also exhibited by the nonlinear simplex method, due to its use of multiple points at each iteration and its ability to overcome minor local minima.

A brief description of the LLSSIM method follows. Assume that a set of values for the input-to-hidden layer weights w^h_{ji} is given. Then the hidden node outputs $y_j(p)$ can be calculated from equation (3). Now $F(w^h,w^o)$ is a function of hidden-to-output weights w^o_j alone; however, the optimization problem to be solved is still a problem of nonlinear least squares. In order to obtain explicit optimal values for hidden-to-output weights, the LLSSIM method defines an approximately equivalent problem of linear least squares. Specifically: (a) the target values $t(p)$ are transformed backward through the logistic function of the output node, i.e. $\tilde{t}(p) = \ln\left(\frac{t(p)}{1-t(p)}\right)$ are obtained from $t(p)$, and (b) a new error function $F_1(w^o)$ is defined in terms of the transformed target values $\tilde{t}(p)$:

$$F_1(w^o) = \frac{1}{2}\sum_{p=1}^{m}\left[\tilde{t}(p) - \sum_{j=0}^{n_h} w^o_j y_j(p)\right]^2 \tag{6}$$

Now, minimization of $F_1(w^o)$ with respect to hidden-to-output weights w^o_j is a problem of linear least squares whose optimal solution $\hat{w}^o = [\hat{w}^o_1, \hat{w}^o_2, \cdots, \hat{w}^o_{n_h}]^T$ is explicitly obtained by solving the following set of linear equations:

$$\sum_{p=1}^{m} y_i(p) \sum_{j=0}^{n_h} y_j(p)\hat{w}^o_j = \sum_{p=1}^{m}\tilde{t}(p)y_i(p) \qquad i = 1,\cdots,n_h \tag{7}$$

Of course, both the hidden node outputs $y_j(p)$ and the solution $\hat{w}^o$ of equation (7) depend on the input-to-hidden layer weights w^h; therefore, one writes $\hat{w}^o(w^h)$ for the optimal solution of equation (6).

The nonlinear part of the LLSSIM method can now be derived by replacing the optimal hidden-to-output weights $\hat{w}^o(w^h)$, obtained from equation (7), into equation (5):

$$\tilde{F}(w^h) = F(w^h, \hat{w}^o(w^h)) = \frac{1}{2}\sum_{p=1}^{m}\left[t(p) - f\left(\sum_{j=0}^{n_h}\hat{w}^o_j(w^h) f(\sum_{i=0}^{n_0} w^h_{ji}x_i(p))\right)\right]^2 \tag{8}$$

The error function $\tilde{F}(w^h)$ now depends only on input-to-hidden layer weights w^h and the resulting nonlinear least squares problem is solved by the Nelder & Mead (1965) in this reduced dimensional space.

RUNOFF MODELLING IN CLIMATICALLY VARIED REGIMES

The Mesochora catchment, drained by the Acheloos River (Fig. 2), was selected for performance exploration of a three-layer feedforward ANN model under climatically varied regimes. As discussed, the SMA model of the US NWS is also considered in this study for comparison purposes. The basic criterion for the catchment selection was its geographical and hydrological significance due to the partial diversion of the river for irrigation and hydropower purposes. The Mesochora catchment, with an area of 633 km^2, lies in the central mountain region of Greece and extends nearly 32 km from

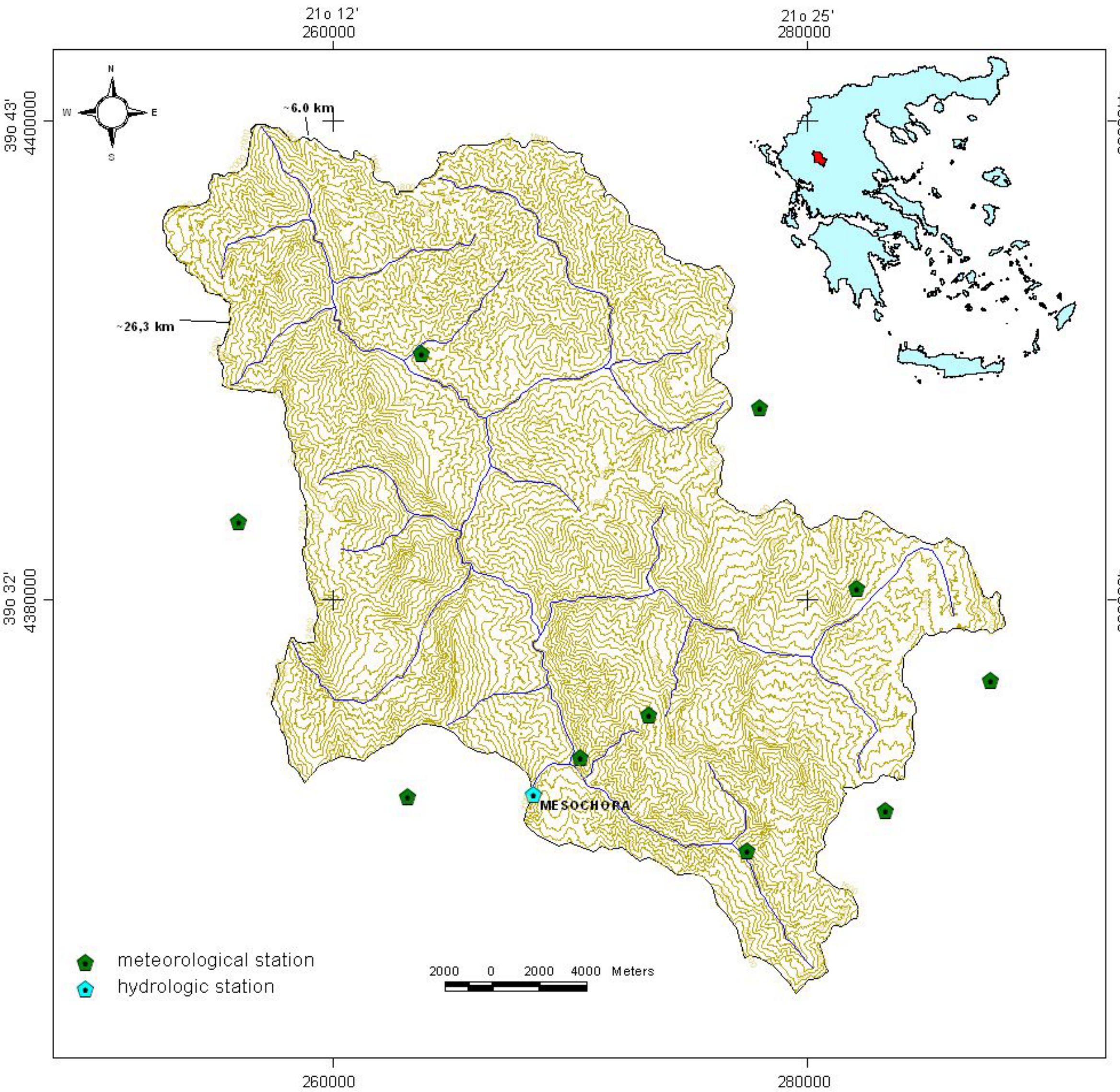
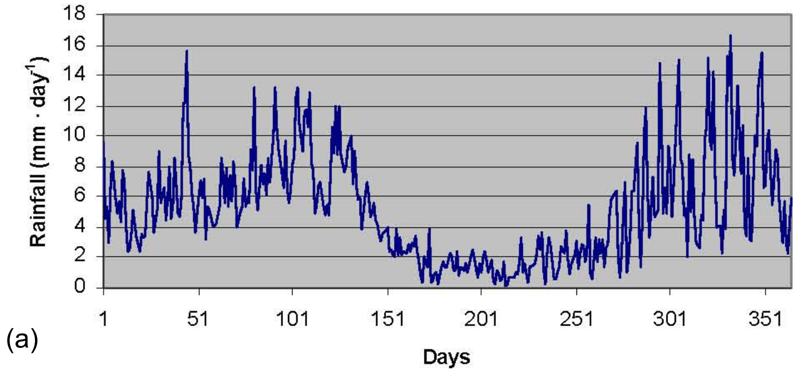


Fig. 2 The Mesochora catchment area and hydrometeorological stations.

north (39°42′) to south (39°25′) with an average width of about 20 km. The climate in the Mesochora catchment is elevation-dependent, with hot summers and mild winters at low elevations and mild summers and cold winters at high elevations. Due to the high mean elevation (1390 m a.s.l.), the catchment hydrology is controlled by snowfall and snowmelt. The catchment mean annual precipitation (weighted average over elevation bands) is about 1900 mm and the mean annual runoff is 1170 mm (or 23.5 $m^3 s^{-1}$). The annual cycle of rainfall (rain plus melt) and observed runoff of the catchment for the 15-year study period is shown in Fig. 3, reflecting a seasonal hydrological cycle with low discharge values during the summer and much higher values during the winter.

Fifteen consecutive years of daily rainfall (rain plus melt generated from SAA model) and runoff data for the Mesochora catchment were selected for model development and testing. As a first step, the first five years of data (1972–1976) were used for model calibration (training), while the remaining ten years (1977–1986) were used for



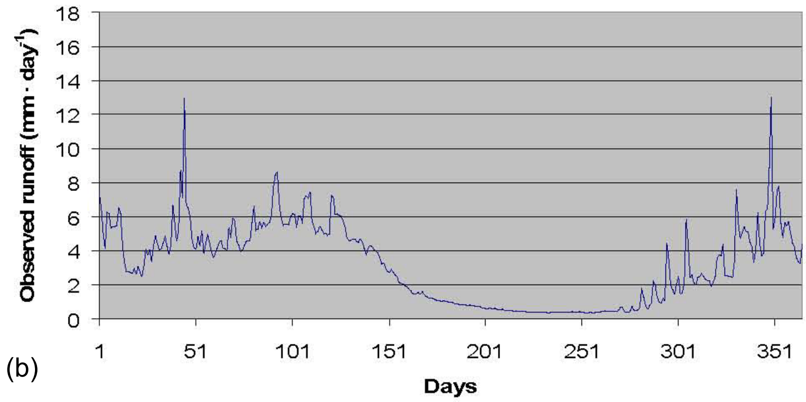


Fig. 3 Mesochora catchment long-term annual cycles in: (a) pseudo-precipitation and (b) runoff.

ANN model validation. The whole period (1972–1986) was used for SMA model calibration. This process was used to select the best ANN model with respect to the number of nodes in the input, output and hidden layers. Training and calibration periods were long enough to extract representative results of overall catchment behaviour, compared to the study by Hsu *et al.* (1995), in which the identification period was only one wet year. The selected ANN model was tested for its ability to perform under different climate conditions. The plot of the long-term annual pseudo-precipitation (rain plus melt) over the catchment (Fig. 4) reflects three segments with varying climate trends, as discussed in the introduction. The ANN model was calibrated for each of the three segments and validated on the other two. In reality, the first segment includes the major portion of data used for ANN model determination (identification); thus, with a small tolerance, the period for ANN identification and that for model calibration on the clearly descending segment of pseudo-precipitation should be assumed identical. In this way, the calibrated ANN model was examined for climate transferability according to the relative principles of Klemes (1985). For completeness of the study, both the ANN and SMA models were compared for the same time horizon (15-year period).

A flood or low-flow forecasting model can offer detailed knowledge when it operates in a sub-daily time step, such as 4, 12, or 18 hours with current (t = 0 h) and antecedent (t = –1, –2, …–n h) conditions and/or when it includes meteorological multi-step forecasting (e.g. Imrie *et al*., 2000; Tingsanchali & Gautam, 2000; Abrahart & See,

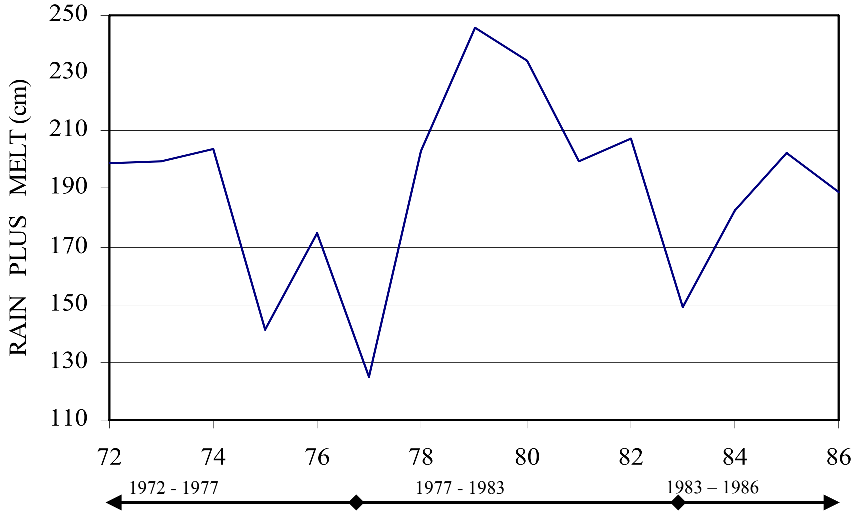


Fig. 4 Long-term annual pseudo-precipitation (rain plus melt) of the Mesochora catchment.

2000; Kim & Barros, 2001). Floods and low flows (occurrences, duration, magnitude, threat scores, etc.) are examined after a threshold is defined (e.g. Panagoulia & Dimou, 1997b; Kim & Barros, 2001). In this study, the daily time step in the modelled time series was kept for exploration of high and low flows at a preliminary stage in investigation of model transferability for high/low flows to varied climate conditions. It should be noted that, despite the evidence of inter-annual variability in runoff (as seen in Fig. 3), the epoch data were not pre-classified into low or high values for ANN model calibration or validation. This approach has been adopted by other authors (e.g. See *et al.*, 1997) and is a topic requiring further research for different climates.

Regarding the evaluation of neural network solutions, the use of "goodness-of-fit" statistics can give no real indication of what the network is getting right or wrong. The neural network solutions are designed to minimize global measurements (Abrahart & See, 2000). Furthermore, global evaluation statistics do not provide specific information of model performance at high and low flow levels. Thus, four global statistical measures were adapted for flows above and below the mean daily flow along with graphical displays. These are: (a) the root mean squared error (*RMSE*), defined as the sum of the squares of the differences of the forecasts and observations; (b) percentage volume in errors (bias) (%*VE*), defined as the differences between observed and simulated hydrographs; (c) the percentage error in peak (%*MF*), defined as the difference between forecasts and observations; and (d) the correlation coefficient (*CORR*), which describes the strength of the linear relationship between forecasts and observations.

Finally, bearing in mind that no comparison would be complete without a plot of observed and forecast hydrograph (Green & Stephenson, 1986), the hydrographs of observed *vs* modelled flows and their associated scatter plots were examined.

ANN MODEL IDENTIFICATION

A first step in building an ANN model for the rainfall–runoff relationship is data normalization. Since sigmoid functions are used in ANN models, the hydrological data should be normalized to the range [0,1]. Indeed, in this study, the rainfall data were normalized to the range [0,1] before applying the LLSSIM algorithm. However, following Hsu *et al.* (1995), the runoff data series was normalized to the range [0.1, 0.9] in order to avoid the problem of output signal saturation, sometimes encountered in ANN applications (Smith, 1993).

Next, a nonlinear discrete-time dynamical model is assumed for the rainfall–runoff process. Specifically (see also Hsu *et al.*, 1995), it is assumed that, at time t, the ANN output $z(t)$ is related to past rainfall measurements $r(t-j)$, $j = 1, \ldots, n_a$ and to past runoff measurements $\tilde{t}(t-j), j = 1, \ldots, n_b$, i.e.:

$$\tilde{t}(t) = g\big(r(t-1), r(t-2), \cdots, r(t-n_a), \tilde{t}(t-1), \tilde{t}(t-2), \cdots, \tilde{t}(t-n_b)\big) + e(t) \tag{9}$$

where the unknown nonlinear mapping $g(\cdot)$ is to be approximated by the ANN model by minimizing the mapping error $e(t)$. Here n_a and n_b are the numbers of past input and output measurements, respectively, contributing to the current output. Thus at time t the input vector $\boldsymbol{x}(p) = [r(t-1), r(t-2), \cdots, r(t-n_a), \tilde{t}(t-1), \tilde{t}(t-2), \cdots, \tilde{t}(t-n_b)]^T$ is presented to the input layer nodes of the ANN. This input vector together with the current output (runoff) measurement $\tilde{t}(p) = \tilde{t}(t)$ constitutes the pth input–output training pattern for the ANN model.

The notation ANN(n_a, n_b, n_h) is used henceforth to denote the above model structure, where $n_a + n_b = n_0$ is the number of nodes in the input layer, n_h is the number of nodes in the hidden layer, and the output layer consists of a single node. Thus an ANN model is identified by selecting values for n_a, n_b and n_h, and by using the LLSSIM algorithm in order to estimate network weights w^h_{ji} and w^o_j which minimize the prediction error function $F(w^h, w^o)$ of equation (5).

In the current study, several combinations of n_a, n_b and n_h were examined in order to make a preliminary assessment of the suitability of ANN models to represent the rainfall–runoff process in mountainous catchments. Selected results of this preliminary exploration are given in Table 1; these results are fully discussed in Panagoulia & Maratos (2003a,b). The same table gives a summary of the results for three representative combinations (n_a, n_b, n_h) listed as ANN models A, B and C, followed by a parenthesis indicating the (n_a, n_b, n_h) combination. From the statistical performance of the ANN models calibrated for a five-year period and validated for the remaining ten-year period, respectively, as reflected in Table 1, and from the corresponding graphs presented in Panagoulia & Maratos (2003a,b) it is deduced that the ANNC(3,5,4) model performs best. The performance of this model, subsequently denoted as ANNI with $I = 0, 1, 2, 3$ denoting the calibration/validation period, is further investigated in order to assess the model ability to represent high and low flows in varied climate conditions.

Table 1 Calibration and validation statistics for the five-year calibration study.

No.	Model	*RMSE*		*%VE*		*%MF*		*CORR*	
		Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
1	ANNA(2,3,3)	16.785	18.170	–2.267	–2.019	–67.300	–71.100	0.8165	0.8411
2	ANNB(5,4,3)	13.190	16.277	–1.410	–1.790	–23.800	–53.700	0.8895	0.8730
3	ANNC(3,5,4)	12.440	16.185	–1.230	–1.480	–3.200	–57.210	0.9100	0.8750

RMSE: root mean square error; %*VE*: percentage volume error; %*MF*: percentage error of maximum flow; and *CORR*: correlation statistic.

SAA AND SMA MODEL IDENTIFICATION

The SAA model was developed by Anderson within the US National Weather Service Hydrologic Research Laboratory (Anderson, 1973) and has been tested in a number of mountainous catchments in the western USA, Mediterranean countries and elsewhere. This is a deterministic, continuous conceptual model consisting of a set of mathematical formulations, which explicitly describe the change in storage of water and heat in the snowpack, based on data for precipitation and temperature at 6-hourly intervals.

It has been widely used for snowmelt purposes (e.g. Anderson, 1973; Lettenmaier & Gan, 1990; Panagoulia, 1992a,b; Georgakakos & Bae, 1994). In this study, the SAA model is applied over three elevation bands and is calibrated concurrently with the SMA rainfall–runoff model, which accepts as input the "rain plus melt water" provided from the SAA model.

The SMA model was developed in the US National Weather Service (US NWS) Sacramento, California River Forecast Center by Burnash *et al.* (1973) and forms the basis of the US NWS catchment hydrological response model for operational forecasting. At first it was used for the Sacramento basin simulation, and since then it has been widely used (e.g. Burnash *et al.*, 1973; Peck, 1976; Kitanidis & Bras, 1980a,b; Lettenmaier & Gan, 1990; Panagoulia, 1992a,b). This is a deterministic, lumped parameter, conceptual model, which explicitly accounts for the flux of soil moisture between five storage zones. Transfer of water between the soil moisture zones controls the runoff response. Direct runoff from impervious areas and water surfaces, surface runoff, interflow from the upper zone free water and the primary and supplemental baseflows from the lower zone generate streamflow. The inputs to the SMA model are the "rain plus melt" provided by the SAA model and the potential evapotranspiration. In this study, the "rain plus melt" was the average quantity weighted over three elevation bands of the catchment. The streamflow generated from the SMA model was included in the study. The coupled models were calibrated manually by using the 15-year period that has a sufficient length for calibration of conceptual models under plentiful data conditions.

RESULTS AND DISCUSSION

In this section, the performance of the ANN (3,5,4) model is further investigated for high and low flows in periods with trends of various climate conditions. This network model, henceforth denoted as ANN0, ANN1, ANN2 and ANN3, relates $n_a = 3$ past rainfall measurements and $n_b = 5$ past flow measurements to the current output that has $n_h = 4$ hidden nodes. The ANN0 model is calibrated for the entire 15-year period. The ANN1, ANN2 and ANN3 models are calibrated and validated for climatically varied conditions, namely, the clearly descending period from 1972 to 1977, the rising period from 1978 to 1983, and the moderately descending period from 1984 to 1986, as seen in Fig. 4. The ANN1 model is calibrated for the first six years (1972–1977) and is validated over: (a) the six following years (1978–1983), denoted as validation period A in Table 2 and graphs; and (b) the last three years (1984–1986), denoted as validation period B in Table 2 and graphs. The ANN2 model is calibrated for the last three years (1984–1986) and is validated over: (a) the six first years (1972–1977), denoted as vali-

Table 2 Calibration and validation statistics for above- and below-mean flows under climatically varied periods (clearly descending, rising, and moderately descending).

Model	Period	Above-mean flow: $RMSE_A$	$\%VE_A$	$\%VF_A$	$CORR_A$	Below-mean flow: $RMSE_B$	$\%VE_B$	$\%VF_B$	$CORR_B$
Hydrological	Calibration: 1–5479 days	37.3100	–4.4200	24.3822	0.6853	8.2300	21.1200	0.1877	0.7028
ANN0	Calibration: 1–5479 days	23.5211	–3.7468	–41.9630	0.8493	3.8388	6.6412	40.4652	0.8737
ANN1	Calibration: 1–2192 days	14.2780	–2.9530	–4.3890	0.8953	2.1975	21.9310	–60.7480	0.6222
	Validation A: 2193–4383 days	22.2960	–3.4480	–38.4130	0.8515	6.4690	28.5970	–17.6700	0.3689
	Validation B: 4384–5479 days	15.2970	–2.7000	–25.6230	0.9087	1.8220	39.4960	14.9070	0.4243
ANN2	Calibration: 4384–5479 days	10.7090	–2.4790	0.3210	0.9562	2.1600	35.9020	–9.8600	0.2401
	Validation A: 1–2192 days	20.4220	–4.5870	–41.3280	0.7714	4.2710	28.8030	1.4790	0.3373
	Validation B: 2193–4383 days	24.7290	–2.9310	–32.3650	0.8138	7.1826	17.5190	1.5390	0.3049
ANN3	Calibration: 2193–4383 days	17.3910	–2.2500	–25.9400	0.9124	3.5750	15.8330	4.2040	0.5780
	Validation A: 1–2192 days	20.9420	–3.7672	–72.2680	0.7568	2.5440	17.5300	–5.5590	0.5746
	Validation B: 4384–5479 days	18.1410	–2.5810	–64.8800	0.8678	2.1110	25.1560	–0.4560	0.3635

dation period A in Table 2 and graphs; and (b) the six following years (1978–1983), denoted as validation period B in Table 2 and graphs. The ANN3 model is calibrated for the six middle years (1978–1983), and is validated over: (a) the six first years (1972–1977), denoted as validation period A in Table 2 and graphs; and (b) the last three years (1984–1986), denoted as validation period B in Table 2 and graphs. The SMA model (hydrological model) is calibrated over the entire 15-year period. The statistics (*RMSE*, %*VE*, %*MF*, *CORR*) computed separately for flows above and below the 15-year mean (23.41 $m^3 s^{-1}$) over all the calibration/validation periods of the ANN*I* models are summarized in Table 2. The analysis of results is presented below beginning with the statistical values of Table 2 and graphs, and continued with hydrographs and scatter plots.

The $RMSE_A$ and $RMSE_B$ statistic measures the residual variance for above-mean and below-mean flows, respectively; the optimal value is 0.0. The ANN0 model has the smallest $RMSE_A$ and $RMSE_B$ against those of the hydrological model, bearing in mind that both models are calibrated for the overall 15-year period. In comparison with the subsequently examined ANN1, ANN2 and ANN3 models calibrated for different periods (smaller in length), the ANN0 model $RMSE_A$ and $RMSE_B$ performance is the worst. This may be a significant finding as it may be related to the ability of "nonlinear" model structure to handle better the high/low flows in classified climate types. The ANN2 and the ANN1 models have the smallest *RMSE* ($RMSE_A$ and $RMSE_B$) during calibration and validation periods respectively, which are the last three years (1984–1986) climatically expressing the moderately descending segment of "rain plus melt" in Fig. 4. However, the ANN2 model validation $RMSE_A$ and $RMSE_B$ performance over the climatically rising segment (second period of six years in Fig. 4) is the worst of the three models. On average, between calibration and validation periods, the ANN1 model performs best when it is calibrated for clearly decreasing pseudo-precipitation and validated for moderately decreasing pseudo-precipitation as measured by $RMSE_A$ and $RMSE_B$ statistics.

The $\%VE_A$ and $\%VE_B$ statistic calculates the percentage volume error (bias) for above-mean and below-mean flows, respectively, under the observed and simulated hydrographs, summed over the data period. The value 0.0 is best; positive values indicate overestimation, and negative values underestimation. The ANN0 model performance is superior to that of SMA for above-mean and below-mean flows, but it is worst for smaller calibration periods for above-mean flows. For below-mean flows, the SMA performance is better for the ANN1 and ANN2 calibration periods. The comparison among ANN1, ANN2, and ANN3 models showed that, during calibration, the ANN3 model performs best for the rising segment of "rain plus melt" over above- and below-mean flows, while, during validation, the ANN3 model performs best for the moderately descending segment over above-mean flows and for the clearly descending segment under below-mean flows correspondingly. The worst performance for the ANN1 model (overestimation) occurs when it is validated for the moderately descending segment under below-mean flows. The best performance for the examined climates is that of ANN3.

The $\%MF_A$ and $\%MF_B$ statistic calculates the percentage error in matching the maximum (peak) flow for above-mean and below-mean flows, respectively, of the data record. The optimal value is 0.0; positive values indicate overestimation, and negative values underestimation. Again, the ANN0 performance is worse than the other ANN models with smaller calibration data periods for above-mean flows. For below-mean flows, the performance continues to be worse except for the case of ANN1 calibration. The comparison among smaller periods showed that, during calibration, the ANN2 and ANN3 models match the peak flow very well for the moderately descending segment over above-mean flows and the rising segment under below-mean flows, respectively. During validation, $\%MF_A$ and $\%MF_B$ present the smallest deterioration for the ANN1 and ANN3 models, respectively, for the moderately descending segment. On average, between calibration and validation periods, the ANN3 model performs best when it is calibrated to increasing pseudo-precipitation and validated to moderately decreasing pseudo-precipitation as measured by $\%MF_A$ and $\%MF_B$ statistics.

The correlation (*CORR*) statistic calculates the linear correlation between the observed and simulated flows for above-mean ($CORR_A$) and below-mean ($CORR_B$) flows, respectively, with an optimal value of 1.0. The ANN0 *CORR* value is slightly worse than that of the other models under calibration for above-mean flows, but this model performance is steadily better compared to that of smaller calibration periods. The *CORR* value is expectedly worse (smaller) during validation than during calibration for all of the models over above- and below-mean flows. During calibration, the best value of *CORR* is given by the ANN2 model for the moderately descending segment over above-mean flows, and correspondingly by the ANN1 and ANN3 models for the clearly descending and rising segments under the below-mean flows. During validation, the ANN1 model correlates best the observed and simulated flows for the moderately descending segment over the above-mean flows, while the ANN3 model correlates less well the observed and simulated flows for the clearly descending segment under the below-mean flows. On average, it may be concluded that the ANN3 model performs best when it is calibrated for increasing pseudo-precipitation and validated for clearly decreasing pseudo-precipitation as measured by the correlation statistic.

To examine these results in more detail, Figs 5 and 6 present the $RMSE_A$ and $RMSE_B$ statistics for each model (SMA, ANN0, ANN1, ANN2, and ANN3), computed separately for each of the 15 years and presented as a function of the mean flow for the given year. Subsequently, Figs 7 and 8 present the $\%VE_A$ and $\%VE_B$ statistics for each model, also computed separately for each of the 15 years and presented as a function of the mean flow for the given year.

The $RMSE_A$ statistic is presented in Fig. 5(a)–(d). In all four cases, the lower mean high-flow years appear to have the tendency for lower model $RMSE_A$. In Fig. 5(a) the $RMSE_A$ is presented for the SMA and ANN0 models, both of which are calibrated for the entire 15-year period, including a variety of values in high flows (more or less high). Clearly, the $RMSE_A$ performance of the SMA model is worse than that of the ANN0 model for all years. This may be attributed to the inability of a "less nonlinear" SMA model structure to handle more variable conditions. With this in mind, and trying to decrease the large number of computations needed for this study, separation of the SMA model performance into calibration and validation periods has not been carried out as it was for the ANN models. However, a comparison of SMA and calibrated/validated ANN model performance is performed within this work. It should be noted that in Fig 5(b), (c) and (d), the ANN*I* models present a nearly linear relationship of their ability to match the calibration and validation data (residual variance computed by $RMSE_A$) with the wetness of the year. This tendency is smaller for the ANN0 model calibration data, as shown in Fig. 5(a)–(d). The ANN3 and ANN1 models present the smallest $RMSE_A$ on calibration and validation period A, respectively. On average, with a small difference between ANN1 and ANN3, the best fitting of linear correlation and smallest $RMSE_A$ values is provided by the ANN1 model, which is calibrated on the clearly descending flows and validated on the increasing flows. Note that the calibration/validation period for the ANN2 model is determined by three years, thus increasing the uncertainty, especially when discussing issues of linearity.

Figure 6(a)–(d) shows the $RMSE_B$ statistic. In all four cases, the lower mean low-flow years do not appear to have a clear tendency of lower model $RMSE_B$ as is the case for $RMSE_A$ for mean high flows. For below-mean flows, the $RMSE_B$ performance of the SMA model (Fig. 6(a)) is worse than that of the ANN0 model for all years, while there is no significant difference in performance among ANN0, ANN1 and ANN3 calibrated models. In Fig. 6(b), (c) and (d), the ANN*I* models keep the tendency of a linear relationship between calibration and validation data (residual variance computed by $RMSE_B$) with the wetness of the year in an unclear way. However, the ANN1 and ANN3 model performance is best, as noticed for the $RMSE_A$ case for the calibration/validation periods.

The $\%VE_A$ statistic is shown in Fig. 7. The results indicate that the SMA model has a poor performance as compared to the ANN*I* models, as does the ANN0 when compared to the ANN1 and ANN3 calibrated models. The ANN3 and ANN1 models present the smallest bias during calibration (rising flows) and validation period A (rising flows), correspondingly. In Fig. 7(b), (c) and (d), the ANN*I* models do not present any clear linear relationship between the ability to match the calibration and validation data (residual variance computed by $\%VE_A$) with the wetness of the year.

Figure 8 shows the $\%VE_B$ statistic. The results indicate that the ANN*I* models consistently have smaller bias as compared to the SMA model. The ANN0 model presents equally small bias as compared to the ANN1 and ANN3 calibrated models with a slight

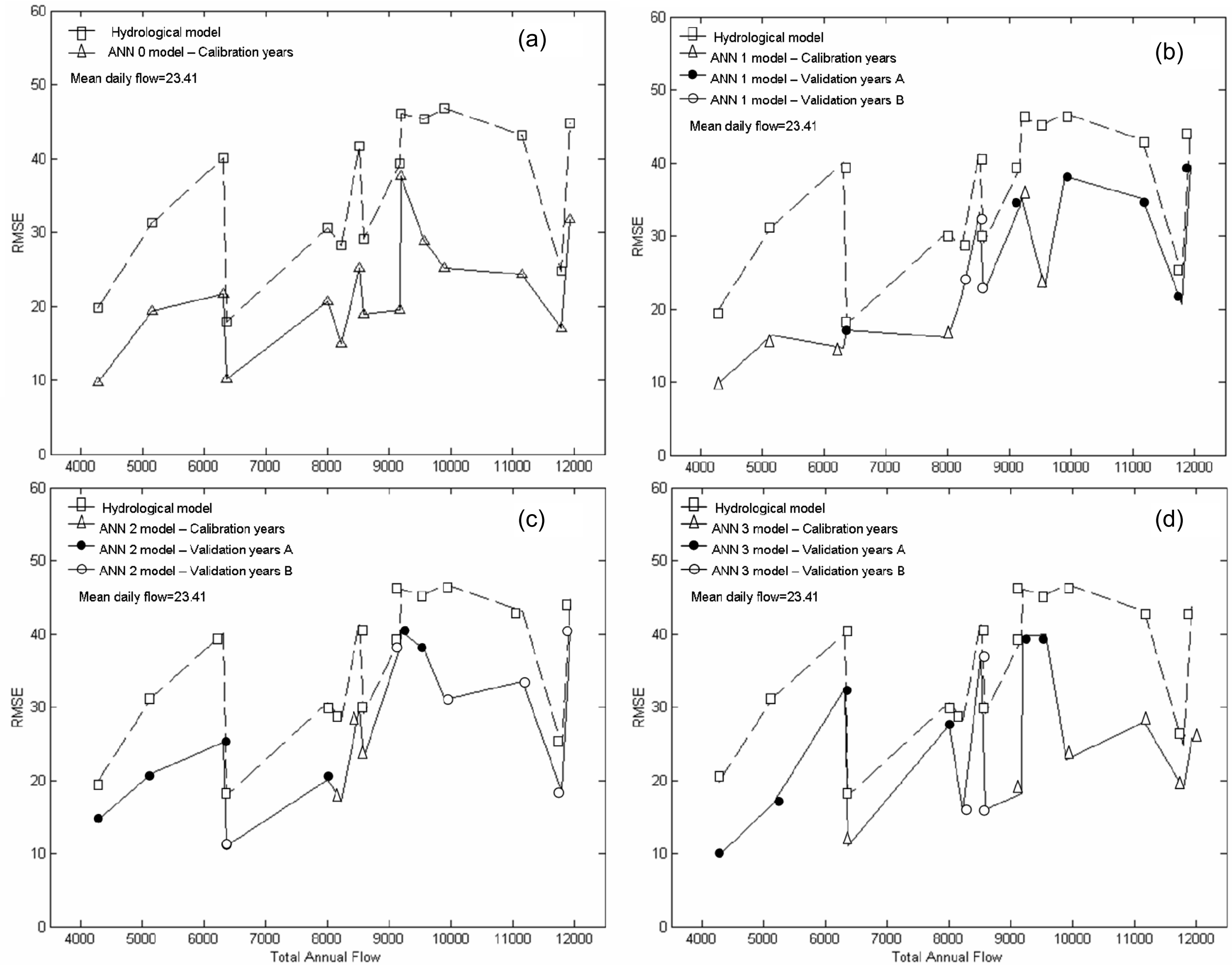


Fig. 5 Annual performance of *RMSE* statistics for above mean flow for each model (SMA, ANN0, ANN1, ANN2 and ANN3) plotted against total flow for each data year.

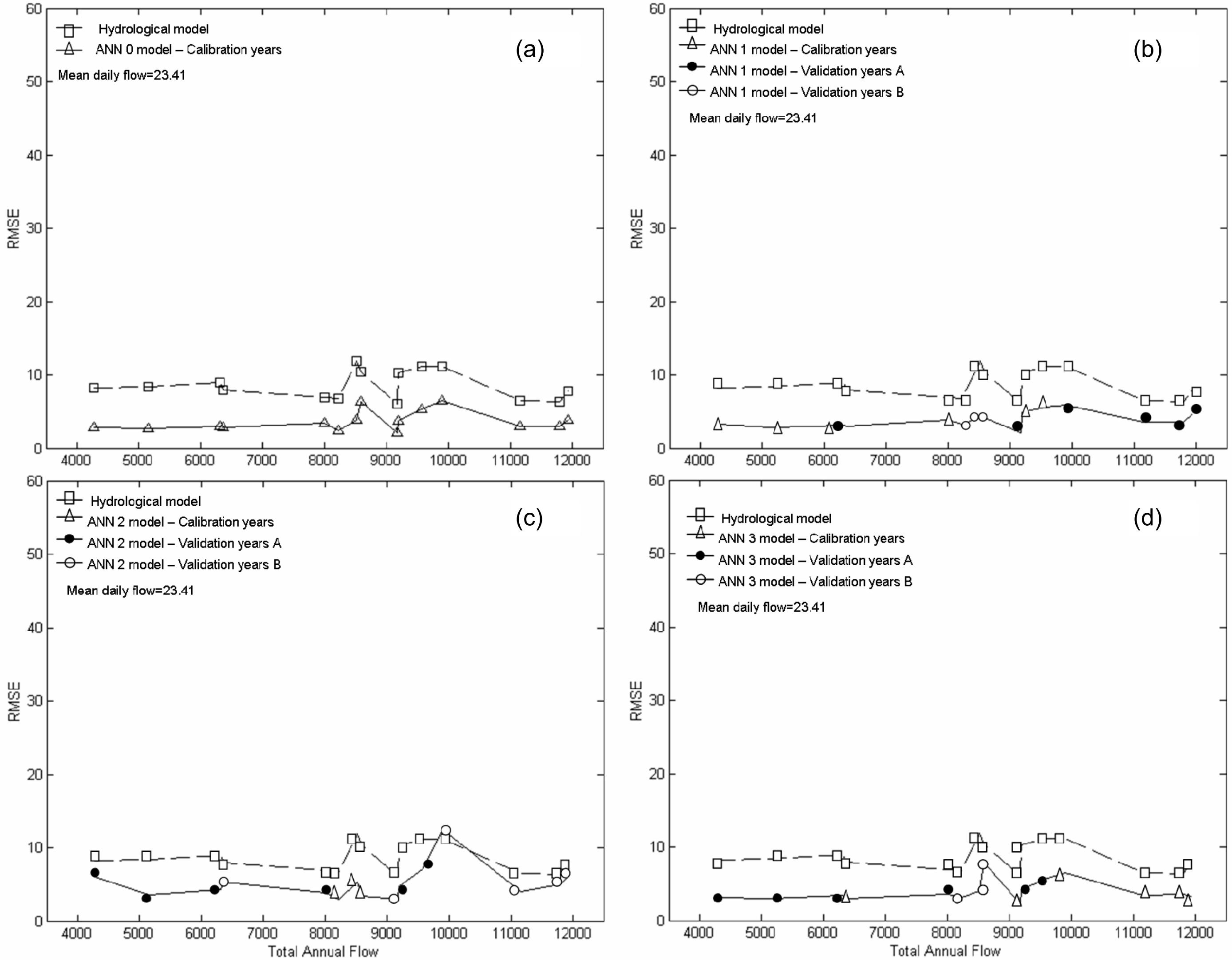


Fig. 6 Annual performance of *RMSE* statistics for below mean flow for each model (SMA, ANN0, ANN1, ANN2 and ANN3) plotted against total flow for each data year.

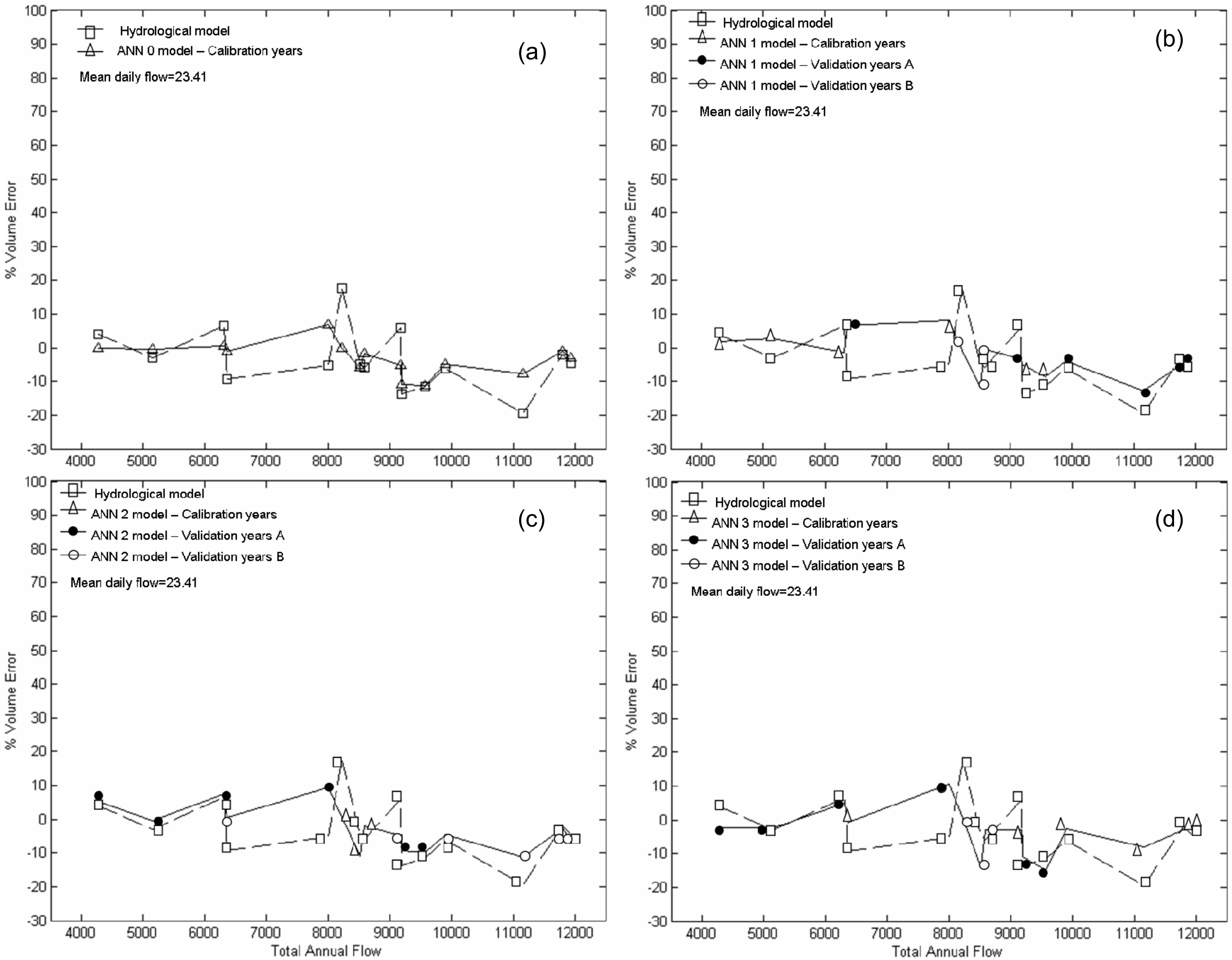


Fig. 7 Annual performance of %*VE* statistics for above mean flow for each model (SMA, ANN0, ANN1, ANN2 and ANN3) plotted against total flow for each data year.

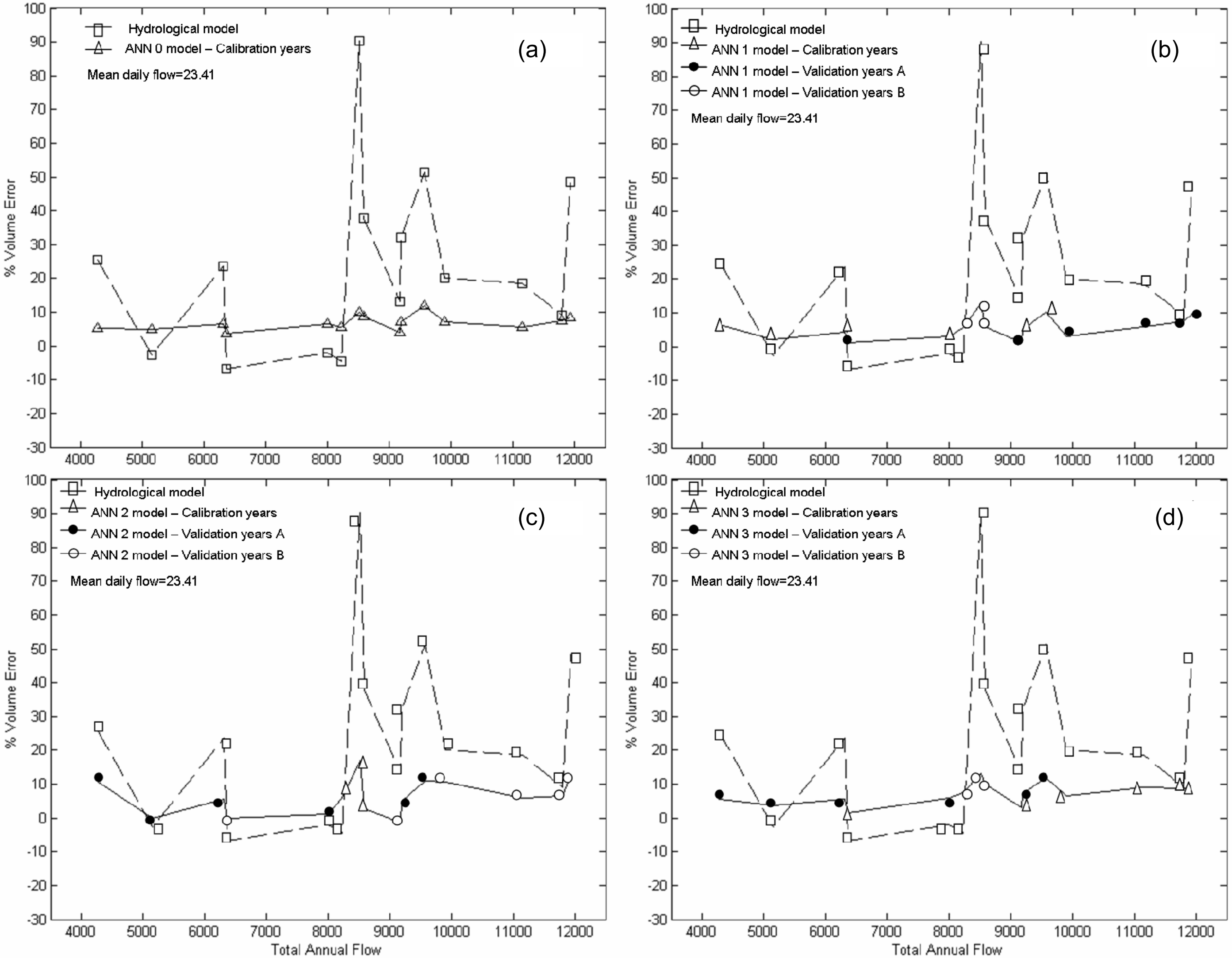


Fig. 8 Annual performance of %*VE* statistics for below mean flow for each model (SMA, ANN0, ANN1, ANN2 and ANN3) plotted against total flow for each data year.

superiority of ANN3 model. The ANN3 and ANN1 models present the smallest bias during the calibration period (rising flows) and the validation period A (rising flows), respectively. In Fig. 8(b), (c) and (d), the ANN*I* models present a somewhat linear tendency between the ability to match the calibration and validation data (residual variance computed by %VE_B) with the wetness of the year, particularly with ANN3 and ANN1 models for the examined calibration/ validation periods.

Figure 9 shows the ability of the SMA (hydrological) and ANN0, as well as the ANN1 and ANN3 models to match the hydrographs focusing on high and low flows for

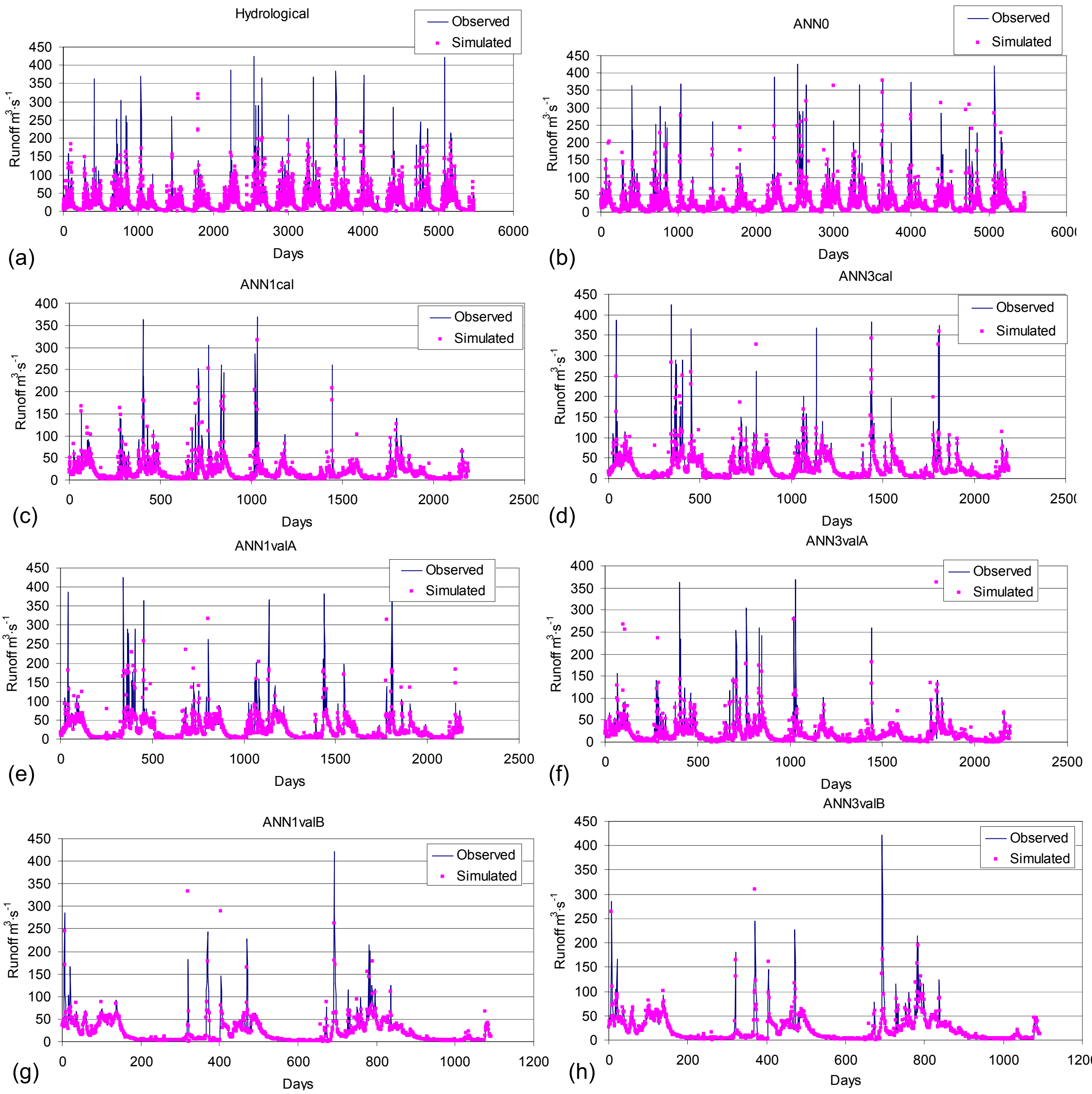


Fig. 9 (a)–(h) Daily hydrographs simulated by SMA, ANN0, ANN1 and ANN3 models for calibration and validation periods.

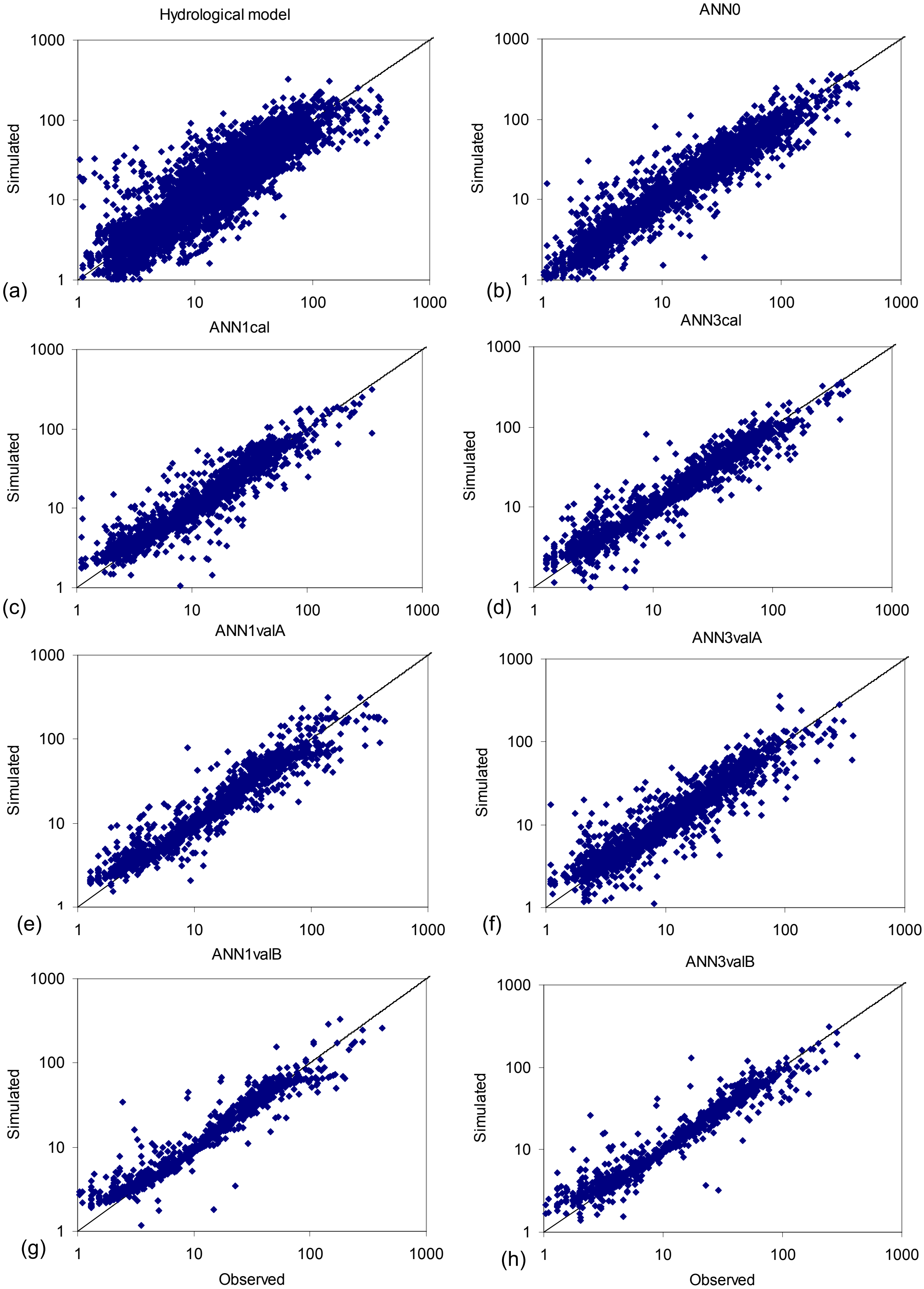


Fig. 10 (a)–(h) Scatter plots comparing simulated and observed flows of SMA, ANN0, ANN1 and ANN3 models for calibration and validation periods.

both calibration and validation periods. The ANN2 model has been excluded in this further investigation as it has been consistently shown to underperform. Figure 9(a) and (b) shows the matching of observed and estimating hydrographs for the SMA and ANN0 models, which are both calibrated on all years. The plot indicates that the ANN0 model can better model the high and low flows than the SMA model, while the same model (ANN0) exhibits a worse performance than the ANN1 and ANN3 calibrated models (Fig. 9(c) and (d)). Figure 9(c) and (d), related to the calibration periods of the ANN1 and ANN3 models, respectively, shows that both models can capture most of the peaks and particularly the ANN3 model that must estimate the highest flows (rising rainfall segment) in the entire 15-year period. The validated models ANN1 and ANN3 over the moderately descending segment (Fig. 9(g) and (h)) fit the high and low flows quite well. This feature may be attributed to the fact that both models are calibrated to intensely variable flows (low and high) and validated to medium values, as reflected by the moderately descending segment of pseudo-precipitation.

Estimated *vs* observed flow plots for the SMA (hydrological) model, ANN1 and ANN3 models for calibration and validation periods are shown in Fig. 10. It is noted that the models were calibrated/validated to flows while the data are presented using logs. Figure 10(a) and (b) indicates that the SMA model tends to have the largest deviation from the 1:1 line. The ANN0 model has the closest matching of simulated and observed flows, indicating that the ANN models are implicitly doing a better job in representing the nonlinearities in partitioning precipitation into precipitation excess. However, the ANN3 calibrated for a smaller period against the ANN0 model presents a closer matching of observed and simulated data. During smaller calibration periods (Fig. 10(c) and (d)), the ANN3 model presents the smallest deviations from the 1:1 line, expressing the rising segment of pseudo-precipitation (increasing flows). The performance of the ANN1 model for this period is worse. In the validation periods, the ANN1 model validated for period B (Fig. 10(g)) over the moderately descending segment of pseudo-precipitation shows the closest match of simulated and observed flows. An almost equally good matching is exhibited by the ANN3 model validated for the same period B (Fig. 10(h)).

SUMMARY AND CONCLUSIONS

The ability of artificial neural network models to simulate high and low flows in various climate conditions over a medium-sized mountainous catchment has been examined. An efficient procedure (LLSSIM, Hsu *et al.*, 1995) for estimating the weights (parameters) of a three-layer ANN was adopted. The nonlinear ANN was employed to simulate the daily flows in three climatically different periods, described by trends of clearly descending, rising and moderately descending segments in the long-term annual pseudo-precipitation (rain plus melt) plot over the Mesochora catchment in central Greece for a 15-year period. The ANN model was calibrated on flows for the entire period and for each of three climatically varied periods, and was validated for the others. The flows above and below the mean daily flow over the 15-year period were considered as high and low flows, respectively, in a preliminary phase of this study. All the applied performance measures and graphs demonstrated the robustness of the ANN model to simulate quite well the high and low flows under various climate regimes. Undoubtedly, there are deviations among evaluation results, as was to be expected, due to the different measures that were employed (see Hsu *et al.*, 1995; Kim & Barros, 2001).

However, the following findings can be reported:

1. The ANN model can simulate high flows quite well when it is calibrated for increasing values of pseudo-precipitation and is validated for moderately decreasing values.
2. Low flows are predicted to acceptably good levels by the ANN model when it is calibrated for increasing values of pseudo-precipitation and validated for moderately, or clearly decreasing, values of pseudo-precipitation.
3. The ANN model can match the calibration and validation residual variance data (measured by $RMSE_{\mathrm{A,B}}$ and $\%VE_{\mathrm{A,B}}$) of high and low flows with the wetness of the year in a somewhat linear relationship when it is calibrated for clearly decreasing values of pseudo-precipitation and validated for increasing values.
4. The daily hydrographs and scatter plots of high and low flows are best represented by the ANN model when it is calibrated for increasing values of pseudo-precipitation and it is validated for moderately decreasing values.
5. On the average statistics and graphs, it could be concluded that the ANN model can simulate well the daily high and low flows, when it is calibrated for increasing pseudo-precipitation and validated for moderately decreasing pseudo-precipitation.
6. For the entire period, the ANN model provided a better representation of the daily high and low flows than the hydrological (SMA) model. Furthermore, the ANN model is more efficient in capturing such flows when it works for classified climates with smaller calibration periods.

Finally, since the ANN model has no physically realistic components and parameters, it is by no means a substitute of the SMA model. Nevertheless, due to model flexibility, the ANN approach provides a viable and effective alternative for daily high- and low-flow simulation and forecasting in climatically varied regimes in cases that do not require explicit analysis of the internal dynamics of the catchment or in cases where insufficient calibration data are available.

Acknowledgements The author wishes to thank the editor and anonymous reviewers for their important suggestions with respect to this manuscript.

REFERENCES

Abrahart, R., J. & Kneale, P. E. (1997) Exploring neural network rainfall–runoff modeling. In: *BHS Sixth National Hydrology Symposium* (Salford, UK), 9:35–9:44.

Abrahart, R. I. & See, L. (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrol. Processes* **14**, 2157–2172.

Anderson, E. A. (1973) National Weather Service River Forecast System: Snow accumulation and ablation model. NOAA Tech Memo. NWS HYDRO 17, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA.

Booy, C. & Lye, L. M. (1989) A new look at flood risk determination. *Water Resour. Bull.* **25**(5), 933–943.

Box, G. E. & Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control.* Holden-Day, Oakland, California, USA.

Bradley, A. A. (1998) Regional frequency analysis methods for evaluating changes in hydrologic extremes. *Water Resour. Res.* **34**(4), 741–750.

Bras, R. L. & Rodriguez-Iturbe, I. (1985) *Random Functions and Hydrology.* Addison-Wesley, Reading, Massachusetts, USA.

Brion, G. M. & Lingireddy, S. (1999) A neural network approach to identifying non-point sources of microbial contamination. *Water Resour. Res.* **33**, 3099–3106.

Burnash, R. J. E., Ferral, R. L. & McGuire R. A. (1973) A generalized streamflow simulation system. Report. 220, Fed-State River Forecast. Centre, Sacramento, California, USA.

Campolo, M., Andreussi, P. & Soldati, A. (1999) River flood forecasting with a neural network model. *Water Resour. Res.* **35**(10), 1191–1197.

Campolo, M., Soldati, A. & Andreussi, P. (2003) Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* **48**(3), 381–398.

Cigizoglu, H. K. (2003) Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrol. Sci. J.* **48**(3), 349–361.

Crawford, N. H. & Linsley, R. K. (1980) Digital simulation in hydrology: Stanford watershed model IV. Tech. Report 39, Dept Civil Engng, Stanford Univ., Stanford, California, USA.

Crespo, M. & Mora, E. (1993) Drought estimation with neural networks. *Advance Engng Software* **18**, 107–170.

Dawson, C. W. & Wilby, R. (1998) An artificial neural network approach to rainfall–runoff modeling. *Hydrol. Sci. J.* **43**(1), 14–66.

Dettinger, M. D. & Cayan, D. R. (1995) Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California. *J. Climate* **8**, 606–623.

Eltahir, E. A. (1996) El Niño and the natural variability in flow of the Nile River. *Water Resour. Res.* **32**(1), 131–137.

Franchini, M. & Pacciani, M. (1991) Comparative analysis of several conceptual rainfall–runoff models. *J. Hydrol.* **122**, 161–219.

Gan, T. Y. & Burges, S. J. (1990) An assessment of a conceptual rainfall–runoff model's ability to represent the dynamics of small hypothetical catchments: models, model properties, and experimental design. *Water Resour. Res.* **26**(7), 1595–1604.

Georgakakos, K. P. & Bae, D. H. (1994) Climate variability of soil water in the American Midwest: Part 2: Spatio-temporal analysis. *J. Hydrol.* **162**, 379–390.

Giustolisi, O. & Laucelli, D. (2005) Improving generalization of artificial neural networks in rainfall–runoff modelling. *Hydrol. Sci. J.* **50**(3), 439–457.

Green, I. R. A. & Stephenson, D. (1986) Criteria for comparison of simple event models. *Hydrol. Sci. J.* **31**(3), 395–411.

Gupta, V. K. (1984) The identification of conceptual watershed models. PhD Dissertation, Case Western Reserve Univ., Cleveland, Ohio, USA.

Hirsh, R. M., Alexander, R. B. & Smith, R. A. (1991) Selection of methods for the detection and estimation of trends in water quality. *Water Resour. Res.* **27**(5), 803–813.

Hirsh, R. M., Helsel, D. R., Cohn, T. A. & Gilroy, E. J. (1993) Statistical analysis of hydrologic data. In: *Handbook of Hydrology* (ed. by D. R. Maidment), 17.28–17.29. McGraw-Hill, New York, USA.

Hsu, K., Gupta, H. V. & Sorooshian, S. (1995) Artificial neural network modeling of the rainfall–runoff process. *Water Resour. Res.* **31**(10), 2517–2530.

Hu, T. S., Lam, K. C. & Ng, S. T. (2001) River flow time series prediction with a range-dependent neural network. *Hydrol. Sci. J.* **46**(5), 729–745.

Hu, T. S., Lam, K. C. & Ng, S. T. (2005) A modified neural network for improving river flow prediction. *Hydrol. Sci. J.* **50**(2), 299–318.

Hurst, H. E. (1951) Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engng* **116**, 770–808.

Imrie, C. E., Durucan, S. & Korre, A. (2000) River flow prediction using neural networks: generalisation beyond the calibration range. *J. Hydrol.* **233**, 138–153.

Kahya, E. & Dracup, J. A. (1993) US streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resour. Res.* **29**(8), 2491–2503.

Karunanithi, N. W., Grenney, D. ,Whitley, D. & Bovee, K. (1994) Neural networks for rivel flow prediction. *J. Comp. Civil Engng* **8**, 201–209.

Kim, G. & Barros, A. P. (2001) Quantitative flood forecasting using multisensor data and neural networks. *J. Hydrol.* **246**, 45–62.

Kitanidis, P. K. & Bras, R. L. (1980a) Adaptive filtering through detection of isolated transient errors in rainfall–runoff models. *Water Resour. Res.* **16**(4), 740–748.

Kitanidis, P. K. & Bras, R. L. (1980b) Real-time forecasting with a conceptual hydrological model, 1. Analysis of uncertainty. *Water Resour. Res.* **16**(6), 1025–1033.

Klemes, V. (1974) The Hurst phenomenon: a puzzle? *Water Resour. Res.* **10**(4), 675–688.

Klemes, V. (1982) Effects of hypothetical climate change on reliability of water supply from reservoirs. A study report by the National Hydrology Research Institute, Ottawa, Ontario, Canada.

Klemes, V. (1985) Sensitivity of water resource systems to climate variations. A study report by the National Hydrology Research Institute, Environment Canada. WMO Report WCP-95, World Meteorological Organization, Geneva, Switzerland.

Lall, U. & Mann, M. E. (1995) The Great Salt Lake: a barometer of interannual climatic variability. *Water Resour. Res.* **31**(10), 2503–2515.

Lettenmaier, D. P. & Gan, T. Y. (1990) Hydrologic sensitivities of the Sacramento-San Joaquin River Basin, California, to global warming. *Water Resour. Res.* **26**(1), 69–86.

Lins, H. F. (1985) Streamflow variability in the United States: 1931–78. *J. Clim. Appl. Met.* **24**, 463–471.

Lisi, F. & Villi, V. (1997) Statistical considerations on the randomness of annual maximum daily rainfall. *J. Am. Water Resour. Assoc.* **33**(2), 431–442.

Loke, E., Warnaars, E. A., Jacobsen, P., Nelsen, F., & do Ceu Almeida, M. (1997) Artificial neural networks as a tool in urban storm drainage. *Water Sci. Tech.* **36**, 101–109.

Mandelbrot, B. B. & Wallis, J. R. (1968) Noah, Joseph and operational hydrology. *Water Resour. Res.* **4**(5), 909–918.

Mimikou, M., Kouvopoulos, Y., Cavadias, G. & Vayianos, N. (1991) Regional hydrological effects of climate change. *J. Hydrol* . **123**, 119–146.

Maier, H. R. &. Dandy, O. C. (1996) The use of artificial networks for the prediction for water quality parameters. *Water Resour. Res.* **32**, 1013–1022.

Minns, A. W. & Hall, M. J. (1996) Artificial neural networks as rainfall–runoff models. *Hydrol. Sci. J.* **41**(3), 399–417.

Minns, A. W. & Hall, M. J. (1997) Living with the ultimate black box: more on artificial neural networks. In: *BHS Sixth National Hydrology Symposium* (Salford, UK), 9.45–9.49.

Nelder, A. J. & Mead, R. (1965) A simplex method for function minimization. *Comput. J.* **7**(4), 308–313.

Nemec, J. & Schaake, J. (1982) Sensitivity of water resource systems to climate variation. *Hydrol. Sci. J.* **27**(3), 327–343.

Nicholls, N. (1989) Sea surface temperature and Australian winter rainfall. *J. Climate* **2**, 966–973.

Nielsen, S. A. & Hansen, E. (1973) Numerical simulation of the rainfall–runoff process on a daily basis. *Nordic Hydrol.* **4**, 171–190.

Panagoulia, D. (1991) Hydrological response of a medium-sized mountainous catchment to climate changes. *Hydrol. Sci. J.* **36**(6), 525–547.

Panagoulia, D. (1992a) Impacts of GISS-modeled climate changes on catchment hydrology. *Hydrol. Sci. J.* **37**(2), 141–163.

Panagoulia, D. (1992b) Hydrological modeling of a medium-sized mountainous catchment from incomplete meteorological data. *J. Hydrol.* **137**(1/4), 279–310.

Panagoulia, D. &. Dimou, G. (1997a) Linking space-time scale in hydrological modeling with respect to global climate change. Part 1. Models, model properties, and experimental design. *J. Hydrol.* **194**, 15–37.

Panagoulia, D. & Dimou, G. (1997b) Sensitivity of flood events to global warming. *J. Hydrol.* **191**, 208–222.

Panagoulia, D. & Maratos, N. (2003a) Artificial neural network modeling of the rainfall–runoff extreme events. In: *Proc. Eighth Int. Conf. on Environmental Science and Technology* (ed. by I. P. Lekkas) (Lemnos Island, Greece, 8–10 September), 48–56. Aegean University Press, Athens, Greece.

Panagoulia, D. & Maratos, N. (2003b) Rainfall–runoff process in mountainous catchment with artificial neural network. In: *Proc. XXX IAHR Congress, Theme D: Hydroinformatics and Advanced Data Technology in Engineering Practice* (ed. by J. Ganoulis & P. Prinos) (AUTh, Thessaloniki, Greece, August 2003), 39–46. AUTh Press, Thessaloniki, Greece.

Peck, E. L. (1976) Catchment modeling and initial parameter estimation for National Weather Service River Forecast System. NOAA Tech. Memo NWS HYDRO-31, Office of Hydrol., National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA.

Raman, H. & Sunilkumar, R. (1995) Multivariate modeling of water resources time series using artificial neural networks. *Hydrol. Sci. J.* **40**(2), 145–163.

Redmond, K. T. & Koch, R. W. (1991) Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. *Water Resour. Res.* **27**(9), 2381–2399.

Richey, J. E., Nobre, C. & Deser, C. (1989) Amazon River discharge and climate variability: 1903–85. *Science* **246**, 101–103.

Rumelhart, D. E., Hinton, E. & Williams, J. (1986) Learning internal representation by error propagation. In: *Parallel Distributed Processing*, vol. 1, 318–362. MIT Press, Cambridge, Massachusetts, USA.

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. (1980) *Applied Modeling of Hydrologic Time Series.* Water Resources Publications, Littleton, Colorado, USA.

Scalero, R. S. & Tepedelenlioglu, N. (1992) A fast new algorithm for training feed-forward neural networks. *IEEE Trans. Signal Processes* **40**(1), 203–229.

See, I. Corne, S. Dougherty, M. & Openshaw, S. (1997) Some initial experiments with neural network models of flood forecasting on the River Ouse. In: *Geo Comp. (97)* (Proc. Second Int. Conf. on Geo Computation **(**Univ. of Otago, Dunedin, New Zealand, 26–29 August). University of Otago Press, Dublin, New Zealand.

Singh, V. P. (1988) *Hydrologic Systems: Rainfall–runoff Modeling,* vol. I. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Smith, M. (1993) *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold, New York, USA.

Sugawara, M. (1974) *Tank Model and Snow Component*. National Research Center for Disaster Prevention, Kyoto, Japan.

Todini, E. (1988) Rainfall–runoff modeling: past, present and future. *J. Hydrol.* **100**, 341–352.

Tomasino, M., Zanchettin, D. & Traverso, P. (2004) Long-range forecasts of River Po discharges based on predictable solar activity and a fuzzy neural network model. *Hydrol. Sci. J.* **49**(4), 673–684.

Trenberth, K. E., Branstator, C. W. & Arkin., P. E. (1988) Origins of the 1988 North American drought. *Science* **242**, 1640–1645.

Tingsanchali, T. & Gautam, M. R. (2000) Application of Tank, NAM, ARMA and neural network models to flood forecasting. *Hydrol. Processes* **14**, 2473–2487.

Vemuri, V. R. & Rogers, R. D. (eds) (1994) *Artificial Neural Network Forecasting Time Series*. IEEE Comp. Soc. Press, Los Alamos, California, USA.

Wall, D. J. & Englot, M. (1985) Correlation of annual peak flows for Pennsylvania streams. *Water Resour. Bull.* **21**(3), 459–464.

Wilby, R. L., Abrahart, R. J. & Dawson, C. W. (2003) Detection of conceptual model rainfall–runoff processes inside an artificial neural network. *Hydrol. Sci. J.* **48**(2), 163–181.

Young, P. & Wallis, S. (1985) Recursive estimation: a unified approach to the identification, estimation, and forecasting of hydrological systems. *Appl. Math. Comput.* **17**, 299–334.

Zhang, Q. & Stanley, S. J. (1997) Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling. *Water Resour Res.* **31**, 2340–2350.

Received 16 July 2004; accepted 9 May 2006