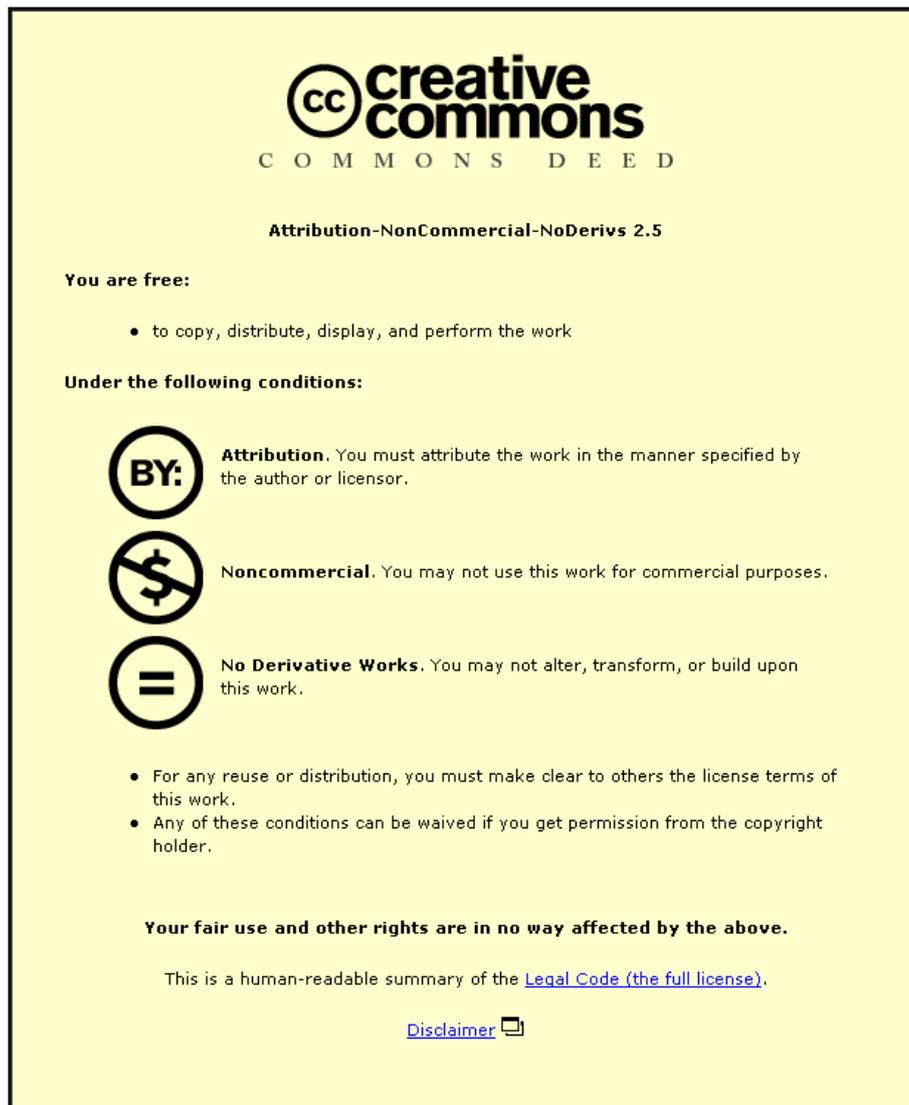


This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

# Flood estimation at ungauged sites using artificial neural networks

Dawson, C.W.<sup>1</sup>, Abrahart, R.J.<sup>2</sup>, Shamseldin, A.Y.<sup>3</sup> and Wilby, R.L.<sup>4,5</sup>

<sup>1</sup> Department of Computer Science, Loughborough University, Loughborough, LE11 3TU, UK

<sup>2</sup> School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK

<sup>3</sup> Department of Civil and Environmental Engineering, The University of Auckland, Private Bag 92019, Auckland, New Zealand

<sup>4</sup> Environment Agency, Trentside Offices, Nottingham, NG2 5FA, UK

<sup>5</sup> Department of Geography, Lancaster University, Lancaster, LA1 4YB, UK

Contact Details:

Christian W Dawson

Email: [c.w.dawson1@lboro.ac.uk](mailto:c.w.dawson1@lboro.ac.uk)

Tel: 01509 222684

Fax: 01509 211586

## Abstract

Artificial neural networks (ANNs) have been applied within the field of hydrological modelling for over a decade but relatively little attention has been paid to the use of these tools for flood estimation in ungauged catchments. This paper uses data from the Centre for Ecology and Hydrology's Flood Estimation Handbook (FEH) to predict T-year flood events and the index flood (the median of the annual maximum series) for 850 catchments across the UK. When compared with multiple regression models, ANNs provide improved flood estimates that can be used by engineers and hydrologists. Comparisons are also made with the empirical model presented in the FEH and a preliminary study is made of the spatial distribution of ANN residuals, highlighting the influence that geographical factors have on model performance.

## Keywords

Artificial neural networks, flood estimation, ungauged catchments

## 1. Introduction

The UK Flood Estimation Handbook (FEH) notes that “many flood estimation problems arise at ungauged sites for which there are no flood peak data” (Reed and Robson, 1999:12). In such cases, the hydrologist is faced with the difficult task of estimating flood event magnitudes from catchment properties and/or regional climatology. The FEH recommends that, wherever possible, such estimates should be based on the transfer of analogous data from sites that are hydrologically similar in terms of catchment area, rainfall and soil type i.e. 'donor sites'. However, it is not always possible to establish an appropriate set of donor sites, and classification of sites into similar groupings can be problematic. Even though attempts have been made to classify catchments (for example, with artificial neural networks; Thandaveswara and Sajikumar, 2000) the FEH notes that there may be fundamental differences between sites that would result in [a] the transfer of inappropriate information and [b] the production of inaccurate flood estimates.

Regionalisation techniques enable the extrapolation of properties of flow regimes across homogeneous regions and the estimation of flow statistics at ungauged sites (Institute of Hydrology, 1980). To date, one of the most extensive studies to regionalise flows in Western Europe was conducted within the framework of the FRIEND (Flow Regimes from International Experimental and Network Data) project (Gustard, 1993). This project, and subsequent studies, highlighted the value of catchment characteristics (such as hydrogeology and soil properties) as descriptors of flows at ungauged sites (Gustard and Irving, 1994). The three most widely applied regionalisation techniques involve: [1] fitting a probability distribution to a flow series, or parameters to a flow duration curve, and then relating the model parameters to physical catchment characteristics (e.g., Smakhtin et al., 1997; Tucci et al., 1995; van der Wateren–de Hoog, 1995); [2] relating index flows with specific return periods (e.g., the mean or median annual flood) to physical catchment characteristics (e.g., NERC, 1975; Schreiber and Demuth, 1997; Vogel and Kroll, 1992); or [3] deriving the parameters of an intermediate conceptual rainfall–runoff model from physical catchment characteristics and then simulating the required discharge sequences (e.g., Ibrahim and Cordery, 1995; Pirt, 1983; Post and Jakeman, 1996; Sefton and Howarth, 1998).

The FEH involves the use of an index flood procedure to derive the flood frequency curve at ungauged sites. The index flood is a middle-sized flood for which the mean or median of the flood data series is typically used (Grover et al., 2002). This procedure is based on the assumption that donor sites have the same flood frequency distribution but differ in terms of the index flood. The flood frequency distribution at the ungauged site is obtained from multiplying the pooled growth curve (dimensionless frequency derived from the data of the donor sites) with the index flood of the ungauged site. In this context the index flood can be viewed as a scaling factor for the growth curve. The FEH uses the median flood to represent the index flood.

It is possible, with standard statistical regression techniques, to produce index flood estimations based on catchment descriptors – for example, derived from catchment area, wetness and base flow index. The FEH also provides algorithms for calculating the index flood for a given site and offers different algorithms for rural and urban catchments. However, Reed and Robson (1999) state that flood estimates ‘made from catchment descriptors are, in general, grossly inferior, to those made from flood peak data’.

The aims of the present investigation are thus threefold: (1) to explore the potential application of artificial neural network (ANN) solutions to the problem of flood estimation in ungauged catchments; (2) to compare ANN model prediction skill with that of the two conventional statistical approaches referred to earlier; and (3) to evaluate possible spatial biases in ANN model output error.

ANNs have been used to perform hydrological modelling operations for over a decade. Since the advent of effective training algorithms for neural networks in the mid 1980s (Rumelhart and McClelland, 1986), neural solutions have been applied to a wide range of hydrological problems, such as rainfall-runoff modelling and river discharge (or stage) forecasting (for a review of forecasting applications see Abraham et al., 2004; Dawson and Wilby, 2001; Govindaraju, 2000). There have, however, been relatively few studies involving the application of ANNs to flood estimation at ungauged sites. For example, at the regional scale, Liong et al. (1994) investigated flood quantile prediction for ungauged catchments in Quebec and Ontario; Muttiah et al. (1997) investigated two-year peak storm discharge predictions for river basins in the United States; Hall and Minns (1998) related the scale and location parameters of the Extreme Value Type 1 (EV1 or Gumbel) distribution for annual floods to six catchment characteristics in two

flood regions of the UK. In subsequent experiments, Hall et al. (2000) used between four and twelve input catchment characteristics to predict the same two EVI parameter outputs using data from sites in Sumatra and Java; whereas Dastorani and Wright (2001) found that seven catchment inputs were sufficient to predict the index flood for selected catchments in the UK. This paper discusses the application of ANNs to predict the index flood for a much larger sample of selected catchments in the UK. It also considers the estimation of 10-, 20- and 30-year flood event magnitudes at such sites. Given the range of record lengths available, the 20-year flood event was chosen for further discussion as it is a convenient metric that is often used for the purposes of comparison in other studies (for example, see Reynard et al., 2004).

The remainder of this paper is arranged as follows. Section 2 provides a brief introduction to ANNs with particular reference to the Multi-Layer Perceptron (MLP). Section 3 describes the data sets and Section 4 the methods that have been applied for flood estimation at ungauged sites. Section 5 considers the error measures that were used to evaluate model performance and Section 6 the results, including a discussion of the geographical distribution of model residuals. Finally, Section 7 provides conclusions and recommendations for further work.

## **2. Artificial neural networks**

Artificial neural networks were first introduced in the 1940s (McCulloch and Pitts, 1943). Interest grew in these tools until the 1960s when Minsky and Papert (1969) showed that networks of any practical size could not be trained effectively. It was not until the mid-1980s that ANNs once again became popular with the research community when Rumelhart and McClelland (1986) rediscovered a calibration algorithm that could be used to train networks of sufficient sizes and complexities to be of practical benefit. Since that time research into ANNs has expanded and a number of different network types, training algorithms and tools have evolved.

Given sufficient data and complexity, ANNs can be trained to model any relationship between a series of independent and dependent variables (inputs and outputs to the network respectively). For this reason ANNs are considered to be a set of *universal approximators* and have been usefully applied to a wide variety of problems that are difficult to understand, define, and quantify – for example, in finance, medicine, engineering, etc. In the context of this paper, ANNs are trained to represent the

relationship between a range of catchment descriptors and associated flood event magnitudes. There is no need for the modeller in this case to fully define the intermediate relationships (physical processes) between catchment descriptors and flood event magnitudes – the ANN identifies these during the "learning process". However, future work may involve ‘drilling’ into network models to extract and interrogate such relationships (e.g., Wilby et al. (2003), Jain et al. (2004) and Sudheer and Jain (2004)) – something that is beyond the scope of the current paper.

Although there are now a significant number of network types and training algorithms, this paper will focus on the Multi-Layer Perceptron (MLP). Figure 1 provides an overview of the structure of this network. In this case the ANN has three layers of neurons (nodes) – an input layer, a hidden layer and an output layer. Each neuron has a number of inputs (from outside the network or the previous layer) and a number of outputs (leading to the subsequent layer or out of the network). A neuron computes its output response based on the weighted sum of all its inputs according to an activation function (in this case the logistic sigmoid). Data flows in one direction through this kind of network – starting from external inputs into the first layer (the predictors), that are transmitted through the hidden layer, and then passed to the output layer from which the external outputs (predictands) are obtained. The network is trained by adjusting the weights that connect the neurons using a procedure called error backpropagation. In this procedure the network is presented with a series of training examples (predictors and their associated predictands) and the internal weights are adjusted in an attempt to model the predictor/predictand relationship. This procedure must be repeated many times before the network begins to model the relationship. Interested readers are directed to neural network texts such as Bishop (1995) for more detailed coverage of such topics.

### **3. Catchments data set**

#### **3.1 Introduction**

The data used in this investigation were obtained from the FEH CD-ROM (Reed and Robson, 1999). The FEH CD-ROM contains data for 1000 sites on drainage paths in mainland Britain, Northern Ireland, the Isle of Wight and Anglesey, which have catchment areas of at least 0.5 km<sup>2</sup>. These data are provided in the form of three separate files for each site. File #1 contains the annual maximum series (AMS), File #2 the peaks-

over-threshold series (sometimes covering a different period to the AMS), and File #3 a set of catchment descriptors for each site. The AMS covers a range of years, some files containing well over 100 years of data from the mid 1800s to the 1990s, while others contain only five or six years of data – usually from the 1970s and 1980s. These data were processed in two stages. First, catchment descriptors were extracted for each site. Second, the AMS was used to estimate [a] the index flood and [b] selected T-year flood events for each catchment.

### 3.2 Catchment descriptors

The FEH CD-ROM contains a number of site descriptors for each catchment, although closer inspection revealed that not all descriptors were available for each catchment. The sixteen descriptors shown in Table 1 were chosen as predictors for this study as they were available for all catchments and provided quantitative representations of catchment characteristics (for information this table also provides the mean value for each descriptor for all 850 catchments used in this study).

### 3.3 Estimation of at-site flood magnitudes

The AMS for each site was extracted from the data and T-year flood events were estimated based on the method of Shaw (1994) assuming a Gumbel Type 1 distribution. It is noted that other distributions could be used but from experience most distributions yield comparable results. As the purpose of this study was to evaluate the effectiveness of ANNs in modelling T-year flood events it did not matter which of the comparable distributions was selected as the ANNs would in all cases be modelling a pseudo T-year flood event.

The annual maximum for a return period of T-years is thus calculated as:

$$Q_T = \bar{Q} + K(T)S_Q \quad (1)$$

$$K(T) = -\frac{\sqrt{6}}{\pi} \left( \gamma + \ln \ln \left[ \frac{T(X)}{T(X) - 1} \right] \right) \quad (2)$$

In which  $\bar{Q}$  is the mean of the annual maximums,  $S_Q$  is the standard deviation of these maximums,  $K(T)$  is a frequency factor and  $T(X)$  is the return period in years.

To increase confidence in the modelling of the T-year flood event the analysis was restricted to a consideration of catchments that had ten or more years of annual maximum data. Several catchments that had significant amounts of missing descriptive data were also removed from the database reducing the number of catchments available in the final modelling operation from 1000 to 850.

The index flood was also calculated for each catchment as the median of the AMS. In cases with an even number of values the index flood was taken as the average of the two middle values. The index flood is a moderate flood event that occurs on average once every two years but is, in contrast, derived directly from the actual data set. It does not need to be estimated from a theoretical frequency distribution which, therefore, removes one potential source of error.

Table 1 shows the correlation between the catchment descriptors and the estimated 20-year flood event and the index flood at each site. As one would expect, characteristics such as *catchment drainage area*, *longest drainage path*, and *mean distance between each node and catchment outlet* are strongly correlated with both the 20-year flood event and the index flood. The similarity of the results also implies a very strong correlation between the 20-year and index floods.

#### **4. Tools and methods**

Four different types of tool are compared in this study. Two data-driven model building strategies were used to develop working neural network flood event predictors based on the use of split-validation and cross-sample methodologies. Two sets of statistical solutions were also developed using step-wise multiple linear regression and the FEH model. These were intended to act as 'benchmark standards'. The first set of neural network solutions developed on the full data set are compared with the step-wise multiple linear regression outputs. The second set of neural network solutions developed on urban and rural partitions of the full data set are compared with the FEH model outputs.

##### **4.1 Neural network split-validation**

The split-validation method (sometimes referred to as cross-validation in the ANN literature) provides a rigorous test of ANN skill (Dawson and Wilby, 2001). It involves dividing available data into three sets: a training set, a validation set, and a test set. The training set is used to fit ANN model weights (for a number of different network

configurations and training cycles), the validation set is used to select the model variant that provides the best level of generalisation, and the test set is used to evaluate the chosen model against unseen data. In this case the 850 data patterns that were available for analysis were split randomly as follows; 424 (50%) catchments for training, 213 (25%) for validation, and 213 (25%) for final testing. The process of random selection produced a reasonable sample of different catchment types and sizes in each sub-set. Table 2 reports the minimum, mean and maximum values of selected catchment properties for the three sub-sets compared with the full data set. Table 2 also indicates that random splitting might not provide the most severe test of model skill since the test data might not contain the most extreme flood events for both the index flood and the 20-year flood event.

Separate networks were trained to predict the 10-, 20-, and 30-year flood events and the index flood. From previous experience network configurations consisting of 3, 5, 10, 15, 20, and 30 hidden neurons were trained using between 100 and 5000 epochs (in steps of 100 epochs) in each case (e.g., Dawson and Wilby, 2001). The training algorithm was 'backpropagation of error', with a low learning rate of 0.1, and a high momentum value of 0.9 (the maximum setting for each parameter is 1.0). Following previous studies, each predictor and predictand was standardised to [0.1, 0.9], such that extreme flood events which exceeded the range of the training data set could be modelled between the boundaries [0, 1] during validation and testing.

## **4.2 Neural network cross-sampling**

To correct for deficiencies in the random division of the sample data sets and to address potential biases arising from urban and rural sub-sets a cross-sampling technique was also employed (sometimes referred to as cross-training). In this case, the whole data set is split into  $S$  segments on a random basis such that each segment contains the same number of data points. Each ANN is trained on  $S-1$  of these segments and tested against the remaining, unseen segment. This procedure is repeated  $S$  times so that each data point in the data set is modelled as an unseen test case once and no points are ignored. Following Schalkoff (1997), ten segments were used. The final solution is in each case evaluated on a full set of segments which means that output statistics can not be directly compared with the split-validation training method.

### 4.3 The benchmark models

Two further approaches were used to provide a standard measure of performance based on conventional and established methods. First, a step-wise multiple linear regression (SWMLR) model was developed on the split-validation data sets using a mixture of forward and backward elimination procedures. This model was designed to predict the 10-, 20, 30-year flood events and index flood. It was developed on the training data set and evaluated on the test data set. The validation data set was not used. The results of these experiments are presented in Section 6.1.

Second, the index flood was derived from catchment descriptors using algorithms provided in the FEH. Models were developed for both urban catchments (those with an urban extent >0.025%) and rural catchments. The skill of these models is compared to the results for the cross-sampled ANN applied to urban and rural partitions in Section 6.2.

## 5. Error measures

Because flood event magnitudes vary significantly between catchments, the following dimensionless error measures were employed in the evaluation of the models: the Mean Squared Relative Error (MSRE), Mean Percent Relative Error (MPRE), the Relative Bias (RB), and the Coefficient of Efficiency (CE). The Standard Error of the Estimate (SE) was also used as this provides an indication of the spread of errors produced by a model (measured in cumecs). The six error measures are calculated according to the following equations:

$$\text{MSRE} = \frac{1}{n} \sum_{i=1}^n \left( \frac{Q_i - \hat{Q}}{Q_i} \right)^2 \quad (3)$$

$$\text{MPRE} = \frac{100}{n} \sum_{i=1}^n \text{abs} \left( \frac{Q_i - \hat{Q}}{Q_i} \right) \quad (4)$$

$$\text{RB} = \frac{1}{n} \sum_{i=1}^n \left( \frac{Q_i - \hat{Q}}{Q_i} \right) \quad (5)$$

$$CE = 1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \quad (6)$$

$$SE = \sqrt{\frac{\sum_{i=1}^n (E - \bar{E})^2}{n}} \quad (7)$$

Where  $Q$  is the observed flood event,  $\hat{Q}$  is the modelled flood event,  $\bar{Q}$  is the mean of the observed flood events,  $E$  is the error (i.e.  $Q - \hat{Q}$ ),  $\bar{E}$  is the mean of the errors, and  $n$  is the number of flood events that have been modelled.

The MSRE and MPRE provide an indication of the relative absolute accuracy of the models while RB provides an idea of whether a model is over- or under-predicting the flood event magnitudes. CE provides an indication of how good a model is at predicting values away from the mean. In this context CE provides some indication of how well the models perform in catchments that possess either particularly low or particularly high flood event magnitudes. The MSRE ranges from 0 for a perfect model to  $\infty$ , and values between 0 and 0.5 would be considered acceptable. MPRE also ranges from 0 for a perfect model to  $\infty$ . RB ranges from  $-\infty$  to  $+\infty$  (negative values indicate a general over-estimation while positive values indicate a general under-estimation of the model) and CE ranges from  $-\infty$  in the worst case to +1 for a perfect model. Shamseldin (1997) suggests a CE value of 0.9 or above to be ‘very satisfactory’, whereas above 0.8 is ‘fairly good’ and below 0.8 is ‘unsatisfactory’.

## 6. Results and Discussion

### 6.1 Model development based on all data

#### 6.1.1 Neural network split-validation method

The results for the split-validation method are provided in Tables 3 and 4. Having tested a number of ANNs on the training set, those configurations shown in Table 3 (*Most accurate models*) were found to be most accurate when evaluated against the independent

validation set using the MSRE and CE statistics. Figure 2 shows the comparative accuracy of the network configurations for both the validation and test data sets using the CE statistic for the 20-year flood event model. While the 20 hidden-node ANN (trained for 2800 epochs) provides the most accurate model for the validation data, a 10 hidden-node ANN (also trained for 2800 epochs) proves to be most accurate at modelling the test data (shown by the two *maximum* indicators in Figure 2). This lends weight to the argument that it is prudent to select parsimonious models that are more likely to be able to generalise than over-parameterised models that may become tuned to noise within the training data. However, although Figure 2 shows that a 10 hidden-node model is more accurate for the test data, in this case it would be wrong to choose this model at this stage as it is in conflict with the split-validation approach (i.e. selection based solely on the validation data). This argument can also be extended to the number of epochs for which a network is trained. Training a network for too long may mean the network has become highly tuned to the training data leading to an inability to generalise.

The general rule-of-thumb is to ensure that there are ‘many more’ training data points than connection weights. This implies that networks should be chosen with as few hidden nodes as possible, and trained for a limited period. Applying this rule to the validation data leads to the selection of the alternative network configurations shown in Table 3 (*Most accurate parsimonious models*). These networks were then evaluated using the independent test set and the results are presented in Table 4. The ANN T-year flood event models are ‘fairly good’ according to Shamseldin’s (1997) criteria with respect to the CE statistic, and the index flood model is ‘very satisfactory’.

Although the training data contains 20-year flood events ranging from 0.61 cumecs to 1288.80 cumecs, the 90<sup>th</sup> percentile of these data is 373.14 cumecs. That is, the majority of the training data contains relatively low magnitude flood events. Given the nature of the data, one would expect this kind of distribution as the data set will be dominated by smaller catchments. Thus, during training the models become ‘fine tuned’ to lower level flood events while higher flood events are rarer. This problem is encountered in any data set containing extremes, for example, river flow forecasting where data are dominated by the lower flow flood events while the extremes (those flood events that one is perhaps more interested in modelling) are less common. Techniques to overcome this problem include resampling from higher-level flood events or restructuring the data set, by eliminating a proportion of the lower flood event data, so that a more even

spread of flood events are included. An alternative is to develop a number of models based on different characteristics in the data set (such as catchment size, flood event size etc.). This approach is investigated later by partitioning into urban and rural catchments. Sivakumar (2005) refers to these kinds of partitions as 'thresholds'. An alternative is to use a network that pre-classifies data into different sets using a clustering technique such as self-organizing maps (Hsu et al., 2002).

Figure 3 shows the accuracy of the 20-year flood event model for the test data. There is one obvious outlier identified as the River Severn at Haw. This is one of the largest catchments in the data set with an area of 9884 km<sup>2</sup>. It is unusual for a catchment of this size to be classed as urban (urban extent is 0.0263%) so one would expect much greater flood events to occur than are actually recorded. However, there were only 17 years of data in the AMS available for the years 1976 to 1992. This period includes some notable droughts; 1976, 1984 and 1988-1992; and, as a consequence, yields a relatively low estimated 20-year flood event. In addition, the flow regime is modified by an impounding reservoir, by abstractions for public, industrial and agricultural supply, and by effluent return (Institute of Hydrology, 1993). All these factors lead to unexpected variations in river flow compared with unregulated, natural catchments with otherwise similar geological characteristics.

The problem in this case seems to be related to the unique behaviour of an individual large catchment for which there is only limited data within the training set. The model has generalised in the case of limited high-magnitude flood events but has been unable to reconcile this extreme case.

### **6.1.2 SWMLR method**

The SWMLR models developed for the different return periods consistently selected the following predictors; drainage area (DTM AREA), standard percentage runoff (SPRHOST), soil moisture deficit (SMDBAR), longest drainage path (LDP) and invariability of slope directions (ASPVAR). In addition, for the index flood the model also selected base flow index (BFIHOST) and proportion of time when SMD < 6mm (PROPWET). Firm conclusions can not be drawn from these selections as the nature of the SWMLR model means that other (quite valid) predictors may be excluded because they are strongly correlated with those selected.

The poor results of this model are presented in Table 5. These findings are particularly disappointing, especially when compared to the high accuracies of the neural solutions. In an attempt to improve the performance of this method in predicting the 20-year flood event the data were logged to reduce the affect of extreme flood events. However, this led to extremely poor results, including some grossly inaccurate predictions. The SWMLR model is also somewhat naïve in assuming linear relationships between variables and potentially useful variables may have been discarded.

## **6.2 Model development based on urban and rural partitions of the data set**

To explore the potential power of data stratification and to make more effective use of limited hydrological records, ANN models were developed for the 20-year flood event and index flood using urban and rural splits using a cross-sampling method. ANN model results for the index flood are compared with those of the FEH model.

### **6.2.1 Neural network cross-sampling method (20-year flood)**

Having identified, with the split-validation approach, the most ‘appropriate’ network model (i.e. a network with 5 hidden neurons trained for 1800 epochs for the 20-year flood event model), this structure was then used in a 10-fold cross-sampling experiment. In this case the data were further split into rural (those with an urban extent of less than 0.025%) and urban catchments to see if any improvement could be made by tuning models to particular catchment types. Table 6 presents catchment statistics for the rural and urban data used in the cross-sampling approach. The results of the cross-sampled 20-year flood event models are presented in Table 7 while Figure 4 shows scatter diagrams of ANN model performance. In Table 7 and Figure 4 *Urban* denotes the ANN model trained and evaluated on urban catchment data only; *Rural* denotes the model trained and evaluated on rural data only; and *All* denotes the model developed and evaluated on all the data (urban and rural catchments combined). In all cases Table 7 shows that the 20-year flood event models are ‘fairly good’ according to the CE statistic for all catchment types.

Figure 4a shows the performance of the urban model during testing, and highlights four notable outliers: two at relatively high values – the River Severn at Haw and the River Ribble at Jumbles Rock; and two at relatively low values – the River Cynon at Abercynon and the River Colne at Denham.

The 20-year flood event for the River Severn was over-estimated in the same way as described in the split-validation approach above. Conversely, the River Ribble is the one notable outlier that has been underestimated by the model. In this case the estimated 20-year flood event was 954 cumecs while the ANN modelled 20-year flood event was 574 cumecs. This catchment has an area of 1049 km<sup>2</sup>, an urban extent of 0.0259%, and so just falls within the urban category. 24 years of AMS data were available for this catchment from 1970 to 1993, so one can assume that the estimated 20-year flood event is a reasonable approximation to the observed flood event. However, examination of the Hydrometric Register (Institute of Hydrology, 1993) indicates that this catchment is a regulated river with an impounding reservoir and is used for public water supplies. When one compares the estimated 20-year flood event for this catchment with a similar catchment it is perhaps not surprising that the model has underestimated this flood event. For example, one such similar catchment is the River Wear at Chester le Street. This catchment has an area of 1005 km<sup>2</sup> and an urban extent of 0.0247%. It is not used for storage or public water supplies but the derived 20-year flood event is 363 cumecs – which is much lower than that of the River Ribble.

At lower levels the 20-year flood events for the River Cynon at Abercynon and the River Colne at Denham have been notably over-estimated by the urban ANN model. The River Cynon is a small catchment (103 km<sup>2</sup>) with a relatively high average annual rainfall of 1766 mm (base flow index: 0.422; longest drainage path: 28.69 km; mean slope: 145.76 m/km; urban extent: 0.0388%; mean altitude above sea level: 270 m) and is described as having 17% forest and with open-cast coal extraction in headwaters. Thirty-two years of AMS data were available for this catchment. A similar urban catchment to this is the River Irwell at Bury Bridge which has a drainage area of 156 km<sup>2</sup>. In this case the estimated 20-year flood event is 302 cumecs which is more in line with prediction made by the model.

The River Colne is a medium sized catchment (733 km<sup>2</sup>) with considerable suburban development in the middle and lower reaches (base flow index: 0.623; average annual rainfall: 703 mm; longest drainage path: 68.5 km; mean slope: 43.67 m/km; urban extent: 0.0754%). It does appear to have a particularly low derived 20-year flood event of 16.26 cumecs (based on 41 years of AMS data from 1953 to 1993). For comparison, the River Aire at Armley is of a similar size (686 km<sup>2</sup>) and urban extent (0.0743%), yet has a derived 20-year flood event of 194 cumecs – more in line with the 264 cumecs

predicted by the model. There are clearly some other influences at work here that require further investigation.

For rural catchment models there are two notable outliers that have been underestimated – the River Findhorn at Forres and the River Lochy at Camisky (Figure 4b). The River Findhorn is a medium sized catchment (781 km<sup>2</sup>) with an urban extent of 0.0001% (base flow index: 0.434; average annual rainfall: 1065 mm; longest drainage path: 100.13 km; mean slope: 119.83 m/km; and has extensive blanket peat cover that drains the Monadhliath Mountains). Thirty-two years of AMS data were available for this catchment and it is classified as *natural* (Institute of Hydrology, 1993). In these circumstances one would expect the observed data to be sufficient to provide a reasonable estimation of the 20-year flood event. A similar catchment to this – the River Dee at Polhollick - with an area of 697 km<sup>2</sup> and urban extent of 0.0001% (base flow index: 0.458; average annual rainfall: 1231 mm; longest drainage path: 62.68 km; mean slope: 224.44 m/km; and described as being a mountain, moorland and pastoral catchment) has a 20-year flood event of 501 cumecs compared with 1171 cumecs for the River Findhorn. This is also described by the Hydrometric Register as *natural* and thus provides a good comparison of the flood magnitude that might be expected.

The River Lochy also appears to have an estimated 20-year flood event that is higher than expected. This catchment has an area of 1256 km<sup>2</sup> and an urban extent of 0.0003% (base flow index: 0.386; average annual rainfall: 2188 mm; longest drainage path: 83.14 km; mean slope: 249.63 m/km; and is described as comprising mainly rough grazing and moorland with some afforestation). There were only 13 years of AMS data for this site covering the period 1981 to 1993 and there were four annual maxima over 1000 cumecs in this limited period (the estimated 20-year flood event is thus open to some uncertainty). For example, there was a recorded flood event in January 1992 of 1540 cumecs – significantly higher than the smallest annual maximum recorded here of 449 cumecs in 1988. The catchment is also subject to the artificial influences of a reservoir. Compared with catchments of a similar size (ranging from 1100 km<sup>2</sup> to 1400 km<sup>2</sup>), the next highest recorded 20-year flood event is for the Wye at Erwood (980 cumecs) while the average 20-year flood event for all catchments between 1100km<sup>2</sup> and 1400km<sup>2</sup> is 493 cumecs. These outliers perhaps show the dangers of a) using donor catchments to predict flood events at unseen sites, b) estimating T-year flood events from

a limited number of data points, c) highly localised extreme events that are not captured by the annual rainfall statistics listed in Table 1.

The remaining row in Table 7 (*All*) enables comparisons to be made with the split-validation approach in the previous section. In this case the model has been trained on all catchment types and evaluated against all catchment types. Figure 4c shows the results of this model when compared with the estimated 20-year flood event. Note that the same outlier rural catchments are again under-estimated by the ANN model.

Comparing the results of this model with the 20-year flood event split-validation method in Table 4 there is some worsening of model performance across all statistics. This is due to the fact that the cross-sampled model is being tested against the entire data set. This is a far more stringent test of model performance than the smaller test subset used in the split-validation approach, which did not include such extreme values (see Table 2).

The results show that there are still occasional anomalies in model performance leading to some significant over- or under-estimates. This may be attributed to the limited data for estimating the 20-year flood event. Conversely, it highlights the dangers of using donor catchments that may provide significantly different estimates of flood events than observed, particularly if artificial influences are not considered in the comparisons.

### **6.2.2 Neural network cross-sampling method (index flood)**

The FEH approach provides a method for estimating the index flood from catchment descriptors. The index flood is first calculated for rural catchments as a function of area, base flow index, standard percentage runoff, flood attenuation index attributable to reservoirs and lakes, and average annual rainfall. This can then be adjusted for urban catchments by further calculations involving standard percentage runoff and urban extent. Table 8 compares the performance of the urban and rural algorithms with the index flood estimated directly from the AMS. The results show the urban model (*Urban-FEH*) provides ‘very satisfactory’ results while the rural model (*Rural-FEH*) is ‘fairly good’ according to the CE statistic.

Table 8 also presents the results of the ANN index flood models produced using the cross-sampling approach (*Rural-ANN*, *All-Rural-ANN*, *Urban-ANN*, *All-Urban-ANN*). The *Rural-ANN* model was trained and evaluated using rural catchment data only and the

*Urban-ANN* model was trained and evaluated using urban catchment data only. In order to see if training networks using the entire data set could make improvements, two further models were developed. The *All-Rural-ANN* model was trained on all available catchment data but evaluated on rural catchments only; the *All-Urban-ANN* model was trained on all available catchment data and evaluated on urban catchments only. The *All-Rural-ANN* model involved training 10 models using all the urban data and 90% of the rural data before testing on the unseen 10% of the rural data. This was repeated 10 times so that all the rural data were eventually tested as unseen. The same procedure was adopted for the *All-Urban-ANN* model.

In the case of the rural models the ANN has outperformed the FEH model according to both the CE and SE statistics. This implies that the ANN model is performing well across the range of index flood magnitudes but less so for smaller flood events as evidenced by the MSRE.

Figure 5a highlights a problem with the FEH approach. While the FEH model performs reasonably well for low magnitude flood events, flood events above 100 cumecs are consistently under-estimated, and generally appear to worsen as the magnitude increases. This conflicts with the findings of Ashfaq and Webster (2002) who modelled 88 representative catchments and reported that in general the FEH method over-estimated flood quantiles. This problem was also found to be more pronounced for higher return periods and most pronounced in catchments that experienced less than 800 mm average annual rainfall i.e. in the south east. However, the RB statistic of the *Rural-FEH* model is negative (-0.0424), which implies that lower level flood events are in general terms being over-estimated in compensation (closer inspection of these results highlighted some particularly large individual relative over-estimates at lower levels leading to this negative RB statistic). This may reflect the non-linear nature of the function that is better captured with the non-linear ANN. There is some improvement in skill for intermediate and large floods by the *Rural-ANN* and *All-Rural-ANN* models (Figures 5b and 5c), but this is at the expense of the relative accuracy of the model according to the MSRE statistic.

There is one notable outlier from the two ANN models for the rural data set: the River Ouse at Skelton. This is a large rural catchment of 3302 km<sup>2</sup> with mixed geology (base flow index: 0.439; average annual rainfall: 899 mm; longest drainage path: 149.96 km; mean slope: 70.17 m/km; urban extent 0.0103%). The River Tweed at Sprouston is

of similar size (3352 km<sup>2</sup>) and smaller urban extent (0.0028%) yet has a much higher computed index flood than the River Ouse (739 cumecs compared with 357 cumecs for the River Ouse). The anomaly for the River Ouse could be explained in terms of gauging errors for peak flows or an observed record containing relatively few major floods.

The *Urban-FEH* model performs relatively well and is classed as a ‘very satisfactory’ model according to the CE (Table 8). Although the *Urban-ANN* model has a smaller RB than the FEH model (Table 8), according to other diagnostics it appears to be performing less well. This is probably due to the limited amount of data that were available for training the ANN model: with 190 urban catchments available and a 10-stage cross-sampling approach, only 171 data points were available for training. To overcome the problem of small sample sizes another ANN model was trained, this time using all the available data (*All-Urban-ANN*). This meant that from 850 data points, the network was trained using 831 points (i.e. all the rural data plus 90% of the urban data). This led to a marked improvement in ANN performance according to the CE statistic (now in the ‘very satisfactory’ category at 90.59%) and SE statistic, but a reduction in relative performance according to the MSRE and MPRE. This is explained by the different nature of the urban and rural catchment data sets as shown in Table 6. The mean index flood for urban catchments is 594.94 cumecs, compared with 951.06 cumecs for rural catchments. The inclusion of rural data in training the urban model reduces the influence of smaller flood events by including a greater number of large flood events. Thus, the *All-Urban-ANN* has become less sensitive to smaller flood events (the MSRE has increased) while its overall performance has improved (CE has increased). Figure 6c shows that there has been some deterioration in the estimate for the River Severn at Haw - an urban catchment with a relatively large index flood. Because the *All-Urban-ANN* model has been trained on a much larger data set consisting of (now) mainly rural data, there is a decline in performance for this urban catchment.

Figure 6 (a–c) also shows that all models appear to under-estimate the index flood for the River Ribble at Jumbles Rock. The 20-year flood event for this catchment was also under-predicted by the ANN models. The characteristics of this catchment are such that the index flood is somewhat higher than one would expect. It is not surprising, therefore, that all three models have under-estimated the index flood as they are basing their estimates on these characteristics.

### **6.3 Geographical analysis of index flood predictions**

The index flood predictions for the 850 catchments were used to construct Thiessen polygon maps of the model residuals. Error maps were developed on the IHDTM geographical coordinate pairings related to each catchment centroid – as provided in the FEH. Two initial problems were experienced. Following visualization and testing operations IHDTM coordinates were used instead of NGR coordinates due to the requirement for a unique set of catchment input points. Northern Ireland catchment centroids were also found to be problematic and had to be re-projected in a GIS: 39 of the 850 catchment coordinates were registered to the Irish National Grid – as opposed to the GB National Grid.

Figure 7(a-c) shows index flood error maps developed on the FEH model predictions; the neural network split-validation model predictions; and a combined map of both urban and rural neural network cross-sampled model predictions. The maps are standardised to a common scale and the spatial pattern on the different maps appears to be in broad agreement. Low errors occur throughout baseflow dominated catchment regimes of the South East. Relatively large errors occur in North and South Wales and in Northern England and the Scottish Highlands. This distinction equates to the wetter and higher altitude regions of the UK. The size and spread of individual catchments across the map also reveals a disproportionate distribution of input records with relatively few polygons in the most challenging regions with highest rainfall. Thus, the nature and extent of the residuals can be explained in terms of broad scale geological and climatological gradients suggesting that additional descriptors are needed to complement those in Table 1.

## **7. Conclusions**

The results of this study show that ANNs can be used to estimate flood statistics for ungauged catchments. The ANNs reproduce the index flood with comparable accuracy to that obtained by the FEH models. It should be noted that while ANNs have been trained in this study to model T-year flood magnitudes derived from the Gumbel distribution, they could just as easily be trained to model floods derived from any other distribution.

Although it is possible to use conventional statistical approaches to build models for predicting T-year flood events (such as SWMLR), the ANN proved to be superior in this study. However, there are a few caveats to be noted. First, the ANN is heavily data

dependent. This was highlighted by improvements in skill achieved by training ANNs on the full available data set instead of a limited (urban) data set. Second, the ANNs cannot explicitly account for physical processes, reducing confidence in model predictions. Finally, despite limiting the analysis to those sites that had at least ten years of record, the limited data at certain sites meant that some T-year flood events and index floods could be grossly under- or over-estimated. This is exacerbated when the data include periods of long-term drought or above average long-term rainfall. In these cases, the ANN may be predicting the T-year flood event accurately, but, with only limited observed data, evaluation of skill can be problematic.

While this study demonstrates the feasibility of using ANNs to model flood events in ungauged catchments, there are still a number of areas of further work. First, it would be useful to investigate different ways of partitioning the data into categories other than rural and urban (see Sivakumar, 2005); for example, based on geology, size or climatic region (as highlighted by the geographical analyses). This would lead to a series of models tuned to the idiosyncrasies of particular catchment types. Second, in catchments where the models appear to be significantly over- or under-predicting estimated flood events, it would be worth exploring anomalies in relation to a wider set of catchment characteristics. Third, other ANN model configurations could be evaluated alongside the backpropagation feedforward network used herein (e.g., radial basis function networks and support vector machines). Finally, an investigation of ANN parameters could yield further insights into the relationships between catchment properties and flood estimation in ungauged catchments.

### **Acknowledgements**

We thank the reviewers for their constructive remarks. The views contained in this paper reflect those of the authors and are not necessarily indicative of the position held by the Environment Agency.

## References

- Abrahart, R.J. Kneale, P.E. and See, L. (eds) (2004) *Neural Networks for Hydrological Modelling*, Taylor & Francis, London.
- Ashfaq, A. and Webster, P. (2002) 'Evaluation of the FEH Rainfall-Runoff Method for Catchments in the UK', *Journal of the Chartered Institution of Water and Environmental Management*, **16**, 223 - 228.
- Bishop, C.M. (1995) *Neural networks for pattern recognition*, Clarendon Press, Oxford, UK.
- Dastorani, M.T. and Wright, N.G. (2001) 'Application of artificial neural networks for ungauged catchment flood prediction', Floodplain Management Association Conference, San Diego, CA, March.
- Dawson, C.W. and Wilby, R.L. (2001) 'Hydrological modelling using artificial neural networks', *Progress in Physical Geography*, **25(1)**, 80 - 108.
- Govindaraju, R.S. (2000) 'Artificial neural networks in hydrology II. Hydrological applications', *Journal of Hydrologic Engineering*, **5(2)**, 124 - 137.
- Grover, P.L., Burn, D.H. and Cunderlik, J.M. (2002) 'A comparison of index flood estimation procedures for ungauged catchments', *Canadian Journal of Civil Engineering*, **29(5)**, 734 - 741
- Gustard, A. (ed.) (1993) *Flow Regimes from International Experimental and Network Data (FRIEND)* Vol. I *Hydrological Studies*, Vol. II *Hydrological Data*, Vol. III *Inventory of Streamflow Generation Studies*. Institute of Hydrology, Wallingford, UK.
- Gustard, A. and Irving, K.M. (1994) Classification of the low flow response of European soils. In Seuna, P., Gustard, A., Arnell, N.W. and Cole, G.A. (eds) *FRIEND: Flow Regimes from International Experimental and Network Data* (Proc. 2nd FRIEND Conf., Braunschweig, Germany, October 1993). IAHS Publ. No. 221.
- Hall, M.J. and Minns, A.W. (1998) 'Regional flood frequency analysis using artificial neural networks', *Proc. 3rd Int. Conf. on Hydroinformatics*, **2**, Balkema, Rotterdam, 759 - 763.

- Hall, M.J., Minns, A.W. and Ashrafuzzaman, A.K.M. (2000) 'Regionalisation and data mining in a data-scarce environment', *BHS 7th National Hydrology Symposium*, Newcastle, UK, 3.39 - 3.43.
- Hsu, K., Gupta, H.V., Gao, X., Sorooshian, S., Imam, B. (2002) 'Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis', *Water Resources Research*, **38(12)**, 38.1 – 38.17.
- Ibrahim, A.B. and Cordery, I. (1995) 'Estimation of recharge and runoff volumes from ungauged catchments in eastern Australia' *Hydrological Sciences Journal*, **40**, 499 – 515.
- Institute of Hydrology (1980) *Low flow studies*. Wallingford, UK.
- Institute of Hydrology (1993) *Hydrological data UK*, Wallingford, UK.
- Jain, A. Sudheer, K.P. and Srinivasulu, S. (2004) 'Identification of physical processes inherent in artificial neural network rainfall runoff models' *Hydrological Processes*, **18**, 571 – 583.
- Liong, S.Y., Nguyen, V.T.V., Chan, W.T. and Chia, Y.S. (1994) 'Regional Estimation of Floods for Ungauged Catchments with Neural Networks', In: Cheong, H-F., Shankar, N.J., Chan, E-S. and Ng, W-J. (Eds.) *Developments in Hydraulic Engineering and their impact on the Environment, Proceedings Ninth Congress of the Asian and Pacific Division of the International Association for Hydraulic Research, Singapore*, 24-26 August, 372 - 378.
- McCulloch, W.S. and Pitts, W. (1943) 'A logical calculus of the ideas imminent in nervous activity', *Bulletin and Mathematical Biophysics*, **5**, 115 - 133.
- Minsky, M.L. and Papert, S.A. (1969) 'Perceptrons', MIT Press, Cambridge, MA.
- Muttiah, R.S. Srinivasan, R. and Allen, P.M. (1997) 'Prediction of Two-Year Peak Stream Discharges Using Neural Networks', *Journal of the American Water Resources Association*, **33(3)**, 625 - 630.
- NERC (1975) *The Flood Studies Report (5 volumes)*. Natural Environment Research Council, Wallingford, UK.
- Pirt, J. (1983) 'A simple model to simulate catchment response to rainfall', *Department of Geography Occasional Paper Number 7*, Loughborough University, Loughborough.

- Post, D. and Jakeman, A. (1996) 'Relationships between the physical attributes and hydrologic response characteristics in small Australian mountain ash catchments', *Hydrological Processes*, **10**, 877-892.
- Reynard, N., Crooks, S., Wilby, R.L. and Kay, A. (2004) 'Climate change and flood frequency in the UK' *Proceedings of the 39<sup>th</sup> Defra Flood and Coastal Management Conference*, University of York, UK.
- Reed, D.W. and Robson, A.J. (1999) *Flood estimation handbook*, Volume 3, Centre for Ecology and Hydrology, UK.
- Rumelhart, D.E. and McClelland, J.L. (Eds.) (1986) 'Parallel Distributed Processing: Explorations in the Microstructures of Cognition', **1**, MIT Press, Cambridge.
- Schalkoff, R.J. (1997) *Artificial neural networks*, McGraw-Hill, Singapore.
- Schreiber, P. and Demuth, S. (1997) 'Regionalization of low flows in southwest Germany', *Hydrological Sciences Journal*, **42**, 845–858.
- Sefton, C.E.M. and Howarth, S.M. (1998) 'Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales', *Journal of Hydrology*, **211**, 1-16.
- Shamseldin, A.Y. (1997) 'Application of a neural network technique to rainfall-runoff modeling', *Journal of Hydrology*, **199**, 272 - 294.
- Shaw, E.M. (1994) 'Hydrology in practice' (3rd edition), Chapman & Hall, London.
- Sivakumar, B. (2005) 'Hydrologic modeling and forecasting: role of thresholds', *Environmental Modelling & Software*, **20**, 515 - 519.
- Smakhtin, V.Y., Hughes, D.A. and Creuse–Naudin, E. 1997. Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrological Sciences Journal*, **42**, 919–936.
- Sudheer, K.P. and Jain, A. 2004. Explaining the internal behaviour of artificial neural network river flow models. *Hydrological Processes*, **18**, 833-844.
- Thandaveswara, B.S. and Sajikumar, N. (2000) 'Classification of river basins using artificial neural networks', *Journal of Hydrologic Engineering*, **5(3)**, 290 – 298.
- Tucci, C., Silveira, A., Sanchez, J. and Albuquerque, F. (1995) 'Flow regionalization in the upper Paraguay basin, Brazil', *Hydrological Sciences Journal*, **40**, 485–497.
- van der Wateren–de Hoog, B. (1995) 'The effect of climate variability on discharge as dependent on catchment characteristics in the Upper Loire basin, France', *Hydrological Sciences Journal*, **40**, 633–646.

- Vogel, R.M. and Kroll, C.N. (1992) 'Regional geohydrologic–geomorphic relationships for the estimation of low–flow statistics', *Water Resources Research*, **28**, 2451–2458.
- Wilby, R.L. Abrahart, R.J. and Dawson, C.W. (2003) 'Detection of conceptual model rainfall-runoff processes inside an artificial neural network', *Hydrological Sciences Journal*, **48(2)**, 163 - 181.

Abbreviation	Parameter	Mean for all catchments	Correlation with 20-year flood event	Correlation with index flood
DTM AREA	Catchment drainage area (km <sup>2</sup> )	410.77	0.61	0.62
BFIHOST	Base flow index	0.50	-0.23	-0.23
SPRHOST	Standard percentage runoff	36.86	0.27	0.27
FARL	Index of flood attenuation attributable to reservoirs and lakes	0.97	-0.11	-0.10
SAAR	Standard period (1961-1990) average annual rainfall (mm)	1084.76	0.25	0.27
RMED-1D	Median annual maximum one-day rainfall (mm)	39.11	0.18	0.19
RMED-2D	Median annual maximum two-day rainfall (mm)	51.85	0.21	0.22
RMED-1H	Median annual maximum one-hour rainfall (mm)	10.73	-0.15	-0.14
SMDBAR	Mean Soil Moisture Deficit for 1941 – 1970 (mm)	25.21	-0.43	-0.43
PROPWET	Proportion of time when Soil Moisture Deficit<6mm during 1961 - 1990	0.46	0.39	0.39
LDP	Longest drainage path (km)	39.95	0.67	0.68
DPLBAR	Mean distance between each node (on a regular 50m grid) and catchment outlet (km)	21.48	0.67	0.67
ALTBAR	Mean altitude of catchment above sea level (m)	207.47	0.36	0.35
DPSBAR	Mean of all inter-nodal slopes in catchment (m/km)	97.71	0.30	0.30
ASPVAR	Invariability of slope directions	0.18	-0.38	-0.38
URBEXT1990	Extent of urban and suburban land cover in 1990 (%)	0.03	-0.13	-0.13

**Table 1** FEH catchment descriptors

	<b>Catchment Attributes</b>						
	<b>Area</b> (km <sup>2</sup> )	<b>Base</b> <b>Flow</b> <b>Index</b>	<b>Average</b> <b>Annual</b> <b>Rainfall</b> (1961-1990) (mm)	<b>Longest</b> <b>Drainage</b> <b>Path</b> (km)	<b>Urban</b> <b>Extent</b> (1990) (%)	<b>Index</b> <b>Flood</b> (cumecs)	<b>20-Year</b> <b>Flood Event</b> (cumecs)
<b>Minimum</b>							
Full data set	1.07	0.17	547	2.41	0.000	0.32	0.61
Training set	1.07	0.18	547	2.69	0.000	0.32	0.61
Validation set	3.10	0.17	557	3.83	0.000	0.37	0.61
Test set	2.30	0.18	555	2.41	0.000	0.43	1.36
<b>Mean</b>							
Full data set	409.00	0.50	1082	39.88	0.027	87.86	149.70
Training set	409.58	0.50	1088	40.13	0.023	82.18	138.56
Validation set	411.81	0.49	1080	39.44	0.026	103.19	177.08
Test set	405.16	0.49	1074	39.84	0.026	82.80	142.02
<b>Maximum</b>							
Full data set	9951	0.97	3473	280.96	0.432	951.06	1533.94
Training set	9951	0.97	3473	273.09	0.432	751.11	1288.80
Validation set	7490	0.97	2808	157.86	0.424	951.06	1533.94
Test set	9895	0.96	2576	280.96	0.424	572.23	1075.34

**Table 2** Statistics for selected catchment descriptors in split-validation data sets

<b>Model</b>	<b>Hidden Neurons</b>	<b>Epochs</b>
<b>Most accurate models</b>		
<b>10-year flood</b>	20	3200
<b>20-year flood</b>	20	2800
<b>30-year flood</b>	20	2600
<b>Index flood</b>	10	2400
<b>Most accurate parsimonious models</b>		
<b>10-year flood</b>	5	1800
<b>20-year flood</b>	5	1800
<b>30-year flood</b>	5	1800
<b>Index flood</b>	5	2200

**Table 3** Optimal ANN configurations for each flood event evaluated on validation data set for split-validation approach

<b>T-Year</b>	<b>MSRE</b>	<b>CE</b>	<b>SE</b> <b>(cumecs)</b>	<b>MPRE</b>	<b>RB</b>
<i>10-Year</i>	2.26	87.09	57.26	77.75	-0.1084
<i>20-Year</i>	2.50	85.60	68.11	80.51	-0.0945
<i>30-Year</i>	2.66	84.77	74.84	81.91	-0.0842
<i>Index flood</i>	1.98	90.48	34.14	70.76	0.0480

**Table 4** ANN performance for flood events evaluated on split-validation test data set

<b>T-Year</b>	<b>MSRE</b>	<b>CE</b>	<b>SE (cumecs)</b>	<b>MPRE</b>	<b>RB</b>
<i>10 Year</i>	91.38	66.39	93.08	249.26	1.2606
<i>20 Year</i>	86.64	65.23	106.50	244.48	1.2119
<i>30 Year</i>	84.99	63.86	115.97	242.78	1.1732
<i>Index flood</i>	90.33	71.19	59.42	260.02	1.1442

**Table 5** SWLMR model performance for flood events evaluated on split-validation test data set

<b>Catchment Attributes</b>							
	<b>Area (km<sup>2</sup>)</b>	<b>BFIHost</b>	<b>SAAR (mm)</b>	<b>LDP (km)</b>	<b>URBExt (1990)</b>	<b>Index Flood (cumecs)</b>	<b>20-Year Flood Event (cumecs)</b>
<b>Rural Catchments (660 data points)</b>							
Minimum	1.07	0.23	547	2.41	0.000	0.32	0.61
Mean	374.30	0.50	1139	39.76	0.007	95.94	162.54
Maximum	6853.22	0.97	3473	265.52	0.025	951.06	1533.94
<b>Urban Catchments (190 data points)</b>							
Minimum	9.93	0.17	555	5.40	0.025	0.43	0.66
Mean	527.32	0.49	883	39.70	0.096	58.84	103.26
Maximum	9951.00	0.87	2183	280.96	0.432	594.94	953.65

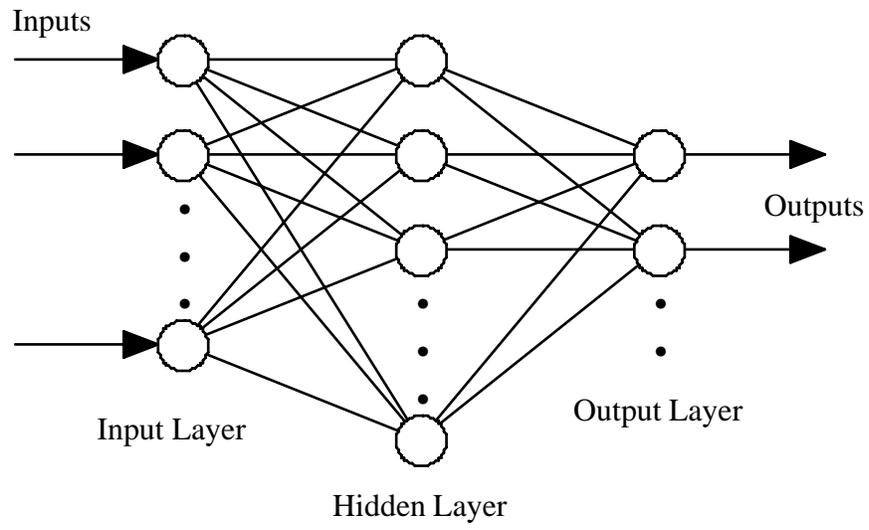
**Table 6** Statistics for selected catchment descriptors in cross-sampling data sets

<b>Catchments</b>	<b>MSRE</b>	<b>CE</b>	<b>SE (cumecs)</b>	<b>MPRE</b>	<b>RB</b>
<i>Urban</i>	5.21	83.94	63.25	92.25	-0.3347
<i>Rural</i>	18.27	83.37	92.51	129.50	-0.7394
<i>All</i>	15.39	83.03	87.24	145.02	-0.8505

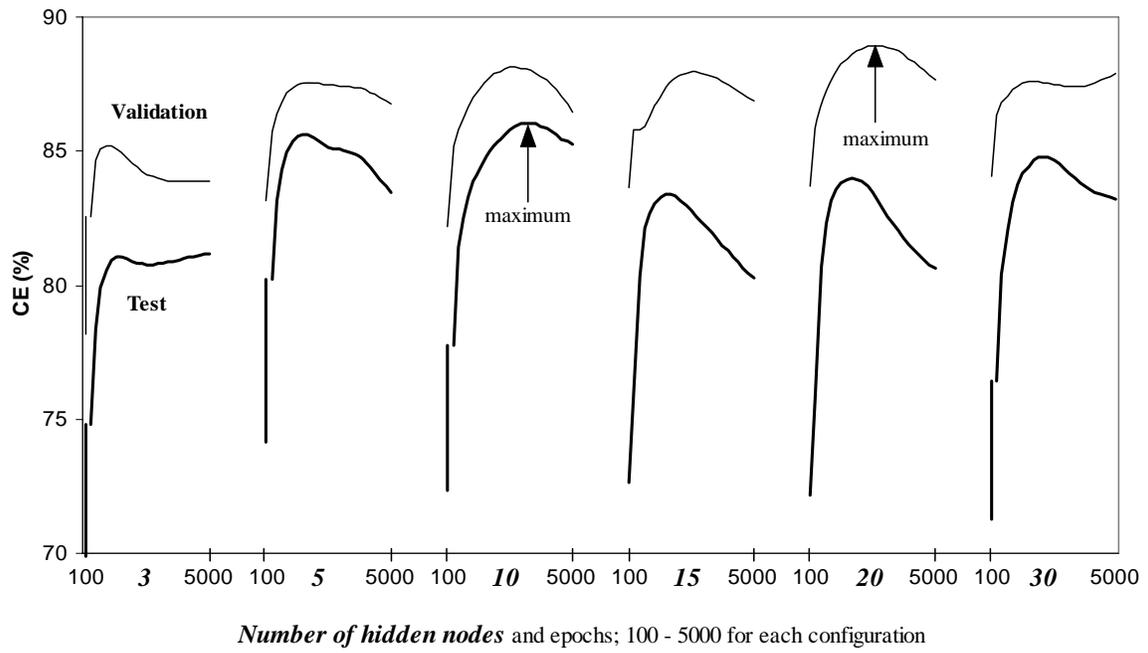
**Table 7** Comparison of cross-sampled ANN models for 20-year flood events with derived values computed on annual maximum series

<b>Catchments</b>	<b>MSRE</b>	<b>CE</b>	<b>SE (cumecs)</b>	<b>MPRE</b>	<b>RB</b>
<i>Rural-FEH</i>	0.9755	80.66	57.23	38.92	-0.0424
<i>Rural-ANN</i>	19.6984	88.47	45.14	137.97	-0.9828
<i>All-Rural-ANN</i>	12.7302	87.82	46.30	130.20	-0.9491
<i>Urban-FEH</i>	1.5264	91.81	26.64	55.13	-0.3548
<i>Urban-ANN</i>	2.7217	84.54	36.87	73.45	-0.0366
<i>All-Urban-ANN</i>	6.6685	90.59	28.60	114.61	0.0462

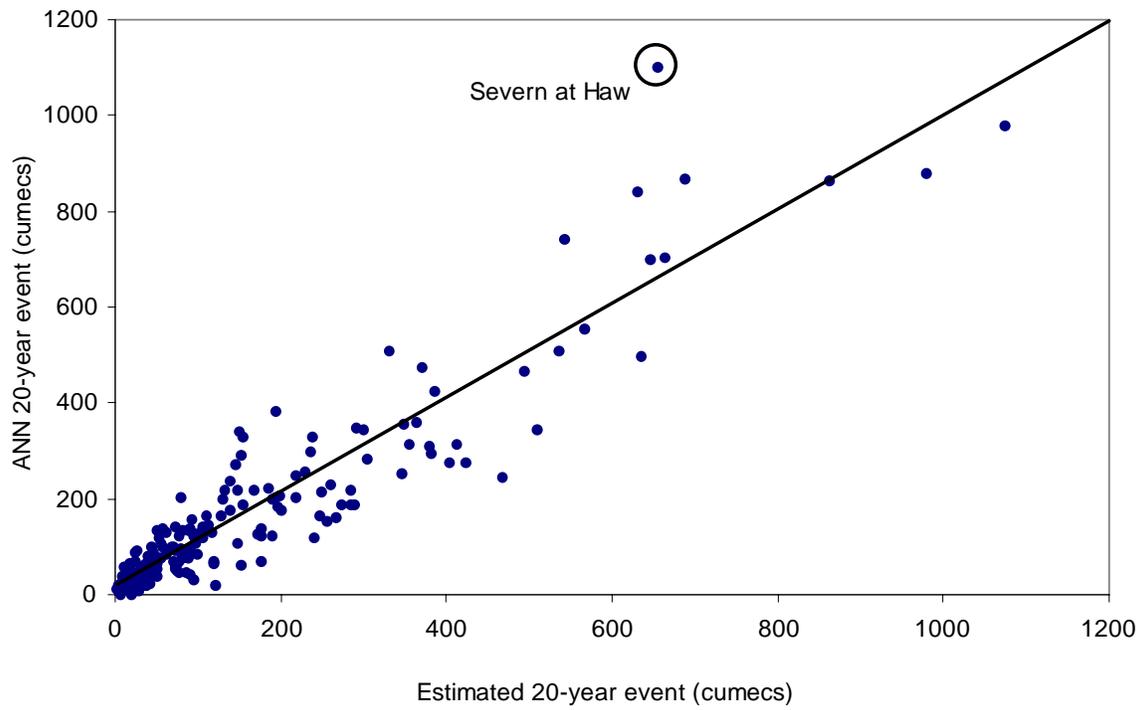
**Table 8** Skill of FEH and ANN models at estimating the index flood in rural and urban catchments



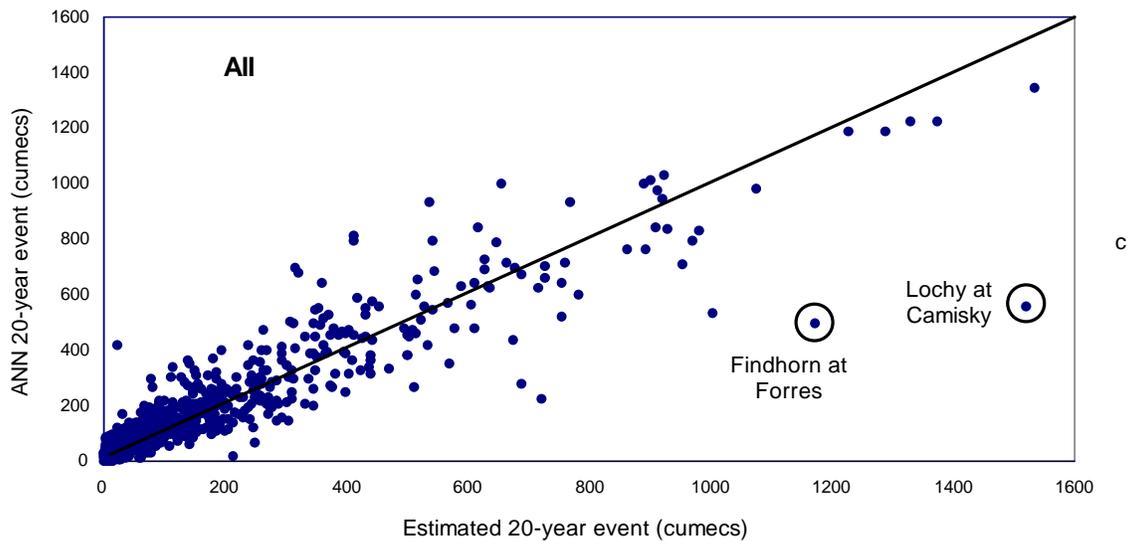
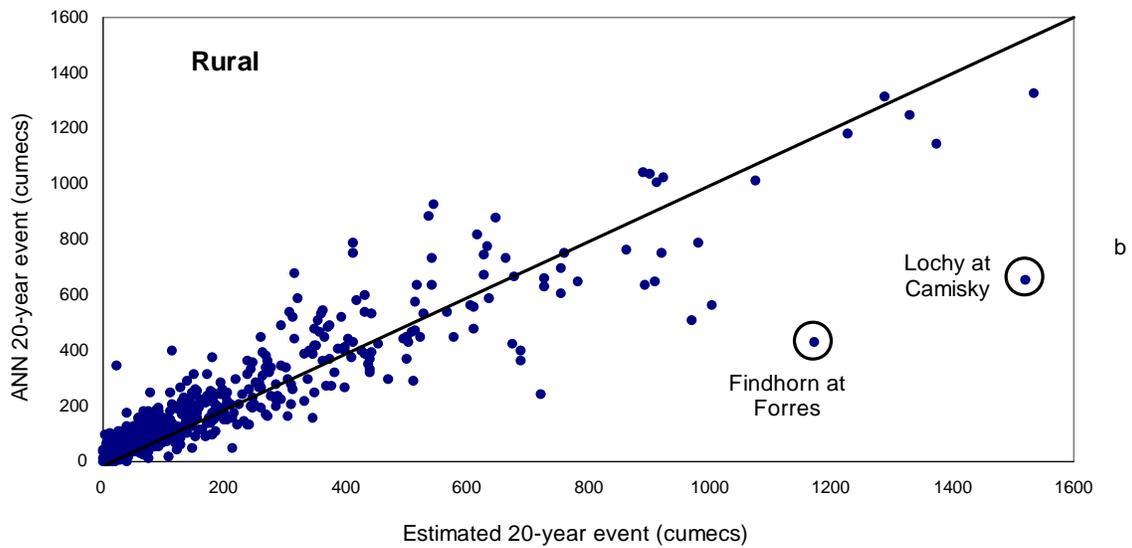
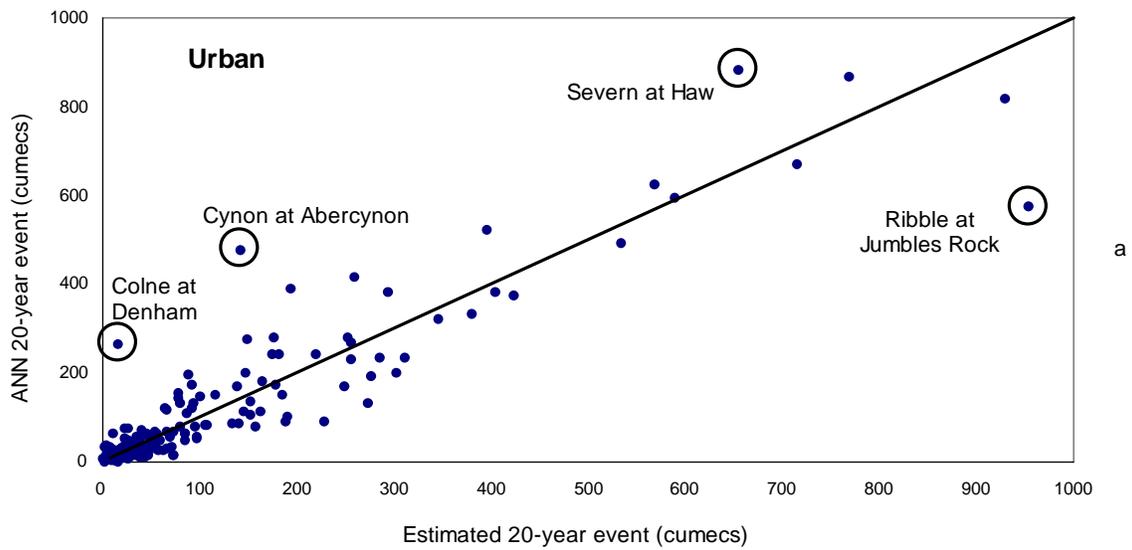
**Figure 1** Multi-layer perceptron



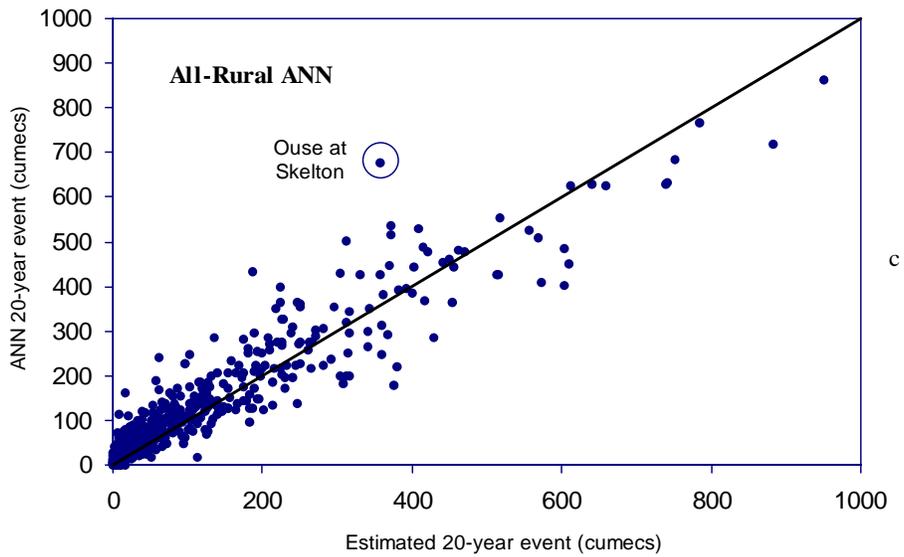
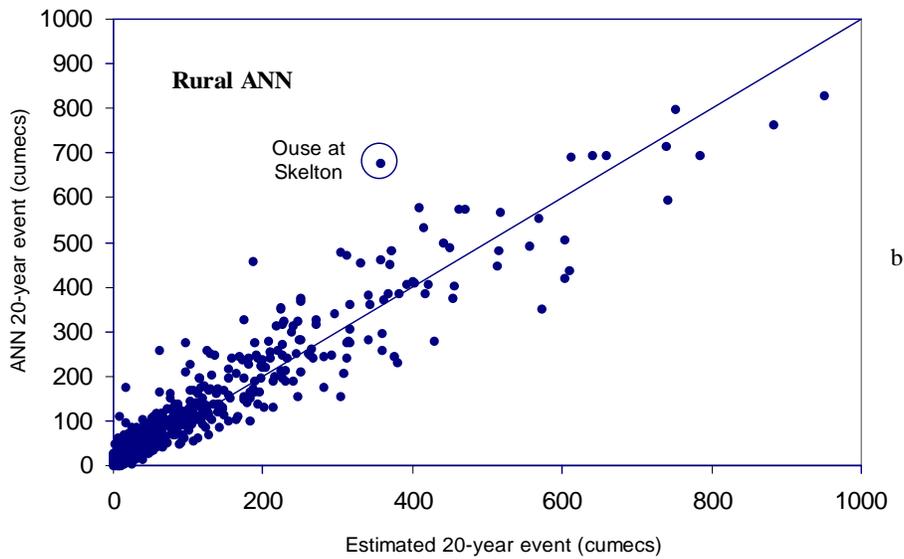
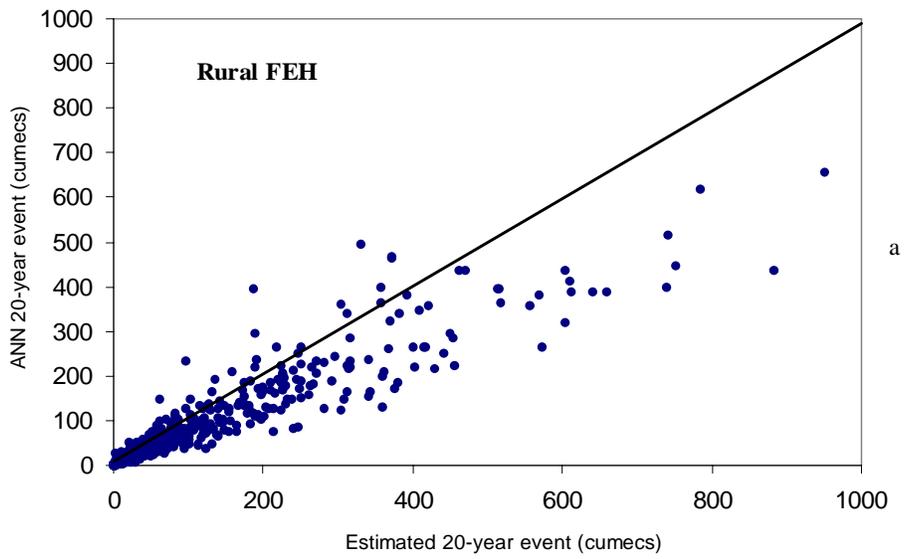
**Figure 2** Comparative performance of different networks during validation and testing of estimated 20-year flood event



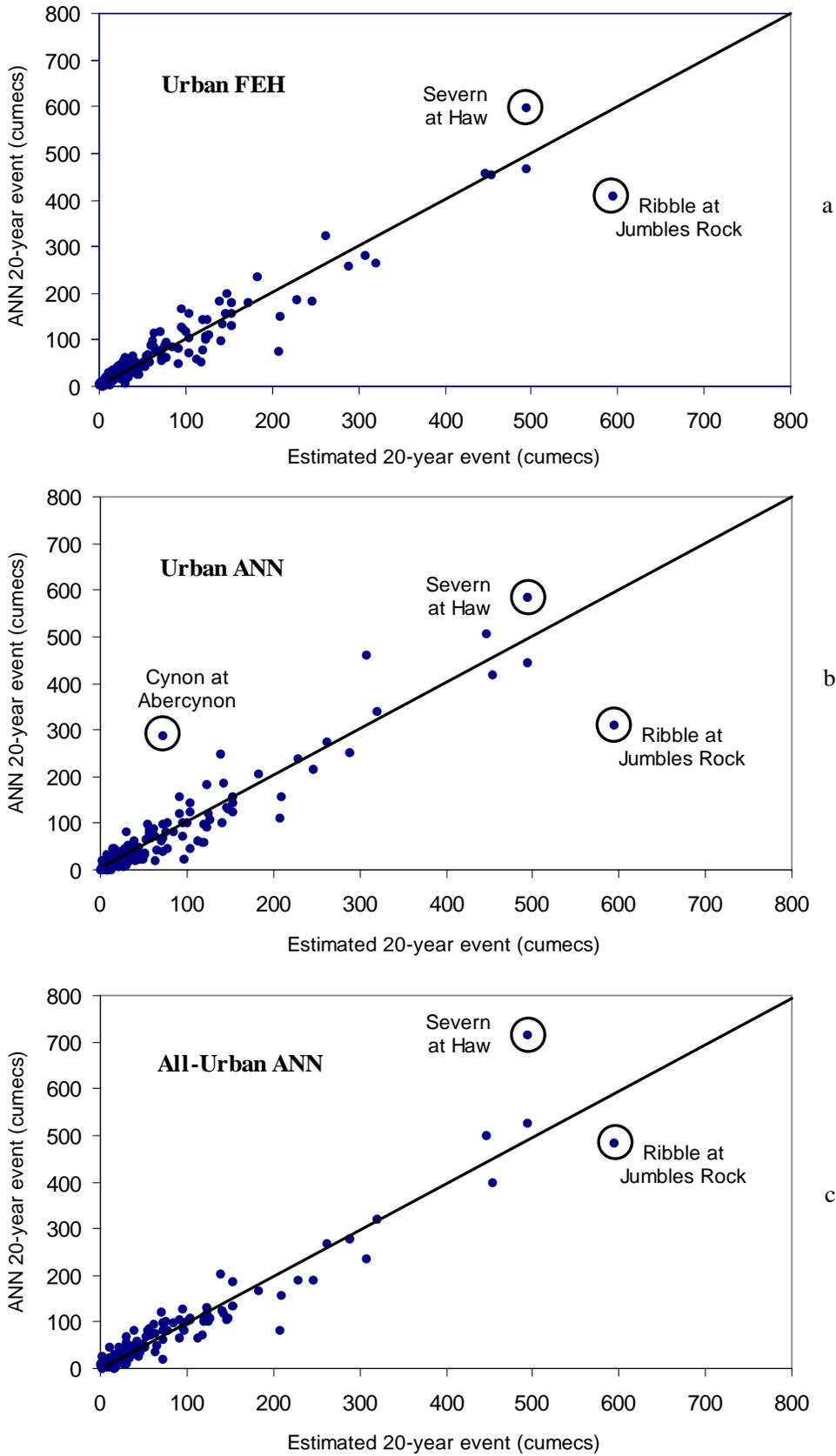
**Figure 3** ANN model of 20-year flood events compared with test data set



**Figure 4** ANN 20-year flood events modelled



**Figure 5** Index flood event models for rural catchments



**Figure 6** Index flood event models for urban catchments