Hydrology and
Earth System
Sciences

Estimation of flood warning runoff thresholds in ungauged basins with asymmetric error functions

Elena Toth

Department DICAM, School of Engineering, University of Bologna, Bologna, Italy

Correspondence to: Elena Toth (elena.toth@unibo.it)

Abstract. In many real-world flood forecasting systems, the runoff thresholds for activating warnings or mitigation measures correspond to the flow peaks with a given return period (often 2 years, which may be associated with the bankfull discharge). At locations where the historical streamflow records are absent or very limited, the threshold can be estimated with regionally derived empirical relationships between catchment descriptors and the desired flood quantile. Whatever the function form, such models are generally parameterised by minimising the mean square error, which assigns equal importance to overprediction or underprediction errors.

Considering that the consequences of an overestimated warning threshold (leading to the risk of missing alarms) generally have a much lower level of acceptance than those of an underestimated threshold (leading to the issuance of false alarms), the present work proposes to parameterise the regression model through an asymmetric error function, which penalises the overpredictions more.

The estimates by models (feedforward neural networks) with increasing degree of asymmetry are compared with those of a traditional, symmetrically trained network, in a rigorous cross-validation experiment referred to a database of catchments covering the country of Italy. The analysis shows that the use of the asymmetric error function can substantially reduce the number and extent of overestimation errors, if compared to the use of the traditional square errors. Of course such reduction is at the expense of increasing underestimation errors, but the overall accurateness is still acceptable and the results illustrate the potential value of choosing an asymmetric error function when the consequences of missed alarms are more severe than those of false alarms.

1 Introduction

In the operation of flood forecasting systems, it is necessary to determine the values of threshold runoff that trigger the issuance of flood watches and warnings. Such critical values might be used for threshold-based flood alert based on real-time data measurements along the rivers (WMO, 2011) or for identifying in advance, through a rainfall-runoff modelling chain, the rainfall quantities that will lead to surpass such streamflow levels, as in the Flash Flood Guidance Systems framework (Carpenter et al., 1999; Ntelekos et al., 2006; Reed et al., 2007; Norbiato et al., 2009).

A runoff threshold should correspond to a *flooding flow*, which is at a value that may produce flood damage and is very difficult to determine on a regional or national scale: it may be defined as a flow that just exceeds bankfull conditions, but in practice, both in gauged and in ungauged river sections, such conditions are arduous to quantify due to the lack of local information (Reed et al., 2007; Hapuarachchi et al., 2011).

In the absence of more sophisticated physically based approaches, based on detailed information of each specific cross section that is rarely available due to limited field surveys, the literature suggests to estimate the bankfull flow as the flood having a 1.5- to 2-year return period (Carpenter et al., 1999; Reed et al., 2007; Harman et al., 2008; Wilkerson, 2008; Hapuarachchi et al., 2011; Cunha et al., 2011; Ward et al., 2013) and a flow that is slightly higher than bankfull may be identified with the 2-year return period flood (Carpenter et al., 1999; Reed et al., 2007).

Many operational systems all around the world adopt a statistically based definition of the flooding flow and the flows associated with given return periods are used as threshold stages for activating flood warning procedures.

The 2-year recurrence is used by many river forecast services in the United States, as suggested by Carpenter et al. (1999), also due to the fact that "the good national coverage of the 2-yr return period flows that the U.S. Geological Survey (USGS) maintains nationwide supports its use" (Ntelekos et al., 2006), as well as in British Columbia (Canada).

However, the floods with different annual exceedance probabilities, associated with different levels of risk, are also frequently adopted in operational real-time flood warning systems: for example in the Czech Republic, flood watch usually corresponds to a 1- to 5-year flow return period (Daňhelka and Vlasák, 2013). In Italy, where a national directive issued in 2004 introduces a system articulated on at least two levels of flow thresholds, many regions have identified the alert levels as flood quantiles with return periods of 2, 5 or 10 years (e.g. the Abruzzo, Lombardia, Puglia Regions). In southern France, the AIGA (Adaptation d'Information Géographique pour l'Alerte) flood warning system compares real-time peak discharge estimated along the river network (on the basis of rainfall field estimates and forecasts) to flood frequency estimates of given return periods (with three categories: yellow for values ranging from 2- to 10-year floods, orange for between the 10 and the 50-year floods and red for peaks exceeding the 50-year flood) in order to provide warnings to the national and regional flood forecasting offices (Javelle et al., 2014).

For river sections where the streamflow gauges are newly installed or where historical rating curves are not available, the observations of the annual maxima are absent or very limited and it is not possible to obtain a reliable estimate of flood quantiles on the basis of statistical analyses of series of observed flood peak discharges.

For these ungauged or poorly gaged basins, the peak flow of a given frequency to be associated with the watch/warning threshold can be estimated transferring information from data-rich sites to data-poor ones, as it is done in the corpus of methodologies applied in RFFA (Regional Flood Frequency Analysis) at ungauged sites, which have always received considerable attention in the hydrologic literature (Bloeschl et al., 2013). Among the possible approaches (statistical and process based) to predict floods in ungauged basins, many researchers have traditionally applied regression-like regionalisation methods for (i) the estimation of the index flood (Darlymple, 1960), usually defined as either the mean or the median (that is the 2-year return period quantile) of the annual maximum flood series, or for (ii) the direct estimate of other quantiles of annual maxima in ungauged basins (Stedinger and Lu, 1995; Salinas et al., 2013). Such methods are based on the assumption that there is a relationship between catchment properties and the flood frequency statistics and are implemented through a regression-type model that relates the flood quantile or the index flood to a number of relevant morpho-climatic indexes. Linear or power (often linearised through a log-transformation) forms, with either a multiplicative or additive error term, are the most commonly used functions (see e.g. Stedinger and Tasker, 1985; GRE-HYS, 1996; Pandey and Nguyen, 1999; Brath et al., 2001; Kjeldsen et al., 2001, 2014; Bocchiola et al., 2003; Merz and Bloeschl, 2005; Griffis and Stedinger, 2007; Archfield et al., 2013; Smith et al., 2015).

In order to allow for more flexibility to the model structure (whose true form is of course not known), the international literature has recently proposed methods based on the use of artificial neural networks (ANNs), providing a nonlinear relationship between the input and output variables without having to define its functional form a priori. Successful applications of ANNs for the estimation of index floods or flood quantiles at ungauged sites are reported in Muttiah et al. (1997), Hall et al. (2002), Dawson et al. (2006), Shu and Burn (2004), Shu and Ouarda (2008), Singh et al. (2010), Simor et al. (2012) and Aziz et al. (2013).

Both the traditional power form or linear regression methods and the neural networks models are generally parameterised by minimising the mean or root mean of the squared errors, i.e. a symmetric function assigning the same importance to overestimation and underestimation errors.

Nevertheless, the consequences of under or overestimating the runoff threshold when used for early warning are extremely different.

Adopting a watch threshold that is higher than the runoff/stage that actually produces flooding damages would in fact lead to missing such events, failing to issue an alarm. Underestimating the runoff threshold may instead determine the issue of false alarms.

False alarms may certainly lead to money losses and also "undermine the credibility of the warning organisation but are generally much less costly than an unwarned event" (UCAR, 2010): in fact the costs of failing to issue an alarm grow rapidly in a real emergency, since a totally missed event has strongly adverse effects on preparedness. The costs of false warnings not only are commonly much smaller than the avoidable losses of a flood, but also cannot match up to indirect and/or intangible flood damages such as loss of lives or serious injuries (Pappenberger et al., 2008; Verkade and Werner, 2011).

Furthermore, regarding the effects of false alarms, "in opposition to 'cry wolf' effect, for some they may provide an opportunity to check procedures and raise awareness, much like a fire practice drill." (Sene, 2013)

Overall, false alarms have usually a higher level of acceptance than misses and this entails that the estimate of flood warning thresholds should be cautionary, so as to conservatively reduce the number of missed alarms.

For the development of watches and warnings it is therefore important to obtain estimates as accurate as possible, minimising both positive and negative errors, but considering that an error will always be present; it is better underpredicting rather than overpredicting the threshold estimate, for safety reasons.

To obtain a conservative estimate of the thresholds, penalizing more the predictions that exceed the "observed" values (in the present case represented by the quantile estimate based on the statistical analysis of measured flow peaks) than those that underestimate them, in the present work it is proposed, for the first time to the Author's knowledge, a parameterisation algorithm that weights asymmetrically the positive or negative errors, in order to decrease the consistency of overestimation and therefore the risk of missing a flood occurrence.

It is important to underline that the proposed asymmetric error function is here applied for optimising a neural network model for predicting the 2-year return period flood (due to its association with the bankfull conditions) but it might be used to improve any other kind of methodology for the estimate of flood warning thresholds associated with any return period.

Section 2 presents the asymmetric error functions; Sect. 3 describes the information available in a database covering the entire country of Italy and the identification of the subsets to be used for a rigorous cross-validation approach. Section 4 presents the implementation of the models for estimating the 2-year return period flood in ungauged catchments, consisting of artificial neural networks calibrated using respectively the symmetric square error and the asymmetric error functions. The results are presented and then discussed in Sect. 5 and Sect. 6 concludes.

2 The asymmetric error function

The scientific literature on forecasting applications, in any scientific area, adopts almost exclusively an objective function based on the sum or mean of the squared discrepancies, i.e. a symmetric quadratic function, due to the well-established good statistical properties of the minimum mean square error estimator.

On the other hand, in economics as well as in engineering and many other fields, there are cases where the forecasting problem is inherently non-symmetric and, in the financial forecasting literature, the use of mean squared error, even if still widely applied, is nowadays not always accepted.

Error (or loss) functions devised to keep account of an asymmetric behaviour have been proposed, such as the linear exponential, the double linear and the double quadratic (Christoffersen and Diebold, 1996; Diebold and Lopez, 1996; Granger, 1999; Granger and Pesaran, 2000; Elliot et al., 2005; Patton and Timmerman, 2006). In particular, Elliot et al. (2005) recently presented a family of parsimoniously parameterised error functions that nests mean squared error loss as a special case (Patton and Timmerman, 2006).

Such function, adapted from Elliot et al. (2005) and defining the error ε as the prediction minus the observed value (i.e. a negative error corresponds to underestimation, a positive one to overestimation), reads as follows:
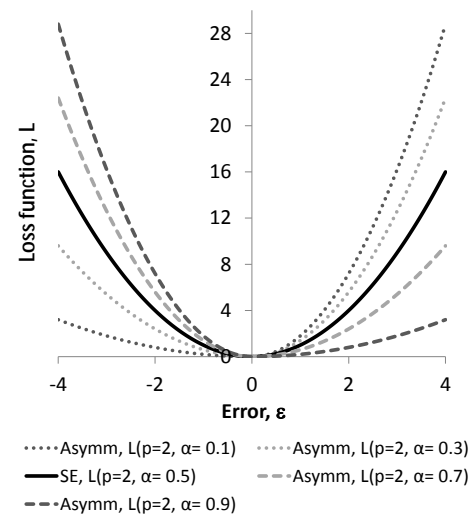
Figure 1. Asymmetric quad–quad loss function (with α varying from 0.1 to 0.9) compared with the squared error (SE).

$$L(p, \alpha) = 2 \cdot [\alpha + (1 - 2\alpha) \cdot 1\{\varepsilon > 0\}] \cdot |\varepsilon|^p, \qquad (1)$$

where $1(\cdot)$ is a unit indicator, equal to 1 when $\varepsilon > 0$ and zero otherwise; p is a positive integer that amplifies the larger errors (corresponding to a quadratic error when equal to 2) and $\alpha \in (0, 1)$ is a parameter representing the degree of asymmetry.

For $\alpha < 0.5$ the function penalises more the overestimation errors ($\varepsilon > 0$), while for $\alpha > 0.5$ more weight is given to negative forecast errors (under-predictions); for $\alpha = 0.5$ the loss weights symmetrically positive and negative errors.

When $p = 2$ and $\alpha \neq 0.5$, the error becomes the asymmetric double quadratic (quad–quad) loss function (see Christoffersen and Diebold, 1996), which is used in the present work for a fair comparison with the traditional mean square error estimator. When $p = 2$ and $\alpha = 0.5$, Eq. (1) corresponds in fact to the traditional, symmetric, square error:

$$L(2, 0.5) = \varepsilon^2. \qquad (2)$$

Figure 1 shows the asymmetric quad–quad loss function (with α varying from 0.1 to 0.9) compared with the squared error (SE).

In the water engineering field, the asymmetric Elliot error function with quadratic amplification ($p = 2$) has been recently applied to parameterise a model for estimating the expected maximum scour at bridge piers, in order to obtain safer design predictions through the reduction of underestimation errors by Toth (2015).

It should be noted that the proposed methodology is a deterministic one, where an optimal point forecast is obtained by minimising the conditional expectation of the future loss; such a framework does not have the advantages of a probabilistic one in terms of quantification of the uncertainties of

the prediction, but it aims to identify the optimal value for the threshold in terms of operational utility.

In Sect. 4, the asymmetric quadratic error function is proposed for optimising the parameters of an input–output model, based on artificial neural networks, between the input variables summarising a set of catchment descriptors (obtainable also for ungauged river sections) and the 2-year return period flood, thus warranting that overestimation errors, which would increase the risk of missing flood warnings, are weighted more than underestimation ones.

3   Available information: the national data set of Italian catchments

The case study refers to a database of almost 300 catchments scattered all over the Italian Peninsula, compiled within the national research project "CUBIST – Characterisation of Ungauged Basins by Integrated uSe of hydrological Techniques" (Claps and the CUBIST Team, 2008).

3.1   Input and output variables

The 12 geomorphological and climatic descriptors are listed in Table 1. The data set unfortunately lacks information on other hydrological properties (e.g. on soils, land-cover, vegetation) and the climatic characterisation is very limited (for example information on extreme rainfall would be extremely important), but the CUBIST set is currently the only database available in the Italian hydrologists community at a national scale.

The data set is described in Di Prinzio et al. (2011), where, following a catchment classification procedure based on multivariate techniques, the descriptors were used to infer regional predictions of mean annual runoff, mean maximum annual flood and flood quantiles through a linear multi-regression model.

As described in such work, in order to reduce the high-dimensionality of the geomorphological and climatic descriptors set, a Principal Components (PC) analysis was applied, obtaining a set of derived uncorrelated variables. The PC variables are as many as the original variables, but they are ordered in such a way that the first component has the greatest variability, the second accounts for the second largest amount of variance in the data and is uncorrelated with the first, and so forth. In the present data set, the first three principal components explain more than three-quarters of the total variance (see Di Prinzio et al., 2011) and these first three PCs are chosen here as input variables to the models described in the following, assuming that they may adequately represent, in a parsimonious manner, the main features of the study catchments.

The database, in addition to the morpho-pluviometric information, includes the annual maxima flow records for periods ranging from 5 to 63 years, whose median values, cor-

Table 1. Geomorphological and climatic descriptors of the CUBIST database of Italian catchments.

1	Long – UTM longitude of catchment centroid
2	Lat – UTM latitude of catchment centroid
3	A – catchment drainage area
4	P – catchment perimeter
5	z_{max} – maximum elevation of the catchment area
6	z_{min} – elevation of the catchment outlet
7	z_{mean} – mean elevation of the catchment area
8	L – length of the maximum drainage path
9	SL – average slope along the maximum drainage path
10	SA – catchment average slope
11	Φ – catchment orientation
12	MAP – mean annual precipitation

responding to the 2-year return period, represent the output variable to be simulated by the models. Only 9 of the 300 locations had less than 8 years of data and therefore, all records were deemed to be sufficient for the purposes of this study.

The data set covers a great diverseness of hydrological, physiographic and climatic properties and in order to partially reduce such heterogeneity, it was decided to limit the analysis to catchments that have a 2-year flood in the range of 10–1000 $m^3 s^{-1}$, i.e. 267 over the original 296 basins.

3.2   Identification of balanced cross-validation subsets with SOM clustering of input data

As will be detailed in Sect. 4, the database is to be divided in three disjoint subsets (called training, cross-validation and test sets) in order to allow for a rigorous independent validation and also to increase the generalisation abilities of the model when encountering records different from those used in the calibration (or *training*) phase, following an *early stopping* parameterisation procedure.

The way in which the data are divided may have a strong influence on the performance of the model and it is important that each one of the three sets contains all representative patterns that are included in the data set. As proposed in the recent literature (Kocjancic and Zupan, 2001; Bowden et al., 2002; Shahin et al., 2004) a self-organising map (SOM) may be applied to this aim. The SOM is a data-driven classification method based on unsupervised artificial neural networks that may be applied for several clustering purposes (for hydrological applications see, for example, Minns and Hall, 2005; Kalteh et al., 2008).

In the recent years, SOMs were also successfully applied for catchments classification either based on geo-morpho-climatic descriptors (Hall and Minns, 1999; Hall et al., 2002; Srinivas et al., 2008; Di Prinzio et al., 2011) or based on hydrological signatures (Chang et al., 2008; Ley et al., 2011; Toth, 2013); however, it is important to underline that the clustering is not carried out here in order to identify a pooling group of similar catchments for developing a region-
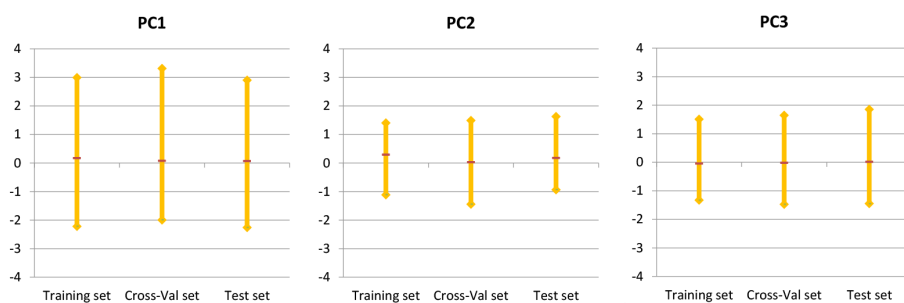
Figure 2. Mean value (red dash) and the bars comprised between the 90th and 10th percentiles of the resulting training, cross-validation and testing sets for each of the three input variable (PC1, PC2 and PC3).

specific model, but for the optimal division of the available data for the parameterisation and independent testing of a single model to be applied over the entire study area.

The SOM is in fact used to cluster similar data records together: an equal number of data records is then sampled from each cluster, ensuring that records from each class (that is catchments with different features) are represented in the training, validation and test sets, which, as a result, have similar statistical properties (Bowden et al., 2002; Shahin et al., 2004).

A SOM (Kohonen, 1997) organises input data through non-linear techniques depending on their similarity. It is formed by two layers: the input layer contains one node (neuron) for each variable in the data set. The output-layer nodes are connected to every input through adjustable weights, whose values are identified with an iterative training procedure. The relation is of the competitive type, matching each input vector with only one neuron in the output layer, through the comparison of the presented input pattern with each of the SOM neuron weight vectors, on the basis of a distance measure (here the Euclidean one). In the trained (calibrated) SOM, all input vectors that activate the same output node belong to the same class.

In the present application, the dimension of the input layer is equal to 3 (that is, the first three principal components of the catchments descriptors); as far as the output layer is concerned, there is not a predefined number of classes; a parsimonious output was chosen that is formed by three nodes in a row, each one corresponding to a call, to ensure the resulting sets were not too dissimilar.

The three resulting clusters are respectively formed by 121, 70 and 76 catchments; each cluster is then divided into three parts, and one-third is assigned to the training, validation and test sets respectively. Overall, the training, validation and test sets are therefore equally numerous (91, 88 and 88 records respectively) and formed by the same proportion of catchments belonging to each of the clusters, having eventually a similar information content, as shown by the similar statistics of the three variables in the three sets represented in Fig. 2.

4 Development of symmetric and asymmetric artificial neural networks models for estimating the 2-year return period flows at ungauged sites

4.1 Feedforward artificial neural networks

Artificial neural networks are massively parallel and distributed information processing systems, composed of nodes, arranged in layers, which are able to infer a non-linear input–output relationship. ANN, in particular feedforward networks, have been widely used in many hydrological applications (see for example the recent review papers by Maier et al., 2010 and by Abrahart et al., 2012) and the readers may refer to the abundant literature for details on their characteristics and implementation.

Three different layer types can be distinguished: input layer (connecting the input information), one or more hidden layers (for intermediate computations) and an output layer (producing the final output); adjacent layers are connected through multiplicative weights and, in each node, the sum of weighted inputs and a threshold (called bias) is passed through a non-linear function known as an activation.

The models applied here are networks formed by one hidden layer, with tan-sigmoid activation functions, and a single output node (corresponding to the estimated flood with 2-year return period), with a linear activation function.

The identification of the network's weights and biases (called training procedure) is carried out with a non-linear optimisation, searching the minimum of an error (or learning) function measuring the discrepancy between predicted and observed values, and feedforward networks are generally trained with a learning algorithm known as backpropagation (Rumelhart et al., 1994), based on the steepest descent or on more efficient quasi-newton methods.

In order to avoid overfitting, which degrades the generalisation ability of the model, the early stopping or optimal stopping procedure was applied (see, for example, Coulibaly et al., 2000). For applying early stopping, the available data have been divided into three disjoint subsets with a similar information content, as described in Sect. 3.2: a training set, an early stopping validation set and a test set. While the net-

work is parameterised minimising the error function on the training set, the error function on the early stopping validation set is also monitored; if the error function on such second set increases continuously for a specific number of iterations, this is a sign of overfitting of the training set: the training is then stopped and network parameters at the lowest validation error are returned. The third set (test set) is not used in any way during the parameterisation phase, but it is used for out-of-sample, independent evaluation of the resulting models.

4.2 Implementation of the symmetric model

Neural networks, including those applied in the recent hydrological literature for the estimation of index floods or flood quantiles at ungauged sites, are traditionally trained minimising the square error function, which is symmetrical about the y axes and negative or positive discrepancies of the same magnitude result in the same function value.

In the present work, the results obtained by a network trained with a conventional square error function are compared with those obtained when parameterising the network through the minimisation of an asymmetric loss function, which takes into account both over and underestimation discrepancies but penalises more the overprediction errors, since the consequences of missing alarms are more severe than those of false alarms.

For both type of models, the output values (2-year flood values) are rescaled as a function of the overall minimum and maximum values to the $[-0.95, +0.95]$ range, to facilitate the optimisation algorithms and also avoid saturation problems by accommodating possible extreme values occurring outside the range of available data (Dawson and Wilby, 2001). Each implemented architecture is randomly initialised 10 times to help avoiding local optima: the parameter set that results in the minimum error function on the early stopping validation data (second set) is chosen as the trained network.

The first implemented model is obtained through the minimisation of the traditional, symmetric mean squared error, applying the quasi-Newton Levenberg–Marquardt backpropagation algorithm (Hagan and Menhaj, 1994), widely applied and regarded as one of the most efficient neural network training algorithms.

The input variables are the first three principal components of the catchment descriptors, so the input layer is formed by three nodes; the output node corresponds to the estimated flood with 2-year return period; as far as the dimension of the hidden layer is concerned, there is, unfortunately, no definitive established methodology for its determination because the optimal network architecture is highly problem-dependent. Different architectures with a number of hidden nodes varying from 2 to 6 were set up: the mean squared error of the estimates over the third, independent set resulted the minimum one with the model having three hidden nodes.

The architecture with three input nodes, three hidden nodes and 1 output node, represented in Fig. 3, is therefore
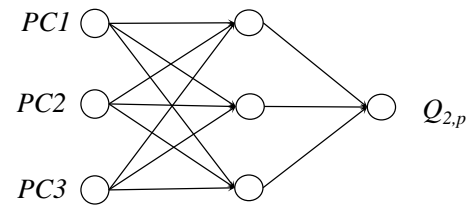
Figure 3. Architecture of the chosen network, with three input nodes, three hidden nodes and 1 output node.

the network finally chosen; the network parameterised minimising the symmetric mean square error function will be denoted as ANN-Symm, and in Sect. 5 its results will be compared with those of the asymmetric models having the same architecture but parameterised with a different error function.

4.3 Implementation of asymmetric models with varying degree of asymmetry

The quad–quad loss function described in Sect. 2 is here applied for calibrating the network parameters of the asymmetric models. The learning function to be minimised is therefore the average value of the double quadratic errors (mean quad–quad error, MQQE), obtainable averaging the M (number of records in the set) errors given by Eq. (1) when $p = 2$:

$$\text{MQQE} = \frac{2}{M} \sum_{j=1}^{M} [\alpha + (1 - 2\alpha) \cdot 1\{\varepsilon > 0\}] \cdot |\varepsilon_j|^2. \quad (3)$$

The value of α, corresponding to the degree of asymmetry of the loss function, cannot be fixed a priori, since such choice should be based on a location-specific cost-benefit analysis, keeping into account the avoidable losses (i.e. the direct and indirect losses, provided they may be quantifiable, which may be prevented by mitigation actions following an alarm issue) and the cost of the mitigation measures themselves. Such analysis is acknowledged to be extremely difficult, especially since it also involves intangible costs, such as life losses, but also warning credibility issues; furthermore, the costs may change over time and are also dependent on the warning lead time (see e.g. Martina et al., 2006; Verkade and Werner, 2011, Montesarchio et al., 2011).

For this reason, in the present application, different asymmetric networks, with α varying from 0.4 to 0.1, are implemented in order to compare the results obtainable with a different asymmetry degree, which is a different extent of importance given to over/underestimation errors. Such asymmetrically trained networks are in the following denoted as Asymm-0.4, Asymm-0.3, Asymm-0.2 and Asymm-0.1.

The training of the four asymmetric networks, based on the minimisation of the MQQE, is carried out through the generalisation of the backpropagation algorithm proposed by Crone (2002) and applied by Silva et al. (2010), which may

be used for parameterising artificial neural networks with any differentiable (analytically or numerically) error function.

5 Results and discussion

5.1 Goodness-of-fit measures and plots

As described in Sect. 4.2, the neural networks are trained over the standardised (rescaled) output values of the training and cross-validation sets and they are successively used for predicting the output over the independent test set: such ANN output values are then scaled back, obtaining the predictions $Q_{2,\mathrm{p}}$.

The performances of the models are therefore evaluated through a set of indexes that describe the prediction error (ε), which is the difference between predictions ($Q_{2,\mathrm{p}}$) issued by the models (as a function of morpho-climatic attributes only) and the observed 2-year flood values (the median of historical annual maxima; $Q_{2,\mathrm{o}}$) on the third set (test set), formed by $N = 91$ catchments distributed all over the country, whose data have not been used in any capacity in the models' development.

The following error statistics have been computed:
MAE (mean absolute error)

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |\varepsilon(i)|}{N}, \tag{4}$$

RMSE (root mean square error)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\varepsilon(i))^2}{N}}. \tag{5}$$

MAE and RMSE both represent a symmetric accuracy, corresponding to the distance of the predictions from the observations independently of the error sign (and the RMSE, being quadratic, emphasises more the larger errors).

In order to keep into account the differences in sign of the errors, representing the extent of overpredictions as compared to underpredictions, the overall percentage of positive errors (Over %) is computed:
Over % (percentage of overestimates)

$$\mathrm{Over\%} = \frac{\left\{i = 1, \ldots, N \,|\, Q_{2,\mathrm{p}}(i) > Q_{2,\mathrm{o}}(i)\right\}}{N}. \tag{6}$$

Such a metric shows the general tendency of the model to overestimate (or to underestimate, since $100 - \mathrm{Over\,\%}$ represents, conversely, the proportion of underpredictions), but these indexes do not distinguish among errors of different magnitude, since they also count predictions that may be only barely above (or below) the targets, i.e. very good predictions, with minimum errors.
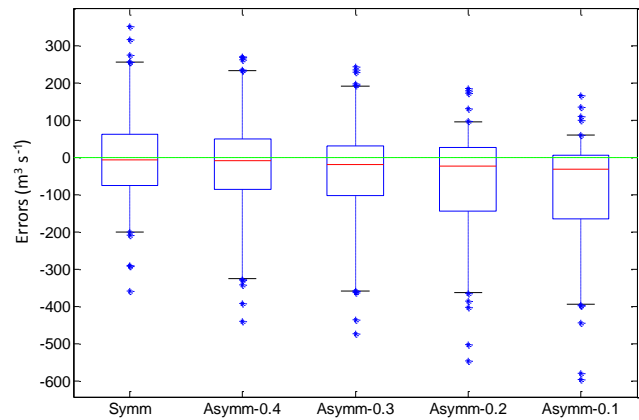
Figure 4. Parallel box plots of the errors ($\varepsilon = Q_{2,\mathrm{p}} - Q_{2,\mathrm{o}}$) of the 2-year floods estimates obtained by symmetric and asymmetric networks on the independent test set of catchments. The bottoms and tops of the rectangular boxes are respectively the lower and the upper quartiles, the horizontal segment inside the box is the median and the whiskers represent the 5th and 95th percentiles.

Therefore, the extent of overestimation is also evaluated through the number of high errors, keeping into account only the more relevant, and therefore potentially more dangerous, overpredictions. An estimate that is more than 30 % higher than the corresponding target value was considered here as high overprediction:
OverH % (percentage of high overprediction errors)

$$\mathrm{OverH\%} = \frac{\left\{i = 1, \ldots, N \,|\, Q_{2,\mathrm{p}}(i) > 1.3 \cdot Q_{2,\mathrm{o}}(i)\right\}}{N}. \tag{7}$$

The more conservative the threshold estimate, the lower the value of OverH %.

On the other hand, even if – as discussed – generally less crucial in terms of consequences, the number of high underestimation errors should also be monitored, since excessively low values imply the tendency of the model to establish thresholds leading to the issuance of too many false alarms.
UnderH % (percentage of high underprediction errors):

$$\mathrm{UnderH\%} = \frac{\left\{i = 1, \ldots, N \,|\, Q_{2,\mathrm{p}}(i) < 0.7 \cdot Q_{2,\mathrm{o}}(i)\right\}}{N}. \tag{8}$$

In addition to the goodness-of-fit measures (reported in Table 2), the box plot of the errors (predicted minus observed quantiles) is shown in Fig. 4.

The results may be evaluated also through the scatter plots of predicted (y axis) vs. observed (x axis) quantiles, shown in Fig. 5 that shows every prediction $Q_{2,\mathrm{p}}$ in respect to the corresponding observation $Q_{2,\mathrm{o}}$.

5.2 Discussion of the results

The box plot (Fig. 4) allows to visually assess both the accuracy and the tendency to over/underestimate of the models:
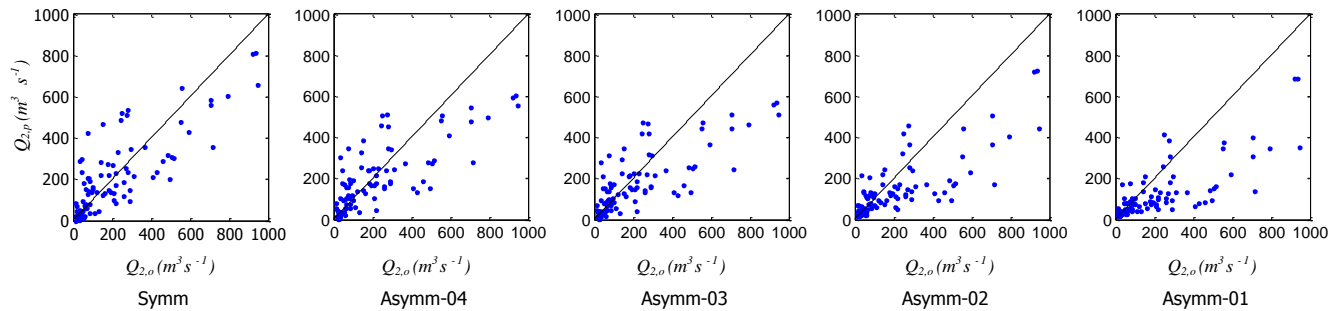
Figure 5. Scatter plots of the predicted (*y* axis) vs. observed (*x* axis) 2-year floods estimates on the independent test set of catchments, for the symmetric and asymmetric models.

the boxes should be compact and close to the dotted line representing zero error but at the same time it is better if the data lie below such a line, thus indicating that the method does not tend to overpredict the thresholds and the warning system is therefore less subject to miss a potentially dangerous flood.

It may be seen in Fig. 4 that for the network that was trained minimising the traditional square error (ANN-Symm), the box and whiskers are centred on the zero-error line and the quantiles (top/bottom of the box, top/bottom whiskers) are at a similar distance from such a line, showing that the errors are equally distributed among overestimation and underestimations. The box is compact, demonstrating the good accurateness of the method for a substantial part of the test set, but, due to the symmetric disposition of the errors, many overestimation errors, also remarkably high, are issued, as shown by the position of the upper whisker.

Analysing Table 2, the relatively good accuracy of the ANN-Symm model is demonstrated by the values of the MAE and RMSE, which are the lowest among the implemented models. The symmetric distribution of the overall errors is shown by an Over % close to 50 % and the similar values of the OverH % (34 %) and UnderH % (32 %) confirm that also the high relative errors are equally split among over and underestimates.

Such results were expected since the training is based on a symmetric loss function, but the consequence is that the ANN-Symm model issues a remarkable number of significant overprediction errors, in fact for about one-third of the test catchments, the estimates are more than 30 % higher than the observations.

The analysis of Table 2 shows that the asymmetrically trained networks tend, for decreasing α values, to reduce the number of overestimations (positive errors). For the overall errors this is shown by the different proportion of over-/underestimations, which moves from a value that corresponds, approximately, to a balance, to a much more skewed distribution of positive vs. negative errors, with Over % decreasing up to 31 %.

At the same time, and more importantly, the number of positive (overestimation) errors larger than 30 % substan-

Table 2. Goodness-of-fit criteria of the 2-year floods estimates obtained by the symmetric and asymmetric networks on the independent test set of catchments.

Model/ index	MAE ($m^3 s^{-1}$)	RMSE ($m^3 s^{-1}$)	Over %	OverH %	UnderH %
Symm	98	133	46 %	34 %	32 %
Asymm-04	104	147	42 %	32 %	35 %
Asymm-03	105	152	41 %	30 %	37 %
Asymm-02	108	162	36 %	27 %	41 %
Asymm-01	115	178	31 %	18 %	47 %

tially decreases with α, with OverH % reaching a value that is much lower than that of the ANN-Symm model when α arrives at 0.1 (18 % vs. 34 %).

Conversely, as expected, the more asymmetric the network, the higher the underprediction errors, as shown by the values of UnderH %: the number of significant negative errors gradually increases from one-third up to 47 % of the total.

Also the accuracy (given by the total amount of the discrepancies independently of their sign) deteriorates when the asymmetry is more pronounced, but the drop is moderate and the RMSE and MAE values are not so far from those of the ANN-Symm network.

Looking at the parallel box plots (Fig. 4), it may be seen that the boxes become less compact and, as expected, their position shifts downwards with increasing asymmetry. The length of the upper whiskers substantially decrease with α but the length of the lower whiskers does not increase at the same rate, thus compensating for the fact that the boxes are taller for the more asymmetric models. It follows that the global distances from 95th and 5th percentiles (given by the distance between the ends of the top and bottom whiskers) are very close for the symmetric (ANN-Symm) and for the two most asymmetric networks, thus showing that the variability of the errors for the vast majority (middle 90 %) of the data is similar. On the other hand, overall, the errors are moving towards the underestimation side for increasing asymmetry (as confirmed also by the corresponding median values)

and for Asymm-01, the upper part of the box indicates that only about one-quarter of the errors are overestimations.

It may be noted, in particular from the scatter plots (Fig. 5), that, for both symmetric and asymmetric models, the errors are not negligible: this is due to the shortcomings of the available data set but mainly to the intrinsic limitations of a regional approach applied to the extreme variability of the study area. As already underlined in Sect. 3.1, the national data set lacks important information that may help to characterise the hydrological behaviour and the phenomena governing formation of extreme flows. On top of the unavoidable risk of erroneous data, the absence in the database of additional influences certainly further hampers the possibility to obtain a reliable relationship with the flood quantiles. Most importantly, the data set covers the entire Italian Peninsula, characterised by extremely different hydro-climatic settings (from Alpine to Mediterranean ones) and this high heterogeneity is certainly an additional reason that limits the performance.

Notwithstanding the limitations of the data set, which equally affect all the proposed models, the results demonstrate that the use of the double quadratic error function, even if at the expense of more substantial underestimation errors, can substantially decrease the number and extent of overestimation errors, if compared to the use of the traditional square errors.

In the application to a specific cross section, the degree of asymmetry might be identified as proportional to the risk averseness of the situation: when the impact of false alarms is, at least comparatively, small, the decision-makers are reluctant to the consequences (economic and social) of a flood and, rather than risking a missed alarm, can accept many cases of false alarm with the associated costs.

6 Conclusions

A crucial issue in the operation of flood forecasting/warning systems at ungauged locations is how to assess the possible impacts of the forecasted flows, i.e. the identification of streamflow values that may actually cause flooding, to be associated with thresholds that trigger the issuance of flood watches and warnings. The values that may produce damaging conditions (or flooding flows), when in absence of detailed local information on each cross section, are in many parts of the world estimated as the peak floods having a certain return period, often 2 years, which is generally associated with the bankfull discharge.

For locations where the gauges are new or where historical rating curves are not available, the series of past annual flow maxima are absent or very limited, and the peak flow of given frequency to be associated with the watch/warning threshold can be estimated with regionally derived empirical relationships, such as those that may be applied for the estimation of the index flood at ungauged sites. Such regression-like methods consist in a relation between a set of catchment descriptors that may also be obtained for ungauged sites and the desired flood quantile; linear or power forms are the most commonly used functions, but recent studies have successfully applied artificial neural network models, due to their flexibility, to flood quantile and index flood estimation.

Whatever the function form, such models are generally parameterised by minimising the mean square error, which assigns equal weight to overprediction or underprediction errors, whereas, instead, the consequences of such errors are extremely different when the estimates are to be used as warning threshold. In fact, false alarms (due to an underprediction of the warning threshold) generally have a much higher level of acceptance than misses (which would derive from an overestimated threshold).

For this reason, in the present work, the regression model (a feedforward neural network) is parameterised minimising an asymmetric error function (of the double quadratic type) that penalises more the overestimation than the underestimation discrepancies. The predictions of models with increasing degree of asymmetry are compared with those of a traditional (trained on the symmetric mean of square errors) neural network, in a rigorous cross-validation experiment referred to as a database of catchments covering the entire country of Italy.

The results confirm, as expected, that the more asymmetric the network, the more numerous and higher the underprediction errors, and the less numerous and less severe the overestimation errors. As also expectable, the symmetric accuracy decreases when the asymmetry is more pronounced, but the drop is moderate and the RMSE and MAE values are not so far from those of the traditionally trained network.

Undoubtedly, the nature of the regional approach, as well as the shortcomings of the data set and the extreme heterogeneity of the study area, generate errors much greater than those obtainable with detailed local studies. On the other hand, where no alternatives exist, the proposed methodology may provide a preliminary estimate of the threshold runoff that do not overestimate the actual *flooding flow*.

Notwithstanding the acknowledged limitations of the data set, which affect equally all the proposed models, the analysis shows that the use of the asymmetric error function substantially reduces the number and extent of overestimation errors, if compared to the use of the traditional square errors. Of course such reduction is at the expense of increasing underestimation errors, but the overall precision is still acceptable and the study highlights the potential benefit of choosing an asymmetric error function when the consequences of missed alarms are more severe than those of false alarms.

Minimising the asymmetric error function has the purpose of optimising the threshold from an operational point of view, in a deterministic framework: future analyses may be devoted to investigate the uncertainty of the issued predictions, since a probabilistic approach (provided that the methodology is able to include all sources of uncertainty and its quality may be objectively assessed) may provide very valuable insights

for a more complete evaluation of the model, supplementing the information provided by point-value predictions.

It is important to highlight that the asymmetric error function is used, in this study, to parameterise a neural network, but of course it might be used to optimise any other model or equation, when aiming to obtain conservative estimates, for safety reasons.

The appropriate degree of asymmetry might be identified depending on the risk averseness of the specific flood-prone context. The quantification of risk aversion is extremely difficult and case specific and it should keep into account that the perception of society may be very different from a technical appraisal of the involved costs. In addition, it should also include indirect, intangible and long-term impacts. More research on the societal perception in different contexts would greatly improve the process of risk-based decision-making (Merz et al., 2009), including the choices concerning flood-warning thresholds. Hopefully, in the next years, a more direct collaboration between the hydrologic and socio-economic research communities, as advocated in the new Panta Rhei science initiative (Montanari et al., 2013; Javelle et al., 2014), in particular with regard to data-driven modelling (Mount et al., 2016), will provide a progress in this direction.

References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., and Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting, Prog. Phys. Geogr., 36, 480–513, doi:10.1177/0309133312444943, 2012.

Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., and Kiang, J. E.: Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach?, Hydrol. Earth Syst. Sci., 17, 1575–1588, doi:10.5194/hess-17-1575-2013, 2013.

Aziz, K., Rahman, A., Fang, G., and Shreshtha, S.: Application of Artificial Neural Networks in Regional Flood Frequency Analysis: A Case Study for Australia, Stoch. Environ. Res. Risk A., 28, 541–554, doi:10.1007/s00477-013-0771-5, 2013.

Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. (Eds): Runff prediction in ungauged basins: Synthesis across processes, places and scales, Cambridge University Press, New York, USA, 490 pp., 2013.

Bocchiola, D., De Michele, C., and Rosso, R.: Review of recent advances in index flood estimation, Hydrol. Earth Syst. Sci., 7, 283–296, doi:10.5194/hess-7-283-2003, 2003.

Bowden, G. J., Maier, H. R., and Dandy, G. C.: Optimal division of data for neural network models in water resources applications, Water Resour. Res., 38, 1010, doi:10.1029/2001WR000266, 2002.

Brath, A., Castellarin, A., Franchini, M., and Galeati, G.: Estimating the index flood using indirect methods, Hydrolog. Sci. J., 46, 399–418, 2001.

Carpenter, T. M., Sperfslage, J. A., Georgakakos, K. P., Sweeney, T., and Fread, D. L.: National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems, J. Hydrol., 224, 21–44, 1999.

Chang, F. J., Tsai, M. J., Tsai, W. P., and Herricks, E. E.: Assessing the Ecological Hydrology of Natural Flow Conditions in Taiwan, J. Hydrol., 354, 75–89, 2008.

Christoffersen, P. F. and Diebold, F. X.: Further results on forecasting and model selection under asymmetric loss, J. Appl. Econ., 11, 561–571, 1996.

Claps and the CUBIST Team: Development of an Information System of the Italian basins for the CUBIST project, Geophys. Res. Abstr., 10, EGU2008-A-12048, 2008.

Coulibaly, P., Anctil, F., and Bobee, B.: Daily reservoir inflow forecasting using artificial neural networks with stopped Training Approach, J. Hydrol., 230, 244–257, 2000.

Crone, S.F.: Training Artificial Neural Networks using Asymmetric Cost Functions, in: Vol. 5, IEEE Proceedings of the 9th International Conference on Neural Infomation Processing (ICONIP'OZ), 18–22 November 2002, Singapore, 2374–2380, 2002.

Cunha, L. K., Krajewski, W. F., and Mantilla, R.: A framework for flood risk assessment under nonstationary conditions or in the absence of historical data, J. Flood Risk Manage., 4, 3–22, 2011.

Dalrymple, T.: Flood frequency analyses, Water Supply Paper 1543-A, US Geological Survey, Reston, Virginia, USA, 80 pp., 1960.

Daňhelka, J. and Vlasák, T: Evaluation of Real-time Flood Forecasts in the Czech Republic, 2002–2012, Czech Hydrometeorological Institute Report, http://www.chmi.cz/files/portal/docs/poboc/CB/pruvodce/vyhodnoceni_en.html (last access: 17 June 2015), 2013.

Dawson, C. W. and Wilby, R.: Hydrological modelling using artificial neural networks, Prog. Phys. Geogr., 25, 80–108, 2001.

Dawson, C. W., Abrahart, R. J., Shamseldin, A. Y., and Wilby, R. L.: Flood estimation at ungauged sites using artificial neural networks, J. Hydrol., 319, 391–409, 2006.

Diebold, F. X. and Lopez, J. A.: Forecast Evaluation and Combination, in: Vol. 14, Handbook of Statistics, edited by: Maddala, G. S. and Rao, C. R., North-Holland, Amsterdam, 241–268, 1996.

Di Prinzio, M., Castellarin, A., and Toth, E.: Data-driven catchment classification: application to the pub problem, Hydrol. Earth Syst. Sci., 15, 1921–1935, doi:10.5194/hess-15-1921-2011, 2011.

Elliott, G., Komunjer, I., and Timmermann, A.: Estimation and Testing of Forecast Rationality under Flexible Loss, Rev. Econ. Stud., 72, 1107–1125, 2005.

Granger, C. W. J.: Outline of Forecast Theory Using Generalized Cost Functions, Spanish Econ. Rev., 1, 161–173, 1999.

Granger, C. W. J. and Pesaran, M. H.: A Decision Theoretic Approach to Forecast Evaluation, in: Statistics and Finance: An Interface, edited by: Chan, W. S., Li, W. K., and Tong, H., Imperial College Press, London, 261–278, 2000.

GREHYS – Groupe de recherche en hydrologie statistique: Presentation and review of some methods for regional ?ood frequency analysis, J. Hydrol., 186, 63–84, 1996.

Griffis, V. W. and Stedinger, J. R.: The use of GLS regression in regional hydrologic analyses, J. Hydrol., 344, 82–95, 2007.

Hagan, M. T. and Menhaj, M. : Training feedforward networks with the Marquardt algorithm, IEEE T. Neural Netw., 5, 989–993, 1994.

Hall, M. J. and Minns, A. W.: The classification of hydrologically homogeneous regions, Hydrolog. Sci. J., 44, 693–704, 1999.

Hall, M. J., Minns, A. W., and Ashrafuzzaman, A. K. M.: The application of data mining techniques for the regionalisation of hydrological variables, Hydrol. Earth Syst. Sci., 6, 685–694, doi:10.5194/hess-6-685-2002, 2002.

Hapuarachchi, H. A. P., Wang, Q. J., and Pagano, T. C.: A review of advances in flash flood forecasting, Hydrol. Process., 25, 2771–2784, 2011.

Harman, C., Stewardson, M., and DeRose, R.: Variability and uncertainty in reach bankfull hydraulic geometry, J. Hydrol., 351, 13–25, 2008.

Javelle, P., Demargne, J., Defrance, D., Pansu, J., and Arnaud, P.: Evaluating flash flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system, Hydrolog. Sci. J., 59, 1390–1402, 2014.

Kalteh, A. M., Hjorth, P., and Berndtsson, R.: Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environ. Model. Softw., 23, 835–845, 2008.

Kjeldsen, T. R., Smithers, J. C., and Schulze, R. E.: Flood frequency analysis at ungauged sites in the KwaZulu-Natal Province, South Africa, Water SA, 27, 315–324, 2001.

Kjeldsen, T. R., Jones, D. A., and Morris, D. G.: Using multiple donor sites for enhanced flood estimation in ungauged catchments, Water Resour. Res., 50, 6646–6657, 2014.

Kocjancic, R. and Zupan, J.: Modeling of the river flowrate: the influence of the training set selection, Chemom. Intell. Lab. Syst., 54, 21–34, 2000.

Kohonen, T.: Self-Organizing Maps, 2nd Edn., Springer, Berlin, 362 pp., 1997.

Ley, R., Casper, M. C., Hellebrand, H., and Merz, R.: Catchment classification by runoff behaviour with self-organizing maps (SOM), Hydrol. Earth Syst. Sci., 15, 2947–2962, doi:10.5194/hess-15-2947-2011, 2011.

Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P.: Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, Environ. Model. Softw., 25, 891–909, doi:10.1016/j.envsoft.2010.02.003, 2010.

Martina, M. L. V., Todini, E., and Libralon, A.: A Bayesian decision approach to rainfall thresholds based flood warning, Hydrol. Earth Syst. Sci., 10, 413–426, doi:10.5194/hess-10-413-2006, 2006.

Merz, B., Elmer, F., and Thieken, A. H.: Significance of "high probability/low damage" versus "low probability/high damage" flood events, Nat. Hazards Earth Syst. Sci., 9, 1033–1046, doi:10.5194/nhess-9-1033-2009, 2009.

Merz, R. and Bloschl, G.: Flood frequency regionalisation – Spatial proximity vs. catchment attributes, J. Hydrol., 302, 283–306, 2005.

Minns, A. W. and Hall, M. J.: Artificial neural network concepts in hydrology. in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G. and McDonnell, J. J., John Wiley and Sons, Chichester, UK, 307–320, 2005.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Bloschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: Panta Rhei-Everything Flows: Change in hydrology and society-The IAHS Scientific Decade 2013–2022, Hydrolog. Sci. J., 58, 1256–1275, 2013.

Montesarchio, V., Ridolfi, E., Russo, F., and Napolitano, F.: Rainfall threshold definition using an entropy decision approach and radar data, Nat. Hazards Earth Syst. Sci., 11, 2061–2074, doi:10.5194/nhess-11-2061-2011, 2011.

Mount, N. J., Maier, H. R., Toth, E., Elshorbagy, A., Solomatine, D., Chang F.-J., and Abrahart, R. J.: Data-driven modelling approaches for socio-hydrology: opportunities and challenges within the Panta Rhei Science Plan, Hydrolog. Sci. J., 61, 1192–1208, doi:10.1080/02626667.2016.1159683, 2016.

Muttiah, R. S., Srinivasan, R., and Allen, P. M.: Prediction of two year peak stream discharges using neural networks, J. Am. Water Resour. Assoc., 33, 625–630, 1997.

Norbiato, D., Borga, M., and Dinale, R.: Flash flood warning in ungauged basins by use of the flash flood guidance and model-based runoff thresholds, Meteorol. Appl., 16, 65–75, doi:10.1002/met.126, 2009.

Ntelekos, A. A., Georgakakos, K. P., and Krajewski, W. F.: On the uncertainties of flash flood guidance: Towards probabilistic forecasting of flash floods, J. Hydrometeorol., 7, 896–915, doi:10.1175/JHM529.1, 2006.

Pandey, G. R. and Nguyen, V.-T.-V.: A comparative study of regression based methods in regional flood frequency analysis, J. Hydrol., 225, 92–101, 1999.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H., Buizza, R., and de Roo, A.: New dimensions in early flood warning across the globe using grand-ensemble weather predictions, Geophys. Res. Lett., 35, L10404, doi:10.1029/2008GL033837, 2008.

Patton, A. J. and Timmermann, A.: Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, J. Econometr., 140, 884–918, 2007.

Reed, S., Schaake, J., and Zhang, Z.: A distributed hydrologic model and threshold frequency based method for flash flood forecasting at ungauged locations, J. Hydrol., 337, 402–420, 2007.

Rumelhart, D. E., Widrow, B., and Lehr, M. A.: The basic ideas in neural networks, Commun. ACM, 37, 87–92, 1994.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci., 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.

Sene, K.: Flash floods: forecasting and warning, Springer, Dordrecht, p. 385, 2013.

Shahin, M., Maier, H., and Jaksa, M.: Data Division for Developing Neural Networks Applied to Geotechnical Engineering, J. Comput. Civ. Eng., 18, 105–114, 2004.

Shu, C. and Burn, D. H.: Artificial neural network ensembles and their application in pooled flood frequency analysis, Water Resour. Res., 40, W09301, doi:10.1029/2003WR002816, 2004.

Shu, C. and Ouarda, T. B. M. J.: Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system, J. Hydrol., 349, 31–43, 2008.

Silva, D. G. E., Jino, M., and de Abreu, B. T.: Machine learning methods and asymmetric cost function to estimate execution effort of software testing, in: IEEE Proc. Third International Conference on Software Testing, Verification and Validation (ICST), 7–9 April 2010, Paris, 275–284, 2010.

Simor, V., Hlavcova, K., Silvia Kohnova, S., and Szolgay, J.: Application of Artificial Neural Networks for estimating index floods, Contrib. Geophys. Geodesy, 42/4, 295–311, 2012.

Singh, K. K., Pal, M., and Singh, V. P.: Estimation of Mean Annual Flood in Indian Catchments Using Backpropagation Neural Network and M5 Model Tree, Water Resour. Manage., 24, 2007–2019, doi:10.1007/s11269-009-9535-x, 2010.

Smith, A., Sampson, C., and Bates, P.: Regional flood frequency analysis at the global scale, Water Resour. Res., 51, 539–553, doi:10.1002/2014WR015814, 2015.

Srinivas, V. V., Tripathi, S., Rao, A. R., and Govindaraju, R. S.: Regional flood frequency analysis by combining self-organizing feature maps and fuzzy clustering, J. Hydrol., 348, 148–166, 2008.

Stedinger, J. R. and Lu, L.: Appraisal of regional and index flood quantile estimators, Stoch. Hydrol. Hydraul., 9, 49–75, 1995

Stedinger, J. R. and Tasker, G. D.: Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared, Water Resour. Res., 21, 1421–1432, 1985.

Toth, E.: Catchment classification based on characterisation of streamflow and precipitation time series, Hydrol. Earth Syst. Sci., 17, 1149–1159, doi:10.5194/hess-17-1149-2013, 2013.

Toth, E.: Asymmetric Error Functions for Reducing the Underestimation of Local Scour around Bridge Piers: Application to Neural Networks Models., J. Hydraul. Eng., 141, 04015011, doi:10.1061/(ASCE)HY.1943-7900.0000981, 2015.

UCAR – University Corporation for Atmospheric Research: Flash Flood Early Warning System Reference Guide 2010, http://www.meted.ucar.edu/communities/hazwarnsys/ffewsrg/FF_EWS.pdf (last access: 17 June 2015), 2010.

Verkade, J. S. and Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning, Hydrol. Earth Syst. Sci., 15, 3751–3765, doi:10.5194/hess-15-3751-2011, 2011.

Ward, P. J., Jongman, B., Sperna Weiland, F. C., Bouwman, A., van Beek, R., Bierkens, M. F. P., Ligtvoet, W., and Winsemius, H. C.: Assessing flood risk at the global scale: Model setup, results, and sensitivity, Environ. Res. Lett., 8, 44019, doi:10.1088/1748-9326/8/4/044019, 2013.

Wilkerson, G. V.: Improved bankfull discharge prediction using 2-year recurrence-period discharge, J. Am. Water Resour. Assoc., 44, 243–258, doi:10.1111/j.1752-1688.2007.00151.x, 2008.

WMO: Manual on flood forecasting and warning, WMO Series No. 1072, 142 pp., http://www.wmo.int/pages/prog/hwrp/publications.php (last access: 17 June 2015), 2011.