PAI-OFF: a new strategy for online flood forecasting in mountainous catchments

J. CULLMANN, G. H. SCHMITZ & W. GÖRNER

Institute of Hydrology and Meteorology, Dresden University of Technology, Würzburger Str 46, D-01187 Dresden, Germany Johannes.Cullmann@mailbox.tu-dresden.de

Abstract We present PAI-OFF (Process Modelling and Artificial Intelligence for Online Flood Forecasting), which combines the reliability of physically based, sophisticated modelling with the operational advantages of Artificial Neural Networks (ANN). Thus we are able to improve ANN performance in the flood forecasting context by detailed process modelling. Low computation times and robustness are the key features of ANN models and also form the basic requirements for flash flood forecasting. After presenting the theory of the new methodology, the results of a catchment related meteorological analysis for generating storm scenarios serve as the input to a coupled hydrological/hydraulic model, which is set up for a mountainous catchment in east Germany. Along these lines we operate the catchment model for all realistically possible constellations of flood formation. This results in a database consisting of corresponding input/output vectors. We complete the database for training the ANN by adding yet more flood relevant data for characterizing the hydrological and meteorological catchment situation prior to a storm event. After this preparatory step, the ANN is applied for online flash flood forecasting in the considered catchment using an "unseen" storm event, i.e. one which did not feature in the training process. The convincing agreement between the predicted and observed flood hydrograph underlines the application potential of the new PAI-OFF methodology for online flood forecasting even in smaller catchments.

Key words artificial neural networks; flood forecasting; network training

INTRODUCTION

Each and every flood event is the direct consequence of complex hydrological processes which, in turn, are a result of rainfall characteristics as well as the catchment specific topographic and soil hydrological properties (nonlinearity and pronounced dynamics are typical features of this phenomenon). Mountainous catchment areas—with their steep slopes and short flow paths, high nonlinearity and pronounced dynamics—restrict the performance of current flood forecasting models and, thus, only allow for short warning periods. In order to extend the flood warning period we have to take into consideration a quantitative precipitation forecast and a detailed and physically based description of the rainfall—runoff process and wave propagation in the river. Liu & Todini (2002) do indeed consider the quantitative precipitation forecast in their TOPKAPI forecast model; however, the operation requirements unfortunately lead to a significant simplification of their modelling concept. Moreover, their approach does not include rigorous hydraulic modelling for taking backwater effects into account. Barbero *et al.* (2001) attempt to introduce hydrodynamic numerical

modelling for flood forecasting; however, these have the disadvantage of enormous computational requirements and their use necessitates considerable experience in numerical modelling. Trying to get around the problem, Hsu *et al.* (2002) compared a forecast algorithm on the basis of the ARMA concept (Box-Jenkins, 1976) with ANN solutions, for considering only one step ahead forecast strategies. In view of these shortcomings a considerable amount of research has been invested for adapting the theory of Artificial Neural Networks (ANN) as the basis for flood forecasting. In addition to the rather general observations concerning ANN and their role in hydrology (ASCE 1 and ASCE 2), Dawson (2001) provides a more specific overview regarding ANN and their application with respect to rainfall–runoff modelling.

In this context, the attempt made by Sajikumar & Thandaveswara (1999) and also Shamseldin (1997), to predict flood waves using solely ANN approaches cannot be used for general flood forecasting. The reason being that the training data (observed rainfall–runoff hydrographs) never covers the full range of possible flood peaks. ANN, due to their empirical character, are not able to provide reliable results where extrapolation is required (Minns & Hall, 1996), and therefore fail to extrapolate extreme events from historical data alone. Research has also been carried out on the various ANN types in view of their respective abilities to predict discharge time series (Castellano-Mendez, 2004). These attempts aim to improve the potential forecasting performance of ANN by combining different types of ANN with methods of time series analysis or fuzzy logic approaches (Nayak *et al.*, 2004; Rajurkar *et al.*, 2004).

Unfortunately these strategies are likewise built upon pure empirical approaches and, thus, feature the same shortcomings—i.e. they cannot reliably predict a rare extreme flood event if it was not part of the training data. Despite significant progress in this field during the last few years, none of the currently available hydrological/hydraulic models is able to offer a robust and computationally efficient solution to the operational flood forecasting problem when taking backwater effects, or even weather forecast uncertainties, into account.

The new PAI-OFF methodology offers a way out of the dilemma. We overcome the ANN deficiencies as regards poor extrapolation capability, by comprehensively enlarging the training database for exploiting catchment specific topographic and soil hydraulic properties with the help of physically based modelling.

THE PAI-OFF METHODOLOGY

ANN are computationally highly efficient tools that can straightforwardly approximate any nonlinear function to an arbitrary accuracy; they are, however, a prisoner of their training data (Minns & Hall, 1996). In the context of flood forecasting this prohibits the direct use of ANN due to the fact that observation (training) data never contain all the possible constellations of extreme flood events. Bearing this in mind, PAI-OFF employs, in its three preparatory steps (Fig. 1), a physically based catchment model for transferring catchment specific information into the ANN. In the first step a rigorous catchment model, which consists of a rainfall–runoff as well as of a hydraulic module, is set up. This model is calibrated and validated with the available meteorological data and the hydrographs at catchment internal points as well as at the catchment outlet (Fig. 1 "Physically based catchment modelling"). We then generate all realistically possible rainstorm events based on the KOSTRA study (1997). In the second step, this ensemble of rainstorms is the transferred into the catchment response by means of the validated catchment model. Thus, a scenario database is established which contains the full range of constellations with respect to all locally possible rainstorm events, the initial catchment conditions and the corresponding catchment response at the outlet (Fig. 1 "Generation of training database"); here catchment internal points are not yet included. In a third step of the preparatory phase of PAI-OFF, the ANN is trained to reproduce all the input (generated rainstorms) output (catchment model response at outlet) relationships. This way the ANN portrays the hydrological/hydraulic model as a black box. It represents all the functional relationships of the transformation of rainfall to runoff covered by the database; we thus replace the catchment model with the ANN core of PAI-OFF.

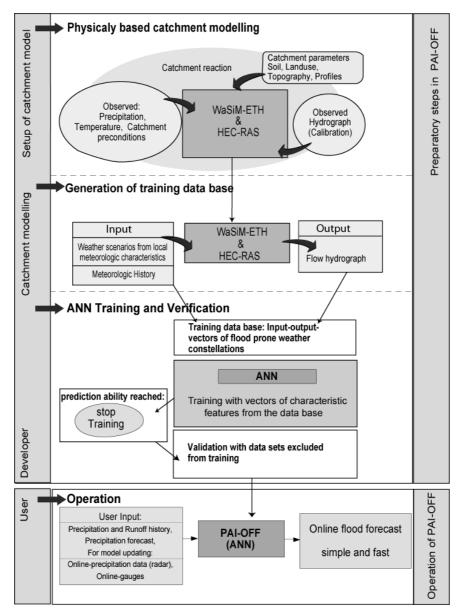


Fig. 1 The PAI-OFF methodology.

The steps involved in the preparatory phase only have to be carried out once for a particular catchment area. As soon as the preparatory steps are completed (Fig. 1), PAI-OFF can be routinely operated as an efficient tool for flood forecasting, requiring the input shown in Fig. 1 and this with negligible computational effort. When the forecast is required, all flood relevant information is thus already pre-processed and stored in the system. In this respect, PAI-OFF is to be seen as a modern, high dimensional equivalent to the Nomogrammes of former times; a tool of the latest generation which is even capable of learning by doing.

THE HYDROLOGICAL/HYDRAULIC MODELLING

As far as PAI-OFF is concerned, any and all types of reliable hydrological/hydraulic models can be employed for portraying the catchment behaviour. After a comparative analysis we selected WaSiM-ETH (Gurtz *et al.*, 2000) for the subsequent demonstration of the PAI-OFF principle due to its efficient use of the Richards' equation for adequately describing the soil moisture distribution prior to a flood.

We use Grass-GIS as a parameterization tool with respect to the parameters of the digital elevation model, soil characteristics and type of land use. The meteorological parameters, i.e. precipitation, temperature, air humidity, global radiation and wind velocity are processed for the model on an hourly basis using a grid of 1 km². The interpolation with respect to the mesh points employs external drift kriging for temperature and precipitation data and ordinary kriging for air humidity, wind and global radiation.

The modules for flood routing are based on two different modelling techniques: in steep and narrow river reaches we use the translation diffusion approach which is included in WaSiM-ETH, and in areas with significant backwater effects as well as in the vicinity of river junctions we apply HEC-RAS, a one dimensional hydrodynamic model based on the numerical solution of the Saint-Venant equations.

For coupling, the results of WaSiM-ETH are used for the upper boundary conditions of HEC-RAS. This interface between WaSiM and Hec-Ras always lies significantly upstream of the backwater-influenced reaches. Thus, we avoid iterative procedures when coupling the hydrological and hydraulic models. On the basis of observed data, automatic calibration using the SCE-UA algorithm according to Duan *et al.* (1994) is used to back up the thorough manual calibration of both models.

FLOOD RELEVANT METEOROLOGIC/HYDROLOGICAL FEATURES PRIOR TO A FLOOD EVENT

For the reliability of a forecast (Fig. 2) it is of crucial importance that all type of flood relevant, commonly available data, be taken into consideration by the flood forecasting system. In this context, ANN have many advantages: with the aid of a vector of characteristic features, relevant information as to the meteorological/hydrological behaviour which was in existence prior to the event can easily be transferred into the ANN as shown in Fig. 2. Here T represents time series of temperature, R Rainfall, Q Runoff and I any other time series that could be of potential interest (e.g. vegetation period).

60

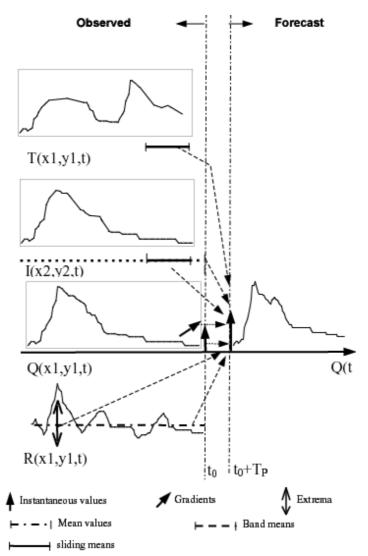


Fig. 2 principle of transformation of meteorological input for the ANN.

This feature vector can be formed on the basis of pre-rainfall index (*R*-series in Fig. 2), preceding discharge at relevant gauges (*Q*-series in Fig. 2) and/or sliding means as for e.g. the moving average temperature of the previous 3 days (Fig. 2 T series). The selected characteristic features of the vector need to satisfy two basic requirements: firstly, they must refer exclusively to data, which is commonly available at the flood alarm centre, and secondly they should portray a maximum of information relevant for the formation of flood events. This latter aspect also serves to reduce the required number of input neurons and, thus, assists in restricting the overall dimensions of the ANN.

We achieve this goal by adopting the unit hydrograph concept for defining selected members of the feature vector. On the basis of this concept we incorporate the principle of superposition into the ANN by not predicting the absolute discharge value but rather its alteration relative to the preceding model results.

Along these lines we also exploit the fact that the discharge alteration ΔQ can be derived from a weighted sum of the rainfall within a relatively small time window.

While the weights are determined according to the unit hydrograph concept, the time window is defined by the prediction interval T_p and the time of concentration T_t , i.e. the time needed for the rainfall in the most distant catchment areas to become effective as regards the river discharge.

The usage of the concentration time also allows subdividing the catchment according to the isochrone's principle so as to be able to work with integrated rainfall values for each of the resulting catchment areas. The soil moisture distribution existing in the catchment prior to an event can be evaluated by continuously operating a water balance model. The PAI-OFF methodology offers a way around this inconvenience by providing the option of estimating the catchment conditions on the basis of relevant meteorological and hydrological data prior to the event as for e.g. the weighted sum of daily and/or weekly precipitation, the minimum discharge values and climate data as shown in Fig. 2.

TRAINING OF THE ANN

Sets of corresponding input–output data represent the foundation for teaching the ANN (Fig. 1). The input vector is formed by the time series of rainfall generated in accordance with the local meteorological situation (together with the prevalent meteorological/hydrological features in existence prior to the flood).

The response of the physically based catchment model, i.e. the discharge hydrograph at the reference gauge, defines the corresponding output vector.

A successful training requires a sufficiently great number of neurons for an optimal representation of the considered process. Likewise, sufficient training data needs to assure that all constellations of the rainfall–runoff/flood wave propagation phenomena are fully portrayed. The reliability of the ANN's forecast may significantly suffer from a substantial violation of these requirements due to the imperfect training.

Keeping this in mind, we subsequently apply a problem adapted learning algorithm for training the selected MLFN (Multi Layer Feed Forward Net). After the successfully validated training, the next step in the PAI-OFF methodology requires the verification of the forecast performance of the trained MLFN. After successful validation, i.e. testing the ANN with data which was not featured in the training, the preparatory phase is complete.

Now, PAI-OFF is ready to be applied as a simple and computationally highly efficient and robust tool for online flood forecasting

ONLINE FLOOD FORECASTING WITH PAI-OFF

The routine operation of PAI-OFF uses the meteorological data provided by the national/regional weather forecast, i.e. quantitative precipitation forecast, data on snow melting, temperature, radiation and wind velocity in the available spatial resolution. Moreover, PAI-OFF allows exploitation of information from meteorological/hydro-logical time series, i.e. discharge hydrographs at reference gauges, and the rainfall pattern of the previous days/weeks. All this information is generally available at flood alarm centres.

Moreover, online data from existing rain and/or river gauges can be used for continuously updating the PAI-OFF input data and, thus, will further improve the forecast reliability.

APPLYING PAI-OFF TO A MOUNTAINOUS CATCHMENT

A catchment in the Ore Mountains (Erzgebirge, east Germany) serves as the basis for a first test application of the new methodology. Our investigations focus on the Schwarze Pokau River, which is a tributary of the Mulde River, and we use the Zöblitz gauge as a reference. The rolling hills of the catchment attain 900 m above m.s.l. The land use mainly consists of forest (35%), agricultural land (30%) and the remainder is fallow.

Catchment modelling

As already outlined, the data provided by the digital elevation model is used in a grid format of 1×1 km. The model takes into account the transient character of seasonal variation with respect to plant growth and its impact on soil water dynamics and evapotranspiration. Due to the rapid catchment response to a rainstorm event we use a time discretization of 1 h for the flood relevant processes. This refers especially to rainfall series, wind velocity, radiation, temperature and air humidity. The soil hydrological parameters were derived from the official soil map of the area (Bük 400) and Corine (2000) provided the data for the land use. The considered reach of the Zöblitz River is characterized by a steep bottom slope without backwater effects. This leads us to employ the translation–diffusion approach for the flood routing in this test case.

On the basis of the aforementioned data processing strategy we parameterized the rainfall–runoff model. The successful calibration of the catchment model employed time series from January 1972 to June 1974 and the subsequent validation referred to data from July 1974 to December 1976 (Fig. 3). The excellent result has to be emphasized in the light of the fact that the flood peak of the validation flood exceeds the maximum flood peak during the calibration period by more than 100% (Fig. 3). Here it is worth noting that the RR-modelling which leads to the training database, despite its good validation, could be further refined. In this paper we put emphasis on the performance of the ANN to reproduce the database, we do not intend to improve the structure of the underlying RR model.

Meteorological characterization of the catchment location

The quality of the rainfall data in its spatial and temporal resolution is a decisive factor for forecast reliability. We therefore analysed the flood relevant rainfall patterns of the catchment on the basis of a 40-year time series of meteorological data recorded on an hourly basis, courtesy of the German National Weather Service.

We firstly employed our catchment model using this data for a continuous simulation of the rainfall–runoff dynamics. The results were analysed and allowed for

J. Cullmann et al.

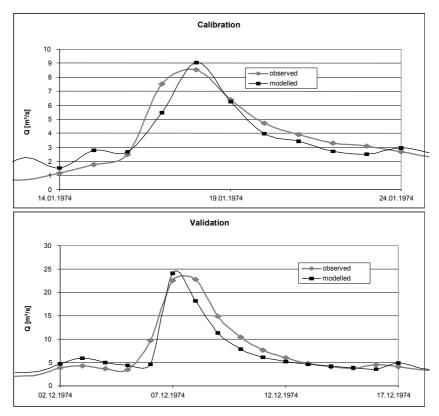


Fig. 3 Example from calibration and validation of the catchment modelling.

classification of the rainfall events with respect to the impact of pre-flood catchment conditions. In so doing we came upon the potential combinations of pre-flood catchment conditions and rainstorm events which initiate flooding. We then used the results of the meteorological catchment analysis together with the original data for setting up a stochastic rainstorm generator. It provided rainstorm scenarios from given rainfall quantities and given storm durations, which correspond to the catchment's typical meteorological behaviour. This means that the parameters of the generated rainfall hydrograph portray the local situation with respect to:

- (a) the hydrograph shape;
- (b) the hydrograph skewness;
- (c) the drift direction and velocity for advective storm events;
- (d) the location of rainstorm centre and radius for convective events;
- (e) the probabilistic term;
- (f) the overall rainfall intensities of the events were derived from the probable maximum precipitation documentation of the area (Kostra-Dwd, 1997).

Generation of training data and setting up of the ANN

The stochastic rainstorm generator is now applied to generate 55 locally typical storm scenarios. Each of the 55 scenarios is related to a hydrological/meteorological situation in existence prior to the event. In Table 1 some of the characteristics of the generated

	REAL	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	SC_7	SC_8	SC_9	SC_10
MEAN	3.69	1.15	2.36	2.04	2.39	1.19	1.43	1.19	0.99	1.70	3.86
MAX	26.26	23.43	5.79	16.05	29.83	28.27	51.06	40.34	28.27	5.55	16.43
STD	5.21	4.03	2.09	4.33	6.13	4.86	7.87	5.20	4.49	1.72	5.59

Table 1 Rainstorm characteristics of a real (2002) and 10 of the generated storms.

storms are compared to the 3-day rainstorm event that triggered the 2002 flood (a rare event in the order of 150 years) in the observed basin.

The spatial distribution of the rainfall was organized using 129 grid cells throughout the catchment. According to the procedure outlined in the preceding, we now use the isochrone's principle for defining 4 sub areas where the amount of time needed for the rainfall to reach the river is practically the same within a selected tolerance limit. Within these sub areas we work with an average rainfall distribution. The feature vectors for characterizing the catchment condition were:

- (a) a moving average of daily / weekly / monthly rainfall;
- (b) The minimum discharge of the previous months;
- (c) The weighted sum of the previous $10 \times \text{hourly rainfall}$ and temperature.

In order to assure in our first test application the comparability with Dawson & Wilby (2001), Garcia-Bartual (2002), Zhang & Govindaraju (2003), all of which started their prediction on the basis of measured rainfall, we selected a forecast period of 5 h. This corresponds to the overall concentration time of the catchment. If not only the prediction of a single discharge after the forecast period is required but also its development, the forecast has to be repeated according to the desired resolution of the discrete flood hydrograph which is to be predicted. For enlarging the forecast period, the meteorological input has simply to be extended in accordance with the meteorological forecast data. In both instances no problems arise for the operation of PAI-OFF due to its negligible computation time.

On the basis of the input scenarios the WaSiM-ETH produces the resulting flood hydrograph at the Zöblitz reference gauge. This results in input–output data sets which altogether form the training database. A problem adapted learning algorithm supplies the ANN with this flood characteristic behaviour of the catchment along with the preevent features. Organising the feature vectors, their structure and the generated rainfall data in a database allows an efficient organization of several subcatchments in various test versions and their straightforward hierarchical arrangement within the global catchment. The training of the ANN provided excellent results throughout the training period, which is documented in Fig. 4, and shows an arbitrarily selected sample section of the complete forecast period. These extremely satisfying training results were achieved for all the characteristically different 55 flood events, which formed the full training programme.

Thus, the overall training of the MLFN demonstrated its astonishing capability for portraying the whole range of discharges, i.e. to portray indirectly, on the basis of the training database obtained by the hydrological modelling, all flood relevant hydrological processes in the catchment. Table 2 gives an idea of some performance criteria of both models. The setting up of the validation database is performed analogously, i.e. the rainstorm generator provides a second set of realizations of the

J. Cullmann et al.

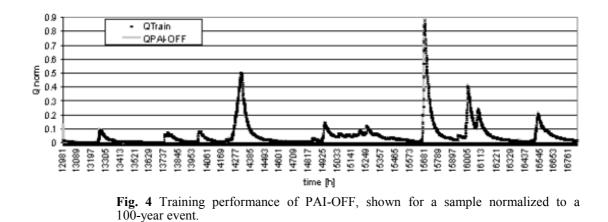


Table 2 Performance criteria of the catchment model and the PAI-OFF core.

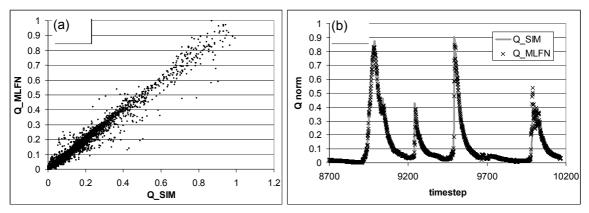
Criteria	PAI-OFF core training	PAI-OFF core validation	WaSiM-ETH calibration	WaSiM-ETH validation
Nash-Sutcliff	0.97*	0.95*	$0.5 - 0.85^{\dagger}$	$0.4 - 0.8^{\dagger}$
RMSE (peaks)	0.006	0.007	0.047	0.09

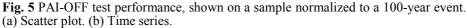
* Overall performance.

[†]Based on the validation of single events.

stochastic rainfall process. It represents a second pool of rainfall–runoff data, which does not form part of the training programme, i.e. the ANN is exposed to this data only for the purpose of assessing the forecast reliability of PAI-OFF.

Besides this data pool, which only forms one part of the test database, the physically based catchment model completes the test. Before confronting the ANN with this data pool, it is first used by the catchment model for computing the resulting flood hydrograph at the Zöblitz gauge. Together with the storm specific feature vector, PAI-OFF now uses the same input data to predict the corresponding flood hydrograph as a reference for assessing its prediction performance. Figure 5 illustrates both the hydrographs obtained from WaSiM and PAI-OFF. They show a convincing agreement,





which is also to be seen in the light of the fact that (other than in real life conditions) no updating techniques on the basis of online measurements were employed.

SUMMARY AND CONCLUSIONS

PAI-OFF (Process Modelling and Artificial Intelligence for Online Flood Forecasting) is the new tool for ideally satisfying the most important requirements for flash flood forecasting, i.e. low computation times, complete robustness and simple operation together with the high predictive reliability of detailed catchment modelling. The first part of this contribution explains in detail the consecutive preparatory steps required to set up PAI-OFF for a given catchment and provides insight into the theory of the new methodology with its different modules. We selected a typical small watershed in the Erzgebirge (east Germany) in order to perform a first test application of PAI-OFF. After parameterization of a hydrological/hydraulic catchment model, we perform a meteorological analysis of the region in order to include as much "core" information as possible. The results finally provide storm scenarios that cover all the realistically possible meteorological constellations of the catchment area. In a subsequent step, these serve as input to the catchment model for generating the resulting flood scenarios considering all flood relevant initial catchment conditions. Operating the catchment model on the basis of this input data leads to a database of corresponding input/output vectors, representing all the realistically possible constellations of flood formation. We complete the database for training the ANN by adding yet more flood relevant data for characterizing the hydrological and meteorological catchment situation prior to a storm event (like flood hydrographs, average rainfall of the last month, week, day, etc.). After this preparatory step, PAI-OFF is applied for online flash flood forecasting in the considered catchment using an "unseen" storm event, i.e. one that did not feature in the training process. The convincing agreement between the predicted and observed flood hydrograph underlines the application potential of the new PAI-OFF methodology for online flood forecasting even in smaller catchments. The advantages of the method presented are:

- (a) different types of catchment models can be integrated (e.g. RR and hydraulic models);
- (b) different parameter sets are integrated within the ANN model (e.g. runoff generation parameters for medium and large flood events);
- (c) the model is extremely fast (several thousands of scenarios can be processed in less than a few seconds on a standard PC) which makes online uncertainty analysis possible in future applications.

REFERENCES

- ASCE-1 (2000) Artificial neural networks in hydrology. I: Preliminary concepts by the ASCE task committee on application of artificial neural networks in hydrology. J. Hydrol. Engng 5(2), April 2000, 115–123.
- ASCE-2 (2000) Artificial neural networks in hydrology. II: Hydrologic applications by the ASCE Task committee on application of artificial neural networks in hydrology. *J. Hydrol. Engng* **5**(2), April 2000, 124–137.
- Barbero, S., Rabuffetti, D., Buffo, M. & Høst-Madsen, J. (2001) Development of a physically based flood forecasting system In: MIKE FLOOD Watch (Fourth DHI Software Conference, Elsinore, 6–8 June, 2001).

J. Cullmann et al.

- Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, rev. edn. Holden-Day, San Fransisco, California, USA.
- Castellano-Mendez, M., Gonzalez-Manteiga, W., Febrero-Bande, M., Prada-Sanchez, J. M., Lozano-Calderon, R. (2004) Modelling of the monthly and daily behaviour of the runoff of the Xallas River using Box-Jenkins and neural networks methods. J. Hydrol. 296, 38–58.
- CORINE (2000) http://www.caf.dlr.de/caf/anwendungen/projekte/projekte_nutzung/corine/.
- Dawson, C. W. & Wilby, R. L. (2001) Hydrological modeling using artificial neural networks *Progr. Phys. Geogr.* 25(1), 80–108.
- Duan, Q., Gupta, V. K. & Sorooshian, S. (1994) http://www.eckhardt.metcon-umb.de/kalibrierung/sceua1.pdf.
- Gurtz, J., Zappa, M., Jasper, K., Lang, H., Badoux, A., Verbunt M. & Vitvar. T. (2000) Modelling and validation of runoff and its components in Swiss Pre-Alpine and Alpine catchments. In: *International Workshop on Runoff Generation* and *Implications for River Basin Modelling* (Freiburg i.Br. 9–12 October 2000, Freiburger Hefte).
- Hsu, K. L., Gupta, H. V., Gao, X., Sorooshian, S. & Imam, B. (2002) Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modelling and analysis. *Water Resour. Res.* 38(12), 1–17.
- KOSTRA-DWD (1997) http://www.itwh.de/S_kostra.htm, enthält den vollständigen im KOSTRA-Atlas 1997.
- Liu, Z. & Todini, E. (2002) Towards a comprehensive physically based rainfall-runoff model. *Hydrol. Earth System Sci.* **6**, 859–881.
- Minns, A. W. & Hall, M. J. (1996) Artificial neural networks as rainfall-runoff models. Hydrol. Sci. J. 41, 399-417.
- Nayak, P. C., Sudheer, K. P., Rangan, D. M. (2004) A neuro-fuzzy computing technique for modelling hydrological time series. J. Hydrol. 291, 52–66.
- Rajurkar, M. P., Kothyari, U. C. & Chaube, U. C. (2004) Modelling of the daily rainfall-runoff relationship with artificial neural network. J. Hydrol. 285, 96–113.
- Sajikumar, N. & Thandaveswara, B. S. (1999) A nonlinear rainfall-runoff model using an artificial neural network. *J. Hydrol.* **216**, 32–55.
- Shamseldin, A. Y. (1997) Application of a neural network technique to rainfall-runoff modelling. J. Hydrol. 199, 272-294.