WATER RESOURCES RESEARCH, VOL. 35, NO. 11, PAGES 3547–3552, NOVEMBER 1999

Forecasting river flow rate during low-flow periods using neural networks

Marina Campolo, Alfredo Soldati,[1] and Paolo Andreussi[2]

Centro Interdipartimentale di Fluidodinamica e Idraulica, Università di Udine, Udine, Italy

Abstract. The pollution in the river Arno downstream of the city of Florence is a severe environmental problem during low-flow periods when the river flow rate is insufficient to support the natural waste assimilation mechanisms which include degradation, transport, and mixing. Forecasting the river flow rate during these low-flow periods is crucial for water quality management. In this paper a neural network model is presented for forecasting river flow for up to 6 days. The model uses basin-averaged rainfall measurements, water level, and hydropower production data. It is necessary to use hydropower production data since during low-flow periods the water discharged into the river from reservoirs can be a major fraction of total flow rate. Model predictions were found to be accurate with root-mean-square error on the predicted river flow rate less then 8% over the entire time horizon of prediction. This model will be useful for managing the water quality in the river when employed with river quality models.

1. Introduction

The region of the river Arno downstream of the city of Florence is highly polluted, and actions to reclaim and to protect the environment are needed. In this region the sources of pollution are agricultural and farming activities, untreated urban sewage (mainly around Florence), and industry (paper mills and textiles and leather factories). Legal threshold values for pollutants in the wastewater do not take into account flow rate variations of the receiving water body. Therefore pollution discharge associated with a particularly dry summer may produce severe environmental damage. A possible means of controlling pollution in the river could be the systematic use of the reservoirs along the river which are employed by the National Power Agency (ENEL) to produce electricity. An optimal schedule for their discharge could help to maintain the river flow rate at a level sufficient to promote the natural pollution remediation mechanisms, that is, mixing, transport, and degradation of pollutants. During low-flow periods the contribution of reservoir discharge can range from 25% to 80% of the total flow rate of the Arno.

Broadly speaking, two different approaches can be used to predict the flow rate of a river [see *Buchtele et al.*, 1996; *Woolhiser*, 1996]: physical models and statistical models. Physical models are based on the solution of a complex set of differential equations, and their extensive use has proved their accuracy and reliability [see, e.g., *Garrote and Bras*, 1995; *Refsgaard and Knudsen*, 1996; *Todini*, 1995]. However, these models are complex, computationally intensive, and data demanding. In contrast, statistical models are based on the analysis of the statistical properties of the time series of physical parameters and have the advantage of being simpler from the mathematical point of view. They are, however, often less accurate [*Mukherjee and Mansour*, 1996; *Hsu et al.*, 1995].

Recently, a number of complex processes have been modeled with the aid of neural networks [*Bishop*, 1994; *Maier and Dandy*, 1996; *Cheng et al.*, 1995], which may be properly classified as nonlinear statistical models. Models based on the use of neural networks have been used to predict rainfall-runoff transformation [see *Chakraborty et al.*, 1992; *Zhu et al.*, 1994; *Hsu et al.*, 1995; *Raman and Sunilkumar*, 1995; *Smith and Eli*, 1995; *Mason et al.*, 1996; *Minns and Hall*, 1996; *Shamseldin*, 1997] and to forecast river floods during heavy rainfall [*Campolo et al.*, 1999]. Models based on neural networks have the advantage of being simple and reasonably accurate. When a physical phenomenon is controlled by several parameters, some of which can be difficult to estimate, neural networks can predict system evolution (flow rate in the river) using a limited set of the variables involved in the phenomenon (rainfall, water level, and dam discharge in the present case).

During low-flow periods the flow rate of the river Arno is determined by the following significant variables: discharge from the reservoirs, occasional rainfall, and current state of the basin. In this work we used two years of data (1992–1993) to develop a model based on neural networks to forecast the river flow rate. The neural model predicts the river flow rate for 6 upcoming days. The forecast of the river flow rate obtained can be used further by water quality models [see *Walton and Webb*, 1994; *Li et al.*, 1992] to predict water quality evolution in the river and, possibly, to reduce the impact of critical environmental conditions.

2. Site and Data

2.1. Site

The Arno basin, shown in Figure 1, includes the valley of Valdarno Superiore and the subbasins of the tributaries (Chiana, Ambra, and Sieve upstream of Florence and Bisenzio, Ombrone, Pesa, Elsa, and Era downstream from Florence). Two reservoirs, La Penna and Levane, holding 13 and 3 million m^3, respectively, are located just upstream of Florence.

---

[1]Also at Dipartimento di Scienze e Tecnologie Chimiche, Università di Udine, Udine, Italy.

[2]Now at Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Pisa, Italy.

Paper number 1999WR900205.
0043-1397/99/1999WR900205$09.00

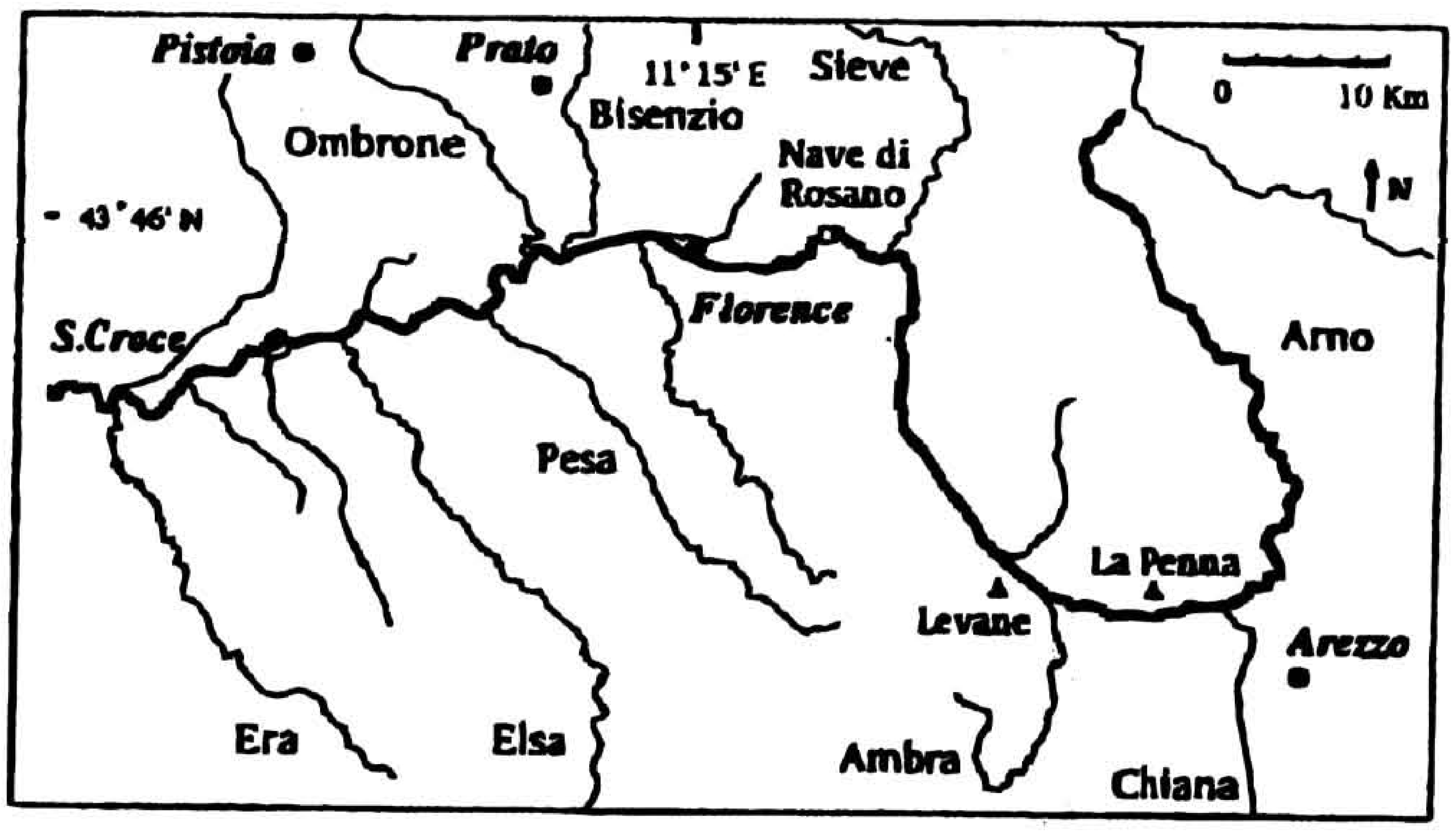

Figure 1. Basin of river Arno: location of control section, hydrometer site, and location of reservoirs.

The flow rate of the Arno is highly variable during the year, with minima less than 1 m^3/s and maxima about 4000 m^3/s. During high-flow periods the aim of water management policy is to control and, if possible, to prevent floods. During low-flow periods the aim of the water management policy is to ensure the availability of water for domestic, agricultural, and industrial uptake and to protect the water quality.

2.2. Available Data and Identification of Low-Flow Periods

During low-flow periods the river flow rate depends mostly on the following variables: state of the basin, that is, soil moisture content and groundwater recharge, occasional incoming rainfall, and water discharged from reservoirs. For the river Arno, information about the state of the basin and rainfall input was obtained from the large number of rain gauges and hydrometers managed by the Ufficio Idrografico e Mareografico of Pisa. The basin contributing to the flow of the river at the reference section of Nave di Rosano, located some kilometers upstream of Florence, covers about 4000 km^2. In this area, rainfall and water level are monitored continuously, with an acquisition interval of 15 min. Automatic data acquisition is made possible by a remote sensing system. The water level can be converted into flow rate by stage-discharge relationships.

In general, during low-flow periods, rainfall is only a small fraction of runoff. Rainfall falling over an unsaturated basin takes a larger time to become runoff than in saturated or partially saturated conditions [*Todini*, 1996]. Therefore the small quantity of rainfall and the increased time-lag characteristic of low-flow periods suggest that rainfall information is not crucial to make predictions and can be accounted for through averaged quantities. Rainfall data were obtained by averaging rainfall measurements from different rain gauge stations in the basin. The rainfall-flow rate time lag in unsaturated condition may be about a day or more, comparable with the resolution of predictions. In these instances the flow contribution from different regions may not be recognized in a daily prediction.

The water level was used to describe the current state of the basin. The recent trend of the water level carries information on the effects of soil moisture content and groundwater recharge on runoff production when information about the forcing inputs producing runoff are also given.

Data on the hydroelectric power production of the Levane dam with a time interval of 1 hour were provided by ENEL. Power production is related to the water discharged by the following relation:

$$P = kQH, \tag{1}$$

where P is power, Q is the flow rate discharged, and H is the height of the waterfall, that is, the difference between the level of water in the reservoir and the level of the tail water. Since the level of the reservoir is not continuously recorded, the flow rate is estimated with some uncertainty.

The total flow rate of the river at Nave di Rosano, as estimated from level data using the available stage-discharge relationship, and the flow rate discharged from the dam at Levane, calculated with (1) are shown in Figure 2. Water level measured by the hydrometer is the abscissa. Negative values of water level correspond to water level beyond the zero of the hydrometer, not corresponding to the bottom of the river section. The error bars on the flow rate discharged from the dam are caused by uncertainty regarding H. It is apparent that for water levels less than 1.5 m a significant contribution to river flow is given by the flow rate discharged from the dam, whereas for larger water levels it is the rainfall contribution, represented by the difference between the curves, that influences river flow rate. We assumed that the river is in conditions of low flow when the level at Nave di Rosano was less than 1.5 m for at least 6 consecutive days.

2.3. Data Used

Data input to the model were (1) basin-averaged rainfall data, (2) water level at Nave di Rosano, and (3) hydropower production at the Levane dam. All data refer only to low-flow periods. Considering daily data from 1992 and 1993 year series, 601 days corresponding to low-flow periods were found (see Table 1).

Neural network models require subdivision of the data set into different subsets for calibration and validation. Data from January 1992 to August 1993 were used to form the calibration set. Remaining data were used to form the validation set. We divided the data set into two subsets taking care that the statistical properties of the data of each subset were most similar (mean, variance, maximum, and minimum value of the water level). The data set used for calibration was further subdivided into two ensembles, the training set and the testing set, using the same criteria as before. Training data were used to find the optimal set of calibration parameters, and testing data were used to decide whether to stop calibration or not.

Rainfall, power, and water level data were normalized to the range [0.1:0.9] using linear rescaling. Such normalization is

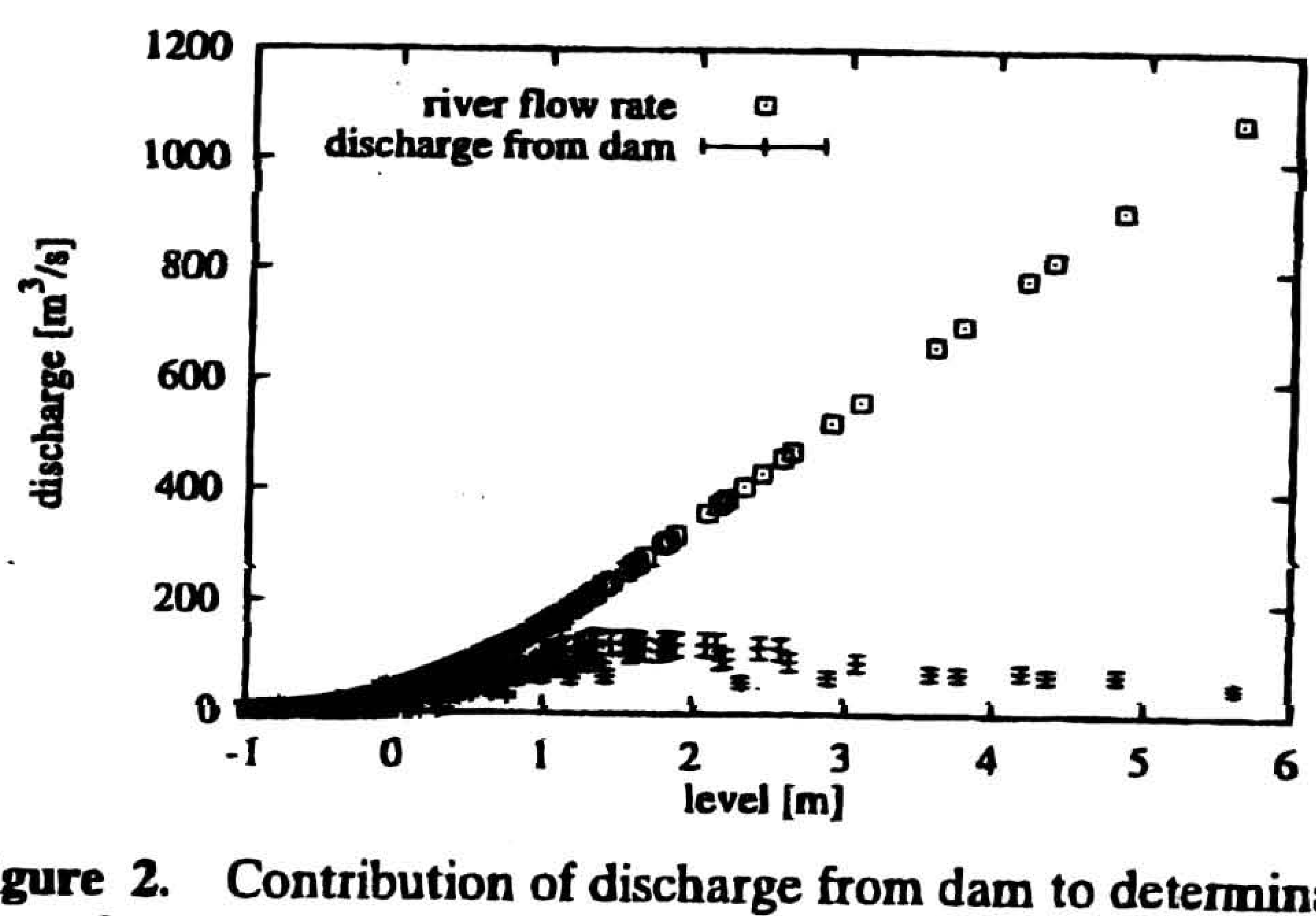

Figure 2. Contribution of discharge from dam to determination of total flow rate in river.

useful whenever input variables represent physical quantities that may have very different ranges of variation [*Bishop*, 1994].

3. Methodology

3.1. Neural Network Model

The river can be considered a complex, nonlinear system through which input variables, i.e., rainfall, water level, and power data, are transformed by nonlinear relationships into output variables, i.e., water levels for the following days. Neural network models can be profitably used whenever the phenomenon to be modeled can be described as a nonlinear mapping between multidimensional input and output domains. The arbitrarily complex mapping, even of unknown form, is extrapolated from pairs of input-output data, each one representing an example of the transformation to be modeled. Transformation of the input array of data, $\bar{I} = \{x_i, i = 1, N_I\}$, with N_I components into the output array, $\bar{O} = \{o_k, k = 1, N_O\}$, with N_O components is obtained by combining a sequence of simple computations performed in parallel by a number of elementary processing units or nodes. Components of the input array x_i are weighted, summed, and then transformed by each node using a nonlinear activation function f,

$$h_j = f\left(\sum_{i=1}^{N_I} w_{i,j}x_i - \sigma_j\right) \qquad j = 1, \cdots, N_H. \tag{2}$$

The result of this operation is an array, $H = \{h_j, j = 1, N_H\}$, the number of component N_H being equal to the number of processing internal nodes, that can be processed by a subsequent layer of nodes in the same way as the input array:

$$o_k^c = f\left(\sum_{j=1}^{N_H} w_{j,k}h_j - \sigma_k\right) \qquad k = 1, \cdots N_O. \tag{3}$$

The array calculated by the last layer of nodes, $\bar{O}_c = \{o_k^c, k = 1, N_O\}$, is the output array $\bar{O}$. The weighting parameters, $w_{i,j}$ and $w_{j,k}$, the thresholds σ_j and σ_k, and the number of processing units of internal layers N_H are the adjustable parameters of the model. During the model building phase, training, these parameters are modified (see *Haykin*, [1995] for details) and finally fixed in order to make the calculated output $\bar{O}_c$ as close as possible to the target, $\bar{O}$, thus minimizing the error of the model. Generally, and in this work, the correction of the weights is made proportional to the actual error of the model (standard back propagation error algorithm). Once the building phase is over, that is, when the error of the model decreases below a specified threshold, the model can be used to compute transformation of new series of input data.

The software Stuttgart Neural Network Simulator (SNNS), developed at the University of Stuttgart [*Stuttgart Neural Network Simulator*, 1995], (http://inf.informatik.uni-stuttgart.de:80/ipvr/bv/projekte/snns/UserManual/node1.html), was used in this work to implement and calibrate the neural model. A logistic activation function was selected for both hidden and output nodes. Training of the model was performed using the standard back propagation algorithm, decreasing the learning rate (from 0.5 to 0.05) as the training proceeded.

3.2. Water Level Forecasting Model

The aim of this work was to develop a model that on the basis of the dam release scheduling for the next 6 days predicts water levels of the river for the same time period. The structure of the neural network exploits 15 input nodes and six output nodes. In the 15 input nodes the rainfall, water level, and power data of the last 3 days and power data for the next 6 days are given. In the six output nodes the water levels for the 6 upcoming days are given. Embedding the input and output time series segments into the input and output arrays of the model was found useful to achieve good model performance at low cost [see *Cheng et al.*, 1995].

Table 1. Low-Flow Periods During 1992 and 1993

Year	Period	Number of days
1992	January 1 to March 24	83
	April 12 to September 17	158
	October 11 to October 14	3
	October 27 to November 4	8
	November 19 to December 31	42
1993	January 1 to May 14	133
	June 5 to October 7	124
	October 18 to November 7	20
	November 22 to December 23	30

Patterns, that is, couples of related input-output, were generated from the training, testing, and validation sets according to the structure of the model. A total of 400, 120, and 81 patterns were generated from the training, testing, and validation sets, respectively. The validation set was reduced with respect to the other sets to enhance the performance of the model [*Anfossi et al.*, 1995]. The number of hidden nodes for the model was fixed through a trial and error process. Starting with a model with 15 hidden nodes, organized in a single layer, different neural models were calibrated, progressively reducing the number of hidden nodes. Comparing performances obtained from these different models, the optimal configuration, corresponding to minimum error statistics, was found for the net with eight hidden nodes.

4. Results

The output of the model is the predicted water level of the river for the next 6 days. During calibration of the model the error on the output vector, representing the error in prediction over the entire time horizon, was minimized. However, to evaluate the performance of the model as the time advance of prediction increases, we considered the error computed for prediction with the same time advance. Distribution of the model error as a function of time advance of prediction was used to judge whether information presented to the model is adequate or a limit in time advance exists beyond which prediction becomes inaccurate.

Predictions of river flow rate are compared with measured data for the 1-day forecast in Figure 3a, and for the 6-day forecast in Figure 3b. Flow rates were obtained from the water levels through a stage-discharge relationship. A prevailing distribution of points in the range [0:40 m^3/s] may be noted. Lower flow conditions, more significant for the risk of pollution, appear to be frequent for the river. We further observed that the error of prediction increases as flow rates increase. In particular, the flow rate is slightly underestimated. This might be attributed to lack of information about future rainfall.

We evaluated the accuracy of the model using the root-mean-square error (RMSE), correlation coefficient, and maximum percent error. The behavior of these parameters is pre-

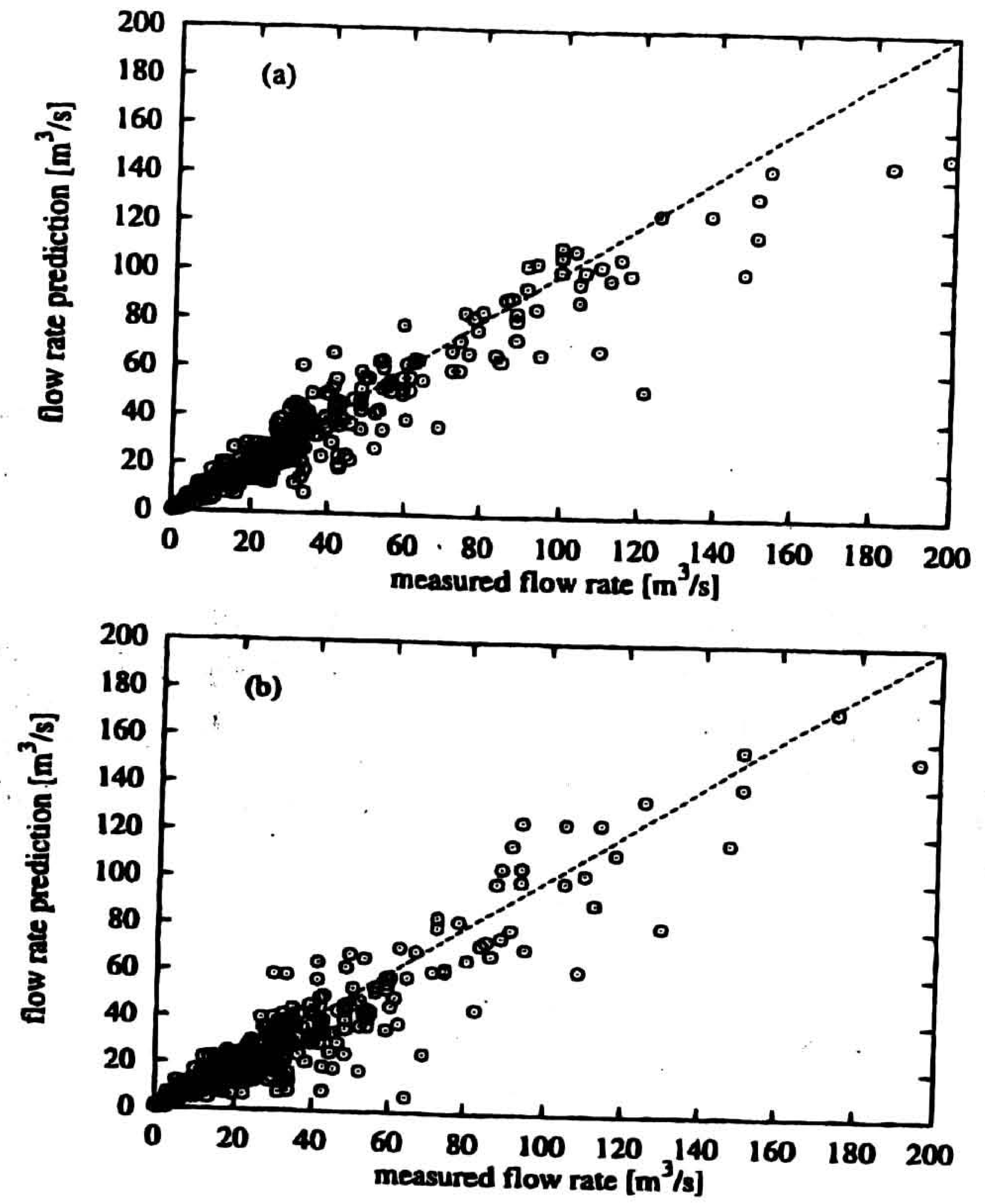

Figure 3. Comparison between predictions of river flow rate and measured data for the (a) 1-day forecast and (b) 6-day forecast.

sented in Figure 4. The RMSE, the correlation coefficient, and the maximum error percent were calculated separately on data from the calibration and validation set to assess the performance achieved during calibration, the training and testing sets, and expected from the model working on independent data, the validation set. The RMSE statistic measures the residual variance of the model and is used to quantify the error between predicted and measured flow rates. Figure 4a shows that the RMSE for the calibration set is around 5 m^3/s for prediction up to 5 days in the future. A slight increase is found on prediction at the sixth day. The value of the RMSE can be considered satisfactory since, as previously noted from Figure 3, predictions when the flow is lower, up to 20 m^3/s, are affected by a small error, whereas the error increases for larger flow rates. The RMSE computed on the validation set is 3 times larger, about 15 m^3/s, than that on the calibration set, with a steady trend around this value within the entire time horizon of prediction. Therefore the response of the model appears to be unaffected by time advance of prediction, proving that the input information is sufficient to compute flow rate within the entire time horizon. Furthermore, stability of the RMSE value within the time horizon of prediction for the different data sets, calibration and validation, proves that the model is not overtrained; that is, the law underlying the calibration data has been extracted without fitting the residual error present in the data. Figure 4b plots the correlation coefficient between calculated and measured flow rates against time advance of prediction. The correlation for the calibration set remains about 98% and never falls below 86% for the validation set. Finally, Figure 4c reports the maximum error in prediction. This was evaluated and related to the correct value of the flow rate to quantify the magnitude of the error if the predicted flow rate is used instead of the corresponding real value. Generally, but not always, a slight increase in the value may be noted for each set of data as the time of prediction increases. The error on the validation set is, as expected, greater than that on the calibration. The absolute value of the maximum error percent is rather high. However, predictions for which the error percent is near to the maximum concern a small fraction of the data base; that is, error percentages larger than 80% of the maximum value are found for 3% of all patterns only.

In Figure 5 the time series of the measured flow rate is compared with the 1-day advance, 3-day advance, and 6-day advance prediction. Periods refer to training (Figure 5a) and validation (Figure 5b). The largest errors affect those periods during which large variations of the flow rate occur. During very low flow periods the prediction seems to be satisfactory for each time delay. Mean error as a percent of predicted flow rate is about 8% within the entire time horizon of prediction.

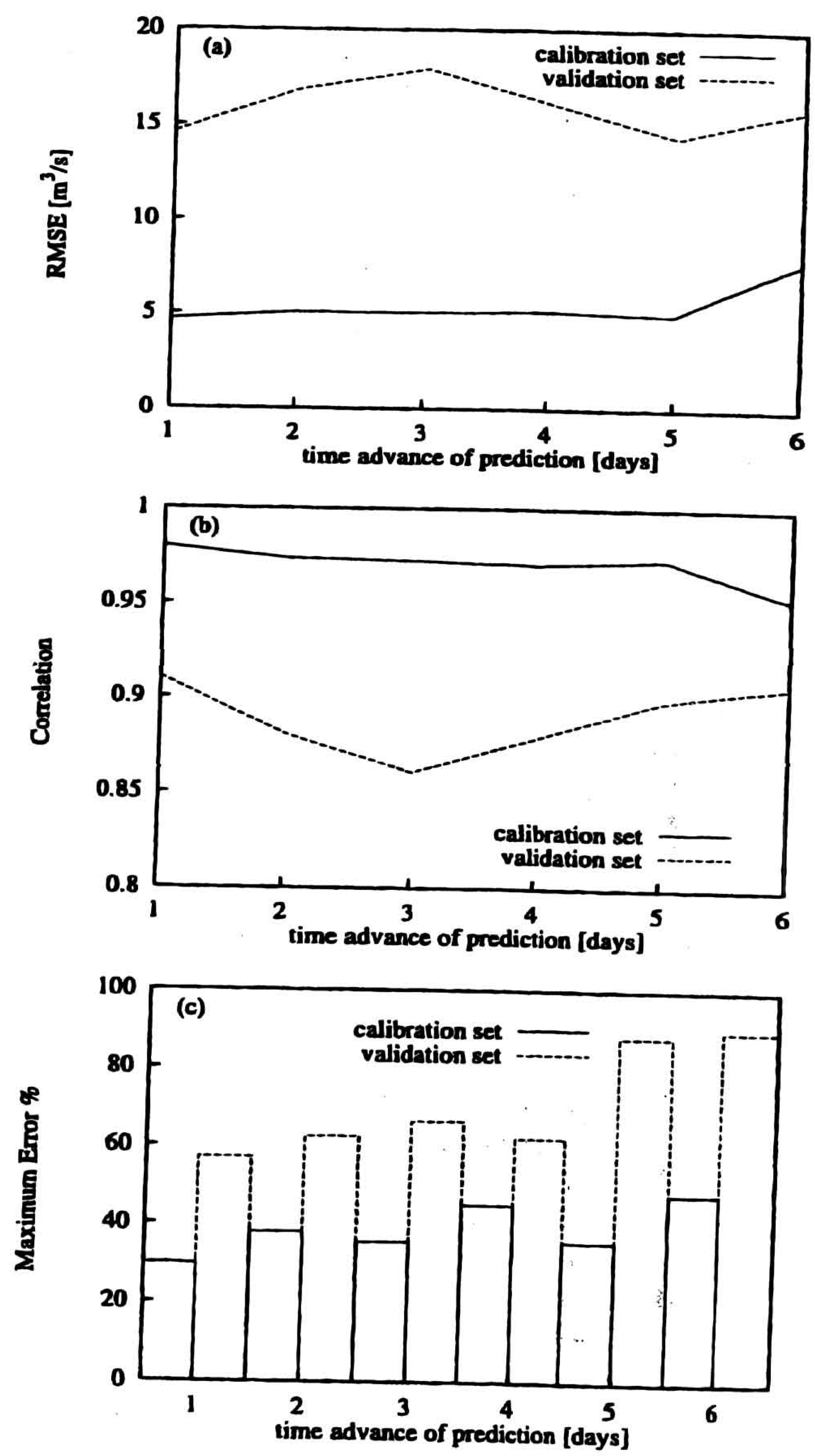

Figure 4. Performance of the model: (a) root-mean-square error, (b) correlation coefficient, and (c) maximum error percent versus time advance of prediction.

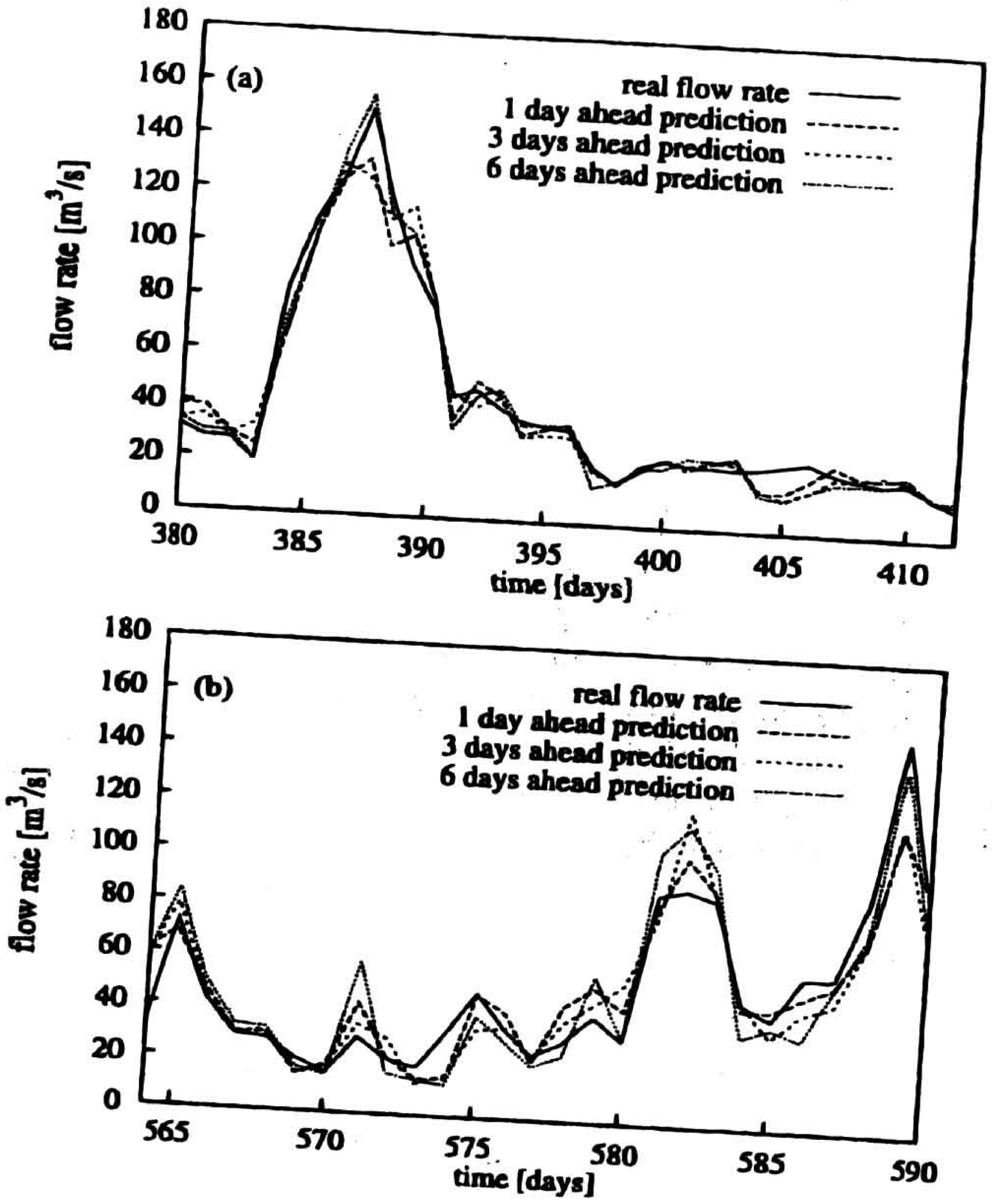

Figure 5. *Time series of river flow rate at Nave di Rosano measured and computed as 1, 3, and 6 days forward prediction for low-flow periods from (a) training set and (b) validation set.*

Results obtained during model calibration and evaluation of nodel performance on the validation set show that the model can be used to forecast the flow rate, given power production scheduling. Then, effects on flow rate of different power production schedules can be evaluated from multiple running of he forecasting model, using the same initial conditions, that is, ast water levels, rainfall, and power production data, and lifferent power production schedules.

. Conclusions

Severe pollution-related environmental consequences in the iver Arno occur especially during low-flow periods. During hese periods, river flow is not sufficient to ensure adequate ispersion and transformation rates for pollutants. A possible olution to this problem is the ad hoc use of the reservoirs resent along the course of the river, exploited for hydropower roduction. It is thus crucial to have a reliable model to forest the river flow rate to manage and control pollution in the er. The model developed in this work gives a forecast of ater level evolution for the following 6 days starting from own initial conditions for the basin and using information put rainfall and hydropower production plans. The model, sed on a neural network, has the advantages of low cost and nplicity with respect to complex physically based models. ta used in this work, collected in the basin of the Arno in the ars 1992 and 1993, are daily-averaged quantities.

Prediction of river flow rate for the following 6 days obtained m the model are satisfactory. The RMSE obtained on the lidation set is about 8% within the entire time horizon of prediction. The error on predicted flow rate is smaller at very low flow and increases with flow rate value.

The model developed in this work can be usefully exploited to select power production schedules that are optimal to manage low-flow periods responsible for an intolerable rise in pollution. Further work is in progress to show how predicted flow rates can be used to infer water quality evolution and how feedback control on scheduled power production can be implemented to optimize water resource management at basin level, ensuring adequate standards of water quality in all conditions.

Acknowledgments. Rain gauges and hydrometer data were acquired from Ufficio Idrografico e Mareografico of Pisa for 1992 and 1993. Hydropower data for the same years were obtained from ENEL.

References

Anfossi, D., C. Cappa, and S. Alessio, An application of neural nets to the Rome precipitation series, *Nuovo Cimento Soc. Ital. Fis. C.*, *18*, 547–561, 1995.

Bishop, C. M., Neural network and their applications, *Rev. Sci. Instrum.*, *65*, 1803–1832, 1994.

Buchtele, J., V. Elias, M. Tesar, and A. Herrman, Runoff components simulated by rainfall-runoff models, *J. Hydrol. Sci.*, *41*, 49–60, 1996.

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

Chakraborty, K., K. Mehootra, C. K. Mohan, and S. Ranka, Forecasting the behavior of multivariate time series using neural networks, *Neural Networks*, *5*, 961–970, 1992.

Cheng, Y., T. W. Karjala, and D. M. Himmelblau, Identification of nonlinear dynamic processes with unknown and variable dead time using an internal recurrent neural network, *Ind. Eng. Chem. Res.*, *34*, 1735–1742, 1995.

Garrote, L., and R. L. Bras, A distributed model for real-time flood forecasting using digital elevation models, *J. Hydrol.*, *167*, 279–306, 1995.

Haykin, S., *Neural Networks: A Comprehensive Foundation*, Macmillan, Indianapolis, Indiana, 1992.

Hsu, K.-L., H. V. Gupta, and S. Sorooshian, Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, *31*, 2517–2530, 1995.

Li, S.-G., F. Ruan, and D. McLaughlin, A space-time accurate method for solving solute transport problems, *Water Resour. Res.*, *28*, 2297–2306, 1992.

Maier, H. R., and D. Dandy, The use of artificial neural network model for the prediction of water quality parameters, *Water Resour. Res.*, *32*, 1013–1022, 1996.

Mason, J. C., A. Tem'me, and R. K. Price, A neural network model of rainfall-runoff using radial basis function, *J. Hydrol. Res.*, *34*, 537–548, 1996.

Minns, A. W., and M. J. Hall, Artificial neural networks as rainfall-runoff models, *J. Hydrol. Sci.*, *41*, 399–417, 1996.

Mukherjee, D., and N. Mansour, Estimation of flood forecasting errors and flow-duration joint probabilities of exceedance, *J. Hydrol. Eng.*, *122*, 130–140, 1996.

Raman, H., and N. Sunilkumar, Multivariate modelling of water resources time series using artificial neural networks, *J. Hydrol. Sci.*, *40*, 145–163, 1995.

Refsgaard, J. C., and J. Knudsen, Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*, 2189–2202, 1996.

Shamseldin, A. Y., Application of a neural network technique to rainfall-runoff modeling, *J. Hydrol.*, *199*, 272–294, 1997.

Smith, J., and R. N. Eli, Neural network models of the rainfall-runoff process, *J. Water Resour. Plan. Manage.*, *121*, 499–508, 1995.

Stuttgart Neural Network Simulator, User manual, version 4.0, *Rep. 6/95*, Inst. for Parallel and Distrib. High Perform. Syst., Stuttgart, Germany, 1995.

Todini, E., The ARNO rainfall-runoff model, *J. Hydrol.*, *175*, 339–382, 1996.

Walton, R., and M. Webb, QUAL2E simulations of pulse loads, *J. Environ. Eng.*, *5*, 1017–1031, 1994.
Woolhiser, D. A., Search for physically based runoff model-A hydrologic El Dorado?, *J. Hydrol. Eng.*, *122*, 122–129, 1996.
Zhu, M. L., M. Fujita, and N. Hashimoto, Application of neural networks to runoff prediction, in *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, edited by K. W. Hipel et al., pp. 205–216, Kluwer Acad., Norwell, Mass., 1994.

P. Andreussi, Dipartimento di Chimica e Chimica Industriale, Università di Pisa, via Risorgimento 2, 56100 Pisa, Italy. (p.andreussi@cpr.it)

M. Campolo and A. Soldati, Centro Interdipartimentale di Fluidodinamica e Idraulica, Università di Udine, Via del Cotonificio 108, 33100 Udine, Italy. (marina@euterpe.dstc.uniud.it; alfredo@euterpe.dstc.uniud.it)

(Received September 22, 1998; revised June 29, 1999; accepted June 30, 1999.)