Input–output dynamic neural networks simulating inflow–outflow phenomena in an urban hydrological basin

Orazio Giustolisi

ABSTRACT

In this paper neural networks have been studied as a tool to realise a single-input single-output nonlinear dynamic system simulating rainfall-runoff transformation in a urban hydrological basin. The aim is to test the performance, in simulation and real time forecasting, of these models when compared to single-input single-output linear dynamic systems with a stochastic process as forecasting component. For this reason, the impulse unit hydrograph, the transfer function of the deterministic component of such linear models, and the stochastic process have been calculated by means of the experimental data (59 events of rainfall-runoff) and, similarly, the identification procedure of the best nonlinear model was performed. The comparison between linear and nonlinear models was achieved by computing the *estimated mean generalisation error* and by performing statistical tests by means of *cross-correlation* and *auto-correlation functions*, using *cross-validation* techniques.

Key words | rainfall-runoff modelling, neural networks, nonlinear dynamic systems

Orazio Giustolisi
Polytechnic of Bari,
II Faculty of Taranto,
Via Turismo no 8,
Paolo VI,
74100 Taranto,
Italy
E-mail: *o.giustolisi@poliba.it*,
oraziogiustolisi@libero.it

INTRODUCTION

The inflow–outflow transformation phenomenon in urban basins is an interesting issue in technical hydrology because it summarises every hydrological and hydraulic process that affects the discharge in a generic section of the urban drainage network due to the inflow in the upstream basin. One category of mathematical models simulates rainfall-runoff transformation by means of the Impulse Unit Hydrograph (IUH). These models belong to the Single-Input Single-Output Linear Time Invariant (SISO LTI) causal dynamic systems class (Giustolisi 1998). Using these linear models to simulate physical inflow–outflow transformation implies the approximation that physical phenomena in urban basins are linear and time independent.

This means that a single Unit Hydrograph is able to describe the hydrological processes during all the possible events of rainfall, neglecting the instant physical conditions of the urban basin. It is known, however, that such a physical system is quite complex and the previous approximations can be introduced only as useful tools to build a simplified mathematical model of the rainfall-runoff transformation.

In the last few years various authors (Giustolisi & Mastrorilli 1994; Hsu *et al*. 1995; Smith & Eli 1995; Minns & Hall 1996; Mason *et al*. 1996; ASCE Task Committee 2000; Minns 2000; Abrahart *et al*. 1999) modelled hydrological systems by means of neural networks. In the present paper the author uses Input–Output Dynamic Neural Networks as a special Single-Input Single-Output NonLinear Time Invariant (SISO NLTI) dynamic system (Haykin 1999). For this reason, a special subclass of neural networks, reported in the scientific literature as ARX neural networks (NARX) (Billings & Chen 1992; Sjöberg *et al*. 1994), was tested using data of rainfall-runoff events measured in the experimental urban basin of Luzzi in the South of Italy (Calomino *et al*. 1993). In building the model, the choice of the neural network structure is a critical point since the range of options is wider than for linear dynamic systems. Additionally, the parameters estimation of neural networks, by means of the error function

as a goal function, is a rather complex inverse problem because the surface of the error function always presents local minima. The selection of the NARX structure was justified by its linear regressor as input and by the fact that it is not a recursive structure (the computed output is not applied as input). These characteristics make the NARX structure, among all possible architectures of neural networks, easier to handle from the mathematical point of view.

SISO LTI DYNAMIC SYSTEMS

At first, linear dynamic systems with a stochastic forecasting component have been chosen (Ljung 1987; Åström & Wittenmark 1995; Haykin 1996) to model rainfall-runoff phenomena in a urban catchment:

$$\begin{cases} q_M(n) = \sum_{k=1}^{\infty} h(k) p_M(n-k) + v(n) = \dots \\ \dots = \sum_{k=1}^{\infty} IUH(k) p_M(n-k) + v(n) = q_D(n) + v(n) \\ v(n) = s(n) + \sum_{k=1}^{\infty} g(k) s(n-k) \end{cases} \quad (1)$$

In Equations (1), $q_M(n)$ and $p_M(n)$ are the measured discharge and rainfall at time step n. The discharge, in the discrete domain, is calculated as the sum of the two values:

- the first addendum (deterministic framework) is the convolution sum between $h(n)$, the impulse response of linear system, then the unit hydrograph (IUH), and the measured rainfall;
- the second addendum (probabilistic framework for prediction) is a stochastic process; $g(n)$ is its impulse response and $s(n)$ is an input sequence of independent random variables distributed according to a gaussian probability density function with zero mean and variance λ (white noise).

The disturbance $v(n)$ globally expresses the fraction of the actual runoff which cannot be modelled by the deterministic framework, also due to the hypothesis of linear time invariant behaviour (Figure 1).

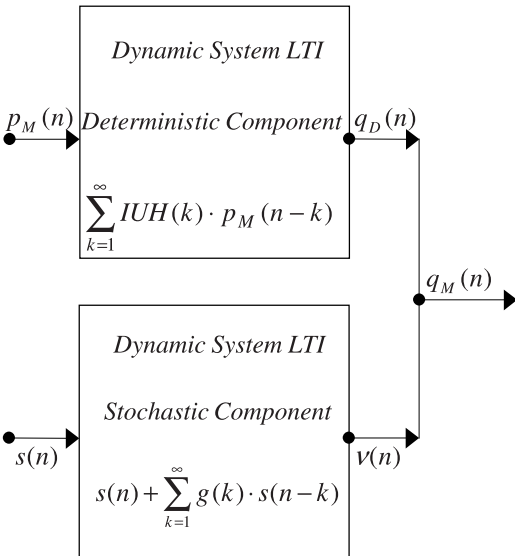

Figure 1 | Linear dynamic system having the stochastic process to forecast $v(n)$.

Parametrised form

Usually the transfer functions $IUH(n)$ and $g(n)$ are parametrised by means of rational functions and the general mathematical representation of such a parametric form is (Ljung 1987)

$$\begin{aligned} &A(z,\theta) q_M(n) = \frac{B(z,\theta)}{F(z,\theta)} p_M(n) + \frac{C(z,\theta)}{D(z,\theta)} s(n) \\ &A(z,\theta) = 1 + \sum_{i=1}^{na} a_i z^{-i} \qquad B(z,\theta) = \sum_{i=1}^{nb} b_i z^{-i-nk+1} \\ &C(z,\theta) = 1 + \sum_{i=1}^{nc} c_i z^{-i} \qquad D(z,\theta) = 1 + \sum_{i=1}^{nd} d_i z^{-i} \\ &F(z,\theta) = 1 + \sum_{i=1}^{nf} f_i z^{-i} \end{aligned} \quad (2)$$

Equations (2) are written in the z-transform domain, see the expressions of polynomials, and changing the number of parameters in the polynomials A, B, C, D, F, parametrised in θ [the vector of parameters: see, for example, Equation (4)] it is possible to describe different model structures (Ljung 1987; Giustolisi 1998), see Table 1, which depend on the dynamic relationship between the probabilistic and deterministic components [note that the z transform is consistent with the q backward shift operator in Box & Jenkins (1989)].

Table 1 | Some model structures of a linear dynamic system with stochastic forecasting. FIR (Finite Impulse Response), MA (Moving Average), AR (AutoRegressive), X (eXogeneous variable or eXtra input).

na	*nb*	*nc*	*nd*	*nf*	Model
$=0$	>0	$=0$	$=0$	$=0$	FIR
$=0$	$=0$	>0	$=0$	$=0$	MA
>0	$=0$	$=0$	$=0$	$=0$	AR
>0	$=0$	>0	$=0$	$=0$	ARMA
>0	>0	$=0$	$=0$	$=0$	ARX
>0	>0	>0	$=0$	$=0$	ARMAX
$=0$	>0	$=0$	$=0$	>0	Output-Error
$=0$	>0	>0	>0	>0	Box–Jenkins

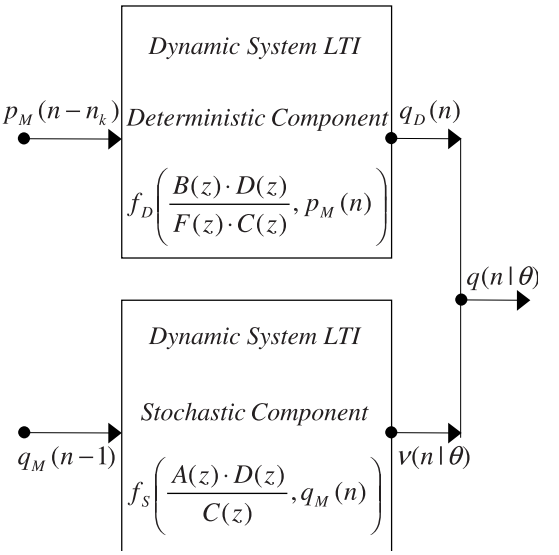

Figure 2 | General linear one-step-ahead prediction model (SISO LTI) which performs online forecasting by means of discharge $q_M(n-1)$ measured up to $n-1$, and rainfall $p_M(n-nk)$ measured up to $n-nk$.

The parameter nk represents the delay between input p_M and output q_M of the linear models formalised by Equations (2).

One-step-ahead prediction of SISO LTI models

The general representation of the one-step-ahead prediction of linear black-box models reported in Table 1 is (Ljung 1987)

$$q(n|n-1,\theta)=q(n|\theta)=\frac{B(z,\theta)D(z,\theta)}{F(z,\theta)C(z,\theta)}\,p_M(n)+\left[1-\frac{A(z,\theta)D(z,\theta)}{C(z,\theta)}\right]q_M(n) \qquad (3)$$

$$\varepsilon(n,\theta)=q_M(n)-q(n|\theta)$$

The models of Equation (3) (Figure 2) represent a special case of k-step-ahead prediction (Ljung 1987; Giustolisi 1998) and they are useful in estimating parameters minimising an error function of $\varepsilon(n,\theta)$ which is the difference between the measured discharge and the one-step-ahead prediction of discharge $q(n|\theta)$. These forecasting performances are generally obtained by improving the model online, Figure 2, thanks to the measured q_M.

ARX is a particularly useful structure of Table 1 for estimating parameters in an easy way, because the mean square one-step-ahead error surface is parabolic with no local minima.

This special structure is a linear regression among measured data:

$$q(n|\theta)=B(z,\theta)p_M(n)+[1-A(z,\theta)]q_M(n)$$

$$\varphi(n)=[-q_M(n-1)\ \ldots\ -q_M(n-na)\quad p_M(n-nk)\ \ldots\ p_M(n-nb+nk-1)] \qquad (4)$$

$$\theta=[a_1\ a_2\ \ldots\ a_{na}\ b_1\ b_2\ \ldots\ b_{nb}]^T$$

$$q(n|\theta)=\varphi(n)\theta \qquad \varepsilon(n,\theta)=q_M(n)-\varphi(n)\theta$$

and for this reason the regressor vector $\varphi(n)$ depends on known data and not on parametrisation (parameters in θ).

SISO NLTI: NEURAL NETWORKS

Looking at Equations (3) and (4), it is possible to write a general form of the one-step-ahead prediction model (Ljung 1987):

$$q(n|\theta) = \varphi(n,\theta)\theta = W_p p_M(n) + W_q q_M(n) \quad (5)$$

where transfer functions W_p and W_q are selected by the first of Equations (3), which represents a pseudo-linear regression because the regressor vector $\varphi(n)$ depends on the parameters in θ. Equation (5) can be fully generalised:

$$q(n|\theta) = G\{\varphi[p_M(n-nk),\, q_M(n),\, f(\theta)],\, \theta(n)\} \quad (6)$$

Equation (6) is the general nonlinear time varying formalisation of the one-step-ahead prediction model in which G depends on the regressor vector φ, function of measured data and the parameters in θ, and on the same vector θ which varies in time step n (Sjöberg *et al*. 1994).

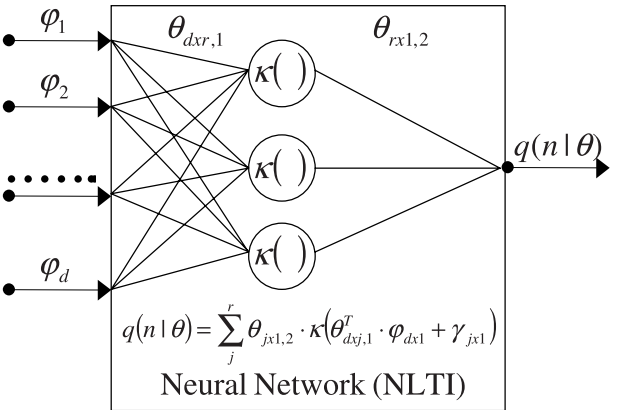

Figure 3 | General nonlinear one-step-ahead prediction model (SISO NLTI) constructed by a neural network.

NARX models

Equation (6) can be simplified by choosing, as regressor vector, the pseudo-linear regressor from linear model structures of Table 1 and working with time invariant systems (Billings & Chen 1992; Sjöberg *et al*. 1994):

$$q(n|\theta) = G\{\varphi(n,\theta),\, \theta\} \quad (7)$$

The further step is to define the description of G by means of neural networks.

One possible choice is the structure:

$$q(n|\theta) = G\{\varphi(n,\theta),\theta\} = \sum_j^r \alpha_j \varkappa(\varphi,\beta_j,\gamma_j) \quad (8)$$

where the third term of Equation (8) is the expansion of G by means of a mother basis function $\varkappa(\ )$ dependent on the regressor vector φ, r scale parameters β and r translation parameters, γ, as reported in Figure 3 where, for example, $r = 3$.

Moreover, the NARX one-step-ahead prediction model (Figure 4) has been carried out by means of a neural network using:

- regressor vector $\varphi(n)$ of ARX structure [Equations (4), Table 1] as input ($nk = 1$);
- parameters θ as synaptic weights;
- r transfer functions, in the neurons of the hidden layer, representing the r basis functions which can be different;
- the prediction $q(n|\theta)$ as a single linear output.

In this investigation the NARX model was chosen because, as is evident in Figure 4, this kind of neural network is not recurrent, the predictor having no feedback from output to input (future inputs will not depend on present and future outputs). This makes it easier to estimate parameters, the weights of the neural network, because the model is more stable during the training phase and use.

NARX PARAMETERS ESTIMATION

Parameters estimation of NARX models has been performed by minimising the mean square error function of the one-step-ahead prediction with a *Tikhonov regularisation* (Tikhonov 1963; Haykin 1999) term given by a function of the squared distance of the parameters vector from the origin:

$$W_N(\theta) = V_N(\theta) + \frac{\alpha|\theta|^2}{2N} = \sum_{n=1}^{N} \frac{[q_M(n) - q(n|\theta)]^2}{2N} + \frac{\alpha|\theta|^2}{2N} \quad (9)$$

where N is the number of data and α is a parameter which prevents overtraining due to overparametrisation.

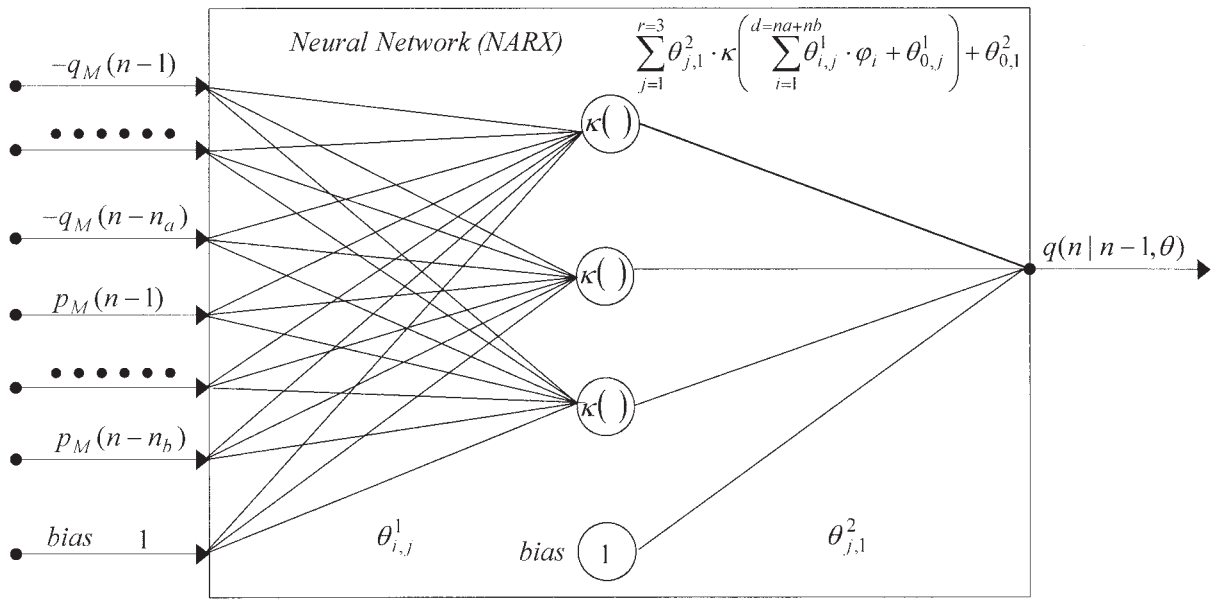

Figure 4 | NARX NLTI one-step-ahead prediction model constructed by a neural network. $p_M(n)$ and $q_M(n)$ are mea sured sequences and '1' represents a bias which makes the model more flexible.

Regularisation has a smoothing effect on the criterion of Equation (9); in this way, several local minima are often removed by it. The weights have been computed as

$$\hat{\theta}=\arg\min_{\theta}\left(\sum_{n=1}^{N}\frac{[q_M(n)-q(n|\theta)]^2}{2N}+\frac{\alpha|\theta|^2}{2N}\right) \tag{10}$$

An iterative search of the solution by the Gauss–Newton-based Levemberg–Marquardt method has been performed:

$$\begin{aligned}
&\theta^{i+1}=\theta^i+\mu f^i=\theta^i-[V_N^{ll}(\theta^i)]^{-1}V_N^{l}(\theta^i)\\
&V_N^{ll}(\theta)=\frac{1}{N}\sum_{n=1}^{N}\psi(n,\theta)\psi^T(n,\theta)+\lambda I\\
&V_N^{l}(\theta)=-\frac{1}{N}\sum_{n=1}^{N}\psi(n,\theta)\varepsilon(n,\theta)\\
&\psi(n,\theta)=\frac{d}{d\theta}q(n|\theta)
\end{aligned} \tag{11}$$

The parameter λ, in the second of Equations (11), avoids the Levemberg–Marquardt approximation of the Hessian, by first order derivative terms, to be singular. The choice of the λ term was adaptively performed (Fletcher 1987; Nørgaard 1996), obtaining:

- the search direction approaching the steepest descent direction and $V^{ll}(\theta^i)^{-1}$ (step size) decreasing while increasing λ;
- the search direction approaching the Gauss–Newton direction and $V^{ll}(\theta^i)^{-1}$ (step size) increasing while decreasing λ.

An alternative to the regularisation method to avoid overtraining was tested. It is known as Early Stopping. It consists of stopping the iterative parameters searching before the minimum is reached. This can be achieved by computing an error function of the validation set, which is a subset of the measured data that are not used to estimate parameters.

Details on estimation phase

The training or estimation subset of data has been presented, coherently with the hypothesis of time invariant

behaviour, lining up sequences of runoff-rainfall events separated by zeros to avoid changes, due to interference among sequences of events, of the initial zero conditions (Giustolisi 1998). Moreover, during the estimation phase, data were scaled and translated in the range [0:1] and the weights were re-scaled after training, allowing original data to be used directly in the neural network.

Bias of the model was not used by zeroing their weights, because the physical initial conditions of rainfall-runoff events did not indicate that they were required.

This estimation of parameters does not guarantee global minima reaching.

For this reason, the choice of initial weights is very important. These were assigned randomly in the range $[-0.5:0.5]$ and the neural network was trained twice with different initial random configurations of weights.

Finally, the parameters estimation of ARX models, indeed, is not iterative. It was obtained by means of *Gauss normal equations* performing the inversion of a square matrix based on the regressor. In this case the overparametrisation causes ill conditioning of the problem.

IDENTIFICATION AND VALIDATION

The identification of the best ARX and NARX models to be compared has been performed by means of *cross-validation*. The whole set of data was divided into two subsets: the estimation or training subset, and the validation or test subset. The cross-validation, as previously said, allowed us to perform Early Stopping during the iterative parameter estimation of the NARX model.

Model selection

Among several ARX and NARX models, varying the number of parameters, the selection of the best ARX model and the best NARX model was performed by means of a validation set.

This set has been used to evaluate the *mean generalisation error of the simulated output* (as input of the models substituting measured with computed discharges) and to execute *correlation based validation*.

ARX

Parameters were estimated by varying na and nb in the range of [1:10] and the best model was selected by computing the mean generalisation error of the simulated output by validation data. Delays between rainfall and runoff for each event were evaluated selecting the best Output-Error model ($nb = 2$, $nf = 2$, $nk = [1:20]$). Then, during computations, the parameter nk was always picked unitary (both in ARX and NARX models) by translating each of the measured runoff sequences backward thanks to the estimated delay.

NARX

The regressor $\varphi(n)$ has been chosen assuming na and nb in the range of [1:10], and varying hidden neurons in the range of [1:10]. Then 100 regressors and 1,000 NARX models were tested. The hyperbolic tangent was the transfer function of the hidden neurons and the output neuron was linear.

By computing, at each iteration, the mean generalisation error of the simulated output, thanks to the validation data, estimation of the parameters has been stopped early.

In fact, the best NARX model was selected by means of validation data computing generalisation error of the simulated output and by means of correlation based validation.

As an alternative to Early Stopping, *regularisation* has also been tested without yielding cross-validation.

Correlation-based validation

The correlation-based methods perform validation estimating some particular correlation functions of the residuals vector $\varepsilon(n,\theta)$ and the input vector $p_M(n)$. The idea was to perform statistical tests on the hypothesis that

such functions belong asymptotically to a *normal distribution* $N(\ )$ which has its variance to be estimated and *zero* mean.

Residuals analysis

The first test regards the auto-correlation sequence of the one-step-ahead prediction error $\varepsilon(n,\theta)$ (residuals) of the ARX and NARX models. In fact, if the deterministic component of the model is good, residuals should be mutually independent if the stochastic component is able to give a correct description of disturbance $v(n)$. This means that the auto-correlation function of residuals, $R_{\varepsilon\varepsilon}$, should be zero for lags τ greater than zero, and equal to the variance λ at lag *zero*. The statistical test was achieved by estimating the auto-correlation from lag *zero* to M, see the first of Equations (12). Then, the auto-correlation function at lag *zero* allowed us to estimate the variance of the *normal distribution*, see the third of Equations (12). Finally, the test was performed by controlling that the auto-correlation of residuals at lags greater than zero belongs to the confidence interval $(1-\alpha)$, see the fourth of Equations (12). This level depends on the significance level α of the test (the probability of rejecting the hypothesis of independence among residuals) which is the level of the *normal distribution*:

$$\begin{aligned}
&R_{\varepsilon\varepsilon}(\tau,\theta)=\sum_{n=\tau+1}^{N}\frac{\varepsilon(n,\theta)\varepsilon(n-\tau,\theta)}{N} \qquad \tau\in[0:M]\\
&\sqrt{N}R_{\varepsilon\varepsilon}(\tau,\theta)\in AsN(0,P)\\
&P=R_{\varepsilon\varepsilon}(0,\theta)R_{\varepsilon\varepsilon}(0,\theta)=\lambda^2\\
&R_{\varepsilon\varepsilon}(\tau,\theta)\in\sqrt{\frac{P}{N}}N_{\alpha}(0,1)[-1:1] \qquad \forall\tau>0
\end{aligned} \tag{12}$$

Input-residuals analysis

The second test regards the cross-correlation sequence between residuals $\varepsilon(n,\theta)$ of the ARX and NARX models and input vector $p_M(n)$.

In fact, if the deterministic component of the model is able to reproduce the deterministic fraction of the output, residuals and input should be independent because otherwise there would be more in the output that originates from the input and that the model has not picked up. This means that the cross-correlation function $R_{\varepsilon p}$ between residuals and input vectors should be zero for lags τ greater than zero (delay *nk* was picked to be unitary by translating rainfall-runoff events). There is no hypothesis on whiteness of residuals.

Similarly to the previous statistical test, the cross-correlation was estimated by the first of Equations (13) from lag 1 to M.

Then the auto-correlation of residuals and of input sequences allowed us to estimate the variance of the *normal distribution*, see the fifth of Equations (13). Finally, the test was performed by controlling that the cross-correlation at lags greater than zero belong to the confidence interval (1-α), see the sixth of Equations (13). This level depends on the significance level α of the test (the probability of rejecting the hypothesis of independence between residuals and input sequences) which is the level of the *normal distribution*:

$$\begin{aligned}
&R_{\varepsilon p}(\tau,\theta)=\sum_{n=\tau+1}^{N}\frac{\varepsilon(n,\theta)p(n-\tau)}{N} \qquad \tau\in[1:M]\\
&R_{\varepsilon\varepsilon}(\tau,\theta)=\sum_{n=\tau+1}^{N}\frac{\varepsilon(n,\theta)\varepsilon(n-\tau,\theta)}{N} \qquad \tau\in[-M:M]\\
&R_{pp}(\tau)=\sum_{n=\tau+1}^{N}\frac{p(n)p(n-\tau)}{N} \qquad \tau\in[-M:M]\\
&\sqrt{N}R_{\varepsilon p}(\tau,\theta)\in AsN(0,P)\\
&P=\sum_{\tau=-M}^{M}R_{\varepsilon\varepsilon}(\tau,\theta)R_{pp}(\tau)\\
&R_{\varepsilon p}(\tau,\theta)\in\sqrt{\frac{P}{N}}N_{\alpha}(0,1)[-1:1] \qquad \forall\tau>1
\end{aligned} \tag{13}$$

RESULTS

The best NARX and ARX models were realised to simulate inflow–outflow phenomena in the urban basin of Luzzi in the South of Italy, see Table 2 [for more details see the book by Calomino *et al.* (1993)]. For this reason,

Table 2 | Some characteristics of the Luzzi basin. *Sr*, *Si*, *Sp* are the roof, impervious and pervious parameters. *T* are statistical parameters about duration.

Area km^2	Sr %	Si %	Sp %	Tmean min	Tmax min	Tmin min
0.02	63	28	9	54	140	8

Table 3 | Global mean generalisation errors $k=\infty$, 5 of the validation set.

na	nb	nk	r	SSE_∞	SSE_5	Model
1	9	1	/	0.1978	0.2372	ARX
3	2	1	5	0.1397	0.1297	NARX

59 measured rainfall-runoff events from that urban catchment area have been used.

Table 3 reports the characteristics of the best ARX and NARX models and the corresponding mean generalisation error of the simulated output (k-step-ahead prediction with $k=\infty$), computed from the validation subset by

$$SSE_{k=\infty} = \sum_{n=1}^{Nv} \frac{[q_M(n) - q_D(n)]^2}{2Nv} \tag{14}$$

where Nv is the length of the validation subset and $q_D(n)$ represents the sequence that was computed as the simulated output (the deterministic component of Figure 2). It was calculated by substituting the measured input $q_M(n)$ with the computed $q(n|\theta)$, see Figure 4.

Tables 3 and 4 show that the mean generalisation errors of the NARX model are smaller than the ARX model but simulation of some events of the validation set is not improved. It is important to underline that many other NARX models (among the 1,000 estimated) had this characteristic.

Figures 5 and 6 report the correlation based validation computed by the validation subset. They show that the best NARX model is better than the best ARX model also from this point of view, because residuals are: less autocorrelated (the prediction component is good); and rather independent from the rainfall sequence. This means that, within the level of significance of the tests, the NARX model is a good model.

In Table 3, moreover, it is reported that the mean generalisation error of the five-step-ahead prediction model (Ljung 1987), obtained from Equation (14) by substituting $q_D(n)$ with $q(n|n-k=5, \theta)$, the five-step-ahead prediction of runoff. The choice of $k=5$ (sampling interval of data was 1 minute) was related to the fact that the mean calculated delay among the events was about 5 minutes. This generalisation error shows that the NARX model is also good in online prediction. The same computed for ARX is high; in fact, it was selected by means of the best simulated output error because the aim was not to have a good stochastic component.

Finally, Figure 7 reports the comparison among measured and computed runoff by the ARX and NARX

Table 4 | Mean generalisation errors of the 17 rainfall-runoff events of the validation set.

Event	1	2	3	4	5	6	7	8	9
ARX	0.2621	0.2544	0.2570	0.0315	0.0140	0.0486	0.2009	1.6191	0.0611
NARX	0.1831	0.2101	0.1562	0.0153	0.0060	0.0313	0.4031	1.1659	0.1230
Event	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	
ARX	2.4814	0.2989	0.0899	0.0692	0.0489	0.0326	0.0558	0.0059	
NARX	0.3797	0.5293	0.1145	0.0383	0.0345	0.0307	0.0571	0.0045	

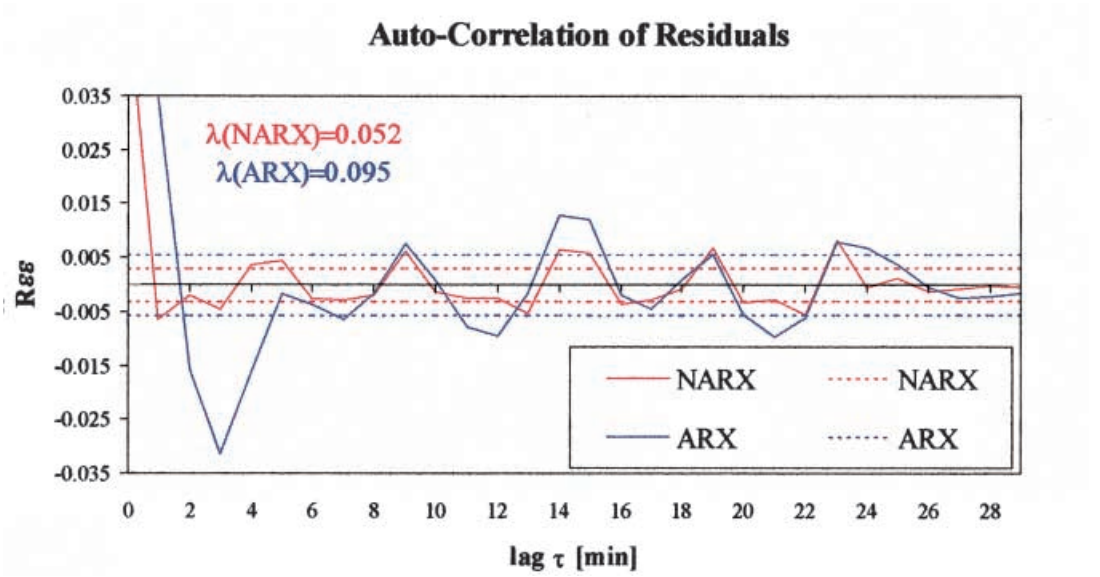

Figure 5 | Auto-correlation of residuals of the best NARX and ARX models. Horizontal lines indicate confidence level assumed at 99%.

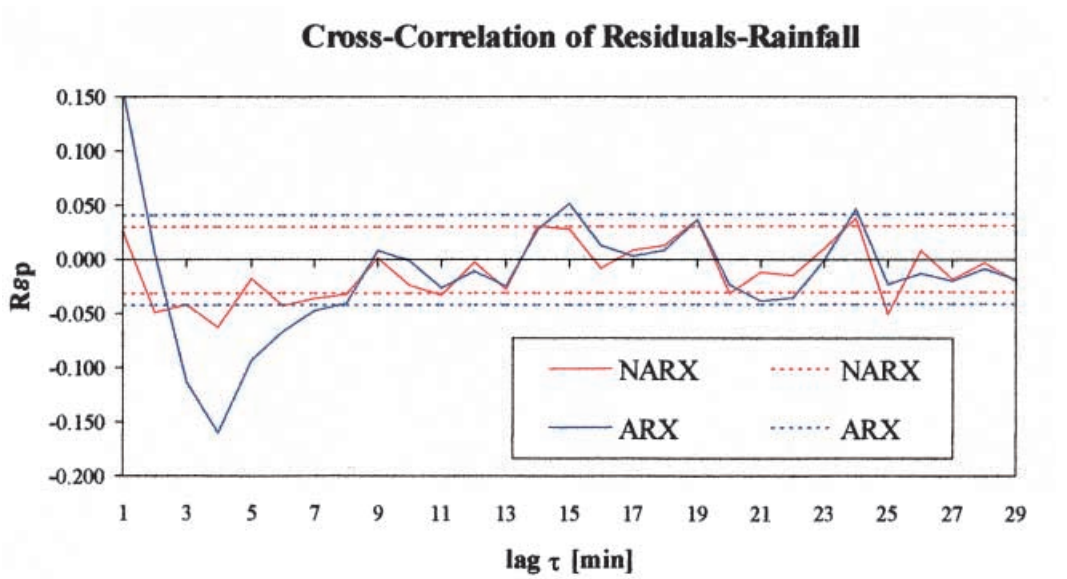

Figure 6 | Cross-correlation between residuals, of the best NARX and ARX models, and measured rainfall sequence. Horizontal lines indicate confidence level assumed at 99%.

models. Four representative events belonging to the validation subset were chosen. Figure 7 points out the capacity of the NARX model of simulating the single event and, particularly, its improved efficiency in predicting higher values of discharge. From a computational point of view, it is important to stress that these results were carried out estimating models by means of Early Stopping that was performed thanks to the validation subset, as previously said.

The Early Stopping technique also showed better performance than Tikhonov regularisation because the data set was large and it was possible to select a large validation

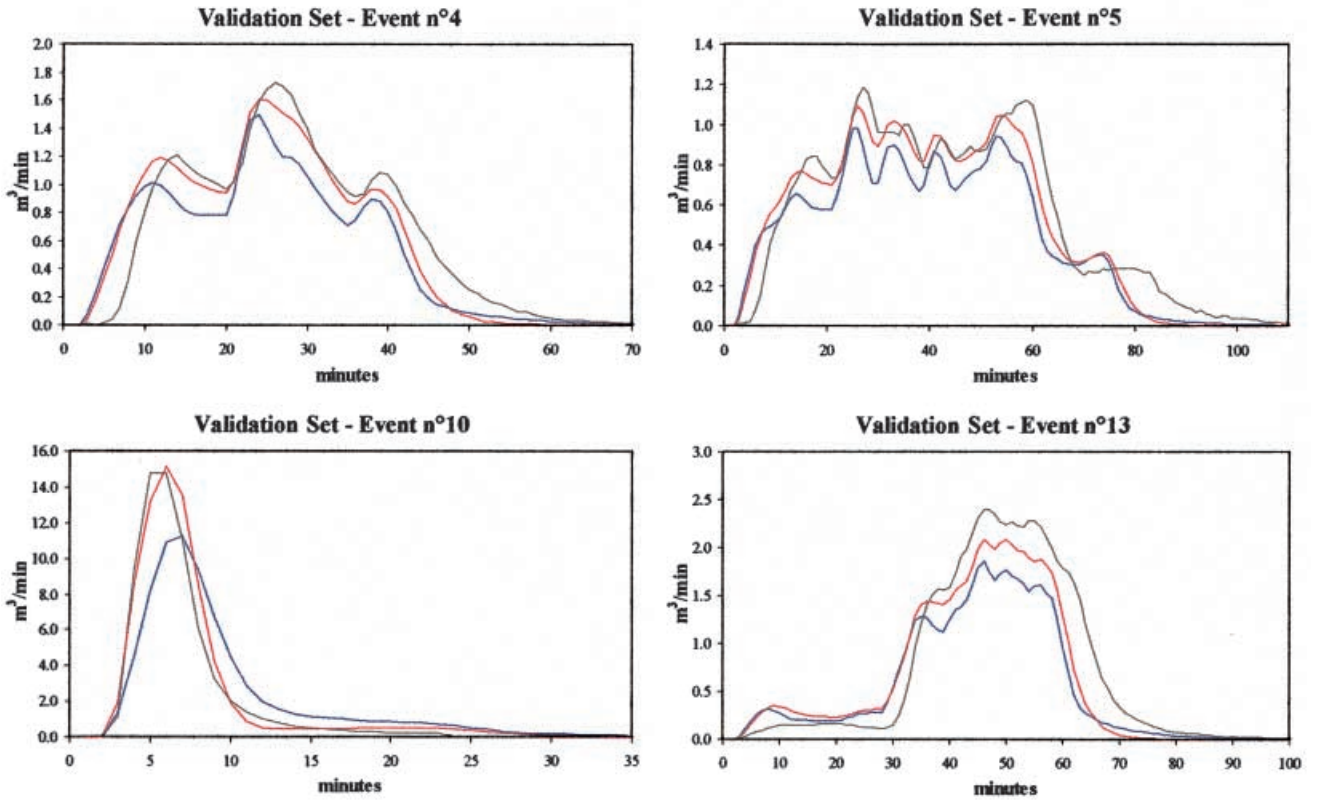

Figure 7 | Simulation of the validation set, NARX (red line), ARX (blue line) and data (black line).

subset to efficiently estimate the mean generalisation error during parameters searching. In fact, the validation subset was fixed at 30% of the whole data set. Seventeen rainfall-runoff events were used to calculate the simulated output for Early Stopping, and then to estimate the mean generalisation errors and correlation functions for validation and model selection among NARX systems.

CONCLUSIONS

In conclusion, neural networks have showed good performance in simulating nonlinear physical phenomena of a urban catchment area. The comparison between linear (ARX) and nonlinear (NARX) dynamic models was important because the second type are more flexible but also more difficult to estimate and identify. For this reason, it underlines that the validation phase is very important in selecting the best NARX structure and the correlation-based analysis is a powerful tool for this aim.

It is useful to stress that linear modelling, by means of one single unit hydrograph describing the physical inflow–outflow phenomena in the hydrological basin, must always be the first choice and the nonlinear modelling the second choice to improve model performance when non-linear effects, such as soil absorption, evapotranspiration, etc., are very significant. Actually, linear modelling in the Luzzi basin (Giustolisi 1998) is efficient because the basin has a large impermeable surface, which was the reason for choosing this basin to have a powerful test of nonlinear modelling by neural networks.

It has to be emphasised that in this work only a particular neural network structure with the ARX regressor was tested and Table 4 shows the non-constant simulation capabilities of such models while the

cross-correlation function of Figure 6 seems to indicate, at lower lags, that this nonlinear simulation picked better than linear the deterministic behaviour but it could be improved. The first choice of NARX structure was done because it is easy to estimate and more stable than other more general and flexible structures. In fact, it was verified (perhaps also thanks to Early Stopping in the estimation phase), during the estimation and simulation phases, that NARX models never exhibited unstable behaviour.

Other neural network structures (different regressors from Table 1 or different types of neural networks) could be more flexible but more complex to deal with; here we remark that the NARX model has the same regressor of Output-Error model (Table 1) when used in simulation.

Moreover, it is to underline that this kind of Input–Output Dynamic Neural Networks are a flexible mathematical tool to realise a compact model useful, for example, in online prediction for real time control.

However, the author thinks that, before selecting one specific neural network, it is necessary to study the various structures from a mathematical point of view, and then their relationship with physical behaviour. In fact, a common idea about neural networks of being a 'magic tool' that is able to solve every modelling problem without physical insight is not realistic. Physical insight about the phenomena to model always remains important in order to choose the best structure. For this reason the future goal should be to insert directly physical knowledge into the structure of neural networks because this will also be an improvement in parameters estimation and validation phase.

Finally, the scope is to choose, by means of input–output dynamic neural networks, among infinite nonlinear dynamics what is the best for the specific physical phenomenon and this is done through experimental data. For this reason it is possible to infer that the quality of these data should play a very important rule in these kinds of models.

REFERENCES

Abrahart, R. J., See, L. & Kneale, E. 1999 Using pruning algorithms to optimise network architectures for forecasting inputs in a neural networks rainfall-runoff model. *J. Hydroinformatics* **1**(2), 103–114.

ASCE Task Committee on Application of Neural Networks in Hydrology 2000 Artificial neural networks in hydrology. II: hydrologic applications. *J. Hydrol. Engng* **5**(2), 124–137.

Åström, K. J. & Wittenmark, B. 1995 *Adaptive Control* 2nd edn. Addison-Wesley: Reading, MA.

Billings, S. & Chen, S. 1992 Neural networks for non linear dynamic system modelling and identification. *Int. J. Control* **51**(6), 1191–1214.

Box, G. E. P. & Jenkins, G. M. 1989 *Time Series Analysis, Forecasting and Control*. Holden-Day: San Francisco, CA.

Calomino, F., Caputo, V., Galasso, L. & Piro, P. 1993 *Il Bacino Sperimentale Urbano di Luzzi*. Editorial Bios s.a.s.: Cosenza, Italy.

Fletcher, R. 1987 *Practical Methods of Optimization* 2nd edn. Wiley: New York.

Giustolisi, O. 1998 Simulating transformation inflow-outflow phenomena in technical hydrology by linear time-invariant models. *Proceedings of the 7th International Conference on Hydraulic Engineering Software, Hydrosoft '98, Como, Italy, 16–18 September 1998*, pp. 383–394. WIT press, Computational Mechanics Publications, Southampton, U.K.

Giustolisi, O. & Mastrorilli, M. 1994 Realizzazione di un modello connessionista applicato ad un problema di idrologia urbana. *XXIV Convegno di Idraulica e Costruzioni Idrauliche, Napoli, Italia, 20–22 settembre 1994*, vol 2, pp. T3a13–T3a2 CUEN s.r.l. Napoli, Italy.

Haykin, S. 1996 *Adaptive Filter Theory* 3nd edn. Prentice-Hall: Englewood Cliffs, NJ.

Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation* 2nd edn. Prentice-Hall: Englewood Cliffs, NJ.

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

Ljung, L. 1987 *System Identification: Theory for the User*. Prentice-Hall: Englewood Cliffs, NJ.

Mason, J. C., Price, R. K. & Temme, A. 1996 A neural network model of rainfall-runoff using radial basis functions. *J. Hydraulic Res.* **34**(4), 537–548.

Minns, A. W. 2000 Subsymbolic methods for data mining in hydraulic engineering. *J. Hydroinformatics* **2**(1), 3–13.

Minns, A. W. & Hall, M. J. 1996 Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.* **41**(3), 399–419.

Nørgaard, M. 1996 System identification and control with neural networks. *PhD Thesis*. Department of Automata, Technical University of Denmark.

Sjöberg, J., Hjalmerson, J. & Ljung, L. 1994 Neural networks in system identification. *10th IFAC Symposium on SYSID, Copenhagen, Denmark*, vol 2, pp. 49–71.

Smith, J. & Eli, R. N. 1995 Neural-network models of rainfall-runoff process. *J. Wat. Resour. Plan. Mngnt* **121**(6), 499–508.

Tikhonov, A. N. 1963 Solution of incorrectly formulated problems and the regulariztion method. *Dokl. Akad. Nauk SSSR* **151**, 501–504.