PERGAMON

AVAILABLE AT www.ComputerScienceWeb.com POWERED BY SCIENCE DIRECT®

Neural Networks 16 (2003) 375–387

Neural Networks

www.elsevier.com/locate/neunet

2003 Special Issue

Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting

Daniela Baratta[a], Giovambattista Cicioni[b], Francesco Masulli[a,c,*], Léonard Studer[d]

[a]*Istituto Nazionale per la Fisica della Materia, Via Dodecaneso 35, Genova I-16146, Italy*
[b]*Istituto di Ricerca sulle Acque, Consiglio Nazionale delle Ricerche, Via Reno 1, Roma I-00198, Italy*
[c]*Dipartimento di Informatica, Università di Pisa, Corso Italia 40, Pisa I-56125, Italy*
[d]*Institut de Physique des Hautes Énergies, Université de Lausanne, Dorigny CH-1015, Switzerland*

Abstract

In previous work, we have proposed a constructive methodology for temporal data learning supported by results and prescriptions related to the embedding theorem, and using the singular spectrum analysis both in order tso reduce the effects of the possible discontinuity of the signal and to implement an efficient ensemble method. In this paper we present new results concerning the application of this approach to the forecasting of the individual rain-fall intensities series collected by 135 stations distributed in the Tiber basin. The average RMS error of the obtained forecasting is less than 3 mm of rain.

Keywords: Time series learning; Ensemble methods; Singular spectrum analysis; Embedding theorem; Daily rainfall forecasting

1. Introduction

Learning a mapping on the basis of a (possibly small) data set of examples is an ill-posed inverse problem (Haykin, 1999). Concerning temporal time series learning, noise, ambiguity of the mapping, and discontinuity of the signal affect the generalization performance of the learning machines.

A popular way to reduce ill-posedness in temporal data learning consists in assuming an input scale (Dietterich, 2000) suitable to alleviate the mapping ambiguity problem. To this aim we should find the optimal dimension of the input vector and the time lag between its elements. After the setting of the mapping input vector and of other design issues, the temporal data can be learned by a machine. In particular, accurate learning of a continuous mapping is supported by the Universal Function Approximation property holding for some classes of learning machines including, e.g. Multi Layer Perceptrons (MLP), Radial Basis Functions Nets, and Fuzzy Basis Functions Nets (Cybenko, 1989; Poggio & Girosi, 1990; Wang & Mendel, 1992). However, for small data set, simple learning machines exhibit better generalization capabilities (Vapnik, 1995).

A constructive framework for the design of time series learning machines has been proposed by Studer and Masulli, (1995) and Masulli, Parenti, and Studer (1999). In particular, it has been suggested to apply results and prescriptions related to the delay-embedding theorem (Mañé, 1981; Takens, 1981) to the design of learning machines of continuous mappings of temporal data. A decompositive ensemble method based on the Singular-Spectrum Analysis (SSA) (Vautard, You, & Ghil, 1992) has been applied by Masulli, Cicioni, and Studer (2000) and Cicioni and Masulli (in press) in order to extend the constructive approach to the learning of discontinuous and/or intermittent signals.

The proposed toolbox has been successfully applied to the design of MLP and Neuro-Fuzzy systems for simulated non-linear and chaotic signal forecasting (Studer & Masulli, 1995), system identification (Masulli et al., 1999), and daily rainfall forecasting (Masulli et al., 2000; Cicioni & Masulli, in press).

The latter application has a strong implication in water quantity and quality management. Our study concerns the learning of the data set of the daily rainfall intensities series collected by 135 stations located in the Tiber river basin (Fig. 1) in the period 01/01/1958–12/31/1967. Using

* Corresponding author. Address: Istituto Nazionale per la Fisica della Materia, Universita di Genova, Via Dodecaneso 35, Genova I-16146, Italy. Tel.: +39-10-353-6604; fax: +39-10-353-6699.

E-mail address: masulli@disi.unige.it (F. Masulli).

0893-6080/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0893-6080(03)00022-4

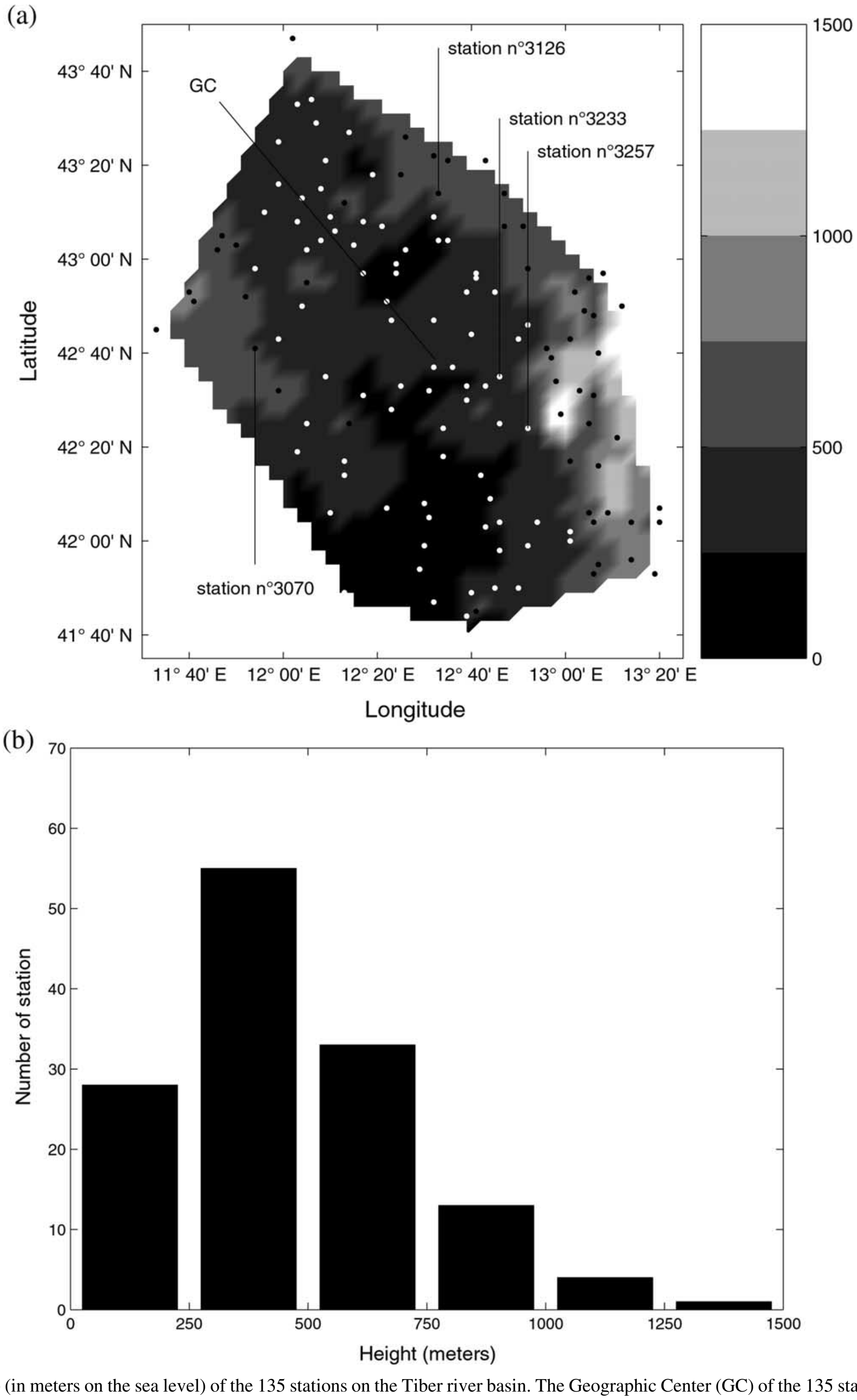

Fig. 1. (a) Height map (in meters on the sea level) of the 135 stations on the Tiber river basin. The Geographic Center (GC) of the 135 stations, the two stations more correlated to MS, and the two stations less correlated to it (Table 3) are shown on the map; (b) histogram of station's height.

the decompositive ensemble method based on the SSA, we obtained a Root Mean Square (RMS) error of 0.95 mm of rain on the daily forecasting of the series of the Mean Station (MS), defined as the average of all 135 rainfall intensity series (Masulli et al., 2000; Cicioni & Masulli, in press). In this paper we extend the experimentation to the daily rainfall forecasting of the individual stations, testing four alternative approaches.

The paper is structured as follows. In Section 2, we discuss the principal approaches to Quantitative Rainfall

Forecasting (QRF) and the meaning of the application presented in this paper. In Section 3, we summarize the main aspects of the constructive approach to temporal data learning and we present the main experimental results already obtained for the MS. Section 4 presents the four methods to the forecasting of the series of individual stations. In Section 5 the principal experimental results are shown, together with their discussion. The section of conclusions summarizes the results presented in the work.

2. Quantitative rainfall forecasting

The incorporation of QRF plays a key role in catchment management and flood warning systems. Accurate forecasts of the spatial and temporal distribution of rain are useful for both water quantity and quality management. For example, a flood warning system for fast responding catchments may require a quantitative rainfall forecast to increase the lead-time for warning. Similarly a rainfall forecast provide advance information for many water quality problems (Luk, Ball, & Sharma, 2001; Toth, Brath, & Montanari, 2000).

It is widely recognized that obtaining a reliable QRF is not an easy task, rainfall being one of the most difficult elements of the hydrological cycle to forecast (French, Krajewski, & Cuykendall, 1992). There are two possible approaches to forecast rainfall. The first approach involves the study of the rainfall processes in order to model the underlying physical laws (Brath, 1999; Luk et al., 2001). However, numerical weather forecasting based on the physical modeling approach may not be feasible because

- rainfall is an end product of a number of complex atmospheric processes which vary both in space and time;
- the data that is available to assist the definition of control variable for the process models, such as rainfall intensity, wind speed, and evaporation, etc. are linked in both the spatial and temporal dimensions;
- even if the rainfall can be described concisely and completely, the volume of calculations involved may be prohibitive; and
- the temporal and spatial resolution provided by this approach is not accurate enough for many hydrologic applications.

A second approach to forecast rainfall makes use of non-parametric models based on statistics and/or machine learning. With this approach, QRF is obtained inferring from features derived from remote sensing observations or from historical rainfall patterns, with no consideration of the physics of the rainfall processes (Luk et al., 2001), even if, when available, the incorporation in the training data series of other variables, such as pressure, temperature and wind speed and direction, can, in principle, improve the forecasting of rainfall.

We point out that, although remote sensing observations (e.g. radar data and satellite images) provide useful information on the precipitation pattern, they do not allow a satisfactory assessment of rain intensities yet (Kryszto-fiwicz, 1995). In addition, radar detection is particularly difficult in mountainous regions, because of ground occultation and altitude effects (Toth et al., 2000).

The standard non-parametric approaches to QRF by means of time-series analysis, is based on statistical techniques that assume linear relationship among variables or reproduce the precipitation time series only in statistical sense (Bodri & Cermák, 2000; Burlando, Rosso, Cadavid, & Salas, 1993; Sharma, 2000a,b; Sharma, Luk, Cordery, & Lall, 2000). Then, in principle, machine learning models, such as artificial neural networks, can improve the forecasting results obtained using models based on standard non-parametric approaches.

Maier and Dandy (1999) presented an extensive survey of the applications of neural networks to the forecasting of water resource variables, that included applications to QRF. Published contributions on the applications of neural networks to forecasting of true (not synthetic) rainfall series on the basis of the information embedded in the past rainfall depths, fall into two principal classes. The first one contains some papers on long-term scale rainfall forecasting, such as the one by Bodri and Cermák (2000) who used MLPs for extreme precipitation forecasting on yearly scale, while the other class includes several papers concerning short-term scale forecasting (hourly or shorter periods). We mention in this second group the paper by Toth et al. (2000) who compared short-term rainfall forecasts (1–6 h forward) for a case study on Apennines mountains (Italy) obtained using MLPs (that give the best results), linear stochastic Auto-Regressive Moving-Average (ARMA) models, and non-parametric nearest-neighbours method. They have been reported also good results on a shorter term range rainfall forecasting (15 min ahead for 16 gauges concurrently) obtained using either MLPs, Partial Recurrent Neural Networks (Elman, 1990), and Time Delay Neural Networks (Waibel, 1990), over an urban catchment in western Sydney (Australia) (Luk, Ball, & Sharma, 2000, 2001).

With respect to the previous referred papers, the study presented in this paper concerns the forecasting of rainfall in a large geographical area constituted by the Tiber river (Italy) basin, involving 135 stations distributed in the Apennines mountains close Rome and other smaller towns, and in a different (daily) time scale that has strong relevance in many hydrologic applications, including flood warning systems.

3. Constructive approach to time series learning

3.1. Embedding theorem and SSA

A constructive approach to shaping a supervised neural model of a non-linear process was proposed by Studer and

Masulli (1995), Masulli et al. (1999, 2000), and Cicioni and Masulli (in press), which can be based on the results and prescriptions related to the Embedding Theorem (Mañé, 1981; Takens, 1981). The input layer of a MLP predictor can be sized as the embedding dimension of the dynamical system computed, e.g. using the Global False Nearest Neighbors (FNN) method (Abarbanel, 1996), while the time lag of input can be selected as the first minimum of the average mutual information of the signal (Abarbanel, 1996; Fraser, 1989; Fraser & Swinney, 1986; Vastano & Rahman, 1989).

As shown by Studer and Masulli (1995), the estimation for the time lag based on mutual information is not supported from theory, and must be validated experimentally. On the other hand, for limited data sets, the best generalization can be obtained with learning machines of limited complexity (Vapnik, 1995), sometime leading to select MLP with input layers smaller that the embedding dimension of the dynamical system. Anyway, the FNN technique gives a reasonable starting point for the search of the optimal structure of the predictor.

Even if this constructive approach has been successfully applied to many cases (Masulli et al., 1999; Studer & Masulli, 1995), it cannot be directly applied to forecasting discontinuous or intermittent signals, such as the rainfall signal that is the target of this study, as the universal function approximation theorems for neural networks (Cybenko, 1989) and fuzzy systems (Wang & Mendel, 1992) require the continuity of the function to be approximated.

In Masulli et al. (2000) and Cicioni and Masulli (in press), we extended to the case of discontinuous or intermittent signals, by implementing an ensemble method based on the SSA (Broomhead & King, 1986; Kumaresan & Tuffs, 1980; Lisi, Nicolis, & Sandri, 1995; Vautard et al., 1992).

In SSA, the state vector $\mathbf{S}_i = (s_i, s_{i+1}, \ldots, s_{i+M-1})$ is a temporal window (augmented vector) of a series s of length N, made up of a given number of samples M.

The cornerstone of SSA is the Karhunen–Loève expansion, or Principal Component Analysis (PCA) (Therrien, 1989), that is based on the eigenvalue problem of the lagged covariance matrix Z_s. Z_s has a Toeplitz structure, i.e. constant diagonals corresponding to equal lags:

$$\begin{pmatrix} c(0) & c(1) & \cdot & \cdot & \cdot & c(M-1) \\ c(1) & c(0) & c(1) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & c(1) \\ c(M-1) & \cdot & \cdot & \cdot & c(1) & c(0) \end{pmatrix} \quad (1)$$

In the absence of prior information about the signal it has been suggested (Vautard et al., 1992) to use the following estimate for Z_s :

$$c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} s_i s_{i+j} \quad (2)$$

The original series can be expanded with respect to the orthonormal basis corresponding to the eigenvectors of Z_s

$$s_{i+j} = \sum_{k=1}^{M} p_i^k u_j^k, \qquad 1 \le j \le M, \ 0 \le i \le N-M \quad (3)$$

where p_i^k are called *principal components* (PCs) and the eigenvectors u_j^k are called the *empirical orthogonal functions* (EOFs), and the orthornomality property

$$\sum_{k=1}^{M} u_j^k u_l^k = \delta_{jl}, \qquad 1 \le j \le M, \ 1 \le l \le M \quad (4)$$

holds. It is worth noting that SSA does not resolve periods longer than the window length M. Hence, if we want to reconstruct a strange attractor, whose spectrum includes periods of arbitrary length, the large M the better, avoiding to exceeding $M = N/3$ (otherwise statistical errors could dominate the last values of the auto-covariance function).

Concerning the application of SSA to forecasting, it is supported by the following argument (Vautard et al., 1992): Since the PC are projections on EOFs, they are filtered versions of the signal (i.e. weighted moving averages with weights given by u_l). They are band limited and their observed behavior is more regular than that of the raw series. Hence, they are more easily forecast.

Following Vautard and Ghil (Ghil & Vautard, 1991; Vautard & Ghil, 1989; Vautard et al., 1992), suppose we want to reconstruct the original signal s_i starting from a SSA subspace $\mathfrak{A}$(of k eigenvectors. By analogy with Eq. (3), the problem can be formalized as the search for a series $\hat{s}$ of length N, such that the quantity

$$H_{\mathfrak{A}}(\hat{s}) = \sum_{i=0}^{N-M} \sum_{j=1}^{M} \left(\hat{s}_{i+j} - \sum_{k \in \mathfrak{A}} p_i^k u_j^k \right)^2 \quad (5)$$

is minimized. In other words, the optimal series $\hat{S}$ is the one whose augmented version $\hat{S}$ is the closest, in the least-squares sense, to the projection of the augmented series S onto EOFs with indices belonging to $\mathfrak{A}$.

The solution of the least-squares problem of Eq. (5) is given by

$$\hat{s}_i = \begin{cases} \dfrac{1}{M} \displaystyle\sum_{j=1}^{M} \sum_{k \in \mathfrak{A}} p_{i-j}^k u_j^k, & \text{for } M \le i \le N-M+1 \\ \dfrac{1}{i} \displaystyle\sum_{j=1}^{i} \sum_{k \in \mathfrak{A}} p_{i-j}^k u_j^k, & \text{for } 1 \le i \le M-1 \\ \dfrac{1}{N-i+1} \displaystyle\sum_{j=i-N+M}^{M} \sum_{k \in \mathfrak{A}} p_{i-j}^k u_j^k, & \text{for } N-M+2 \le i \le N \end{cases} \quad (6)$$

When $\mathfrak{A}$ consists on a single index k, the series $\hat{s}$ is called the kth *reconstructed component* (RC), and will be denoted by $\hat{s}^k$. RCs have additive properties, i.e.:

$$\hat{s} = \sum_{k \in \mathfrak{A}} \hat{s}^k \tag{7}$$

In particular, the series s can be expanded as the sum of its RCs:

$$s = \sum_{k=1}^{M} \hat{s}^k \tag{8}$$

Note that, despite its linear aspect, the transform changing the series s into $\hat{s}^k$ is, in fact, non-linear, since the eigenvectors u^k depend non-linearly on s.

If we truncate this sum to an assigned number of RCs, the explained variance of the related augmented vector $\hat{S}$ is the sum of the eigenvalues associated to those RCs, while the estimation of the resulting reconstruction error is the sum of the eigenvalues corresponding to the remaining RCs. As a consequence, it is suitable to order the RCs following the value of the eigenvalues.

Let $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_L$ be L independent subspaces with cartesian product spanning the full space of M eigenvectors. We define a *reconstructed wave* (RW) Ω_l as:

$$\Omega_l = \sum_{k \in \mathfrak{A}_l} \hat{s}^k. \tag{9}$$

Then, from Eqs. (8) and (9), we can obtain

$$s = \sum_{l=1}^{L} \Omega_l, \tag{10}$$

that says that the original series s can be recovered as the sum of all the individual RWs.

3.2. Ensemble method

In order to design a learner for complex signals, such as discontinuous and/or intermittent signals, we can apply the following approach that combines an unsupervised step and one supervised one, building-up in such a way an ensemble of learning machines:

- Unsupervised decomposition: Using the SSA, decompose the original signal S in RWs, corresponding to sub-spaces with equal explained variance.
- Supervised learning: Prepare a predictor for each RW as shown in the previous sub-section.
- Operational phase: The forecasting of the original signal S is then obtained as the sum of the forecasts of individual RWs, i.e. using Eq. (10).

It is worth noting that, sometime the most complex waves (in general those corresponding to the low eigenvalues) cannot be satisfactory forecasted, using the available data. Following the criteria of the *best prediction* (Lisi et al., 1995) in Eq. (10) we can excluded them if, when if enclosed in the sum, make worse the overall forecasting.

3.3. Previous results on the MS series

One challenging application of this ensemble method, carried out by our research group, is the forecasting of the daily rainfall intensities series. The considered data are the daily rainfall intensities series collected by 135 stations located in the Tiber river basin (Fig. 1) in the period 01/01/1958–12/31/1967.

The data analysis reported in depth by Masulli et al. (2000) and Cicioni and Masulli (in press) was related to the series of the MS, defined as the average of all 135 daily rainfall intensity series. Fig. 2 shows a window on the period 07/01/66–12/30/66 that enlightens the discontinuity and intermittence of the MS signal.

We report here a sketch of the methodology and of some experimental results that are relevant for the analysis of the daily rainfall series of the individual stations.

A preliminary work concerned the design of a MLP predictor of raw MS data using the constructive approach. The forecasting results obtained in this way were very poor, due to the discontinuity of the signal.

Then, in order to reduce the effects of the discontinuities, we applied the SSA to the first 3000 samples of the MS series. We shall consider in the following the results obtained with an SSA using a window length $M = 182$ days.

Some RW and the original signal for period 07/10/1966–12/30/1966 are shown in Fig. 3.

Using the SSA, from the original MS series we obtained 10 waves (RW1, …, RW10) reconstructed from 10 disjoint sub-spaces, each of them representing a 10% of the explained variance (Table 1). The best results for each

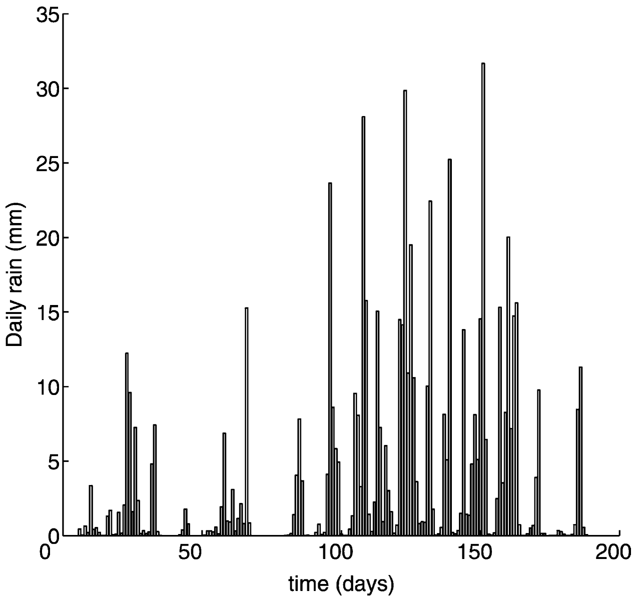

Fig. 2. Mean station: daily rain millimeters; period: 07/01/66–12/30/66.

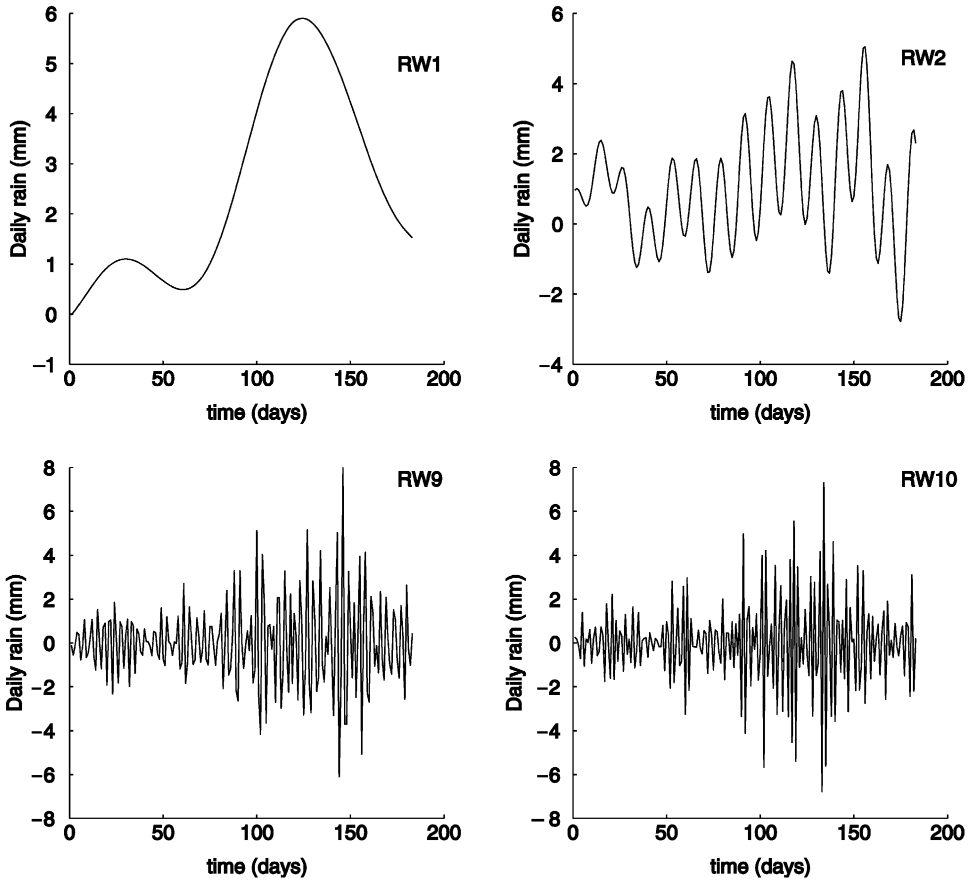

Fig. 3. Some RWs of the MS series; period: 07/01/1966–12/30/1966.

Table 1
Mean Station: RWs from disjoint SSA subspaces (each of them explaining 10% of the variance) and corresponding RCs

	RW									
	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
RCs	1–4	5–11	12–19	20–28	29–39	40–52	53–70	71–93	94–126	127–182

The SSA is performed using a window of 182 days.

RW have been obtained using as inputs windows of 5 consecutive elements and two hidden layers with dimensions shown in Table 2.

As each wave contains 3652 daily samples, in our case for each wave we obtained a data set of 3646 associative couples, each of them consisting of a window of 5 consecutive elements, as input, and the next day rainfall intensity, as output.

Each MLP was trained using the first 2000 associative couples (training sets) and error back-propagation algorithm

Table 2
Mean Station: Size of the hidden layers (L1 and L2), RMS error and Maximum Absolute (MAXA) error on the test set for each RW

	RW									
	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
L1	6	8	6	8	8	8	4	6	3	3
L2	4	5	4	4	5	4	4	4	4	4
RMS	0.02	0.03	0.04	0.04	0.06	0.15	0.15	0.64	0.75	0.29
MAXA	0.05	0.12	0.15	0.11	0.14	0.40	0.38	1.92	2.40	0.90

The size of MLPs input layer is 5.

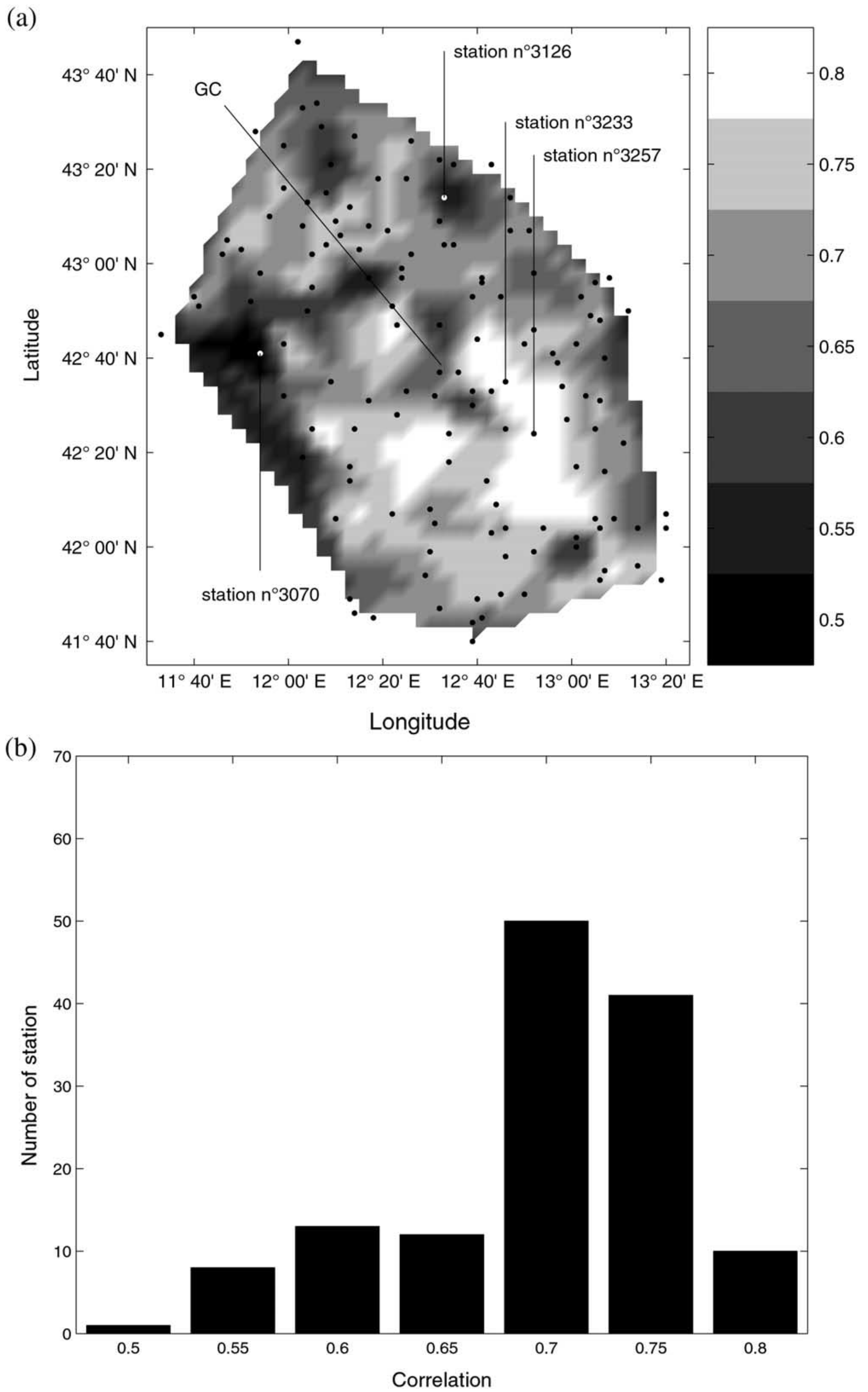

Fig. 4. (a) Correlation to MS map of the 135 stations on the Tiber river basin. The Geographic Center (GC) of the 135 stations, the two stations more correlated to MS, and the two stations less correlated to it (Table 3) are shown on the map; (b) histogram of station's correlation to MS.

with momentum and batch presentation of samples (Vogl, Mangis, Rigler, & Zink, 1988). The following 1000 associative couples (validation sets) were used in order to implement an early stopping of the training procedures. The remaining 646 were used for measuring the quality of the forecasting of the RW (test sets).

Using a window of 182 days for the SSA, the best forecasting results were obtained using MLPs with five

Table 3
Correlation to MS (Corr), RMS error on the test set obtained using the Approach C (RMS-C) and the Approach D (RMS-D) for the two stations more correlated to MS and for the two stations less correlated to it

Rank	Station	Corr	RMS-C	RMS-D
1	3257	0.82	1.93	2.40
2	3233	0.81	1.49	1.72
134	3126	0.53	1.56	2.31
135	3070	0.45	2.35	4.51

inputs and two hidden layers. Details on the size of hidden layers and on the forecasting results are given in Table 2. The sum of the forecasts of the 10 waves at 1 day ahead is very satisfactory, as the resulting MS forecasting on the test set has a RMS error of only 0.95 mm of rain. We also report that the Maximum Absolute error is 6.47 mm. Note that the forecasted signal is clamped to zero.

4. Learning the individual stations

The daily rainfall series of individual stations are more discontinuous than the series of the MS, but are well correlated to it. In Fig. 4(a) we plot the correlation map of the daily rainfall series of the individual stations to the MS, while the corresponding histogram is presented in Fig. 4(b). The average linear correlation coefficient is 0.7. Moreover, using the Fisher–Snedecor test, we find a linear dependence

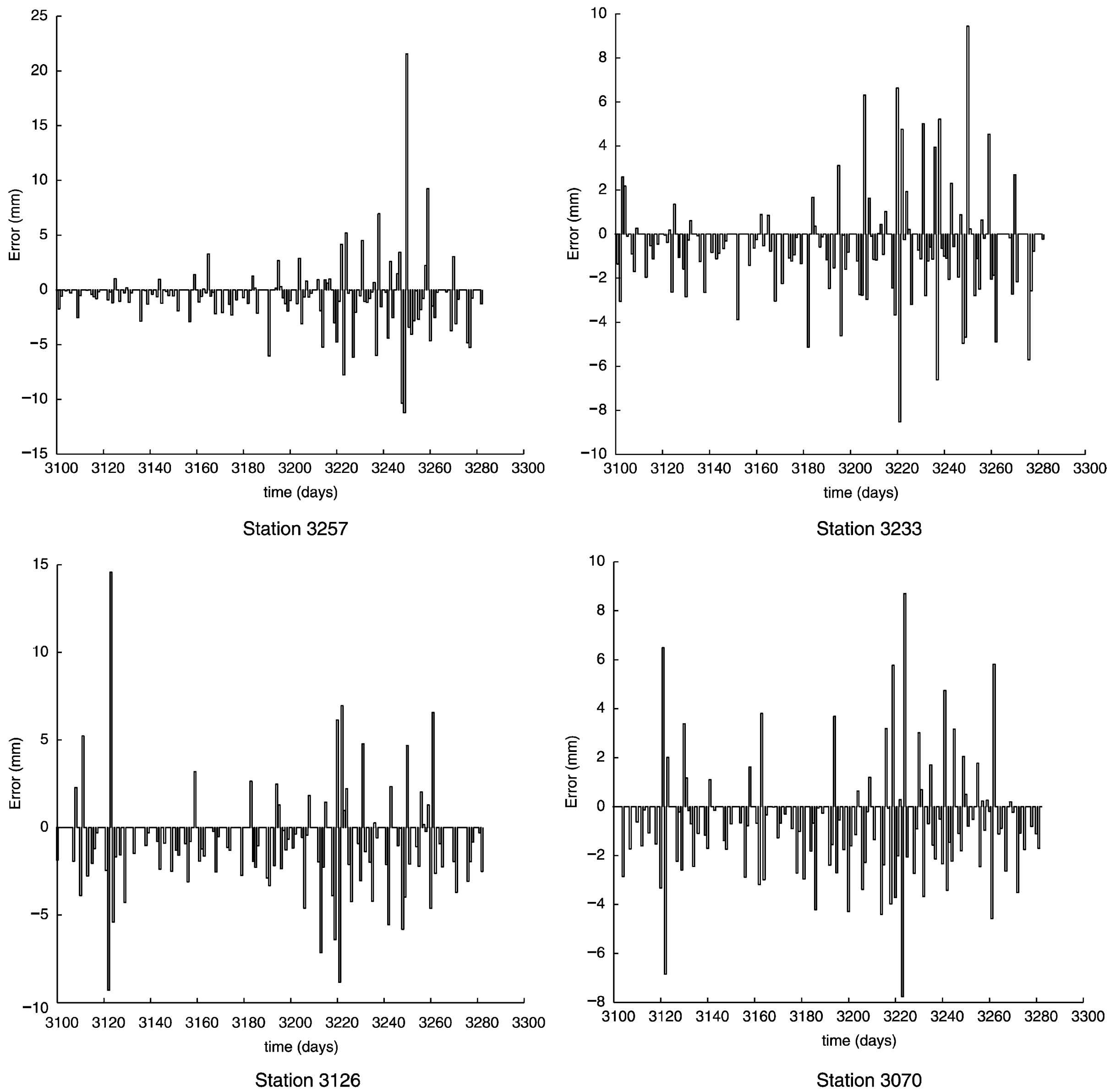

Fig. 5. Errors in the period 07/01/1966–12/30/1966 using ensembles of 10 MLPs with 5 inputs. The plots are relative to the two stations more correlated to MS, and to the two stations less correlated to it (Table 3).

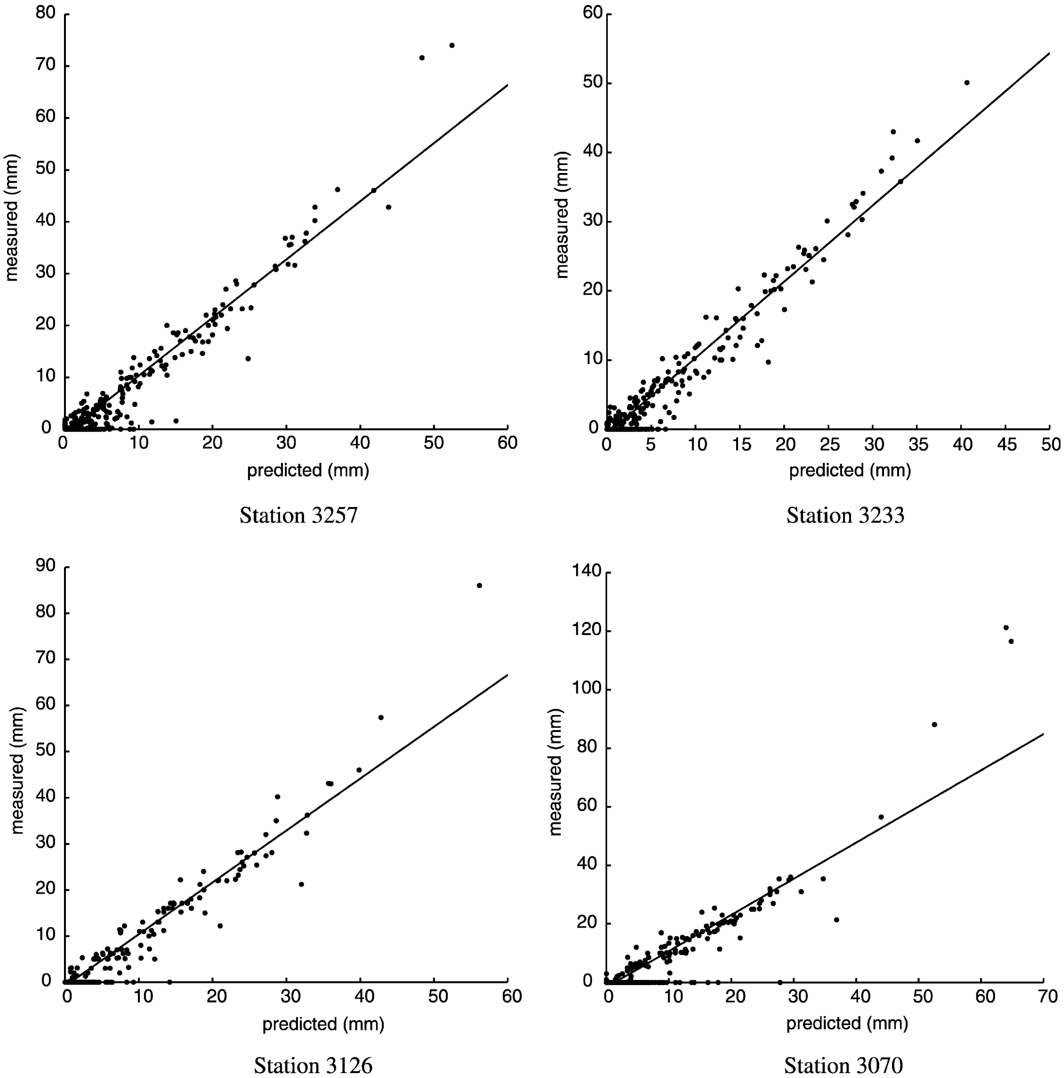

Fig. 6. Scatter plots on the test sets using ensembles of 10 MLPs with 5 inputs. The plots are relative to the two stations more correlated to MS, and the two stations less correlated to it (Table 3).

at the 0.01 level of station's correlation versus the distance from Geographical Center (GC) of the 135 stations. GC is defined as the average position of the 135 stations. Its longitude with respect to Greenwich is: $42\ 38'\ 88.8''$ E, its latitude is: $12\ 32'\ 51.0''$, and its height is 473.1 m.

Various problems have to be faced in forecasting the individual time series of the stations. For example, we could expect that the limited number of data points could lead to some uncertainty in setting the parameters of our predictors. Moreover, we could expect a common background meteorological behavior for most stations because of their correlation to the MS.

Let us mention other dilemmas. Efficient predictors can be based on the characteristic derived from the MS or from individual stations directly. The SSA could be applied to the time series of the MS or directly to the time series of the individual stations. A MLP could be trained on the time series of the MS, or MLPs could also be trained on time series from individual stations, eventually decomposed by SSA, and so on.

In this study, among all possible alternatives, we explored the four following approaches to the forecasting of individual time series:

Approach A. Design of a single neural predictor for each station, sizing of its input layer using the measurement of the average mutual information and the method of Global False Nearest Neighbors.

Approach B. Implementing the unsupervised decompositive ensemble method based on SSA for each station, following the same approach previously presented for the MS.

Approach C. Projecting the series of each individual station on the EOFs of the SSA already performed on

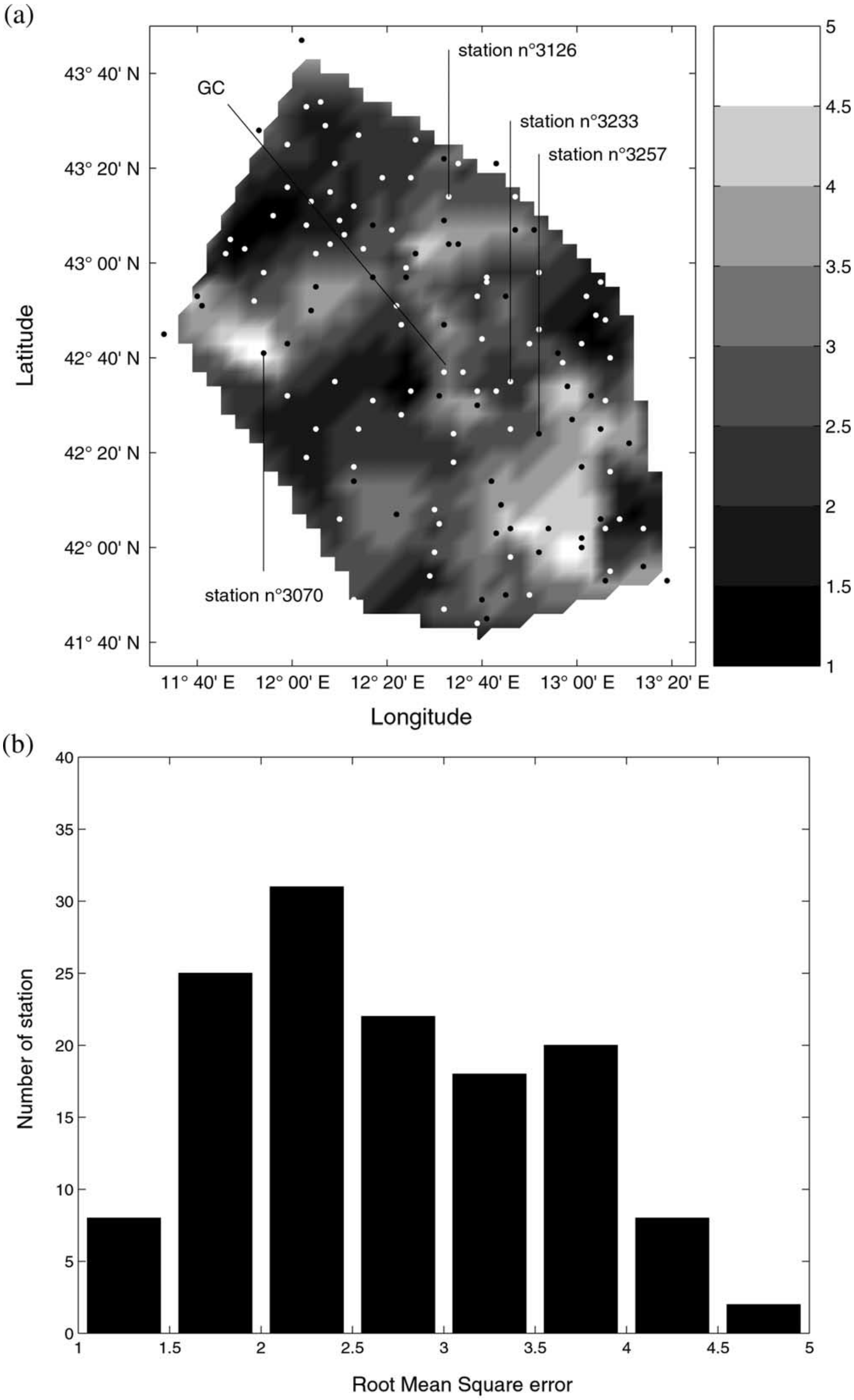

Fig. 7. (a) One day head forecasting RMS error map for the 135 stations on the Tiber river basin. The Geographic Center (GC) of the 135 stations, the two stations more correlated to MS, and the two stations less correlated to it (Table 3) are shown on the map; (b) histogram of station's one day head forecasting RMS error.

the MS, calculating the RCs, aggregating the RCs in 10 RWs following Table 1, and then training one MLP for each RW and for each projected individual station. The forecasting of the station's series will be the sum of the forecasts of the 10 RWs.

Approach D. Projecting the series of each individual station on the EOFs of the SSA already performed on the MS, calculating the RCs, and aggregating the RCs in 10 RWs following Table 1. The forecasting of the station's series will be the sum of the forecasts of the 10 RWs obtained using the MLPs already trained for the MS (with hidden layer dimensions shown in Table 2).

Note that the Approaches B–D are ensemble methods based on the SSA decomposition of the signal, implementing different alternatives: In Approach B for each individual station we perform the SSA and we train an ensemble of MLPs on the obtained RWs. In Approach C we obtain the RWs from the projection of the series of an individual station on the EOFs of the SSA of the MS and we train one MLP for each RWs. In Approach D we have only to obtain the RWs from the projection of the series of an individual station on the EOFs of the SSA of the MS, as we will use to forecast the weight of MLPs already computed for the MS.

5. Results and discussion

From our experimentation, the Approach A is unable to give useful results for any individual station, as well as for the MS station.

The Approach B, while is the most computationally expensive, at the same time leads to poor results, that we could ascribe to ill-conditioning in the SSA due to the significant presence of noise in the series of an individual station.

The Approaches C and D give similar good results. Using the Approach D (that is less computationally expensive than Approach C) the average RMS error for all the stations is about 2.71 mm of rain for all the stations, that is a good value in QRF. The results obtained with Approach C are often slight better than those of Approach D.

In Table 3, we show the RMS errors obtained for the two stations more correlated to the MS, i.e. Rieti (code 3257) and Arrone–Terni (code 3233), and for the two stations less correlated to MS, i.e. Scritto–Perugia (code 3126) and San Lorenzo Nuovo–Viterbo (code 3070), while in Figs. 5 and 6 we present the errors and the scatter plots on the test set for the stations 3070, 3126, 3233, and 3257. All results have been obtained following the Approach D, with the exception of those of station 3257 (that is the less correlated with MS) that have been obtained using the Approach C.

In Fig. 7 the one day ahead forecasting RMS error for the 135 stations is presented in form of as a geographic map and as an histogram.

6. Conclusions

In order to design a predictor for rainfall forecasting in the Tiber basin we applied a constructive methodology proposed in Studer and Masulli (1995), Masulli et al. (1999, 2000), and Cicioni and Masulli (in press) that leads to the design of efficient predictors even for complex signals, such as discontinuous or intermittent signals.

The approach followed by us to design the rainfall forecaster is an ensemble method that combines an unsupervised and a supervised step:

- *Unsupervised decomposition*. The original signal is decomposed in RWs, using the Singular Spectrum Analysis.
- *Supervised learning*. For each RW we design and train a MLP predictor using suggestions from dynamical systems theory.

In the operational phase the forecasting of the original signal is obtained as the sum of the forecasts of individual RWs. As reported in Masulli et al. (2000 and Cicioni and Masulli (in press), the daily rainfall forecasts of MS are very satisfactory, with a RMS error equal to 0.95 mm of rain.

In this paper we have extended the methodology to the learning of individual stations. The approach that gives the best results is based on the following steps:

(1) Decompose the series of the station using the SSA already performed on the MS; calculate the RCs and aggregate the RCs in a number of RWs.
(2) Train one MLP for each RW.

The forecasting of the station's series is the sum of the forecasts of the RWs.

We have also shown that it is possible to skip step 2 and use in forecasting the MLPs already trained for the MS. The daily rainfall forecasts on individual station obtained with this latter approach show an average RMS error of 2.71 mm of rain.

Note that, in the present study we tried to exploit the potentialities of the proposed ensemble method based on SSA decomposition and we have chosen to use as base learner a standard MLP trained with simple cross-validation. As a consequence, the good quality of the results in the forecasting of the station's series we have obtained is to be ascribed to the ensemble method.

The next step is (a) to investigate more powerful learning techniques for MLPs (e.g. Optimal Brain Damage (LeCun, Denker, & Solla, 1990), Brain Surgery (Hassibi, Stork, & Wolff, 1992) or Weight Decays (Krogh & Hertz, 1992)), other base learners (e.g. Partial Recurrent Neural Networks (Elman, 1990), and Time Delay Neural Networks) (Waibel, 1990), Finite Impulse Response Neural Networks (Wan, 1994), Non-linear Auto-Regressive models with exogenous input Neural Networks (van Zyl & Omlin, 2001), etc.), and ensemble methods (e.g. boosting (Freund, Iyer, Schapire, & Singer, 1999)) that can be introduced to improve the forecasting of RWs, and (b) to study their effectiveness on the overall improvement of the forecasting of the station series.

Acknowledgements

This works was partially supported by CNR Project 'Neuro-fuzzy systems to forecast hydrological variables', INFM, University of Genova and University of Pisa.

References

Abarbanel, H. (1996). *Analysis of observed chaotic data*. New York: Springer.

Bodri, L., & Cermák, V. (2000). Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. *Advances in Engineering Software*, *31*, 311–321.

Brath, A (1999). On the role of numerical weather prediction models in real-time flood forecasting. *Proceeding of the International Workshop on River Basin Modeling, Monselice, Italy* (pp. 249–259).

Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, *20*, 217–236.

Burlando, P., Rosso, R., Cadavid, L. G., & Salas, J. D. (1993). Forecasting of short-term rainfall using ARMA models. *Journal of Hydrology*, *144*, 193–211.

Cicioni, G., & Masulli, F. A software toolbox for time series prediction and its application to daily rainfall forecasting in a geographic basin. *Economics & Complexity* (in press).

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals, and Systems*, *2*, 303–314.

Dietterich, T. G. (2000). The divide-and-conquer manifesto. *Proceedings of the Eleventh International Conference on Algorithmic Learning Theory*, New York: Springer, pp. 13–26.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fraser, A (1989). *Information theory and strange attractors*. Tech. Rep. PhD thesis. University of Texas, Austin.

Fraser, A., & Swinney, L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review*, *33*, 1134–1140.

French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space time using a neural network. *Journal of Hydrology*, *137*, 1–31.

Freund, Y., Iyer, R., Schapire, R. E., & Singer, Y. (1999). *An efficient boosting algorithm for combining preferences*. *Machine Learning: Proceedings of the Fifteenth International Conference*, San Francisco: Morgan Kaufmann.

Ghil, M., & Vautard, R. (1991). Rapid disintegration of the wordie ice shelf in response to atmospheric warning. *Nature*, *350*, 324.

Hassibi, D., Stork, D. G., & Wolff, G. J. (1992). Optimal Brain Surgeon and general network pruning (Vol. 1). *IEEE International Conference on Neural Networks, San Francisco*, p. 293–299.

Haykin, S. (1999). *Neural networks. A comprehensive foundation* (2nd ed). Upper Saddle River, NJ: Prentice Hall.

Krogh, A., & Hertz, J. A. (1992). A simple weight decay can improve generalization. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in Neural Information Processing Systems 4* (pp. 450–957). San Mateo, CA: Morgan Kaufmann.

Kryzsztofiwicz, R. (1995). Recent advances associated with flood forecast and warning systems, US National Report to IUGG, 1991–1994. *Reviews in Geophysics*, *33*(2), 1139–1147.

Kumaresan, R., & Tuffs, D. (1980T). Data-adaptive principal component signal processing. *IEEE Proceedings of the Conference on Decision and Control*, Albuquerque, USA: IEEE, pp. 949.

LeCun, Y., Denker, J. S., & Solla, S. A. (1990). Optimal brain damage. In D. S. Touretzky (Ed.), (pp. 598–605). *Advances in Neural Information Processing Systems 2*, San Mateo, CA: Morgan Kaufmann.

Lisi, F., Nicolis, O., & Sandri, M. (1995). Combining singular-spectrum analysis and neural networks for time series forecasting. *Neural Processing Letters*, *2*, 6–10.

Luk, K. C., Ball, J. E., & Sharma, A. (2000). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology*, *227*, 56–65.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling*, *33*, 683–693.

Maier, H. R., & Dandy, G. C. (1999). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software*, *15*, 101–124.

Mañé, R. (1981). On the dimension of the compact invariant sets of certain non-linear maps. In D. Rand, & L. S. Young (Eds.), *Dynamical systems and turbulence* (*Vol. 898*) (pp. 230–242). *Lecture notes in mathematics*, Berlin: Springer.

Masulli, F., Cicioni, G., & Studer, L. (2000). Discontinuous and intermittent signal forecasting: A hybrid approach, *Technical report DISI-TR-00-4*, Department of Computer and Information Sciences, University of Genoa, Genova, Italy (http://www.disi.unige.it/person/MasulliF/papers/DISI-TR-00-4-masulli.pdf).

Masulli, F., Parenti, R., & Studer, L. (1999). Neural modeling of non-linear processes: relevance of the Takens–Mañé theorem. *International Journal on Chaos Theory and Applications*, *4*(2/3), 59–74.

Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. *Proceedings of the IEEE*, *78*, 1481–1497.

Sharma, A. (2000a). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management. Part 1-A strategy for system predictor identification. *Journal of Hydrology*, *239*, 232–239.

Sharma, A. (2000b). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management. Part 3-A nonparametric probabilistic forecast model. *Journal of Hydrology*, *239*, 249–258.

Sharma, A., Luk, K. C., Cordery, I., & Lall, U. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part2-predictor identification of quarterly rainfall using ocean–atmosphere information. *Journal of Hydrology*, *239*, 240–248.

Studer, L., & Masulli, F. (1995). On the structure of a neuro-fuzzy system to forecast chaotic time series. *Fuzzy Systems and A.I., Reports and Letters*, *4*, 31–37.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand, & L. S. Young (Eds.), *Dynamical systems and turbulence* (*898*) (pp. 366–381). *Lecture notes in mathematics*, Berlin: Springer.

Therrien, C. W. (1989). *Decision, estimation, and classification: an introduction to pattern recognition and related topics*. New York: Wiley.

Toth, E., Brath, A., & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology*, *239*, 132–147.

van Zyl, J., & Omlin, C. W. (2001). Prediction of seismic events in mines using neural networks. *International Joint Conference on Neural Networks, Washington, DC*, *2*, 1410–1414.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.

Vastano, J., & Rahman, L. (1989). Information transport in spatio-temporal chaos. *Physical Review Letters*, *72*, 241–275.

Vautard, R., & Ghil, M. (1989). Singular-spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, *35*, 395–424.

Vautard, R., You, P., & Ghil, M. (1992). Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. *Physica D*, *58*, 95–126.

Vogl, T., Mangis, J., Rigler, A., Zink, W., & Alkon, D. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, *59*, 257–263.

Waibel, A. (1990). Modular construction of time-delay neural networks for speech recognition. *Neural Computation*, *1*, 39–46.

Wan, E. (1994). Time series prediction by using a connectionist network with internal delay lines. In A. Weigend, & N. Gerhenfeld (Eds.), *Time series prediction. Forecasting the future and understanding the past* (pp. 195–217). Reading, MA: Addison-Wesley.

Wang, L., & Mendel, J. (1992). Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, *5*, 807–814.