

# Daily Rainfall Forecasting using an Ensemble Technique based on Singular Spectrum Analysis

Francesco Masulli, Daniela Baratta  
Istituto Nazionale per la Fisica della Materia and  
DISI-Dipartimento di Informatica e Scienze dell'Informazione  
Università di Genova, Via Dodecaneso 35, I-16146 Genova, Italy  
E-mail: masulli@disi.unige.it

Giovambattista Cicioni  
Istituto di Ricerca sulle Acque del  
Consiglio Nazionale delle Ricerche  
Via Reno 1, I-00198 Roma, Italy  
E-mail: cicioni@irsa.cnr.rm.it

Léonard Studer  
Institut de Physique des Hautes Énergies  
Université de Lausanne, CH-1015 Dorigny, Switzerland  
E-mail: leonard.studer@iphe.unil.ch

## Abstract

In previous work, we have proposed a constructive methodology for temporal data learning supported by results and prescriptions related to the Takens-Mañé theorem and using the Singular Spectrum Analysis in order to reduce the effects of the possible discontinuity of the signal. In this paper we present some new results concerning the application of this approach to the forecasting of the individual rainfall intensities series collected by 135 stations distributed in the Tiber basin.

## 1 Introduction

Learning a mapping on the basis of a (possibly small) data set of examples is an ill-posed inverse problem [4]. 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 [3] 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, Radial Basis Functions Nets, and Fuzzy Basis Functions Nets [2, 9, 14]. However, for small data set, simple learning machines exhibit better generalization capabilities [12].

In [10, 8], a constructive framework for the design of time series learning machines has been proposed. In particular, it has been suggested to apply results and prescriptions related to the delay-embedding theorem [11, 6] to the design of learning machines of continuous mappings of temporal data. In [8, 7] a decompositional ensemble method based on the Singular-Spectrum Analysis (SSA) [13] has been applied 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 Multi Layer Perceptrons and Neuro-Fuzzy systems for simulated non-linear and chaotic signal prediction [10], system identification [8], and daily rainfall forecasting [7].

The latter application concerns the learning of the data set of the daily rainfall intensities series collected by 135 stations located in the Tiber river basin (see Fig. 1) in the period 01/01/1958 - 12/31/1967.



**Figure 1:** Distribution of the 135 stations on the Tiber river basin.

In [7], we obtained a Mean Square Error (MSE) of .95 mm of rain on the daily prediction of the series of the Mean Station (MS), defined as the average of all 135 rainfall intensity series.

In this paper we extend this work to the daily rainfall prediction of the individual stations. The preliminary results give MSE of about 2 mm of rain for the daily prediction of the individual stations.

The paper is structured as follows. In the next section we summarize the main aspects of the constructive approach to temporal data learning. Sect. 3 presents the experimental results obtained for the MS and for the series of individual stations, while the section of conclusions summarizes the main results of this work.

## 2 Constructive Approach to Time Series Learning

In [10, 8, 8, 7], we proposed a constructive approach to shaping a supervised neural model of a non-linear process can be based on the results and prescriptions related to the Takens-Mañé theorem [11, 6].

The input layer of the Multi Layer Perceptron (MLP) predictor can be sized as the embedding dimension of the dynamical system computed, e.g., using the Global False Nearest Neighbors (FNN) method [1], while the time lag of input can be selected as the first minimum of the average mutual information of the signal [1].

As shown by Studer and Masulli [10], 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 [12], leading, sometime, 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 [10, 8], it can not 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 [2] and fuzzy systems [14] require the continuity of the function to be approximate.

In [7], we extended to the case of discontinuous or intermittent signals, by implementing an ensemble method based on the Singular-Spectrum Analysis (SSA) [13].

The cornerstone of SSA is the Karhunen-Loève expansion or Principal Component Analysis (PCA) that is based on the eigenvalues problem of the lagged covariance matrix. Concerning the application of SSA to prediction, it is supported by the following argument: Since the principal component (PC) are filtered version of the signal and typically band-limited, their behavior is more regular than that of the raw series, and hence more predictable.

In our approach [7], in order to reduce the computational costs, the raw series is decomposed in reconstructed waves (RWs) corresponding to SSA subspaces equivalent to similar explained variance and we predict them using Multi-Layer Perceptrons combined with independent evaluation of time lag using the first minimum of mutual information and embedding dimension using False Nearest Neighbors method [1]. The prediction of the original series can be recovered as the sum of those of all the individual series components.

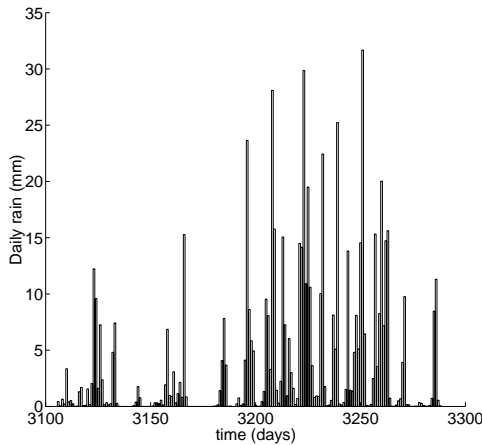
### 3 Application to Rainfall Forecasting

#### 3.1 Data Set

We applied the constructive forecasting method to the daily rainfall intensities series collected by 135 stations located in the Tiber river basin in the period 01/01/1958 - 12/31/1967.

#### 3.2 Learning the Mean Station

**3.2.1 Data Processing.** The data analysis started by considering the series of the Mean Station (MS), defined as the average of all 135 rainfall intensity series. Fig. 2 illustrates a window on the period 07/01/66 - 12/30/66 that enlightens the discontinuity and intermittence of the MS signal.



**Figure 2:** Mean Station: Daily rain millimeters. Period 07/01/66 - 12/30/66

A preliminary work concerned the design of a MLP predictor of raw MS data using the constructive approach. The first minimum of the MS's average mutual information gives an estimation of the time lag as  $T = 7$ . This time lag has been used for the computation of Global False Nearest Neighbors, obtaining an estimated embedding dimension  $d_E = 6$  [1]. The prediction 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.

Using the method illustrated in Sect. 2, from the original MS series we obtain 10 waves (RW1, ..., RW10) reconstructed from 10 disjoint sub-spaces, each of them representing a 10% of the explained variance (see Tab. 1).

**Table 1:** Reconstructed waves (RWs) from disjoint SSA subspaces (each of them explaining 10% of the variance) and corresponding reconstructed components (RCs). The SSA is performed using a window of 182 days.

RW	RCs
$\Omega_1$	1-4
$\Omega_2$	5-11
$\Omega_3$	12-19
$\Omega_4$	20-28
$\Omega_5$	29-39
$\Omega_6$	40-52
$\Omega_7$	53-70
$\Omega_8$	71-93
$\Omega_9$	94-126
$\Omega_{10}$	127-182

The best results for each RW have been obtained using as inputs windows of 5 consecutive elements and two hidden layers with dimensions shown in Tab. 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 with momentum and batch presentation of samples. 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 reconstructed waves (test sets).

**3.2.2 Results and Discussion.** Using a window of 182 days for the SSA, the best prediction results were obtained using MLPs with five inputs and two hidden layers. Details on the size of hidden layers and on the prediction results are given in Tab. 2. The sum of the predictions of the 10 waves at 1 day ahead is very satisfactory, as the resulting MS prediction on the test set has a Mean Square Error (MSE) of only .95 mm of rain. We report also that the Maximum Absolute Error is 6.47 mm (see Fig. 3). Note that the predicted signal is clamped to zero.

**Table 2:** Size of the hidden layers (L1 and L2), Mean Square Error (MSE) and Maximum Absolute Error (MAE) on the test set for each reconstructed wave (RW). The Size of MLPs Input Layer is 5.

RW	L1	L2	MSE	MAE
$\Omega_1$	6	4	.02	.05
$\Omega_2$	8	5	.03	.12
$\Omega_3$	6	4	.04	.15
$\Omega_4$	8	4	.04	.11
$\Omega_5$	8	5	.06	.14
$\Omega_6$	8	4	.15	.40
$\Omega_7$	4	4	.15	.38
$\Omega_8$	6	4	.64	1.92
$\Omega_9$	3	4	.75	2.40
$\Omega_{10}$	3	4	.29	.90

### 3.3 Learning the Individual Stations

**3.3.1 Data Processing.** We explored the following alternative approaches to the design of a predictor for an individual station:

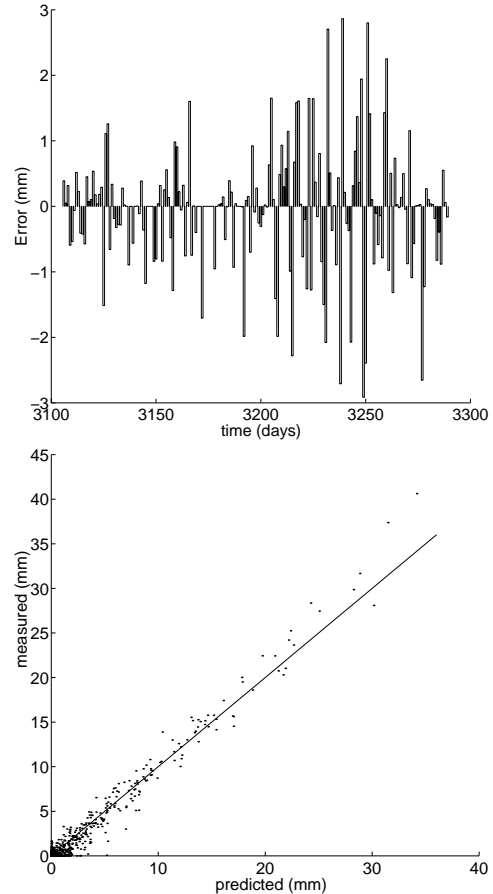
*Approach A:* Design of a single neural predictor for each station, sizing of its input layer using the measurement of the average mutual information and the method of Global False Nearest Neighbors.

*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:* Decomposing the series of a station using the SSA already performed on the MS, calculating the RCs, aggregating the RCs in 10 RWs following Tab. 1, and then training one MLP for each RW. The prediction of the station’s series will be the sum of the predictions of the 10 RWs.

*Approach D:* Decomposing the series of a station using the SSA already performed on the MS, calculating the RCs, aggregating the RCs in 10 RW following Tab. 1. The prediction of the station’s series will be the sum of the predictions of the 10 RWs obtained using the MLPs trained for the MS.

Note that the Approaches B, C, and D are ensemble methods based on the SSA decomposition of the signal, with different flavors.



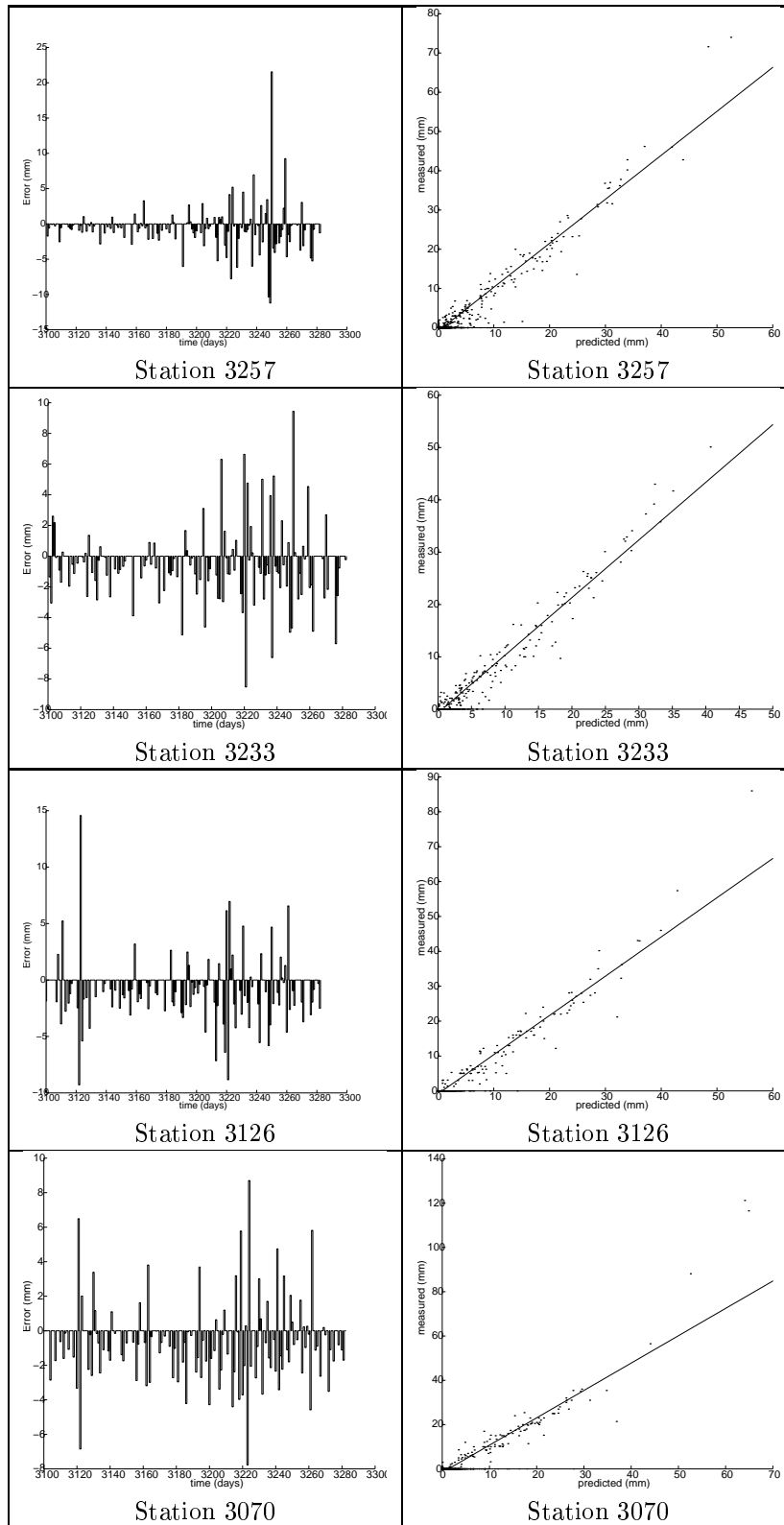
**Figure 3:** Mean Station: 1 day ahead forecasting. Errors in the period 07/01/66 - 12/30/66 and scatter plot on the complete test set (MLPs with 5 inputs).

**3.3.2 Results and Discussion.** We show here the preliminary results obtained for the two stations more correlated to the MS, i.e., Rieti (code 3257) and Arrome-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) (see Tab. 3).

Like for the MS station, the Approach A is unable to give useful results for any individual 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. The obtained MSEs were about 2 mm of rain for all the stations we have considered, as shown in Tab. 3. In Fig. 4 we present the absolute errors and the scatter



**Figure 4:** Errors in the period 07/01/1966 - 12/30/1966 (left) and scatter plots (right) on the test set using ensembles of 10 MLPs with 5 inputs.

**Table 3:** Correlation to MS (Corr), Mean Square Error on the test set obtained using the Approach C (MSE-C) and the Approach D (MSE-D) for the 2 stations more correlated to MS and for the two stations less correlated to MS.

Rank	Station	Corr	MSE-C	MSE-D
1	3257	.82	2.40	1.93
2	3233	.81	1.72	1.49
134	3126	.53	2.31	1.56
135	3070	.45	4.51	2.35

plots on the test set for the stations 3070, 3126, 3233, and 3257. All those results have been obtained following the Approach D, with the exception of station 3257 that is the less correlated with MS.

#### 4 Conclusions

In order to design a predictor for rainfall forecasting in the Tiber basin we applied a constructive methodology proposed in [7, 8, 10] 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 reconstructed waves (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 prediction of the original signal is obtained as the sum of the predictions of individual RWs.

The daily rainfall predictions of MS are very satisfactory, with a Mean Square Error equal to .95 mm of rain. Good results have been also obtained for the individual stations, where the Mean Square Errors are about 2 mm of rain.

#### Acknowledgments

This work was partially supported by Project of CNR: "Neuro-fuzzy systems to forecast hydrological variables", INFN, and University of Genoa. We thank Massimo Riani for helpful suggestions.

#### References

- [1] H.D.I. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, New York, USA, 1996.
- [2] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals and Systems*, 2:303-314, 1989.
- [3] T. G. Dietterich. The divide-and-conquer manifesto. In *Proc 11th Int Conf Algorithmic Learning Theory*, pages 13-26, New York, 2000. Springer-Verlag.
- [4] S. Haykin. *Neural Networks. A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, 1999.
- [5] S. Haykin and J. Principe. Making sense of a complex world: Using neural networks to dynamically model chaotic events such as sea clutter. *IEEE Signal Processing Magazine*, 15, 1998.
- [6] R. Mañé. On the dimension of the compact invariant sets of certain non-linear maps. In *Lecture Notes in Mathematics*, volume 898, pages 230-242, 1981. Springer-Verlag, Berlin.
- [7] F. Masulli, Gb. Cicioni, and L. Studer. Discontinuous and intermittent signal forecasting: A hybrid approach. Technical report DISI-TR-00-4, Dept Computer and Information Sciences, Univ Genoa, Italy, 2000. (<http://www.ge.infn.it/~masulli/papers/DISI-TR-00-4.pdf>).
- [8] F. Masulli, R. Parenti, and L. Studer. Neural modeling of non-linear processes: Relevance of the Takens-Mañé theorem. *Int J Chaos Theory and Applications*, 4, 2-3:59-74, 1999.
- [9] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78:1481-1497, 1990.
- [10] L. Studer and F. Masulli. On the structure of a neuro-fuzzy system to forecast chaotic time series. *Fuzzy Systems & A.I., Reports and Letters*, 4:31-37, 1995.
- [11] F. Takens. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics*, volume 898, pages 230-242, 1981. Springer-Verlag, Berlin.
- [12] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [13] R. Vautard, P. You, and M. Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, 58:95-126, 1992.
- [14] L. Wang and J.M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE T Neural Networks*, 5:807-14, 1992.