Available online at www.sciencedirect.com
ScienceDirect
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
Journal of Computational and Applied Mathematics 221 (2008) 302–309
www.elsevier.com/locate/cam

Wavelets and Elman Neural Networks for monitoring environmental variables

Patrizia Ciarlini*, Umberto Maniscalco

Istituto per le Applicazioni del Calcolo "Mauro Picone", Viale del Policlinico 137, 00161 Roma, Italy

Received 9 January 2007; received in revised form 23 May 2007

Abstract

An application in cultural heritage is introduced. Wavelet decomposition and Neural Networks like virtual sensors are jointly used to simulate physical and chemical measurements in specific locations of a monument. Virtual sensors, suitably trained and tested, can substitute real sensors in monitoring the monument surface quality, while the real ones should be installed for a long time and at high costs. The application of the wavelet decomposition to the environmental data series allows getting the treatment of underlying temporal structure at low frequencies. Consequently a separate training of suitable Elman Neural Networks for high/low components can be performed, thus improving the networks convergence in learning time and measurement accuracy in working time.

Keywords: Wavelet pre-processing; Recursive Neural Network

1. Introduction

In a monitoring process of a monument several sensors must be installed on it to measure physical and chemical quantities. They have to be maintained for years to repeat sample campaigns. This process is invasive, reduces the enjoyment of the monument itself and has high costs. The solution of a *not-invasive* monitoring of a monument via Neural Networks has been given in [3]: the author showed that physical atmosphere variables, such as temperature or humidity, can be indirectly measured at points on the surface of the roman theater in Aosta city, by using specific virtual sensors based on Neural Networks (NN), having as input measurements acquired by an Air Ambient Measurement Station (AAMS) working nearby the theater (ground level).

A rich environmental database was available. It contains measurements of several ambient variables along many months, characterized by complex relationships and correlations in space and time. It was used for training the Neural Networks [4], thus obtaining several virtual sensors able to predict values for the environmental variables on the monument, with good standard errors. However, an unsatisfactory behavior of the baseline in the data simulated by the Neural Networks in a day period, has been observed: a drift could be due to the difficulty of the used Elman NN (ENN) in learning the low frequency components characterizing the physical signal during the day.

* Corresponding author.
E-mail addresses: p.ciarlini@iac.cnr.it (P. Ciarlini), maniscalco@iac.cnr.it (U. Maniscalco).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.10.040

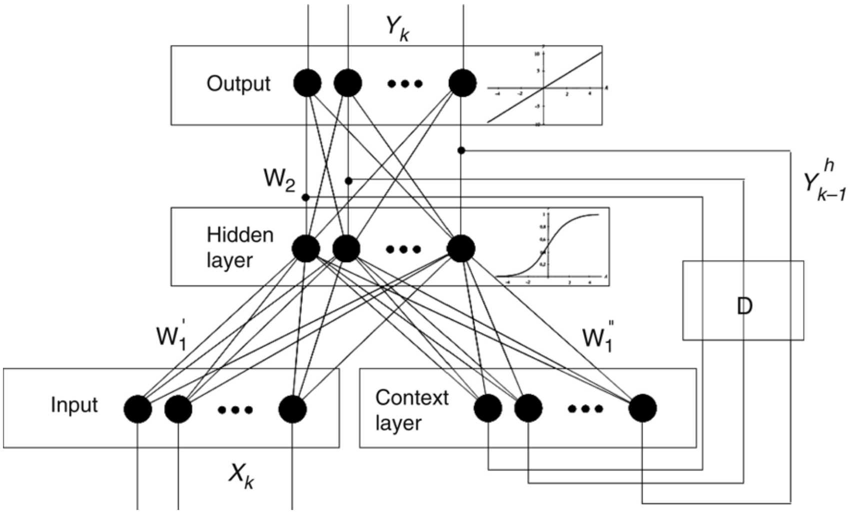

Fig. 1. Scheme of a recursive Elman Neural Network [2] (one hidden layer) and multidimensional input/output data.

In this work we firstly apply the wavelet decomposition to process the time series data that will be used for monitoring an ambient variable via recursive ENN. Wavelet transform have already been used, coupled to the Neural Network mechanism, to process air pollutant data [6], in order to recover low frequency temporal structures in an accurate way. In fact, the wavelet post-processing step is a fusion step, which allows us to correct the temporal drift and to increase the accuracy and the reliability of the virtual measuring system.

We want to underline that this monitoring application is not a classical forecast problem in time series as the one in [1], since the Neural Network has to predict a value approximating the ambient variable at the same time in which an input information is given to the NN, but in another location and it must provide accurate estimates for a long time after the training period.

Since these virtual sensors are here viewed as instruments that give measurements in an indirect way, the accuracy of the virtual measures will be estimated and validated via a statistical procedure introduced in [4].

The next section shortly recalls the operational mode of a NN and introduces the monitoring problem in cultural heritage as solved by means of Neural Networks of recursive type. In Section 4 the time series data pre-processing by wavelet is described. Section 5 briefly describes the original procedure that is used to validate the virtual sensor performances in a statistical way. Moreover, the experimental results obtained by applying the wavelet processing of data series with different wavelet bases are compared and discussed. Conclusions are finally given.

2. Monitoring process of a monument via Neural Networks

A NN is configured for a specific application, through a learning process of several neurons about complex relationship between different variables and it is able to derive meaning from imprecise data. It has two modes of operation: the training mode and the working mode. In the first mode, NN is trained to give a known output value (target) for a given input, depending on specific inner structure (coded as number of neurons, weights and biases); in the working mode, the NN produces an output depending on both the input and the learnt structure of weights and biases.

In Fig. 1 the topology of a recursive ENN is given (the recursive mechanism is indicated by D); it corresponds to a not-linear system, where $\mathbf{W}_1'$ is the matrix of weights that connects the input layer to the hidden layer, $\mathbf{x}_k$ is the multidimensional input of the ENN at kth item, $\mathbf{y}_{k-1}^h$ is the output of the hidden layer for the $(k-1)$th item, $\mathbf{W}_1''$ is the matrix of the weights that connects the context layer to the hidden layer, $\mathbf{W}_2$ is the matrix of the weights that connect the hidden layer to the output layer and $\mathbf{y}_k$ is the kth multidimensional output of the network. All weights and biases can simply be included in a *weight* structure that corresponds to a vector for a single input/single output Elman NN in the case of one hidden layer, otherwise to a matrix, whose dimensions depend on the number of neurons at each layer. A training dataset of N items of ($\mathbf{x}$ and $\mathbf{y}$), and the Levenberg–Marquad algorithm are used to identify the number of neurons and to estimate the weights of the associated not-linear system, starting from a random initial state [2].

In the application of monitoring the ambient variables at locations on the Aosta theater, let us consider for example the Elman virtual sensor developed to associate the input variable *air temperature* $a(t)$, acquired at ground level, to the variable *contact temperature* $r(t)$, acquired at the West locations on the monument: this specific ENN has been trained to simulate a virtual temperature value $\tilde{r}(t)$ at the West location on the monument for any new input value $a(t)$, measured at the same time t by the AAMS. The association to be learnt in the training phase is multivalues: a temperature value measured at ground level may correspond to several values as measured in the specific monument location at the same hour, but in different days. A training-set of size N has series of N ordered pairs of measurements of both the atmosphere variables ($\mathbf{a}$; $\mathbf{r}$), with $\mathbf{a} = (a(t_1), \dots, a(t_N))'$ as for $\mathbf{r}$.

In a working phase, the virtual sensor has to predict values at the West location on the monument in place of the real sensor and for a long period. Therefore the quality of the ENN response is to be accurately validated, i.e. by analyzing the errors $r(t) - \tilde{r}(t)$ using a new dataset, a test set having size N' and $N \cong N'$.

Let us simply describe our virtual sensor in the working phase with the following compact equation:

$$s(j) = f(a(j), \Delta(a(j-1)); \tilde{\mathbf{W}}) \quad j = 1, \dots \tag{1}$$

where $f(.)$ represents the specific trained Elman Neural Network that estimates the target contact temperature $r(t)$: $s(j) = \tilde{r}(j)$ for any novel input $a(j)$; $\Delta(.)$ is the output of the hidden layer representing the recursive network behavior and $\tilde{\mathbf{W}}$ contains all the estimated weights and biases. We underline that the implicit model between two environmental variables, learnt during the training, will not change with the time. Therefore the virtual sensor $s(t)$ will simulate the same values $\mathbf{s}$ for the same input $\mathbf{a}$ and same initial state, but it is also able to answer in new cases ever learnt before.

3. Data series pre-processing by DWT

The multiresolution analysis allows us to get a good treatment of signals with very rapid variation in time, as the ones regarding the temperature (Fig. 3, upper plot). Let us describe in filter notation the wavelet procedures that must be applied to the data series of N items in the training phase and of N' in the working phase.

Say $\mathbf{y}^{(0)}$ a generic data series of N items, $\mathbf{H} = (h_j)$ and $\mathbf{G} = (g_j)$ the low-pass and high-pass filters, which we assume to be compactly supported. Then, at level i of decomposition the following two filtered sequences are obtained for low coefficients:

$$\mathbf{y}_L^{(i)} = \mathbf{H} * \mathbf{y}_L^{(i-1)} \tag{2}$$

where $*$ is the usual convolution operation between vectors, and for high coefficients:

$$\mathbf{y}_H^{(i)} = \mathbf{G} * \mathbf{y}_L^{(i-1)}, \quad i = 1, \dots, I. \tag{3}$$

Two wavelet data processing procedures are displayed in Fig. 2 for level of decomposition $I = 1$: wavelet transforms of input/output data series in the training phase, data prediction and data post-processing by IDWT in the working phase.

In the training phase DWT is firstly applied to decompose, in terms of time and frequencies, both the input series $\mathbf{a}$ and the target ones $\mathbf{r}$ of length N, according to Eq. (2) and $\mathbf{y}^{(0)} = \mathbf{a}$ or $\mathbf{r}$. Two filtered sequences of low-pass coefficients and two of high-pass coefficients are obtained, say ($\mathbf{a}_L^{(1)}$; $\mathbf{r}_L^{(1)}$) and ($\mathbf{a}_H^{(1)}$; $\mathbf{r}_H^{(1)}$).

Then a specific ENN must be designed and trained for the low components and another one for the high components, in the classical manner, i.e. using these training sets of size $N/2$: the low coefficients ($\mathbf{a}_L^{(1)}$; $\mathbf{r}_L^{(1)}$) are provided to the ENN in order to compute the specific weights $\tilde{\mathbf{W}}_L$ that will be used in the working phase; analogous procedure is needed to get $\tilde{\mathbf{W}}_H^{(1)}$.

If the resolution level is defined for values $I > 1$, the following DWT operations are applied:

– DWT is applied to both the data series $\mathbf{a}$ and $\mathbf{r}$ to get two filtered sequences of low-pass components and two of high-pass components for $i = 1$;
– DWT is again applied to the low components, going ahead along the usual wavelet tree having length I;

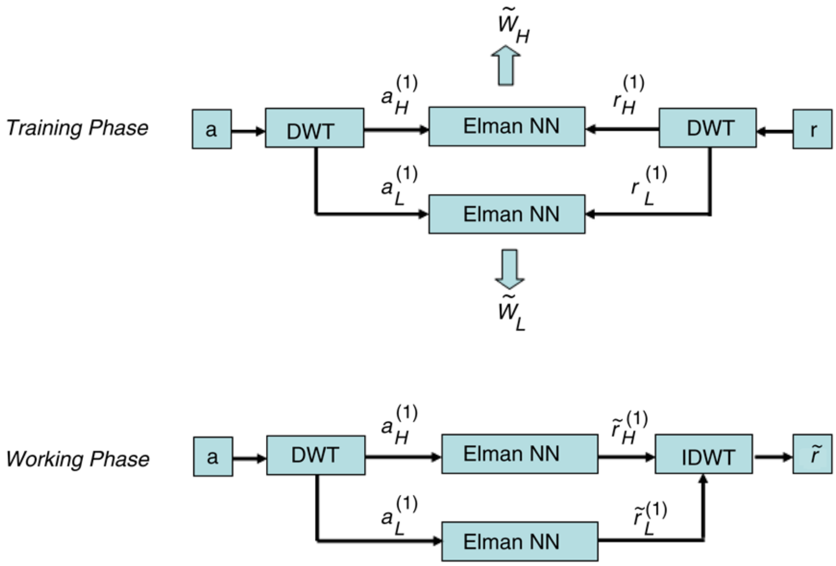

Fig. 2. Wavelet procedures at decomposition of level 1 as applied in the training and working phase.

– several ENNs are to be designed and trained: $i = 1, \ldots, I$ different ENNs, each one having $(\mathbf{a}_H^{(i)}; \mathbf{r}_H^{(i)})$ as input and target couples of high components with halved length at each ith level, to estimate $\tilde{\mathbf{W}}_H^{(i)}$; only one ENN having as input couples of low components obtained at the last level I, to compute the ENN weights $\tilde{\mathbf{W}}_L$.

In the working phase, the DWT is applied only to a new input data series $\mathbf{a}$ of given length N'. Then several values are predicted by the trained ENNs for low/high components at different levels, i.e. $\tilde{\mathbf{r}}_L^{(1)}$ by using $\tilde{\mathbf{W}}_L$ and $\tilde{\mathbf{r}}_H^{(i)}$ by using $\tilde{\mathbf{W}}_H^{(i)}$ (Fig. 2 (bottom)). The output of each ENN are the predicted coefficients having the required length $N'/2^i$.

In the wavelet reconstruction step, IDWT fuses all these obtained predictions at the appropriate scale levels to finally output N' simulated values $\tilde{\mathbf{r}}$ of the target quantity $\mathbf{r}$.

For the choice of the wavelet basis the following considerations are of some relevance. Usually the symmetry feature of the wavelet basis is not required in time series processing. The observed environmental signals have low regularity, then the correction of their time drift can be achieved by wavelet bases with compact support and with low regularity order. Moreover, in the wavelet decomposition we get a downsampled sequence at each level, thus reducing the information that are provided to each ENN. This implies that small values of I should be chosen otherwise the convergence of the ENN could not be achieved. This convergence problem, due to the downsampling operations, might be overcome by using the stationary wavelet decomposition. The not-decimated wavelet transform, often adopted in de-noising application, could also be used in the data pre-processing jointly with the ENN mechanism. However, this method heavily increases the computational costs of the training phase, since not only the number of items in each filtered sequence maintains the same value at every i, but also the number of ENNs that must be trained increases rapidly with i: for $I = 1$ we should train four ENNs (two ENNs for low components and even positions, two for high ones and odd positions) instead of two in the not-stationary case.

4. Validation of the virtual sensor and the experimental results

The virtual sensor performance is usually validated by a statistical procedure, based on the substitution error, and using a new dataset, called *test set*, usually regarding a period as long as the train period. The procedure in [4] uses several statistical evaluators, besides the usual overall statistical estimator as the *standard error*. These statistical evaluators allow us to assure the accuracy of the computed predictions at any hour of the day and in all the observed temperature range. The substitution error for the virtual sensor s of Eq. (1) at the kth time is given by

$$e_s(k) = s(k) - r(k), \quad k = 1, \ldots, N \tag{4}$$

where $s(k)$ is the predicted signal (the dependency on the weights is omitted) and $r(k)$ is the target signal, i.e. measurement of the ambient variable by the real sensor working in the same monument location at the same time. In the validation procedure, these virtual errors are analyzed by means of several estimators similar to that used for a real instrument. For $\mathbf{e}_s$ the following evaluators are here computed using a test set containing N' data couples of new data:

- ratio RR_s between the ranges of **s** and **r**;
- minimum, maximum, mean and standard deviation values for $\mathbf{e}_s$;
- CC_s, correlation coefficient [**s**, **r**];
- mean M^c and standard deviation values Std^c for each $\mathbf{e}_s^c$, $c = 1, \ldots, C$.

The evaluators in the last item allow us to analyze more accurately the substitution error behavior in a given observed range: they are computed by subdividing the range of an observed variable (t or T) in subintervals I_c, $c = 1, \ldots, C$, of equal length. The N' items of the test sets are then subdivided in C subsets, according to the chosen variable. For each subinterval I_c, the mean M^c and the standard deviation Std^c of the substitution errors $\mathbf{e}_s^c$ can be computed and analyzed. For example, if we analyze the error behavior during the day we divide the time range in $C_t = 24$ intervals of one hour length, if in the temperature range in $C_T = 40$ of one degree length.

To validate the predictions we have also taken into account that the AAMS is a particular virtual sensor, where no physical modeling feature is involved: it predicts values at the monument location equal to those measured at the ground level. In this case the monitoring operation could be achieved without the placement of any virtual or hard sensors on the monument, but only by using the AAMS surveys. This particular virtual sensor is named *null-sensor* and its corresponding substitution error $e_0(t) = a(k) - r(k)$ can be used as comparison value to get the gain of a virtual sensor (as the ratio between standard errors).

In the study-case of the roman theater in Aosta city we used datasets having $N = N' = 2048$ measurement couples, where the first one is acquired hourly by an AAMS station that is located 40 m nearby the theater and the second one by the west real sensors installed on a pillar, for the train and test periods. In order to identify the wavelet basis that allows us to provide a prediction model more accurate in terms of the given statistical evaluators, a large experimental study has been developed: hundreds of ENN have been designed and trained and only those providing the best overall standard error are chosen for the comparison study; all the virtual sensors are realized by the Elman NN with one hidden layer of thirty neurons, one input neuron and one output neuron, as in [4].

In the wavelet processing we chose Haar, Daubechies bases of minimum order (2 and 4), say db2 and db4 in MATLAB notation, and for each bases, we applied $I = 1, 2$ level of decomposition. These choices are due to the necessity of maintaining the computational time in the training phase low, since $(I + 1)$ ENNs are to be trained and tested.

In our application both the train and the test set are very large, therefore the problem of boundaries handling in the wavelet procedures can be solved in different ways, since the first predicted items, being not accurate, can be disregarded. Fig. 3 reports the wavelet decomposition (high and low coefficients) with db2 at level 2 of the input signal acquired by the AAMS.

For each particular choice, the specific ENN for the low/high coefficients has been initially designed with a number of 30 neurons as in [4] and trained using the Matlab toolbook [5]. Among the twenty trained ENNs, the best one, i.e. with the best standard error, has been chosen for simulating a specific low/high coefficient series.

In Fig. 4 the target signal, measured by the sensor installed on the monument, is the upper plot; the second plot refers to the substitution error for the virtual sensor realized as pure ENN; the third one is the substitution error for the virtual sensor with db2 (1 level) pre-processing; the bottom one with db2 (2 level) pre-processing. We also tried db2 at higher levels, but severe problems in convergence arose in training the ENN for high frequencies, due to successive downsampling operations.

The statistical results have been compared with those obtained without pre-processing and with the null sensor (see table in Fig. 5). For every choice (excepting Haar, 2 level) the overall Std has been improved as compared with Elman Std. Best values of standard error and of correlation are given for db2, 1 level. This pre-processing provided a gain of 0.30% as comparing the Std with the null sensor Std, instead of 0.26% provided by comparing the simple ENN sensor. Moreover, the plots of the standard errors $\mathbf{e}_s^c$ subdivided in day hours (Fig. 6) or in temperature range (Fig. 7) showed better behavior, excepting highest values, when compared with those obtained by simply using the ENN sensor. We can summarize these results saying that the application of the wavelet decomposition and the separate training of

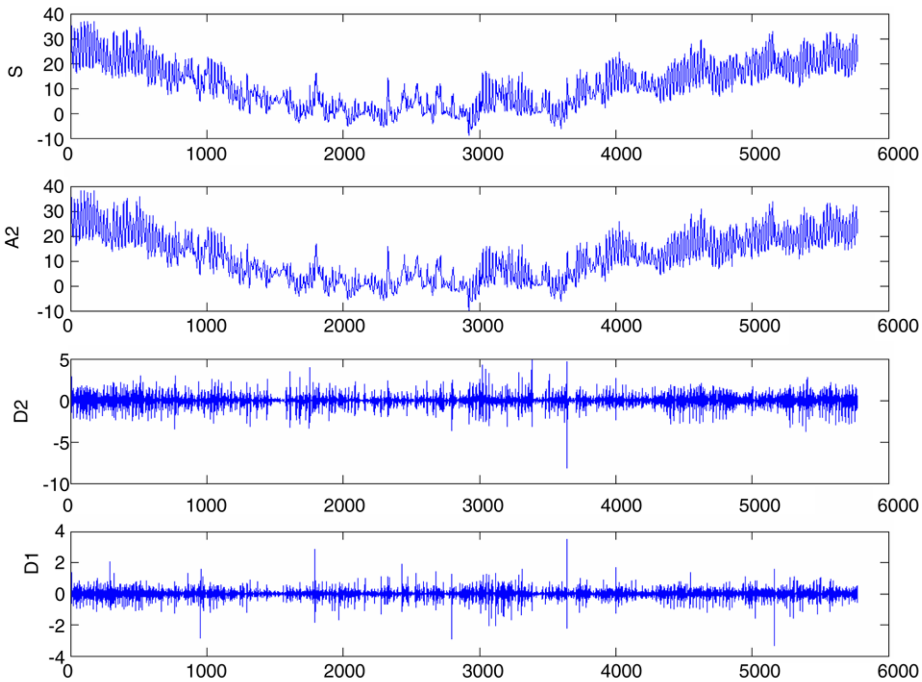

Fig. 3. Wavelet analysis of the input air temperature signal (ground location) with db2 basis (2 level).

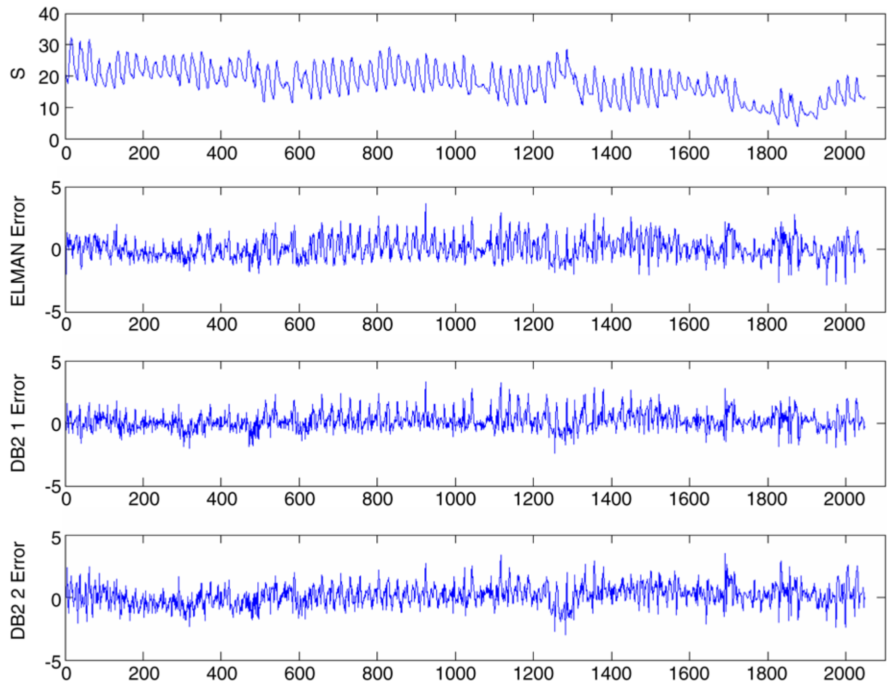

Fig. 4. Real signal and substitution errors for ENN virtual sensor, with db2 (1 level) pre-processing, with db2 (2 level) pre-processing.

several ENNs for high/low components allow us to improve the convergence in the learning time and the effectiveness in the working time, but with an increasing of the number of Neural Networks to be trained and tested at each level.

	Mean	Std	*RR*	*CC*
Daubechies 4 1th	0.13	0.81	0.9295	0.9936
Daubechies 4 2nd	0.15	0.83	0.9435	0.9927
Daubechies 2 1th	0.22	0.72	0.9933	0.9944
Daubechies 2 2nd	0.16	0.81	0.9586	0.9928
Haar 1th	0.24	0.81	1.0427	0.9936
Haar 2nd	0.14	1.43	0.9967	0.9803
Elman	0.10	0.85	0.9993	0.9928
Null Sensor	1.74	2.8	1.3473	0.9898

Fig. 5. Comparison among different virtual sensors (statistical evaluators for the substitution error and a test set of size $N = 2048$).

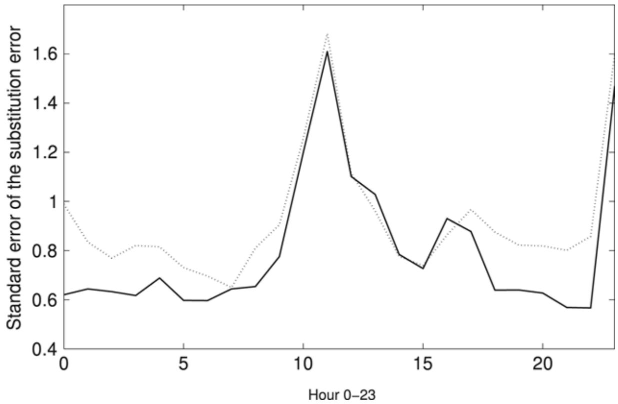

Fig. 6. Comparison of the substitution errors during the day for $C_t = 24$ (test set of $N = 2048$): Elman sensor (dotted line), Elman sensor with db2 pre-processing (1 level).

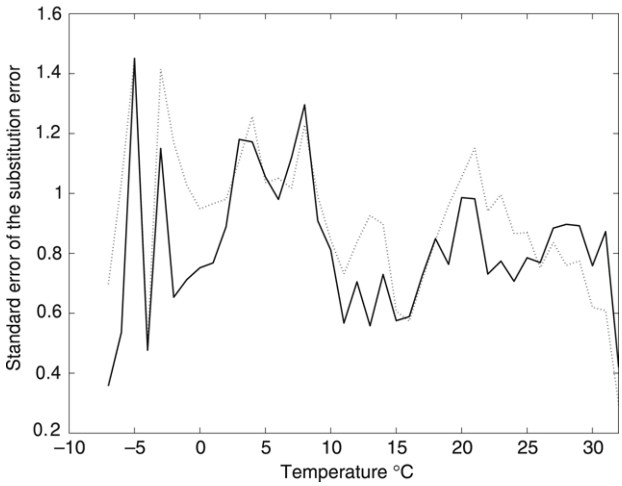

Fig. 7. Comparison of the substitution errors in the observed temperature range for $C_T = 40$ (test set of $N = 2048$): Elman sensor (dotted line), Elman sensor with db2 pre-processing (1 level).

5. Conclusions

Wavelet decomposition and information fusion by Elman Neural Network have been fruitful adopted in modeling complex ambient phenomena, such as air temperature in locations on a monument. The data pre-processing via wavelet analysis, besides the learning capability of Neural Networks, allowed obtaining a more accurate modeling of time varying ambient variables. We have proved its effectiveness in a cultural heritage application where the monitoring of ambient variables must be performed for long periods. The prediction accuracy of virtual sensor estimates with wavelet processing for a long working period has been improved of 12.5%, when its standard error has been compared with that given by using ENN only. This benefit can be due to a better treatment of low frequency anomalies in the two observed ranges, time and temperature and to the combination of several dynamical recurrent Neural Networks aiming at capturing the specific dynamics of the multiresolution versions of datasets. However, the wavelet processing increases the number of Neural Networks to be trained and tested for each virtual sensor to be realized.

The achieved standard errors have been considered negligible by the cultural heritage expert for the monitoring task. Therefore our soft sensors, working in place of the hard sensors, will realize a not-invasive monitoring of the environmental variables around a monument.

References

[1] A. Aussem, F. Murtagh, Combining Neural Network forecasts on wavelet-transformed time series, Connection Science 9 (1) (1997) 113–121.
[2] J. Elman, Finding structure in time, Cognitive Science 14 (1991) 179–211.
[3] U. Maniscalco, A connectionist system of virtual sensors to support the monitoring of cultural heritage damage, in: Proc. of WIRN 2004, Morlacchi Ed., Perugia, Italy, 2004.
[4] U. Maniscalco, P. Ciarlini, G. Regoliosi, Validation of soft sensors in monitoring ambient parameters, in: Ciarlini, et al. (Eds.), Advanced Mathematical and Computational Tools in Metrology VII, in: Series on Advanced in Mathematics for Applied Sciences, vol. 72, World Scientific, Singapore, 2006, pp. 142–150.
[5] http://mathworks.com/products.
[6] F. Morabito, M. Versaci, Neural Networks, in: Proc. IJCNN'02, Vol. 1, 2002. pp. 432–437.