13th Workshop on ADC Modelling and Testing
Sep. 22-24, 2008, Florence, Italy

Wavelet and Mixture of Soft Sensors to improve the Monitoring of Environmental Parameters By Neural Network

Patrizia Ciarlini[1], Umberto Maniscalco[2]

[1] *Istituto per le Applicazione del Calcolo -CNR, Viale del Policlinico 137 00161 Rome, Italy*
ciarlini@iac.rm.cnr.it
[2] *Istituto DI Calcolo e Reti ad Alte Prestazioni-CNR, Viale delle Scienze Palazzina 11, Palermo, Italy*
maniscalco@pa.icar.cnr.it

Abstract- - Soft sensors, based on Elman NNs, have been developed to provide virtual measurements at different locations on the monument surface using as input source only the measurements acquired by an Air Ambient Monitor Station located nearby. Simulation of measurements by trained NN is a useful computational tool to monitor the physical or chemical conditions of the composing materials in a not invasive way, but their accuracy has to be high as analyzed from a metrological and statistical point of view. Two different mathematical and computational tools can be adopted to improve the accuracy of the virtual measurements: a wavelet preprocessing of times series data and the mixture of soft sensors to fuse several input sources**..**

I. Introduction

Neural networks are able to model ambient parameters and air pollution with more advantages over general statistical methods, due to their ability to recover not linear behaviors of air dynamic phenomenon [1]. Due to their properties, such as flexibility, they have been used to develop soft sensors, to be used instead of real sensors, to get measurements in complex experimental conditions or particular locations in an innovative way.

This paradigm was firstly adopted in an application in cultural heritage to realize a non-invasive monitoring the physical or chemical conditions of materials composing a monument [2, 3]. Monitoring is a long-term process, which obliges to maintain several real sensors on the monument surface for a long time to periodically repeat the sample campaigns at high costs and reducing the enjoyment of the monument itself.

The paper describes two different strategies used to to improve the accuracy of these virtual measurements as applied to the case study of Roman Theater in Aosta City. They are base on different mathematical and computational tools: a wavelet decomposition when times series data are available; the mixture-of-experts when several input sources are available for train neural networks (NN). The first tool refers to combining wavelet and NN mechanism to better recover low frequencies characterizing the physical signals: it can be used when data are acquired uniformly for long time, both in train and test phase. The wavelet approach has been used by the authors in predicting data of physical parameters related to the surface of a monument [5].

The second tool, the mixture technique, allows designing multi-sensor systems, which are interesting either when the measurement accuracy provided by a single soft sensor is not sufficient or when the information required depend on more than a single parameter. This tool can be used in the cases of having several sources of input data.

In section II the recursive soft sensors are described as applied in constructing a modular system for the ambient parameters monitoring; the procedure to validate their performances from a metrological point of view is also briefly described. In section III the wavelet preprocessing is described as applied to the environmental time series. In section IV the tool mixture-of-experts is described to fuse input data from two Air Ambient Monitor Station (AAMS). The performances of the soft sensors, developed according to all these different strategies, are finally analyzed in section V and compared by means of a two-phase procedure [4].

II. Soft sensors for monitoring environmental variables

The modular system of soft sensors [2], developed to predict ambient parameter values (air temperature *Ta*, contact temperature *Tc* and humidity, *H*) in four locations on the monument surface, consists in 12

soft sensors of Elman type. Here the monument and the physical/chemical parameters, characterizing the atmosphere, are considered to be a unique system An environmental rich database (available in Aosta) was used as input to the system, containing measurements (*Ta, H*) acquired hourly by an AAMS located at the ground level nearby the monument and measurements (*Ta*, *T*c, *H*) acquired hourly by four real sensors placed on four point on the a pilar. Some data may be missing in these database.
We have adopted the recursive Elman NN for designing and training the soft sensors, since we showed that it enables to provide best performances in comparison with others commonly used [3].
Let's consider the soft sensor that learns the relation between *air temperature* at a certain the time (*hour),* and *contact temperature,* in a specific point of the monument (say 3 the point in the West face of the pilar): this association mapping is multi-values, since to an input value temperature *Ta* may correspond several values of the output temperature *Tc3,* being acquired in different hours of several days for a long observation period (several months).
Afterwards, in the working time the soft sensor must be able to predict the $Tc_3(t)$, to a novel air temperature input from the AAMS, say *a(t).* Let's define the soft sensor of Elman type and number 3, in the following compact notation:

$$S_3(t) = f(a(t), t, \psi(a(t-1)); \mathrm{W}_3) \tag{1}$$

where *f(.)* represents the specific trained Elman neural network that approximates the physical phenomenon, $\boldsymbol{W}_3$ the vector of all the inner model parameters, such as the number of neurons, the weight matrices and biases, the ambient temperature *a(t) is* measured at time *t* by the AAMS, Ψ gives the output of the hidden layer to get the recursive network behavior.
In [4] a statistical procedure to analyze and validate the response of the soft sensors also from a metrological point of view has been defined. This complex validation procedure is based on the substitution error, since it considers the soft sensor as a virtual instrument, which must be substitute a hard sensor in measuring with good accuracy for long periods.
This procedure firstly computes specific estimators, similar to the ones used to give specifications of a real instrument (data sheet); then performs the validation by comparison of the substitution errors, in order to assure the obtained gain in the range of interest. The substitution error for the soft sensor *S* at time t_j is defined as:

$$\mathrm{E}_S(t_j) = S(t_j) - r(t_j) \tag{2}$$

where $r(t_i)$ is the measurement acquired by an hard sensor at a location on the monument for the train/test period, $j=1,\dots,N$ (called target). Our statistical procedure characterizes the behavior of E_s in the observed range of each ambient variable. Let's subdivide the range of an observed variable in *C* subintervals I_c, $c = 1,..., C$ of equal length; consequently the test set of N couples (a_j, r_j) is subdivided in *C* subintervals and for each I_c the substitution errors E_s^c are computed. The overall standard deviation of the substitution error, σ_S, is also computed for the test set. Our procedure validates the soft sensor performances in the whole range of interest, by applying statistical estimators to E_S and to the two sequences of type E_s^c, either by considering the input space subdivided in $C_t = 24$ subintervals (each-one equal to 1 hour), or the temperature space subdivided in $C_T = 45$ (each one equal to 1 °*C*).

III. Wavelet preprocessing

The multi-resolution analysis allowed getting a good treatment of signals with very rapid variation in time, as the ones regarding the temperature signals treated in our case study. The idea of combining wavelet and the NN mechanism has been proposed by several authors, to model non-linear data, mainly in two different approaches: wavelet used as the neuron's activation function; in a preprocessing phase for the extraction of features from the time series data. The second approach was followed in the monitoring application [6], where times series data where available both for the train and for the test period. In this respect we call time-series the data used in this preprocessing and they must be without missing values since in the wavelet decomposition dilation and translation operations must follow the regular grid structure.
Input signal, say *a(t),* is represented as a linear combination of translation and dilations of a mother wavelet. Computationally, the discrete version of a wavelet transform is used, where the frequency space is sampled by using power of 2. DWT is implemented using low-pass and high-pass filters to compute the coefficients (approximation and details coefficients) and it can be iteratively applied to low-pass components for higher levels of wavelet decomposition. Fig. 1 displays the wavelet

procedures for the first level of decomposition., coupled to the NN mechanism, either in the train phase to estimate the NN weights, or in the working phase, to simulate prediction at a novel input.
In the train phase DWT is applied to transforms input data series and known output data series (the target). In the working phase, data prediction and data post processing is performed by IDWT. It must be underlined that this preprocessing at a level of decomposition greater than 1 has a very high computational cost. In fact, for each environment parameter a specific NN must be designed and trained both for the low components and for the high components (training sets are of size $N/2$, if N ,a power of 2, is the size of the input times series) according to the following procedure:

- input the low coefficients ($\boldsymbol{a}_L^{(1)}$, $\boldsymbol{r}_L^{(1)}$) to the NN to compute the weights $\boldsymbol{W}_L$; analogous computation for $\boldsymbol{W}_H^{(1)}$

The DWT operations for decomposition level $I > 1$ are:

- apply DWT to both data series $\boldsymbol{a}$ and $\boldsymbol{r}$ to get two filtered sequences of low-pass components and two of high-pass components, for i=1;
- apply DWT to the low components, going ahead along the usual wavelet tree that has length I;
- train several NNs, i=1,..., I, each NN having ($\boldsymbol{a}_H^{(i)}$, $\boldsymbol{r}_H^{(i)}$) as input and target couples of high components with halved length at each i-th level, to estimate the NN weights $\boldsymbol{W}_H^{(i)}$;
- train only one NN having as input couples of low components obtained at the last level I , to compute the NN corresponding weights.

In the working phase, the DWT is applied only to a new input data series $\boldsymbol{a}$ of given length. In the wavelet reconstruction step, IDWT fuses all these computed predictions at the appropriate scale levels to finally output simulated values of the target quantity $\boldsymbol{r}$.
As regard the choice of the wavelet basis, we considered that usually environmental signals have low regularity. The correction of their time drift can be achieved at low computational costs by wavelet bases with compact support and with low regularity order. The comparison of the performances of different wavelet bases having small size support showed [5] that Daubechies bases of order 2 (at decomposition level 1) gave the best standard error. At higher decomposition levels convergence problem arose for the downsampling operations. To overcome this problem the not-decimated wavelet transform could also be used instead. However, this method, jointly used with the recursive Elman mechanism, heavily increases the computational cost of the training phase, for two reasons: the number of items in each filtered sequence maintains the same value N at every i, the number of NNs that must be trained increases rapidly with i.

IV. Multiple sources sensors and Mixture-of-experts

The design of suitable multi-source soft sensors allows improving the prediction performances when several input sources are available as in our case study where two AAMS, placed not too far from the monument, are working. Indeed, we can realizes the data fusion at different levels and in different ways according to the information also of probabilistic type that can be known.
Averaging repeated measurements, as it is known, enables to reduce the uncertainty of a measured value when each measurement can be considered acquired independently of others and in the same environment conditions. Differently when a soft sensor is used as an “instrument”, being of deterministic type, it provides infinite precision measurements: by repeating the operation in the same conditions it always output the same value. The idea of “averaging” here must be applied differently. We average simulated outputs, if the outputs are obtained by using different virtual instruments but having homogeneous variances (in the sense that all the soft instruments must pertain to the same class). In this metrological framework, we are assessing that we use several soft sensors for measuring the same quantity in reproducibility conditions.
Being available two input sources of the same quantities, say $AAMS_1$ and $AAMS_2$, we construct two soft sensors to measure the same ambient parameter at a given location by using again the Elman network technology and output two measures for the same parameter at the same time. These simulated measures are linear independent because obtained by not-linear recursive mechanism realized using different train set. Thus we can average these two output values to improving the accuracy in measurement.

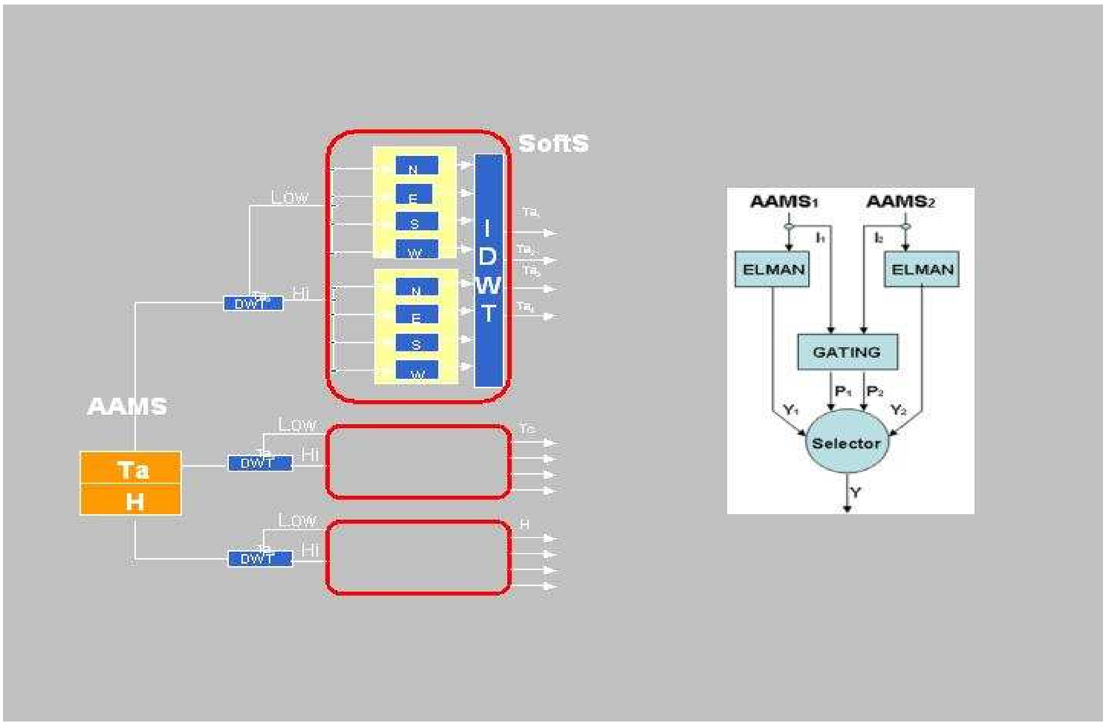

Fig. 1. (Left) The connectionist system topology with the wavelet decomposition (completely displayed only for air temperature parameter): twelve soft sensors of Elman type are used to simulate *Ta*, *Tc* and humidity *H* in four locations, with and post wavelet processing. (Right) The mixture of two local experts whose input sources are two AAMS, with a gating NN.

Besides this natural strategy in building multi-source soft sensors by averaging, the concept of "mixture-of-experts" neural network can be also adopted in the general framework of data fusion.
The "mixture" paradigm [6] allows reducing the complexity of a problem by decomposing the input space (the learning tasks) and the variance of the output by combining multiple sensor predictions in a probabilistic way. Some experts and a gating network characterize a system developed according to this architecture. Each expert, or specialist network, is a neural network and all the experts receive the same inputs and have the same number of output. The gating network can reduce the fitting errors, since it is also a neural network that gives the probabilities of selecting each expert. The benefit of this approach is evident when the learning structure can be well identified, for example using prior knowledge or clustering methods.
In [8] the use of the mixture-of-experts was proposed to achieve a complex mapping function by specialized neurons, since the learning tasks are already divided: the two AAMSs give obviously separate and different inputs, being $AAMS_1$ placed fourth meters far from the monument, $AAMS_2$ some hundreds of meters far. Moreover the association mapping is multi-values for each soft sensor.
Fig.1 (Right) shows a soft sensor with two experts, one gating network and one selector. The two Elman neural NNs are to be independently trained and tested using different train and test sets.
By the use of the two test sets we obtain two error vectors $E_{S1}(t_i)$ and $E_{S2}(t_i)$ (one for each Elman NN). These two vectors are useful to build the output part of the train set for the gating network. In fact, we train the gating network using as input the *i*-th survey obtained from both the two AAMSs and as output a couple of values that represent the normalized conditional probability p_{1i} and p_{2i}:

$$p_{1i} = \frac{p(Y_{1i} \mid I_{1i}, I_{2i})}{p(Y_{1i} \mid I_{1i}, I_{2i}) + p(Y_{2i} \mid I_{1i}, I_{2i})}$$

$$P_{2i} = \frac{p(Y_{2i} \mid I_{1i}, I_{2i})}{p(Y_{1i} \mid I_{1i}, I_{2i}) + p(Y_{2i} \mid I_{1i}, I_{2i})}$$

We can use as p_{1i} and p_{2i}, $1/E_{S1}(t_i)$ and $1/E_{S2}(t_i)$ respectively. Thus, we realize the gating network that

gives two values of conditional probability. These two values represent, substantially, the accuracy with which the two experts have supplied the result in similar input condition during the train phase.
The selector acts as a multiple input, single output stochastic switch; the probability that the switch will select the output from an expert is linked to the its conditional probability returned by the gating network. The results of these two strategies to built multi-source sensors, here outlined as averaging and mixture-of-experts, will be compared from a statistical and metrological point of view in the next section.

V. Results

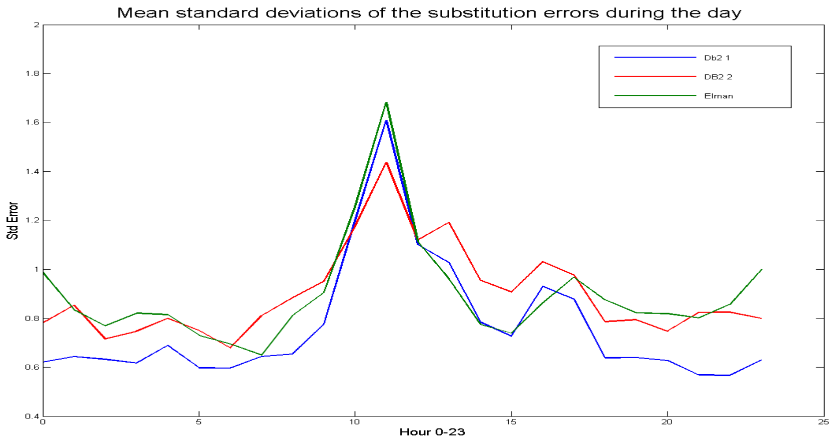

Fig. 2. Plots of the standard deviations of the substitution error (as subdivided for hour) for air temperature: simple Elamn soft sensor (green line);.with wavelet preprocessing Db2 at level 1 (blue line); Db2 at level 2 (red line).

Table 1. Overall standard deviation errors for twelve air ambient parameters: P and Q are single input soft sensors with different input Station, ***A***(Q,P) is the multi-sensor obtained by averaging and ***M***(Q,P) is the mixture-of-experts (performance for each parameter the best standard value in each row is given in bold).

	Q	***P***	***A(Q,P)***	***M*(Q,P)**
Ta1	0.7075	0.6209	0.6233	**0.6012**
Ta2	0.6529	0.6050	0.5919	**0.5701**
Ta3	0.6529	0.5986	0.5920	**0.5810**
Ta4	0.7046	0.6369	0.6289	**0.6025**
Tc1	1.8813	1.8602	1.8083	**1.6920**
Tc2	0.6859	0.6639	0.6367	**0.6102**
Tc3	0.6651	0.5934	0.5792	**0.5601**
Tc4	2.6363	2.6409	2.6172	**2.2451**
H1	4.6086	4.0235	4.0122	**3.8921**
H2	3.9134	3.6341	3.5820	**3.4932**
H3	4.1744	3.9343	3.8079	**3.8002**
H4	5.8507	4.2076	4.4568	**4.1982**

We discuss the accuracy improvements that have been achieved following the previous two strategies, in case study of monitoring the environmental parameters around the Roman Theater. Comparison results for each strategy are performed analyzing differences with the ones by soft sensors based only on the Elman mechanism, with 1 inner layer and a short delay, 30 neurons. Indeed, this model has been shown correct in [3, 4] in constructing an implicit physical model, substantially without bias and in the

same range (temperature, time).
For the wavelet strategy, we showed [5] that the prediction accuracy can be improved of 12.5%, but with high computational costs expecially for the train phase.
For the mixture strategy we have performed an experimental setup with the same train set of size N = 1800 (missing data for some hours are possible), the rule of training fifty Elman NNs and taking the ones that provide similar behavior and standard deviation of substitution error in the range [0.806,0.830].
Table 1 gives the numerical results for the twelve environmental parameters, as simulated by two single input soft sensors and the double input soft sensors of mixture type: ***Q*** represents the Elman soft sensor given by using as input $AAMS_1$, ***P*** the one given by $AAMS_2$; $\boldsymbol{A_{(Q,P)}}$ the soft sensor given by the averaging strategy to design a mixture of two experts; $\boldsymbol{M_{(Q,P)}}$ the mixture soft sensor with the gating network. The columns contain the standard deviations of the substitution error given by the two single soft sensors ***Q*** and ***P*** and by the multiple-source soft *sensors* $\boldsymbol{A_{(Q,P)}}$**)**, $\boldsymbol{M_{(Q,P)}}$.
It can be observed that the concept of mixture (see $\boldsymbol{M_{(Q,P)}}$ results) unables to realize the fusion of two different soft sensors and of different type, and to obtain a new sensor having the best performances in every cases (for each ambient parameter better results are reported in bold).

In Fig. 2 the results of wavelet preprocessing using the Daubechies bases are analyzed during the day (24 hours). The plots of mean value of the substitution errors in time, E_s^c, related to the trained soft sensors with wavelet preprocessing Db2 at level 1 (blue line), with Db2 at level 2 (red line) and without (green line). The mean standard deviations for air temperature are computed having subdivided the 1800 substitution errors in 24 subintervals (1 hour length).

VI. Conclusions

Different strategies for error variance reduction in measuring via soft sensors have been proposed: an averaging procedure that can be use when only one input source is available; the mixture-of-experts and the direct data fusion with a gating NN, when input from different sources are available. Both the strategies have been validated as applied in our cultural heritage study-case.
Every type of implemented soft sensors implemented the Elman recursive mechanism,. Each soft sensor was trained on a rich test set and tested on a similar dataset.
The two types of multi-source soft sensor, designed to make the fusion of two different input sources, ware statistically analyzed and their performances are better than those of single source soft sensors. However the mixture soft sensors, taking advantages on a stochastic switch, predicted values with the most accurate standard error for every ambient parameter. The mixture paradigm can be adopted for combining soft sensors of different types and in different way, according to the available knowledge on probabilities to be associated to each expert.

References

1. A. Fasso’, I. Negri “ Non-linear statistical modeling of high frequency ground ozone data”, *Environmetrics* 13 (3), pp. 225-241, 2002.
2. U. Maniscalco, “A Connectionist System of Virtual Sensors to Support the Monitoring of Cultural Heritage Damnage”, WIRN 2004, Perugia, Italy.
3. A. Chella, P. Ciarlini, U. Maniscalco “Neural Networks as Soft Sensors: a Comparison in a Real World Application. Neural Networks”, in *Proc. of IJCNN '06 International Joint Conference*, p. 2662 – 2668, 2006.
4. U. Maniscalco , P. Ciarlini, G. Regogliosi , “Validation of Soft Sensors in Monitoring Ambient Parameters”, In *Advanced Mathematical and Computational Tools in Metrology VII*, Series on Advanced in Mathematics for Applied Sciences, World Scientific, Singapore. Ciarlini, Felipe, Forbes, Pavese, Perruchet, Siebert Eds, **72**, pp. 142-150, 2006
5. P. Ciarlini U. Maniscalco, “Wavelets and Elman Neural Networks for monitoring environmental variables”, *JCAM* (to appear 2008)
6. M. I. Jordan R. A. Jacobs, “ Adaptive mixture of local Expert Neural Computation“, *Neural Comput.* vol 3, 1991.
7. M. Farinas, C. Pedreira, “ Mixture of Experts Local-Global Neural Networks“, Proc. ESANN’2003, d-side Publ. ISBN2-93037-03-X, pp. 331-336, 2003.