MIXTURE OF SOFT SENSORS FOR MONITORING AIR AMBIENT PARAMETERS

Patrizia Ciarlini, Umberto Maniscalco

[1] Istituto per le Applicazione del Calcolo -CNR, Viale del Policlinico 137 00161 Rome, Italy
{maniscalco,ciarlini}@iac.rm.cnr.it

Abstract Monitoring the physical or chemical conditions of the materials composing a monument can be achieved in a not invasive way by using trained neural networks. Soft sensors based on Elman neural networks have been developed to provide virtual measurements at locations of the monument surface using only the measurements acquired by an Air Ambient Monitor Station located nearby the monument. Here we improve the accuracy of the virtual measurements by using averaging techniques or mixture of such soft sensors**.** The accuracy of these virtual instruments is analyzed and compared from a metrological and statistical point of view.

Keywords: soft sensors; Elman neural network; mixture-of-experts; cultural heritage; statistical data analysis.

1. INTRODUCTION

Soft sensors constitute an innovative paradigm for obtaining measurements in complex experimental conditions or in particular locations [1-2].

This paradigm was firstly adopted for solving a problem in cultural heritage in [3-5]. In this field it is very important to have not invasive tools for monitoring the physical or chemical conditions of materials composing a monument. 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. So doing, this process has high costs and becomes invasive, since it reduces the enjoyment of the monument itself.

A modular system of twelve soft sensors [3], based on recursive neural network of Elman type, was developed to predict ambient parameter values (such as air temperature, contact temperature and humidity) in four locations on the monument surface. The input to the system are real measurements (air temperature, humidity) acquired at the same time by an Air Ambient Monitor Station (AAMS), located nearby the monument.

We underline that in this application a soft sensor is viewed as an “instrument” that provides indirect measurements, in fact the output is in a different geographical position from the location of the input and can be a different ambient parameter.

The aim of this work is to improve the accuracy of the virtual measurements simulated by our soft sensors by designing multi-sensor systems using averaging techniques or the concept of mixture of soft sensors. We address the problem in two different cases: when data acquired by two AAMS (placed nearby the monument in different location) are available and when they are not.

Multi-sensor are commonly developed either because the measurement accuracy provided by a single soft sensor is not sufficient or because the information required cannot be obtained by measuring a single parameter.

The design of suitable multi-source soft sensors can solve the problem of improving the performances in the case of several AAMS, placed not too far from the monument, which provide several inputs of the same quantities. 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.

In section 2 the Elman soft sensors are described as applied in constructing a modular system for monitoring the ambient parameters; the procedure to validate their performances from a metrological point of view is also briefly described.

In order to produce virtual measurements with higher accuracy when real input data comes from one AAMS source only, an averaging procedure of several soft sensors, having homogeneous variances, is defined in section 3. In section 4 the tool mixture-of-experts is proposed to achieve the data fusion from two AAMS, besides the construction of a soft sensors simply having a multi-source input.

The performances of the soft sensors, developed according to all these different strategies, are finally analyzed in section 5 and compared by means of a two-phase procedure [4].

2. ELMAN NEURAL NETWORK FOR SOFT SENSORS

2.1. A modular system for monitoring ambient parameters.

In [4-5] a description of the Elman methodology for designing and training the modular system of twelve soft sensors (Fig. 1), is given. In the monitoring application each sensor had to learn a complex predictive model for an air ambient parameter at a particular location on the surface of a monument. Here the monument and the physical/chemical parameters, characterizing the atmosphere, have been considered to be a unique environmental system.

A soft sensor learns the complex relation between a specific physical or chemical quantity, measured by the AAMS, and the quantity measured in a specific location on the monument surface by a real sensor: for example the input vector contains *air temperature* and the *hour*, the output vector contains *contact temperature*, which have been observed at that hour. The association mapping is multi-values, since it has learnt that to an input value might correspond several values of the output ambient parameter in a specific location on the monument, at the same hour but in different days.

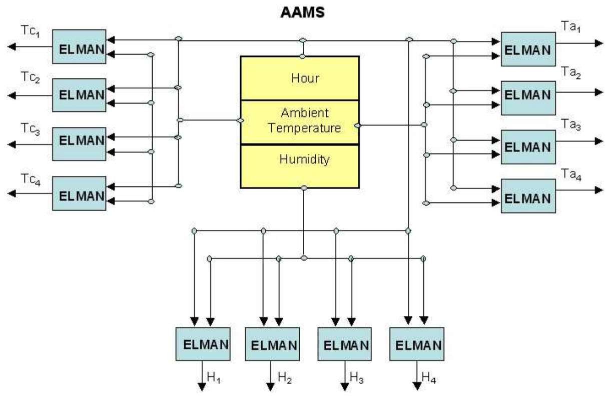

Fig. 1. The connectionist system topology: twelve soft sensors of Elman type to measure contact temperature, humidity and air temperature in four locations of a monument (labels 1-4 correspond to the 4 locations).

Afterwards, in the working time the soft sensor will be able to predict, for example, the *contact temperature* at the same location of the surface, $Tc_3(t)$, to a novel air temperature input from the AAMS, $a(t)$. Let's define the soft sensor of Elman type, which estimates the contact temperature values in the third location (corresponding to the West cardinal position) on the monument surface, in the following compact notation:

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

where *f(.)* represents the specific trained Elman neural network that approximates the physical phenomenon, $\boldsymbol{W}_3$ the matrix of all the inner model parameters, such as the number of neurons, the weight matrices and biases. In this example S_3(t) approximates Tc_3(t). In *f(.)* the input are the ambient temperature $a(t)$, measured at time t by a AAMS, the corresponding day hour (h) at which a is acquired, the function Ψ that gives the output of the hidden layer to get the recursive behavior of the network.

2.2. The procedure to validate the soft sensor performance

Usually in the validation of neural networks only statistical evaluators that compute overall values are used to assess the success of the predictions for a new dataset (Test set). We defined instead a more complex validation procedure [4] that considers the soft sensor as a virtual instrument, which must be able to substitute a hard sensor in measuring an ambient parameter with good accuracy in the observed domain.

The statistical analysis of the response of the soft sensors in the modular system in Fig.1 is then achieved by means of a new procedure, which has two phases: the statistical phase based on specific estimators, similar to the ones used in making specifications of a real instrument (data sheet); the validation by comparison of the substitution errors, which allows to assure the gain of the soft sensor.

The substitution error (or prediction errors) for the soft sensor S at time t_i is given by:

$$E_S(t_i) = S(t_i) - r(t_i) \qquad (2)$$

where $r(t_i)$ is the measurement acquired by an hard sensor that is working at the location of the monument for the training/testing period (i = 1,…, N). The procedure statistically characterizes the behavior of E_s in the observed range of each ambient variable: the range of an observed variable is subdivided in C subintervals I_c, c = 1,..., C of equal length; consequently the Test set is subdivided in C subsets and for each I_c the mean value and the standard deviation are computed for the corresponding substitution errors E_s^c. The overall standard deviation σ_S of the substitution error is also computed for all the N item in the Test set. Our procedure allows to validate the soft sensor performance in the whole domain of interest by applying several estimators to E_S and to the two sequences E_s^c, which are obtained by subdividing the input space in C = 24 for the variable h, or C = 45 for the variable a.

For any of the twelve sensors in Fig.1 substitution errors were analyzed and zero-mean Gaussian-like distributions ware shown. The procedure assessed that each soft sensor based on the Elman recursive mechanism, with a short delay, have constructed an implicit physical model that is correct, since the predicted measurements are substantially without bias, in the same range (*temperature, hour*) of the real data *r(t)* and without any particular trend in *temperature* or during the day.

It must be underlined that if two AAMSs are working nearby the monument, we could predict the contact temperature by means of two different soft sensors and they would probably provide different estimates of the same quantity, say S'_3 and S''_3, as it will be discussed in section 4.

3. SINGLE INPUT SOURCE AND MULTISENSOR SYSTEM

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 the others and in the same environment conditions. Differently when a soft sensor, viewed as an "instrument" that provides indirect measurements, is used, being of deterministic type, we have infinite precision measurements: by repeating the operation in the same conditions we always obtain the same value.

To gain in variance reduction of the errors, when input from only one AAMS is available, the following averaging strategy can be adopted, i.e. by using several soft sensors trained in the same conditions to simulate the same quantity.

The idea of averaging different simulated outputs can be viewed as the averaging operation of repeated measurements, obtained by using different instruments pertaining to the same class (in the sense that all the instruments must have homogeneous error variance). In this metrological framework, we are assessing that we use several soft sensors for measuring a quantity in reproducibility conditions.

The implementation of this idea is driven from the consideration that the estimates for the parameters W_3 in the Elman model are obtained by solving not linear systems in the training phase and that they depend on the chosen starting point and some thresholds.

This means that different soft sensors could be constructed by choosing different starting and working conditions, therefore several soft sensors can be trained for an ambient parameter using the same Training set, but different numerical strategies.

We can assure that all these virtual instruments are of the same type, since they have been generated using the same neural structure and the same Training set. Moreover, they are to be considerate independent, because the relations among them are not linear and the specific values of weights and biases $\boldsymbol{W}$ are different, being recursively estimated, for each neural network. Indeed, the recursivity and the not linearity of the Elman mechanism in Eq. (1) guarantee different evolutions of the states, thus reaching different neural networks.

As it is known, the variance is reduced when repeated data to be averaged are corrupted by zero-mean random uncorrelated errors. In the context of soft sensors the statistical assumption of no correlation is not fully satisfied, hence the error reduction using the average operator may be less than usual, say less than k, the total number of the averaged items.

The averaging procedure to build such *average* soft sensor for a specific ambient parameter is the following:

- for the same Training set, generate several soft sensors, choosing different initial points;
- among them, choice the ones pertaining to the same class, in the sense that they give almost the same standard deviation (homogeneous sensors), say k this total number;
- activate these k soft sensors and work for the Test set;
- average their k outputs to get the output $P^{(k)}(t_i)$ for each item of the set.

4. MULTIPLE SOURCES AND MIXTURE-OF-EXPERTS

In order to obtain the variance reduction in the case of several input sources the neural network mechanism can be directly use to do the data fusion in a natural way.

Being available two inputs of the same quantities, say $AAMS_1$ and $AAMS_2$, we can construct a new sensor to measure an ambient parameter at a given location by using again the Elman network technology, but now with three inputs. For example to simulate the contact temperature in eq. (1) we build $U(t)$, with $a_1(t)$, $a_2(t)$, $h(t)$ as inputs (consequently also in the recursive part Ψ), where $a_1(t)$, $a_2(t)$ are ambient temperature values acquired by $AAMS_1$ and $AAMS_2$, respectively. This new soft sensor $U(t)$ is a *multi-source* soft sensor that associates contact temperature values to a tri-dimensional input space.

Besides this natural strategy in building multi-source soft sensors, the concept of "mixture-of-experts" neural network can be adopted in this general framework of data fusion.

The "mixture" paradigm [6] allows to reduce the complexity of a problem by decomposing the input space (the learning tasks) and the variance of the output by combining multiple sensor predictions. A system developed according to this architecture is characterized by some experts and a gating network. 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 [7] the use of the mixture-of-experts was proposed to achieve a complex mapping function by specialized neurons.

In our application the learning tasks are already divided, since the two AAMSs give obviously separate and different inputs: one $AAMS_1$ is placed fourth meters far from the monument, while $AAMS_2$ some hundreds of meters far. Moreover the association mapping is multi-values for each soft sensor.

To realize a generic soft sensor we consider a system with two experts, one gating network and one selector, as in Fig. 2.

The two Elman neural networks are independently trained and tested using different train and test as described in section 2.

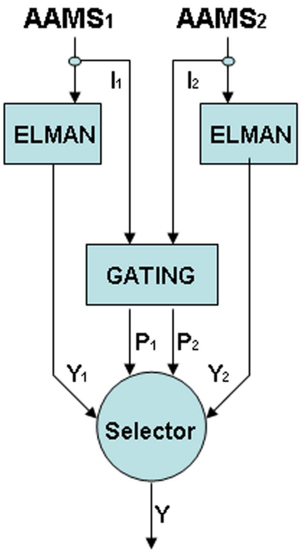

Fig. 2. The mixture of local expert system.

By the use of the two Training sets two vectors of errors $E_{S1}(t_i)$ and $E_{S2}(t_i)$ (one for each Elman Neural Network) are computed. These two vectors are useful to build the output part of the Train set for the gating network. In fact, we must train the gating network using as input the surveys obtained from the two AAMSs and as output a couple of values that represent the normalized conditional probability p_{1i} and p_{2i}:

$$
\begin{aligned}
p_{1i} &= \frac{p(Y_{1i} \mid I_{1i}, I_{2i})}{p(Y_{1i} \mid I_{1i}, I_{2i}) + p(Y_{2i} \mid I_{1i}, I_{2i})} \\
p_{2i} &= \frac{p(Y_{2i} \mid I_{1i}, I_{2i})}{p(Y_{1i} \mid I_{1i}, I_{2i}) + p(Y_{2i} \mid I_{1i}, I_{2i})}
\end{aligned}
\tag{3}
$$

We can use as p_{1i} and p_{2i}, $1/\ E_{S1}(t_i)$ and $1/\ E_{S2}(t_i)$ respectively. Finally, the gating network that gives two values of conditional probability can be obtained. These two values represent, substantially, the accuracy with which the two experts have supplied the results in similar condition of input during the train phase.

The selector works like 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.

If information about the prior distribution is not known or it is difficult to train the gating network, an uniform distribution as a not informative prior can be adopted. So doing, an equal weight 1\2 is given to each expert to mean that both the experts contribute in the same way.

Let us name Q the soft sensor given by using the input $AAMS_1$ (*single-input sensor)* and P the single-input soft sensor given by $AAMS_2$; $P^{(k)}$ the *average* soft sensor given by the averaging procedure in section 3 (for k soft sensors having homogeneous behavior) when the single source $AAMS_2$ is available; U the soft sensor built when multiple data (both from $AAMS_1$ and from AAM S_2) are available and used as inputs to be directly fused by the Elman recursive mechanism; $M_{(Q,P)}$ the *mixture soft sensor* of two Elman sensors Q and P. The last two type of soft sensors can be viewed as *multi-source input sensors*.

5. RESULTS

The strategies, outlined in the previous sections, will be compared from a statistical and metrological point of view as applied to solve the monitoring problem of ambient parameters in different locations on the roman theater in Aosta city (Italy).

We want to compare the accuracy improvement that has been achieved in predicting ambient parameters by the several type of soft sensors.

The air temperature parameter in the first location (in the East side of the theater) will be deeply analyzed, however we have trained and validated a huge quantities of soft sensors for the twelve ambient parameters introduced in section 2.

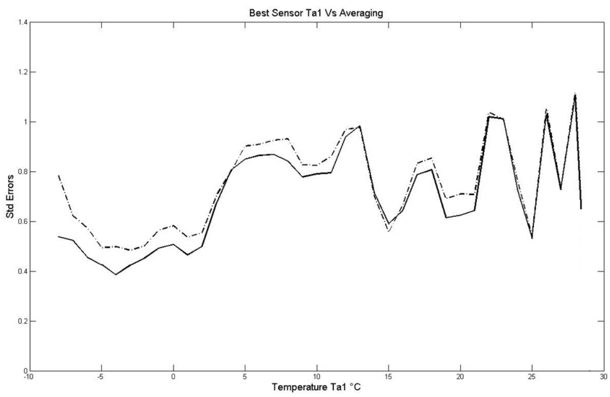

Fig. 3. Plot of the standard deviations of the substitution error (as subdivided for *temperature*) of the best trained soft sensor *P* (dash-dot line) among the 9 soft sensors used in the averaging procedure; plot of the average soft sensor $P^{(9)}$ (solid line).

For the averaging procedure we have performed an experimental setup, training fifty Elman Neural Networks, using the same Training set (size N = 1348) from $AAMS_1$ and choosing among them the ones that provide similar behavior to the best one (here we are considering results relative to the air temperature T_{a1}). Finally we have obtained k = 9 soft sensors, which have a standard deviation of the substitution error in the range [0.806,0.830]. We have analyzed the results given by the soft sensor $P^{(9)}$ according to the two-phase procedure and the Test set of almost N surveys, belonging to the same half year of the Training set: the measures of the ambient parameters were taken hourly, but missing data for some hours are possible in both sets.

In Fig. 3 the plot related to the soft sensor P having the smallest overall σ_P = 0.806 is compared with the one of $P^{(9)}$, which provides a reduced value $\sigma_{P(9)}$ = 0.75. The standard deviation errors of $P^{(9)}$ and of the best sensor P correspond to the standard deviations of the substitution error for a Test of size 1800 that has been subdivided in C = 40 subsets of the *temperature* range (a similar behavior is given also in *hour* range). The averaging procedure reduces the standard deviation for the most part of the surveys.

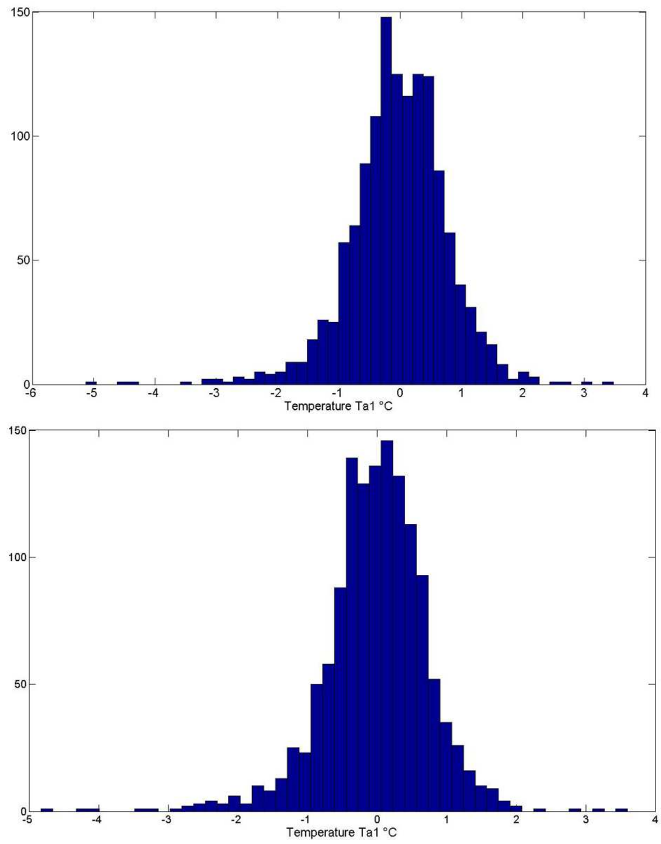

Fig. 4. Histogram of the standard deviations for the substitution error (for temperature T_{a1}) of the best soft sensor *P* (top) among the 9 soft sensors used in averaging procedure and the average soft sensor $P^{(9)}$ (bottom).

Figure 4 shows that both the best sensor P and the average sensor $P^{(9)}$, give errors having a Gaussian-like distribution but with a slightly longer left tail. Moreover, the error distribution for $P^{(9)}$ is more centered than for P.

As to concern the mixture strategy applied with equal probability weights we built two different soft sensors: $\boldsymbol{M_{(Q,P)}}$ as the mean of the outputs of the single soft sensors $\boldsymbol{Q}$ and $\boldsymbol{P}$ and as the mean of three outputs of soft sensors of two different types $\boldsymbol{Q}$, $\boldsymbol{P}$ and $\boldsymbol{U}$. Fig.5 shows that the error distribution for $\boldsymbol{M_{(Q,P,U)}}$ has similar tails, but a bimodal shape, which is probably due to the mixture mechanism.

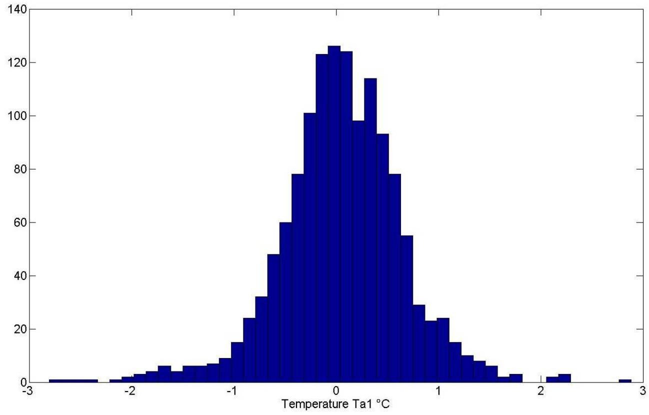

Fig. 5. **Histogram of the standard deviations for the substitution error (for temperature T_{a1}) for the mixture sensor $M_{(Q,P,U)}$ providing the best performances in Table 1.**

Table 1 reports the numerical results for three air ambient parameters (air ambient temperature, air contact temperature and air humidity) in four points of the monument achieved by different types of soft sensors (single input type, multi-source type, mixture-of-experts), which have been trained for a larger set (N = 1800) in several ways (in the Table 1 the soft sensor having the best σ value for an ambient parameter is reported in bold).

Table 1. Overall standard deviation errors for twelve air ambient parameters: P and Q are single input soft sensors, U is a tri-dimensional input soft sensor, $M_{(Q,P)}$ and $M_{(Q,P,U)}$ are mixture-of-experts (performance for each parameter the best σ is given in bold).

	Q	***P***	***U***	$M_{(Q,P)}$	$M_{(Q,P,U)}$
Ta1	0.7075	0.6209	0.6631	0.6233	**0.6148**
Ta2	0.6529	0.6050	0.6096	0.5919	**0.5785**
Ta3	0.6529	0.5986	0.6295	0.5920	**0.5899**
Ta4	0.7046	0.6369	0.9940	**0.6289**	0.6846
Tc1	1.8813	1.8602	1.8986	**1.8083**	1.8088
Tc2	0.6859	0.6639	0.6577	0.6367	**0.6268**
Tc3	0.6651	0.5934	0.5977	0.5792	**0.5564**
Tc4	2.6363	2.6409	2.5715	2.6172	**2.5632**
Ha1	4.6086	4.0235	4.0168	4.0122	**3.8597**
Ha2	3.9134	3.6341	4.0497	3.5820	**3.5734**
Ha3	4.1744	3.9343	4.0774	3.8079	**3.7398**
Ha4	5.8507	**4.2076**	4.7761	4.4568	4.3150

It can be observed that the concept of mixture allows to realize the fusion of three different soft sensors and of different type (column $\boldsymbol{M_{(Q,P,U)}}$), and to obtain the best variance reduction. In fact the performances for every ambient parameter are the best for ten ambient parameters out of twelve.

6. CONCLUSIONS

Different strategies for error variance reduction in measuring ambient parameters have been proposed in building new soft sensors: an averaging procedure of single-source soft sensors with homogeneous behavior to be used when only one input source is available; the mixture-of-experts and the direct data fusion by the Elman mechanism having a tri-dimensional input space, when input from multi sources are available.

Every type of soft sensors implemented the Elman recursive mechanism and was trained on a rich Training set and tested on a similar Test set to validate their performances in the cultural heritage application.

The averaging procedure, applied to single–input soft sensors having similar standard deviation, succeeded in reducing the standard errors.

The multi-source soft sensor, which directly makes the fusion of two different input sources, was designed, implemented and statistically analyzed. Its performance is similar and not much better than the one of the best single source soft sensor.

The mixture paradigm was 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.

The mixture of three soft sensors (two single-input sensors and a multi-input sensor) predicted values with improved accuracy for almost every ambient parameter. The validation procedure applied to this new sensor substantially revealed a superior behavior in monitoring the ambient parameters at each hour of the day and in all temperature range for a long period. However, the averaging procedure of several of such soft sensors could be again applied to this type of multi-source sensor, based on the mixture paradigm, to further improve the variance reduction.

ACKNOWLEDGMENTS

We would like to express our best thanks to the Dr. Lorenzo Appolonia, ``Soprintendenza della Regione Valle d'Aosta", and to Prof. Antonio Chella, Department of Computer Science, University of Palermo, for their fundamental contributes.

REFERENCES

[1] H. Ressom, H. Voos, L.Litz, P. Schmitt, “On-line Estimation of Key Quality Parameters in Nonwoven Production,”.Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, pp. 1745-1749, Nashville, TN, 2000.

[2] J. Sevilla, C. Pulido, “Virtual industrial sensors trough neural networks. Demonstration examples in nuclear power plants”, Instrumentation and Measurement Technology Conference. Conference Proceedings. IEEE Vol. 1, pp. 293-297, 1998.

[3] A. Chella, U. Maniscalco, “Virtual Sensors to Support the Monitoring of Cultural Heritage Damage”, IEEE Tran. on Instrum. and Measurement. (submitted).

[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] U. Maniscalco, “A Connectionist System of Virtual Sensors to Support the Monitoring of Cultural Heritage Damnage”, WIRN 2004, Perugia, Italy, to be appear in Neural Nets (Series: Lecture Notes in Computer Science).

[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.