



HAL
open science

Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals

Piero Baraldi, Antonio Cammi, Francesca Mangili, Enrico Zio

► **To cite this version:**

Piero Baraldi, Antonio Cammi, Francesca Mangili, Enrico Zio. Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals. *IEEE Transactions on Nuclear Science*, 2010, 57 (2), pp.793-806. 10.1109/TNS.2010.2042968 . hal-00609155

HAL Id: hal-00609155

<https://centralesupelec.hal.science/hal-00609155>

Submitted on 25 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals

Piero Baraldi, Antonio Cammi, Francesca Mangili, and Enrico E. Zio

Abstract—Sensors are placed at various locations in a production plant to monitor the state of its components and accordingly operate its control and protection. For the plant state monitoring to be effective, the sensors themselves must be monitored for detecting anomalies in their functioning and for reconstructing the correct values of the signals measured. In this work, the task of sensor monitoring and signal reconstruction is tackled with an ensemble of Principal Component Analysis (PCA) models. The novelty of the work consists in the investigation of local fusion (LF) strategies for the aggregation of the outcomes of the different models of the ensemble. In the reconstruction of a signal, each model of the ensemble is assigned a weight and a bias related to the error committed in the reconstruction of training patterns similar to the one under reconstruction. Iteration of the reconstruction procedure and use of past measurements of the signals are introduced for improved performance.

The proposed methodology is applied to a case study concerning the reconstruction of seven signals in the pressurizer of a Pressurized Water Reactor (PWR) nuclear power plant.

Index Terms—Local fusion, pressurizer, random feature selection ensemble, signal monitoring, signal reconstruction.

I. INTRODUCTION

SENSORS contribute to the safe and productive operation of a nuclear power plant by conveying information on its state to the automated controls and the operators. For this to be effective, it is important to timely detect sensor malfunctions and reconstruct the incorrect signals before using them in the operation, control and protection of the plant [1], [2].

An approach to sensor signal validation relies on auto-associative models [3]–[5] e.g. based on Principal Component Analysis (PCA) [6], AutoAssociative Kernel Regression (AAKR) [7] and AutoAssociative Neural Network (AANN) [4] methods. The practical problem, however, is that a single reconstruction model cannot handle the multiplicity of signals measured on a real plant [8]–[11]. A possible way to overtake this limitation is to subdivide the signals into small overlapping groups, develop an ensemble of reconstruction models, one for each group, and finally combine their outcomes. Key to building of the ensemble is the diversity of the individual models. In the approach investigated in this work, diversity is promoted by randomly generating the signals groups according to the Random Feature Selection Ensemble (RFSE) technique [12]; this is a completely random

technique in which no optimization of the composition of the individual groups is sought, i.e., no relevance is given, for example, to the correlation between the signals in the groups or to their capability of reconstruction. The groups thereby created are then used to develop a corresponding number of signal reconstruction PCA models [13]–[16].

The investigation of different methods for aggregating the outcomes of the individual models of the ensemble is the main objective of the present work. In particular, LF methods are explored, in which the aggregation is guided by the local performance of each model, i.e., its reconstruction accuracy on signal patterns of training similar (and for this reason also called neighbors) to those to be reconstructed (also called test patterns) [17]. These methods rely on the idea that each model can perform well in some regions of the signals space and poorly in others.

The general LF process is based on the following three steps:

- 1) retrieve neighbors of the test pattern from the set of training patterns;
- 2) associate a weight and a bias to each individual model of the ensemble depending on its reconstruction accuracy on the retrieved neighbors;
- 3) aggregate the outputs, accounting for the models weights and bias.

With respect to step 1), two strategies have been compared for the identification of the neighborhood of a test pattern: the first one considers all the patterns of the training set located in a hyper-rectangle centered on the test pattern of interest; the second one includes only its k nearest neighbors.

With respect to step 2), for a given test pattern the bias correction of each model in the ensemble is taken equal to the local mean error (me) which the model makes on the neighbor patterns of training, whereas different weighting strategies have been considered:

- a) weight proportional to the inverse of the mean absolute error (mae) made by the model on the neighboring patterns of training;
- b) weight proportional to the logarithm of the inverse of the mae ;
- c) the borda-count method [18].

With respect to step 3), the output aggregation is performed by a weighted average with the weights computed at step 2 modifying the model outcomes by subtracting the bias.

In order to enhance the robustness of the method, a variation of these strategies has been investigated in which the mae is computed on training patterns to which a noise has been added.

Also, to improve the accuracy of the reconstruction, past signal measurements are used as further input to the reconstruction models and the reconstruction of the faulty signals is iterated until satisfactory convergence.

Manuscript received November 05, 2009; revised January 21, 2010 and January 29, 2010. Current version published April 14, 2010.

The authors are with the Department of Energy, Polytechnic of Milan, 20133 Milano, Italy (e-mail: piero.baraldi@polimi.it; antonio.cammi@polimi.it; francesca.mangili@mail.polimi.it; enrico.zio@polimi.it).

Digital Object Identifier 10.1109/TNS.2010.2042968

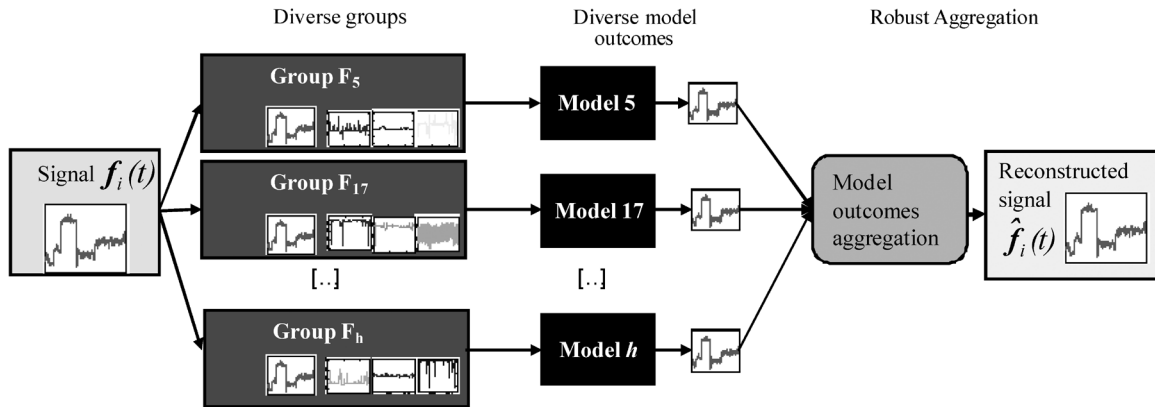


Fig. 1. Approach to the reconstruction of signal f_i contained in groups F_5, F_{17}, \dots, F_h .

The proposed aggregation techniques are compared with the classical Simple Mean (SM) method of signal aggregation [10], [19] on a case study concerning the reconstruction of seven signals simulated by a SIMULINK model of a pressurizer of a Pressurized Water Reactor (PWR) nuclear power plant. Faults have been added to the simulated signals in order to test the robustness of the Local Fusion (LF) and its effectiveness in fault detection and signal reconstruction.

The remaining part of the paper is organized as follows. Section II states the problems of fault detection and signal reconstruction; Section III recalls briefly the RFSE technique; Section IV describes the methodologies for fusing multiple models outcomes, while Section V presents the results from a set of experiments based on the reconstruction of the signals measured by the sensors of a simulated PWR pressurizer; finally, Section VI presents the conclusions and describes potential future work.

II. FAULT DETECTION AND SIGNAL RECONSTRUCTION

A sensor measurement f_i provides an estimate of the true value f_i^T of a physical quantity; ideally, it should be $f_i = f_i^T$, but in practice the measurement f_i is affected by random noise and sometimes it also deviates from the true value f_i^T because of a sensor fault.

Under normal conditions, sensor measurements are strongly correlated by the physical relations governing the process. Then, validation of sensor measurements is possible by taking advantage of the redundant information coming from the correlated measurements; this allows detecting incorrect measurements, identifying faulty sensors and reconstructing the correct values of measurements. More specifically, fault detection and identification aim at detecting if a measured signal f_i is no longer a reliable estimation of the true signal value f_i^T , due to an identified sensor malfunctioning. The detection of a sensor fault is often achieved by comparison between the actual sensor measurement f_i and the measurement estimate \hat{f}_i provided by a physical or empirical model reproducing the response of the sensor in non-faulty conditions; many techniques, like the sequential probability ratio test (SPRT) technique [20], [21], are based on the monitoring of the residual signal $\varepsilon_i(t) = f_i(t) - \hat{f}_i(t)$. Faulty signal reconstruction uses the information contained in other correlated signals to infer a more satisfactory estimate of

the true value f_i^T of the physical quantity: we call this estimate the reconstructed value \hat{f}_i^T of the faulty signal f_i .

The main objective of the present paper is the investigation of techniques for reducing the deviation between the estimate \hat{f}_i of the quantity measured by a faulty sensor and its real value f_i^T . The improvement of the estimate \hat{f}_i brings two main benefits:

- i. a more reliable estimate of the residual $\varepsilon_i(t)$ which leads to a more prompt and accurate detection of the sensor faults;
- ii. a more accurate reconstruction \hat{f}_i^T of the faulty signal in case of sensor fault.

The first beneficial effect above mentioned is not further developed here, since the definition of an explicit procedure for sensor faults detection and identification is outside the scope of the present work.

III. THE RANDOM FEATURE SELECTION ENSEMBLE APPROACH

Fig. 1 reports a sketch of the flow of modeling for the reconstruction of a faulty signal. A set of n sensors' signals f_i , $i = 1, 2, \dots, n$, is available for building a model for the reconstruction of a detected faulty signal. Given the typically large value of n , a single model cannot perform the reconstruction task with the desired accuracy and reliability. Then the signal set must be partitioned into subsets, for each of which a reconstruction model is built. Actually in the ensemble approach here proposed the subsets of signals are overlapping (i.e., two subsets may contain signals which are the same), a reconstruction model is built for each subset and the reconstruction of a signal contained in different models is obtained by aggregating their outcomes, within an ensemble approach.

In this work, the individual models employed in the ensemble for signal reconstruction are based on the PCA technique; the motivation is the need for a trade-off between computational time and accuracy of the reconstruction. Other techniques such as AANN and AAKR have given satisfactory accuracies [22] but are unfit to be used in an ensemble approach, due to their very high computational cost (long training time for AANN, long on-line execution time for AAKR). PCA is based on the transformation of the data to a smaller set of variables which are linear combinations of the original variables; the transformation is made so as to retain as much information as possible. The principal components correspond to the directions in which the

projected observations have the largest variance: the largest variance comes to lie on the first principal component, the second largest variance on the second principal component, and so on. Principal components are included in the transformation until the cumulative variance associated to the retained components reaches a fixed fraction F_{Var} of the total variance. Appendix B gives some details on the application of PCA to signal reconstruction.

In this work, signal grouping is performed by the RFSE technique which consists in randomly sampling from the n available signals, with replacement, H subsets \mathbf{F}_h , $h = 1, 2, \dots, H$, each constituted by m signals [12]. This guarantees high signal diversity in the overlapping groups upon which the PCA models are built and allows for rapid construction of the signal groups. Furthermore, randomly selecting the signals in the groups with a reasonable choice of the group size parameters m and H , can basically guarantee coverage of all the signals in the ensemble with adequate redundancy [10], [19].

As mentioned above, the H diverse signal groups generated are used as bases for developing a corresponding number of PCA reconstruction models. To do this, the data set \mathbf{X} of N signal patterns available is partitioned into a training set \mathbf{X}_{TRN} (made of N_{TRN} patterns) and a test set \mathbf{X}_{TST} (made of N_{TST} patterns). The former is used to train the individual models, whereas the latter is used to verify the ensemble performance in the signal reconstruction task.

Each signal i is present in a number H_i of groups and thus a corresponding number of individual PCA models provide its reconstruction. Different methods can be used to aggregate the outcomes of these individual models in the ensemble, to get the final reconstructed value. Some of these will be discussed in the next Section IV.

After the ensemble of PCA models has been trained, its performance is verified on the test data: the entire set of n signals is fed into the ensemble of models which returns their reconstructed values. In particular, each model h receives in input and reconstructs only the signals belonging to the corresponding subset \mathbf{F}_h .

IV. LOCAL FUSION STRATEGIES FOR MODELS OUTCOME AGGREGATION

A. Motivations

Let $\hat{f}_i^{\mathbf{X}_{TRN}}$ be the reconstruction of the i -th signal of a test pattern made by a generic model built on the training set \mathbf{X}_{TRN} , and f_i^T its real value. Since the model depends on the actual data used to build it, the randomness of the training set \mathbf{X}_{TRN} implies that the response $\hat{f}_i^{\mathbf{X}_{TRN}}$ of the model at any point is a random variable. In order to evaluate in a probabilistic sense the quality of $\hat{f}_i^{\mathbf{X}_{TRN}}$ as estimator of f_i^T the mean squared error (*mse*) over all possible training datasets is commonly used:

$$mse = E_{\mathbf{X}_{TRN}} \left[\left(\hat{f}_i^{\mathbf{X}_{TRN}} - f_i^T \right)^2 \right] \quad (1)$$

The *mse* can be decomposed in a variance term plus a bias component [23]:

$$mse = E_{\mathbf{X}_{TRN}} \left[\left(\hat{f}_i^{\mathbf{X}_{TRN}} - \bar{f}_i \right)^2 \right] + \left(\bar{f}_i - f_i^T \right)^2 \quad (2)$$

where $\bar{f}_i = E_{\mathbf{X}_{TRN}}[\hat{f}_i^{\mathbf{X}_{TRN}}]$ is the expected value of $\hat{f}_i^{\mathbf{X}_{TRN}}$ over all possible training sets \mathbf{X}_{TRN} . Notice that the bias component represents the mean squared error that is obtained in the ideal situation where the outcomes of infinite models built over all possible training sets are combined by means of the arithmetic average, since this combination technique compensates the variance component, i.e., the fluctuations around the mean value of the reconstructions obtained by models built on different datasets.

In [24], two issues are raised in this regard:

- 1) although an ensemble of models built on different training sets actually boosts the accuracy of the reconstruction by reducing the variance component, the simple average fusion strategy of the single model outcomes cannot effectively reduce the bias;
- 2) the mean squared error also depends on the test pattern under reconstruction since all terms in (2) depend on it. Therefore, different patterns in the signals space will usually have different error profiles of the model reconstructions.

In this context, a procedure is defined in [24] for the aggregation of an ensemble of models trained over different datasets generated through the bootstrap technique. The proposed fusion strategy aims at favoring the best performing models in each region of the input space and, at the same time, effectively correcting not only the variance component of the error, but also the bias.

In this work, this local strategy for the aggregation of multiple outcomes is applied in the context of multiple models in which diversity is injected through the RFSE technique, since the considerations about the dependence of the model error on the input pattern remains true for each model h built on the signal subset \mathbf{F}_h .

To further delve into the aspect of local variation of the error profile, Fig. 2 and Fig. 3 show some examples of models reconstructions for the water level signal in the PWR pressurizer application.

Fig. 2 represents the time evolution of the water level in the pressurizer during a plant transient (top) and the residual error ($\hat{f}_L^{12} - f_L^T$) made by model 12 in its reconstruction (bottom). Notice that the reconstruction error varies in the different time regions and seems correlated to the signal value. For example, the dotted boxes highlight regions corresponding to flat peaks of the level measurement signal, to which correspond similar trends of the residual error. Thus, in order to correct the error in the reconstruction of a test pattern falling in the region of a flat peak, it would seem more effective to compute the mean error of model 12 considering only the similar (neighbor) training patterns of the highlighted regions, instead of using the mean error on all the available training patterns.

Fig. 3 shows the performances of models 8 and 11 in the reconstruction of the same signal of Fig. 2; although

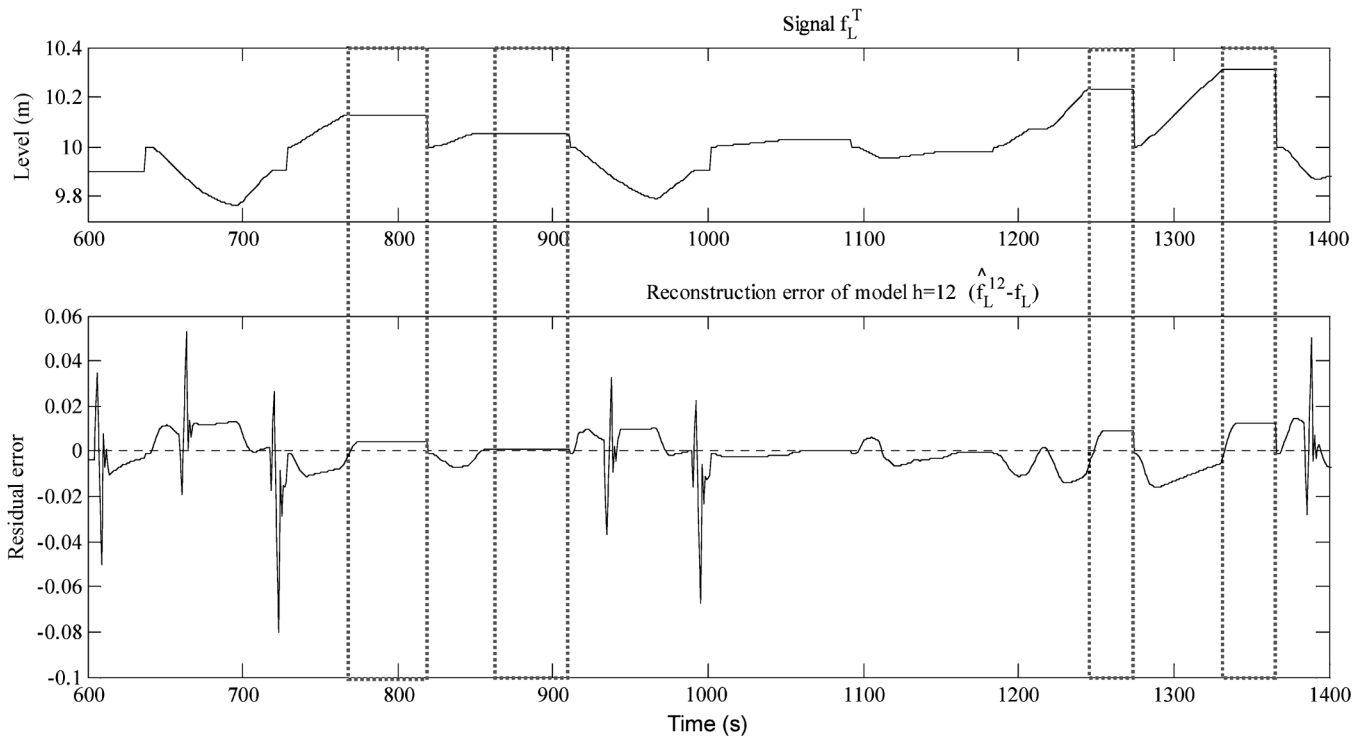


Fig. 2. Top, true signal value; bottom, reconstruction error of model 12.

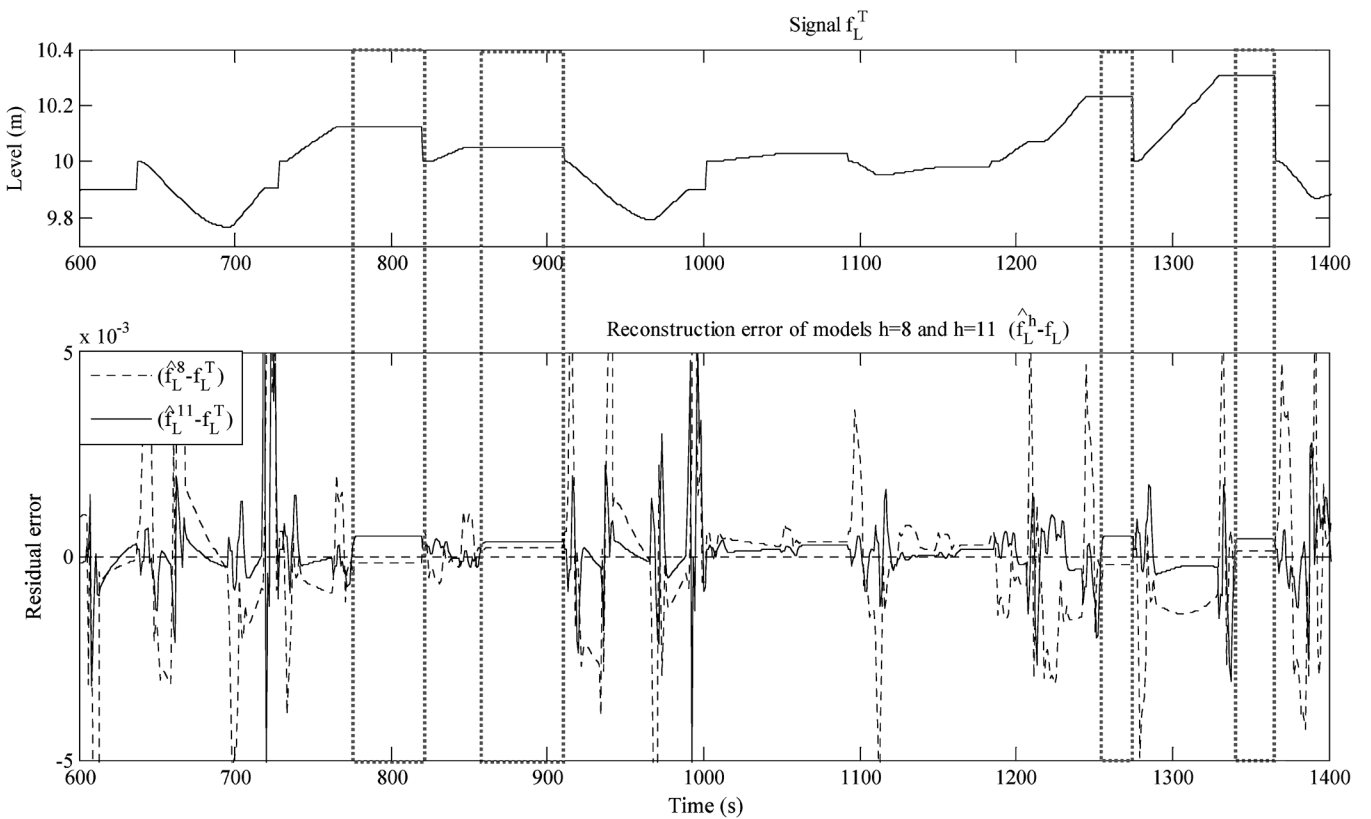


Fig. 3. Comparison between model 8 and model 11 reconstruction errors. Dotted boxes highlight some regions where the model built on the 11th group of signals performs better than the one built on the 8th group.

according to the mean absolute error (*mae*) model 11 is globally performing better than model 8, there are some time regions (highlighted by dotted boxes) in which model 8 is

more accurate. Also this example points out the importance of assigning local weights to the models in the aggregation of their outcomes.

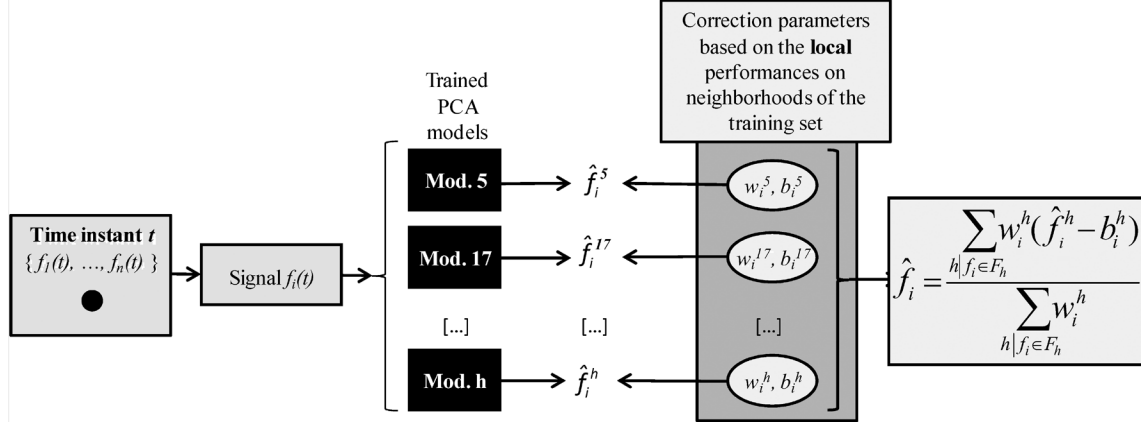


Fig. 4. Framework of the locally weighted fusion of model outcomes.

B. Models Outcome Aggregation

In general, the aggregation of the models outcomes requires to associate a weight w_i^h and a bias correction b_i^h to the reconstruction \hat{f}_i^h of each model h . The idea is to correct the values of \hat{f}_i^h by subtracting the estimated bias b_i^h and to combine \hat{f}_i^h with the other models estimates by means of a weighted average:

$$\hat{f}_i = \frac{\sum_{h|f_i \in F_h} w_i^h \cdot (\hat{f}_i^h - b_i^h)}{\sum_{h|f_i \in F_h} w_i^h} \quad (3)$$

According to the considerations of Section IV-A, the ensemble performance can be increased if both the bias corrections b_i^h and the weights w_i^h are computed locally. Fig. 4 shows the overall framework for the reconstruction of the generic signal f_i .

The following two subsections provide some details on how to select the neighboring training patterns to be used for the computation of the bias and weight parameters and how to associate them to the models of the ensemble.

1) *Neighborhood of Patterns for Bias and Weight Computation:* Let Q_{TRN} be the set of training patterns to be used for the computation of the models biases and weights in the reconstruction of a given test pattern. With respect to this set, two aggregation approaches can be distinguished: global and local. In the former, Q_{TRN} is constituted by the entire training set \mathbf{X}_{TRN} ; in the latter approach, only a subset of the training set is considered and one talks about “Local Fusion” (LF). The main idea behind LF is to take into consideration the location of the test pattern in the signal space and its neighborhood in the training set. To this aim, two different techniques are here considered for the identification of the test pattern neighborhood by: a) an hyper-rectangle and b) the k -nearest neighbors (k -nn).

a) *Hyper-rectangle-based neighborhood:* For each model h , an hyper-rectangle centered in the test pattern is considered in the m -dimensional space of the signals contained in the subset \mathbf{F}_h (Fig. 5). The length of each side of the hyper-rectangle is defined as a fraction Hrf of the range of values of the corresponding signal. The patterns of the training set contained in this hyper-rectangle make up the set Q_{TRN} . A limitation of this approach is that the cardinality $N_{Q_{TRN}}$ of Q_{TRN} varies depending

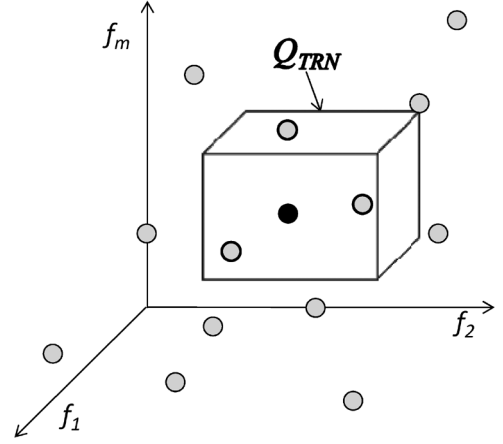


Fig. 5. Hyper-rectangle-based neighborhood. The crossed circle represents the test pattern; the other circles are training patterns.

on the density of the training patterns in the neighborhood of the test pattern.

b) *k -nn-based neighborhood:* Within the k -nn-based neighborhood approach, Q_{TRN} is formed by the k training patterns nearest to the test pattern (Fig. 6). Notice that in this way the cardinality $N_{Q_{TRN}}$ of Q_{TRN} is fixed. A drawback of this approach is that when the neighborhood of the test pattern has a low density of patterns, its dimensions increase and the precision of the estimated parameters is reduced.

2) *Local Weights and Bias Estimation:* In order to evaluate the performance of the generic model h in the reconstruction of the i -th signal of a given test pattern, the model reconstructions \hat{f}_i^h of the training patterns in the neighborhood set Q_{TRN} are compared to the measured signal f_i , which, for a fault-free training set, is supposed to be equal to the true signal value f_i^T , except for the presence of noise. The true signal value f_i^T cannot be used in real applications, since it is not available. To this purpose, the following two indicators are considered:

- the local mean error

$$me_{i,Q_{TRN}}^h = \sum_{Q_{TRN}} (\hat{f}_i^h - f_i) \quad (4)$$

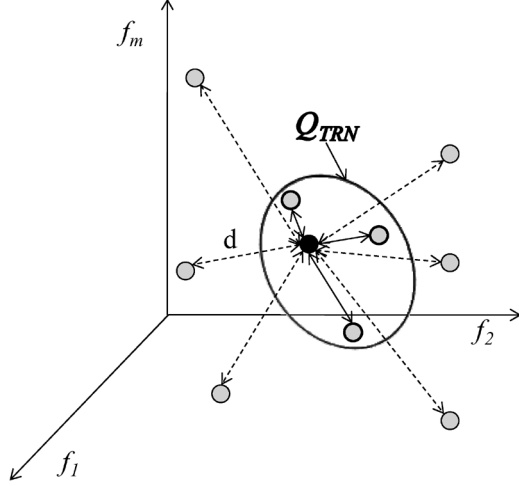


Fig. 6. k -nn-based neighborhood. The crossed circle represents the test pattern; the other circles are training patterns.

- the local mean absolute error

$$mae_{i,Q_{TRN}}^h = \Sigma_{Q_{TRN}} \left| \hat{f}_i^h - f_i \right| \quad (5)$$

Since the local mean error $mae_{i,Q_{TRN}}^h$ represents the mean bias which affects the reconstructions \hat{f}_i^h produced by model h of the training patterns closer to the test pattern, it can be considered an estimate of the bias which affects the reconstruction \hat{f}_i^h of the test patterns. Thus, the local bias affecting model h in the reconstruction of the i -th signal with respect to the test pattern is set equal to:

$$b_i^h = mae_{i,Q_{TRN}}^h \quad (6)$$

The local mean absolute error $mae_{i,Q_{TRN}}^h$ provides information about the performance of model h in the reconstruction of the patterns of the training set which are closer to the test pattern. Therefore, it can be considered an estimation of the error that will affect the reconstruction \hat{f}_i^h by model h and thus be used as estimate of the weight w_i^h to be associated to model h in the reconstruction of signal i . Since the weight must reflect the contribution of the model to the fusion, the most reliable models must have the highest weights, and vice versa.

In this work three different relationships between the weights and the local mean error are investigated:

- 1) The weight is the inverse of the mean absolute error:

$$w_i^h = \frac{1}{mae_{i,Q_{TRN}}^h} \quad (7)$$

- 2) The weight is proportional to the logarithm of the inverse of the normalized mean absolute error:

$$w_i^h = \log \left[\frac{\max_{X_{TRN},h} (\hat{f}_i^h - f_i)}{mae_{i,Q_{TRN}}^h} \right] \quad (8)$$

The normalization factor $\max_{X_{TRN},h} (\hat{f}_i^h - f_i)$ is the maximum value of the residual error over all patterns of the training set X_{TRN} and all models $h = 1, \dots, H_i$; being always greater than the $mae_{i,Q_{TRN}}^h$, it forces the argument of the logarithm to be greater than 1, thus maintaining the value of the weight w_i^h positive.

This strategy, which is inspired by the assignment of the weights in the adaboost algorithm [25], is here employed to attenuate the difference in the weights of groups with very different local mean absolute errors; this can be useful, in particular, when an outstanding performance of a group on the training set is not due to the real quality of the model, but derives from overfitting.

- 3) The Borda Count method [18]. The estimated local error is used to make a ranking of the different models and to assign them a score s_i^h , $1 < s_i^h < H_i$, according to their position in the ranking, being 1 the score associated to the worst performing model and H_i that to the best performing one:

$$w_i^h = s_i^h \quad (9)$$

Another possibility to overcome the overfitting problem is to compute the model performance not on the training data but on a different dataset. In this respect, an artificial noise has been added to the training data before their reconstruction so that the training error used to estimate the weights will also account for the robustness of the model, reducing in this way the risk of giving high weights to models that are overfitting.

All the three proposed strategies defined to assign a weight to the different models ((7), (8), (9)) can be used with the local mean absolute error computed without bias and with the disturbed training error instead of the mae .

V. APPLICATION

The case study used to test the LF strategies presented in the previous Sections concerns the reconstruction of signals measured in a controlled PWR pressurizer. The pressurizer maintains the pressure in the primary system of the plant at an opportune value; in particular, for safety and operational reasons, it is fundamental to have a pressure sufficiently high to avoid boiling of the coolant while the plant is operating.

The data for the analysis (training and test patterns) were generated by means of a SIMULINK model purposely developed. The SIMULINK model is based on a system of non-linear differential equations derived from the mass and energy conservation laws applied to the two regions of vapour and liquid of the pressurizer; exchanges between the two regions, due to evaporation of liquid and condensation of steam, are taken into account. The water and steam properties used in the SIMULINK model are based on the Industrial Formulation IAPWS-IF97 [26].

The variables controlled in the pressurizer are the level and the pressure. Their control is achieved by acting on the heaters, the charging/letdown flows and the relief valve (Fig. 7). The relief valve is a safety device which opens only when the pressure exceeds a limit value (here set equal to 165 bar). Since this limit

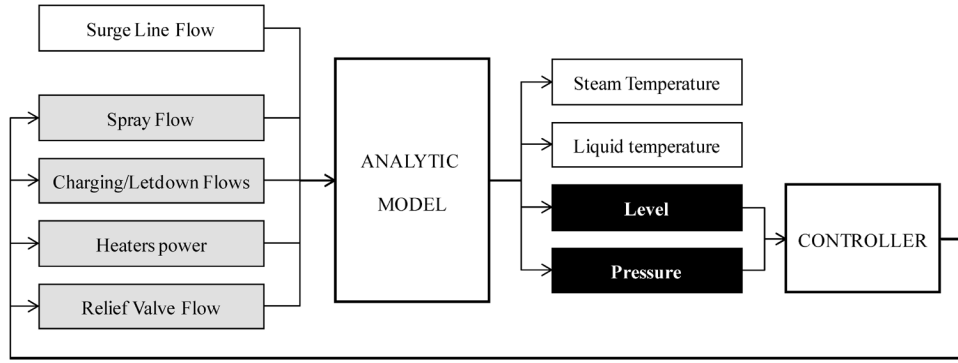


Fig. 7. Input, output and control variables of the SIMULINK model of the pressurizer.

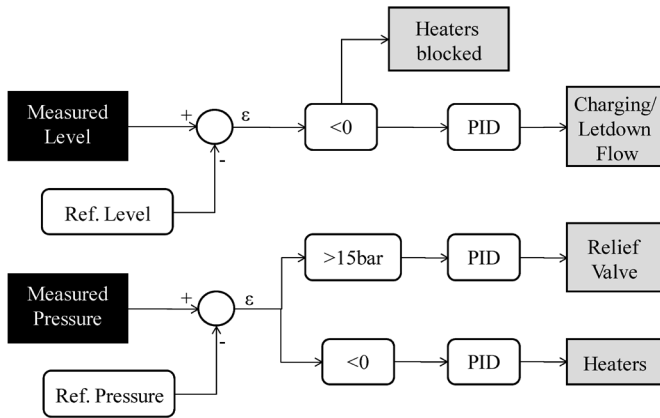


Fig. 8. Pressurizer control scheme.

TABLE I
INITIAL CONDITIONS OF THE PRESSURIZER

	Equilibrium condition
LEVEL	10m
LIQUID TEMPERATURE	322.16°C
STEAM TEMPERATURE	362.16°C
PRESSURE	150 bar

pressure value has never been reached during the simulated transients, the relief valve flow signal has not been considered in this work.

Fig. 8 reports the control scheme adopted. The measured level is compared with the level set point: in case of low level, the charging/letdown flows correct the mismatches and the heaters are turned off independently from the value of the pressure. Besides, when the measured pressure is lower than its set point, the heaters are turned on. The control signal sent to charging/letdown flows, relief valve and heaters is obtained by using a proportional-integral-derivative (PID) controller.

Starting from the initial conditions reported in Table I, 50 transients of 100 time instants each have been simulated by randomly changing the value of the surge line flow entering the pressurizer. Fig. 9 reports the evolution of the signals in one of the simulated transients.

The surge line flow depends on the reactor power; in particular, when the Primary Heat Transport (PHT) system pressure

rises, the higher pressure pushes some coolant from the PHT into the pressurizer (in-surge flow); on the contrary, in case of low PHT system pressure, some liquid moves from the pressurizer into the PHT system circuit (out-surge flow). In this application, the transients have been generated by assuming surge flows in the range of $[-40; +40]$ kg/s, positive values indicating in-surge flow and negative values out-surge flow.

During the simulations, the following seven signals are recorded:

- the level in the pressurizer
- the liquid temperature
- the steam temperature
- the pressure
- the charging/letdown flows
- the surge line flow
- the heaters power.

In order to test the method on pseudo-realistic data, white noise has been added to each signal according to engineering considerations on the sensors accuracy [27]–[29]. Table II reports the standard deviations of the considered noises.

The reconstruction task has been carried out both for clean and noisy data.

A. Sensor Measurement Reconstruction

In order to catch the dynamic evolution of the system, the set of input measurements f_i which are given as input to the reconstruction model is constituted not only by the current values of the 7 signal measurements but also by a sliding window that contains measurements at previous times: in particular, for each one of the 7 signals, the current and the previous 9 measurements have been considered, leading to a total number of 70 input measurements (Fig. 10).

In this application, 2730 patterns, corresponding to 30 transients, have been used to train the PCA models and to compute the single model local mean errors and local mean absolute errors which are the quantities used for the settings of the ensemble local weights and bias, whereas the remaining 1820 patterns, taken from 20 different transients, are used to test its performance. Notice that, in order to avoid any overfitting of the data, the ensemble performance is evaluated on test data that are different from the training data used to evaluate the single model performances which are necessary for setting the ensemble parameters.

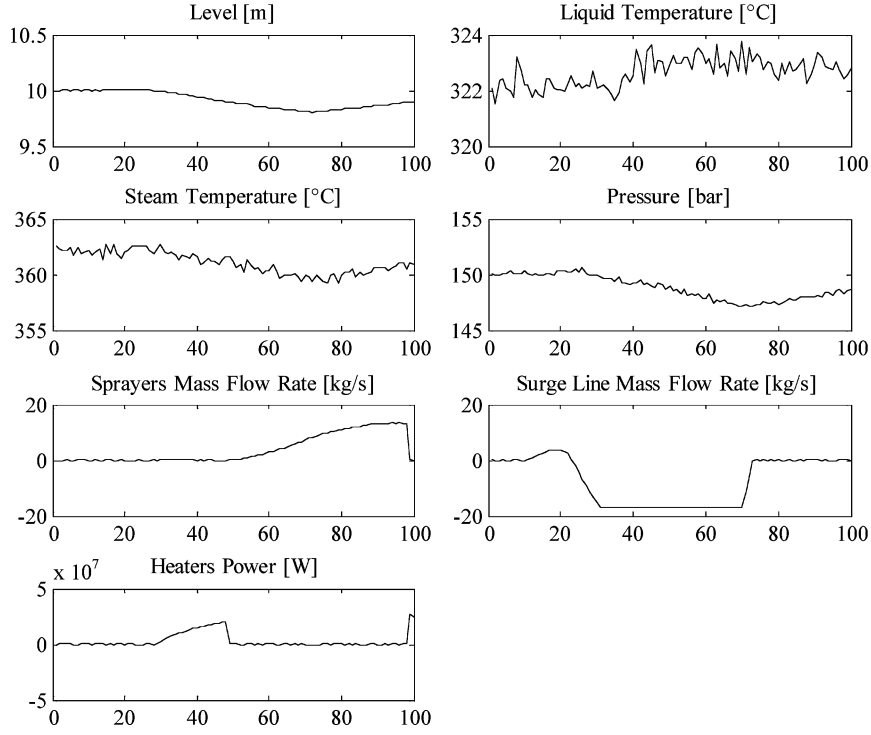


Fig. 9. Evolution of the seven signals during a plant transient.

LEVEL										LIQUID TEMPERATURE									STEAM TEMPERATURE																				
t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9										
PRESSURE										CHARGING/LETDOWN FLOWS									SURGE LINE FLOW									HEATERS POWER											
t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9

Fig. 10. Input pattern used for the reconstruction of the current signal measurement.

TABLE II
STANDARD DEVIATION OF SENSORS NOISE

	Noise standard deviation
LEVEL SENSOR	± 0.01 m
PRESSURE SENSOR	± 0.5 bar
FLOW SENSOR	± 0.2 kg/s
POWER SENSOR	± 50 kW

The optimal number of models to be used in the ensemble H and the optimal number of input signals in each PCA model m have been established through optimization of the performances on fault-free noisy validation data (different from both the training and test data). The best results have been obtained by using 20 PCA models, each one with 25 input measurements. In this way, all signals are reconstructed by at least 7 models. Notice that in order to reduce the computational time, redundancies higher than 7 have not been considered although they may increase the ensemble performance.

The fraction of retained variance for each PCA model is $F_{Var} = 0.99$.

Finally notice that in the evaluation of the performance of the ensemble model, only the reconstruction of the 7 signal measurements at the present time and not their past values have been considered for the estimation of the mean absolute error. The values refer to the signals normalized in the range [0.2; 1].

The following Section V-B considers the problem of reconstructing the true signal values in the case in which there are no failures in the sensors. Since the comparison between the reconstructed signal values \hat{f}_i and the sensor measurements f_i is the base for the identification of anomalies in the sensors, the reconstruction is fundamental for a correct sensor fault detection. In Section V-C, the problem of signal reconstruction in the case of sensor failure will be tackled.

In both cases, the use of a set of signals containing also previous measurements (9 in this work) has demonstrated to be effective. In case of noisy data the mae of a single PCA model built on the 7 present signals is $1.157 \cdot 10^{-3}$, while considering the past measurements the obtained mae is $1.094 \cdot 10^{-3}$, with a reduction of the error of $+5.5\%$. Moreover, as shown in Fig. 11, referring to the case of a constant bias fault of the pressure sensor, a significant improvement is obtained in the signals reconstruction during the first instants after the initiation of a

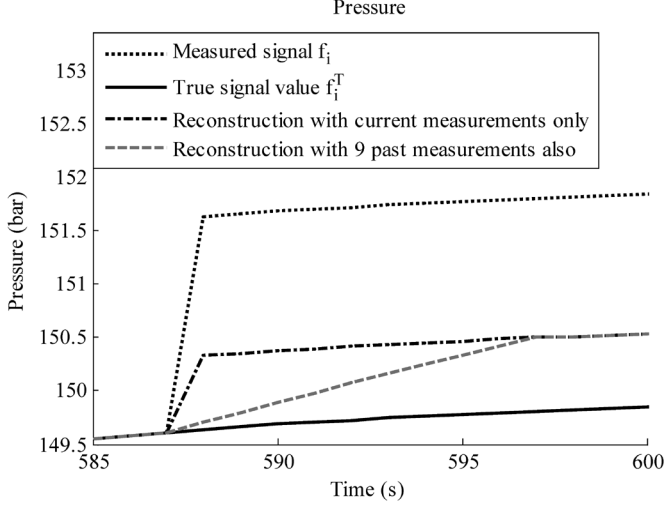


Fig. 11. Comparison of the reconstruction of the pressure signal undergoing a constant bias fault, obtained from a single PCA model built on the current values of the 7 signals and the one built on the set containing also the 9 most recent measurements.

fault. When past information is considered, the signals used for the reconstruction, in the 9 instants just after the fault, have in part been measured before the beginning of the faulty condition; this allows an improvement in the reconstruction of the faulty signal, which facilitates the fault detection and identification. Finally, if the pressure signal is reconstructed by the iteration procedure explained in Section V-C, a mae of $2.691 \cdot 10^{-3}$ is obtained using the past measurements in input, which is 60% lower than the mae of $6.658 \cdot 10^{-3}$ obtained when only the current measurements are used.

B. Signal Validation for Sensor Fault Detection

In this Section, the results obtained by using the proposed LF strategies are compared with those obtained by using two global approaches: 1) the Simple Mean (SM) and 2) the Globally Weighted Average (GWA). The latter is realized by assigning to the h -th model reconstruction of signal i a weight proportional to the inverse of its mae_i^h , computed on the whole training dataset [24].

Table III shows the performances obtained in the reconstruction of clean (top) and noisy (bottom) data with the different fusion strategies.

Several runs have been carried out to set the values of the hyper-rectangle length Hrf and the number of nearest neighbors $N_{Q_{TRN}}$, in order to get the best results for fair comparison. The value of Hrf has to be high enough to avoid empty neighborhoods in correspondence of any test pattern. Notice that in the case of clean data, the Hrf value which leads to the best results is the smallest one respecting the condition. This is in accordance with the fact that the optimal number of nearest neighbors is very low ($N_{Q_{TRN}} = 4$) in case of clean data.

When noise is added to the test data measurements, the optimal values of Hrf and $N_{Q_{TRN}}$ increase. This is due to the fact that, when noise is present, the error committed by a model h in the reconstruction of signal i of a generic pattern is the result of two contributions: the modeling error $\varepsilon_i^h = \hat{f}_i^h - f_i^T$ and the

TABLE III
MEAN ABSOLUTE ERROR (mae) OBTAINED BY USING DIFFERENT COMBINATION STRATEGIES

CLEAN DATA		
Fusion strategy	MAE (10^{-3})	
1. Simple Mean	2.1130	
2. Globally Weighted Average	1.1399 (+46%)	
	$Hrf = 0.25^a$	4-nn ^b
3. Constant common weight	1.599 (+24%)	0.8572 (+59%)
4. Local Weight + Bias correction	0.6272 (+70%)	0.3754 (+82%)
NOISY DATA		
Fusion strategy	MAE (10^{-3})	
1. Simple Mean	10.319	
2. Globally Weighted Average	10.045 (+2.7%)	
	$Hrf = 0.3^a$	90-nn ^b
3. Constant common weight	10.144 (+1.7%)	10.112 (+2.03%)
4. Local Weight + Bias correction	9.944 (+3.6%)	9.895 (+4.12%)

The reported percentages refer to the improvement of the performance with respect to the SM. ^a Hyper-rectangle-based neighborhood; ^b k-nn-based neighborhood.

noise error ε_n , i.e. the error resulting from having inaccurate inputs to the model. The ε_n error component cannot be properly estimated using the training data because of the random nature of the noise. The $me_{i,Q_{TRN}}^h$ results from the average of the local errors made on the $N_{Q_{TRN}}$ neighbors ((10)).

$$me_{i,Q_{TRN}}^h = E_{Q_{TRN}} [\varepsilon_i^h + \varepsilon_n] = E_{Q_{TRN}} [\varepsilon_i^h] + E_{Q_{TRN}} [\varepsilon_n] \quad (10)$$

When we consider a large number of neighbors, the last term of (10) can be approximated to 0, so that the contribution of noise to the error estimate is eliminated. However, as a consequence of the increased value of $N_{Q_{TRN}}$, the local character of the procedure is reduced because a wider region of the input space is considered; thus the quality of the parameters estimation is decreased and the performance of local and global techniques tend to become more similar.

In order to distinguish the contribution to the improvement of the performance of the bias correction from that of the local weights, Table III reports in the third row the performance obtained by assigning a constant weight equal to $1/H_i$ to all the models and correcting the bias locally, according to (3). In the fourth row, Table III reports the results obtained by applying the complete LF scheme of (3), which includes bias correction and local weighting, using a weight proportional to the inverse of the mae ((7)). Notice that the main contribution to the improved performances of the local fusion is due to the local weights, whereas the bias correction conveys only a small amelioration, particularly in case of noisy signals.

Finally, notice that the best strategy for the selection of the neighbors seems to be the k -nn-based neighborhood, given that it leads to lower reconstruction errors than the hyper-rectangle-based neighborhood, in both cases of clean and noisy data.

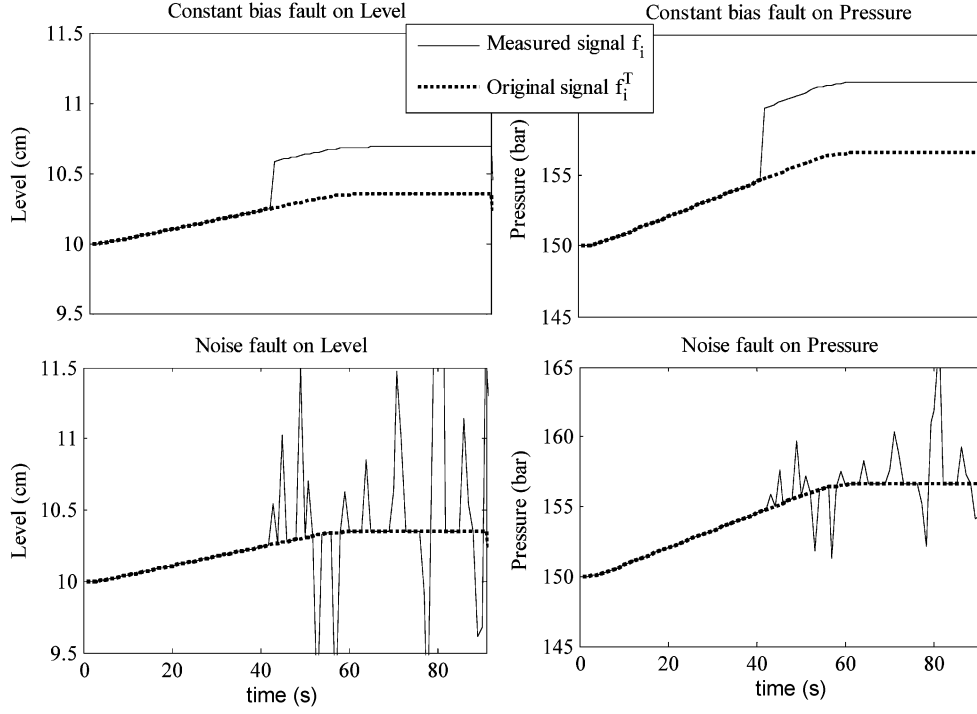


Fig. 12. Examples of signals in fault conditions.

C. Reconstruction in Case of Faulty Sensor

When one of the plant sensors is faulty, a faulty signal is sent in input to the PCA models which include that signal. The ensemble should however still provide a good estimate of the true value of the signal, by exploiting the correlated information coming from the non-faulty signals. To verify the performance of the different local fusion techniques in dealing with sensor faults, a number of tests have been performed as described in this Section.

Two faults types are considered: constant bias faults (CF) and noise faults (NF), affecting both the pressure and level sensors. These two sensors have been chosen since their measurements influence the controller decisions and thus are particularly critical for plant operation. The faults are supposed to start in all the transients at $t = 50$ s, with various intensities. Fig. 12 shows some examples of sensor faults analyzed.

The LF signal reconstruction approach based on the k -nn-based neighborhood (which has given the best results in the previous Section), has been considered with $N_{QTRN} = 4$ in case of clean training data and $N_{QTRN} = 90$ in case of noise. The different techniques introduced in the previous Section IV-B-2 for the estimation of the local weights have been investigated and compared to the SM global approach. Table IV reports the mae obtained in the reconstruction of the true value f_i^T of the signal measured by the faulty sensor.

The results show that in the case of faults affecting the pressure sensor, the SM performs better than the LF techniques. Also, among the local techniques, the best performing weighting mechanism is the one based on logarithmic weights. The reason of this is that some models achieve a very high performance in the reconstruction of the training data just because they tend to exactly reproduce the input measurements and

TABLE IV
 $mae(10^{-3})$ OF THE FAULTY SIGNAL RECONSTRUCTION

		PRESSURE FAULT				
		MAE (10^{-3})				
		SM	LF			
			$1/\varepsilon$	Log	Borda	NT
CF	Clean data	24.41	25.36 (-3.9%)	24.69 (-1.2%)	24.77 (-1.5%)	23.44 (+4.0%)
	Noisy data	23.92	24.34 (-1.8%)	24.22 (-1.3%)	25.60 (-7.0%)	24.05 (-0.54%)
NF	Clean data	6.713	6.784 (-1.1%)	6.447 (+4.0%)	6.609 (+1.6%)	6.439 (+4.1%)
	Noisy data	7.365	7.391 (-0.36%)	7.347 (+0.24%)	8.029 (-9.2%)	7.385 (-0.27%)
		LEVEL FAULT				
		MAE (10^{-3})				
		SM	LF			
			$1/\varepsilon$	Log	Borda	NT
CF	Clean data	29.86	29.06 (+2.7%)	29.01 (+2.9%)	30.68 (-2.8%)	28.05 (+6.1%)
	Noisy data	31.47	31.06 (+1.3%)	31.06 (+1.3%)	31.44 (+0.10%)	30.85 (+2.0%)
NF	Clean data	31.20	31.76 (-1.8%)	30.34 (+2.8%)	31.96 (-2.4%)	28.15 (+9.8%)
	Noisy data	33.69	32.60 (+3.2%)	33.00 (+2.1%)	34.94 (-3.7%)	32.46 (+3.7%)

SM = simple mean; LF = local fusion, k -nn-based neighborhood; $1/\varepsilon$ = weights computed with eq.(7); Log = weights computed with eq.(8); Borda = weights computed with eq.(9); NT = weights computed with eq.(7), adding noise on training data; CF = constant fault, NF = noise fault.

thus they are associated with very high weights; however, such models are not robust in the case of sensor faults; nevertheless, according to the good performance on the training set, the LF

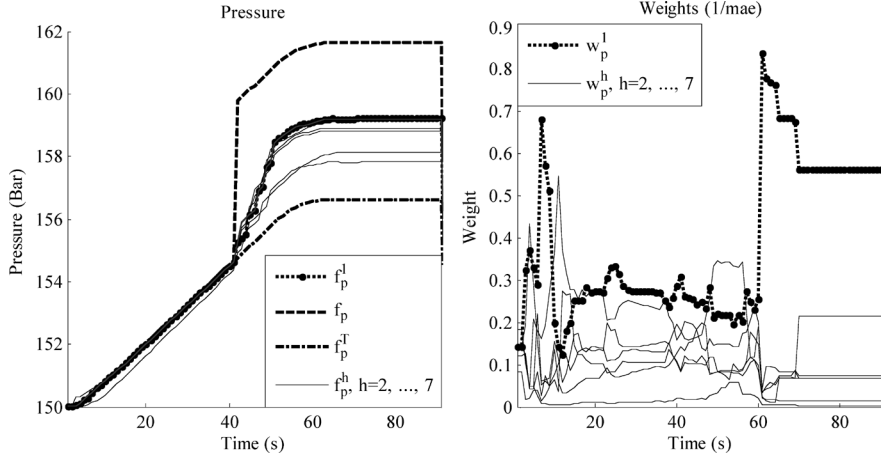


Fig. 13. Weights associated to the 7 models of the ensemble using the training data and corresponding reconstruction of the pressure during a plant transient.

mechanism will assign to them weights excessively high compared to the other models. A logarithmic weight mechanism reduces the weights associated to these non robust models and thus improve the local fusion performance with respect to the $1/\varepsilon$ technique, but its performance is still worse than the SM weighting.

In order to investigate this effect, Fig. 13 (left) shows the reconstruction of the pressure in a transient obtained by the seven models of the ensemble which use as input measurements the actual pressure, while Fig. 13 (right) shows the weights assigned to the models according to their performance on the training dataset. Notice that from approximately $t = 60$ s, model 1 has an exceptionally high weight even if, when the fault is present, it is the worst performing one. In this example, the use of logarithmic weights permits to reduce the predominance of model 1 in the reconstruction, thus improving the ensemble performance. However, even with a logarithmic weight mechanism, the weight associated to model 1 would be greater than the weight associated to the actual best performing model.

To overtake this limitation of the local fusion techniques, it is possible to estimate the actual robustness of the models by computing the reconstruction error not on the training dataset but on a new data set obtained by adding noise on the training data. The results shown in the 4th and 8th columns of Table IV confirm our expectation: the new weights assigned to the pressure reconstructions of Fig. 13 (left) are shown in Fig. 14.

Finally, notice that once the fault has been detected, in order to correctly reconstruct the true signal value, it is possible to repeat the reconstruction of the faulty signal N_{iter} times, using as input to the ensemble model at the n -th iteration the value \hat{f}_i^{n-1} of the reconstructed measurement at the previous iteration [22]. This technique has proven very effective in the reconstruction of the measurements of faulty sensors as it can be seen in Table V, which reports the results obtained by repeating the reconstruction $N_{iter} = 5$ times. First of all, notice the reduction of the mae obtained by repeating the reconstruction. Furthermore, it is interesting to observe that computing the weight by adding a noise on the training data (NT technique) leads to the best signal reconstruction in the case of noisy data, whereas, as expected,

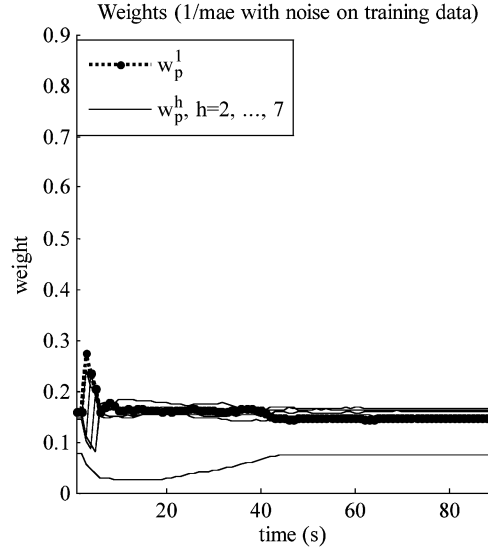


Fig. 14. Weights associated to the reconstructions of Fig. 13, obtained adding a noise to the training data.

it is not effective in case of clean data where the best results are obtained by the logarithmic weighting technique.

D. Results Discussion

The results reported in Sections V-B and V-C show that the local fusion approach performs better in the signal validation task, than the two global techniques considered, SM and GWA.

Concerning the choice of the local neighborhood for the fusion, the best results are obtained using a nearest neighbors approach. When data are affected by noise, the number of neighbors to be considered increases significantly.

In case of sensor failure, the best results are obtained by assigning the model weights proportional to the inverse of the absolute error computed on training data to which a noise has been added. This strategy permits to avoid assigning high weights to overfitting models, i.e., models that perform very well on the

TABLE V
 $mae(10^{-3})$ OF THE FAULTY SIGNAL RECONSTRUCTION AFTER 5 ITERATIONS

		PRESSURE FAULT			
		MAE (10^{-3})			
		SM	LF		
			$1/\varepsilon$	Log	NT
CF	Clean data	3.855	3.296 (+15%)	3.103 (+20%)	3.608 (+6.4%)
	Noisy data	5.009	4.675 (6.7%)	4.683 (+6.5%)	4.431 (+12%)
NF	Clean data	3.855	3.181 (+18%)	2.909 (+25%)	3.657 (+5.2)
	Noisy data	5.009	4.573 (+8.7%)	4.575 (+8.7%)	4.463 (+11%)

		LEVEL FAULT			
		MAE (10^{-3})			
		SM	LF		
			$1/\varepsilon$	Log	NT
CF	Clean data	3.589	2.608 (+27%)	2.475 (+31%)	3.572 (+0.47)
	Noisy data	6.262	5.048 (+19%)	5.056 (+19%)	5.080 (+20%)
NF	Clean data	3.589	2.527 (+30%)	2.422 (+33%)	3.422 (+4.7)
	Noisy data	6.262	5.181 (+17%)	5.194 (+17%)	5.156 (+18%)

SM = simple mean; LF = local fusion; $1/\varepsilon$ = weights computed with eq.(7); Log = weights computed with eq.(8); NT = weights computed with eq.(7), adding noise on training data; CF = constant fault, NF = noise fault.

training data but are not suitable for the reconstruction of faulty signals.

A significant improvement of the results can also be obtained by repeating the reconstruction task iteratively several times, each time feeding the reconstructed value of the previous iteration.

VI. CONCLUSIONS

The ensemble approach has been shown to be an effective way for increasing the accuracy and robustness of signal reconstruction.

The goodness of the results depends on the methods used for aggregating the multiple outcomes of the ensemble models. This work has shown that local approaches can increase the performance of the ensemble in the signal reconstruction task. The best results have, in fact, been obtained when the parameters of the aggregation procedure dynamically change depending on the values of the signals to be reconstructed, i.e., the location of the test pattern in the signal space relative to the set of training patterns.

Realistic noise and sensor faults have been considered in order to show that the proposed local fusion approach is effective for boosting model prediction accuracy, in practical settings.

Furthermore, the use of a set of signals containing also previous measurements has been demonstrated to be an effective way to reduce the detrimental contribution of noise or faults.

In the case of sensor faults, the proposed procedure of faulty signal reconstruction, based on iterations of the reconstruction step and memorization of the reconstructed signal for use at the next iteration, improves the accuracy of the reconstruction of the value of the signal.

Although the methodology has been demonstrated in the context of multiple models based on PCA and constructed on groups of signals selected by the random feature selection technique, it seems to be broadly applicable to other empirical modeling paradigms. Investigations in this direction could be of interest, considering in particular other fast empirical regression techniques.

Finally, the local performance identification and parameters computation have to be performed online; this could be a costly procedure that might not be suitable when response time is an issue. Future work will be devoted to the set up of a procedure for the offline assignment of the local fusion parameters values.

APPENDIX A ACRONYMS AND NOMENCLATURE

k -nn	k -nearest neighbors
LF	Local Fusion
mae	Mean Absolute Error
me	Mean Error
mse	Mean Squared Error
PCA	Principal Components Analysis
PHT	Primary Heat Transport
PID	Proportional Integral Derivative
PWR	Pressurized Water Reactor
RFSE	Random Features Selection Ensemble
SM	Simple Mean
SPRT	Sequential Probability Ratio Test
f_i	Sensor measurement
\hat{f}_i	Estimate of the i -th sensor measurement provided by a model
f_i^T	True value of the physical quantity measured by the i -th sensor
\widehat{f}_i^T	Final reconstruction of f_i^T
n	Total number of sensors' signals available
\mathbf{X}	Data set
\mathbf{X}_{TRN}	Training data subset
\mathbf{X}_{TST}	Test data subset
N	Total number of signal patterns
N_{TRN}	Number of training patterns
N_{TST}	Number of test patterns
\mathbf{F}_h	Subset of signals
m	Number of signals in each subsets F_h
H	Total number of models
H_i	Number of models reconstructing the i -th signal
\hat{f}_i^h	Reconstruction made by the h -th model of the i -th signal
\bar{f}_i	Expected value of \hat{f}_i^h over the H_i models built on groups which contain the i -th signal
$\varepsilon_i(t)$	Residual error $f_i(t) - \hat{f}_i(t)$

ε_i^h	Part of the reconstruction error $\hat{f}_i^h - f_i(t)$ due to modeling errors
ε_n	Part of the reconstruction error $\hat{f}_i^h - f_i(t)$ due to the presence of a noise on the input
Hrf	Length of each side of the hyper-rectangle expressed as a fraction of the range of values of the corresponding signal
Q_{TRN}	Subset of training patterns similar to the test pattern, also called neighborhood of the test pattern
$N_{Q_{TRN}}$	Cardinality of the neighborhood Q_{TRN}
$mae_{i,Q_{TRN}}^h$	Mean absolute error of the h -th model in the reconstruction of the i -th signal of the patterns in the set Q_{TRN}
$me_{i,Q_{TRN}}^h$	Mean error of the h -th model in the reconstruction of the i -th signal of the patterns in the set Q_{TRN}
b_i^h	Bias correction associated to the h -th model for the reconstruction of the i -th signal
w_i^h	Weight associated to the h -th model for the reconstruction of the i -th signal
s_i^h	Score assigned by the Borda Count method to the h -th model for the reconstruction of the i -th signal.

APPENDIX B

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS (PCA) TO SIGNAL RECONSTRUCTION

In this Appendix, the procedural steps of Principal Component Analysis (PCA) are briefly sketched [9].

Let $\mathbf{X} \equiv (N, m)$ be the data set matrix with row vector $\mathbf{f}_t \equiv (f_1(t) \dots f_i(t) \dots f_m(t))$, $t = 1, \dots, N$, which gathers the N signals patterns measured at different time instants t . The purpose of the PCA is to identify a λ -dimensional ($\lambda < m$) subspace $\mathfrak{R}^\lambda \subset \mathfrak{R}^m$ in which the most of the data set variation is retained and the least information is lost.

Let $\mathbf{P} \equiv (m, m) \in \mathfrak{R}^m$ be a matrix constituted by m orthonormal principal vectors which are the eigenvectors associated to the eigenvalues Λ_i of the covariance matrix \mathbf{V} of \mathbf{X} :

$$\mathbf{V} = (\mathbf{X} - \bar{\mathbf{X}})^T (\mathbf{X} - \bar{\mathbf{X}}) \equiv \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T \quad (\text{B1})$$

where it is assumed, without loss of generality, that $\mathbf{\Lambda} = \text{diag}(\Lambda_1, \dots, \Lambda_i, \dots, \Lambda_m)$ is a diagonal matrix with diagonal elements in decreasing order of magnitude.

\mathbf{P} represents an orthonormal basis for the data set and the transformation

$$\mathbf{U} = \mathbf{X} \cdot \mathbf{P} \equiv (N, m) \quad (\text{B2})$$

projects the data in the new basis. The data set has now two representations: when intended in the original basis, the t -th pattern is the vector \mathbf{f}_t with components $f_{ti} = f_i(t)$; when intended in the orthonormal \mathbf{P} basis, the same t -th pattern is represented by

the t -th rows \mathbf{u}_t of the principal component matrix \mathbf{U} with component u_{ti} along \mathbf{p}_i . To get the inverse relation, we right-multiply by \mathbf{P}^T to obtain the data set \mathbf{X} in the original basis:

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{P}^T \quad (\text{B3})$$

Eqs. (B2) and (B3) represent an optimal linear transformation of the data matrix \mathbf{X} in the sense that the new set of variables, the principal components (PCs), are uncorrelated and ordered so that the first few retain the most of the variation present in all of the original variables [13]. Up to this point the equations are exact and the data values are transformed in both senses without any loss of information.

The PCA approximation consists in mapping the observations in a subspace $\mathfrak{R}^\lambda \subset \mathfrak{R}^m$ identified by $\lambda < m$ vectors. For that, the matrix \mathbf{P} is partitioned as follows:

$$\mathbf{P} = (\mathbf{P}_\lambda \mathbf{P}_{m-\lambda}) \quad (\text{B4})$$

where $\mathbf{P}_\lambda \equiv (m, \lambda)$ is the submatrix constituted by the first λ columns of \mathbf{P} and $\mathbf{P}_{m-\lambda} \equiv (m, m - \lambda)$ is the submatrix constituted by the last $m - \lambda$ columns of \mathbf{P} . The column vectors in \mathbf{P}_λ and $\mathbf{P}_{m-\lambda}$ constitute the bases of the two mutually orthogonal subspaces \mathfrak{R}^λ and $\mathfrak{R}^{m-\lambda}$ in which \mathfrak{R}^m has been divided. In terms of the above submatrices, matrix \mathbf{X} can be rewritten as:

$$\begin{aligned} \mathbf{X} &= \mathbf{X} \cdot (\mathbf{P}_\lambda \mathbf{P}_{m-\lambda}) \begin{pmatrix} \mathbf{P}_\lambda^T \\ \mathbf{P}_{m-\lambda}^T \end{pmatrix} \\ &= \mathbf{X} \cdot (\mathbf{P}_\lambda \mathbf{P}_\lambda^T + \mathbf{P}_{m-\lambda} \mathbf{P}_{m-\lambda}^T) \\ &= \tilde{\mathbf{X}} + \mathbf{X} \mathbf{P}_{m-\lambda} \mathbf{P}_{m-\lambda}^T \end{aligned} \quad (\text{B5})$$

where $\tilde{\mathbf{X}}$ is the principal part of the data explained by the first λ eigenvectors:

$$\tilde{\mathbf{X}} = \mathbf{X} \mathbf{P}_\lambda \mathbf{P}_\lambda^T \equiv (N, m) \quad (\text{B6})$$

The t -th row of $\tilde{\mathbf{X}}$, namely $\tilde{\mathbf{f}}_t$, is the orthonormal projection of \mathbf{f}_t onto \mathfrak{R}^λ .

If all the information about the data set \mathbf{X} essentially lies in a λ -dimensional space \mathfrak{R}^λ (apart from small components in $\mathfrak{R}^{m-\lambda}$ given by $\mathbf{X} \mathbf{P}_{m-\lambda} \cdot \mathbf{P}_{m-\lambda}^T$ as stated in (B5), then each observation vector $\mathbf{f}_t \in \mathfrak{R}^m$ can be approximated by its orthonormal projection $\tilde{\mathbf{f}}_t \in \mathfrak{R}^\lambda$ plus a residual vector \mathbf{r} in $\mathfrak{R}^{m-\lambda}$ which is postulated to be independent of t :

$$\tilde{\mathbf{f}}_t^{appx} = \tilde{\mathbf{f}}_t + \mathbf{r} \cdot \mathbf{P}_{m-\lambda}^T = \mathbf{f}_t \mathbf{P}_\lambda \mathbf{P}_\lambda^T + \mathbf{r} \cdot \mathbf{P}_{m-\lambda}^T \quad (\text{B7})$$

The best residual vector is the one which, on average, minimizes the absolute value E of the square error between the real $\{\mathbf{f}_t\}$ and approximated $\{\tilde{\mathbf{f}}_t^{appx}\}$ data patterns, i.e.:

$$E = \frac{1}{2} \sum_{t=1}^N \left\| \mathbf{f}_t - \tilde{\mathbf{f}}_t^{appx} \right\|^2 \quad (\text{B8})$$

The calculations for the minimization can be found in [9]; the resulting optimal residual vector takes the form:

$$\mathbf{r} = \frac{1}{N} \sum_{t=1}^N \mathbf{f}_t \cdot \mathbf{P}_{m-\lambda} = \left(\frac{1}{N} \sum_{t=1}^N \mathbf{f}_t \right) \cdot \mathbf{P}_{m-\lambda}^T = \bar{\mathbf{f}} \cdot \mathbf{P}_{m-\lambda} \quad (\text{B9})$$

where the vector $\bar{\mathbf{f}}$ is the arithmetic average of the observation vectors, i.e. the average value of the signals. In particular, the i -th component of $\bar{\mathbf{f}}$ is the arithmetic average of the i -th column of \mathbf{X} .

Then, from (B7), the expression for the PCA approximation of the data pattern \mathbf{f}_t is given by:

$$\tilde{\mathbf{f}}_t^{appx} = \tilde{\mathbf{f}}_t + \bar{\mathbf{f}} \cdot \mathbf{P}_{m-\lambda} \mathbf{P}_{m-\lambda}^T \quad (\text{B10})$$

or in matrix form,

$$\tilde{\mathbf{X}}^{appx} = \tilde{\mathbf{X}} + \bar{\mathbf{X}} \cdot \mathbf{P}_{m-\lambda} \cdot \mathbf{P}_{m-\lambda}^T \quad (\text{B11})$$

As demonstrated and explained in details in [13], [14] and [30], the minimum error corresponding to the optimal residual vector \mathbf{r} can be written as:

$$E_{min} = \frac{1}{2} Tr [\mathbf{P}_{m-\lambda}^T \cdot \mathbf{V} \cdot \mathbf{P}_{m-\lambda}] \quad (\text{B12})$$

and the fact of choosing \mathbf{P} as the eigenvectors of the covariance matrix \mathbf{V} corresponds to minimizing the minimal error E_{min} of (B12).

Finally, for the specific problem of signal validation and reconstruction the orthonormal basis \mathbf{P} has been obtained by (B1) from the covariance matrix \mathbf{V} of the pairwise signal correlations between the m signals. The number λ of eigenvectors to be retained is determined as follows: components corresponding to the largest variance of the data are retained until the cumulative variance associated to the retained components reaches a fraction F_{Var} of the total variance.

Furthermore, in order to simplify the calculations, the time trends of the signals have been previously normalized so that their mean is zero and their standard deviation equals 1. This allows skipping the computation of the residuals, since $\bar{\mathbf{f}} = \mathbf{0}$ and, according to (B10)

$$\tilde{\mathbf{f}}_t^{appx} = \left(\hat{f}_1(t) \dots \hat{f}_i(t) \dots \hat{f}_m(t) \right) = \tilde{\mathbf{f}}_t \quad (\text{B13})$$

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their constructive comments which have allowed improving the paper.

REFERENCES

- [1] M. Hoffmann, On-Line Monitoring for Calibration Reduction HWR-784, *OECD Halden Reactor Project*, Oct. 2005.
- [2] M. Hoffmann, "Signal grouping algorithm for an improved on-line calibration monitoring system," in *Proc. FLINS*, Genova, Italy, Aug. 2006.
- [3] K. E. Holbert and B. R. Upadhyaya, "An integrated signal validation for nuclear power plants," *Nucl. Technol.*, vol. 92, no. 3, pp. 411–427, 1990.
- [4] M. Hoffmann, "Signal grouping algorithm for an improved on-line calibration monitoring system," in *Proc. FLINS*, Genova, Italy, Aug. 2006.
- [5] D. Roverso, M. Hoffmann, E. Zio, P. Baraldi, and G. Gola, "Solutions for plant-wide on-line calibration monitoring," in *Proc. ESREL*, Stavanger, Norway, 2007, vol. 1, pp. 827–832.
- [6] Y. Tharrault, G. Mourou, J. Ragot, and D. Maquin, "Fault detection and isolation with robust principal component analysis," *Int. J. Appl. Math. Comput. Sci.*, vol. 18, no. 4, pp. 429–442, 2008.
- [7] R. M. Singer, K. C. Gross, J. P. Herzog, R. W. King, and S. W. Wegerich, "Model-based nuclear power plant monitoring and fault detection: Theoretical foundations," in *Proc. 9th Intl. Conf. Intelligent Systems Applications to Power Systems*, Seoul, Korea, 1996.
- [8] E. Zio, P. Baraldi, G. Gola, D. Roverso, and M. Hoffmann, "Genetic algorithms for grouping of signals for system monitoring and diagnostics," in *Proc. ESREL*, Stavanger, Norway, 2007, vol. 1, pp. 833–840.
- [9] P. Baraldi, E. Zio, G. Gola, D. Roverso, and M. Hoffmann, "Genetic algorithms for signal grouping in sensor validation: A comparison of the filter and wrapper approaches," *J. Risk Reliab.*, vol. 222, pp. 189–206, 2008, Proc. IMechE.
- [10] P. F. Fantoni, M. I. Hoffmann, R. Shankar, and E. L. Davis, "On-line monitoring of instrument channel performance in nuclear power plant using PEANO," *Prog. Nucl. Energy*, vol. 43, no. 1–4, pp. 83–89, 2003.
- [11] P. F. Fantoni and A. Mazzola, "Multiple-failure signal validation in nuclear power plants using artificial neural networks," *Nucl. Technol.*, vol. 113, no. 3, pp. 368–374, 1996.
- [12] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: Improving accuracy of classifiers ensembles by using random feature subsets," *Pattern Recognit.*, vol. 36, pp. 1291–1302, 2003.
- [13] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer, 2002.
- [14] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: Wiley, 1996.
- [15] B. Scholkopf, A. Smola, and K. R. Muller, "Kernel principal component analysis," *Adv. Kernel Methods-Support Vector Learn.*, 1999.
- [16] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. 26, no. 1, Feb. 1981.
- [17] P. Baraldi, E. Zio, G. Gola, D. Roverso, and M. Hoffmann, "A procedure for the reconstruction of faulty signals by means of an ensemble of regression models based on principal components analysis," in *Proc. NPIC-HMIT Topical Meeting*, Apr. 5–9, 2009.
- [18] R. Polikar, "Ensemble based system in decision making," *Circuits Syst. Mag.*, pp. 21–45, 3rd quarter 2006.
- [19] M. A. and A. F. Atiya, "Novel methods for the feature subset ensembles approach," *Int. J. Artificial Intell. Machine Learn.*, vol. 6, no. 4, 2006.
- [20] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
- [21] H. Y. Yang, S. H. Lee, and M. G. Na, "Monitoring and Uncertainty Analysis of Feedwater Flow Rate Using Data-Based Modeling Method."
- [22] R. Chevalier, D. Provost, and R. Seraoui, "Assessment of statistical and classification models for monitoring EDF's assets," in *Proc. NPIC-HMIT Topical Meeting*, April 5–9, 2009.
- [23] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, pp. 1–58, 1992.
- [24] P. P. Bonissone, F. Xue, and R. Subbu, "Fast meta-models for local fusion of multiple predictive models," *Appl. Soft Comput. J.*, 2008.
- [25] R. Polikar, "Bootstrap-inspired techniques in computational intelligence," *IEEE Signal Process. Mag.*, vol. 24, no. 4, Jul. 2007.
- [26] [Online]. Available: <http://www.iapws.org>
- [27] H. M. Hashemian, *Sensor Performance and Reliability*. Research Triangle Park, NC: ISA, 2004.
- [28] H. M. Hashemian, *Maintenance of Process Instrumentation in Nuclear Power Plants*. New York: Springer, 2006.
- [29] F. Johnson, *Nuclear Reactor Controls and Instrumentation*. Palm Springs, CA: Wexford College Press, 2008.
- [30] M. Marseguerra, *Lecture Notes on Principal Components Analysis (PCA)* Polytechnic of Milan, Milan, Italy.