



**HAL**  
open science

## **Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems**

Piero Baraldi, Enrico Zio, Giulio Gola, Davide Roverso, M. Hoffmann

### ► **To cite this version:**

Piero Baraldi, Enrico Zio, Giulio Gola, Davide Roverso, M. Hoffmann. Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems. *Annals of Nuclear Energy*, 2011, 38 (2-3), pp.212-220. <10.1016/j.anucene.2010.11.007>. <hal-00609545>

**HAL Id: hal-00609545**

**<https://centralesupelec.hal.science/hal-00609545v1>**

Submitted on 27 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.



HAL Authorization

# Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems

P. Baraldi<sup>1</sup>, E. Zio<sup>1,\*</sup>, G. Gola<sup>2</sup>, D. Roverso<sup>2</sup>, M. Hoffmann<sup>2</sup>

<sup>1</sup>*Department of Energy, Polytechnic of Milan, Via Ponzio 34/3, 20133, Milano, Italy*

*\*enrico.zio@polimi.it*

<sup>2</sup>*Institutt for energiteknikk, OECD Halden Reactor Project, 1751, Halden, Norway*

## Abstract

Detecting anomalies in sensors and reconstructing the correct values of the measured signals is of paramount importance for the safe and reliable operation of nuclear power plants. Auto-associative regression models can be used for the signal reconstruction task but in real applications the number of sensors signals may be too large to be handled effectively by one single model. In these cases, one may resort to an ensemble of reconstruction models, each one handling a small group of sensor signals; the outcomes of the individual models are then combined to produce the final reconstruction. In this work, three methods for aggregating the outcomes of a feature-randomized ensemble of Principal Components Analysis (PCA)-based regression models are analyzed and applied to two case studies concerning the reconstruction of 215 signals monitored at a Finnish nuclear Pressurized Water Reactor (PWR) and 920 simulated signals of the Swedish Forsmark-3 Boiling Water Reactor (BWR). Based on the insights gained, two novel aggregation procedures are developed for optimal signal reconstruction.

## 1. Introduction

Accurate monitoring of the health state of systems, structures and components can contribute significantly to the safe and efficient functioning of nuclear power plants for it allows the timely detection of malfunctions and anomalies during operation.

The monitoring relies on the signals collected by a large number of sensors placed at various locations in the plant. The measured signals are transmitted to the control and protection systems and to the operators, to evaluate the plant health state and eventually take corrective or emergency actions for safely steering critical situations and preventing accidents. For efficient monitoring, early detection of sensor failures and eventual correction of the delivered measurements is fundamental to avoid that misleading information be used for plant operation, possibly leading to unsafe and/or inefficient states (Hoffmann, 2005; Hoffmann, 2006).

This work investigates the problem of reconstructing signals in realistic applications in which the number of measured signals is very large and cannot be handled effectively by a single reconstruction model (Hoffmann, 2006; Roverso et al., 2007; Zio et al., 2007; Baraldi et al. 2008a). The problem is tackled by resorting to an ensemble-based signal reconstruction procedure. The ensemble approach is founded on the subdivision of the set of sensor signals into small overlapping groups; a reconstruction model is developed for each group of signals and the outcomes of the individual models are eventually

aggregated to generate the reconstructed signal (Perrone and Cooper, 1993; Krogh and Vedelsby, 1995; Sharkey, 1996; Gola et al., 2007; Baraldi et al., 2008b; Gola et al., 2008) (Figure 1).

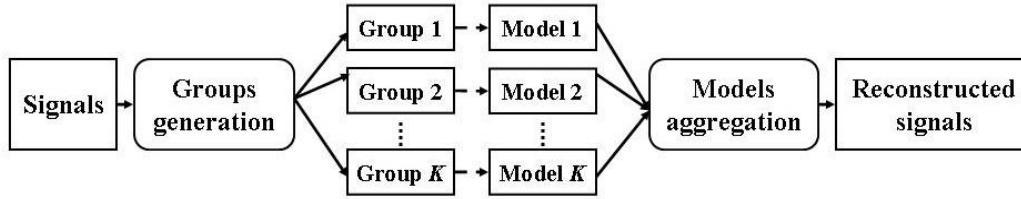


Figure 1. The multi-group ensemble approach to signal reconstruction

Two issues are central to the ensemble approach to signal reconstruction: (1) how to define the signal groups and (2) how to aggregate the outcomes of the models built on such groups.

In this work, the signals groups are randomly generated by the Random Feature Selection Ensemble (RFSE) technique (Aly and Atiya, 2006; Bryll et al., 2003; Baraldi et al., 2009; Polikar, 2006) which allows obtaining highly diverse signal groups and, correspondingly, diverse signal outcomes from the individual models. The regression models based on the signal groups are developed by Principal Component Analysis (PCA) (Jolliffe, 2002; Diamantaras and Kung, 1996; Scholkopf et al., 1999; Moore, 1981).

The issue of how to compute the ensemble-aggregated signal reconstruction output is the focus of this paper. The random composition of the signal groups is such that some of the individual models might provide largely incorrect signal reconstructions which negatively affect the ensemble aggregate. Thus, discarding the outcomes of some models can enhance the accuracy and robustness of the randomized ensemble approach. In this view, three different aggregation methods based on the distribution of the models' outcomes are analyzed: Simple Mean (SM), Median (MD) and Trimmed Mean (TM) (Polikar, 2006). The first amounts to using all the available outcomes of the models in a simple average; the second considers only the central outcome in the distribution of the outcomes of the individual models; the last one discards the outcomes falling in the tails of the distribution. Furthermore, two novel procedures for aggregating the models outcomes based on the combination of the Median and Trimmed Mean approaches are proposed to fully exploit the benefits of the two methods.

The paper is organized as follows. In Section 2, the randomized approach for generating diverse groups of signals is briefly summarized and the methods of model outcomes aggregation are described. In Section 3, the performance of the aggregation methods is analyzed with reference to two case studies: one concerns the reconstruction of a data set of 215 signals measured at a Finnish nuclear Pressurized Water Reactor (PWR) located in Loviisa; the other addresses the reconstruction of a data set of 920 simulated signals obtained with the HAMBO code (Karlsson et al., 2001) which simulates the Forsmark-3 Boiling Water Reactor (BWR) located in Sweden. In Section 4, two novel aggregation procedures are introduced and verified on both case studies. Conclusions on the advantages and limitations of the proposed methods are drawn in the last Section.

## 2. Model ensemble construction and aggregation

Given a set of  $n \gg 1$  signals  $f_i$ ,  $i = 1, 2, \dots, n$ , measured in the plant, the problem is to define a set of  $K$  groups of  $m \ll n$  signals, with characteristics favourable for their ensemble aggregation. In particular,

diversity in the groups composition is important, balanced with a suitable redundancy of the signals in the groups (i.e. the same signal must be included in more than one group) so that their ensemble reconstruction can be based on several (diverse) models outcomes. In other words, groups must partially overlap in order to have some signals in common, while being sufficiently diverse among one another.

An effective method to pursue these characteristics is offered by the Random Feature Selection Ensemble (RFSE) technique (Aly and Atiya, 2006; Bryll et al., 2003; Baraldi et al., 2009; Polikar, 2006), which consists in randomly sampling, with replacement, from the  $n$  signals,  $K$  groups, each one constituted by  $m$  signals. The procedure is simple and easily applicable in large scale applications. Furthermore, a reasonable choice of the ensemble parameters  $m$  and  $K$  allows the desired diversity among the groups and signal redundancy across the groups (Baraldi et al., 2009).

On the basis of the  $K$  signal groups an equal number of PCA regression models are constructed, based on an available data set  $\mathbf{X}$  of  $N$  signal patterns, partitioned into a training set  $\mathbf{X}_{TRN}$  (made of  $N_{TRN}$  patterns) and a test set  $\mathbf{X}_{TST}$  (made of  $N_{TST} = N - N_{TRN}$  patterns); the former is used to train the individual PCA models, whereas the latter is used to verify the signal reconstruction performance of the ensemble.

For the generic  $t$ -th test pattern, the generic  $k$ -th model based on the  $m$  signals of group  $k$  (randomly selected by RFSE) gives as outcome the prediction  $\hat{f}_i^k(t)$ ,  $t = 1, 2, \dots, N_{TST}$ ,  $i = 1, 2, \dots, m$ . The generic signal  $i$  is included in a suitable number of groups  $K_i$  and the outcomes of the corresponding models are aggregated into the final ensemble outcome.

Three different methods are here considered to aggregate the individual model outcomes: Simple Mean (SM), Median (MD) and Trimmed Mean (TM).

Simple Mean (SM) amounts to combining all the available model outcomes in a simple arithmetic average (Baraldi et al., 2009; Polikar, 2006; Kuncheva, 2004; Kittler et al., 1998):

$$\hat{f}_i^{E,SM}(t) = \frac{1}{K_i} \sum_{k=1}^{K_i} \hat{f}_i^k(t) \quad i = 1, 2, \dots, n \quad (1)$$

where  $\hat{f}_i^{E,SM}(t)$  is the ensemble aggregate of the reconstructions  $\hat{f}_i^k(t)$  of pattern  $t$  of signal  $i$  provided by the individual models  $k = 1, 2, \dots, K_i$  containing signal  $i$ .

Nevertheless, due to the random nature of the groups composition upon which the signal reconstruction models are built, some individual models might provide largely incorrect signal reconstructions, which compromise the simple average-aggregated ensemble outcome<sup>1</sup>.

---

<sup>1</sup> In principle, if the outcomes of a very large number of (randomized) models were available for each signal (i.e., the signals are highly redundant), simple averaging would mitigate the effects of the group randomness and provide an accurate ensemble reconstruction. In practice, imposing a high redundancy ( $K_i$ ) for each signal  $i = 1, 2, \dots, n$  while, at the same time, requiring a small group size ( $m$ ) for model effectiveness forces to increase the number of groups ( $K$ ) in the ensemble (Baraldi et al., 2009), with consequent increase of the computational cost associated to the development of a corresponding number of models.

Alternative aggregation methods can be devised for preventing the inclusion of highly incorrect outcomes in the ensemble aggregate. The methods are based on the randomness of the models outcomes, which, if unbiased, are expected to distribute around the correct (unknown) signal value. In this view, the outcomes lying around the centre of the distribution are conjectured to be close to the correct signal value, whereas those lying on the tails of the distribution are considered fairly incorrect.

The Median (MD) approach amounts to considering for the generic pattern  $t$  the single outcome  $\hat{f}_i^{k_c}(t)$  lying in the centre of the distribution of the outcomes for that pattern, i.e.:

$$\hat{f}_i^{E,MD}(t) = \hat{f}_i^{k_c}(t) \quad i = 1, 2, \dots, n \quad (2)$$

where  $k_c$  denotes the index of the model whose outcome is central with respect to the reconstructed values of the  $K_i$  models including signal  $i$ .

The Trimmed Mean (TM) approach amounts, instead, to discarding a fraction  $\mathcal{G}_{TM}$  of the  $K_i$  outcomes of signal  $i$   $\{\hat{f}_i^k(t), k = 1, 2, \dots, K_i\}$  at each tail of the distribution. The ensemble reconstruction of the generic pattern  $t$ ,  $\hat{f}_i^{E,TM}(t)$ , is then computed by simply averaging the  $K_i^{TM} = (1 - 2\mathcal{G}_{TM})K_i$  remaining outcomes:

$$\hat{f}_i^{E,TM}(t) = \frac{1}{K_i^{TM}} \sum_{k=1}^{K_i^{TM}} \hat{f}_i^k(t) \quad i = 1, 2, \dots, n \quad (3)$$

In this sense, the TM represents a compromise between the SM and MD methods, for it allows discarding the tails of the distribution while still considering multiple (central) outcomes to reconstruct the signal<sup>2</sup>.

Finally, to evaluate the performance of the ensemble aggregates, first the absolute signal reconstruction error is computed<sup>3</sup>:

$$\varepsilon_i^{E, (SM, MD, TM)} = \frac{1}{N_{TST}} \sum_{t=1}^{N_{TST}} \left| f_i(t) - \hat{f}_i^{E, (SM, MD, TM)}(t) \right| \quad (4)$$

Then, the ensemble performance index is retained as the average of the absolute signal reconstruction errors (Eq. 4):

$$\eta^{E, (SM, MD, TM)} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^{E, (SM, MD, TM)} \quad (5)$$

### 3. Application of the aggregation methods and analysis of their performance

---

<sup>2</sup> Notice that  $\mathcal{G}_{TM} \in \left[ 0, \frac{1}{2} \left( 1 - \frac{1}{K_i} \right) \right]$ . The lower limit corresponds to considering all the available outcomes (SM), whereas the upper limit represents the case of using the single central outcome (MD).

<sup>3</sup> In the application that follows, each signal of the validation set has been previously normalized in the range [0.2, 1], for convenience.

Two case studies are considered to analyze the performance of the aggregation methods. The first concerns the reconstruction of 215 signals measured at a nuclear Pressurized Water Reactor (PWR) located in Loviisa, Finland; the second addresses the reconstruction of 920 simulated signals of the Swedish Forsmark-3 Boiling Water Reactor (BWR). The main characteristics and parameters adopted in the two cases studies are reported in Table 1. The PCA models have been constructed with the code <http://lib.stat.cmu.edu/multi/pca>, adapted to perform the signal reconstruction task of interest here.

Case study		Loviisa Pressurized Water Reactor (PWR)	Forsmark-3 Boiling Water Reactor (BWR)	
Signals	Number of signals, $n$	215	920	
	Type of signals	Measured <i>in situ</i>	Simulated with the HAMBO code	
Data set	Number of patterns available, $N$	12713	5463	
	Number of training patterns, $N_{TRN}$ , in $X_{TRN}$	8000	3600	
	Number of test patterns, $N_{TST}$ , in $X_{TST}$	4713	1863	
	Type of measurements <sup>4</sup>	Normal operation and transients related to two outages	Start up, normal operation and shut down	
	Sampling rate	1 hour	1 hour	
Wrapper-optimized ensemble parameters	Signal redundancy, $K_i$	7	7	
	Group size, $m$	38	70	
	Ensemble size <sup>5</sup> , $K$	40	92	
Aggregation parameters	TM fraction, $\mathcal{G}_{TM}$	0.2	0.2	
	Number of groups (i.e. models) used to reconstruct each signal	SM ( $K_i$ )	7	7
		MD	1	1
		TM ( $K_i^{TM}$ )	5	5

Table 1. Main characteristics and parameters of the two case studies

An effective ensemble scheme of signal reconstruction must be capable of reconstructing the signals from faulty sensors, e.g. affected by random noises or offsets. Within the proposed ensemble approach, a faulty sensor sends a faulty signal in input to the PCA models which include that signal; in this situation, the ensemble must be capable of providing a good estimate of the true value of the signal by correlating the information coming from the non-faulty signals in the models of the ensemble.

To verify this, disturbs are randomly introduced in the patterns of the test set. More precisely, the signals of a test pattern are randomly affected either by noise (with probability  $p^{RN} = 0.12$ ) or by their value being set equal to the offset value of the corresponding sensor (with probability  $p^{OF} = 0.08$ ); with probability 0.8 the signals remain unmodified.

<sup>4</sup> In the Forsmark-3 case study both  $X_{TRN}$  and  $X_{TST}$  include signal measurements related to start up, normal operation and shut down conditions, whereas in the Loviisa case study transients related to the first outage are included in  $X_{TRN}$ , while those of the second outage are in  $X_{TST}$ .

<sup>5</sup> Once  $m$  and  $K_i$  are set, the number of groups to generate with the RFSE procedure is obtained from the relation  $mK = nK_i$ . For a more detailed explanation of the procedure, the interested reader may refer to Baraldi et al. (2009).

Table 2 reports the results obtained by the three aggregation methods for the two case studies, in terms of ensemble reconstruction performance indexes (Eq. 5) on test patterns made of undisturbed ( $\eta^{E,U}$ ) and disturbed ( $\eta^{E,D}$ ) signals. In general, the results show the benefits of excluding some outcomes from the aggregation: overall, the MD ensemble provides the most accurate and robust signal reconstruction in both applications, with a slight improvement with respect to the TM technique.

$\times 10^{-2}$	Loviisa Pressurized Water Reactor (PWR)			Forsmark-3 Boiling Water Reactor (BWR)		
	SM	MD	TM	SM	MD	TM
$\eta^{E,U}$	0.977	0.956	0.957	2.091	1.878	1.894
$\eta^{E,D}$	6.221	5.169	5.261	8.172	7.607	7.638

Table 2. Ensemble reconstruction performance indexes obtained with SM, MD and TM aggregations on undisturbed and disturbed signals for the Loviisa and Forsmark-3 case studies

To further delve into the aggregation approaches, further analyses have been carried out on the Loviisa and Forsmark-3 data.

First of all, to show the effects of the three aggregation methods on the reconstruction, signal 163 of the Loviisa set is considered within the ensemble setting of Table 1 (i.e.,  $K_i = 7, \forall i$ ). Figures 2 and 3 illustrate the reconstruction errors obtained by the  $K_{163} = 7$  groups including signal 163 and those achieved by the SM, MD and TM aggregation methods when the signal is undisturbed (Figure 2) and disturbed (Figure 3). Table 3 reports the numerical values of the signal reconstruction errors (Eq. 4) on the undisturbed ( $\varepsilon_{163}^{E,U}$ ) and disturbed ( $\varepsilon_{163}^{E,D}$ ) signal, respectively, obtained by the SM, MD and TM techniques.

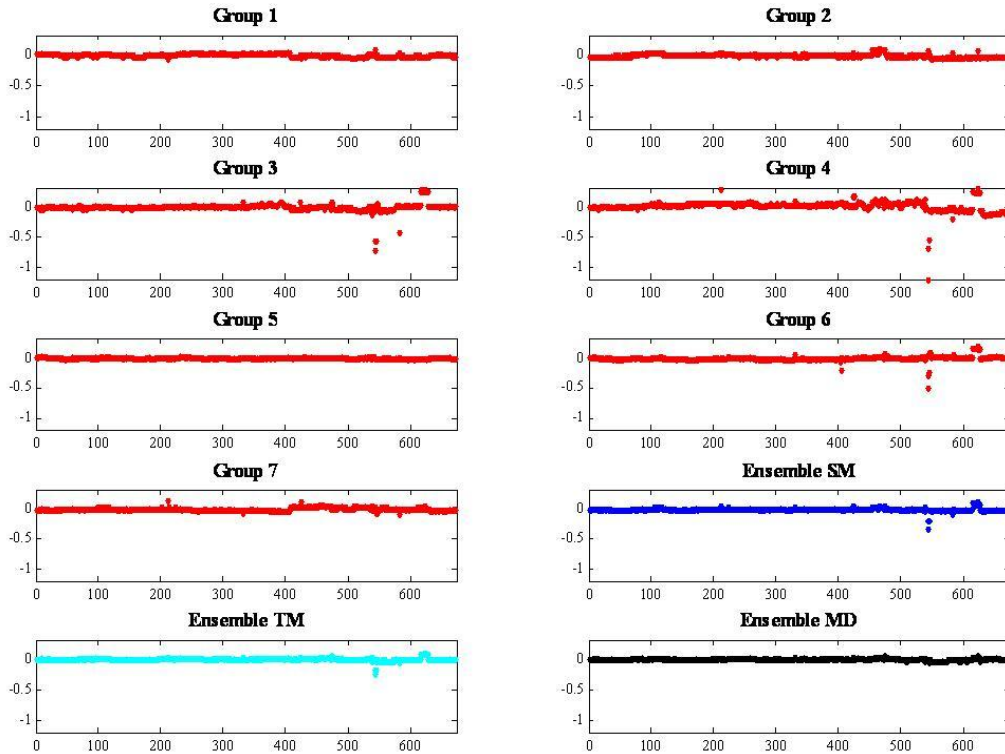


Figure 2. Groups and ensembles reconstruction errors for signal 163 (undisturbed)

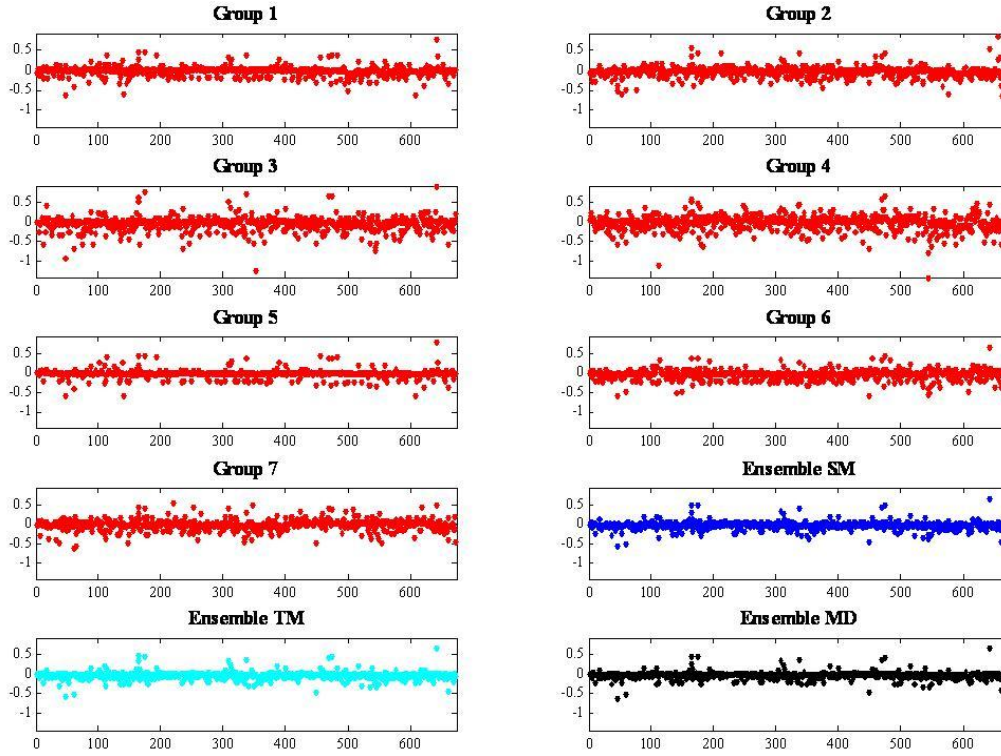


Figure 3. Groups and ensembles reconstruction errors for signal 163 (disturbed)

As shown in Figure 2, the effects of the largely incorrect outcomes on some test patterns provided by some groups (e.g. patterns 543, 544, 583 and 617-627 for group 3, patterns 543-545, 586 and 618-627 for group 4 and patterns 543, 545 and 617-626 for group 6) are still evident (although damped after averaging) in the SM and TM ensemble reconstructions, while they are discarded by the MD method, which results the most accurate (as reported in Table 3).

In case of disturbed signals, resorting to the ensemble approach generally allows reducing the reconstruction errors of the individual groups, eventually resulting in a less *noisy* ensemble-aggregate signal reconstruction. Nevertheless, discarding the outlying predictions (i.e., in the MD and TM ensembles) enhances such effect and provides a more robust signal reconstruction (Table 3).

$\times 10^{-2}$	SM	MD	TM
$\varepsilon_{163}^{E,U}$	1.298	1.172	1.240
$\varepsilon_{163}^{E,D}$	7.334	4.954	5.554

Table 3. Loviisa case study: ensemble performances on signal 163 (undisturbed and disturbed) obtained with SM, MD and TM aggregations

The three aggregation methods have also been evaluated with an augmented signal redundancy  $K_i = 15, \forall i$ , i.e. with more models' outcomes used in the ensemble to reconstruct the signals. With the increased signal redundancy and the optimal group size  $m = 38$ , the number of groups to generate with the RFSE procedure is  $K = 85$ <sup>6</sup>. The fraction  $\mathcal{G}_{TM}$  used in the TM aggregation method has been kept equal to 0.2, which

<sup>6</sup> See Note 5.

corresponds in this case to adopting 15, 1 and 9 models for the reconstruction of each signal by the SM, MD and TM methods, respectively.

Table 4 reports the relative percentage improvement achieved using the higher signal redundancy (from  $K_i = 7$  to  $K_i = 15$ ). The improvement index is simply defined as  $I^{E, (U, D)} = 100 \left( 1 - \eta_{K_i=15}^{E, (U, D)} / \eta_{K_i=7}^{E, (U, D)} \right)$  for the case of undisturbed and disturbed signals, respectively.

%	SM	MD	TM
$I^{E,U}$	6.24	6.65	6.99
$I^{E,D}$	6.38	11.18	12.16

Table 4. Loviisa case study: relative percentage improvements when increasing the signal redundancy obtained with SM, MD and TM aggregation on undisturbed and disturbed signals

In general, increasing the number of outcomes upon which the ensemble reconstruction of the signal is based improves the performance in all cases. This effect is more evident when using the MD and TM methods on the disturbed signal set.

Furthermore, the reconstruction capabilities of the MD and TM approaches on each signal have been analyzed, with reference to the data of the Forsmark-3 case study, in terms of the distance between the  $K_i^{TM}$  PCA model outcomes used to produce the TM aggregate and the single PCA model outcome  $\hat{f}_i^{k_c}(t)$  used as ensemble aggregate in the MD approach. In particular, the correlation between the distribution of the outcomes and the reconstruction performances of the two aggregation methods is analyzed.

To compute the average distance between the outcomes used to reconstruct the generic signal  $i$  by ensemble aggregation, first the absolute point-wise outcome distance  $\delta_i^t$  for the generic training or test pattern  $t = 1, 2, \dots, N_{TRN(TST)}$  is computed:

$$\delta_i^t = \frac{1}{K_i^{TM}} \sum_{\substack{k=1 \\ k \neq k_c}}^{K_i^{TM}} \left| \hat{f}_i^k(t) - \hat{f}_i^{k_c}(t) \right| \quad (6)$$

Small values of  $\delta_i^t$  are obtained when the  $K_i^{TM}$  PCA models considered in the ensemble for the aggregation provide very similar outcomes, whereas if  $\delta_i^t$  is large the  $K_i^{TM}$  PCA model outcomes for pattern  $t$  are more spread. Then, the average model outcome distance associated to the PCA models reconstructions of signal  $i$  is simply:

$$d_i^{TRN(TST)} = \frac{1}{N_{TRN(TST)}} \sum_{t=1}^{N_{TRN(TST)}} \delta_i^t \quad (7)$$

Figure 4 plots the average model outcome distances  $d_i$ ,  $i = 1, 2, \dots, n$ , vs. the difference in the ensemble average reconstruction errors  $\varepsilon_i^{E, TM} - \varepsilon_i^{E, MD}$ ,  $i = 1, 2, \dots, n$ , computed using the training (left) and test (right) sets, respectively. In both cases, the ensemble models are built using the undisturbed training set. In the left graph of Figure 4, the models

have been fed with the same undisturbed training patterns and with training patterns disturbed according to the procedure previously described, whereas in the right graph of Figure 4 the ensemble of models have been fed with undisturbed and disturbed test patterns.

In general, if  $d_i$  is small, i.e.  $d_i \leq 0.02$  on both training (left graph) and test (right graph) patterns, the error difference  $|\varepsilon_i^{E, TM} - \varepsilon_i^{E, MD}| \leq 0.005$  and therefore the difference in adopting MD or TM is negligible. This concerns the reconstruction of all the undisturbed training signals (black dots in the left graph) and most of the undisturbed test signals (black dots in the right graph). Nevertheless, the increase of  $d_i$  (i.e.  $d_i > 0.02$ , a situation mostly occurring when signals are disturbed – light dots in left and right graphs) manifests the advantage of using either MD or TM to aggregate the models outcomes since the difference  $|\varepsilon_i^{E, TM} - \varepsilon_i^{E, MD}|$  can reach values much larger than 0.005.

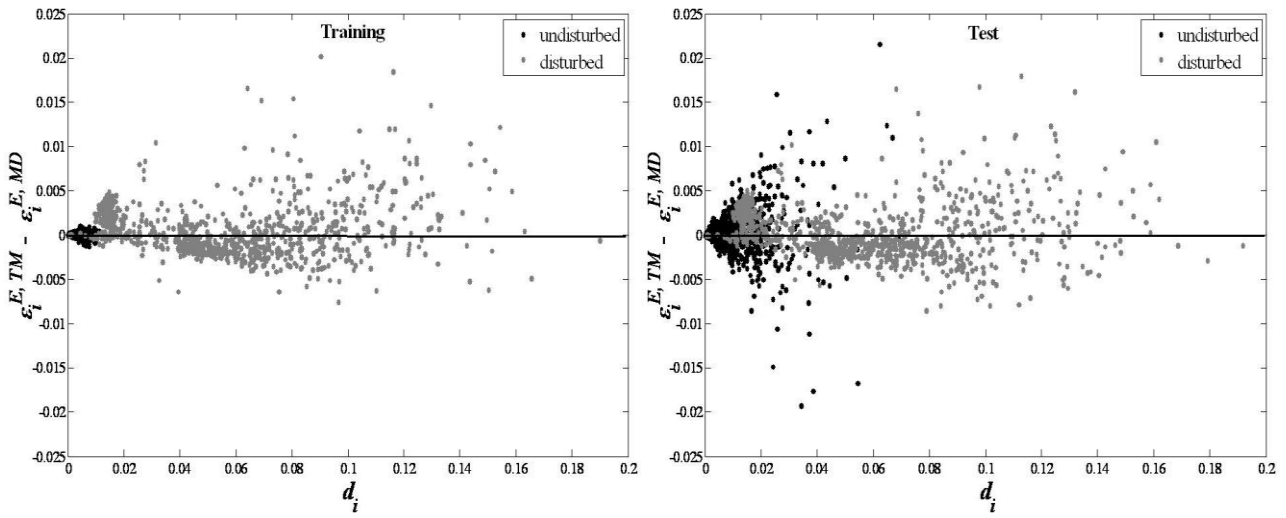


Figure 4. Forsmark-3 case study: average PCA model outcomes distances versus difference between ensemble reconstruction errors by TM and MD, computed using the training (left) and test (right) sets, respectively, on undisturbed (dark dots) and disturbed (light dots) signals

For the Forsmark-3 and Loviisa applications, respectively, Tables 5a and 5b summarize numerically the comparison between MD and TM in terms of the number of signals for which one method outperforms the other ( $\varepsilon_i^{E, TM} - \varepsilon_i^{E, MD} \geq 0$  for MD outperforming TM or viceversa) and the corresponding average error reduction in using the best performing one. For the Forsmark-3 case study (Table 5a), approximately half of the signals are better reconstructed by MD (dots above the black line in the graphs) and half by TM (dots below the black line in the graphs); nevertheless, on disturbed signals, the average error reduction achieved by MD is larger, i.e. the improvements of using MD are more relevant than those obtained with TM, even though on a slightly smaller number of signals. This can also be visualized in the graphs of Figure 4 from the light dots representative of the values of the difference in ensemble reconstruction errors related to disturbed signals. On the other hand, in the Loviisa case study (Table 5b), results show that on disturbed signals MD performs better on a larger number of signals (more than two thirds) with an average error reduction considerably higher than TM.

Aggregation method	Number of signals for which the aggregation method is better / Average error reduction ( $\times 10^{-3}$ )			
	Training		Test	
	Undisturbed	Disturbed	Undisturbed	Disturbed
MD	356 / 0.14	451 / 2.92	474 / 1.58	420 / 2.97
TM	564 / 0.13	469 / 1.60	446 / 1.34	500 / 1.92

Table 5a. Forsmark-3 case study: comparison of MD and TM aggregation methods on undisturbed and disturbed training and test signals

Aggregation method	Number of signals for which the aggregation method is better / Average error reduction ( $\times 10^{-3}$ )			
	Training		Test	
	Undisturbed	Disturbed	Undisturbed	Disturbed
MD	78 / 0.79	155 / 5.68	113 / 4.55	163 / 6.12
TM	137 / 0.73	60 / 2.88	102 / 4.89	52 / 2.91

Table 5b. Loviisa case study: comparison of MD and TM aggregation methods on undisturbed and disturbed training and test signals

For a further understanding of the differences between the two methods, Figure 5 shows the results of a detailed analysis focused on six disturbed test signals, of which three are better reconstructed by MD (namely number 257, 501 and 132, left side of the Figure) and three by TM (namely number 508, 98 and 373, right side of the Figure).

Each graph reports the values (in descending order) of the reconstruction errors  $\varepsilon_i^k$ ,  $k=1,2,\dots,7$  obtained on signal  $i$ , as:

$$\varepsilon_i^k = \frac{1}{N_{TST}} \sum_{t=1}^{N_{TST}} |f_i(t) - \hat{f}_i^k(t)| \quad (8)$$

Furthermore, the graphs show the frequencies  $\phi_{MD}^k$  and  $\phi_{TM}^k$  with which the outcomes of each PCA model  $k=1,2,\dots,7$  are used to reconstruct the signal with MD and TM, respectively:

$$\begin{aligned} \phi_{MD}^k &= N_{TST}^k / N_{TST} \\ \phi_{TM}^k &= N_{TST}^k / K_i^{TM} N_{TST} \end{aligned} \quad (9)$$

where in  $\phi_{MD}^k$ ,  $N_{TST}^k$  is the number of test patterns whose ensemble reconstruction is given by model  $k$  (thus corresponding to the central outcome), whereas in  $\phi_{TM}^k$  it represents the number of test patterns which include model  $k$  in the  $K_i^{TM}$  models used for their reconstruction.

For the three signals on the left of the Figure (better reconstructed by MD), most of the models make low reconstruction errors and the central outcome is more likely to belong to one of these models, as proved by the high value of  $\phi_{MD}^k$  in

correspondence of the accurate models; in such situations, using multiple outcomes might lead to include also those of the less accurate models, eventually compromising the pattern reconstruction.

For the right-hand side signals (better reconstructed by TM), instead, fewer models make low reconstruction errors: in these cases, choosing one single outcome (as in the MD approach) entails the risk of using the one corresponding to a less accurate model, as shown by the peaks of  $\varphi_{MD}^k$  in correspondence of those models; instead, including multiple outcomes increases the chances of considering also those provided by the (few) accurate models, as verified by the more evenly distributed values of  $\varphi_{TM}^k$ .

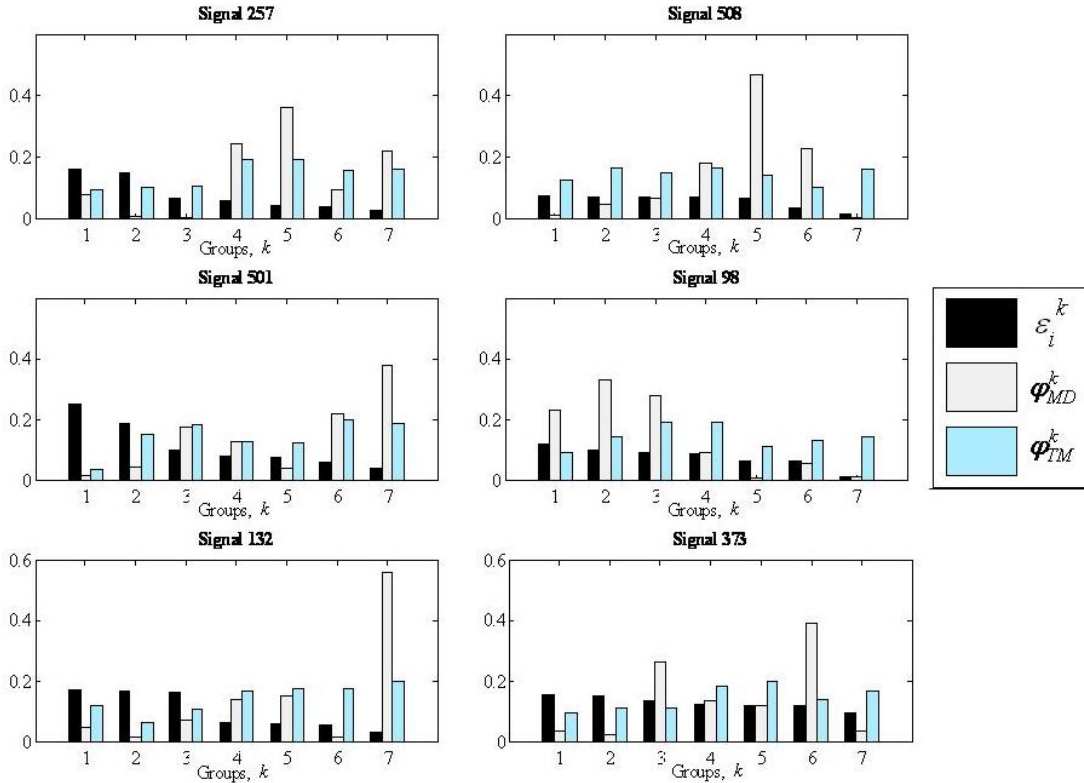


Figure 5. Forsmark-3 case study: model reconstruction errors ( $\varepsilon_i^k$ ) and model outcome inclusion frequency in MD ( $\varphi_{MD}^k$ ) and TM ( $\varphi_{TM}^k$ ) ensembles, for three signals better reconstructed using MD (left) and three using TM (right)

#### 4. Two novel procedures of aggregation

The results of the above analyses show that neither MD, nor TM can clearly outperform the other; on the other hand, procedures for combining MD and TM can be sought, with the aim of exploiting the advantages of both methods. In particular, based on the signal reconstruction errors obtained on the training set, the idea is to define a reconstruction scheme finalized to discern automatically which method is the most effective for reconstructing the test patterns of a signal.

In this respect, two procedures for combining the MD and TM aggregation methods are proposed in this work. To tailor these procedures for robust signal reconstruction and since the error difference  $\varepsilon_i^{E,TM} - \varepsilon_i^{E,MD}$  is quantitatively negligible on undisturbed signals (so that using TM or MD makes really no difference, as shown for the Forsmark-3

case study by the dark dots in the left and right graphs of Figure 4) the procedures are devised with respect to the disturbed training signals (Figure 4, left) for application to the reconstruction of the test signals (be it undisturbed or disturbed).

#### 4.1 Procedure 1

The first combination procedure, sketched in Figure 6, is based on the relationship between  $d_i$  and  $\varepsilon_i^{E, TM} - \varepsilon_i^{E, MD}$ . In fact, according to the results presented in Figure 4 for the Forsmark-3 case study, the value of the signals' average model outcome distances  $d_i$  (Eq. 7) can provide some indications of the best method to adopt.

Operatively, the range of the average outcome distances  $d_i$  is partitioned into a number  $h$  of non-overlapping subsets  $\xi_l$ ,  $l=1,2,\dots,h$ . For each subset  $\xi_l$ , the average error difference  $\langle \varepsilon \rangle_l$  is computed considering only those signals with  $d_i \in \xi_l$ :

$$\langle \varepsilon \rangle_l = \sum_{i: d_i \in \xi_l} \varepsilon_i^{E, TM} - \varepsilon_i^{E, MD} \quad (10)$$

In order to identify specific portions of the range of the average outcome distances in which one method outperforms the other, neighbour subsets characterized by the same sign of  $\langle \varepsilon \rangle_l$ , e.g.  $\{\xi_l, \xi_{l+1} : \text{sign}\langle \varepsilon \rangle_l = \text{sign}\langle \varepsilon \rangle_{l+1}\}$ , are grouped together in one single cluster,

$$C_j = \bigcup_{l: \text{sign}\langle \varepsilon \rangle_l = \text{sign}\langle \varepsilon \rangle_{l+1}} \xi_l \quad j = 1, 2, \dots \quad (11)$$

By doing so, each cluster  $C_j$  is defined by the range of average outcome distances  $[d_{C_j}^{\min}, d_{C_j}^{\max}]$  within which the disturbed training signals have the tendency to be better reconstructed by MD ( $\langle \varepsilon \rangle_l > 0, \forall l: \xi_l \in C_j$ ) or TM ( $\langle \varepsilon \rangle_l < 0, \forall l: \xi_l \in C_j$ ).

These clusters define the reconstruction scheme by which deciding the method to adopt for reconstructing the test signals (Figure 6). In this view, for the generic signal  $i$ , first the average outcome distance  $d_i$  is computed using the (test) outcomes (Eq. 7); then, the (training) cluster  $C_j$  to which  $d_i$  belongs is retained for defining the most effective method for aggregating the outcomes.

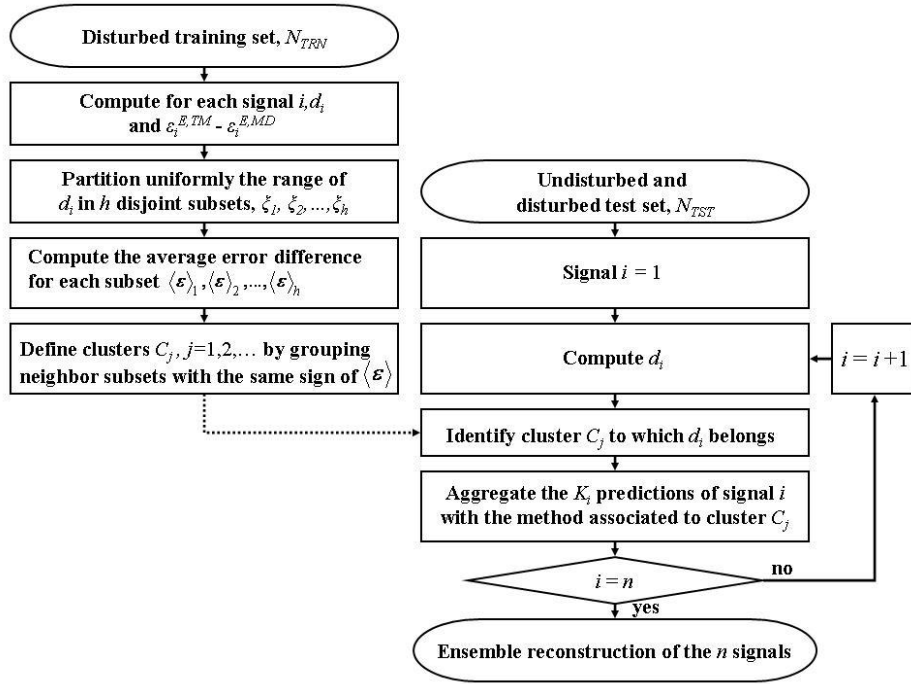


Figure 6. Sketch of the first novel procedure for combining the MD and TM aggregation methods

The proposed method has been applied to the Forsmark-3 case study by setting  $h$  equal to 30. Figure 7 shows the average error difference  $\langle \varepsilon \rangle_l$  for each subset  $\xi_l, l=1,2,\dots,30$  and the four clusters  $C_j, j=1,\dots,4$  obtained with their associated optimal method. Notice that  $C_3$  (corresponding to MD, i.e.  $\langle \varepsilon \rangle_l > 0, \forall \xi_l \in C_3$ ) includes an exception, i.e. a subset whose average error difference is negative, but quantitatively negligible and in between two subsets with positive  $\langle \varepsilon \rangle_l$ .

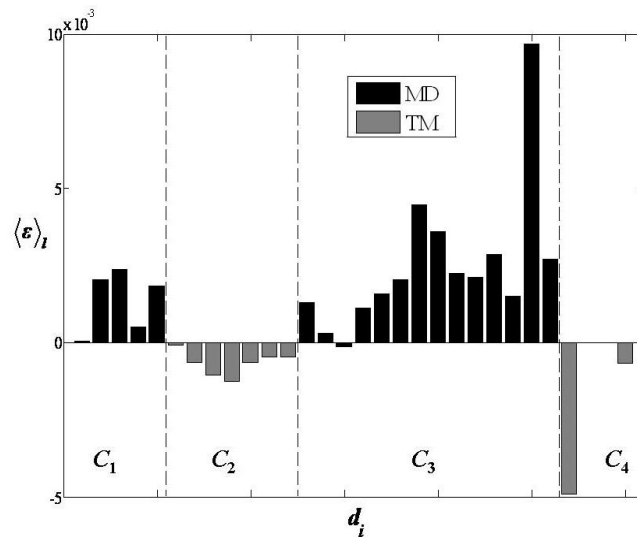


Figure 7. Application of the first combination procedure to the Forsmark-3 case study: average error difference  $\langle \varepsilon \rangle_l$  for each subset and clusters computed using disturbed training signals

Table 6 reports the ensemble reconstruction performances on undisturbed and disturbed test signals obtained by MD, TM and the first proposed mixed approach (MX1). The MX1 approach, based on the reconstruction scheme built using disturbed training signals, provides a more robust reconstruction of the disturbed test set, at the expenses of a negligible loss of accuracy on undisturbed signals.

$\times 10^{-2}$	MD	TM	MX1
$\eta^{E,U}$	1.878	1.894	1.881
$\eta^{E,D}$	7.607	7.638	7.363

Table 6. Application of the first combination procedure to the Forsmark-3 case study: ensemble reconstruction errors obtained with MD, TM and MX1 approaches on undisturbed and disturbed signals

Notice however that, the application of this combination procedure (with  $h = 30$ ) to the Loviisa case study has led to obtaining  $\langle \varepsilon \rangle_l > 0$ ,  $\forall l$ , and thus no performance improvement with respect to using the pure MD approach to reconstruct the signals. This is in spite of the fact that Table 5b reports that in the Loviisa case study some signals are better reconstructed by TM. For this reason, a second combination procedure has been developed, aimed at capturing the signal reconstruction capabilities of the MD and TM methods on each individual signal.

## 4.2 Procedure 2

The proposed procedure (sketched in Figure 8) is based on the ensemble reconstruction errors obtained on the disturbed training patterns of each signal by the MD and TM methods. The conjecture is that if the (disturbed) training patterns of a signal are better reconstructed by one method (MD or TM), then the same method will better reconstruct also the (disturbed) signal's test patterns.

In this view, for each signal  $i$ , first the ensemble reconstruction errors  $\varepsilon_i^{E,MD}$  and  $\varepsilon_i^{E,TM}$  by the MD and TM methods, respectively, are computed on the disturbed training set; then, the method by which reconstructing the test patterns of signal  $i$  is MD if  $\varepsilon_i^{E,TM} > \varepsilon_i^{E,MD}$  or, viceversa, TM if  $\varepsilon_i^{E,TM} < \varepsilon_i^{E,MD}$  (Figure 8).

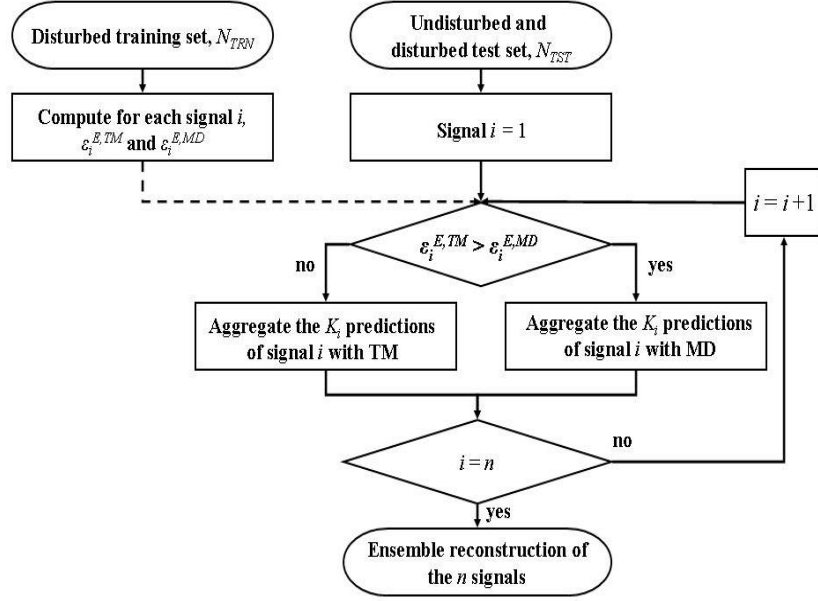


Figure 8. Sketch of the second novel procedure for combining the MD and TM aggregation methods

Table 7 reports the ensemble reconstruction performances on undisturbed ( $\eta^{E,U}$ ) and disturbed ( $\eta^{E,D}$ ) test signals obtained by MD, TM and the second proposed mixed approach (MX2) on the Loviisa and Forsmark-3 case studies. In both, the MX2 approach, based on the reconstruction scheme built using disturbed training signals, provides a more robust reconstruction of the disturbed test set; the (negligible) loss of accuracy on undisturbed signals is indeed due to the fact that the mixed signal reconstruction scheme is based on the disturbed training patterns.

$\times 10^{-2}$	Loviisa Pressurized Water Reactor (PWR)			Forsmark-3 Boiling Water Reactor (BWR)		
	MD	TM	MX2	MD	TM	MX2
$\eta^{E,U}$	0.956	0.957	0.958	1.878	1.894	1.890
$\eta^{E,D}$	5.169	5.261	5.023	7.607	7.638	7.298

Table 7. Application of the second combination procedure to the Loviisa and Forsmark-3 case studies: ensemble reconstruction errors obtained with MD, TM and MX2 approaches on undisturbed and disturbed signals

## 5. Conclusions

This work addresses the problem of reconstructing the correct signal values measured by faulty sensors in nuclear power plants. One practical approach to effectively handle the dimensionality of the problem due to the large number of sensors typically involved is offered by resorting to an ensemble of auto-associative models for signal reconstruction.

The set of signals is first subdivided into small, overlapping groups, made diverse by randomly sampling the signals according to the RFSE procedure. Then, one PCA-based reconstruction model is developed using the signals of each group. Finally, the outcomes of the models are aggregated to provide the ensemble reconstruction.

Three aggregation methods have been investigated with respect to their performance in providing accurate and robust reconstructions: Simple Mean (averaging the outcomes of all individual models), Median (taking as ensemble aggregate the outcome located in the centre of the distribution of the outcomes of the individual models) and Trimmed Mean (discarding the outcomes lying on the tails of the distribution).

The approach has been applied to two case studies concerning the reconstruction of 215 signals measured at a Finnish nuclear PWR and 920 simulated signals of a Swedish nuclear BWR. The accuracy and robustness of the three aggregation methods have been compared on the reconstruction of undisturbed and disturbed signals. In general, the analysis has proved the benefits of reconstructing signals by discarding the outlying outcomes, for they are likely to be fairly inaccurate and thus deteriorate the aggregate. In particular, the ensemble based on the Median aggregation has provided more accurate and robust performances in both applications, especially in reconstructing signals when in presence of disturbs.

On the basis of the insights gained from the investigation, two novel procedures have been developed for combining the Median and Trimmed Mean aggregations in a way to fully exploit their benefits; the novel procedures have shown an increased robustness in reconstructing signals when in presence of disturbs.

## References

- [1] Hoffmann, M., 2005. On-line monitoring for calibration reduction. HWR-784, OECD Halden Reactor Project.
- [2] Hoffmann, M., 2006. Signal grouping algorithm for an improved on-line calibration monitoring system. Proceedings of FLINS Conference, Genova, Italy.
- [3] Roverso, D., Hoffmann, M., Zio, E., Baraldi, P., Gola, G., 2007. Solutions for plant-wide on-line calibration monitoring. Proceedings of ESREL Conference, Stavanger, Norway, 1, 827-832.
- [4] Zio, E., Baraldi, P., Gola, G., Roverso, D., Hoffmann, M., 2007. Genetic algorithms for grouping of signals for system monitoring and diagnostics. Proceedings of ESREL Conference, Stavanger, Norway, 1, 833-840.
- [5] Baraldi, P., Zio, E., Gola, G., Roverso, D., Hoffmann, M., 2008. Genetic algorithms for signal grouping in sensor validation: a comparison of the filter and wrapper approaches. Journal of Risk and Reliability (JRR), Proc. IMechE, 222, Part O, 189-206.
- [6] Perrone M. P., Cooper L. N., 1993. When networks disagree: ensemble methods for hybrid neural networks. In: Mammone, R. J. Eds., 1993. Neural networks for speech and image processing. Chapman & Hall, New York, NY.
- [7] Krogh, A., Vedelsby, J., 1995. Neural network ensembles, cross-validation and active learning. In: Tesauro, G., Touretzky, D. S., Loen, T. K., 1995. Advances in newel information processing systems, 7, 231-238, MIT press, Cambridge, MA.
- [8] Sharkey, A. J. C., 1996. On combining artificial neural nets. Connection Science, 8(3), 299-314.
- [9] Gola, G., Zio, E., Baraldi, P., Roverso, D., Hoffmann, M., 2007. Signal grouping for sensor validation: a multi-objective genetic algorithm approach. HWR-852, OECD Halden Reactor Project.

- [10] Baraldi, P., Zio, E., Gola, G., Roverso, D., Hoffmann, M., 2008. Reconstruction of faulty signals by an ensemble of principal component analysis models optimized by a multi-objective genetic algorithm, Proceedings of FLINS Conference, Madrid, Spain.
- [11] Gola, G., Zio, E., Baraldi, P., Roverso, D., Hoffmann, M., 2008. Reconstructing signals for sensor validation by a GA-optimized ensemble of PCA models. HWR-894, OECD Halden Reactor Project.
- [12] Aly, M. A., Atiya, A. F., 2006. Novel methods for the feature subset ensembles approach. Article available at <http://alumnus.caltech.edu/~amir/FSE5.pdf>.
- [13] Bryll, R., Gutierrez-Osuna, R., Quek, F., 2003. Attribute bagging: improving accuracy of classifiers ensembles by using random feature subsets. Pattern Recognition, 36, 1291-1302.
- [14] Baraldi, P., Zio, E., Gola, G., Roverso, D., Hoffmann, M., 2009. A procedure for the reconstruction of faulty signals by means of an ensemble of regression models based on principal components analysis. Proceedings of the ANS Topical Meeting (NPIC-HMIT), Knoxville, TN.
- [15] Polikar, R., 2006. Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21-45.
- [16] Jolliffe, I. T., 2002. Principal component analysis. Springer Eds.
- [17] Diamantaras, K. I., Kung, S. Y., 1996. Principal component neural networks: theory and applications. John Wiley & Sons Inc., New York, NY.
- [18] Scholkopf, B., Smola, A., Muller, K.-R., 1999. Kernel principal component analysis. In Scholkopf, B., Burges, C. J. C., Smola, A. Eds., 1999. Advances in kernel methods – SV learning. 327-352, MIT Press, Cambridge, MA.
- [19] Moore, B., 1981. Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Transactions on Automatic Control, 26(1), 17-32.
- [20] Karlsson, T., Jokstad, H., Meyer B. D., Nihlwing, C., 2001. The HAMBO BWR simulator of HAMMLAB. HWR-663, OECD Halden Reactor Project.
- [21] Kuncheva, L. I., 2004. Classifier ensembles for changing environments. In: Proc. 5<sup>th</sup> International Workshop on Multiple Classifier systems, 1-15, Springer-Verlag, Berlin, Germany.
- [22] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998. On combining classifiers. IEEE Transaction on Pattern Analysis and Machine Intelligence, 20(3), 226-239.