# Genetic Algorithms for Grouping of Signals for System Monitoring and Diagnostics

E. Zio, P. Baraldi & G. Gola
*Department of Nuclear Engineering, Polytechnic of Milan, Via Ponzio 34/3, 20133, Milan, Italy*

D. Roverso & M. Hoffmann
*Institutt for Energiteknikk, OECD Halden Reactor Project, 1751, Halden, Norway*

ABSTRACT: A robust monitoring system must be able to detect also the faults possibly occurring in the sensors and to reconstruct correctly the signals monitored by the sensors detected as failed. To this aim, it is important to identify the appropriate groups of signals which carry sufficient information for reconstructing the signal(s) of the failed sensor(s) with the required accuracy. In this paper, the problem of finding an optimal grouping of signals for sensor validation is handled by means of a Multi-Objective Genetic Algorithm (MOGA) optimization. Two objectives are considered: the maximization of the correlation between signals and the maximization of the group size. The approach is applied to a data set of 84 signals collected from a boiling water nuclear power plant located in Oskarshamn, Sweden.

## 1 INTRODUCTION

Accurate monitoring of operating conditions bears important feedbacks on the operation of modern plants, with respect to production, accident management and maintenance. Such monitoring is based on a system of sensors. During plant operation, some sensors may experience anomalies (e.g. drifts) which might convey inaccurate or misleading information about the actual plant state to automated controls and to the operators. A robust monitoring system must be able to detect such anomalies and reconstruct correctly the signals of the failed sensors.

On-line sensor monitoring evaluates instrument channel performance by assessing its consistency with other plant indications (signal validation). Information about the condition, performance and calibration state of the channels through accurate and frequent monitoring while the process is in operation provides a basis for determining when signal reconstructions and sensor recalibrations are necessary (Hoffmann 2005).

In real systems, the number of sensor signals involved in the monitoring of the plant state is too large to be handled effectively within a single validation and reconstruction model. One approach to address the problem is to divide the signals into smaller groups and develop one model for each set of signals.

In this paper, signal grouping is carried out by means of Genetic Algorithms (GAs) (Holland 1975, Goldberg 1989, Chambers 1995, Sawaragy et al. 1985, Raymer et al. 2000, Bozdogan 2003). The objective functions of the grouping optimization need to capture several aspects. Most importantly, a good degree of correlation between the signals in a group is required for an effective and accurate signal validation and reconstruction (Hoffmann 2005, Hoffmann 2006).

Once the groups are generated, the signals of each group are fed to a specifically trained Auto-Associative Neural Network (AANN) (Kramer 1992, Hines et al. 1998, Fantoni 2005, Fantoni & Mazzola 1996, Marseguerra et al. 2004), for signal validation and reconstruction. Other estimation modelling techniques may also be adopted, such as the Neural Network Partial Least Squares (NNPLS) algorithm (Hoffmann & Kirschner 2004, Kirschner & Hoffmann 2004).

The paper is organized as follows. Section 2 presents the problem of signal grouping and briefly illustrates the techniques used to solve it. Section 3 focuses on the Genetic Algorithm approach here developed. In Section 4, the approach is applied to a data set of signals measured by 84 sensors at a Boiling Water Reactor (BWR) in Oskarshamn, Sweden. Finally, some conclusions on the advantages and limitations of the proposed grouping technique are drawn in the last Section.

## 2 AN OVERVIEW ON SIGNAL GROUPING

The problem of signal grouping regards the task of discerning out of the several signals of the monitored process or system those to be grouped together within an efficient model for signal validation and reconstruction.

In practical cases, the number of signals monitored is too large to be handled by a single validation and reconstruction model. To overcome this problem, one may decompose the modelling task into a number of models based on subsets of signals, thus reducing the complexity of the modelling and providing a more efficient signal validation and reconstruction.

Formally, given $n >> 1$ sensors' signals $f_i$ to be validated, $i = 1,2,\ldots,n$, it is desirable to group them in $K$ groups $k_j$, $j = 1,2,\ldots,K$, each constituted by $m_j << n$ signals and bearing some required characteristics useful for the signal validation and reconstruction tasks.

The inclusion or not of a signal in a group can be encoded in terms of a binary variable which takes value 1 or 0, respectively. For $n$ signals, the size of the binary vector space in which the search for the $K$ groups is to be carried out is $2^n$. Each group of signals selected during the search must be evaluated with respect to the given objective functions, e.g. the correlation between the signals and the group size. An exhaustive search is impractical unless $n$ is small.

Several methods for signal grouping can be adopted. A simple and direct way for generating groups is based on the pairwise correlation between signals. Setting to $m$ the desired size of the group, the first $m-1$ most correlated signals to each signal are put in the same group together with the signal itself. Letting $m$ vary in a preset range, $n$ groups of signals of different sizes can be generated, each one characterized by the value of the average correlation between the signals. Then, the $K$ most correlated groups can be selected manually and separately used as basis for validating the signals. This approach is expected to work well for the signal of reference of the group, to which all other signals are most highly correlated. However, it cannot ensure that the other $m-1$ signals which are most correlated to the reference signal have also a high degree of correlation between themselves and thus it may not work optimally in their validation and reconstruction. Nevertheless, this approach represents a good basis of comparison for other more refined grouping techniques.

Grouping techniques can be classified into two main categories: filter and wrapper methods (Kohavi & John 1997). In the former, the grouping algorithm functions as a filter to include/discard the signals in the groups based on their characteristics, independently of the specific algorithm used for signal validation and reconstruction (e.g. AANNs). Numerical evaluation functions are used to compare the goodness of the groups of signals selected by a search algorithm, with respect to signal characteristics which are regarded relevant for their validation and reconstruction. At the end of the search, the groups with best values of the evaluation functions are kept for building the models (e.g. AANNs) for signal validation and reconstruction.

Contrary to filter methods, in wrapper methods the signal grouping algorithm behaves as a "wrapper" around the specific algorithm used to validate and reconstruct the signals of the groups. The performance of the validation and reconstruction algorithm itself is directly used to compare the different groups of signals selected by the search algorithm (Kohavi & John 1997).

The filter approach is generally computationally more efficient than the wrapper one because for each set of signals of trial, the computation of the evaluation functions is less time consuming than the development of a complete validation and reconstruction model, as required by the wrapper approach. Indeed, a high number of groups of signals are tested during the search for the optimal ones and the time consumption of a wrapper approach depends mainly on the time necessary for the development of the validation and reconstruction model, e.g. the training and testing of AANNs. Hence, for many practical applications the wrapper approach is feasible only if the validation and reconstruction model is a fast-computing algorithm (Duran & Odell 1974). On the other hand, wrapper approaches are more performing than the filter ones since the former ensure the creation of the groups of signals tailored for the specific validation and reconstruction algorithm used, whereas the latter totally ignore the effects of the created groups of signals on the performance of the validation and reconstruction model that will actually be used.

With respect to the group search algorithms, three approaches are commonly adopted: complete, heuristic and probabilistic (Kohavi & John 1997). In the complete approach, the properties of a pre-defined evaluation function are used to prune the signal search space to a manageable size (Duran & Odell 1974). Only some evaluation functions give rise to a search that guarantees the optimal groups selection without being exhaustive.

The heuristic approach does not guarantee that the best groups of signals are achieved, but is less time consuming than the complete one and may be employed in combination with any evaluation function (Zio et al. 2005). At present, the most employed heuristic methods are greedy search strategies such as the Sequential Forward Selection (SFS) or the Sequential Backward Elimination (SBE) “hill climbing” methods, which iteratively add or subtract signals to the groups and at each iteration the evaluation function is evaluated. The forward selection refers to a search that begins with the empty group and at each step a signal is added to the subspace; on the contrary, the backward elimination refers to a search that begins with the *n*-dimensional signal set and at each step a signal is removed. At each step, the choice of which signal to add or remove is driven by its effect on the evaluation function in the direction of climbing towards its maximum value. The hill-climbing search is usually stopped when adding or removing new signals does not increase the value of the evaluation function or when the number of signals has reached a predefined threshold.

The hill-climbing methods suffer from the so called "nesting effect": if the signals added cannot be removed, a local minimum of the evaluation function may be found. To reduce this effect, it is possible to use the so called plus-*l*-take-away-*r* method (PTA) (Kohavi & John 1997). In this method, after *l* steps of the forward selection, *r* steps of the backward elimination are applied so as to allow escaping from local minima. Still, there is no guarantee of obtaining the absolute optimum.

The probabilistic approach is based on population-based methauristics guided by the goodness of the solutions, such as the Genetic Algorithms presented in this work, or on methods like simulated annealing and tabu search algorithms (Zhang & Sun 2002).

## 3 A GENETIC ALGORITHM APPROACH TO SIGNAL GROUPING

Mathematically, the problem of signal grouping can be formulated as an optimization problem aimed at finding the optimal groups of signals for achieving the best performance of the validation and reconstruction model, i.e. the smallest error in the reconstruction of the signals.

In this Section, a probabilistic approach to signal grouping based on Genetic Algorithm is propounded.

In this respect, the probabilistic search is performed by constructing a sequence of populations of chromosomes, the individuals of each population being the children of those of the previous population and the parents of those of the successive population. A population is constituted by $K$ chromosomes each one constituted by $n$ bits and representing all the possible signals included in a group. The initial population is generated by randomly sampling the bits of all the strings. At each step, the new population is then obtained by manipulating the strings of the old population by repeatedly performing the four fundamental operations of reproduction, crossover, replacement and mutation (all based on random sampling) in order to arrive at a new population hopefully characterized by an increased mean fitness. This way of proceeding enables to efficiently arrive at optimal or near-optimal solutions.

The problem is here framed as a Multi-Objective Genetic Algorithm (MOGA) optimization search (Goldberg 1989, Chambers 1995, Sawaragy et al. 1985, Raymer et al. 2000, Bozdogan 2003, Zio et al. 2006) in a filter configuration.

In the generic $j$-th chromosome coding group $k_j$, the $i$-th bit encodes the presence (1) or absence (0) of the $i$-th signal $f_i$ in the $j$-th group, $i = 1,2,..., n$, $j = 1,2,..., K$ (Fig. 1).

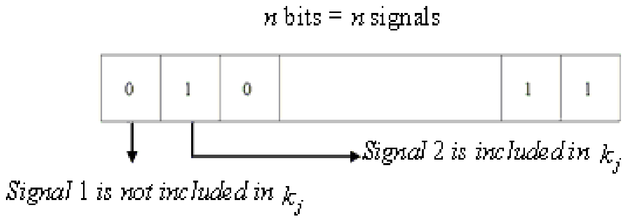


Figure 1. The structure of the generic $j$-th chromosome.

Operatively, the total number $N$ of available *n*-dimensional data patterns are partitioned into a set ($X_{MOGA}$) used for the signal grouping task and a separate set ($X_{AANN}$) of approximately the same size used for signal validation.

Then, a GA can be devised to find an optimal set of binary $n$-dimensional transformation vectors which operate on $X_{MOGA}$ to maximize/minimize a set of optimization criteria regarding the group of signals.

In a multi-objective optimization problem, several possibly conflicting objective functions must be evaluated in correspondence of each chromosome. The comparison of chromosomes (i.e. groups) is achieved in terms of the concepts of Pareto optimality and dominance (Goldberg 1989, Chambers 1995, Sawaragy et al. 1985). The chromosomes that are non-dominated within the entire search space are said to be Pareto optimal and constitute the so called Pareto optimal front.

In this respect, let $m$ be the number of 1's and $n-m$ the number of 0's in a transformation vector: then, a modified set of $m$-dimensional patterns is obtained and used to evaluate, in terms of dominance, the quality of the groups with respect to the objective functions (Raymer et al. 2000, Bozdogan 2003, Zio et al. 2006). On the basis of this feedback, the GA conducts its probabilistic search for the Pareto-optimal groups of signals.

In this paper, two objective functions are considered for the probabilistic optimization: the maximization of the average correlation of the signals in a group and the maximization of the group size.

The motivation of the first objective function is related to the fact that the signals in a generic group will be used to build a model for validating and reconstructing the signals themselves, with the conjecture that highly correlated signals are capable of regressing one another. The measure used to quantify the objective function is the Pearson's correlation coefficient (Lawrence & Kuei 1989, Hunt 1986): considering $N$ measurements of the two signals $f_p(t)$ and $f_q(t)$, $t = 1,2,..., N$, the Pearson's coefficient is:

$$corr_{p,q} = \frac{1}{N-1}\sum_{t=1}^{N}\left(\frac{f_p(t)-\overline{f}_p}{S_{f_p}}\right)\left(\frac{f_q(t)-\overline{f}_q}{S_{f_q}}\right) \quad (1)$$

where $\overline{f}_p$, $S_{f_p}$, $\overline{f}_q$ and $S_{f_q}$ are the mean values and standard deviations of $f_p$ and $f_q$, respectively.

By definition, the value of $corr_{p,q}$ varies from $-1$ to 1, being 0 for statistically independent quantities. Signals that have the same trend (both increasing or both decreasing with respect to the mean) will have positive values of $corr_{p,q}$, whereas those which vary in opposite ways will render $corr_{p,q}$ negative.

To associate a degree of correlation to the generic $k$-th group of $m_k$ signals, $k = 1,2,..., K$, first the average correlation between the $p$-th signal, $p = 1,2,..., m_k$, and the remaining $m_k - 1$, is computed:

$$\langle corr_p \rangle = \frac{1}{m_k - 1}\sum_{\substack{q=1\\ q\neq p}}^{m_k} corr_{p,q} \quad (2)$$

Then, the group correlation $r_k$ is computed as the mean of the average correlations of each signal:

$$r_k = \frac{1}{m_k}\sum_{p=1}^{m_k}\langle corr_p \rangle \quad (3)$$

Finally, the group correlation measure is normalized between 0 and 1.

$$r_k^* = \frac{1+r_k}{2} \quad (4)$$

As for the second objective function, the number of signals $m$ in each group is maximized to ensure that each signal is included in at least one group.

At convergence of the MOGA search, the sets of signals of the two-dimensional Pareto front are identified by the number of signals $m$ and the average correlation $r_k^*$. Among these, $K$ groups upon which to base the signal validation and reconstruction can be manually selected according to the specifics of the problem. The choice depends both on the expert's system knowledge, i.e. selecting those groups including the signals mostly relevant for monitoring the system, and on the complexity, i.e. number of parameters, and desired accuracy proportional to the training time of the adopted validation model. For example, one could choose few large groups which globally include all the signals and therefore training a small number of highly parameterized validation models or a large number of small groups in order to include the signals with a certain redundancy thus using more models and combining the results.

## 4 APPLICATION

The proposed approach is applied to a real case study concerning the grouping of $n$=84 signals which have been collected from a BWR located in Oskarshamn, Sweden (Fig. 2).

The MOGA code here used has been developed by the Laboratorio di Analisi di Segnale e di Analisi di Rischio (LASAR, Laboratory of Analysis of Signals and Analysis of Risk) of the Department of Nuclear Engineering of the Polytechnic of Milan (http://lasar.cesnef.polimi.it ).

The 84 signals have been sampled every 10 minutes from May 31 2005 to January 5 2006 from a corresponding number of sensors, providing a total amount of $N$=30110 time samplings. Each instant recording provides an 84-dimensional pattern $f_1(t), f_2(t),..., f_{84}(t)$, $t = 1,\ldots, N$, identified by the values of the $n$=84 signals at the time instant $t$, as illustrated in Table 1.

Of the $N$ patterns available, $N_{MOGA}$=14996 have been used for the MOGA grouping, $N_{AANN}^{train}$=12000 to train the validation and reconstruction AANN models and $N_{AANN}^{test}$=3114 to test them (Fantoni 2005, Fantoni & Mazzola 1996, Marseguerra et al. 2004). Training and testing of the AANNs has been performed with the AANN code also developed by the LASAR group.

Table 2 shows the GA settings here adopted, with a synthetic explanation of the terms used. Since the correlation function (Eq. 4) is very fast to calculate, the wrapping MOGA search has been run for 30000 generations, for a computing time of approximately 20 minutes. Actual convergence was attained after 2000 generations.

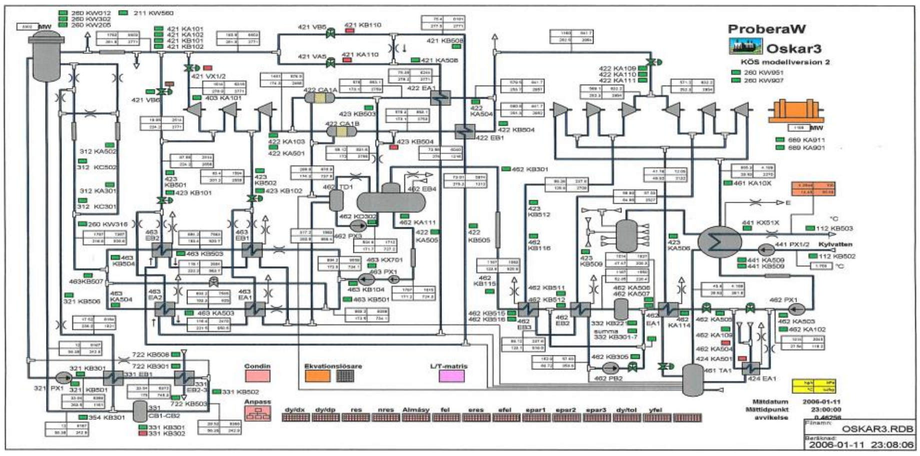


Figure 2. The process diagram of the nuclear power plant in Oskarshamn. The 84 signals are identified by an alpha-numeric code (e.g., 312 KA301, 423 KB509, etc…)

Table 1. Partition of the data set for MOGA grouping and AANN training and test

| Time samplings ($t$ =1,…,30110) | | | Signals ($i$ =1,...,84 ) | | |
|---|---|---|---|---|---|
| MOGA data set ($N_{MOGA}$=14996) | | 1 | $f_1(1)$ | ... | $f_{84}(1)$ |
| | | … | … | | … |
| | | 14996 | | | |
| AANN data set ($N_{AANN}$ =15114) | Training ($N^{train}_{AANN}$ = 12000) | 14997 | | | |
| | | … | | | |
| | | 26996 | | | |
| | Test ($N^{test}_{AANN}$ = 3114) | 26997 | | | |
| | | … | … | | … |
| | | 30110 | $f_1(30110)$ | | $f_{84}(30110)$ |

Table 2. Parameters used for the MOGA searches

| | |
|---|---|
| Selection | *FIT-FIT*: the population, rank-ordered on the basis of the Pareto dominance criterion, is scanned and each individual is parent-paired with an individual of the next fittest Pareto rank class. |
| Replacement | *FITTEST*: out of the four individuals (two parents and two chromosomes) involved in the crossover procedure, the fittest two replace the parents |
| Mutation probability | $10^{-2}$ |
| Population size | 100 |
| Number of Generations | 30000 |

Figure 3 shows the Pareto front and final population obtained at convergence of the MOGA search. Each of the $K$ points on the graph represents a dominant solution, i.e. a group, identified by the number $m$ of signals and the average correlation between themselves $r^*_k$ (Eq. 4), $k = 1,2,..., K$ .

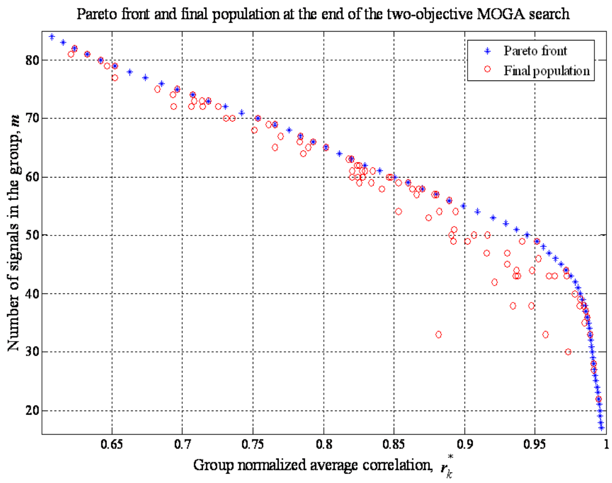


Figure 3. Pareto front and final population at the end of the two-objective MOGA search

As a preliminary validation of the adopted MOGA approach, the optimal groups of the Pareto front obtained by the genetic search are compared with those obtained by grouping the most correlated signals of a given signal (Fig. 4). Since the Pareto front obtained at convergence of the MOGA search is constituted by groups ranging from 17 to 84 signals and always including signal 7, this has been taken as the reference signal for creating groups of $m = 17,18,....,84$ most correlated signals, based on the method explained in Section 2.

A reasonable number of signals which can be handled in a validation and reconstruction model is around 30 (as indicated in Figure 4). For the optimal group of this size, the MOGA finds an average correlation of 0.9901 against the value of 0.9877 by the simple correlation method (Fig. 4).

In order to verify the appropriateness of the correlation between signals as a grouping criterion, two AANNs have been trained to perform the signal validation and reconstruction, based on $N^{train}_{AANN}$ =12000 patterns of 30 signals: one with the 30 highly correlated signals of the optimal Pareto front group and the other with 3 signals of such group, namely number 7, 32 and 56 and the other 27 signals randomly chosen among the remaining 54 of the 84 signals available. Table 3 reports the main features of the architecture adopted for the AANNs and Table 4 the characteristics of the two groups. The trained AANNs are fed with the $N_{AANN}^{test}$=3114 patterns to give in output the reconstructed estimates of the 30 signals received in input.

Figure 5 reports the reconstruction of signal 7 common to the two groups. The reconstruction based on the correlated signals of the MOGA optimal group leads to smaller errors especially in correspondence of steady-states and sharp variations of the signal to be reconstructed. The average relative errors for signals 7, 32 and 56 common to the two different groups of signals achieved using the two groups are reported in Table 5. Since the signals of the optimal group are more correlated than those of the random group, the reconstruction is more accurate

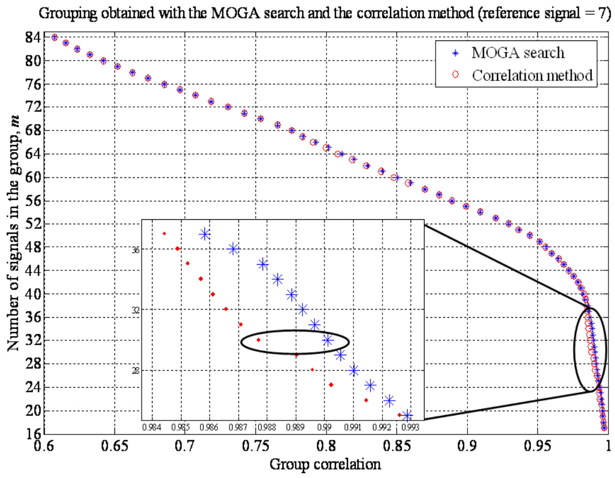


Figure 4. Comparison between the grouping obtained by the MOGA search and the correlation method.

Table 3. The AANNs architecture and training parameters.

| | | |
|---|---|---|
| Number of nodes | Mapping layer | 45 |
| | Bottleneck layer | 15 |
| | De-mapping layer | 45 |
| Learning rate ( $\eta$ ) | | 0.6 |
| Momentum ( $\alpha$ ) | | 0.8 |
| Number of iterations | | $10^6$ |

Table 4. Characteristics of the optimal and random groups. In bold the common signals.

| Group | Numb. of signals | Average correlation | Signal labels |
|---|---|---|---|
| Optimal | 30 | 0.99010 | **7** 8 9 10 11 12 22 31 **32** 33 34 35 41 42 43 44 45 46 **56** 63 64 65 7173 74 76 77 78 79 81 |
| Random | 30 | 0.54789 | 1 4 6 **7** 13 15 17 19 21 24 27 28 **32** 36 37 40 50 51 52 53 54 **56** 57 59 61 66 67 69 75 80 |

Table 5. Average relative errors on the test of $N_{AANN}^{test}$=3114 patterns achieved using the optimal and random groups, for the 3 signals common to both groups.

| Average relative errors | Signal 7 | Signal 32 | Signal 56 |
|---|---|---|---|
| Optimal group | 4.105 x $10^{-3}$ | 6.920 x $10^{-3}$ | 1.069 x $10^{-2}$ |
| Random group | 1.448 x $10^{-2}$ | 2.504 x $10^{-2}$ | 2.820 x $10^{-2}$ |

As previously mentioned, the main purpose of an effective signal monitoring and validation scheme during plant operation is to allow the reconstruction of the signal values when in presence of sensor failures, e.g. drifts. For this, in our case the trained AANNs must be sufficiently ‘‘robust’’ in the sense that they should provide a good estimate of the correct output corresponding to the faulty sensor signal, without deteriorating too much the other outputs of the network associated to correct signals (Fantoni 2005, Fantoni & Mazzola 1996, Marseguerra et al. 2004).

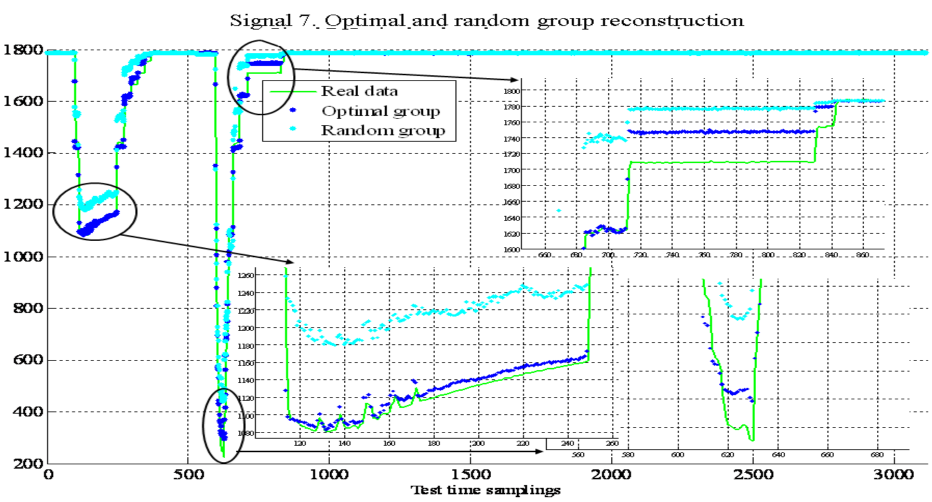


Figure 5. Reconstruction of signal 7 (with details) by the AANN fed with the signals of the optimal (dark dots) and random (light dots) groups.

When one of the sensors of an identified group becomes faulty, it sends in input to the corresponding AANN a faulty signal; if the network is well trained, it should provide as output a good estimate of the true value of the faulty signal, obtained through the correlated information provided by the other signals coming from the non-faulty sensors of the group. To verify this aspect, a linear drift has been imposed to signals 7, 32 and 56 of both the MOGA-optimal and random groups of 30 signals previously considered (Tab. 4). The disturbance has been set so as to decrease linearly the value of the signals up to 75% of their real values.

Figure 6 shows the reconstruction of signal 7 obtained by the AANNs trained with the signals of the optimal and random groups. As illustrated in the highlighted regions of the Figure (corresponding to ramps and peaks where performing a good signal reconstruction in more difficult), when the validation and reconstruction model is tested on the drifted signal the more correlated signals of the MOGA optimal group allow for a superior reconstruction of the real signals with respect to the random group, even though at the tale of the time series, i.e. where the drift gets larger, the random group slightly outperforms the optimal group. Nevertheless, Table 6 reports that globally the average relative errors committed by the model trained with the optimal group is smaller than that obtained using the random group.

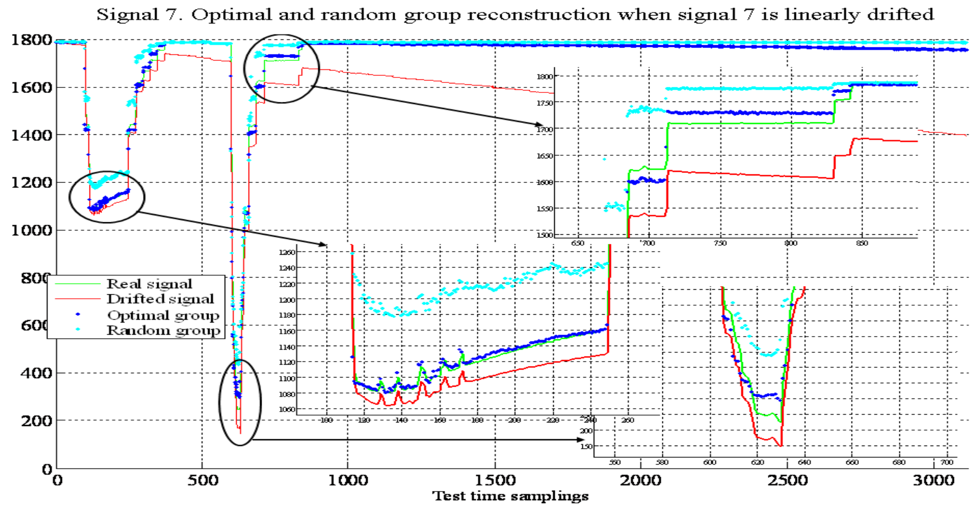


Figure 6. Reconstruction of signal 7 (with details) when linearly drifted (light grey), by the AANNs fed with the signals of the optimal (dark dots) and random (light dots) groups

Table 6. Average relative errors during test achieved using the optimal and random groups, for the three signals common to both groups, in case of drift.

| Average relative errors | Signal 7 | Signal 32 | Signal 56 |
|---|---|---|---|
| Optimal group | 0.1256 | 0.1687 | 0.1780 |
| Random group | 0.1469 | 0.2173 | 0.2255 |

## 5 CONCLUSIONS

This paper addresses the problem of sensor monitoring for signal validation and reconstruction purposes, from the practical point of view.

A major practical issue relates to the fact of dealing with the large number of signals typically collected in real plants. Currently, it is not possible to manipulate all these signals as a whole by an efficient validation and reconstruction model. Then, one must find a way of optimally grouping the signals in smaller sets, which can then be used for developing effective models for signal validation and reconstruction.

In this work, the problem of signal grouping has been tackled by means of a Multi-Objective Genetic Algorithm in a filter configuration. The two objectives of the search are the maximization of the correlation between the signals in a group and the maximization of the group size. The first objective function has been chosen because highly correlated signals are shown capable of re-constructing and regressing one another. The need to ensure that a signal is included in at least one group leads to maximizing the group size as a second objective.

At convergence of the MOGA search, the Pareto front is constituted by a number of groups of highly correlated signals, which are used for developing a corresponding number of signal validation and reconstruction models, in this case Auto-Associative Neural Networks.

The approach has been applied to a real case study concerning signals collected from a nuclear BWR located in Sweden. In order to prove the effectiveness of the correlation as a grouping criterion, the validation and reconstruction capabilities of the AANN modelling technique have been tested with respect to two different groups: one optimal group of the Pareto front found by the MOGA, and thus constituted by highly correlated signals, and one group of the same size mostly constituted by randomly chosen and thus poorly correlated signals. As expected, the AANN based on the signals of the MOGA optimal group has provided a globally more accurate re-construction than the AANN based on random signals.

Finally, a linear drift, simulating a sensor failure, has been imposed to the three signals common to the optimal and random groups. The perturbed signals thereby obtained have been fed in input to the trained AANNs for testing their robustness as reconstruction models, i.e. their capability of exploiting the undrifted signals of the groups for reconstructing the correct values of the corrupted signals. Also in this case, the AANN based on the highly correlated signals of the optimal group have obtained a more precise reconstruction than the AANN based on random signals. Although these differences cannot be considered conclusive and could be due to a case of non optimal training of the AANN on the random group and more extensive tests including several trainined AANNs and several group selections are needed to confirm these first results, the adopted MOGA approach to signal grouping based on the correlation between the signals is worth being further investigated.

## REFERENCES


Bozdogan, H., 2003, Statistical data mining with informational complexity and genetic algorithms, in: H. Bozdogan, “Statistical Data Mining and knowledge discovery”, CRC Press.

Chambers, L., 1995, Practical handbook of genetic algorithms: applications Vol. I; new frontiers Vol. II, CRC Press.

Duran, B., Odell, P., 1974, Cluster analysis: a survey, Springer-Verlag, New York.

Fantoni, P.F., 2005, Experiences and applications of Peano for on-line monitoring in power plants, Progress in nuclear energy 46, 206-225, Elsevier.

Fantoni, P.F., Mazzola, A., 1996, Multiple-failure signal validation in nuclear power plants using artificial neural networks, Nuclear technology, Vol. 113, No. 3, pp. 368-374.

Goldberg, D.E., 1989, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley Publ. Co.

Hines, J.W., Uhrig, R.E., Wrest, D.J., 1998, Use of autoassociative neural networks for signal validation, Journal of Intelligent and Robotic Systems, Vol. 21, No. 2.

Hoffmann, M., 2005, On-line monitoring for calibration reduction, HWR-784.

Hoffmann, M., 2006, Signal grouping algorithm for an improved on-line calibration monitoring system, Proceedings of FLINS, Genova, Italy.

Hoffmann, M., Kirschner, A., 2004, PEANO - Findings from using the NNPLS algorithm and HAMMLAB applications, HWR-690.

Holland, J.H., 1975, Adaptation in natural and artificial systems: an introductory analysis with application to biology, Control and Artificial Intelligence. MIT Press, 4-th edition.

Hunt, R.J., 1986, Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability, Journal of Dental Research, Vol 65, pp. 128-130.

Kirschner, A., Hoffmann, M., 2004, PEANO NNPLS: advancements in 2002-03, HWR-741.

Kohavi, R., John, G., 1997, Wrappers for feature subset selection, Artificial Intelligence 97, pp. 273-324.

Kramer, M.A., 1992, Autoassociative neural networks, Computers & Chemical Engineering, Vol. 16, No. 4, pp. 313-328.

Lawrence, I., Kuei, L., 1989, A concordance correlation coefficient to evaluate reproducibility, Biometrics, Vol. 45, No. 1, pp. 255-268

Marseguerra, M., Zio, E., Marcucci, F., 2004, A soft-computing based classification procedure for the identification of transients in the steam generator of a pressurized water reactor, Annals of Nuclear Energy 31, pp. 1429-1446.

Raymer, M.L., Punch, W.F., Goodman, E.D., Khun, L.A., Jain, A.K., 2000, Dimensionality reduction using genetic algorithms, IEEE Transaction on Evolutionary Computation, vol. 4, No. 2.

Rumelhart, D., McClelland, J., 1986, Parallel distributed processing: explorations in the microstructure of cognition, MIT Press, Cambridge, MA.

Sawaragy, Y., Nakayama, H., Tanino, T., 1985, Theory of multiobjective optimization, Academic Press, Orlando, Florida.

Zhang, H., Sun, G., 2002, Feature selection using tabu search method, Pattern Recognition 35, pp 701-711.

Zio, E., Baraldi, P., Pedroni, N., 2006, Selecting features for nuclear transients classification by means of genetic algorithms, accepted for publication on IEEE Trans. on Nuclear Science.

Zio, E., Baraldi, P., Roverso, D., 2005, An extended classifiability index for feature selection in nuclear transients, Annals of Nuclear Energy 32, pp.1632-1649.