Automatic tuning via Kriging-based optimization of methods for fault detection and isolation
Julien Marzat, Eric Walter, Hélène Piet-Lahanier, Frédéric Damongeot

To cite this version:

HAL Id: hal-00520808
https://centralesupelec.hal.science/hal-00520808
Submitted on 24 Sep 2010

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.
Automatic tuning via Kriging-based optimization
of methods for fault detection and isolation

Julien Marzat, Éric Walter, Hélène Piet-Lahanier, Frédéric Damongeot

Abstract—All the methods for Fault Detection and Isolation (FDI) involve internal parameters, often called hyperparameters, that have to be carefully tuned. Most often, tuning is ad hoc and this makes it difficult to ensure that any comparison between methods is unbiased. We propose to consider the evaluation of the performance of a method with respect to its hyperparameters as a computer experiment, and to achieve tuning via global optimization based on Kriging and Expected Improvement. This approach is applied to several residual-evaluation (or change-detection) algorithms on classical test-cases. Simulation results show the interest, practicability and performance of this methodology, which should facilitate the automatic tuning of the hyperparameters of a method and allow a fair comparison of a collection of methods on a given set of test-cases. The computational cost turns out to be much lower than the one obtained with other general-purpose optimization methods such as genetic algorithms.

I. INTRODUCTION

A fault detection and isolation (FDI) procedure is usually made up of a residual generator, and a method for residual analysis that processes these residuals [1]. This is used to decide whether a fault is present and then which fault.

Each of the many change-detection methods has internal parameters that must be carefully tuned. These parameters, often called hyperparameters, have a strong impact on performance and robustness. The user may thus be at a loss for selecting the most efficient method. This can be achieved by first defining a suitable performance criterion and then finding a way of tuning the hyperparameters of each method in order to optimize this criterion, on a representative set of test-cases.

The main existing tools for the tuning of hyperparameters are cross-validation and its variants (k-fold cross-validation, leave-one-out cross-validation, generalized cross-validation [2]). Cross-validation is used to estimate the performance of a given parameter value of the parameter vector and can then be complemented by an optimization procedure to find the best tuning of these hyperparameters. In [3] and [4], such approaches based on a discretization of hyperparameter spaces have been presented. Another method using Bayesian networks for tuning parameters has been proposed in [5], where prior knowledge consists of the previous simulation runs. All these approaches prove to be extremely computer-intensive and thus not applicable when the simulation budget is severely limited.

This paper describes an optimization procedure that is dedicated to this type of problem, and its application to the automatic tuning of methods for FDI. Following the computer experiment framework [6], we propose to use a global optimization algorithm relying on Kriging and the notion of Expected Improvement to explore real-valued hyperparameter spaces effectively at a limited computational cost.

Section II formally presents the problem and explains the basics of the tuning methodology. Section III describes illustrative test-cases, examples of candidate methods to be tuned and compared, along with classical performance indices in FDI to be used as optimization criteria. Results are reported in Section IV, and conclusions and perspectives in Section V.

II. HYPERPARAMETER-TUNING METHODOLOGY

A. Problem formulation

Assume several FDI methods compete for the same application. The j-th method depends on a vector $x^j \in \mathcal{X}^j \subset \mathbb{R}^{d_j}$ of hyperparameters, where $\mathcal{X}^j$ is the feasible hyperparameter space and $d_j = \dim x^j$. All of these methods are to be compared using the same real-valued performance criterion $y$. This criterion could combine several performance indices, e.g., the trade-off between false-alarm and non-detection rates for change-detection procedures. Tuning the j-th method means looking for the value of $x^j$ that minimizes $y(x^j)$. A possible way to compare methods is then to rank them according to their best value for $y$.

The tuning of a given method is central to the selection of the best of them and will now be considered. For the sake of simplicity, the index $j$ will be omitted in what follows. The cost function is thus a scalar function $y(x)$, where $x \in \mathcal{X} \subset \mathbb{R}^d$. The only available information is the result of previous computer experiments that provides the value of $y(x)$ for given values of $x$. The procedure is recursive and we shall assume that we have already computed $n$ samples forming the vector $y_n = [y_1, ..., y_n]^T$ corresponding to $\mathcal{X}_n = \{x_1, ..., x_n\}$. Since the evaluation of $y(x)$ is expensive, we shall use a simpler prediction $\hat{y}(x)$ of $y(x)$ based on these samples and obtained by Kriging.

B. Basics of Kriging

Kriging has been given this name by the French geostatistician G. Matheron, to recognize the seminal influence of...
the work of D.G. Krige on the gold deposit of the Rand, in South Africa [7]. In Krigeing, the function \( y(\cdot) \) is modeled as a Gaussian process \( Y(\cdot) \) with mean function \( m(\cdot) \) and covariance function \( k(\cdot, \cdot) \) [8]. More specifically, \( Y(\cdot) \) is written as
\[
Y(x) = f^T(x) b + Z(x)
\]
where \( f(x) \) is some known regression function vector (usually chosen constant or polynomial in \( x \)), \( b \) is a vector of unknown regression coefficients to be estimated, and \( Z(\cdot) \) is a zero-mean Gaussian process with known (or parametrized) covariance function \( k(\cdot, \cdot) \). Kriging is then the search for the best linear unbiased predictor (BLUP) of \( Y(\cdot) \) [9].

The actual covariance \( k(\cdot, \cdot) \) is usually unknown. It is expressed as
\[
k(Z(x_i), Z(x_j)) = \sigma^2_k R(x_i, x_j)
\]
where \( \sigma^2_k \) is the process variance and \( R(\cdot, \cdot) \) is a parametric correlation function. Both \( \sigma^2_k \) and the parameters of \( R(\cdot, \cdot) \) must be chosen or estimated from the available data. Under a stationarity assumption, \( R(x_i, x_j) \) depends only on the displacement vector \( x_i - x_j \), denoted by \( h \) in what follows. A frequent choice of correlation function, also adopted in the present paper, is the power exponential correlation function
\[
R(h) = \exp \left( -\sum_{k=1}^{d} \frac{|h_k|^{p_k}}{\theta_k} \right)
\]
where \( 0 < p_k \leq 2 \), and \( h_k \) is the \( k \)-th component of \( h \). Note that with this choice, \( R(h) \) tends to 1 when \( h \) tends to 0. The \( \theta_k \) may be estimated from the data by maximum likelihood, to get what is known as empirical Kriging (this setting has been used for the application reported in Section IV). A wide range of other choices for the correlation function is available [6].

Define \( R \) as the \( n \times n \) matrix such that
\[
R(i, j) = R(x_i, x_j)
\]
and \( r(x) \) as the \( n \) vector
\[
r(x) = [R(x, x_1), ..., R(x, x_n)]^T
\]
and \( F \) as the \( n \times \dim b \) matrix
\[
F = [f(x_1), ..., f(x_n)]^T
\]
In this presentation, we assume, for the sake of simplicity, that the parameters of the covariance matrix are known, but remember that in our application they are estimated by maximum likelihood. The maximum-likelihood estimate \( \hat{b} \) of the regression coefficients \( b \) from the available data \( \{x_n, y_n\} \) is
\[
\hat{b} = (F^T R^{-1} F)^{-1} F^T R^{-1} y_n
\]
The predictor of the mean of the Gaussian process, at \( x \in X \), is then given by
\[
\hat{Y}(x) = f^T(x) \hat{b} + r(x)^T R^{-1} (y_n - F \hat{b})
\]
This predictor is linear in \( y_n \) and interpolates the training data, as \( \hat{Y}(x_j) = y_j \). Another interesting property of Kriging, which is crucial regarding global search, is the possibility to compute the variance of the prediction error [10] at \( x \in X \) by
\[
\hat{\sigma}^2(x) = \sigma^2_k \left( 1 - r(x)^T R^{-1} r(x) \right)
\]

C. Maximizing Expected Improvement

The idea is to use the Kriging predictor \( \hat{Y} \) to find the \((n+1)\)-st point at which a simulation of the complete FDI process will be run. This point is chosen according to a criterion \( J(\cdot) \) that measures the interest of an additional evaluation at \( x \), given the past results \( y_n \) obtained at \( x_n \) and the Kriging prediction of the mean \( \hat{Y}(x) \) and variance \( \hat{\sigma}^2(x) \),
\[
x_{n+1} = \arg \max_{x \in X} J(x, x_n, y_n, \hat{Y}(x), \hat{\sigma}^2(x))
\]
A common choice for \( J(\cdot) \) is Expected Improvement [11]. The best available estimate of the minimum of \( y \) after the first \( n \) evaluations is \( y_{\min} = \min_{i=1,...,n} \{y_i = y(x_i)\} \). With
\[
u = \left( y_{\min} - \hat{Y}(x) \right) / \hat{\sigma}(x)
\]
the Expected Improvement is expressed in closed-form as
\[
\text{EI}(x) = \tilde{\sigma}(x)^2 \Phi(u) + \phi(u)
\]
where \( \Phi \) is the cumulative distribution function and \( \phi \) the probability density function of the normalized Gaussian distribution \( \mathcal{N}(0, 1) \). Maximizing Expected Improvement achieves a trade-off between local search (numerator of \( u \)) and the exploration of unknown areas (where \( \tilde{\sigma} \) is high) and is therefore well suited for global optimization.

D. EGO algorithm

The global optimization procedure that has been used for this study, based on the aforementioned elements, is called EGO, for efficient global optimization [12]. A preliminary sampling is required to obtain the \( n \) points of the initial design \( X_0 \). Latin Hypercube Sampling (LHS) has been chosen to explore \( X \) evenly [13]. The description of EGO is given in Algorithm 1. The algorithm stops either when the maximal number of iterations \( n_{\max} \) (which depends on the simulation budget) is reached or when the Expected Improvement becomes lower than some threshold \( \epsilon \). Our implementation is based on Sasena’s toolbox SuperEGO [14] and uses the DIRECT optimization algorithm [15] to achieve Step 5 of Algorithm 1.

III. ILLUSTRATIVE APPLICATION TO THE CHOICE OF A RESIDUAL-EVALUATION STRATEGY

This section presents the residual-analysis methods that will be tuned and compared, performance indices as goals for the optimization procedure and two classical test-cases. It should be noted that the methodology advocated in this paper can be applied to a much broader class of problems, and that the selection considered here is just for the purpose of illustration. Indeed, EGO is particularly well suited to problems where the evaluation of \( y \) is computationally expensive, as would be the case, for instance, when using cross-validation.
A. Strategies to be evaluated

A scalar residual \( r(t) \) is a signal that should remains negligible as long as there is no fault to which it is sensitive, and that becomes sufficiently large to be noticeable when a fault occurs. We consider residual-evaluation methods that provides a scalar binary decision function, which should return false if the residual is close enough to its initial mean (usually zero) and true if a jump or a drift occurs in the signal. The problem to be solved here is to detect a statistical change in the mean from its initial value zero to an unknown but different value.

Six candidate methods are to be tuned and compared by the proposed methodology. The operating principle of each of them is briefly recalled to highlight the hyperparameters involved, and references are given for further details. As the nominal mean \( \mu_0 \) and variance \( \sigma_0^2 \) of the signal are usually required, we estimate them on the first data for all methods and do not include them in the hyperparameters to be tuned.

1) The “three sigma” rule: This method proposes to choose bilateral fixed thresholds equal to \( \mu_0 \pm \nu \sigma_0 \), where \( \nu \geq 3 \) usually [16], relying on the fact that 99.7% of the points of a Gaussian distribution lie within three standard deviations. The decision takes the value true when the value of the residual falls outside the thresholds, else the decision is false.

2) Student’s t-test: This test checks whether the signal follows a Gaussian distribution \( N(\mu_0, \sigma_0) \), which leads to an automatic thresholding given by Student’s table considering the required confidence level is fixed here at 5% [17]. The test is applied to a sliding window of width \( N \).

3) Generalized Likelihood Ratio (GLR) test: This test is based on the likelihood ratio \( \Lambda(t) \) of the probability that the mean of \( r \) is \( \mu_1 \neq \mu_0 \) to the probability that it is \( \mu_0 \), still assuming that the signal is Gaussian [18], [19]. The generalized version uses the maximum-likelihood estimate \( \hat{\mu}_1 \) of \( \mu_1 \) to allow the detection of a change of unknown magnitude. The practical implementation using a sliding window of width \( N \) and the log-likelihood ratio is given by

\[
\sum_{t=1}^{N} r(t) > \frac{\sigma_0^2}{\mu_0 - \mu_1} \ln(\lambda) + \frac{N(\mu_0 - \hat{\mu}_1)}{2} \implies \text{decide true}
\]

\[
\text{else} \implies \text{decide false}
\]

where the threshold \( \lambda \) is one of the hyperparameters.

4) Sequential Probability Ratio Test (SPRT): The SPRT is very similar to the GLR, as it also uses the likelihood ratio on a sliding window of width \( N \). However, the minimum change detection size \( \mu_1 \) has to be specified, and the threshold \( \lambda \) is determined by the desired false-alarm and non-detection probabilities, respectively \( \alpha \) and \( \beta \) [19]. The following decisions are taken at each step:

\[
\begin{cases}
\Lambda < \frac{\beta}{1-\beta} 
\implies \text{decide false} \\
\Lambda > \frac{\alpha}{1-\alpha} 
\implies \text{decide true} \\
\text{else} 
\implies \text{take no decision}
\end{cases}
\]

5) CUSUM test: No statistical hypothesis is needed here. This two-sided test is expressed as follows [19][20]

\[
\begin{align*}
S_1(t) &= \max\{S_1(t-1) + r(t) - \mu_0 - \delta/2, 0\} \\
S_2(t) &= \max\{S_2(t-1) - r(t) + \mu_0 - \delta/2, 0\}
\end{align*}
\]

where \( \delta \) is the minimal size of the fault to be detected. The decision rule is then

\[
\begin{cases}
(S_1 > \lambda) \text{ or } (S_2 > \lambda) \implies \text{decide true} \\
\text{else} \implies \text{decide false}
\end{cases}
\]

where the threshold \( \lambda \) is one of the hyperparameters.

6) Randomised SubSampling (RSS): This very recent method, proposed in [21], uses \( M \) subsamplings of the signal on a sliding window of width \( N \). The sum of the errors with respect to the expected mean \( \mu_0 \) is computed on each subsample. The decision is false if at least \( q \) of the \( M \) sums are greater than zero and at least \( q \) of the \( M \) sums are smaller than zero, else the decision is true. An interesting property of the test is that the expected probability of false alarm is

\[
\alpha_{\exp} = 2q/M.
\]

Table I summarizes the hyperparameters involved in the methods considered.

<table>
<thead>
<tr>
<th>3-Sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>( N )</td>
<td>( N, \lambda )</td>
<td>( N, \mu_1, \alpha, \beta )</td>
<td>( \delta, \lambda )</td>
<td>( N, q, M )</td>
</tr>
</tbody>
</table>

B. Performance indices

We propose to use some of the quantitative indices defined within the DAMADICS benchmark [22]. Figure 1 shows time zones in the evolution of the Boolean decision function that are the basis of the definition of the performance indices. The value of the function before \( t_{\text{on}} \) and after \( t_{\text{on}} \) is not to be taken into account, while \( t_{\text{from}} \) is the instant at which the fault occurs. The indices that will be used for performance evaluation are

- the detection delay \( t_{\text{df}} \), which is the time elapsed between the fault occurrence time \( t_{\text{on}} \) and the last instant of time at which the decision signal switched from false to true;
They have been generated with a seed equal to 0, and the cost function defined in Section III-B is to minimize some weighted global performance indices from values considered by EGO. The simplest way to achieve thousands of evaluations, as evolutionary algorithms would typically require many more iterations, is to use the cost function of the global optimization problem that is to be solved, which is defined as:

\[ c = w_{td} r_{td} + w_{nd} r_{nd} + w_{fd} t_{fd} + w_{hor} t_{hor} \]

where the \( w_i \) are positive weights to be chosen. As the two indices \( r_{td} \) and \( r_{nd} \) take values in \([0; 1]\), the weights \( w_{td} \) and \( w_{nd} \) can be taken equal to 1, for an unweighted trade-off. The detection delay could also be included in the criterion, but should be normalized to match the range of the two other indices. Two continuous cost functions have been used in this study. The first one avoids the trade-off between false-detection and non-detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.

The feasible hyperparameter search spaces for all methods are indicated in Table II. Note that \( N, q \) and \( M \) are integers.

B. Results

The tuning results obtained on the two test-cases with the cost functions \( c_1 \) and \( c_2 \) for the candidate methods are presented in Tables III, IV, V and VI. The optimal values of the cost and the corresponding ranking of the methods are given, along with the values taken by the three performance indices from Section III-B and the corresponding hyperparameter tuning. Figures 3 and 4 show the decision functions corresponding to the best setting for each method on both test-cases. Explorations of the hyperparameter spaces (those with no more than two hyperparameters) by the global-optimization algorithm EGO are displayed on Figure 5. An acceptable tuning has been successfully found for each method, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

Even if these two test-cases are not sufficient to assess the absolute ability of these methods, some trends can be spotted. It appears that the 3-sigma method is not well suited to detect a change of the same order of magnitude as the standard deviation of the signal. Student’s test and the GLR test perform better if the Gaussian hypothesis stands true. The best results have been obtained with the SPR T test, the RSS approach and especially the CUSUM test. A possible explanation is that the latter two tests are not based on statistical hypothesis and only require the noise to be symmetrically distributed around the mean.

In the case of Student’s test and the GLR test, the detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.

The feasible hyperparameter search spaces for all methods are indicated in Table II. Note that \( N, q \) and \( M \) are integers.

B. Results

The tuning results obtained on the two test-cases with the cost functions \( c_1 \) and \( c_2 \) for the candidate methods are presented in Tables III, IV, V and VI. The optimal values of the cost and the corresponding ranking of the methods are given, along with the values taken by the three performance indices from Section III-B and the corresponding hyperparameter tuning. Figures 3 and 4 show the decision functions corresponding to the best setting for each method on both test-cases. Explorations of the hyperparameter spaces (those with no more than two hyperparameters) by the global-optimization algorithm EGO are displayed on Figure 5. An acceptable tuning has been successfully found for each method, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

Even if these two test-cases are not sufficient to assess the absolute ability of these methods, some trends can be spotted. It appears that the 3-sigma method is not well suited to detect a change of the same order of magnitude as the standard deviation of the signal. Student’s test and the GLR test perform better if the Gaussian hypothesis stands true. The best results have been obtained with the SPR T test, the RSS approach and especially the CUSUM test. A possible explanation is that the latter two tests are not based on statistical hypothesis and only require the noise to be symmetrically distributed around the mean.

The two criteria often (but not always) yield similar results. This is due to the complementary goal shared by the methods, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

In the case of Student’s test and the GLR test, the detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.

The feasible hyperparameter search spaces for all methods are indicated in Table II. Note that \( N, q \) and \( M \) are integers.

B. Results

The tuning results obtained on the two test-cases with the cost functions \( c_1 \) and \( c_2 \) for the candidate methods are presented in Tables III, IV, V and VI. The optimal values of the cost and the corresponding ranking of the methods are given, along with the values taken by the three performance indices from Section III-B and the corresponding hyperparameter tuning. Figures 3 and 4 show the decision functions corresponding to the best setting for each method on both test-cases. Explorations of the hyperparameter spaces (those with no more than two hyperparameters) by the global-optimization algorithm EGO are displayed on Figure 5. An acceptable tuning has been successfully found for each method, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

Even if these two test-cases are not sufficient to assess the absolute ability of these methods, some trends can be spotted. It appears that the 3-sigma method is not well suited to detect a change of the same order of magnitude as the standard deviation of the signal. Student’s test and the GLR test perform better if the Gaussian hypothesis stands true. The best results have been obtained with the SPR T test, the RSS approach and especially the CUSUM test. A possible explanation is that the latter two tests are not based on statistical hypothesis and only require the noise to be symmetrically distributed around the mean.

The two criteria often (but not always) yield similar results. This is due to the complementary goal shared by the methods, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

In the case of Student’s test and the GLR test, the detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.

The feasible hyperparameter search spaces for all methods are indicated in Table II. Note that \( N, q \) and \( M \) are integers.

B. Results

The tuning results obtained on the two test-cases with the cost functions \( c_1 \) and \( c_2 \) for the candidate methods are presented in Tables III, IV, V and VI. The optimal values of the cost and the corresponding ranking of the methods are given, along with the values taken by the three performance indices from Section III-B and the corresponding hyperparameter tuning. Figures 3 and 4 show the decision functions corresponding to the best setting for each method on both test-cases. Explorations of the hyperparameter spaces (those with no more than two hyperparameters) by the global-optimization algorithm EGO are displayed on Figure 5. An acceptable tuning has been successfully found for each method, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

Even if these two test-cases are not sufficient to assess the absolute ability of these methods, some trends can be spotted. It appears that the 3-sigma method is not well suited to detect a change of the same order of magnitude as the standard deviation of the signal. Student’s test and the GLR test perform better if the Gaussian hypothesis stands true. The best results have been obtained with the SPR T test, the RSS approach and especially the CUSUM test. A possible explanation is that the latter two tests are not based on statistical hypothesis and only require the noise to be symmetrically distributed around the mean.

The two criteria often (but not always) yield similar results. This is due to the complementary goal shared by the methods, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

In the case of Student’s test and the GLR test, the detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.

The feasible hyperparameter search spaces for all methods are indicated in Table II. Note that \( N, q \) and \( M \) are integers.

B. Results

The tuning results obtained on the two test-cases with the cost functions \( c_1 \) and \( c_2 \) for the candidate methods are presented in Tables III, IV, V and VI. The optimal values of the cost and the corresponding ranking of the methods are given, along with the values taken by the three performance indices from Section III-B and the corresponding hyperparameter tuning. Figures 3 and 4 show the decision functions corresponding to the best setting for each method on both test-cases. Explorations of the hyperparameter spaces (those with no more than two hyperparameters) by the global-optimization algorithm EGO are displayed on Figure 5. An acceptable tuning has been successfully found for each method, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

Even if these two test-cases are not sufficient to assess the absolute ability of these methods, some trends can be spotted. It appears that the 3-sigma method is not well suited to detect a change of the same order of magnitude as the standard deviation of the signal. Student’s test and the GLR test perform better if the Gaussian hypothesis stands true. The best results have been obtained with the SPR T test, the RSS approach and especially the CUSUM test. A possible explanation is that the latter two tests are not based on statistical hypothesis and only require the noise to be symmetrically distributed around the mean.

The two criteria often (but not always) yield similar results. This is due to the complementary goal shared by the methods, within \( n_{\text{max}} \) runs of the simulation. Although the examples treated here contain no more than four hyperparameters, nothing in the method forbids considering higher-dimensional problems.

In the case of Student’s test and the GLR test, the detection without taking explicitly delay into account, unlike the second one that also seeks for a reduced detection delay.
problem is available and that performance indices are computable. Kriging acts as a surrogate and simple-to-compute approximation of the complicated simulation leading to the evaluation of the performance indices. A global optimization procedure using the Kriging predictor then looks for the best real-valued hyperparameters.

The practicability of the methodology has been successfully illustrated through the selection of a residual-analysis strategy among various change-detection methods. Future work will address the evaluation of whole diagnosis strategies, comprising a residual generator coupled with an analysis algorithm on representative case-studies. These methods will necessarily imply more hyperparameters and the practical applicability of the method to larger dimensions will therefore be addressed. As a more general FDI case-study will involve model and measurement uncertainty, there is also the need to take into account environmental variables [6] (time of occurrence of faults, noise level, model uncertainty level...). Other multiobjective optimization techniques may also be investigated.

This paper employed the most classical method for Kriging-based global optimization, namely EGO. Alternative approaches, such as IAGO [23] could also be considered.

REFERENCES

TABLE II: Hyperparameter spaces for the candidate methods

<table>
<thead>
<tr>
<th>3-sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu \in [0.5; 10] )</td>
<td>( \lambda \in [50; 250] )</td>
<td>( \lambda \in [10; 150] )</td>
<td>( \mu_1 \in [0.1; 190] )</td>
<td>( \alpha \in [0.05; 0.2] )</td>
<td>( \beta \in [0.05; 0.2] )</td>
</tr>
</tbody>
</table>

TABLE III: Gaussian test-case with criterion \( c_1 \)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>3-sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best cost ( c_1 )</td>
<td>( \nu = 1.154 )</td>
<td>( N = 216 )</td>
<td>( N = 72 )</td>
<td>( \lambda = 1.679 )</td>
<td>( \mu_1 = 0.5296 )</td>
<td>( \alpha = 0.1017 )</td>
</tr>
</tbody>
</table>

TABLE IV: Uniform test-case with criterion \( c_1 \)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>3-sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best cost ( c_1 )</td>
<td>( \nu = 1.8304 )</td>
<td>( N = 55 )</td>
<td>( N = 20 )</td>
<td>( \lambda = 2.6364 )</td>
<td>( \mu_1 = 0.4484 )</td>
<td>( \alpha = 0.1248 )</td>
</tr>
</tbody>
</table>

TABLE V: Gaussian test-case with criterion \( c_2 \)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>3-sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best cost ( c_2 )</td>
<td>( \nu = 1.3445 )</td>
<td>( N = 217 )</td>
<td>( N = 72 )</td>
<td>( \lambda = 1.679 )</td>
<td>( \mu_1 = 0.5286 )</td>
<td>( \alpha = 0.1017 )</td>
</tr>
</tbody>
</table>

TABLE VI: Uniform test-case with criterion \( c_2 \)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>3-sigma</th>
<th>Student</th>
<th>GLR</th>
<th>SPRT</th>
<th>CUSUM</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best cost ( c_2 )</td>
<td>( \nu = 1.8478 )</td>
<td>( N = 55 )</td>
<td>( N = 25 )</td>
<td>( \lambda = 2.4794 )</td>
<td>( \mu_1 = 0.5537 )</td>
<td>( \alpha = 0.1583 )</td>
</tr>
</tbody>
</table>