OPTIMIZATION OF STATISTICAL METHODOLOGIES FOR ANOMALY DETECTION IN GAS TURBINE DYNAMIC TIME SERIES

Giuseppe Fabio Ceschini[1], Nicolò Gatta[2], Mauro Venturini[2], Thomas Hubauer[1], Alin Murarasu[1]

[1] Siemens AG, Nürnberg, Germany
[2] Dipartimento di Ingegneria, Università degli Studi di Ferrara, Ferrara, Italy

ABSTRACT

Statistical parametric methodologies are widely employed in the analysis of time series of gas turbine sensor readings. These methodologies identify outliers as a consequence of excessive deviation from a statistically-based model, derived from available observations. Among parametric techniques, the k-σ methodology demonstrates its effectiveness in the analysis of stationary time series. Furthermore, the simplicity and the clarity of this approach justify its direct application to industry. On the other hand, the k-σ methodology usually proves to be unable to adapt to dynamic time series, since it identifies observations in a transient as outliers.

As this limitation is caused by the nature of the methodology itself, two improved approaches are considered in this paper in addition to the standard k-σ methodology. The two proposed methodologies maintain the same rejection rule of the standard k-σ methodology, but differ in the portions of the time series from which statistical parameters (mean and standard deviation) are inferred. The first approach performs statistical inference by considering all observations prior to the current one, which are assumed reliable, plus a forward window containing a specified number of future observations. The second approach proposed in this paper is based on a moving window scheme.

Simulated data are used to tune the parameters of the proposed improved methodologies and to prove their effectiveness in adapting to dynamic time series. The moving window approach is found to be the best on simulated data in terms of True Positive Rate (*TPR*), False Negative Rate (*FNR*) and False Positive Rate (*FPR*). Therefore, the performance of the moving window approach is further assessed towards both different simulated scenarios and field data taken on a gas turbine.

NOMENCLATURE

k	acceptability threshold for the k-σ test criterion
N	number
$\bar{s}$	standard deviation
t	time
w	number of measurements in the window sample
x	measurement in the time series
$\bar{x}$	mean value

Subscripts and Superscripts

b	backward window
false	false
f	forward window
negative	negative
outlier	outlier
positive	positive
reliable	reliable
t	time point t
t+1	time point subsequent to the time point t
true	true

Acronyms

BFMW	backward and forward moving window
GT	gas turbine
FMW	forward moving window
FNR	false negative rate
FPR	false positive rate
O&M	Operations & Maintenance
SD	standard deviation
TPR	true positive rate

INTRODUCTION

Energy market demand sets high requirements to the productivity of gas turbine (GT) units, imposing high availability and efficiency level to achieve cost effectiveness. Furthermore, the complexity of the units implies a high level of insight on the health state and condition of the turbines. For these reasons, O&M companies continuously search for new methodologies to develop systems for monitoring, diagnostics and trending of the health state of the units installed worldwide. Efficient tools are expected to significantly improve availability, support condition-based maintenance and also enhance the reputation of O&M companies towards customers.

Several tools for monitoring, diagnostics and trending of the health state of GT units have been developed since the early '90s. The research in this field is still going on and, in fact, new tools have been developed achieving increasingly higher efficiency and reliability [1, 2, 3, 4]. Despite their different purpose, all these tools rely on direct measurements acquired by different types of sensors mounted in the GT unit. These thermodynamic (e.g. temperatures, pressures, etc.) and mechanical (e.g. vibrations, displacements, etc.) measurements are employed to estimate GT health indices to monitor its current state or to forecast operation ahead in time. A classic example is Gas Path Analysis [5, 6]. Tools based on Artificial Intelligence techniques can also be used for the same purpose [7, 8, 9].

However, the environment in which sensors operate is extremely harsh. High temperatures, enhanced fluid speed and dynamic operation of GT units are only a few of the many phenomena that can severely affect the quality of sensors data and cause hardware degradation and failures of measurement devices [10]. The use of low quality data in GT analysis tools leads to erroneous outputs, which consequently cause false positive calls. False positives constitute a severe issue for O&M companies as they lead to the deployment of maintenance operations not completely required, thus causing unnecessary downtime, extra costs and loss of reputation towards customers. Therefore, if input data quality is poor, health state analysis fails despite the capabilities of software tools employed [11].

For these reasons, the application of methodologies for raw data processing becomes fundamental to enhance input data quality by identifying and removing anomalous observations. In this way, only "reliable" sensor data serve as inputs for further processing, thus empowering the capabilities of the health state analysis [12, 13] and also for health state forecasting [14, 15, 16]. Such methodologies should be applied at an intermediate stage between sensor data collection (raw data) and data processing by means of software tools aimed at GT health analysis.

A fundamental aspect for the effectiveness of field application of raw data processing methodologies is the simplicity of tuning from the point of view of the user. Even if the detection capabilities are somewhat related to the number of model parameters, in most cases this is seen as a source of issues rather than a benefit. Considering their early stage application, straightforward methodologies are more desirable than complex but more refined techniques. This is the main reason why methodologies based on series of heuristic rules tend to fail. Furthermore, general guidelines for tuning should be available, in order to reduce data dependency and enlarge the application field for specific tuning setting. Another key aspect of a successful methodology for anomaly detection is the capability to reduce false positive calls, thus preventing reliable data suitable for further analysis from being discarded. Finally, even if the monitoring analysis usually focuses on steady states as these are the most meaningful for monitoring and diagnostics [11], raw data processing methodologies should be able to detect transient states without a priori flagging them as anomalies. For the reasons listed above, outlier identification constitutes a promising and challenging field for scientific research, especially regarding GT units [17, 18, 19, 20].

Outlier detection methodologies can be classified as parametrical (or statistical) or non-parametrical techniques. Parametrical methodologies perform data reliability assessment on the basis of statistical features of the time series, which are estimated from available data or assumed a priori [21]. These methodologies identify outliers as a consequence of excessive deviation from a statistically-based model derived from available observations, by means of direct statistical inference or derived from autoregressive models. The latter are usually combined with the maximization of likelihood functions [22]. Statistical techniques rely on the main assumption of identically and independently distributed data, whose validity is weak in case dynamic time series are analyzed [21, 23]. Data are usually assumed to be normally distributed, according to the central limit theorem [17]. Even if this hypothesis is well supported in case of GT sensor measurements, outliers certainly detach from such assumption [17].

Classical examples of statistical methodologies for outlier detection are the k-σ rule [24], the Hampel identifier [25], and the Kalman filter [26], which have constituted the starting point for more advanced methodologies. For example, Martin and Thompson [27] developed a modified version of the classic Kalman filter, called MT filter to smooth the effect of outliers in its prediction. A modified version of this algorithm was proposed by Liu et al. [28]. An alternative methodology for data filtering was applied to jet engine health signal by Ganguli [29], by adopting weighted Finite Impulse Response (FIR) median hybrid filters. A combination of Kalman filter and the autoregressive approach applied with a moving window (ARIMA) was proposed by Xu and Wojsznis [23] under the name of Time Series Kalman Filter (TSKF). Even if effective and easy to tune, this algorithm basically tightens the acceptability boundaries until a certain user-specified fraction of time series data are rejected. Therefore, in this case, the user should already be aware of the percentage of observations that will be identified as reliable. This solution can be ineffective in case of severely corrupted datasets. Yamanishi & Takeuchi [30] developed a methodology to build a statistical model of the time series data on the basis of the statistical inference on Sequentially Discounting Auto Regression model, which is able to account for the dynamic behavior of observations. A parametric test is used to rank the probability of observations to be outliers according to a logarithmic score. Other examples of scoring-based outlier detection methodologies were proposed by Takahashi & Tomioka [31] and Bhattacharya *et al.* [32].

However, techniques based on autoregressive models are applicable only to small datasets with a low number of outliers and guidelines for parameter tuning are not available for industrial applications [23]. Furthermore, simplicity of application and clarity are

two fundamental requirements for industrial application that are not often met by these methodologies. On the contrary, the k-σ methodology is considered of particular interest for its straightforwardness, low computational effort and efficiency on stationary data [24].

In its basic form, the k-σ methodology is known to perform poorly during transients, being not capable of adapting to set point changes [6, 24]. Therefore, this study aims at improving the capabilities of the k-σ methodology focusing on the analysis of steady states in dynamic time series. Even though its performance during a transient will be assessed, the main goal is the evaluation of outlier identification capability at steady state conditions according to industry practice. Moreover, this paper aims at setting as a reference in technical literature regarding guidelines for tuning the k-σ methodology parameters for its field application. In fact, tuning guidelines are available in literature [24, 33, 34], but they may result unsuitable for field application.

This study is part of a more comprehensive research activity developed by the authors in [35, 36]. In [35], the benefits of implementing robust statistical estimators are evaluated, by considering three different approaches. Different scenarios are considered to evaluate statistical efficiency, resistance and robustness, by injecting outliers in field data sets. In [36], a comprehensive and straightforward tool for Detection, Classification and Integrated Diagnostics of Gas Turbine Sensors (named DCIDS) is proposed. The tool consists of two main algorithms, i.e. the Anomaly Detection Algorithm and the Anomaly Classification Algorithm. The performance of the tool is assessed by analyzing different types of field measurements taken on several gas turbines in operation.

This study also documents a part of the efforts made by Siemens to continuously improve automated data processing and sensor fault isolation. In fact, nowadays, a large amount of data is collected on a daily basis and the trend is to increase even further. In such a context, analyses that involve frequent human decisions are both error-prone and highly impractical. Therefore, it is of great importance for Siemens to provide automated analyses that are both effective and efficient with regard to the detection of problems.

This paper is organized as follows. The first section describes the features of the standard k-σ approach. Then, two improved applications of the same acceptability rule are proposed. Subsequently, directions for optimal tuning are derived by means of simulated data. Finally, the capabilities of the best performing technique are further investigated by means of field data with injected outliers.

k-σ METHODOLOGY FOR OUTLIER DETECTION

The k-σ methodology processes sensor observations according to a parametric test. The test is performed by inferring sample location and dispersion characteristics, i.e. mean and standard deviation, and comparing the result to a user-specified scalar threshold k. Considering a generic time point t in the time series, the relative sensor observation x_t is considered reliable if the following test criterion is verified:

$$\frac{\left|x_t - \bar{x}\right|}{\bar{s}} < k_b \qquad (1)$$

where $\bar{x}$ and $\bar{s}$ are the mean and standard deviation respectively, calculated on all observations assessed as reliable prior to the one under evaluation (i.e. x_t). Therefore, if considered reliable, the current observation is added to the pool for statistical inference employed in the statistical test performed for x_{t+1}. On the contrary, if x_t is not considered reliable, the observation is flagged as anomalous and neglected while applying the statistical inference test criterion to x_{t+1}. It can also be noticed that, if $\bar{s} = 0$ (i.e. all observations in the sample are equal), the left side of the test criterion in Eq. (1) returns an undetermined value. Therefore, an infinitesimal quantity can be added to the denominator to preserve the mathematical meaning of the test criterion in case $\bar{s} = 0$.

In addition to the straightforwardness and promptness in the calculation of the test parameters, the most important feature of the k-σ methodology is the unambiguous value of the test criterion and, therefore, of the detection outcome. Furthermore, unlike other methodologies that apply a series of heuristic rules to identify outliers, the only parameter requested by the k-σ methodology is the scalar value k. From the test criterion in Eq. (1), it can be noted that this scalar value determines the width of the acceptability boundary expressed in terms of number of standard deviations. However, even if a sole tuning parameter is required, it can be difficult for a field user to assess the value of k, which is clearly influenced by the characteristics of the time series. Suggestions from literature identify the proper range for k being between 3 [33] and 2 [34], which means a less or more selective methodology, respectively. These indications can result vague from a user point of view, especially if dealing with data from different sources.

In the form reported in the test criterion in Eq. (1), the methodology proved to be ineffective towards the analysis of dynamic time series [6]. This poses a considerable limit for GT sensor data analysis. The source of this limitation lies in the structure of the methodology itself. Namely, as the transient proceeds, new measurements gradually detach from the mean value of previous stationary points. Thus, they approach the acceptability boundaries until they are flagged as outliers and consequently rejected. The model does not consider these measurements while inferring the mean value and cannot adapt the acceptability boundaries.

A sample application of the standard k-σ methodology towards dynamic data is reported in Figure 1. In this case, normalized field data of vibration V1, taken on a Siemens gas turbine, are processed by using the test criterion in Eq. (1) with k=3 (value chosen according to literature guidelines). As it can be seen, the algorithm is not able to adapt to set point changes, i.e. it is not suitable for dynamic data. Therefore, alternatives for the application of the k-σ methodology are proposed in this paper to manage transient data as well.
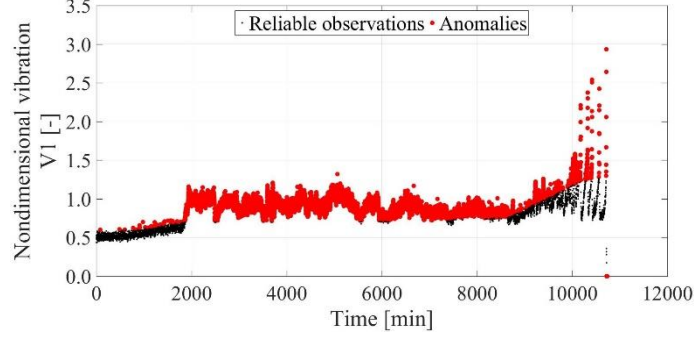
Figure 1 – Field dataset of nondimensional vibration V1

IMPROVED k-σ METHODOLOGIES

The incapability of the k-σ methodology to manage GT set-point changes in the time series derives from the fact that the acceptability test does not consider any observations ahead of the one currently under assessment.

This does not necessary imply that the methodology needs to be applied completely offline. As a certain portion of ahead-in-time observations are considered, the k-σ can be applied with a delay inversely proportional to the sensor sampling frequency. The future observations window is in fact defined in terms of number of data for statistical inference and not in terms of time interval. To provide the test criterion with an insight on the statistics of future observations, a moving window approach is applied in two different forms.

Forward Moving Window (FMW). In this case, the k-σ test criterion is applied in its base form (i.e. applied to *all* past observations) together with a test criterion applied to a forward window including a number w_f of observations taken at subsequent time points. The window moves ahead in the time series and maintains its size w_f fixed as future observations are processed.

The acceptability criterion for considering an observation x_t reliable according to the FMW methodology can be written as:

$$\frac{|x_t - \bar{x}|}{\bar{s}} < k_b \ \ \text{OR} \ \ \frac{|x_t - \bar{x}_f|}{\bar{s}_f} < k_f \tag{2}$$

Similarly to the k-σ approach, observations that are considered reliable are added to the statistical inference pool for the application of the test criterion, while anomalous measurements are neglected.

Backward and Forward Moving Window (BFMW). The k-σ test criterion is applied to a moving window of fixed size defined by w_b past observations and w_f future observations with respect to the current observation. The window moves forward as a new observation is processed, including reliable observations in the w_b past window sample while neglecting anomalies. A generic observation x_t is considered reliable if the following rule is verified:

$$\frac{|x_t - \bar{x}_b|}{\bar{s}_b} < k_b \ \ \text{OR} \ \ \frac{|x_t - \bar{x}_f|}{\bar{s}_f} < k_f \tag{3}$$

Unlike the basic form the k-σ methodology, FMW and BFMW prove to be able to handle transient data. Figure 2 reports FMW and BFMW performance towards vibration V1 measurements, which demonstrated to be troublesome for the standard k-σ methodology in Figure 1. Differently from the standard k-σ methodology, both FMW and BFMW identify outliers which seem in agreement with engineering judgment.

As a further example, Figure 3 shows the performance of FMW and BFMW techniques towards more severe transient maneuvers. The normalized experimental data in Figure 3 refer to a temperature, referred as T1, that ramps down up to 20% of its steady-state value. Some outliers are identified during the transients at t=1800 s and t=2250 s, but most of these data are considered reliable, even though the mean value in the test criterion dramatically changes.

Observations lying in the last w_f positions of the time series cannot be processed as no more future observations are available to apply the test criterion. Assessing the reliability of these measurements only on the basis of past observations would not be meaningful and the same consideration holds if a gradually reducing forward window is applied. In fact, as the end of the time series approaches, the forward sample for statistical inference is reduced, together with its statistical significance. For these reasons, the final w_f observations are marked as "unprocessed". The same criterion is applied to the initial observations in the backward window w_b.

In order to derive tuning guidelines for the FMW and BFMW user-specified parameters (w_b, k_b, w_f, k_f), quantitative measures of methodologies performance are required. For this purpose, simulated data are employed to produce time series with injected outliers in known positions, to assess the detection capabilities of the different tunings.
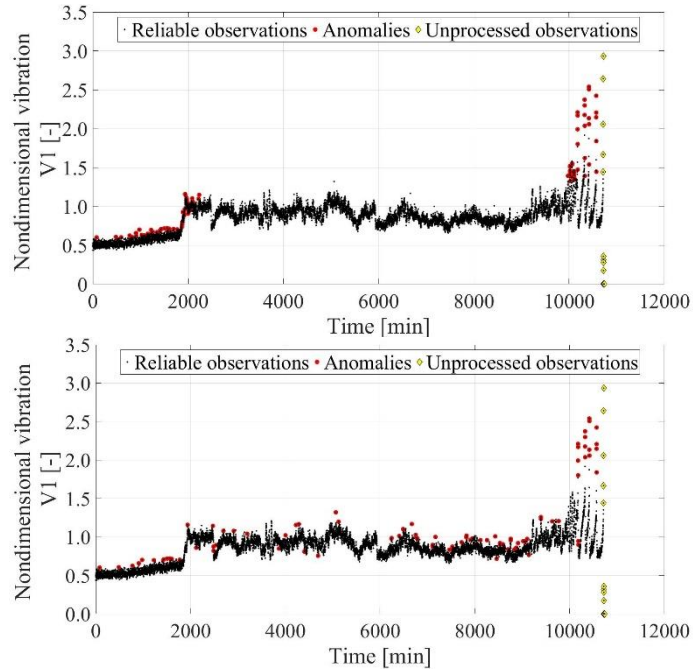
GTP-17-1359 ; Venturini

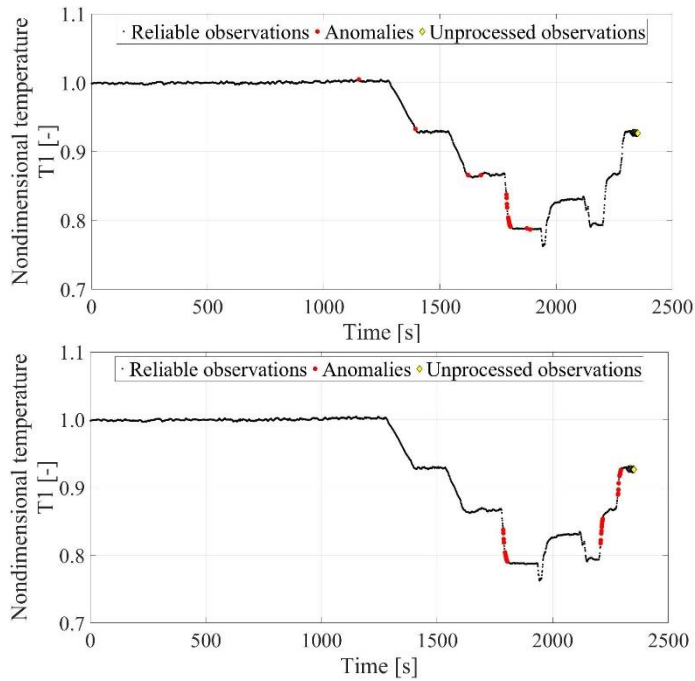**Figure 2** – FMW (above) and BFMW (below) methodology applied to the field dataset of nondimensional vibration V1

Figure 3 – FMW (above) and BFMW (below) applied to the field dataset of nondimensional temperature T1

TUNING OF IMPROVED k-σ METHODOLOGIES

The simulated time series used for tuning the improved k-σ methodologies are composed of four segments, i.e. an initial outlier-free portion, a steady state maneuver followed by a transient maneuver to reach a new steady state value, followed by an additional steady state maneuver. The simulated time series is generated according to the following pattern:

- SS0 = initial outlier-free steady state @ value 1.0;
- SSA = steady state A @ value 1.0;
- TSB = transient B with a step of 10%;

- SSC = steady state C @ value 1.1.

A sample of generated data is provided in Figure 4, with controlled outliers at 7% magnitude. In order to consider different scenarios, different gradients are evaluated (30°, 45°, 60° and 90°). A gradient of 90° with respect to the mean value simulates a step variation. In the following, the results will be reported for the 45° gradient only. Outliers are inserted with a user specified magnitude (±3%, ±4%, ±5% and ±7% with respect to the mean value) at randomly selected time points for each segment SSA, TSB and SSC. Therefore, the number of outliers (5% of the data) and their time coordinate are known and can be used to evaluate the performance of the improved k-σ methodologies. Furthermore, Gaussian noise (1% or 2%) is added to simulate field observations. The noise intensity is expressed as the percentage ratio of SD over the mean value of the steady state (i.e. 1 for SSA and 1.1 for SSC).
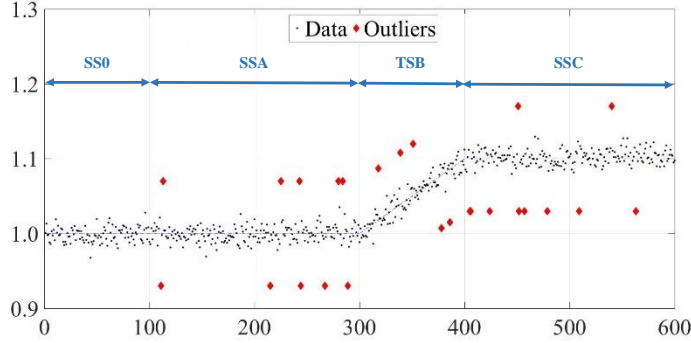

Figure 4 – Simulated data with noise 1%, 10% step, 7% outlier magnitude

Three widely-used performance indices, derived from statistical test theory [37], are considered for the quantitative assessment of the different methodologies.

a) *True Positive Rate (TPR)* i.e. percentage ratio between the true positives identified by the algorithm and the number of total observations flagged as anomalies by the algorithm, i.e. the sum of true positives and false positives:

$$TPR = \frac{N_{\text{true positive}}}{N_{\text{true positive}} + N_{\text{false positive}}} \qquad (4)$$

b) *False Negative Rate (FNR)* i.e. percentage ratio between the number of anomalous data incorrectly flagged as reliable and the number of true outliers

$$FNR = \frac{N_{\text{false negative}}}{N_{\text{true outlier}}} = 1 - \frac{N_{\text{true positive}}}{N_{\text{true outlier}}} \qquad (5)$$

c) *False Positive Rate (FPR)* i.e. percentage ratio between the number of false positive calls and the number of true reliable data

$$FPR = \frac{N_{\text{false positive}}}{N_{\text{true reliable}}} \qquad (6)$$

In order to provide results with statistical meaningfulness, the value of *TPR*, *FNR* and *FPR* for each configuration is averaged over 10^3 simulated time series, similarly to Monte Carlo simulation approach.

Results of the tuning for the parameters of the FMW methodologies and BFMW are provided in the following. The best performing methodology combines the maximum *TPR* with the minimum *FNR* and minimum *FPR*. The range values of the k_b and k_f parameters were derived from literature [33, 34], while the size of the moving windows w_b and w_f was derived from engineering judgment and statistical considerations on sample meaningfulness. In fact, while the k parameters influence the acceptability boundaries, w parameters adjust the sensitivity of the methodologies towards dynamic data behavior. According to the results presented in the following, the most influencing parameters on *TPR* and *FPR* performance prove to be the backward and forward acceptability thresholds k_b and k_f, respectively.

Table 1 reports the different settings evaluated for these parameters. In compliance with the guidelines available in technical literature, the optimization space for the acceptability thresholds k_b and k_f ranges between 2 and 3. In first instance, the sizes of the backward and forward moving windows are maintained constant at 100 and 25 observations, respectively.

The different configurations are tested by using FMW and BFMW schemes towards different gradient scenarios. Namely, 30°, 45°, 60° and 90° (step variation) slopes are considered. Only the results relative to the 45° case are reported in the following subsections for the sake of conciseness. However, it is worth noting that best and worst settings do not change with the gradients, implying the validity of results towards different scenarios. The best configuration is the one that achieves the best balance between high *TPR* and low *FNR*.

In fact, in the following, the results for *FPR* are not reported since *FPR* resulted lower than approximately 0.1% in all the analyzed cases.

Table 1 – Parameter settings analyzed to identify the optimal tuning of the FMW and BFMW schemes

w_b	k_b	w_f	k_f
100	2	25	2
100	2	25	3
100	3	25	2
100	3	25	3

Forward Moving Window (FMW). The results of the tuning procedure for the FMW scheme are reported in Table 2. The results of the performance comparison between the best and the worst tuning are presented in Figure 5. The *TPR* performance is similar for the two configurations, being both higher than 93% as the outlier magnitude ranges from 4% to 7%. A sensible difference is noticed when the magnitude sets at 3%, with the best configuration achieving a *TPR* of 75%, compared to the 32% of the worst performing configuration. However, the most remarkable difference between the two configurations occurs in terms of FNR performance. Namely, the FNR obtained by the best configuration is approximately 20% lower than that achieved by the worst configuration.

Table 2 – Tuning of the FMW methodology

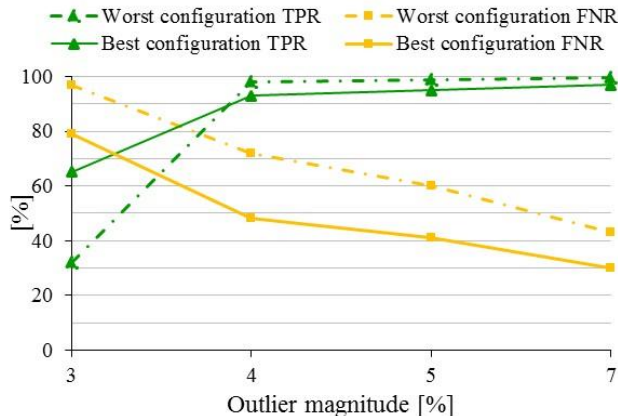|	w_b	k_b	w_f	k_f
Best	100	3	25	2
Worst	100	3	25	3

Figure 5 – Performance comparison between best and worst tunings of the FMW scheme for different outlier magnitudes

Backward and Forward Moving Window (BFMW). The results of the tuning procedure for the BFMW scheme are reported in Table 3. The performance achieved by the BFMW methodology with its best and worst configuration is presented in Figure 6. The *TPR* experiences a sensible benefit from the adoption of the best performing tuning as the outlier magnitude reaches 3%. In this scenario, the *TPR* of the BFMW with the best configuration achieves 62% instead of 20% of the worst performing configuration. Similarly to the results of the FMW, the *TPR* achieved by the worst and best configurations are close to each other for the remaining magnitude scenarios. The *FNR* results are remarkably improved by the adoption of the best performing configuration, with an average decrease of false negatives by 20%.

The at-a-glance comparison between the performance achieved by the FMW and BFMW methodologies with their best configuration is reported in Figure 7. Results in terms of *TPR* are similar for the two methodologies, with the BFMW achieving slightly better performance. However, it should be noticed that the benefit of using the BFMW lies in reducing the *FNR* by an average 15% with respect to the FMW methodology. For these reasons, the BFMW methodology univocally results preferable.

Table 3 – Tuning of the BFMW methodology

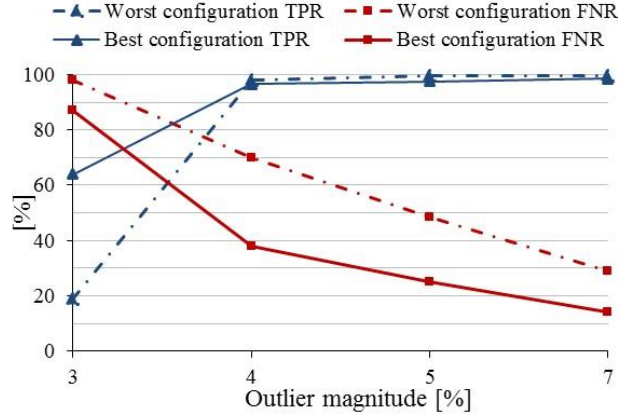|	w_b	k_b	w_f	k_f
Best	100	3	25	2
Worst	100	3	25	3

Figure 6 – Performance comparison between best and worst tunings of the BFMW scheme for different outlier magnitudes
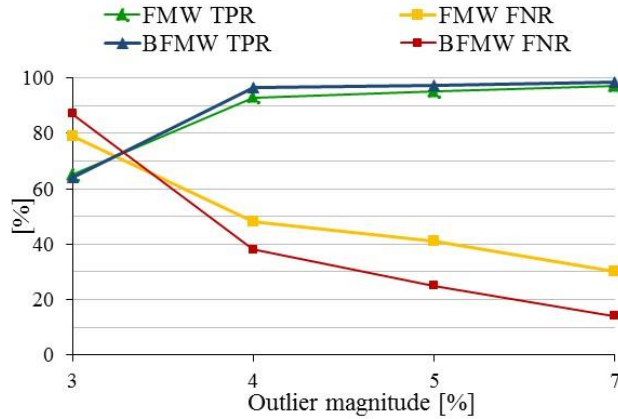

Figure 7 – Performance comparison between FMW and BFMW methodologies with best performing tuning

Optimal tuning of the BFMW methodology. As the BFMW methodology proved to be the most effective scheme, the influence of the backward window size is further investigated. Results show that the reduction of the backward window size from 100 to 50 observations enhances the detection capability of the algorithm in case of transient data.

The setting reported in Table 4 combines the high *TPR* of the strict configuration with k_b =2 and k_f =2, with the limited *FNR* of the setting with k_b =3 and k_f =2. Figure 8 provides a comparison between performance achieved with w_b=100, w_b=50 and w_b=25. As it can be seen, *TPR* results improve by 10% with respect to the optimal tuning reported in Table 3 for all considered magnitude scenarios. The performance achieved by the w_b=25 setting is similar to that obtained with w_b=50, but with a slightly lower *TPR*.

The influence of k_b and k_f values on the performance of the BFMW methodology is also further investigated All combinations of k_b and k_f in the range 1-4 are considered, by assuming w_b = 50 and w_f = 25 according to Table 4.

Table 4 – Optimal tuning of the BFMW methodology

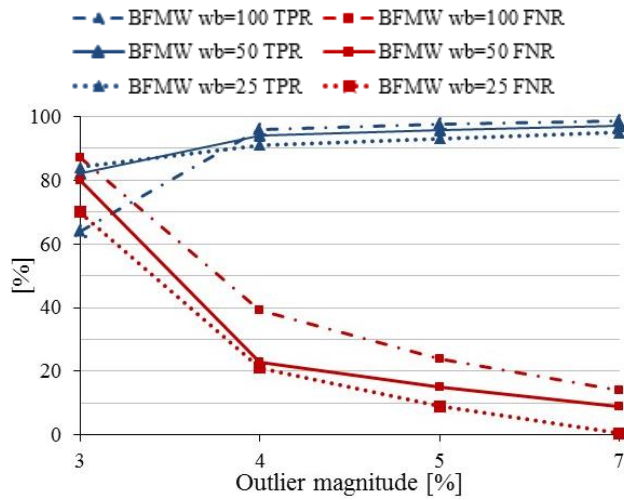| w_b	k_b	w_f	k_f
50	3	25	2

GTP-17-1359 ; Venturini

Figure 8 – Influence of backward window size for different outlier magnitudes

Two outlier magnitudes are investigated, i.e. 7% and 3%, since they are representative of the two extreme cases considered in this paper. The results are shown in Figure 9 and 10 for *TPR* and *FNR*. In general, all settings with k_b or k_f equal to 1 or 4 are characterized by poor performance, i.e. both *TPR* and *FNR* are high or both of them are low, mainly for outlier magnitude equal to 3%. Instead, the preferable combinations include k_b or k_f equal to 2 or 3, according to literature guidelines [33, 34].

The setting with $k_b = 3$ and $k_f = 2$ proves to be the best if outlier magnitude is 7%, since the corresponding value of *TPR* is the highest (97.0%) and FNR is the lowest (8.1%).
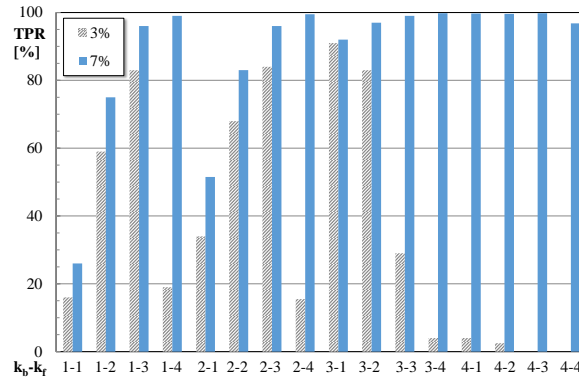

Figure 9 – *TPR* performance of BFMW methodology as a function of k_b and k_f values (outlier magnitude at 3% and 7%)
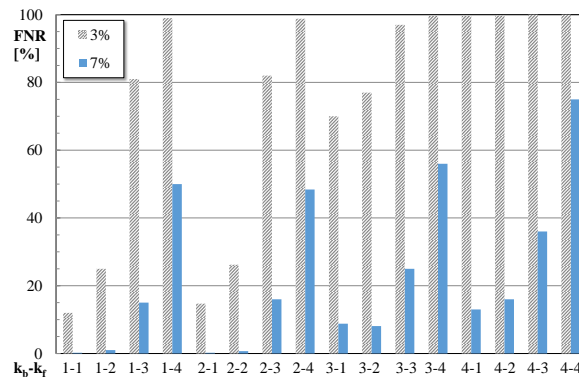

Figure 10 – *FNR* performance of BFMW methodology as a function of k_b and k_f values (outlier magnitude at 3% and 7%)

GTP-17-1359 ; Venturini

Instead, if outlier magnitude is 3%, the best setting is $k_b = 2$ and $k_f = 2$, which allows *TPR* equal to 68.0% and *FNR* equal to 26.2%. However, the setting $k_b = 3$ and $k_f = 2$ still allows good performance, with even higher *TPR* (83.0%), but considerably higher *FNR* (77.0%).

Therefore, the setting $k_b = 3$ and $k_f = 2$ is confirmed as the optimal compromise for outliers with different magnitudes, mainly with the aim of achieving high *TPR* values.

In conclusion, the optimal setting of the BFMW methodology is identified, as reported in Table 4. This configuration is maintained for the subsequent analysis reported in the paper, where the capabilities of the BFMW methodology are assessed towards different simulation scenarios and also field data with injected outliers.

ASSESSMENT OF BFMW k-σ METHODOLOGY

Influence of increased noise. The performance of the moving window algorithm is investigated by considering a doubled noise value (2%). Tests are replicated according to the previously followed procedural pattern. The results presented in Figure 11 show that an increase of noise does not severely affect the detection capabilities of the methodology in the 7% outlier magnitude scenario. In this case, the *TPR* sets at 95% with the *FNR* at 38%. The negative effect of enhanced data dispersion is more evident as the outlier magnitude decreases. In fact, starting from 5% magnitude, the *TPR* drops at 36%, while the *FNR* sets close to 100%, thus proving the difficulties encountered in outlier detection due to enhanced noise. These are mainly caused by the increased value of the standard deviation both in the backward and forward window, which widens the acceptability boundaries of the k-σ test criterion. However, a noise level of 2% affecting all the data in the sample is quite unusual and was selected in this paper with the only aim of challenging the BFMW k-σ methodology.
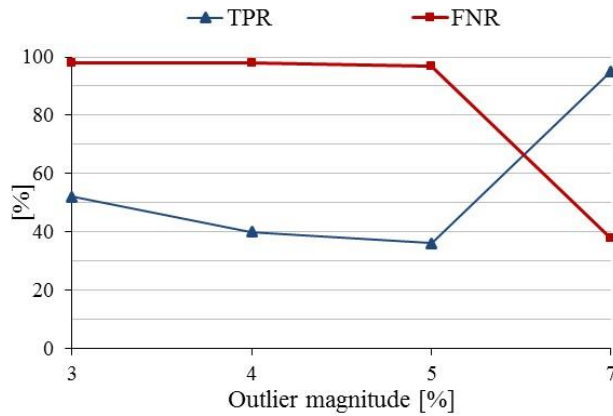

Figure 11 – BFMW methodology performance for different outlier magnitudes at 2% noise

Influence of the step change during the transient. The performance of the BFMW methodology is further investigated by varying the step magnitude of the transient TSB. Therefore, in addition to the +10% step considered so far, +20%, +50% and +100% step changes are also analyzed by using different combinations of k_b and k_f values.

Once again, the best performing configuration is characterized by $k_b = 3$ and $k_f = 2$. Results regarding *TPR* and *FNR* as a function of outlier magnitude and step change during the transient are presented in Figure 12. As it can be noticed, the *FNR* increases as the step change increases. This is mainly due to a masking effect of the BFMW methodology, which can be appreciated from Figure 13. The methodology tends to create a masking effect, making outliers in the lower part of the transient very unlikely to be detected. This phenomenon is embedded in the k-σ methodology itself, as the outliers of the non-detection zone are marked as reliable due to the mean value of the backward acceptability rule. Symmetrical results were obtained by imposing a negative slope to the transient.

As the duration of the transient increases (together with the step magnitude in the 45° case), a higher number of outliers is likely to lie in the non-detection side, thus increasing the *FNR*. The effect of the masking effect in detection is further investigated in the following, by acknowledging the difference in performance when outliers are inserted in the transient only or in steady state only.
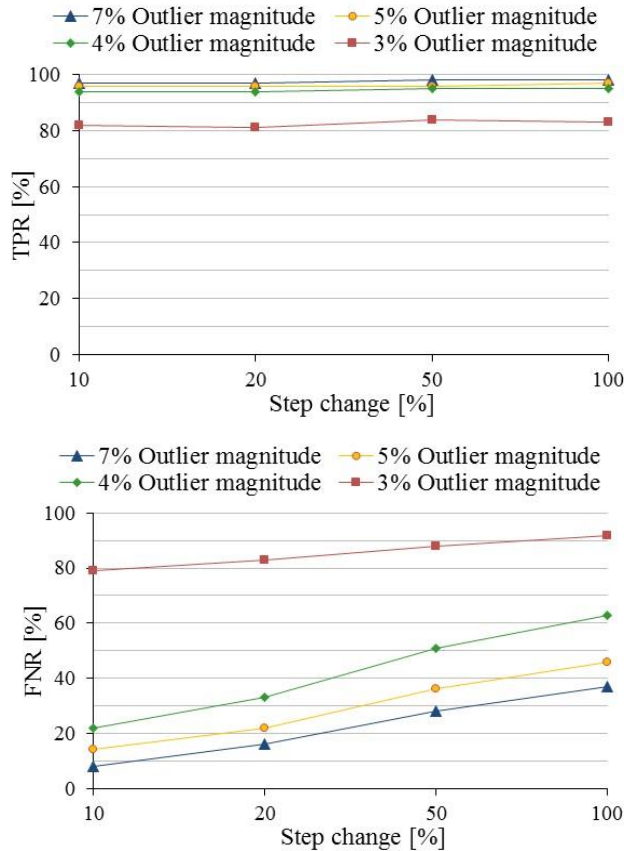
Figure 12 – BFMW methodology *TPR* (above) and *FNR* (below) for different outlier magnitudes and different step changes
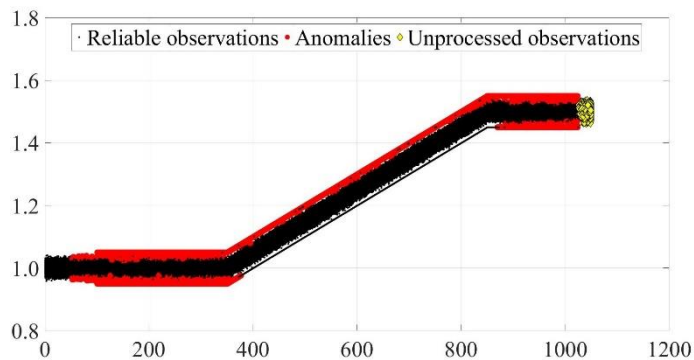

Figure 13 – Example of the masking effect for BFMW methodology

Influence of outliers position in the time series. In order to highlight how the masking effect in the transient TSB affects detection performance, outliers are placed only in particular portions of the time series. This is made by considering two likely operational scenarios, i.e. outliers in the transient TSB only (e.g. due to incorrect measurements during GT set point change) or outliers in the second steady state SSC only (e.g. due to sensors not promptly adapting to a stationary state after a transient).

Results can be analyzed from two different points of view, i.e. by evaluating the dependency of *TPR* and *FNR* towards (i) outlier magnitude at fixed step (10%) and (ii) as a function of the step magnitude and constant outlier magnitude (7%).

As shown in Figure 14, the presence of outliers in the transient affects performance almost linearly, as outlier magnitude decreases. However, the methodology successfully identifies outliers in SSC with TPR higher than 85% and FNR lower than 14% until a 4% magnitude is reached.

GTP-17-1359 ; Venturini

Figure 15 presents *TPR* and *FPR* as a function of the step magnitude and constant 7% outlier magnitude. Results prove the reliability of the methodology in outlier detection even after a severe transient (+100%), reaching a *TPR* of 85% with 3% of *FNR* in the most successful case. The influence of the masking effect becomes more evident as the transient duration increases.
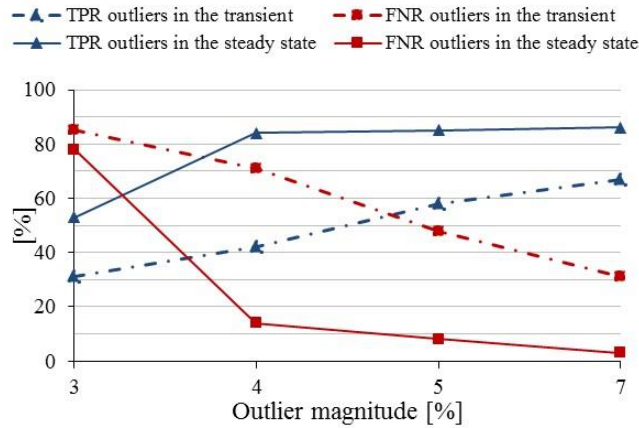

Figure 14 – BFMW methodology performance vs. outlier magnitude, with outliers in TSB only (solid line) or SSC only (dashed line)
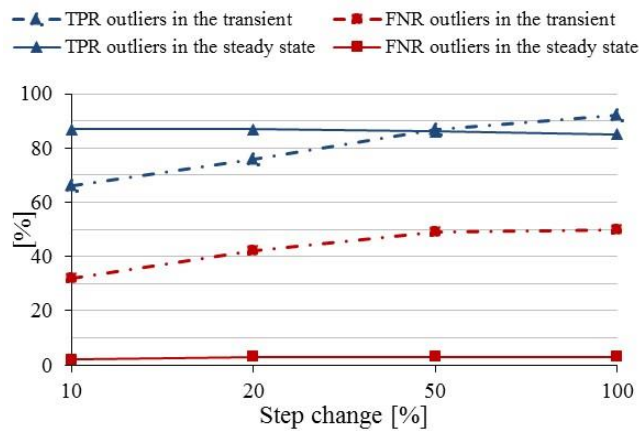

Figure 15 – BFMW methodology performance vs. step change, with outliers in TSB only (solid line) or SSC only (dashed line)

Discussion. Figure 16 gathers *TPR* and *FPR* performance achieved by the BFMW methodology for all the simulated scenarios. BFMW parameters are set according to the optimal values reported in Table 4.

In general, the performance of the BFMW methodology proves to be satisfactory, combining good *TPR* with consistently low *FNR*. However, it can be noticed that performance decreases as the outliers magnitude reaches 3%, which represents a rather challenging scenario considering the structure of simulated data. In fact, being data distributed with 1% Gaussian noise (i.e. SD = 0.01) the probability for them to achieve a value between ± 0.01 and ± 0.03 from the mean value is 31%. In this scenario, the difference with outliers imposed at ± 0.03 becomes consistently narrow, questioning the fact that they are actually outliers and consequently affecting *TPR* and *FNR* values.

Furthermore, outliers with such a reduced magnitude are not frequently observed in field applications. For these reasons, the 3% magnitude case is to be mainly considered for its speculative and challenging meaningfulness.

For a given step magnitude, the decrease of the outlier magnitude from 7% to 4% affects *TPR* performance almost linearly, while causing a more evident increase in the *FNR* values.

Figure 16 also highlights the influence of the masking effect on methodology performance. Namely, it can be noticed that *TPR* values decrease as the step magnitude (together with the duration of the transient) increases. This result has to be interpreted in association with the results reported in Figure 15, which demonstrated that the detection rate of outliers which occur within the transient decreases by 35% as the step magnitude passes from 10% to 100%. Therefore, the decrease in performance illustrated in Figure 16 is due to the fact that i) outliers in the transient are more difficult to detect as the step magnitude increases and ii) outliers are more likely placed in simulated data during the transient as the transient duration increases with respect to the TSA and TSC.

However, as proved by the results in Figure 15, the methodology successfully identifies steady state outliers even after a 100% step change. This is a remarkable achievement considering that outlier identification during a transient was not the primary objective of this study and is not the main goal in many field applications. In fact, the reliable assessment of power plant performance is usually required at steady state conditions. However, the methodology proves to be effective in managing dynamic time series, as it does not identify the observations in the transient as anomalous, despite the step magnitude, as proved by the low values achieved by the *FNR*.
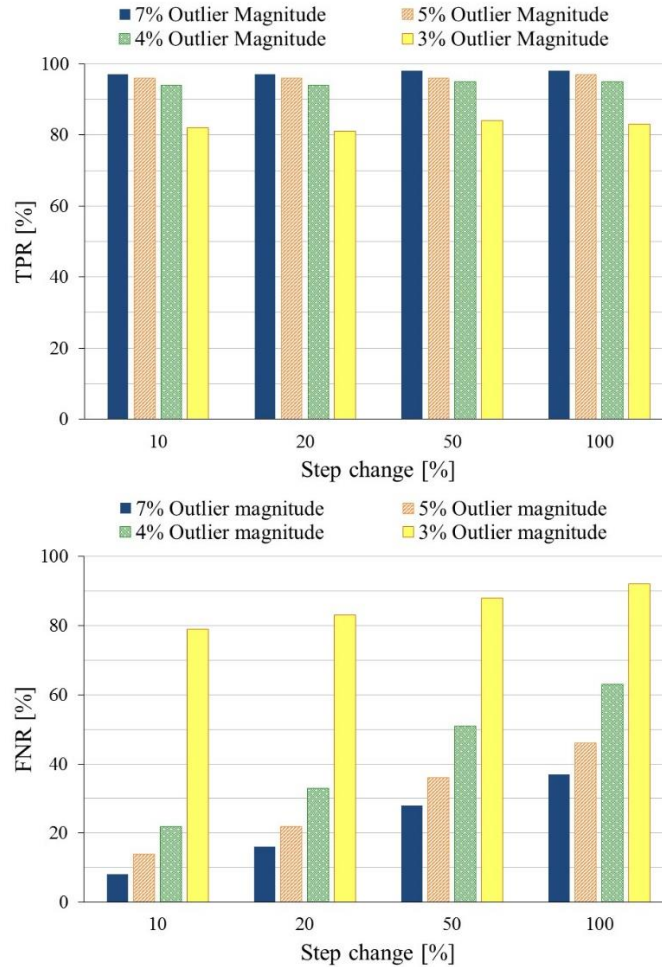

Figure 16 – Overview of BFMW methodology *TPR* and *FNR* performance for different simulated scenarios

ASSESSMENT OF BFMW k-σ METHODOLOGY BY MEANS OF FIELD DATA WITH INJECTED OUTLIERS

To further test the capabilities of the BFMW methodology, field data are poisoned by injected outliers of controlled magnitude, according to the approach adopted e.g. in [38]. As the position of anomalies is known, performance indices can be calculated in a less controlled operational environment with respect to purely simulated data.

The outliers magnitude is expressed in terms of multiples m of the noise factor s, i.e. the SD of each steady state in the time series. Therefore, if an observation x_t is selected to be replaced with an outlier, its value is calculated according to the following expression:

$$x_{t\,outlier} = x_t + ms \qquad (7)$$

The dataset selected for outlier injection contains nondimensional temperature measurements, here referred as dataset T2. A portion of the same time series is analyzed in [35], while the whole time series, together with the readings of the sensors set of the same quantity, is analyzed in [36]. This case study is selected since, as it can be seen in Figure 17, a severe transient occurs at $t = 800$ min. The reliability of the results obtained from the analysis of field data with injected outliers relies on the assumption of absence of previous anomalies in the dataset. To this purpose, the temperature T2 dataset was previously analyzed by means of visual inspections and application of engineering sense. As few isolated spikes occurred in the second steady state, these were replaced with observations equal to the mean value of the steady state itself.

Results are reported in Figure 18 in terms of *TPR* and *FNR*. The BFMW methodology achieves *TPR* values between 60% and 70% as the outlier magnitude ranges from 3% to 7%, while the values of *FNR* are rather low, ranging from 8% to 1% in the same outlier magnitude interval. The results on field data highlight a low sensitivity of the BFMW methodology towards outlier magnitude variation. This is actually an encouraging result, considering the numerical values of the performance indices.

Even if lower than the values observed in simulated data, the *TPR* still ranges between 60% and 70%; this fact, in combination with the extremely reduced values of the *FNR*, is an index of proficient outlier detection skills. Furthermore, the low sensitivity of the indices towards outliers magnitude indicates that this proficiency is not affected by the outliers which are closer to the acceptability boundaries.

The reason behind the difference between *TPR* values obtained on simulated and field data lies in the increase of false positive calls and, consequently, of the *FPR*. The rapid and wide step change of the transient, combined with the low sampling frequency (1 minute), creates a consistent gap in subsequent measurements, thus increasing the probability for observations to trespass the acceptability thresholds imposed by the parametric test. This leads to the increase of false positive calls and consequently, according to Eq. (4), to the decrease of *TPR*. Despite these false positive calls, the shape of the transient is preserved, thus not affecting the information provided by the observations in the time series. Therefore, the methodology proves its capability in outlier detection, also in case of dynamic time series.
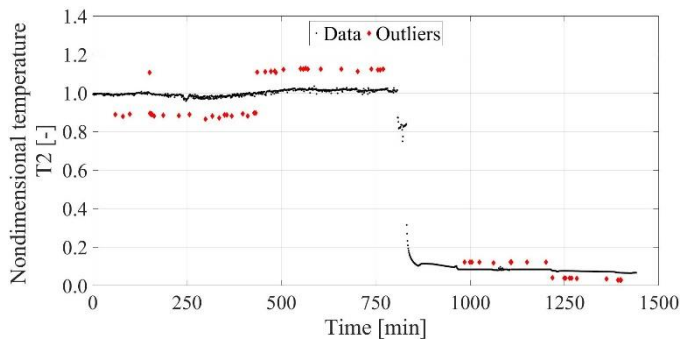

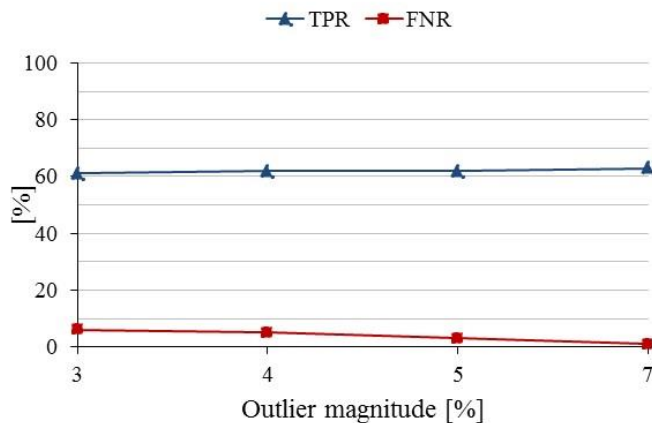**Figure 17** – Field dataset of nondimensional temperature T2 with injected outliers of 7% magnitude

Figure 18 – *TPR* and *FNR* values as a function of outlier magnitude for the field dataset of nondimensional temperature T2

CONCLUSIONS

Inspired by the k-σ methodology, two different approaches were developed by modifying the application of the test criterion of the k-σ methodology into a moving window frame. In addition to steady state data, Forward Moving Window (FMW) and Backward and Forward Moving Window (BFMW) methodologies proved their effectiveness towards the analysis of dynamic data.

Among these two, the BFMW methodology proved to be the best performing by means of simulated data, achieving True Positive Rate (*TPR*) values ranging from 94% to 97% and False Negative Rate (*FNR*) from 20% to 10% as the outlier magnitude ranges from 3% to 7%. The rate of incorrectly discarded anomalies, i.e. the False Positive Rate (*FPR*), sets below 0.1% for all the considered cases.

Tuning guidelines suitable for general application were derived by means of simulated data, i.e. optimal backward window size of 50 observations combined with an acceptability threshold of 3 and a 25 observation window combined with an acceptability threshold of 2 for the forward test criterion.

GTP-17-1359 ; Venturini

An increase of data noise affects the detection capabilities of the BFMW methodology, as the calculation of standard deviation and average values for the two windows becomes troublesome. In fact, by setting measurement noise at 2%, the methodology achieves a *TPR* of 95% at 7% outlier magnitude and drops to 50% as the magnitude reaches 5%. On the contrary, *FNR* sets at 38% in the 7% magnitude scenario and reaches almost 100% as the magnitude reaches 5%. The detection of outliers in the downside of positive slope transient maneuvers (symmetrical results are obtained for negative step changes) is made difficult, because of outlier masking effect. These outliers are marked as reliable due to the mean value of the backward acceptability rule. This phenomenon leads to a decrease in *TPR* of 25% as the step change passes from 10% to 100% (at 45° transient slope). Furthermore, the *TPR* decreases if outliers are placed only in the transient rather than only in the final steady state of the simulated maneuver. This is a consequence of the masking effect as well. However, the main goal in many field applications is the assessment of sensor reliability at steady state conditions.

The capabilities of the methodologies were tested by injecting controlled outliers in a dataset of temperature field measurements acquired from a Siemens gas turbine. The analyzed time series was characterized by the presence of a severe transient. The application of the methodology also proved to be effective, as the values of the *TPR* ranged between 60% and 70% as the magnitude of injected outliers varied between 3% and 7%, respectively. The detection capability was supported by the low values achieved by the *FNR*, which ranged from 8% to 1% in the 3% to 7% outlier magnitude range, respectively.

The main achievement of this paper lies in the development of an improved and optimized algorithm with respect to the standard k-σ methodology that is able to also analyze dynamic time series. Moreover, tuning guidelines for its application on a general basis were extracted.

In order to identify solutions to mitigate the masking effect of clustered outliers, an iterative application of the Backward and Forward Moving Window will be analyzed in the future, mainly for the analysis of small or medium size data sets, in which the impact of the iteration of the methodology on computational time may be low.

ACKNOWLEDGEMENTS

REFERENCES

[1] Scheianu, D., 2014, "Methods And Results In Remote Monitoring And Diagnosing A Fleet Of Industrial Gas Turbines", Proceedings of ASME Turbo Expo 2014, GT2014- 26068

[2] Jiang , X., Foster, C., 2014, "Plant Performance Monitoring And Diagnostics – Remote, Real-Time And Automation", Proceedings of ASME Turbo Expo 2014, GT2014- 27314

[3] Tsoutsanis, E., Meskin, N., Benammar, M., Khashayar K., 2015, "Transient Gas Turbine Performance Diagnostics Through Nonlinear Adaptation of Compressor and Turbine Maps", *J Eng Gas Turb Power* **137**(9), 091201 (Sep 01, 2015)　　　(12 pages)

[4] Cavarzere, A., Venturini, M., 2011, "Application of Forecasting Methodologies to Predict Gas Turbine Behavior Over Time" *J Eng Gas Turb Power* **134**(1), 012401 (Oct 28, 2011) (8 pages)

[5] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., Sebastianelli, S., 2001, "A System for Health State Determination of Natural Gas Compression Gas Turbines", *ASME Paper 2001-GT-223.*

[6] Pinelli, M., Spina, P. R., Venturini, M., 2012, "Gas Turbine Health State Determination: Methodology Approach and Field Application", *International Journal of Rotating Machinery*, vol. 2012, Article ID 142173, 14 pages. doi:10.1155/2012/142173.

[7] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., 2007, "Artificial Intelligence for the Diagnostics of Gas Turbines. Part I: Neural Network Approach", *ASME J Eng Gas Turb Power*, **129**(3), pp. 711-719.

[8] Bettocchi, R., Pinelli, M., Spina, P. R., Venturini, M., 2007, "Artificial Intelligence for the Diagnostics of Gas Turbines. Part II: Neuro-Fuzzy Approach", *ASME J Eng Gas Turb Power*, **129**(3), pp. 720-729.

[9] Courdier, A., Li, Y. G., 2016, "Power Setting Sensor Fault Detection And Accommodation For Gas Turbine Engines Using Artificial Neural Networks", *ASME Paper GT2016- 56304.*

[10] Sarkar, S., Jin, X., and Ray, A., 2011, "Data-Driven Fault Detection in Aircraft Engines With Noisy Sensor Measurements", J Eng Gas Turb Power, 133(8), p. 081602.

[11] Simon, D. and Litt, J., 2011, "A Data Filter for Identifying Steady-State Operating Points in Engine Flight Data for Condition Monitoring Applications", *ASME J Eng Gas Turb Power*, **133**(7), p. 071603.

[12] Roumeliotis, I., Aretakis, N., Alexiou, A., 2016, "Industrial Gas Turbine Health and Performance Assessment with Field Data", *ASME Paper GT2016-57722.*

[13] Simon, D. L., Rinehart, A. W., 2016, "Sensor Selection for Aircraft Engine Performance Estimation and Gas Path Fault Diagnostics", *J Eng Gas Turb Power* **138**(7), 071201 (Feb 17, 2016) (11 pages)

[14] Venturini, M., Puggina, N., 2012, "Prediction Reliability of a Statistical Methodology for Gas Turbine Prognostics", *ASME J Eng Gas Turb Power*, **134**(10), 101601 (9 pages).

[15] Venturini, M., Therkorn, D., 2013, "Application of a Statistical Methodology for Gas Turbine Degradation Prognostics to Alstom Field Data", *J Eng Gas Turb Power* **135**(9), 091603 (10 pages), doi:10.1115/1.4024952.

[16] Hanachi, H., Liu, J., Banerjee, A., Chen, Y., 2016, "Prediction Of Compressor Fouling Rate Under Time Varying Operating Conditions", *ASME Paper GT2016-56242.*

[17] Dewallef, P. and Borguet, S., 2013, "A Methodology to Improve the Robustness of Gas Turbine Engine Performance Monitoring Against Sensor Faults", *J Eng Gas Turb Power*, **135**(5), p. 051601.

[18] Van Paridon, A., Bacic, M., Ireland, P. T., 2016, "Kalman Filter Development For Real Time Proper Orthogonal Decomposition Disc Temperature Model", *ASME Paper GT2016- 56330.*

[19] Hurst, A. M., Carter, S., Firth, D., Szary, A., Van De Weert, J., 2015, "Real-Time, Advanced Electrical Filtering for Pressure Transducer Frequency Response Correction", *ASME Paper GT2015-42895.*

[20] Gutierrez , L. A., Pezzini, P., Tucker, D., Banta, L., 2014, "Smoothing Techniques For Real-Time Turbine Speed Sensors", *ASME Paper GT2014-25407.*

[21] Ben-Gal, I.., Outlier detection, In Maimon, O. and Rokach, L., 2005, Data mining and knowledge discovery handbook, Springer, New York.

[22] Wang, J. and Xiong, S., 2014, "A hybrid forecasting model based on outlier detection and fuzzy time series – A case study on Hainan wind farm of China", Energy, **76**, pp. 526-541.

[23] Xu, S., Baldea, M., Edgar, T., Wojsznis, W., Blevins, T., and Nixon, M., 2014, "An improved methodology for outlier detection in dynamic datasets", AIChE Journal, **61**(2), pp. 419-433.

[24] Pinelli, M., Venturini, M., Burgio, M., 2003, "Statistical Methodologies for Reliability Assessment of Gas Turbine Measurements", *ASME Paper GT2003-38407.*

[25] Hampel, F., 1974, "The Influence Curve and its Role in Robust Estimation", Journal of the American Statistical Association, **69**(346), pp. 383-393.

[26] Gomez, J., 2011, Kalman filtering, Nova Science Publishers, Hauppauge, N.Y.

[27] Martin, R. and Thomson, D., 1982, "Robust-resistant spectrum estimation", Proceedings of the IEEE, **70**(9), pp. 1097-1115.

[28] Liu, H., Shah, S., and Jiang, W., 2004, "On-line outlier detection and data cleaning", Computers & Chemical Engineering, **28**(9), pp. 1635-1647.

[29] Ganguli, R., 2002, "Noise and outlier removal from jet engine health signals using weighted FIR median hybrid filters", Mechanical Systems and Signal Processing, **16**(6), pp. 967-978.

[30] Takeuchi, J. and Yamanishi, K., 2006, "A unifying framework for detecting outliers and change points from time series", IEEE Trans. Knowl. Data Eng., **18**(4), pp. 482-492.

[31] Takahashi, T., Tomioka, R., and Yamanishi, K., 2014, "Discovering Emerging Topics in Social Streams via Link-Anomaly Detection", IEEE Trans. Knowl. Data Eng., **26**(1), pp. 120-130.

[32] Bhattacharya, G., Ghosh, K., and Chowdhury, A., 2015, "Outlier detection using neighborhood rank difference", Pattern Recognition Letters, 60-61, pp. 24-31.

[33] Everitt, B. and Howell, D., 2005, Encyclopedia of statistics in behavioral science, Wiley, Chichester.

[34] Miller, J., 1991, "Short report: Reaction time analysis with outlier exclusion: Bias varies with sample size", The Quarterly Journal of Experimental Psychology Section A, **43**(4), pp. 907-912.

[35] Ceschini, G., Gatta, N., Venturini, M., Hubauer, T., Murarasu, A., 2017, "Resistant Statistical Methodologies for Anomaly Detection in Gas Turbine Dynamic Time Series: Development and Field Validation", *ASME Paper GT2017-63410.*

[36] Ceschini, G., Gatta, N., Venturini, M., Hubauer, T., Murarasu, A., 2017, "A Comprehensive Tool for Detection, Classification and Integrated Diagnostics of Gas Turbine Sensors (DCIDS)", *ASME Paper GT2017-63411.*

[37] Young, G. A. and Smith, R. L., 2005, Essential of Statistical Inference, Cambridge University Press, Cambridge, UK.

[38] Sharma, A., Golubchik, L., and Govindan, R., 2010, "Sensor faults: detection methods and prevalence in real-world datasets", *ACM Transactions on Sensor Networks*, **6**(3), pp. 1-39.

List of figure captions

Figure 1 – Field dataset of nondimensional vibration V1

Figure 2 – FMW (above) and BFMW (below) methodology applied to the field dataset of nondimensional vibration V1

Figure 3 – FMW (above) and BFMW (below) applied to the field dataset of nondimensional temperature T1

Figure 4 – Simulated data with noise 1%, 10% step, 7% outlier magnitude

Figure 5 – Performance comparison between best and worst tunings of the FMW scheme for different outlier magnitudes

Figure 6 – Performance comparison between best and worst tunings of the BFMW scheme for different outlier magnitudes

Figure 7 – Performance comparison between FMW and BFMW methodologies with best performing tuning

Figure 8 – Influence of backward window size for different outlier magnitudes

Figure 9 – *TPR* performance of BFMW methodology as a function of k_b and k_f values (outlier magnitude at 3% and 7%)

Figure 10 – *FNR* performance of BFMW methodology as a function of k_b and k_f values (outlier magnitude at 3% and 7%)

Figure 11 – BFMW methodology performance for different outlier magnitudes at 2% noise

Figure 12 – BFMW methodology *TPR* (above) and *FNR* (below) for different outlier magnitudes and different step changes

Figure 13 – Example of the masking effect for BFMW methodology

Figure 14 – BFMW methodology performance vs. outlier magnitude, with outliers in TSB only (solid line) or SSC only (dashed line)

Figure 15 – BFMW methodology performance vs. step change, with outliers in TSB only (solid line) or SSC only (dashed line)

Figure 16 – Overview of BFMW methodology *TPR* and *FNR* performance for different simulated scenarios

Figure 17 – Field dataset of nondimensional temperature T2 with injected outliers of 7% magnitude

Figure 18 – *TPR* and *FNR* values as a function of outlier magnitude for the field dataset of nondimensional temperature T2

List of table captions

Table 1 – Parameter settings analyzed to identify the optimal tuning of the FMW and BFMW schemes

Table 2 – Tuning of the FMW methodology

Table 3 – Tuning of the BFMW methodology

Table 4 – Optimal tuning of the BFMW methodology