A Comprehensive Approach for Detection, Classification, and Integrated Diagnostics of Gas Turbine Sensors

Giuseppe Fabio Ceschini
Siemens AG,
Nürnberg 90461, Germany

Nicolò Gatta
Dipartimento di Ingegneria,
Università degli Studi di Ferrara,
Ferrara 44122, Italy

Mauro Venturini
Dipartimento di Ingegneria,
Università degli Studi di Ferrara,
Ferrara 44122, Italy

Thomas Hubauer
Siemens AG,
Nürnberg 90461, Germany

Alin Murarasu
Siemens AG,
Nürnberg 90461, Germany

Anomaly detection in sensor time series is a crucial aspect for raw data cleaning in gas turbine (GT) industry. In addition to efficiency, a successful methodology for industrial applications should be also characterized by ease of implementation and operation. To this purpose, a comprehensive and straightforward approach for detection, classification, and integrated diagnostics of gas turbine sensors (named DCIDS) is proposed in this paper. The tool consists of two main algorithms, i.e., the anomaly detection algorithm (ADA) and the anomaly classification algorithm (ACA). The ADA identifies anomalies according to three different levels of filtering based on gross physics threshold application, intersensor statistical analysis (sensor voting), and single-sensor statistical analysis. Anomalies in the time series are identified by the ADA, together with their characteristics, which are analyzed by the ACA to perform their classification. Fault classes discriminate among anomalies according to their time correlation, magnitude, and number of sensors in which an anomaly is contemporarily identified. Results of anomaly identification and classification can subsequently be used for sensor diagnostic purposes. The performance of the tool is assessed in this paper by analyzing two temperature time series with redundant sensors taken on a Siemens GT in operation. The results show that the DCIDS is able to identify and classify different types of anomalies. In particular, in the first dataset, two severely incoherent sensors are identified and their anomalies are correctly classified. In the second dataset, the DCIDS tool proves to be capable of identifying and classifying clustered spikes of different magnitudes. [DOI: 10.1115/1.4037964]

Introduction

Energy market demand sets high requirements to the productivity of gas turbine (GT) units, by imposing high availability and efficiency levels to achieve cost effectiveness. Furthermore, the complexity of the units implies a high level of insight on the health state of the turbines.

In this context, Siemens remote diagnostic service (RDS) monitors oil and gas and industrial rotating equipment in more than eighty countries around the world. The main goal of RDS is to identify potential problems long before they impact operation, thus reducing downtime and maintenance costs. RDS is provided with a platform fulfilling a wide range of responsibilities including the collection of signals and events from customer sites, data storage in the Siemens network, derivation of information and alert analysis, automated data processing, fault isolation, and sending notifications to customers, which include diagnosis and short-term recommendations.

Nowadays, a large amount of data, e.g., in the order of tens of GBs, is collected by RDS 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. Effectiveness refers to the ability to find all the real problems, while reducing the number of false alarms, especially filtrating those generated by sensor malfunctioning. Despite their different purpose, in fact, state-of-the-art analysis tools rely on direct thermodynamic (e.g., temperatures, pressures, etc.) and mechanical (e.g., vibrations) measurements to monitor [1,2] or forecast [3–5] the health state of GT units. However, the extreme conditions in which sensors operate may cause hardware degradation and failure of measurement devices [6], thus generally lowering data quality. For these reasons, outlier identification constitutes a promising and challenging field for scientific research, especially regarding GT units [7–10].

Data processing effectiveness can generally be achieved through careful tuning of parameters. However, 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. This is the main reason why complex methodologies based on heuristic rules tend to fail in field applications. In this regard, a desired feature for RDS analysis algorithms is high usability, meaning that the tuning process should be user-friendly without sacrificing algorithm detection capability. Furthermore, general guidelines for tuning should be available, in order to reduce data dependency and enlarge the application field for a specific tuning setting.

On the other hand, efficiency means that analysis should be concluded in a reasonable amount of time, leaving enough room to react proactively to an identified problem. This stresses the suitability of computational undemanding methodologies rather than complex model based approaches for the development of efficient methodologies for remote diagnostics.

In the framework outlined earlier, this paper aims at the development of a comprehensive tool for detection, classification, and integrated diagnostics of gas turbine sensors (named DCIDS). The ultimate goal is to achieve an optimal balance between effectiveness, efficiency, and usability. This works sets as the final step of a wider research activity undertaken by the authors, which focused in Refs. [11] and [12] on the analysis of statistics-based methodologies for anomalies identification. In Ref. [11], a well-known parametric test methodology, i.e., the k–σ methodology, was improved by adopting a backward and forward moving

Contributed by the Oil and Gas Applications Committee of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received July 14, 2017; final manuscript received July 30, 2017; published online October 25, 2017. Editor: David Wisler.

window scheme, which allowed the methodology to also successfully manage dynamic time series. In Ref. [12], possible improvements deriving from the implementation of robust statistical estimators were evaluated. In fact, in addition to the k–σ methodology, three methodologies (i.e., k-MAD, hybrid σ-MAD, and biweight) were also developed and scenarios to best exploit their respective potential were identified.

Therefore, this paper contributes to technical literature by providing a comprehensive framework for field data assessment and sensor fault classification.

Detection, Classification, and Integrated Diagnostics for Sensor Anomaly Analysis

Despite the effectiveness demonstrated both on simulated and field data, the sole application of statistical methodologies may be not effective in order to provide a comprehensive and reliable analysis of sensor data. In fact, in many field applications, redundant sensors are installed to increase the reliability of the sensing system in case of failures of the single devices. The information provided by redundant sensors is extremely important to evaluate data reliability, as a set of measurements of the same quantity at a specific location of the unit becomes available. This information cannot be neglected when a comprehensive and industry-ready tool for anomalies detection is to be developed, which is the case of the approach adopted in this work.

Moreover, the capabilities of the algorithm should be expanded beyond the sole outlier identification for data quality improvement. Identified anomalies should be classified according to their characteristics into fault classes that reflect the features of typical sensor failure modes. This opens to the implementation of a methodology for the diagnostics of the whole sensing system.

On the basis of these considerations, this paper presents the development of a comprehensive tool for DCIDS of gas turbine sensors. The approach adopted to develop the tool is the same adopted in Refs. [11] and [12], i.e., reducing the number of parameters requested to the user in order to obtain a straightforward, effective, and industry-ready tool. In this way, the algorithm should be able to challenge more complicated heuristics-based methodologies, which often demand the setting of acceptability thresholds (e.g., gradient increase) that might be unfamiliar to the user. To this purpose, generally applicable guidelines are provided in this paper to properly tune all the parameters requested for the analysis.

The structure of the developed tool is reported in Fig. 1. The anomaly detection algorithm (ADA) processes the data of the time series by means of user-specified thresholds and distinguishes between reliable and anomalous observations. The former can serve as enhanced quality inputs for subsequent analysis, while the latter are further analyzed according to their characteristics, which are identified by the ADA as well.

Anomalies, together with the respective features, serve as inputs to the anomaly classification algorithm (ACA), which does not require any additional user specified parameter to perform its analysis. The classification, supported by failure mode databases, can be employed to connect each identified fault class to specific hardware failures and perform complete diagnostics of the sensing system. The results of the classification algorithm also open to the application of machine learning methodologies (e.g., artificial neural networks).

It is worth mentioning that the ADA is also provided with two additional capabilities. The first capability enables the detection of stuck sensor readings, i.e., a series of data values, of which the variation is zero or almost null for a period of time longer than expected [13]. The sensibility of the assessment can be adjusted by the user according to the type of the considered physical quantity, the sampling frequency, and the accuracy of the sensing system. Namely, the user can specify both the maximum difference in measurements and the minimum period of time (expressed in terms of number of observations) required to consider the sensor readings as stuck. The second capability is referred here as sensor muting. In field applications, some sensors may produce anomalous observations with sporadic reliable observations. The presence of these isolated reliable readings in the post processing time series may produce noisy and unclean results which can be difficult to interpret during trend or diagnostic analyses. The sensor muting capability is designed to tackle this issue. After the application of ADA, the user can reject all observations of specific sensors that proved to be faulty for most of the time. In this way, all measurements from "muted" sensors are excluded from the reliable data set, thus enhancing post processing data quality. The percentage ratio between anomalies and total number of observations, called contamination rate (CR), is calculated and provided to the user. This is a useful insight to assess the meaningfulness of observations from each sensor and to identify the sensors to mute.

In the following, the two kernels of the DCIDS tool, i.e., ADA and ACA, are described in detail. Guidelines and best practices for the tuning of each methodology implemented in ADA are provided as well. Finally, the capabilities of the DCIDS tool are tested by analyzing field datasets.

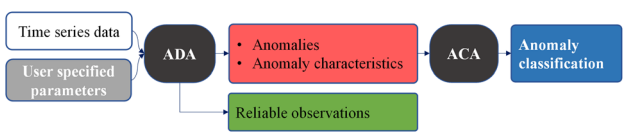

Fig. 1 DCIDS tool structure

Anomaly Detection Algorithm

The architecture of the ADA is illustrated in Fig. 2, while a detailed description of the data flow through ADA is presented in the flowchart in Fig. 3. Gray boxes identify the user-specified parameters required to perform the analysis.

Observations are processed according to a progressively refined analysis, performed at three different levels of judgment. As one level is performed, the values and time points of the anomalies are identified, while observations marked as reliable serve as inputs to the subsequent level. Anomalies characteristics are stored as well, namely distance from the acceptability threshold and detection score (DS), i.e., number of sensors which identified an anomaly at a given time point. This feature can be calculated only in case redundant sensors are available. The execution of the ADA stops when all the three levels have been passed. These are described in detail in the following, by reporting the concept behind their implementation, the user specified parameters required for their proper setting and the evaluation of outlier magnitude.

Physics-Based Thresholds (First Level). At the first level, anomalies are detected according to minimum and maximum acceptability boundaries based on physics-based considerations. Evident anomalies are flagged by imposing thresholds to detect incoherent measurements (e.g., negative temperature) and perform gross outlier identification and sensor check. At this level, the parameters required to the user are two, i.e., minimum and maximum acceptability thresholds (hereafter referred as Min and Max).

These parameters can be set by the user on the basis of engineering sense and/or by considering recommendations from

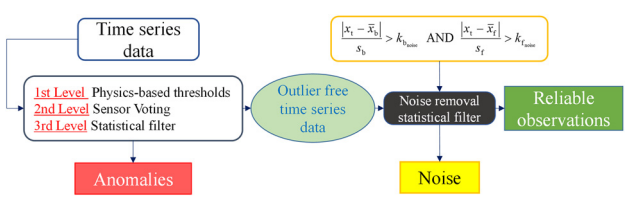

Fig. 2 ADA architecture

sensor datasheets to identify severe hardware degradation. The distance from the acceptability boundaries is expressed in terms of difference between the measured value and the threshold.

Sensor Voting (Second Level). At the second level, anomalies are identified by means of an intersensor analysis hereafter referred as "sensor voting," which can be performed only in case redundant measurements are available. The flowchart reported in Fig. 4 illustrates the sensor voting procedure, which compares the observations of each sensor to the ones collected by the whole sensor set in order to assess their mutual coherence.

At each time point of the time series, the algorithm computes the median value of the sensor set readings and flags the observations exceeding a user specified threshold difference as anomalies. If anomalies are detected, the algorithm excludes them from the analysis and, for the time point under consideration, iterates the process until no further anomalies are detected at that specific time point. This decreases the influence of outliers on median calculation, as the analysis is progressively refined. When no further anomalies are detected, the algorithm proceeds to the next time point until the whole time series is processed. If the available measurements at a specific time point are just two, sensor voting cannot be performed, as the probability for those observations to be reliable is equal. In this case, the algorithm marks such measurements as unprocessed and moves to the subsequent time point.

The sensor voting threshold is expressed in terms of deviation of each observation from the median value of the sensors set at a specific time point. The DCIDS tool allows the user to choose between absolute and relative deviation (AD and RD, respectively) from the median. These are defined according to Eqs. (1) and (2), respectively, in which $x_{i,t}$ is the observation of a single sensor i at time point t, while $x_{set,t}$ is the group of measurements collected by the whole sensor set at the same time point t.

$$AD = |x_{i,t} - \text{med}(x_{set,t})| \tag{1}$$

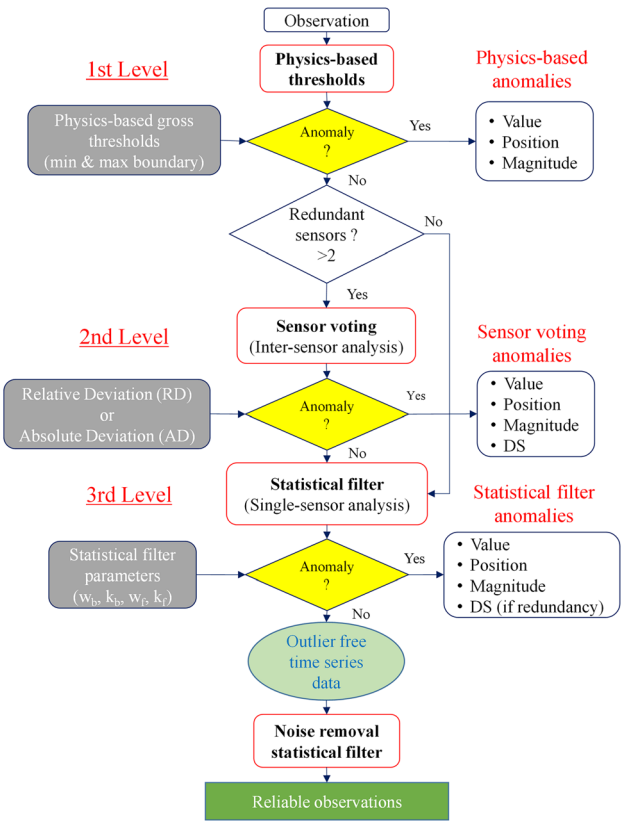

Fig. 3 ADA flowchart

$$RD = \frac{AD}{\text{med}(x_{set,t})} \cdot 100 \tag{2}$$

The values of AD or RD thresholds can be set according to engineering sense and/or manufacturer guidelines. However, it is worth highlighting some general comments about the different behavior of AD and RD. Figure 5 represents the median of temperature measurements from a temperature dataset (referred as T1 dataset) collected on a gas turbine, together with sensor voting acceptability boundaries set at AD = 35 °C and RD = 10%.

By setting the RD acceptability threshold, the width of the allowable boundaries is clearly a function of the observation values. Therefore, it is important to notice that voting is stricter if the data values are low, while it becomes less selective as data values increase. Therefore, a percentage variation might result easier to set according to experience and practical sense of the user, becoming recommendable when the user is not familiar with the analyzed physical quantity and its likely range of variation. However, since RD is proportional to data current values, the threshold is less controllable by the user.

On the contrary, AD threshold is by nature independent of the values of observations. Therefore, this solution is recommended when guidelines on allowable spreads of the analyzed quantity are available. In the case of temperatures, for example, best practices

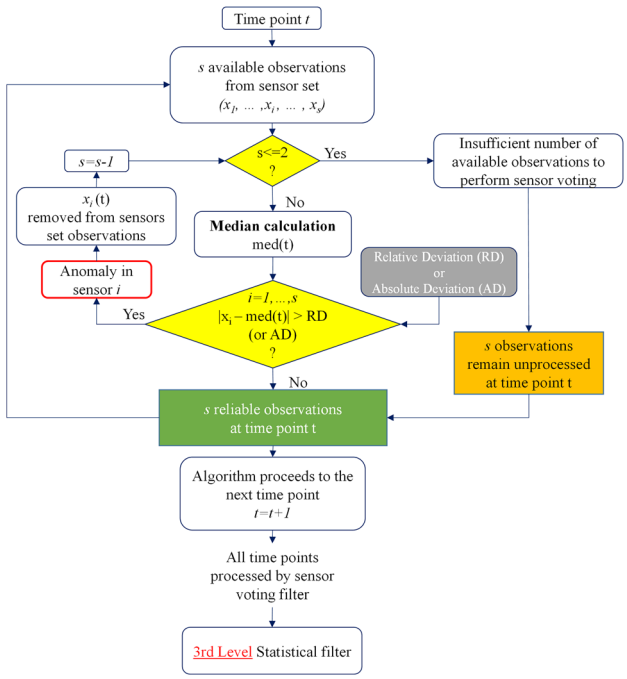

Fig. 4 Sensor voting procedure

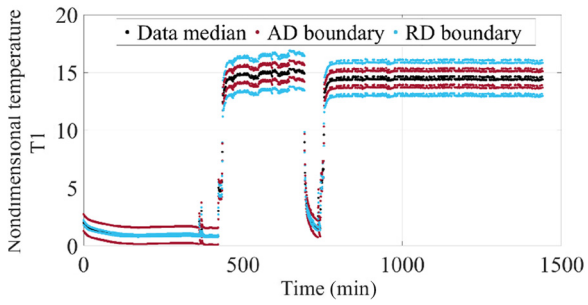

Fig. 5 AD and RD for the nondimensional temperature T1 dataset

suggest a maximum deviance of 35 °C between temperatures measured by adjacent thermocouples [14]. However, general guidelines about the optimal setting of RD or AD cannot be derived, as their efficiency is sensibly data dependent and user-specified. Since RD and AD can be related to the median value of sensor set observations, the respective value of RD, which should be set to achieve acceptability boundaries of the same width, can be calculated in correspondence of a given AD at each time point.

Statistical Filter (Third Level). After removing gross outliers and incoherent measurements in the sensor set at first and second level, respectively, a refined single sensor analysis is performed at the third level by means of the application of the statistical methodologies analyzed in Ref. [12], i.e., k–σ, k-MAD, hybrid σ-MAD, and biweight methodologies. Even if each of the mentioned methodologies can be selected in the DCIDS tool, the application of the k–σ is considered as the primary option in this paper, as this proved to be effective toward different scenarios which dealt with field data [12]. Independently of the selected methodology, the anomaly characteristics are expressed in terms of the absolute difference between the value of k, achieved by any anomalous observation, and the value set by the user for this acceptability threshold.

The parameters requested to the user at this stage are the same as the ones which characterize the statistical moving window methodology, i.e., number of observations in the backward w_b and forward w_f windows and their respective acceptability thresholds k_b and k_f. Optimal values for these parameters were identified in Ref. [12] and are reported in Table 1 for each methodology. Thanks to the results obtained toward both simulated and field data; these values can be recommended for field application.

Noise Removal Filter. This additional statistical filter aims at identifying an intermediate region between reliability and anomaly boundaries, here referred as noise. The statistical filter noise is implemented by applying the statistical methodology of the third level, but its acceptability thresholds are properly adjusted. Starting from the optimal tuning values, a rule of thumb configuration, based on the results achieved in Refs. [11] and [12] and on engineering practice, was conceived and applied in this paper. The values of the noise removal filter parameters are reported in Table 2. As it can be seen, the main difference of the noise removal filter parameters with respect to the four methodologies listed in Table 1 is the value of k_b, which is equal to two. This is coherent with the fact that the noise removal filter performs a refined analysis on data which have been already processed by three levels of judgment. Therefore, in case the noise removal filter is enabled by the user, the third level of the ADA distinguishes between anomalies, noise and reliable observations according to the rules reported in Fig. 6.

DCIDS Tool Set-Up. Table 3 summarizes the parameters requested to the user to run the DCIDS tool. These values, derived by the authors in Refs. [11] and [12], are adopted for the analysis performed in this study.

Anomaly Classification Algorithm

As reported in the scheme in Fig. 1, the ACA follows the ADA in the DCIDS tool. The aim of the ACA is to interpret the information regarding the anomalies collected by the ADA to identify sensor failure patterns and their possible root causes.

Namely, key information about anomalies concerns:

(a) mutual time correlation;
(b) magnitude, i.e., distance from acceptability boundaries;
(c) detection score.

The analysis of these key features allows the discrimination between the following fault classes:

(a) isolated faults versus serial faults;
(b) minor faults versus major faults;
(c) specific sensor(s) fault versus multiple sensor faults.

Anomalous observations are related to the proposed fault classes according to three different criteria, which reflect the previously identified key information:

(a) Time correlation: two anomalies are time correlated if they occupy subsequent slots in the time series (e.g., at time point t and $t+1$).
(b) Major faults: anomalies whose magnitude is outside an interval, whose width is a function of measurement uncertainty. An anomaly that lies inside the same interval is classified as minor fault.
(c) The value of the detection score allows the discrimination among single sensor faults, multiple sensor faults (i.e., at least two sensors are faulty), and the case in which anomalies are detected by all but two sensors, since at least two sensors are required to make sensor voting meaningful.

The classification pattern is reported in Fig. 7, in which 1 indicates that a particular feature is verified, while 0 means the opposite. Symbols S, M, and A for detection score mean that a certain anomaly is detected by single (S), multiple (M), or all but two sensors (A), respectively. The symbol "/" specifies that the considered feature cannot be determined on the basis of the available

Table 1 Statistical filter optimal tuning [11,12]

Methodology	w_b	k_b	w_f	k_f
k–σ	50	3	25	2
k-MAD	50	3	25	3
Hybrid σ-MAD	50	3	25	3
Biweight	50	3	25	3

Table 2 Noise removal filter parameters

w_b	k_b	w_f	k_f
50	2	25	2

$$\frac{|x_t - \bar{x}_b|}{s_b} > k_b \text{ AND } \frac{|x_t - \bar{x}_f|}{s_f} > k_f \quad \text{Anomaly}$$

$$\frac{|x_t - \bar{x}_b|}{s_b} > k_{b_{noise}} \text{ AND } \frac{|x_t - \bar{x}_f|}{s_f} > k_{f_{noise}} \quad \text{Noise}$$

$$\frac{|x_t - \bar{x}_b|}{s_b} \le k_b \text{ OR } \frac{|x_t - \bar{x}_f|}{s_f} \le k_f \quad \text{Reliable observation}$$

Fig. 6 Detection rules of statistical filter and noise removal filter

Table 3 DCIDS tool parameters

Level of detection	Parameter	Value	Ref.
First level	Min	User-defined	—
	Max	User-defined	—
Second level	RD or AD	35 °C[a]	[14]
Third level	w_b	50	[12]
	k_b	3	[12]
	w_k	25	[12]
	k_f	2 or 3	[12]

[a]RD value for temperature dataset.

information. Different anomalies, in fact, provide a different amount of information according to the level at which they are identified by the ADA. For instance, anomalies identified at second and third level (sensor voting and statistical filter, respectively) provide more information than the ones identified at first level (physics thresholds), allowing a deeper analysis during the classification phase.

The proposed fault classes are designed to reflect typical field operation sensor faults. For example, the situation in which the classification changes from serial minor fault to serial major fault of a single sensor (identified in Fig. 7 as 1.0 S and 1.1 S, respectively) is representative of a series of time correlated anomalies with an increasing magnitude. This is the typical pattern of a sensor drift fault, in which sensor measurements gradually deteriorate over time. On the contrary, if no evolution of anomaly magnitude is detected, the occurrence of a bias fault is more likely. In this scenario, in fact, sensors produce anomalous observations, of which the magnitude is constant over time. Another example is the case of isolated major fault(s) of single or multiple sensors, identified in Fig. 7 as 0.1 S and 0.1 M, respectively, which correspond to the occurrence of a spike. The results of the classification algorithm may be elaborated by means of machine learning systems, e.g., artificial neural networks, trained to relate failure modes to sensor hardware failures. Thus, complete diagnostics of the whole sensing system may be performed.

Detection level	Time correlation	Major fault	Detection Score	Fault Classification
1st Level anomalies (Physics-based thresholds)	1	/	/	Serial sensor(s) fault
	0	/	/	Isolated sensor(s) fault
2nd Level anomalies (Sensor Voting)	1	1	S	Serial major fault of a single sensor
			M	Serial major fault of multiple sensors
			A	Serial major fault of all sensors
		0	S	Serial minor fault of a single sensor
			M	Serial minor fault of multiple sensors
			A	Serial minor fault of all sensors
	0	1	S	Isolated major fault of a single sensor
			M	Isolated major fault of multiple sensors
			A	Isolated major fault of all sensors
		0	S	Isolated minor fault of a single sensor
			M	Isolated minor fault of multiple sensors
			A	Isolated minor fault of all sensors
3rd Level anomalies (Statistical Filter)	1	1	S	Serial major fault of a single sensor
			M	Serial major fault of multiple sensors
			A	Serial major fault of all sensors
		0	S	Serial minor fault of a single sensor
			M	Serial minor fault of multiple sensors
			A	Serial minor fault of all sensors
	0	1	S	Isolated major fault of a single sensor
			M	Isolated major fault of multiple sensors
			A	Isolated major fault of all sensors
		0	S	Isolated minor fault of a single sensor
			M	Isolated minor fault of multiple sensors
			A	Isolated minor fault of all sensors
Noise filter (optional)	1	/	/	Measurement noise

Fig. 7 Anomaly classification algorithm

While the criteria for time correlation and maximum detection score are unequivocally defined, the classification of minor and major faults is worth of further discussion. The proposed methodology classifies outliers according to their magnitude by means of measurement uncertainty. This is used to establish two different layers within the anomalous behavior boundaries. The adoption of measurement uncertainty relies on the concept that minor faults are likely to be caused by low accuracy in measurements, while major faults significantly detach from this behavior. Furthermore, by using measurement uncertainty as the discriminant does not require the input of any parameter by the user, as uncertainty can be directly inferred from available observations. This avoids any ambiguity in defining the difference between minor and major anomalies, while at the same time it provides a classification with a physics-based background.

Uncertainty quantification requires a certain amount of information so that only second and third level anomalies (i.e., identified by sensor voting and statistical filter) can be analyzed.

However, these types of anomalies are identified by two completely different methodologies and are consequently defined by different magnitude characteristics. As previously described, sensor voting classifies anomalies in terms of absolute or relative distance from the sensor set median, while the anomalies identified at the third level are classified according to a statistical distance from the acceptability threshold. Therefore, uncertainty, and consequently the width of minor and major fault boundaries, must be defined in two different ways.

Sensor voting anomalies are detected when observations at a specific time point detach from the median value of the whole sensor set more than a relative or absolute threshold. As well known, measurement uncertainty can be estimated by considering the dispersion of repeated measurements of the same quantity. This concept can be extended by considering the measurements collected by the sensor set at a specific time point as if they were repeated measurements. Under this assumption, uncertainty estimation can be performed by adopting a so-called Type A approach [15]. The measurement uncertainty u can be calculated according to the following expression [15], in which s_t is the standard deviation (SD) of reliable observations collected by the n reliable sensors of the sensing set at the time point t

$$u = \frac{s_t}{\sqrt{n}} \tag{3}$$

The uncertainty u can be added to the upper and lower acceptability boundaries of sensor voting, for both absolute and relative

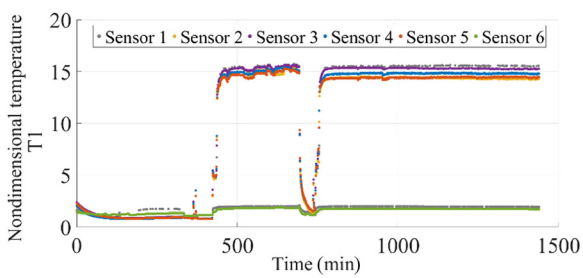

Fig. 8 Nondimensional temperature T1 dataset

difference thresholds to determine minor faults and major fault boundaries. Anomalies identified by the statistical filter are observations whose difference from the location estimate (e.g., mean) is larger than k times the value of the scale estimator (e.g., standard deviation). As the optimal values for the k thresholds were identified to be between two and three depending on the adopted methodology, it is reasonable to set the minor faults band within an interval up to four times the standard deviation. Thus, all anomalies included in this boundary are classified as minor faults, while other anomalies are classified as major faults.

Application of Detection, Classification, and Integrated Diagnostics Tool to Field Data

The detection and classification capabilities of the DCIDS algorithm are evaluated in this section by means of field datasets, taken on gas turbines in operation. First, a visual comparison between the raw time series and the same time series processed by the DCIDS tool is provided. This comparison offers an at a glance insight on the detection capabilities of the tool. Second, the analysis quantifies the detection capability per each level. It should be noted that the physics-based thresholds (first level) were imposed in such a way that no anomalies were identified, so that the analysis only highlights sensor voting and statistical filter capabilities (second and third levels, respectively). For each sensor, the rate of reliable, unprocessed, and anomalous observations is reported. Finally, the results from the ACA document the occurrence of each proposed fault class.

Field Data. In this paper, the results of two field time series (both with redundant sensors) are reported, in order to fully evaluate the capabilities of DCIDS tool.

The first time series, hereafter referred as temperature T1 dataset, contains the readings from six different thermocouples collected with 1 min sampling frequency. The nondimensional time series, represented in Fig. 8, consists of three steady-states connected by two severe transients. As it can be seen, measurement spikes occur for the majority of the sensors at $t = 340$ min. As the dataset refers to temperatures, the sensor voting (second level) is set with AD = 35 °C, according to best practices reported in Ref. [14]. The AD value was subsequently scaled to analyze normalized data. The k–σ methodology is selected for the statistical filter (third level), in agreement with the guidelines reported in Ref. [12].

The second dataset, hereafter referred as temperature T2, contains the readings from thirteen different thermocouples collected with 1 min sampling frequency. The dataset, presented in Fig. 9, consists of two steady-states connected by a significant transient. It can be observed that several concentrated spikes occur in the second steady-state. These constitute a rather challenging test for the k–σ methodology; thus, the k–σ methodology is also compared to the k-MAD methodology for the statistical filter. As in the previous case, the dataset refers to temperatures; thus, the sensor voting (second level) is set with AD = 35 °C, properly scaled to analyze normalized data.

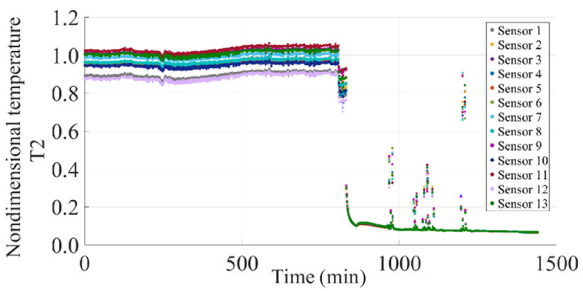

Fig. 9 Nondimensional temperature T2 dataset

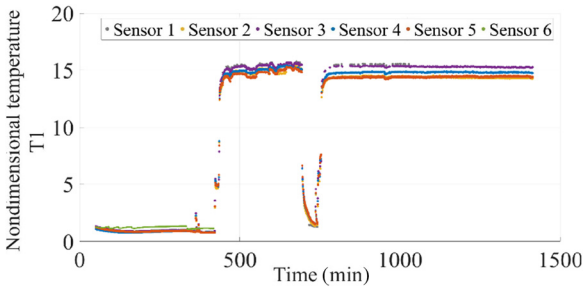

Fig. 10 ADA application to the nondimensional temperature T1 dataset

Temperature T1 Dataset Results. The results of DCIDS processing of temperature T1 measurements are presented in Fig. 10. As it can be seen by comparison to Fig. 8, data are cleaned from the readings of sensors #1 and #6, which produced incoherent measurements with respect to the other sensors. These anomalous observations were detected by sensor voting. Furthermore, the statistical filter (which makes use of the k–σ methodology) is able to detect and isolate the spikes at $t = 340$ min. Some observations are flagged as anomalous at the beginning of the first transient, mainly because of the gap in measurements caused by the sampling frequency. However, the methodology is able to preserve the trend of the transient despite these likely false-positive calls.

This dataset is also particularly meaningful for testing sensor muting capability. Even if faulty for the majority of observations, in fact, sensor #1 (gray) and sensor #6 (green) produce few reliable measurements (see Fig. 11), which are not removed from the postprocessing time series. The meaningfulness of such sporadic reliable observations is rather low and their presence may be difficult to interpret during diagnostic or trend analyses. The DCIDS tool computes the CR of each sensor so that the user can decide whether to mute them or not. The CRs for each sensor are reported in Table 4. Sensors #1 and #6 present a CR of 68% and 70% respectively, denoting a low overall reliability. Therefore, these two sensors may be excluded from the postprocessed time series by means of sensor muting and a neater result may be obtained without losing relevant information.

Figure 11 reports the shares of identified anomalies among the three levels of control, together with the percentages of unprocessed and reliable observations. As it can be seen, the sensor voting (second level) identifies the majority of anomalies, consequent to the faulty measurements from sensors #1 and #6. In particular, the sensor voting can effectively detect the 61%, and 69% of the readings from sensors #1 and #6, respectively.

For the other sensors, the percentage of sensor voting detection is negligible, because of time series morphology. The statistical filter applied at the third level identifies as anomalous nearly the 3% of observations of sensors #2, #3, #4, and #5. Instead, the

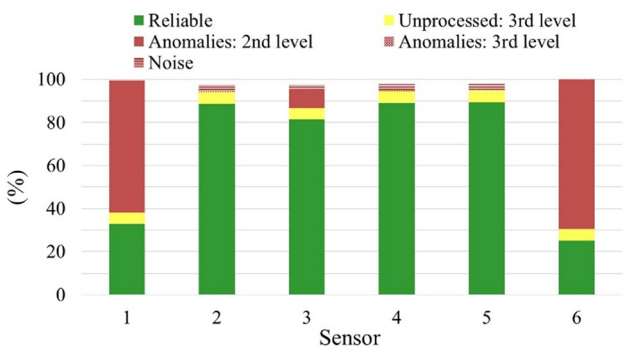

Fig. 11 Share of ADA results for the nondimensional temperature T1 dataset

Table 4 Contamination rates for the temperature T1 dataset

Sensor	#1	#2	#3	#4	#5	#6
CR (%)	68	6	44	6	6	70

statistical filter does not identify any anomaly. Subsequently, the application of the statistical noise filter isolate the 3%, 2%, 3%, and 3% of measurements of sensors #2, #3, #4, and #5, respectively. Instead, sensors #1 and #6 prove to be noise free, as they provide an almost constant output throughout the time series.

The percentage of observations left unprocessed is equal to 5% of total observations for all sensors. This fraction corresponds to the measurements included in the forward window w_f at the end of the time series that are not processed by the k–σ methodology [11].

The results obtainable by means of the ACA are presented in Fig. 12, which reports sensor voting (second level) and statistical filter (third level) anomaly classification shares so that the sum of detected anomalies makes 100%. Even though the results are data dependent, the most common failure classes can be identified.

Considering sensor voting anomalies, it can be observed that serial, i.e., time related, anomalies are remarkably more frequent than time isolated anomalies in the T1 dataset. In particular, sensors #1 and #6 experience serial major faults, due to the sensors breakdown that can be noticed by comparing Figs. 8 and 10. Regarding sensor #3, sensor voting anomalies are classified for the 70% as serial minor faults and for the 30% as isolated minor faults. Serial and isolated minor faults entirely contribute to the classification of sensor voting anomalies for sensors #2 and #4, respectively.

While serial anomalies are more frequent for sensor voting anomalies, this tendency is inverted in the T1 dataset for statistical filter anomalies. In this case, in fact, isolated minor faults are most frequently identified, as a consequence of the presence of minor spikes in the dataset. Almost 30% of statistical filter anomalies for sensor #2 are classified as serial major faults, while no anomalies are identified by the statistical filter in sensors #1 and #6. No anomalies, occurring contemporarily in all the sensors, are detected by both sensor voting and statistical filter.

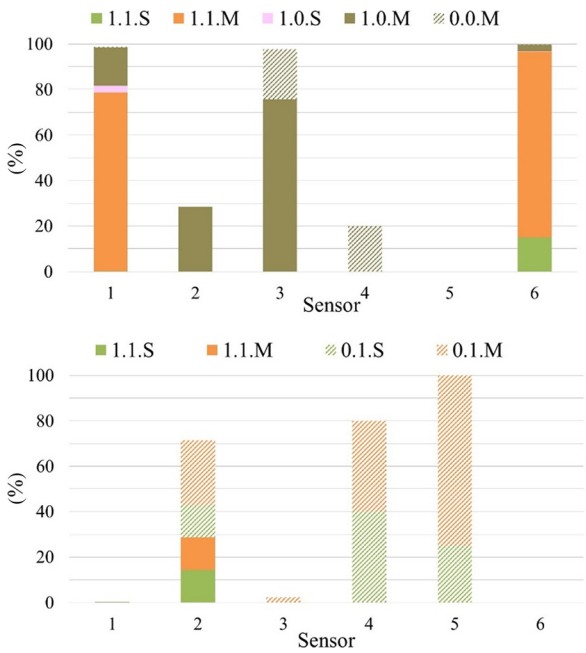

Fig. 12 ACA application to the nondimensional temperature T1 dataset—sensor voting (top) and statistical filter (bottom)

Temperature T2 Dataset Results. The outcome of processing the temperature T2 dataset by means of the DCIDS tool is presented in Fig. 13. As for the temperature T1 time series, according to Ref. [12], these results are obtained by means of the application of the statistical filter in the form of the k–σ methodology. It can be seen that, even if the majority of the spikes occurring after the transient are removed by the combined application of second and third levels of control, some minor spikes can still be present. This is due to the reduced capability of the k–σ methodology to handle clustered outliers, which is the case of the temperature T2 dataset. Thus, the k-MAD methodology is used instead of the k–σ methodology, to provide a neater result. In fact, according to Ref. [12], this solution proves its effectiveness, as it can be noticed from Fig. 14, which reports the results obtained by means of the k-MAD methodology.

The spikes are completely removed from the time series by the statistical filter, at the cost of higher anomalous calls, mainly during the transient. The increase of detection capability is not negligible, but rather contained, as an increase of 5% in the CR of the sensor set is observed by passing from the application of the k–σ methodology to the k-MAD methodology. Furthermore, the shape of the time series is still preserved, thus not compromising the meaningfulness of data but contemporarily achieving an overall enhanced data quality with respect to the results of the k–σ methodology.

Figure 15 reports the share of identified anomalies among the three levels of control, together with the percentage of unprocessed reliable observations. The contribution of sensor voting is marked for sensors #1, #9, #11, and #12 (57%, 23%, 57%, and 57%, respectively), while it is almost negligible for the other thermocouples. Therefore, sensors #1, #11, and #12 may be excluded from the postprocessed time series. On the contrary, the rate of detection for the statistical filter is almost constant for all the thirteen sensors, setting at an average of 6%. The only contribution to the rate of unprocessed observations is due to the observations in the last forward window in the statistical filter.

The results of the application of the ACA to the temperature dataset T2 are presented in Fig. 16. It can be seen that for those sensors that experience a high relative rate of sensor voting detection, i.e., sensors #1, #11, and #12, the most frequent fault class is "serial major faults" occurring in multiple sensors.

This class is immediately followed by the serial specific sensor(s) minor faults, which is the most frequent fault class for sensor #9. Similarly to the dataset T1, no anomalies occurring in all the sensors of the sensing set are identified by sensor voting. Furthermore, isolated minor faults are detected in multiple sensors, i.e., #2, #3, #5, #6, #7, #9, and #13 (5% on average).

Regarding statistical filter anomalies, the most frequent fault class is serial major faults. This occurs in multiple sensors of the sensor set and, in a remarkable percentage of the cases, even throughout the whole sensor set. This distribution between faults occurring in multiple sensors and in the sensor set as a whole, is also noticed for the second most frequent fault class, i.e., isolated

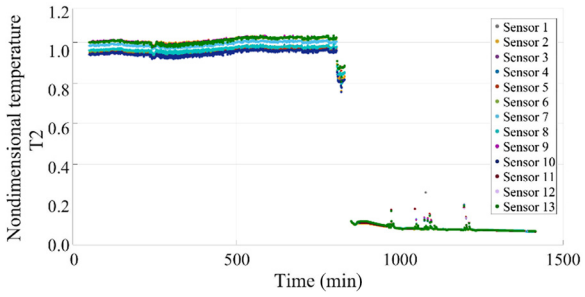

Fig. 13 ADA application to the nondimensional temperature T2 dataset (k–σ methodology at third level)

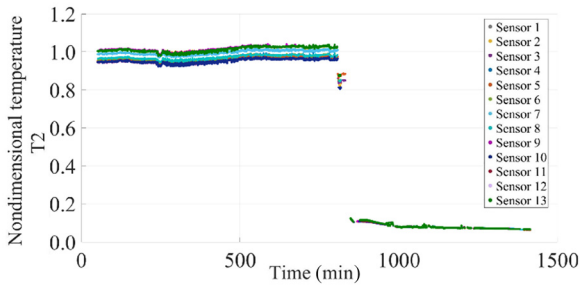

Fig. 14 ADA application to the nondimensional temperature T2 dataset (*k*-MAD methodology at third level)

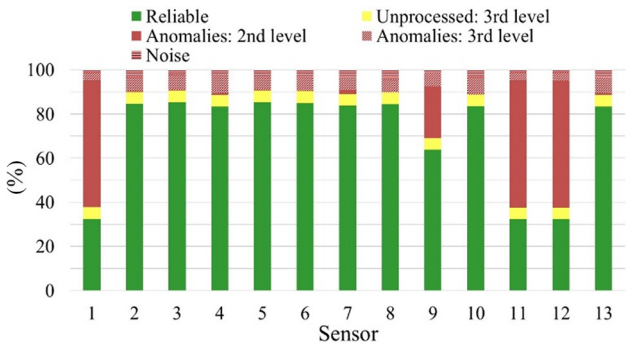

Fig. 15 Share of ADA results for the nondimensional temperature T2 dataset

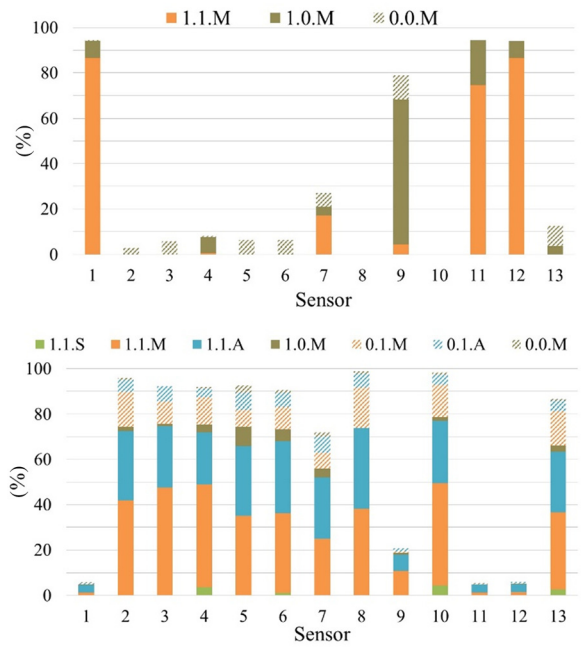

Fig. 16 ACA application to the nondimensional temperature T2 dataset—sensor voting (top) and statistical filter (bottom)

major faults. In fact, the spikes occurring after the transient are cataloged in this class by the ACA.

Conclusions

In this study, a comprehensive methodology for detection, classification, and integrated diagnostics of gas turbine sensors (named DCIDS) was developed and implemented. This consists of two main kernels, i.e., the ADA and the ACA. While the ADA identifies outliers according to a multilevel processing and calculates their characteristics, the ACA exploits available information to classify anomalies according to fault classes corresponding to frequent sensor fault scenarios. Both algorithms are designed to require a low number of parameters to the user, thus enhancing promptness of tuning and generality of application.

In particular, the ADA processes anomalies according to three different levels of analysis, followed by a noise statistical filter. Instead, the ACA is designed to classify anomalies into unambiguous fault patterns, without requiring any further parameter to the user. The results of the ACA can subsequently be used for sensing system diagnostics.

In order to test the capabilities of the DICDS tool, two different temperature field datasets were considered. Even if the results are strictly data dependent, the capabilities of both ADA and ACA can be appreciated. In particular, the DCIDS tool is able to analyze multi sensor time series (namely, 6 and 13 sensors for the first and the second dataset, respectively), successfully identifying anomalies while preserving the time series trend despite the occurrence of severe transients.

In the first dataset, the DCIDS tool successfully identified as anomalous two sensors that produced severely incoherent measurements with respect to the other four composing the sensor set. These anomalies were successfully classified as serial multiple sensor major faults. In the same dataset, minor spikes were detected as well and correctly classified as isolated minor faults occurring both in single and multiple sensors.

The second dataset was a challenging field of application, in particular for the k–σ methodology at the third level of detection, as clustered spikes of different magnitudes occurred in the final part of the time series. The DCIDS tool proved to be effective toward this scenario as well, thanks to the application of the k-MAD methodology at the third level. In this way, detection performance sensibly improved with respect to the k–σ methodology application and all spikes were successfully detected and classified.

Acknowledgment

The authors gratefully acknowledge Siemens for the permission to publish the results.

Nomenclature

k = acceptability threshold for the statistical filter
med = median value
n = number of reliable sensors at a given time point
s = standard deviation of the reliable readings of the sensor set (sensor voting)
$\bar{s}$ = standard deviation in the window sample for the statistical filter
t = time
u = uncertainty
w = number of measurements in the window sample
x = measurement in the time series
$\bar{x}$ = mean value

Subscripts and Superscripts

b = backward window
f = forward window
set = set of sensors
i = time point t

Acronyms

ACA = anomaly classification algorithm
AD = absolute difference threshold (sensor voting)
ADA = anomaly detection algorithm
CR = contamination rate
DCIDS = detection classification and integrated diagnostics

DS = detection score
GT = gas turbine
RD = relative difference threshold (sensor voting)
SD = standard deviation

References

[1] Roumeliotis, I., Aretakis, N., and Alexiou, A., 2016, "Industrial Gas Turbine Health and Performance Assessment With Field Data," ASME Paper No. GT2016-57722.

[2] Simon, D. L., and Rinehart, A. W., 2016, "Sensor Selection for Aircraft Engine Performance Estimation and Gas Path Fault Diagnostics," ASME J. Eng. Gas Turbines Power, **138**(7), p. 071201.

[3] Venturini, M., and Puggina, N., 2012, "Prediction Reliability of a Statistical Methodology for Gas Turbine Prognostics," ASME J. Eng. Gas Turbines Power, **134**(10), p. 101601.

[4] Venturini, M., and Therkorn, D., 2013, "Application of a Statistical Methodology for Gas Turbine Degradation Prognostics to Alstom Field Data," ASME J. Eng. Gas Turbines Power, **135**(9), p. 091603.

[5] Hanachi, H., Liu, J., Banerjee, A., and Chen, Y., 2016, "Prediction of Compressor Fouling Rate Under Time Varying Operating Conditions," ASME Paper No. GT2016-56242.

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

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

[8] van Paridon, A., Bacic, M., and Ireland, P. T., 2016, "Kalman Filter Development for Real Time Proper Orthogonal Decomposition Disc Temperature Model," ASME Paper No. GT2016-56330.

[9] Hurst, A. M., Carter, S., Firth, D., Szary, A., and Van De Weert, J., 2015, "Real-Time, Advanced Electrical Filtering for Pressure Transducer Frequency Response Correction," ASME Paper No. GT2015-42895.

[10] Gutierrez, L. A., Pezzini, P., Tucker, D., and Banta, L., 2014, "Smoothing Techniques For Real-Time Turbine Speed Sensors," ASME Paper No. GT2014-25407.

[11] Ceschini, G., Gatta, N., Venturini, M., Hubauer, T., and Murarasu, A., 2017, "Optimization of Statistical Methodologies for Anomaly Detection in Gas Turbine Dynamic Time Series," ASME Paper No. GT2017-63409.

[12] Ceschini, G., Gatta, N., Venturini, M., Hubauer, T., and Murarasu, A., 2017, "Resistant Statistical Methodologies for Anomaly Detection in Gas Turbine Dynamic Time Series: Development and Field Validation," ASME Paper No. GT2017-63410.

[13] Ni, K., Srivastava, M., Ramanathan, N., Chehade, M., Balzano, L., Nair, S., Zahedi, S., Kohler, E., Pottie, G., and Hansen, M., 2009, "Sensor Network Data Fault Types," ACM Trans. Sens. Networks, **5**(3), p. 25.

[14] Boyce, M., 2011, *Gas Turbine Engineering Handbook*, Gulf Professional Publisher, Houston, TX.

[15] Taylor, B. N., 2009, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, DIANE Publishing, Washington, DC.