



AEROSPACE INFORMATION REPORT

AIR7999™

Issued

2020-10

Diagnostic and Prognostic Metrics
for Aerospace Propulsion Health Management Systems

RATIONALE

With the increasing application of diagnostic and prognostic technologies within the propulsion system community, there is a need for standardized metrics that can be used by developers and end users alike for assessing the performance of algorithms and systems for diagnosing faults and predicting the future health state of a system. The focus of this document is to introduce a variety of metrics that can be used for this purpose. Originally published in 2020, this document consolidates and expands upon the contents of two previous SAE E-32 Committee documents to create a single comprehensive document on diagnostic and prognostic metrics for Aerospace Propulsion Health Management Systems. The details of the other two documents are as follows:

AIR4985 - A Methodology for Quantifying the Performance of an Engine Monitoring System

AIR5909 - Prognostic Metrics for Engine Health Management Systems

With the publication of this document, the two superseded documents, AIR4985 and AIR5909, will be cancelled. This document consolidation strategy was agreed to by the SAE E-32 Committee to reduce the number of documents it is required to maintain.

FOREWORD

Health management technologies are becoming more common and are being tied directly to maintaining aerospace propulsion system reliability and performance. The implementation of these technologies is in turn supporting the more automated and integrated logistics concepts that are being fielded for many commercial and defense systems. Many airframers or end customers are imposing requirements on fault detection and isolation capabilities for propulsion health management systems. This document provides an overview and introduction to the metrics that can be applied for assessing the capability of diagnostic and prognostic technologies for aerospace propulsion systems and for comparing capability to requirements. This will include a discussion of the required definitions, necessary data sources and a comprehensive set of metrics to assess the performance and effectiveness of these methods.

SAE Technical Standards Board Rules provide that: "This report is published by SAE to advance the state of technical and engineering sciences. The use of this report is entirely voluntary, and its applicability and suitability for any particular use, including any patent infringement arising therefrom, is the sole responsibility of the user."

SAE reviews each technical report at least every five years at which time it may be revised, reaffirmed, stabilized, or cancelled. SAE invites your written comments and suggestions.

Copyright © 2020 SAE International

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of SAE.

TO PLACE A DOCUMENT ORDER: Tel: 877-606-7323 (inside USA and Canada)
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
http://www.sae.org

SAE WEB ADDRESS:

Full document details: SAE AIR7999

For more information on this standard, visit
<https://www.sae.org/standards/content/AIR7999/>

Documents

Document Sharing Hub

TABLE OF CONTENTS

1.	SCOPE.....	4
1.1	Purpose	4
2.	REFERENCES.....	4
2.1	Applicable Documents	4
2.1.1	SAE Publications.....	4
2.2	Related Publications	5
2.3	Definitions	5
2.4	Acronyms	6
2.5	Symbols	8
3.	DIAGNOSTIC AND PROGNOSTIC METRICS FOR AEROSPACE PROPULSION HEALTH MANAGEMENT SYSTEMS	9
3.1	Diagnostic and Prognostic Metrics.....	10
3.2	Propulsion System Diagnostics	12
3.3	Propulsion System Prognostics	15
3.4	Required Information and Data for Applying Diagnostic and Prognostic Metrics	17
3.4.1	Introduction of Select Diagnostic Metrics.....	17
3.4.2	Introduction of Select Prognostic Metrics.....	26
3.4.3	Cost Benefit Metrics	31
3.4.4	Computational Performance Metrics.....	32
4.	EXAMPLE APPLICATION OF DIAGNOSTIC METRICS	32
4.1	Description of Simulated Aircraft Engine Gas Path Diagnostic Problem.....	32
4.2	Description of Applied Diagnostic Methods	33
4.2.1	Data Conditioning.....	33
4.2.2	Fault Detection	33
4.2.3	Fault Isolation.....	33
4.3	Applied Diagnostic Metrics and Results.....	34
4.3.1	Fault Detection Metric Results	34
4.3.2	Fault Isolation Metric Results.....	37
5.	EXAMPLE APPLICATION OF PROGNOSTIC METRICS.....	38
5.1	Description of Simulated Aircraft Engine EGT Margin Forecasting Problem	38
5.2	Description of Applied Prognostic Methods	39
5.2.1	Method #1: Fleet Average EGT Margin Forecast	40
5.2.2	Method #2: Third Order Least Squares Fit	40
5.3	Applied Prognostic Metrics and Results	41
6.	USER CONSIDERATIONS FOR SELECTING AND APPLYING AEROSPACE PROPULSION HEALTH MANAGEMENT METRICS	45
7.	NOTES	46
7.1	Revision Indicator.....	46
Figure 1	Diagnostic and prognostic metrics classification	10
Figure 2	Multi-fault diagnostic process.....	12
Figure 3	Fault detection process time history	13
Figure 4	Fault isolation process time history.....	14
Figure 5	Distribution in condition indicator signal under nominal and faulty conditions.....	14
Figure 6	Illustration of intermittent condition indicator signal behavior	15
Figure 7	Illustration of trajectory prediction and RUL calculation.....	15
Figure 8	RUL versus time plot.....	16
Figure 9	Notional illustration of ROC curves.....	21
Figure 10	Illustration of TPR convergence for three different detection methods.....	25
Figure 11	Illustration of FPR robustness to increased signal noise	28
Figure 12	RUL versus time plot including uncertainty distributions of estimated RUL	28

Figure 13	Illustration of prediction uncertainty	29
Figure 14	Illustration of PH based on β -criterion results	30
Figure 15	α - λ accuracy with the accuracy bounds tightening with time	30
Figure 16	Fault detection ROC curves	34
Figure 17	Detection signal histograms and detection threshold placement	35
Figure 18	Fault detection confusion matrices	36
Figure 19	Confusion matrices	37
Figure 20	Illustration of EGT margin forecasting problem	39
Figure 21	Forecasting results of two methods applied to EGT margin forecasting problem	40
Figure 22	EGT margin versus flight number test cases (two individual engines plus the fleet average engine)	42
Figure 23	Estimated RUL versus horizon prior to EoL	43
Figure 24	Engine A PH results	43
Figure 25	Engine B PH results	44
Figure 26	Fleet average PH results	44
Table 1	Distribution in nominal and faulted engines	33
Table 2	ProDiMES sensed output parameters	33
Table 3	ROC AUC and AAC results	35
Table 4	TPR, TNR, FPR, and FNR detection metric results	36
Table 5	Fault detection metrics for WSSR fault detection method	36
Table 6	Fault detection metrics for 3 point median WSSR fault detection method	36
Table 7	CIR and kappa coefficient metric results	37
Table 8	Fault isolation metrics for WLS isolation with WSSR detection	38
Table 9	Fault isolation metrics for WLS isolation with median WSSR detection	38
Table 10	Fault isolation metrics for PNN isolation with WSSR detection	38
Table 11	Fault isolation metrics for PNN isolation with median WSSR detection	38
Table 12	Average bias, MAE, MAPE, and sample standard deviation results	41
Table 13	Estimated RUL results	42
Table 14	α - λ accuracy results	45
Table 15	RA and CRA results	45
APPENDIX A	Comprehensive List of Diagnostics Metrics	47
APPENDIX B	Comprehensive List of Prognostic Metrics	50
APPENDIX C	Cost Benefit and Computational Performance Metrics	55

1. SCOPE

This Aerospace Information Report (AIR) presents metrics for assessing the performance of diagnostic and prognostic algorithms and systems delivering propulsion health management functions.

1.1 Purpose

The purpose of this AIR is to present metrics that can be applied to assess the performance of diagnostic and prognostic methods. Various metrics are presented and discussed. Additionally, the application of select metrics is illustrated through examples.

2. REFERENCES

2.1 Applicable Documents

The following publications form a part of this document to the extent specified herein. The latest issue of SAE publications shall apply. The applicable issue of other publications shall be the issue in effect on the date of the purchase order. In the event of conflict between the text of this document and references cited herein, the text of this document takes precedence. Nothing in this document, however, supersedes applicable laws and regulations unless a finding of Equivalent Level of Safety or a specific exemption has been obtained from the governing regulatory authority.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or +1 724-776-4970 (outside USA), www.sae.org.

AIR5871	Prognostics for Gas Turbine Engines
ARP1587B	Aircraft Gas Turbine Engine Health Management System Guide
ARP5783	Health and Usage Monitoring Metrics Monitoring the Monitor
ARP6275	Determination of Cost Benefits from Implementing an Integrated Vehicle Health Management System

1.1.1 Other Documents

Reference 1: Davison, C. R., and Bird, J. W., "Review of Metrics and Assignment of Confidence Intervals for Health Management of Gas Turbine Engines," GT2008-50849, ASME Turbo Expo 2008: Power for Land, Sea, and Air, June 9-13, 2008, Berlin, Germany.

Reference 2: Davison, C. R., and Drummond, C., "Application of Cost Matrices and Cost Curves to Enhance Diagnostic Health Management Metrics for Gas Turbine Engines," ASME Journal of Engineering for Gas Turbines and Power, Vol. 132, Issue 4, April, 2010.

Reference 3: Gupta, J. S., Trinquier, C., Medjaher, K., and Zerhouni, N., "Continuous Validation of the PHM Function in Aircraft Industry," 2015 Prognostics and System Health Management Conference, October 21-23, 2015, Beijing, China.

Reference 4: Klein, L. A., "Sensor and Data Fusion Concepts and Applications," Second Edition, Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, WA, 1999.

Reference 5: Frederick, D. K., DeCastro, J. A., and Litt, J. S., "User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)," NASA/TM-2007-215026, October 1, 2007.

Reference 6: Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K., "Metrics for Offline Evaluation of Prognostic Performance," International Journal of Prognostics and Health Management, Vol. 1, Issue 1, April, 2010.

Reference 7: Simon, D. L., "Propulsion Diagnostic Method Evaluation Strategy (ProDiMES) User's Guide," NASA/TM-2010-215840, January 2010.

Reference 8: Volponi, A. J., "Gas Turbine Parameter Corrections," Springer, Cham, 2020.

2.2 Related Publications

The following publications are provided for information purposes only and are not a required part of this SAE Aerospace Technical Report.

Ashby, M. J., and Scheuren, W. J., "Intelligent Maintenance Advisor for Turbine Engines", IEEE Aerospace Conference Proceedings, vol. 6, pps 211-219, February 2000, Big Sky MT.

Kacprzynski, G. J., Hess, A.J., Begin, M., "Metrics and Development Tools for Prognostic Algorithms," 2004 IEEE Aerospace Conference, March 6-13, 2004, Big Sky, MT.

Koutsoukos, X., Biswas, G., Mylaraswamy, D. A., Hadden, G. D., Mack, D., Hamilton, D., "Benchmarking the Vehicle Integrated Prognostic Reasoner," 2010 Annual Conference of the Prognostics and Health Management Society, October 10-16, 2010, Portland, OR.

Peirce, C. S., "The Numerical Measure of the Success of Predictions", Science, Vol. 4, No. 93, pp. 453-454, November 1884.

Roemer, M. J., Dzakowic, J., Orsage, R. F., Byington, C. S., Vachtsevanos, G., "Validation and Verification of Prognostic and Health Management Technologies," 2005 IEEE Aerospace Conference, March 5-12, 2005, Big Sky, MT.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., Schwabacher, M., "Metrics for Evaluating Performance of Prognostic Techniques," IEEE International Conference on Prognostics and Health Management, October 6-9, 2008, Denver, CO.

Saxena, A., Celaya, J., Balaban, E., Saha, B., Saha, S., Goebel, K., "Evaluating Algorithm Performance Metrics Tailored for Prognostics," 2009 IEEE Aerospace Conference, March 7-14, 2009, Big Sky, MT.

Saxena, A., Celaya, J., Balaban, E., Saha, B., Saha, S., Goebel, K., "On Applying the Prognostic Performance Metrics," Annual Conference of the Prognostics and Health Management Society (PHM09), September 27—October 1, 2009, San Diego, CA.

Saxena, A., Celaya, J., Balaban, E., Saha, B., Saha, S., Goebel, K., "Evaluating Prognostics Performance for Algorithms Incorporating Uncertainty Estimates," Aerospace Conference, 2010 IEEE, March 6-13 2010, Big Sky, MT.

Simon, D. L., Borguet, S., Léonard, O., Zhang, X., "Aircraft Engine Gas Path Diagnostic Methods: Public Benchmarking Results," ASME Journal of Engineering for Gas Turbines and Power, Vol. 136, 041201-1-10, April 2014.

Uckun, S., Goebel, K., Lucas, P. J. F., "Standardizing Research Methods for Prognostics," IEEE International Conference on Prognostics and Health Management, October 6-9, 2008, Denver, CO.

Wheeler, K. R., Kurtoglu, T., Poll, S. D., "A Survey of Health Management User Objectives in Aerospace Systems Related to Diagnostic and Prognostic Metrics," ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, August 30—September 2, 2009, San Diego, CA.

2.3 Definitions

DIAGNOSTICS: The process of monitoring the current health state of a system, identifying if a fault has occurred, isolating the type and magnitude of any detected faults, and quantifying the level of confidence in the diagnosis declared.

FAILURE: A functional status or physical condition characterized by the inability of a system, piece of equipment, component, or subassembly to fulfill its design purpose; the most severe degree of malfunction or degradation.

FAULT: A condition whereby a system, piece of equipment, component, or subassembly is not able to perform as per specification. A fault may be indicative of an impending failure.

FAULT DETECTION: The process of detecting the occurrence of a fault in a system, piece of equipment, component, or subassembly.

FAULT ISOLATION: The process of identifying the type and magnitude fault in a system, piece of equipment, component, or subassembly.

HEALTH MANAGEMENT: The general discipline or techniques for detection, diagnosis and prognosis of degradation or fault of any systems, piece of equipment, component or subassembly and offering a maintenance action or decision support to address the degradation or fault.

HEALTH MANAGEMENT SYSTEM: A data acquisition, analysis, and information processing system that uses parametric and utilization data to determine the integrity of components and subsystems.

PROGNOSTICS: The forecast of future functional status and condition of a system, piece of equipment, component or subassembly based on current and accumulated inputs. The forecast can be expressed as remaining useful life, or time to reach a specific degradation level, or time to reach a point at which the risk of component failure is unacceptable.

2.4 Acronyms

AAC	Area Above Curve
ARMA	Auto-Regressive Moving Average
AUC	Area Under Curve
CIR	Correct Isolation Rate
C-MAPSS	Commercial Modular Aero-Propulsion System Simulation
CPU	Central Processing Unit
CRA	Cummulative Relative Accuracy
EGT	Exhaust Gas Temperature
EoL	End-of-Life
EoLP	Predicted End-of-Life
FDA	Fault Detection Accuracy
FDR	False Discovery Rate
FMEA	Failure Modes and Effects Analysis
FMECA	Failure Modes, Effects, and Criticality Analysis
FN	Number of False Negative Fault Detections
FNR	False Negative Rate
FOR	False Omission Rate
FP	Number of False Positive Fault Detections
FPR	False Positive Rate

HPC	High Pressure Compressor
HPT	High Pressure Turbine
IFSDR	In-Flight Shutdown Rate
LPT	Low Pressure Turbine
LRU	Line Replaceable Unit
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAR	Mission Abort Rate
MIR	Mis-Isolation Rate
MTBR	Mean Time Between Removals
MTTD	Mean Time To Detect
MTTI	Mean Time To Isolate
NPV	Negative Predictive Value
OEMs	Original Equipment Manufacturers
PH	Prognostic Horizon
PNN	Proababilistic Neural Network
PPV	Positive Predictive Value
ProDiMES	Propulsion Diagnostic Method Evaluation Strategy
RA	Relative Accuracy
ROC	Receiver Operating Characteristic
ROI	Return on Investment
RUL	Remaining Useful Life
SVR	Shop Visit Rate
TAT	Turn Around Time
TN	Number of True Negative Fault Detections
TNR	True Negative Rate
TP	Number of True Positive Fault Detections
TPR	True Positive Rate
UERR	Unscheduled Engine Removal Rate
UUM	Unit Under Monitoring

WLS	Weighted Least Squares
WSSR	Weighted Sum of Squared Residuals

2.5 Symbols

B	Average bias
C_{CI}	Component isolation coverage
C_D	Detection coverage
C_I	Isolation coverage
H	Horizon
i	Index representing time index t_i
$i_{\alpha\beta}$	The first time index when predictions satisfy the β -criterion for a given α -bounds
I_{pp}	Correct isolation of fault type “p” as fault type “p”
I_{pq}	Mis-isolation of fault type “p” as fault type “q”
k	Index representing the k^{th} fault test case
K	Number of fault test cases
l	Index representing the l^{th} UUM
L	Number of UUMs
N_{DF}	Number of detectable fault modes
N_{IF}	Number of fault modes which are only ambiguous with fault modes of the same component
N_{NF}	Number of non-ambiguous fault modes
N_{TF}	Total number of fault modes
p	Index representing the p^{th} fault type
$r_{actual}^l(i)$	Actual RUL at time i for the l^{th} UUM
$r_{predicted}^l(i)$	Predicted RUL at time i for the l^{th} UUM
S	Sample standard deviation
t	Time
t_D	Time index when fault is detected by diagnostic system
t_F	Time index when fault of interest initiates
th	Superscript representing an unspecified ordinal number

t_P	Time index when prognostic prediction begins
t_{EoL}	Actual time index when UUM reaches its end-of-life or failure threshold
t_{EoLP}	Predicted time index when UUM will reach its end-of-life or failure threshold
w	Weight factor
α	Accuracy bounds on prediction
α^+	Maximum allowable positive error
α^-	Maximum allowable negative error
β	Minimum desired probability mass threshold
$\Delta^l(i)$	Error, or difference between calculated and actual end-of-life for the l^{th} UUM at time t_i
κ	Kappa coefficient
λ	Time window modifier such that $t_\lambda = t_P + \lambda(t_{EoL} - t_P)$
ρ	Number of time indices where predictions are made between t_P and t_{EoLP}
ϕ^l	Prediction of variable of interest for the l^{th} UUM
$\phi^l(j i)$	Prediction of variable of interest for the l^{th} UUM at time t_j given data up through the time t_i
$\Phi^l(i)$	Trajectory of future predictions of ϕ^l defined as $\{\phi^l((i i), \phi^l(i+1 i), \dots, \phi^l(EoLP i))\}$
$\varphi(x)$	Non-parameterized probability distribution for any variable x
$\pi[RUL_{predicted}] _{\alpha^-}^{\alpha^+}$	Probability mass of the prediction residing within the defined α -bounds

3. DIAGNOSTIC AND PROGNOSTIC METRICS FOR AEROSPACE PROPULSION HEALTH MANAGEMENT SYSTEMS

Health management systems play a critical role in assisting operators in managing the reliability, availability, and affordability of their propulsion assets, which may also provide an intrinsic enhancement to safety. The functionality provided by a health management system includes the capability to evaluate the current health of the system (diagnostics) as well as the emerging approach of predicting or forecasting the future health of the system (prognostics).

The emphasis and need for effective diagnostic and prognostic methods continue to grow. The desire for condition-based maintenance, optimal scheduling and performance of maintenance, and fleet management are all technology pulls. Operators are facing health management technology design choices, and with these choices they require an effective means to compare the relative performance and benefits of different health management methods. The purpose of this document is to set forth a series of metrics that can be applied to quantify the performance of aerospace propulsion health management systems.

3.1 Diagnostic and Prognostic Metrics

Choice of health management evaluation methods are guided by end user objectives. There are many stakeholders of health management information, each with different goals and performance criteria. Consequently, the performance metrics can also be classified into related categories. In general, based on user objectives, the following stakeholders can be identified:

1. Operations (program manager, plant/fleet manager, mission operator, command and control, and maintainer, owner, insurer)
2. Regulatory (policy makers)
3. Engineering (designer, researcher, OEMs)

Each of these stakeholders find a different subset of metrics interesting to them.

Diagnostic and prognostic metrics can generally be partitioned into three broad categories as depicted in Figure 1, namely:

1. Algorithm and system performance metrics
2. Cost benefit metrics
3. Computational performance metrics

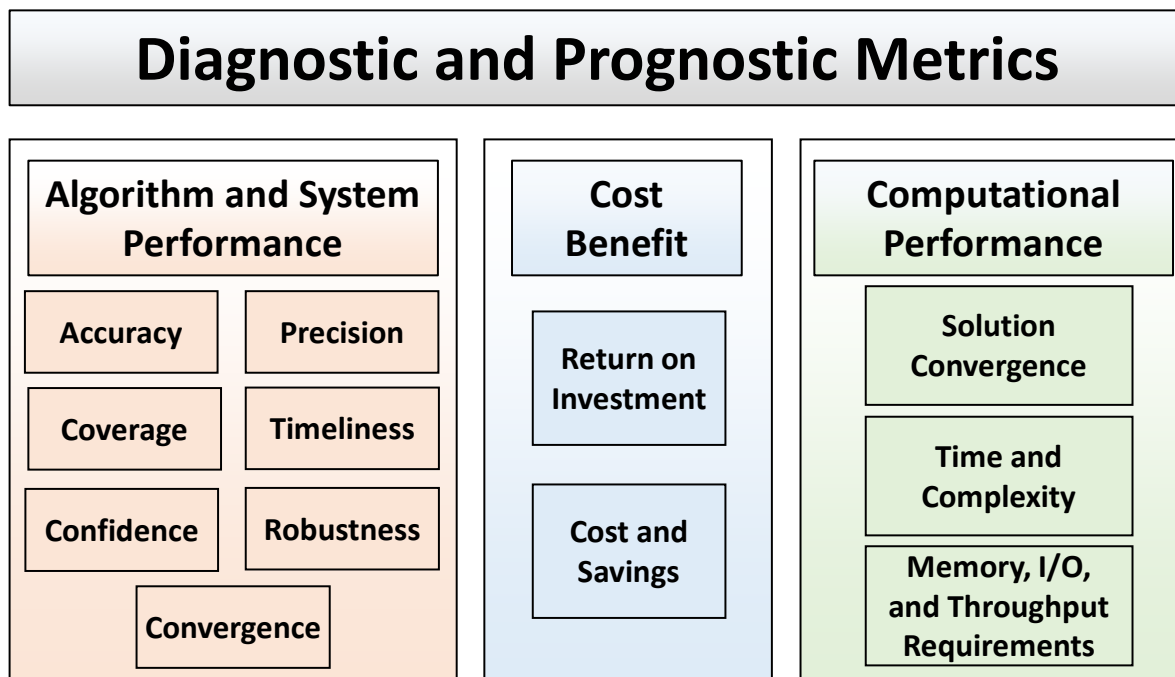


Figure 1 - Diagnostic and prognostic metrics classification

Algorithm and system performance metrics are the core set of metrics that measure the efficacy of a propulsion health management system by measuring accuracy, precision, coverage, and timeliness attributes. Furthermore, its performance is enhanced by the notion of robustness (to various disturbances) and convergence. The cost-benefit metrics are meta metrics that utilize algorithm and system performance metrics to assess effectiveness over an extended period in a larger context of system health management vis-à-vis investments made. Such metrics are interesting to system owners, fleet managers, program managers, and systems engineers. Cost-benefit metrics are useful for assessing effectiveness of diagnostics and prognostics systems, which can include on-board and off-board elements, over traditional maintenance strategies such as scheduled maintenance at the fleet level. These metrics are computed from maintenance records and logs, and are an important indicator of how a health management system is performing. Computational performance metrics offer trade space to optimize between performance, speed, and computational resources available to operate a health management system, and are of obvious interest to system designers.

A health management method is typically tasked with producing two basic results: predicting the current or future health state of the system and providing an associated measure of uncertainty with that prediction. Quantitative metrics of a health management method's performance take both results into account. There is currently no generally accepted agreement as to an appropriate and acceptable set of metrics that can be employed to unequivocally assess the technical performance of a given health management system. The metric classes depicted in Figure 1 provide a broad range of measures that should be adequate for most systems. These general classes can be further sub-divided into sub-categories as noted in Figure 1. The largest category, and indeed, the one that contains the most defined metrics appearing in this document, is algorithm and system performance.

Three performance attributes typically dominate the health management task and should be tracked by any set of metrics; they are correctness (*accuracy*, *precision*, and *coverage*) of predictions, timeliness (too late or well in time) of a prediction, and confidence measure in a prediction. In general, *accuracy* measures the closeness of the inferred health state to the actual health state, *precision* implies how close those predictions are bunched or clustered together, and *coverage* refers to the percentage of fault modes that can be detected or isolated. Most of the traditional metrics are centered around these measures of correctness and are presented in the Appendices for reference. Metrics are also commonly used to assess timeliness and confidence measures, as presented in subsequent discussions. The two other algorithm performance sub-classes, appearing in Figure 1 are meta-metrics, namely robustness and convergence. One way to view robustness is the effect on the sensitivity of predictions to algorithm parameter variations or external disturbances. For example, variation in sensor inputs at a given flight condition, variations in algorithm parameters used to express prior distributions or initial conditions, maintenance actions resulting in a sudden change in measurement data, or a change in the usage profile of the system. Robustness might also be interpreted as a measure of ease in adapting or changing the process to different applications. Convergence performance can be viewed from two perspectives; *prediction convergence* and *solution convergence*. The former is a measure of how any metric, such as accuracy or precision, improves with time as additional measurement data becomes available. In the case of diagnostics, this may reflect any disambiguation that occurs as a fault persists or increases in magnitude. For prognostics, this may reflect how fast an algorithm arrives at a meaningful solution as a system approaches the end of its useful life. The latter measure of convergence, *solution convergence*, arises in processes that are of an inherently iterative nature, where a closed-form solution may not exist, but the algorithm *converges* to a solution over time. In both cases, assurance of convergence i.e., it will not diverge, and how quickly does it arrive at its solution, are measures that need to receive appropriate weight when evaluating any algorithm. One may think of the sub-categories of robustness and convergence as being more qualitative than quantitative. Instead of attributing a numerical score, they might contribute a heuristic or *fuzzy* measure. For example, if a particular method is easily extendable to components or sub-systems other than the one currently under consideration, it might garner a *high* for robustness. Similarly, if a method relied on an iterative process where convergence was sometimes an issue (like some non-linear aero-thermo engine simulations), it may have a *low* or *moderate* rating for Solution Convergence. This makes robustness and convergence relative rather than absolute metrics.

The other elements of cost benefit and computational performance are no less important than the quantitative measures of health management system performance. A costly procedure having too large a memory footprint and high central processing unit (CPU) requirement, may be less desirable than a less accurate and less precise method that poses fewer memory, CPU throughput and cost constraints.

3.2 Propulsion System Diagnostics

Diagnostics entails monitoring the current health state of a system, identifying if a fault has occurred, isolating the type and magnitude of any detected faults, and quantifying the level of confidence in the diagnosis declared. As a point of introduction, consider the multi-fault class diagnostic process illustrated in Figure 2. Shown is a system that can operate in either a nominal state, or in a faulty state where it has encountered one of N potential fault types. Also shown is a diagnostic method applied to produce a diagnostic prediction of the current system state based on acquired sensed measurements. The diagnostic method consists generally of the three-step process of:

1. Data conditioning - processes the acquired system sensor measurements to produce condition indicator signal(s) used for fault detection and isolation purposes.
2. Fault detection - monitors condition indicator signals for threshold exceedance, which classifies the system as being in either a nominal or faulty state.
3. Fault isolation - invoked upon fault detection to classify the system as being one of the N possible fault types, and in some instances to also estimate the fault magnitude and the confidence level associated with the diagnostic inference.

Ongoing research is pursuing the development of empirical diagnostic algorithms based on machine learning techniques. With these more modern algorithms it might be possible to collapse the entire diagnostic process into a single step that produces an isolated fault diagnosis directly from raw measurement data. However, for the foreseeable future the more traditional diagnostic process is expected to dominate. The metrics discussion can be changed appropriately in case these more modern techniques are applied.

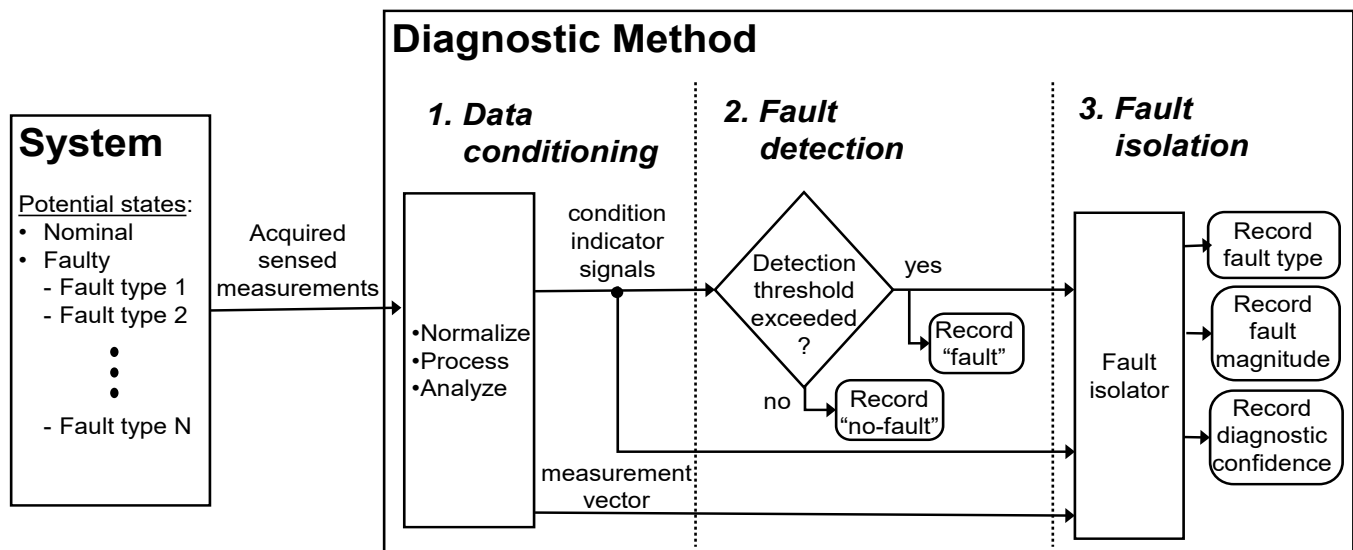


Figure 2 - Multi-fault diagnostic process

The fault detection step of the diagnostic process is further illustrated in Figure 3. The top half of this figure shows an example time history of a system condition indicator signal, produced at discrete time samples by data conditioning logic (Step 1). Fault detection logic compares the provided condition indicator signal against a defined fault detection threshold. If the condition indicator signal exceeds the threshold a system fault is declared by setting a detection signal to 1, otherwise a nominal, or "no fault," condition is indicated by setting the detection signal to 0. In this example, the system experiences a fault at time t_f and a fault is first correctly detected at time t_D . The bottom half of Figure 3 compares the detection signal (circles) against the ground truth condition (plus (+) symbols) of the system. Instances where the circles and plus symbols are concentric denote correct detection inferences. There are four possible detection decision process outcomes including:

- True positive: Fault declared when system is faulty
- False positive: Fault declared when system is nominal

- False negative: No fault declared when system is faulty
- True negative: No fault declared when system is nominal

The example shown in Figure 3 applies a single fault detection threshold, and any exceedance of this threshold results in a fault detection. In actual diagnostic applications, detection logic may be more complex. This may include the application of multiple detection thresholds to provide annunciations of different criticality such as “advisories” versus “alerts” or threshold exceedance persistency requirements to reduce nuisance false alarms. Regardless of the applied fault detection process, the metrics discussed in this document may be applicable.

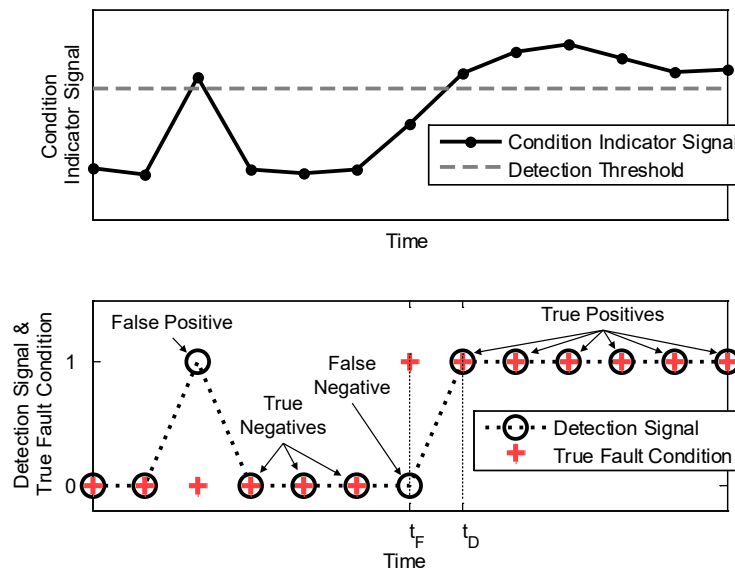


Figure 3 - Fault detection process time history

The fault isolation step of the diagnostic process is illustrated in Figure 4. This shows the isolated fault condition of the system (circles) against the ground truth condition (plus (+) symbols) for a system that can either be operating nominally (fault condition = 0) or experiencing one of N different fault types. Instances where the circles and plus symbols are concentric denote correct isolation inferences. In the example shown, the system experiences fault condition 4 at time t_F . Generally speaking, there are two potential isolation decision outcomes including either the correct identification or the mis-identification of the fault condition. These outcomes can be further partitioned in accordance with the true fault condition at the time of the isolation decision:

- True “No Fault” Condition
 - Correct Detection (True Negative): A no fault condition (0) is isolated when system is nominal (0)
 - False Positive: A fault condition (1 through N) is isolated when system is nominal (0)
- True “Fault” Condition
 - Correct Detection and Isolation: A fault condition (1 through N) is present and correctly isolated
 - Mis-isolation: A fault condition i (1 through N) is correctly detected as being present but incorrectly isolated as fault condition k (1 through N , $k \neq i$)
 - False Negative: A fault condition (1 through N) is present but incorrectly detected as the no fault condition (0)

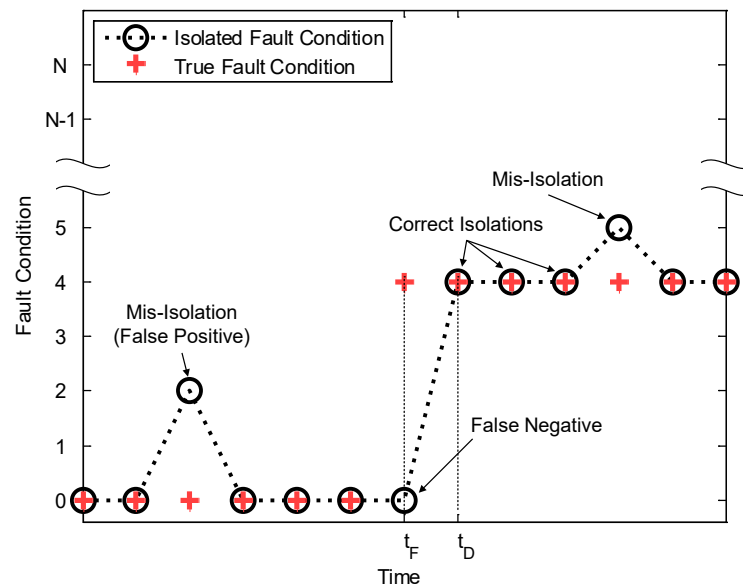


Figure 4 - Fault isolation process time history

A challenge in developing reliable machinery diagnostic methods is that the process is not deterministic. System measurement noise, variations in ambient conditions, operating load, deterioration, and nonlinear dynamics are all factors that contribute random variation to the process. This can lead to incorrect diagnostic predictions. For example, consider the fault detection logic shown in step 2 of Figure 2 that is tasked with performing the binary classification problem of declaring a system as either “nominal” or “faulty” based on a provided signal. The distributions in this signal under nominal and faulty operation are shown in Figure 5. Also shown is the placement of the threshold applied for detecting nominal versus faulty operation. Due to the overlap in the two distributions, it will not be possible to attain 100% fault detection decision accuracy. This is further illustrated in Figure 6 showing the condition indicator signal produced by a fault detection method applied to a system experiencing a fault at time t_F . This particular signal exhibits noisy unstable behavior contributing to intermittent exceedances of the fault detection threshold both before and after t_F . In such instances the implementation of persistency requirements on the signal may help in avoiding nuisance false positives or false negatives although this will add latency to the fault detection process. Detection performance will depend on multiple factors: the separation between the nominal and faulty distributions of the condition indicator signal, the applied detection threshold, and any persistency requirements of the signal for declaring a fault. Increasing the detection threshold and adding persistency requirements will have the desired effect of reducing false positives, but this will come at the expense of increasing false negatives or increasing detection latency. This can pose a dilemma in attempting to compare the merits of candidate diagnostic strategies, and effective metrics are required to assist end users in making such comparisons.

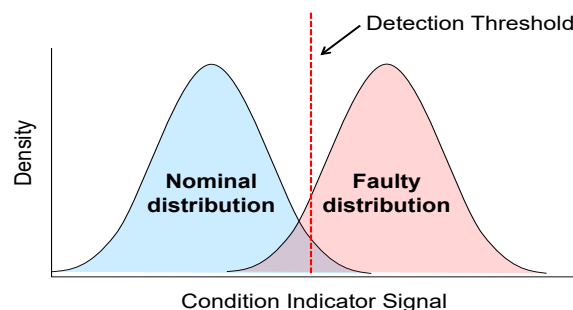


Figure 5 - Distribution in condition indicator signal under nominal and faulty conditions

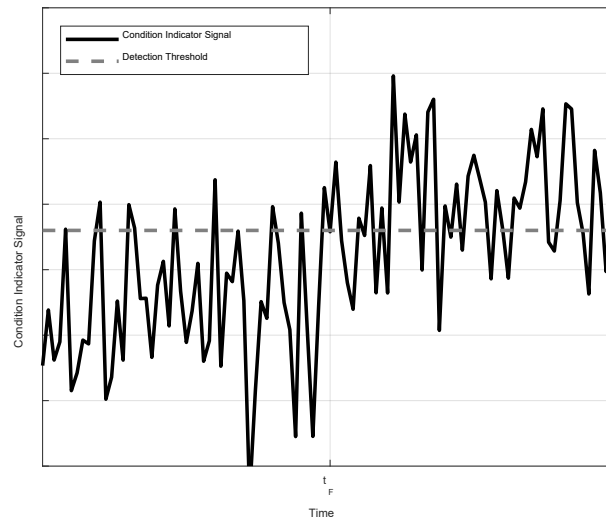


Figure 6 - Illustration of intermittent condition indicator signal behavior

3.3 Propulsion System Prognostics

Prognostics involves the forecasting of future functional status or condition of a system, piece of equipment, component or subassembly based on current and accumulated inputs. While the types of prognostic forecasts vary by application and approach, they inherently involve the calculation and forecasting of a variable of interest representing system health. Examples include a crack propagation length, a component effective cycle count, or remaining exhaust gas temperature margin. The point at which this variable reaches the threshold where an action must be taken is known as the End-of-Life (EoL) condition. The EoL condition does not necessarily equate to system failure: it could be the point where all useful life has been consumed, degradation has reached a specific level, or the risk of failure has become unacceptable. In this document, the remaining time until EoL is encountered is referred to as Remaining Useful Life (RUL). Figure 7 illustrates the prognostic process of forecasting the trajectory of a variable of interest through the point where it intersects its EoL threshold and the resulting RUL prediction. Here, the time index is denoted by i , and information is shown for the l^{th} Unit Under Monitoring (UUM), which may be a component, subsystem, or system. Calculation of the predicted variable of interest, denoted as ϕ^l , initiates at time t_P and is updated each time step as additional measurement data becomes available. $\phi^l(j|i)$ denotes a point prediction of the future value of this variable of interest at time t_j given data acquired through time t_i . A trajectory of future predictions of this parameter from time t_i onward is given as $\Phi^l(i) = \{\phi^l(i|i), \phi^l(i+1|i), \dots, \phi^l(EoLP|i)\}$. This trajectory is updated each time step as additional data becomes available. The difference between predicted EoL time, t_{EoLP} , and current time, t_i , defines predicted RUL: $r_{predicted}^l(i) = t_{EoLP} - t_i$.

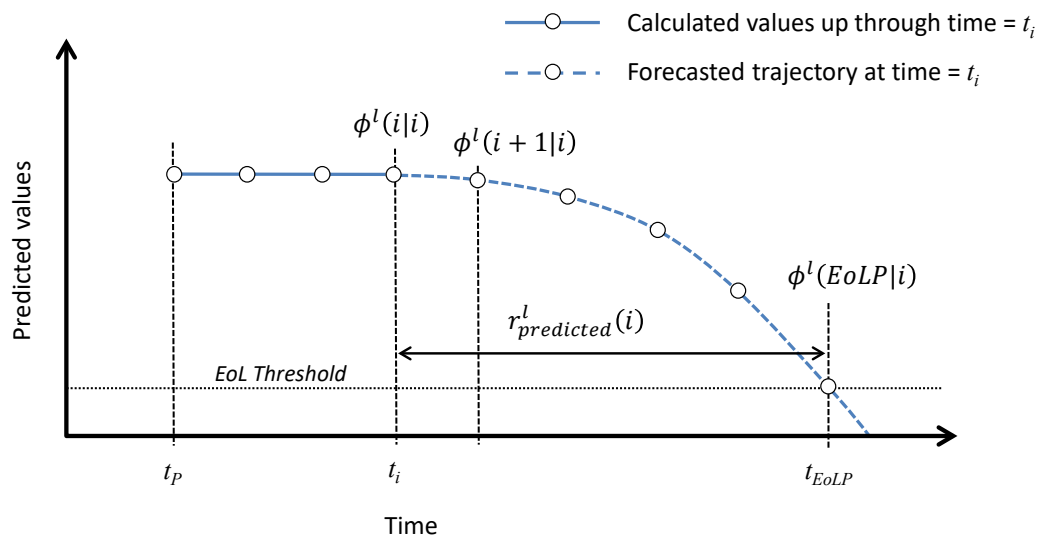


Figure 7 - Illustration of trajectory prediction and RUL calculation



Given the time history of a UUM's predicted RUL along with ground truth knowledge of its EoL, t_{EoL} , a "RUL versus time" plot can be generated as shown in Figure 8. This plot forms the foundation of most of the prognostic metrics presented in this document. Here predicted RUL is shown by the dashed line and actual RUL is shown by the solid line. When generating the RUL versus time plot, RUL data (y-axis) and time data (x-axis) are plotted in the same units (e.g., operating hours, flights, cycles, etc.). This results in actual RUL versus time (solid line) being a straight line with a slope of negative one. Prognostic predicted RUL versus time (dashed line) is updated over time on a regular basis as additional system measurement data becomes available. An individual point along this line, $r_{predicted}^l(i)$, represents predicted RUL at time i for the l^{th} UUM. At the end of its lifespan the UUM will have a deterministic t_{EoL} . This EoL time ground truth is never known *a priori* in real practice. However, for prediction performance assessment purposes, predictions are compared with this ground truth information. Correspondingly, individual points along the straight solid line, $r_{actual}^l(i)$, represent the actual RUL at time i for the l^{th} UUM. Figures 7 and 8 both show the prognostic process initiating at time t_P . This starting time for RUL prediction is application specific. In some prognostic applications RUL prediction may start immediately upon the UUM's introduction into service while in other applications t_P may be preceded by a specific triggering event such as the detection of a fault. In either case, the prognostic method is tasked with updating the RUL calculation from time t_P onward.

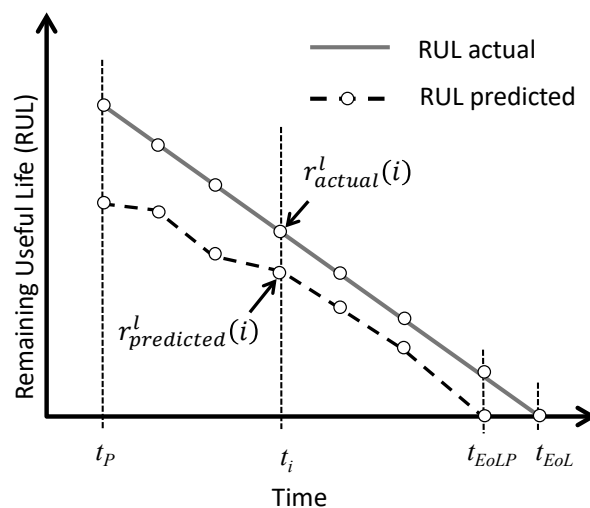


Figure 8 - RUL versus time plot

Another key aspect in prognostics is that of uncertainty in predictions. Any post-prognostic decision, whether safety or cost related, should be taken in light of uncertainty in predictions. Uncertainty in prognostics arises from inaccuracies in modeling, measurement noise, errors in state estimation, and uncertainties in future loading and operational conditions. None of these factors can be completely eliminated and therefore, prognostics algorithms now include adaption routines to characterize, represent, and propagate these uncertainties into the future estimates of system states. Consequently, prognostic performance metrics have become more sophisticated in dealing with prognostic uncertainty beyond initial concepts of precision in RUL probability distribution. For a complete discussion of prognostics for aerospace propulsion systems, readers are referred to AIR5871.

3.4 Required Information and Data for Applying Diagnostic and Prognostic Metrics

A requirement for the application of health management metrics is access to a rich database of system data, containing both nominal and anomalous operating scenarios. This includes data collected over the horizon prior to a fault or EoL event. Additionally, unambiguous knowledge of the ground-truth state of the system should be available. This includes knowing whether a fault exists or has initiated, the fault type and magnitude, and when the system will reach EoL. As aerospace propulsion systems are very reliable, and faults/failures do not occur very often, access to such data may be limited. These systems are also very expensive, which means that such data is opportunistic at best and typically not planned as part of certification testing. This challenge is further compounded by the fact that even when a fault does occur in service, the underlying cause or the maintenance action(s) taken to correct the event may not be fully known or well documented. In such instances it may be possible to conclude that the anomaly resides within a specific ambiguity group (i.e., a grouping of faults that have a similar signature or that occur within the same Line Replaceable Unit (LRU)). If so, one option is to apply metrics at the ambiguity group level as opposed to the individual fault level. If knowledge of the fault type or corrective action is completely unknown, the user may opt to assume that the event resides in an “unknown anomaly” category for the purposes of calculating metrics. It is also prudent to re-apply metrics as additional events are identified and more experience is gained with the use of a health management system. Given the challenges associated with obtaining in-service fault data, alternatives are to include simulated data, seeded fault test data, or run-to-failure data during the verification and validation phases of a propulsion system. Often, life estimates are based on detailed fatigue models which may be constructed from material characterizations in addition to limited and accelerated life testing. Data from seeded fault testing (e.g., damaged balls or races in bearings) are especially useful.

For diagnostics, the meaning of normal versus faulted, and the distinction between fault types, is usually clear. However, there are times when this distinction is not always clear. For example, how does one distinguish normal engine deterioration from a turbine fault? Setting of fault thresholds to indicate “not normal” is therefore important. It may also be important to have access to fault data of varying magnitudes. In situations where the ratio of available fault to no fault test cases does not reflect that expected in service, Bayesian inference may be applied to improve metric accuracy as will be discussed in 3.4.1.1. Often, the distinguishing feature of a gracefully degrading engine would be evident in key parameters trending in expected ways, whereas more sudden shifts in these parameters would indicate faults.

For prognostics, the reader may question the ability to ever know the true EoL condition from which RUL can be calculated and metrics imposed to quantify the efficacy of prognostic algorithms. This apparent conundrum can be relieved by relaxing the *literal* meaning of EoL (e.g., system failure) and instead defining EoL in other terms that can be readily observed either quantitatively or qualitatively. For example, we could invoke Figure 7 and use a measured parameter relative to a pre-defined threshold, wherein, reaching this threshold marks EoL. This is precisely the approach taken in performance monitoring with Hot Day Exhaust Gas Temperature (EGT) margin tracking. A zero margin can be thought of as EoL for certain applications. Likewise, one could use an expert assessment at time of maintenance (from teardown inspection) as to the relative condition of the component in question and the expert's assignment as to how much “life” the component has consumed based on observable damage. This latter example is a bit more subjective than the EGT margin case, but it illustrates how EoL can be re-defined by the end user to fit their situation and application. Once this is done, the metrics can be used and the interpretation modified accordingly to the EoL definition utilized. Further discussion of these points can be found in Reference 3.

3.4.1 Introduction of Select Diagnostic Metrics

This section will introduce and discuss a select number of diagnostic metrics to help familiarize the reader with their formulation and utility. This includes many of the classical diagnostic detection and isolation metrics commonly applied. A more extensive list of diagnostic metrics can be found in Appendix A.

3.4.1.1 Diagnostic Detection Metrics

Detection Confusion Matrix: A 2×2 matrix that reflects an algorithm's ability to discriminate between fault and no-fault cases. The confusion matrix's rows reflect the true fault condition and its columns reflect the diagnosed fault condition, or vice versa. The main diagonal reflects the number of correct predictions (true positives and true negatives) and its off-diagonal elements reflect the number of incorrect predictions (false negatives and false positives).