Electronic Prognostics Innovations for Applications to Aerospace Systems

Matthew Gerdes
Oracle Corporation
matthew.gerdes@oracle.com

Kenny Gross
Oracle Corporation
kenny.gross@oracle.com

Guang Chao Wang*
Oracle Corporation
guang.wang@oracle.com

Abstract—**Oracle has an anomaly detection solution for monitoring time-series telemetry signals for dense-sensor IoT prognostic applications. It integrates an advanced prognostic pattern recognition technique called Multivariate State Estimation Technique (MSET) for high-sensitivity prognostic fault monitoring applications in commercial nuclear power and aerospace applications. MSET has since been spun off and met with commercial success for prognostic Machine Learning (ML) applications in a broad range of safety critical applications, including NASA space shuttles, oil-and-gas asset prognostics, and commercial aviation streaming prognostics. MSET proves to possess significant advantages over conventional ML solutions including neural networks, autoassociative kernel regression, and support vector machines. The main advantages include earlier warning of incipient anomalies in complex time-series signatures, and much lower overhead compute cost due to the deterministic mathematical structure of MSET. Both are crucial for dense-sensor avionic IoT prognostics. In addition, Oracle has developed an extensive portfolio of data preprocessing innovations around MSET to solve the common big-data challenges that cause conventional ML algorithms to perform poorly regarding prognostic accuracy (i.e, false/missed alarm probabilities). Oracle's MSET-based prognostic solution helps increase avionic reliability margins and system availability objectives while reducing costly sources of "no fault found" events that have become a significant sparing-logistics issue for many industries including aerospace and avionics. Moreover, by utilizing and correlating information from all on-board telemetry sensors (e.g., distributed pressure, voltage, temperature, current, airflow and hydraulic flow), MSET is able to provide the best possible prediction of failure precursors and onset of small degradation for the electronic components used on aircrafts, benefiting the aviation Prognostics and Health Management (PHM) system.**

TABLE OF CONTENTS

1. INTRODUCTION

Prognostics and Health Management (PHM) solutions for mission critical systems require a comprehensive methodology for proactively detecting and isolating failure mechanisms, avoiding Type I and Type II errors (false- and missed-alarms), recommending and unambiguously guiding condition based monitoring (CBM) actions, and estimating in real time the remaining useful life (RUL) of critical components and associated subsystems [1]. Conventional surveillance methods that have been applied to mechanical assets (motors, pumps, valves, hydraulics, etc.) have enhanced the availability and serviceability of those assets; but now the elements of least reliability have been pushed into the computing elements, peripherals, and interconnecting networks, areas where conventional PHM approaches have heretofore not been successfully adapted for determining electronic component health, performance tolerances, or quantitative RUL estimation [2,3]. This mismatch in PHM capabilities between mechanical assets and increasingly predominant computing and control elements often results in apparent and actual decreases in overall equipment readiness and test program precision.

A common challenge for operators of complex aerospace and military systems is the early detection of incipient degradation of sensors and the components the sensors monitor in order to avoid unplanned outages, to orderly plan for preventative maintenance activities, and to assure that continued optimal quality-of-service (QoS) levels are attained. In such systems, there are usually a large number of sensors serving many functions, including input to control systems, monitoring of dynamic physical parameters and component performance limits, and system environmental conditions. Oftentimes, an embedded or offline PHM system is used to perform real-time analysis of data from sensors. Most of these diagnostic systems employ simple tests (e.g., threshold, mean value, etc.) that are sensitive only to gross changes in the process mean, or to large steps or spikes that exceed a threshold limit to determine whether or not a failure has occurred [4]. These methods can fail dramatically, especially in situations where noisy data are present or only a slight drift is noted prior to catastrophic failure.

To address these challenges and detect incipient faults in process equipment, at the earliest possible stage of development, it is necessary to analyze the characteristics of the noise carried by sensor telemetry signals (both wired and wireless) monitoring the process in addition to the mean values of these signals. Analyzing the noise is a requirement because small initial disturbances will cause subtle changes in the stochastic properties of sensor signatures, well prior to any measurable changes in the signal mean. The monitored variables are physical variables (distributed internal temperatures, currents, voltages, vibration, acoustics, etc). The telemetry data encompass an extensive array of environmental, performance, and reliability variables sampled at regular intervals. These variables provide a rich foundation for building PHM models for several system components individually as well as for the system as a whole [5–7].

This paper showcases a multivariate pattern recognition technique developed by Oracle for accurate electronic prognostics in aircraft assets. In addition, a data preprocessing system dedicated to time series telemetry data is introduced. It is composed of a set of ancillary techniques that can handle the imperfections in the real world telemetry data, including dramatic differences in sampling frequencies for the multitudes

*Corresponding author
978-1-6654-9032-0/23/$31.00 ©2023 IEEE

of sensors, disparate signal-to-noise ratios, the presence of missing values, and spikiness in classes of sensors. These data imperfections are often observed on the aircraft assets and cause conventional ML prognostic algorithms to perform poorly (i.e. low prognostic accuracy and high false positive and false negative decisions). The remainder of the paper is organized as follows. In Section 2, the multivariate time series anomaly detection technique is introduced. Section 3 introduces the ancillary techniques with detailed examples. Finally, Section 4 concludes the paper.

2. MULTIVARIATE ANOMALY DETECTION TECHNIQUE

Architecture of MSET

Oracle's ML prognostic technique is based on an innovative algorithm called Multivariate State Estimation technique (MSET [8]). MSET was originally designed to monitor all sensors and instrumentation in commercial nuclear power plants to increase safety margins and reduce Operations-and-Maintenance costs through predictive maintenance [5], but then has been spun off to NASA Space Shuttles, commercial avionics, and Navy Destroyers. Over the years, MSET has been evolved and scaled to the big data prognostic applications commonly seen in safety-critical industries including aerospace, utilities, and computer systems [7, 9].

The MSET framework consists of a training phase and a monitoring phase. The training procedure is used to characterize the monitored equipment using historical, error-free operating data covering the envelope of possible operating regimes for the system variables under surveillance. This training procedure evaluates the available training data and selects an optimal subset of the data observations that are determined to best characterize the monitored asset's normal operation. It creates a stored model of the equipment that is used in the monitoring procedure to estimate the expected values of the signals under surveillance. In the monitoring step, new incoming observations for all the asset signals used in conjunction with the trained MSET model to estimate the expected values of the signals.

Fault detection of the monitoring phase is performed by the Sequential Probability Ratio Test module (SPRT [10]). The SPRT technique proves to be sensitive not only to disturbances in signal mean, but also to the subtle changes in the statistical moments of the monitored signals. Instead of threshold limits that trigger faults when a signal increases beyond a nominal value, the SPRT technique is based on user-specified false-alarm and missed-alarm probabilities, allowing the end user to control the likelihood of missed detection or false alarms. For sudden gross failures of sensors or system components, SPRT divulges the disturbance as fast as a conventional threshold limit check; however, for slow degradation that evolves over a long time period, SPRT raises a warning of the incipience or onset of the disturbance before it would be apparent to threshold based rules.

Mathematical Derivation of MSET

The mathematical derivation of the latest MSET algorithm is outlined in this section.

The main objective of MSET is to make a quantitative assessment of the current operation status by using the degree of similarity between historical normal operating data and the current surveillance observations. First, the degree of similarity between two matrices A and B of the same column size is defined by $A \otimes B$, where $\otimes$ represents a proprietary non-linear matrix operator.

Assume the historical data **D** from the monitored system under normal operation consisting of m measurements and n sensors is available. A data subset D, consisting of m' measurements and n sensors that preserves prominent non-linear dynamic and inter-correlations between the sensors, is selected:

$$D = \begin{bmatrix} X_{1,1} & \dots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{m',1} & \dots & X_{m',n} \end{bmatrix} \in \mathbb{R}^{[\gtrdot' \times \ltimes]} \tag{1}$$

The pairwise correlation between the measurements in D can be quantified by:

$$D^\top \otimes D, \tag{2}$$

To minimize the Euclidean norm between the estimated and measured data vectors X^{obs}, a weight w is defined by:

$$w = (D^\top \otimes D)^* (D^\top \otimes X^{\text{obs}}), \tag{3}$$

where sign $*$ indicates pseudoinverse calculation, which can accommodate a singular matrix caused by two or more repeated or highly correlated sensor signals in the dataset (i.e., high collinearity).

To proceed, MSET estimates X^{est} are produced for new observations X^{obs} by:

$$X^{\text{est}} = D(D^\top \otimes D)^* (D^\top \otimes X^{\text{obs}}). \tag{4}$$

The residual errors between the MSET estimates and the actual observations are:

$$e = X^{\text{est}} - X^{\text{obs}}. \tag{5}$$

Finally, the residuals e go through SPRT, which is based on considering the log likelihood ratio (LLR) as a function of observation numbers:

$$\text{LLR} = \log \left[\frac{\prod_{i=1}^{n} P(e_i|\text{normal})}{\prod_{i=1}^{n} P(e_i|\text{abnormal})} \right]. \tag{6}$$

The SPRT algorithm quantifies both mean and variance shifts between the normal distribution and any degraded distribution to flag anomalies.

3. TIME-AWARE MACHINE INTELLIGENCE SYSTEM

The data pre-processing system that functions upstream of MSET is further introduced in this section. They consist of a set of techniques dedicated to the time series telemetry data which we call the Time-Aware Machine Intelligence (TAMI) system. It is designed to solve challenges emerging from the use of distributed-acquisition modules (DAQ) in electronic systems. For example, software clocks that are in widespread use generate a clock signal from MHz "ticks" of a microprocessor chip. Depending on the workload of the CPU or the Dynamic Voltage and Frequency Scaling (which varies the clock rates with instantaneous local temperatures) implemented in all chips, significant clock skew in the software clocks occurs. Another consequential instance of phase variability occurs when packets flow through complex networks in aviation assets. The packets flowing from the same source, to the same target device, through the wireless network mesh can arrive at different times depending on the paths through the network mesh. The arrival time discrepancies are minuscule but can have a substantive effect on the prognostic ML algorithms capacity to discover incipient problems in hardware and software systems.

The reason that TAMI is important to advanced prognostics is that ML Intelligence (in the form of high accuracy diagnostics, prognostics, and root-cause-analyses) is heavily dependent on whether Event A precedes Event B, or is exactly simultaneous with B. Relying on notions of "uniform wall-clock" time is insufficient. Thus, multiple techniques around TAMI were developed to address these issues.

There are three key foundational technologies in the development of the TAMI system [11–13]. These technology areas solve increasingly complex "time aware" challenges in various industries including aerospace, which can all be divided into the following phases of time-aware solution portfolios:

(a) Upsampling and downsampling of telemetry streams with disparate instrumentation sampling rates. It is common that the ML community uses simplistic "interpolation" routines. Conventional analytical interpolation is unsatisfactory for ML prognostics of mission-critical and safety-critical assets, whereas imputation utilizing MSET2 is more robust.
(b) Time asynchronies resulting from clock-mismatch issues and from variable clock skew issues in DAQ aggregators.
(c) Non-uniform flow of the fundamental "time dimension". For many use cases, assuming a uniform flow of "wall-clock time" throughout complex engineering assets is naive. What is required as a foundational principle for advanced Machine Learning prognostics is an internal explicit time dimension as a building block for synchronous network knowledge representation.

In this section, we showcase a few of the most prominent "time aware" challenges observed from the aerospace assets, along with the corresponding solution.

Monitoring Complex Systems with Disparate Sampling Rates

The expansions of dense-sensor applications across many industrial segments are fast growing in the past decade. For example, a typical large commercial airplane has 7,500 sensors. Across avionics and other IoT industries, it is quite common that different sensors (e.g. temperatures, voltages, currents, fan speeds, vibration levels, rotation speeds for rotating machinery, etc) are sampled at drastically different sampling frequencies, yielding a set of telemetry signals with non-uniform sampling rates; however, ML prognostic algorithms cannot tolerate non-uniform sampling rates for signals. The catalyst and foundation for a robust TAMI infrastructure is creating uniform sampling rates for all the signals under surveillance, in a manner that does not influence the accuracy of the ML algorithms.

The conventional approach to address non-uniform sampling rates is to apply interpolation based methods, wherein the non-uniform sampling signals are first upsampled through an interpolation method on a univariate basis to be able to work with any time-series ML technique. Unfortunately, for prognostic health monitoring and proactive/predictive maintenance applications, these signals yield sub-optimal prognostics because each signal goes through an independent interpolation process, which unavoidably creates correlation discrepancies between the individual upsampled signals. Whether an ML user conducts simple linear interpolation (e.g. mean between the last measured value and the next measured value), or utilizes more sophisticated interpolation algorithms such as, exponentially weighted moving averages, cubic splines, and inverse Lagrangian interpolation, the interpolation approaches have zero prognostic value. When the gap between samples is filled in with interpolation, the resulting observation is autocorrelated to the anterior and subsequent observations, but contains no correlation to any event that may have transpired during the time between samples. Therefore, considering that the measurements are for the same assets occurring at the same time, the preferred method is to analyze the underlying correlations among the signals during the interpolation process. Best practice is to assume there is prognostic significance that occurs between samples in the slower sampling rate signals (including irregularly sampled signals). Hence, the impetus for developing an ML based imputation method for upsampling non-uniform sampling-rate signals is utilizing the inferential physical correlations between the signals [14].

Given any dataset, the imputation technique first identifies the fastest sampling rate signal and computes its sampling rate with the timestamps. Then it sorts all of the slower signals and temporarily performs an interpolation-based upsampling technique (Figure 1). The values will be subsequently improved in the systematic iterative imputation procedure to fill in the temporal gaps between the slower signals and the fastest signal. This first step is essentially a shape-preserving piecewise cubic-spline interpolation that fills the missing observations in the slower signals with a temporary starting values. The innovative part of our imputation technique comes after this interpolation. We split the uniform sampling dataset to two subsets A and B of equal size.

For subset A, an MSET model is trained to characterize the internal physical correlations between the signals. The model is then deployed to generate MSET estimates for subset B. Subsequently, the measurements that were generated by the cubic interpolations are updated and replaced with the corresponding MSET estimates, which constitutes a new subset B' which is more accurate and possesses closer alignment with what the measurements would be if the sensors were sampled at a faster speed. With this imputed subset B, second MSET model is trained to produce estimates for subset A. The previously interpolated values in subset A are updated by the corresponding estimated values, resulting in a new subset A', thereby updating subset A with more accurate imputed measurements. Further iterations for cross training and estimation yield better imputation performance, but the incremental improvements are greatest in the early iterations and diminish with each subsequent iteration. It has been determined empirically that 5 iterations converges to optimal accuracy. After the 5th iteration subsets A' and B' are concatenated to reconstitute an upsampled data set with a uniform sampling rate and higher accuracy.

To illustrate this process, an example signal from a dataset with missing values is presented in Figure 1. The figure contains 4 subplots. The first subplot is the example signal, in blue, the missing indexes which are initially interpolate and represented by the orange dot. The first step in the process is illustrated in the second subplot form the top. On the left hand side we see the portion of the signal that will be used for training the MSET model. On the right hand side we see the portion of the signal for which estimates will be made. After the first monitoring step the value's in purple will be updated with the estimated values displayed as yellow markers in the third subplot. In the third subplot the portion of the signal that contains the newly imputed values will be used in the training set where the initial training set is now used during monitoring. The last subplot is the entire signal after all of the missing values have been imputed. The results, when used for training an ML model, contain more prognostic information than simply interpolating on a univariate level. The detailed schematic for this MSET based imputation for non-uniform sampling rate datasets is presented in Figure 2.

Processing Correlated Signals with Uniform Phase Disparities

Use of telemetry associated with physical variables (distributed temperatures, voltages, currents, vibration levels, fan speeds, power, integrated energy), can bring many benefits to aerospace system monitoring, including enhancements to availability, serviceability, performance, capacity planning, QoS, and security. However, the raw telemetry metrics often have disparate sampling rates (may be significantly out of phase), and the phase shifts between and among signals that are time varying with internal system loads. For dynamic time-series signal analyses, any types of ML computations involving differences of signals or ratios of signals incur substantial inaccuracies in the computed results, if there is even a slight lead-lag incoherence present between signals.

As a simple demonstrative example to motivate the subsequent developments: consider two time series variables that are in synchronization and that vary with approximately identical sinusoidal dynamics. The difference of these two variables produces a residual function that is stationary in time. Similarly, the ratio of these two time series produces a quotient time series that is stationary in time. However, if there was a small lag between the variables, caused by latency in wireless sensor and DAQ network, the difference and the ratio become severely contaminated with serial correlation. This serial correlation introduces substantial error terms in analytical computations involving the time series variables, unless it could be fixed "up-stream" of the analytical engine [15].

Time-Domain Synchronization of Telemetry Signatures—This section presents a novel means that combine telemetry time series with a pattern recognition technique called a Parity Space algorithm for optimal synchronization of any number of hardware clocks, with any number of software clocks (typically on the domain side) [16]. One hardware-based solution is to subscribe to commercial clock-synchronization hardware/software systems that will periodically synchronize a network of distributed clocks to a remote highly-accurate atomic clock. This approach became popular in the previous decade, in spite of the significant cost to design in clocks with either network or wireless capability for periodic re-sync updates. While effective this clock-sync methodology has fallen out of favor for many business-critical facilities after a few well-publicized hacks occurred when hackers discovered they could tap into facility critical assets through the clock-sync connections (which require a penetration through the facility-network firewall). Additionally, alternative hardware solutions have been developed that use "GPS clocks" and "radio clocks". Such approaches should eliminate clock-drift problems in future networks that implement these new approaches. Unfortunately, it would be a prohibitively expensive undertaking to retrofit legacy networks, with new clock capability which is the impetus for the investing in a algorithmic solution.

The time-domain technique developed in this section exploits a pattern recognition approach called parity space to produce residual time-series for each hardware/software clock in the system. The residuals are monitored with a "Detector" module–another consumer module–as part of the telemetry framework. The Parity Space Detector generates a time-sync validation flag that is stored along with other telemetry data. Time stamp sequences are used in a Parity Space (PS) algorithm. A PS approach averages N available time series (in this case the N clock signals) and computes a "residual function" for each individual clock by subtracting that signal from the realtime mean.

Upon initialization, the residual functions are normalized to zero. This means that if there is an initial offset among the various system clocks, the telemetries-PS algorithm will not attempt to correct the initial offset(s). It will record and then zero out that offset in the residuals; then from that point forward it will function as a time-sync validation detector. If any of the N residual functions should exceed a configurable threshold, a time-sync flag will be set to zero and a warning will be generated (via the same PA framework as other telemetry generated warnings).

Many variables of interest most often have disparate sampling rates, are not synchronized, and may have time-varying coherence. It is difficult to synchronize the samples from disparate signal sources, and even more difficult to assure phase coherence between/among the data streams. As stated, and solved, in the previous section the first problem is obtaining uniform sampling rates while the second problem is that even if the sampling rates can be made exactly uniform, the processes being monitored can be out of phase. For solving the phase misalignment, there are three options developed by Oracle: correlogram, Cross Power Spectral Density (CPSD), and a Genetic Algorithm (GA). This section discusses the correlogram (correlation coefficient vs lag) analysis, which operates in the time domain. The correlogram determines the optimum lag at which two signals are aligned the closest, i.e., the lag at which the correlation (absolute value) is the highest. One signal is picked as the 'reference anchor' signal, meaning its timestamps will be assumed to be correct. All other signals in the asset or fleet of assets will be empirically aligned to the "reference anchor" signal by computing pairwise cross correlation coefficients, then systematically "adjusting" the lags for the individual signals to optimize the correlation coefficient with respect to the "reference anchor". This step is required even though the signals may have common time stamps due to various reasons mentioned previously.

In Figure 3, two signals with identical sinusoidal dynamics were generated with Fourier composites. One signal, in orange, has an additional lag time inserted to simulate the phase shift that can occur for the myriad reasons expounded upon above. The correlogram analysis is performed matching the out of phase signal #2 (orange), to signal #1(blue) by iteratively measuring the correlation coefficient between signal #1 and shifted copies of signal #2 (middle plot). The non-zero lag corresponding to the highest correlation is captured and used to shift the signal #2, yielding the synced two signals (bottom plot). In a multivariate set of signals the correlogram analysis is performed for every pair of signals

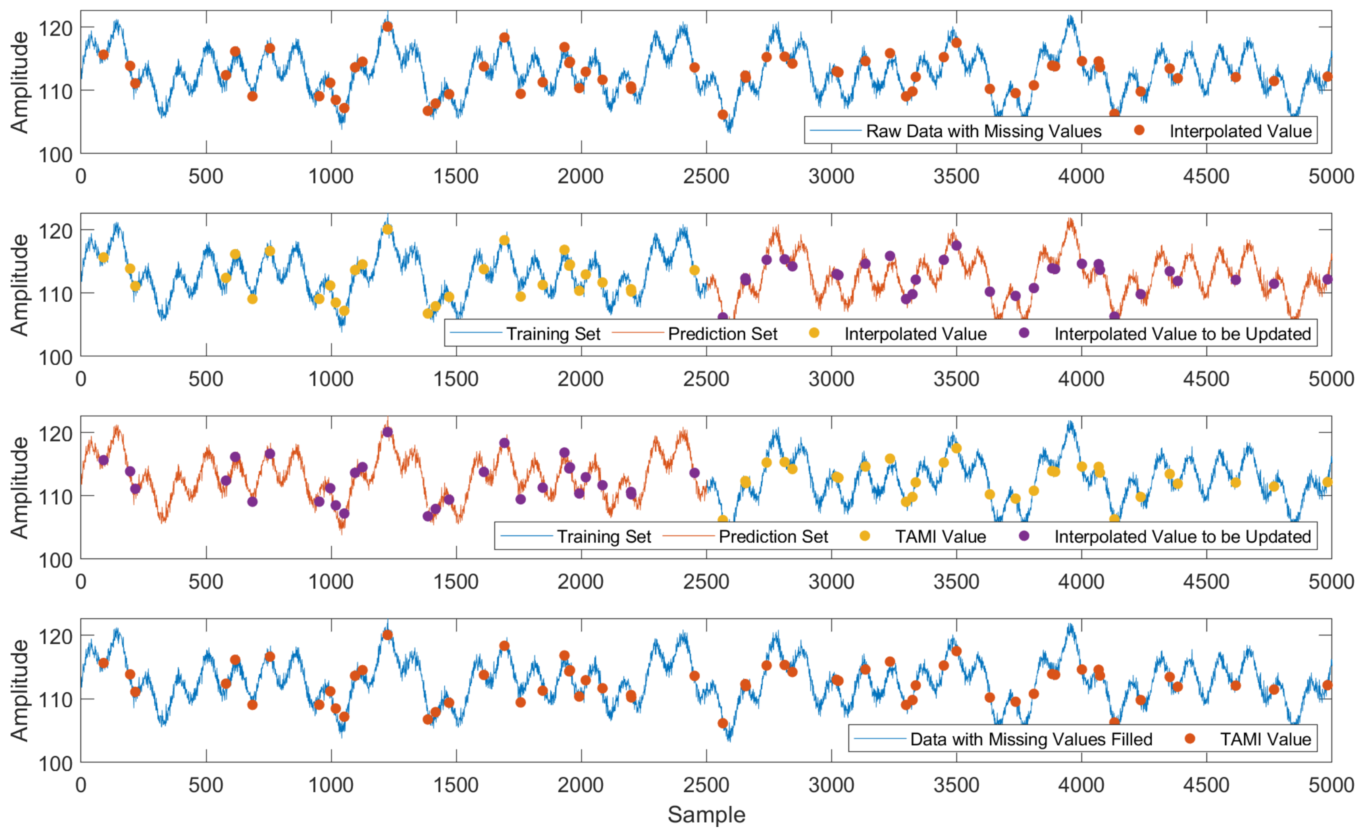


Figure 1: **Illustration of the TAMI imputation process.**

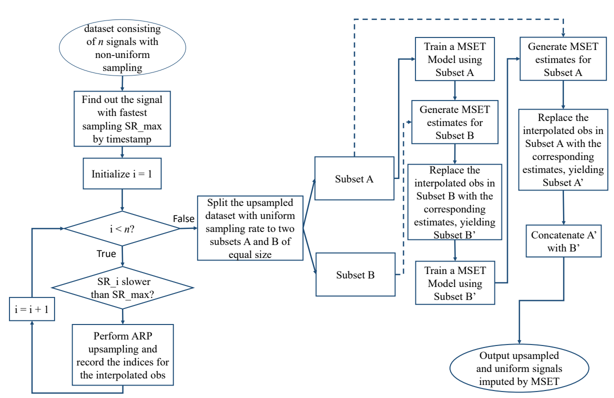


Figure 2: **Flowchart of MSET based imputation for upsampling varying sampling rate dataset.**

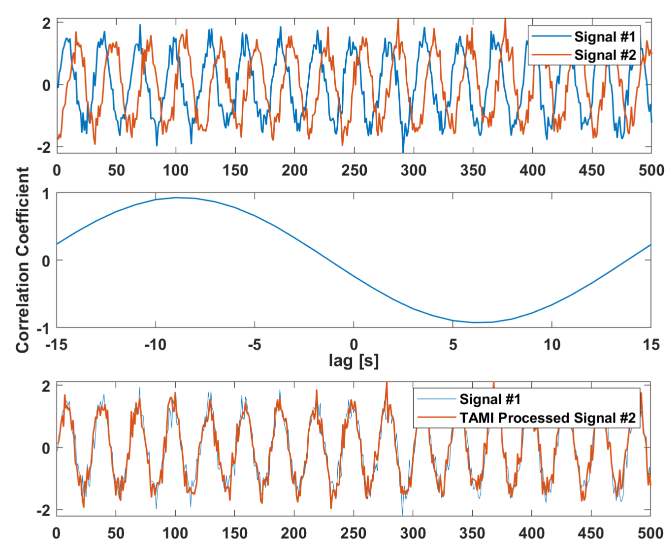
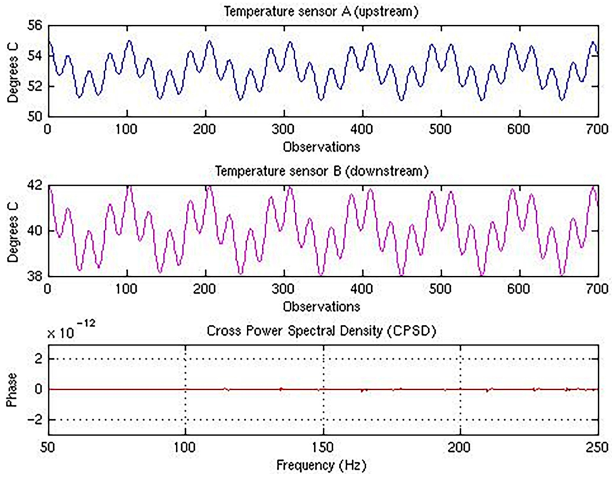
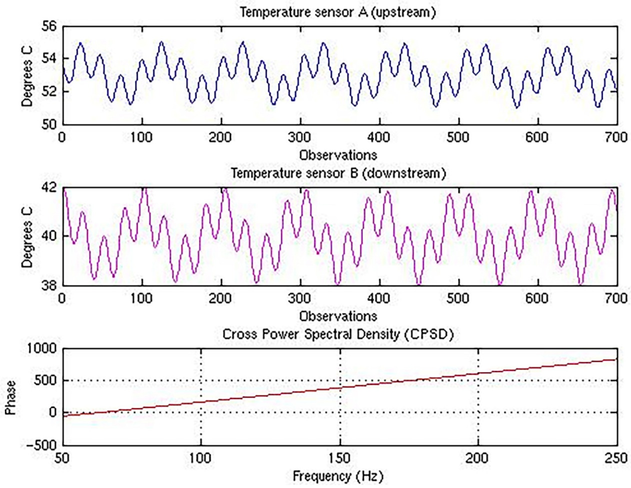


Figure 3: **The phase differences between two correlated signals are corrected by the correlogram technique in the TAMI system.**

Frequency-Domain Synchronization of Telemetry Signatures:— The technique presented in the previous section addresses the challenge of optimizing the alignment of telemetry signals with the analysis performed in the time domain. The technique presented in this section, CPSD, comes at the challenge from a different perspective:a signal transformation into the frequency domain occurs. CPSD is a bivariate frequency-domain technique that exploits a sophisticated Fast Fourier Transform (FFT) computation inferring, with high accuracy, the "phase angle" (in the frequency domain) between two time-series. An algorithm is then employed to compute an optimal estimate of the lag time from the phase angle. Pairwise computations are performed for all signals in the collection, adjusting signals to bring the empirical lag times to zero. This is an alternative approach and one that results in significantly improved accuracy for optimal signal alignment when the raw input telemetry signals contain any degree of "quantization". Most legacy digital assets contain 8-bit A/D converter chips that produce physical variables with low resolution and significant quantization. Conventional time-domain approaches using correlation coefficients result in sub-optimal phase coherence when the input telemetry signals are quantized



(a)



(b)

Figure 4: **A phase shift between signal A and signal B occurs (a), and is corrected after the CPSD analysis (b).**

In this analysis, the phase angle is plotted against frequency, then the slope of the phase-vs-frequency curve indicates whether the original telemetry signals are aligned. A non-zero slope in the phase angle vs frequency line indicates that the two signals are not aligned. The phase angle is treated as an adjustable parameter in a systematic stepwise iterative algorithm that is recursively applied until a zero slope indicates that the signals are perfectly aligned. As with correlogram analysis, CPSD is required even though the signals may have common time stamps. As an illustration of the inventive technique, examples below show the CPSD (phase angle vs frequency) for two signals when there is a non-zero lag (non-zero slope) and when the telemetry metrics are brought into optimal phase alignment with a zero lag (zero slope).

The novel frequency-domain technique has been developed for analytical resampling and phase shift optimization of telemetry signals coming from server telemetry but has been applied to signals from many other industries including aerospace. The advantage that this technique brings is that monitoring can be performed more efficiently and accurately, even when the signals are dynamically varying and even when those signals are contaminated with significant degrees of quantization from low-resolution A/D chips used in most legacy computing systems. Conventional time-domain signal synchronization techniques can presently only measure metrics while all signals are held constant with time. If the signals are varying dynamically, then any small time shifts due to clock skews between the external hardware power meter, the temperature measurements, and the OS throughput and performance variables, causes very large uncertainties with conventional methods.

Genetic Algorithm for Phase Shift Synchronization— The Genetic Algorithm (GA), like the correlogram, is conducted in the time domain but unlike the correlogram it does not rely on brute force. The GA is an iterative process whereas in each iteration, signals are given random perturbations in the positive or negative directions, after which the overall synchronization score for the population is evaluated. Perturbations that result in an improvement are retained for the next "generation," whereas ones that worsen the degree of correlation are moved back to the positions they held in the previous generation. Perturbation sizes are systematically reduced in each generation to ensure convergence to the optimum synchronization value.

Figure 5 displays the sequential steps of the algorithm when applied to an example pair of signals that exhibit correlation. In the top subplot the signals are presented as they were originally recorded, it is visually apparent that they are not synchronous in the time domain. Moreover, they are quantitatively asynchronous as the cross-correlation coefficient (CCF) between the two signals is approximately zero. The second subplot illustrates the signals after the first iteration of the GA. Signal A, in blue, was utilized as the reference signal while Signal B, in red, was randomly shifted forward in time by 5 timestamps. The absolute value of the CCF after the first shift increased to 0.88 indicating an increase in correlation. The next iteration is displayed in the second subplot after signal B has been randomly shifted forward in time again. Once again there is an increase in the absolute CCF. The third step, in the third subplot, randomly shifts signal B backwards in time by one unit in time landing on the maximum absolute CCF. An intermediary step (not pictured) subsequent to the third subplot that shifts signal B both forward and backward in time again. If the absolute CFF decreases in both directions the algorithm determines the maximum has been reached. Once the maximum absolute CCF is located the algorithm checks the sign of the CCF. If the sign is negative the algorithm shifts the signal back in time half a period aligning the signals in the time domain. If the is CCF positive no action is taken. The result of this last step is displayed in the fourth subplot and as such the CCF is positive and close to 1.

To illustrate complexity improvements of the GA, when compared to a brute force approach, a simple experiment was conducted whereby the number of signals was increased to 8. In turn the phase shift window size is 9 and the GA converges in 30 iterations. While by comparison the brute force method, terminates after 240 iterations.

Processing Correlated Signals with Dynamic Phase Misalignment

All operating systems speed up and slow down the CPU executions based on the utilization dynamics [17]. The aerospace industry is not immune to this issue. As the integration of automation and electronic monitoring into aerospace equipment increases so does the increase in CPUs. This issue can be especially problematic for telemetry used in condition based monitoring. For example, if we run a dynamic

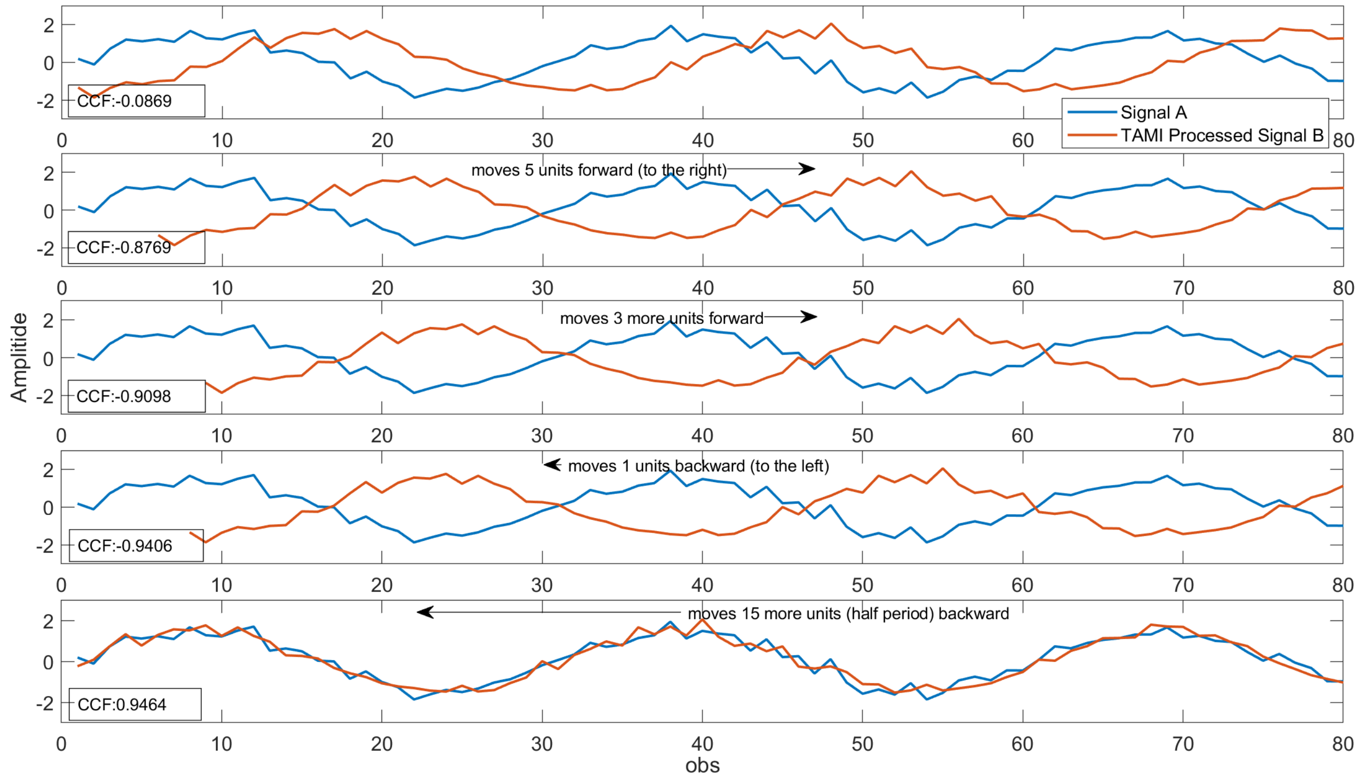


Figure 5: **The sequential steps of the Genetic Algorithm (TAMI Processed) for a pair of signals.**

stress script on two identical DAQs, then overlay the dynamic telemetry time series graphics on the same plot, and if we synchronized the front end of the dynamic profile, the back end of the dynamic signal is not synchronized. Conventional real-time phase synchronization techniques are unable to solve this issue. Figure 6 is a typical example illustrating the artifact of the variable speed up/slow down nature of OS. The whole Signal B is being synchronized with Signal A, it is always the case that part of Signal B observations are aligned with Signal A while the rest will be still out-of-phase in reference to the corresponding parts of Signal A. Then in the downstream process in the ML prognostic algorithm Signal B will generate false-alarms and will leave open the interpretation of those alarms and whether they come from variable time step executions, or from a possible unknown developing fault in the asset.

As demonstrated in the previous sections the synchronization post-processing autonomously corrects for conventional, constant, phase shifts in real-time but is unable to solve the challenge discussed above because the synchronization assumes lockstep temporal executions and consequential a constant lead/lag time between the signals of interest. When there is a dynamically varying lead/lag relationship in which the Linux (and other OS) dynamically speed up / slow down, a different synchronization technique is required.

To overcome the above challenges that increase false alerts and missed alerts in any ML techniques, the "Compression-Dilation Time-series Phrase Synchronization Framework," was developed. The framework transforms the signal of interest by dynamically and iteratively shrinking/expanding and resampling the signal before performing a global real-time phrase synchronization technique in a sliding window fashion. The output is a reconstituted signal of interest which aligns

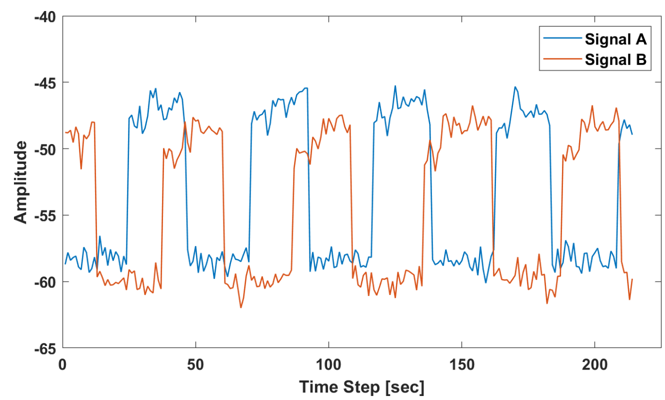


Figure 6: **Signal A and Signal B are produced by two identical variable-step execution OS. Signal B is in varying out-of-phase in reference to Signal A along the time axis.**

with the reference signal throughout the time axis, and both signals now have uniform equally spaced sampling intervals.

The detailed procedure is illustrated in a flowchart in Figure 7. The flowchart begins with reference signal A and varying out-of-phase signal B. The technique then employs a moving time window to segment both signals, and then each segment of signal B will be going through an optimal compression or expansion process by a "transform factor" from 80% to 120% of the segment through phase-synchronization and analytical resampling technique. Real measurements outside the present window segment are used to support the resampling process as needed. The transformed segment will then be shifted by a set of pre-defined time lags through the Correlegram technique and computing at each step the correlations to the corresponding segment of Signal A. The maximum correlation is selected and the associated lead/lag time and transform factor are determined optimally and used to reconstitute that segment of signal B. Then the time window is moved one step ahead and the above process is recursively repeated until the end when all the reconstituted segments are combined to yield the final reconstituted signal B.

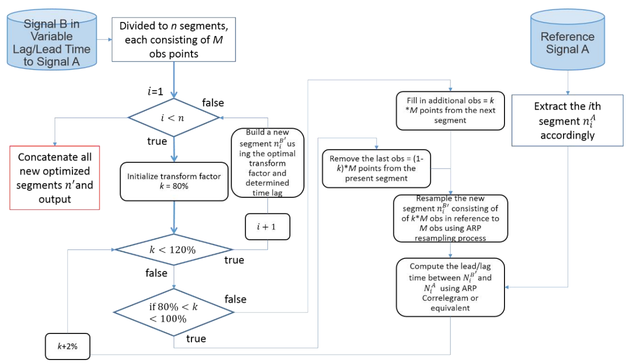


Figure 7: **Flowchart of Compression/Dilation timeseries signal phase synchronization for variable phase signals.**

To demonstrate the performance of the Compression and Dilation technique, Figure 8 compares the results of the algorithm using the same signals in Figure 6 with the outcome of the typical time synchronization technique. The reconstituted signal B (green) by TAMI correctly aligns with the original reference signal A with much better alignment than the common solution (yellow).

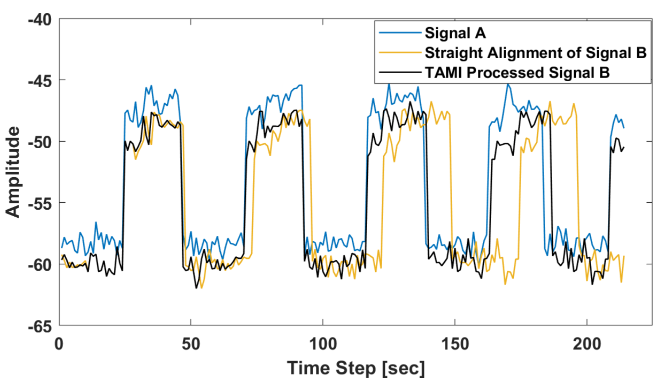


Figure 8: **Signal A and B responding to the same load source which suffers from the variable time lag executions, so they are out of synchronization in varying degree. The simple time synchronization approach aligns the two signals from the front end to the back end, causing inconsistent alignment, while the compression and dilation approach (black) corrects the varying phase shift.**

This variable out-of-phrase situation will certainly cause false alarms and missed alarms in the ML prognostic applications.

4. Conclusions

Anomaly Detection is the critical success factor in the aviation Prognostic and Health Management system that detects the incipience or onset of degradation mechanisms in any monitored on-board subsystems, since it reports actionable prognostics to anticipate when maintenance is required, rather than the classical preventive approach in which activities are planned on a regularly scheduled basis for the maintenance. Moreover, sustaining operations for complex aircraft systems also requires rapid processing and leveraging the sensor data in the avionics and internal aircraft systems to improve operational readiness. A reliable multivariate machine learning prognostic product suite that can handle imperfect data from sensors is crucial. This paper describes Oracle's multivariate pattern recognition technique for prognostic applications coupled with a portfolio of data pre-processing innovations for monitoring time-series telemetry signals for optimal prognostics performance. It provides the best possible prediction of failure precursors and onset of small degradation for the electronic components used on aircraft, benefiting the aviation Prognostics and Health Management system.

REFERENCES

[1] S. Ofsthun, "Integrated vehicle health management for aerospace platforms," *IEEE instrumentation & measurement magazine*, vol. 5, no. 3, pp. 21–24, 2002.

[2] C. Kwan, B. Ayhan, J. Yin, X. Liu, P. Ballal, A. Athamneh, A. Ramani, W. Lee, and F. Lewis, "Real-time system condition monitoring using wireless sensors," in *2009 IEEE Aerospace conference*. IEEE, 2009, pp. 1–8.

[3] K. Gross, K. Whisnant, and A. M. Urmanov, "Prognostics of electronic components: Health monitoring failure prediction time to failure," in *Proc. New Challenges in Aerospace Technology and Maintenance Conf. 2006*, 2006.

[4] K. Shaukat, T. M. Alam, S. Luo, S. Shabbir, I. A. Hameed, J. Li, S. K. Abbas, and U. Javed, "A review of time-series anomaly detection techniques: A step to future perspectives," in *Future of Information and Communication Conference*. Springer, 2021, pp. 865–877.

[5] K. C. Gross and W. Lu, "Early detection of signal and process anomalies in enterprise computing systems." in *ICMLA*, 2002, pp. 204–210.

[6] K. Gross, R. Singer, S. Wegerich, J. Herzog, R. VanAlstine, and F. Bockhorst, "Application of a model-based fault detection system to nuclear plant signals," Argonne National Lab., IL (United States), Tech. Rep., 1997.

[7] K. C. Gross and M. Li, "Method for improved iot prognostics and improved prognostic cyber security for enterprise computing systems," in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. The Steering Committee of The World Congress in Computer Science, 2017, pp. 328–334.

[8] R. M. Singer, K. C. Gross, J. P. Herzog, R. W. King, and S. Wegerich, "Model-based nuclear power plant monitoring and fault detection: Theoretical foundations," Argonne National Lab., IL (United States), Tech. Rep., 1997.

[9] K. Gross and G. C. Wang, "Ai decision support prognostics for iot asset health monitoring, failure prediction, time to failure," in *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2019, pp. 244–248.

[10] A. Wald, *Sequential analysis*. Courier Corporation, 2004.

[11] K. Baclawski, E. S. Chan, D. Gawlick, A. Ghoneimy, K. Gross, and Z. H. Liu, "Self-adaptive dynamic decision making processes," in *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE, 2017, pp. 1–6.

[12] K. Gross, K. Baclawski, E. S. Chan, D. Gawlick, A. Ghoneimy, and Z. H. Liu, "A supervisory control loop with prognostics for human-in-the-loop decision support and control applications," in *2017 IEEE conference on cognitive and computational aspects of situation management (CogSIMA)*. IEEE, 2017, pp. 1–7.

[13] K. Baclawski, E. S. Chan, D. Gawlick, A. Ghoneimy, K. Gross, Z. H. Liu, and X. Zhang, "Framework for ontology-driven decision making," *Applied Ontology*, vol. 12, no. 3-4, pp. 245–273, 2017.

[14] K. C. Gross, K. Vaidyanathan, A. Bougaev, and A. Urmanov, "Round-robin staggered-imputation (rrsi) algorithm for enhanced real-time prognostics for dense-sensor iot applications," in *Proceedings on the International Conference on Internet Computing (ICOMP)*. The Steering Committee of The World Congress in Computer Science, Computer . . . , 2016, p. 169.

[15] K. Baclawski, A. Chystiakova, K. Gross, D. Gawlick, A. Ghoneimy, and Z. H. Liu, "Use cases for evaluation of machine based situation awareness," in *2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE, 2019, pp. 107–113.

[16] G. Wang and K. Gross, "Telemetry parameter synthesis system for enhanced tuning and validation of machine learning algorithmics," in *IEEE 2018 Intn'l Symposium on Internet of Things & Internet of Everything (CSCI-ISOT), Las Vegas, NV*, 2018.

[17] J. Kim, M. M. Sabry, D. Atienza, K. Vaidyanathan, and K. Gross, "Global fan speed control considering non-ideal temperature measurements in enterprise servers," in *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2014, pp. 1–6.

BIOGRAPHY

Matthew Gerdes *received his BSc degree from the California State University Northridge in 2017, and a M.Sc. degree from University of California San Diego in 2020. He is currently an ML researcher at Oracle Labs and his current research interests include time series, anomaly detection, machine learning, and vibration analysis.*

Kenny Gross *is a Machine Learning Architect in Oracle Labs. Kenny has introduced numerous advanced pattern recognition and prognostic machine learning innovations that are now in use for Internet-of-Things applications in the fields of utilities, avionics, manufacturing, enterprise data centers, and life sciences. Kenny has 294 US patents issued and pending on his innovations, 226 scientific publications. Kenny's Ph.D. is in nuclear engineering from the University of Cincinnati.*

Guang Chao Wang *received his M.Sc. degree in Aerospace Engineering at The Ohio State University, and earned his Ph.D. in Energy Science from UC San Diego. He is currently a Machine Learning Researcher in Oracle Labs and has extensive experiences in time series anomaly detection. His other research interests include time series forecasting, distributed energy resources planning, and photovoltaic power modeling.*