A Supervisory Control Loop with Prognostics for Human-in-the-Loop Decision Support and Control Applications

K. C. Gross*, K. Baclawski†, E. S. Chan‡, D. Gawlick‡, A. Ghoneimy‡ and Z. H. Liu‡
*Oracle Corporation, San Diego, CA USA
†College of Computer and Information Science
Northeastern University, Boston, MA USA
‡Oracle Corporation, Redwood City, CA USA

***Abstract*—This paper presents a novel tandem human-machine cognition approach for human-in-the-loop control of complex business-critical and mission-critical systems and processes that are monitored by Internet-of-Things (IoT) sensor networks and where it is of utmost importance to mitigate and avoid cognitive overload situations for the human operators. The approach is based on a decision making supervisory loop for situation awareness and control combined with a machine learning technique that is especially well suited to this control problem. The goal is to achieve a number of functional requirements: (1) ultra-low false alarm probabilities for all monitored transducers, components, machines, systems, and processes; (2) fastest mathematically possible decisions regarding the incipience or onset of anomalies in noisy process metrics; and (3) the ability to unambiguously differentiate between sensor degradation events and degradation in the systems/processes under surveillance. The novel approach that is presented here does not replace the role of the human in operation of complex engineering systems and processes, but rather augments that role in a manner that minimizes cognitive overload by very rapidly processing, interpreting, and displaying final diagnostic and prognostic information to the human operator in a prioritized format that is readily perceived and comprehended.**

***Index Terms*—Prognostics; decision support; machine learning; sensor networks; situation awareness; control**

I. Introduction

When a human operator is controlling a system, it is essential for the operator to have good situation awareness (SA) so that the operator can take appropriate actions. This is especially difficult for complex systems that have a very large number of sensors and control parameters. It is even more difficult, or even impossible, for a human operator to control a complex system during an emergency situation because such situations put considerable pressure on the operator to make decisions very rapidly. Consequently, it is very important to avoid the onset of cognitive overload by human operators. Furthermore emergency situations typically produce multiple faults, numerous alarms, conflicting data and incomplete, inaccurate or missing information due to failing or failed sensors. There are now systems that have been developed to assist human operators to achieve situation awareness. Examples include the Situation Awareness Assistant (SAWA) for battlefield SA [1] and the National Information Sharing Consortium (NISC) for the emergency management community and first-responders [2]. However, these systems do not address all of the issues that arise in complex systems and emergency situations. SAWA, for example, does not have a mechanism for dealing with failing or failed sensors that can produce inaccurate, misleading or no data at all. NISC is only concerned with information sharing and interoperability and does not address cognitive overload or the problem of failing or failed sensors.

In this paper, we present a novel approach to human-machine cognition for human supervisory control applications that addresses both the system complexity and sensor failure issues discussed above. Our approach is intended to assist the operators of complex systems in two ways. First, by helping the operators to deduce the state of the system from the (possibly faulty) sensor data. Second, by providing expert advice on possible actions, even in the face of incomplete knowledge. Our approach is based on situation theory and uses information fusion technologies to help solve these two problems.

This paper is organized as follows. In Section II, we present our approach to supervisory control assistance. This approach involves two main components, one to deal with the complexity problem and another to deal with faulty sensor data. We elaborate on each of these components in the next two sections. The first component is the subject of Section III, and the second is the subject of Section IV. As our approach utilizes machine learning (ML) techniques, we compare the techniques we use to other ML techniques in Section V. The method whereby our approach deals with missing and faulty data is presented in Section VI. We conclude in Section VIII. Our approach has been validated in the case of nuclear power plant supervisory control, and we use this scenario and similar ones to illustrate our approach throughout the paper.

II. The Supervisory Control Loop with Prognostics

Our approach to supervisory control is based on the work of John Boyd who explored situation awareness in the 1960s. John Boyd's analysis was formulated as the OODA loop (Observe, Orient, Decide, and Act) [3] involving human cognitive processes. Boyd's work was done long before the discovery and evolution of many of the modern knowledge intensive modeling, reasoning, and ML technologies. We combined the modern AI technologies with the OODA loop to formulate the Situation Awareness CARE Loop, named after the four modes of reasoning: termed Classification, Assessment, Resolution, and Enactment in the Situation Awareness Loop [4]–[6]. We present the formalization of the OODA loop in Section III.

It has been known since the inception of human-machine cognition science that it is extremely improbable that humans will receive and be confused by 2 or more events creating only one alarm, rather, the situations that initiate human operator confusion and rapidly lead to cognitive overload scenarios, are where a single event creates multiple alarms [7]. The novel system reported in this paper implements an advanced pattern recognition module, called the Multivariate State Estimation Technique (MSET) [8], which is also known in the literature as Similarity Based Modeling [9] This system can rapidly process, interpret, and display final diagnostic and prognostic information to the human operator in a prioritized format that is readily perceived and comprehended. While MSET is an ML technique, MSET differs from other ML techniques, such as neural networks (NNs) [10] and support vector machines (SVMs) [11], with respect to human comprehensibility. This allows the very rapid sequences of hypothesis evaluations to identify and filter invalid/extraneous/faulty alerts (e.g., from degraded, failed, or missing sensors) to be offloaded from the human to the ML decision aid. As a result, the human can maintain the highest state of situation awareness throughout system-upset events. MSET has previously been applied successfully for operator decision aids and for humans administrating complex engineering assets [8], [12]–[14]. In this paper we extend the value proposition for MSET by integrating it with KIDS for avoidance of cognitive overload scenarios for general human-in-the-loop supervisory control of complex hardware-software assets. We present more details about MSET in Section IV.

Model-based reasoning is reasoning about the behavior of a system using a highly accurate nonlinear, non-parametric model based on the structure and function of the system. Ideally, well constructed models will also aid in providing explanations of the state and behavior of the system. This is an essential capability for realizing the activity graph in KIDS ontology. This capability allows a human operator to "scroll back" along the sequences of decisions that were made, whether in real time while events are evolving, or in a post-incident analysis to construct a detailed root-cause-analysis of an episode. Furthermore, this capability is critical for systems that have human safety significance (e.g., nuclear plants, oil/gas production facilities, transportation and avionics systems). The capability is also extremely vital in situations where the asset owners want to learn from degradation scenarios, to better mitigate/avoid similar events in the future. Section VI discusses this capability in more detail.

Expert systems are sophisticated computer programs that manipulate knowledge to solve problems efficiently and effectively in a narrow problem area. An expert system provides high-level expertise to aid in real-time cognitive problem solving as part of enhanced situation awareness for the human operator [3]–[5]. The expertise (knowledge) should be explicit and accessible. Two capabilities of expert systems that are particularly important in this work are predictive modeling and "root cause" sequence-of-event reconstruction. A vital element of the root cause explanation is disambiguation between false alarms (also called Type-I errors in statistical process control), from real anomalous behavior in the monitored systems or processes. A predominant cause of false alarms in conventional machine-learning prognostics is the fact that most conventional surveillance methodology is threshold based. Section VII below shows how threshold-based ML prognostics result in either lower sensitivity for annunciation of anomalies, or higher false alarm rates, or both. MSET is integrated with a sequential probability ratio test (SPRT) and overcomes the endemic problem of lower prognostic sensitivity versus higher false alarm rates, producing a human decision aide that mitigates alarm complexity and minimizes the prospect of cognitive overload for human operators.

It is in these kinds of real-time problem-solving situations that many of the limitations of humans are at their most apparent. Their tendency to overlook relevant information, to respond too slowly and to unavoidably succumb to some degree of panic when the rate of information flow is too great all contribute to lower than desired levels of performance. It is the goal of the research presented in this paper to provide effective decision support in order to transform the environment from an inefficient, data-intense, high cognitive demand situation to an efficient, knowledge intensive, information-rich, high-performance human-machine system. Such decision support is designed to enable a human decision maker to maintain peak situation awareness during an emergency situation.

III. KIDS Situation Awareness

KIDS manages data, knowledge, and processes. KIDS identifies four categories of data (Facts, Perceptions, Hypotheses, and Directives) and four categories of knowledge for reasoning (Classification, Assessment, Resolution, and Enactment) as well as the Relevance reasoning to select the data that form the situations for subsequent steps in the reasoning process. Facts constitute sensor measurements, which are quantitative. Facts are classified to derive perceptions, which are compact qualitative interpretations of facts designed to be easily comprehended by the human brain. Perceptions are assessed to derive one or more hypotheses, which are used to derive directives. Directives are action plans proposed to resolve the symptoms and root causes of the anomalies. Directives are enacted by control systems or agents in the external world. The enactment of the directives will create new facts. The loop starts again until the problem on hand has been solved. The CARE loop is a dynamic system driven by the deductive (classification), abductive (assessment), and inductive (resolution) reasoning processes. The reasoning processes are applied in the proper sequence as shown in Figure 1.

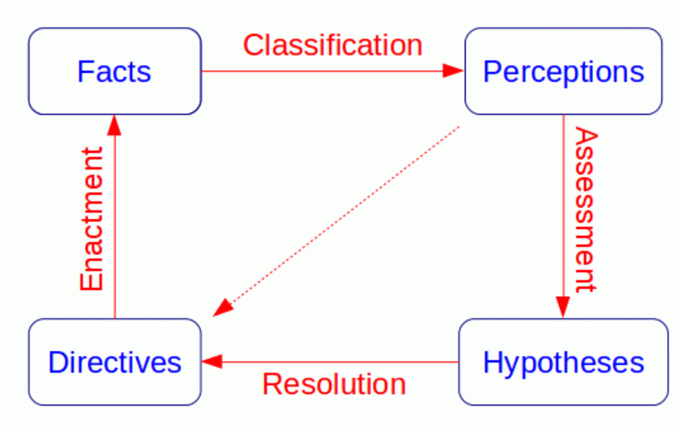

Fig. 1. Diagram of the KIDS CARE loop

The four categories of data require state of the art data management. This starts with the support of structured and unstructured data and with extensibility for domain specific data. Temporal support with provenance and selective long term retention is a must as is real time processing of queries (rules, models). The management of data has to be done with rich declarative interfaces. On-line transaction processing as well as real-time analytics have to be handled concurrently and with great efficiency. The relevance of data for different situations must be filtered efficiently with highly scalable expression filters.

The **Classification** knowledge – transforming facts into perceptions – is primarily represented by **deductive** reasoning. Some Classification knowledge that produces prediction or norm may involve inductive reasoning as well. The computation model for classification includes Auto-Associative Memory Models [15] (including Auto-Associative Neural Networks, Auto-Associative Kernel Regression, Auto-Associative MSET), Support Vector Machines [16], naïve-Bayes classifier, Clustering, Association Rules, Decision Trees, Cognitive computing, etc.

The **Assessment** knowledge – transforming perceptions into hypotheses – is typically implemented by **abductive** reasoning [17] that derives the *Hypotheses* from *Perceptions*. The computation model for assessment includes SPRT [18], Bayesian Belief Network [19], and Least-Squares Optimization or Regression of solutions for inverse problem [20].

The **Resolution** knowledge – transforming hypotheses into directives – involves **inductive** reasoning and **making decisions** under the uncertainty of outcomes by considering the relative merit of the different outcomes and the associated payoffs/costs. The computation model for resolution includes Bayesian Belief Network extended with decision nodes and payoff/cost nodes, known as Influence Diagrams [21], Dempster-Shafer theory [22], [23], Decision Trees, and Prognosis of Remaining Useful Life.

The **Enactment** knowledge – transforming directives into actions in the outside world – consists typically of **control structures** encoded in scripts, plans, schedules, workflows, and business processes. If the capturing of new facts is part of the enactment the fact management will be notified to be able to supervise the timely delivery of the expected facts. The enactment activity involves:

- **Supervise the compliance to directives**: The timely delivery of expected facts is a fundamental requirement to understand the current situation. The task is to supervise simple rules; which is a task that is well understood.
- **Verify the quality of the facts**: Collecting facts is bound to be an erroneous process. Facts may be incomplete (measurements are done but there were problems in the transmission), they may be stale (typically due to transmission delays), may be wrong (due to sensor errors). The task of verifying the quality of the data requires a technology such as MSET and the Mimir data reduction and analysis tools [24]. MSET has been proven to be a powerful tool that can identify issues in sensor data and swap in a highly-accurate inferential variable to correct the data. Big Data systems typically exhibit varying levels of quality (due to missing, delayed, or incorrect data). To derive the high-level Perceptions, the unstructured data is typically transformed into a relational model by Extract-Transform-Load (ETL) operations and by relational join operations. Mimir introduces a database query primitive called lenses, which provide varying degree of precision of the query results in the Adaptive Schema Database [25] or Probabilistic Database [26] settings. Mimir propagates the uncertainty model of the original unstructured data to the query results as the data goes through any number of ETL and join operations in the query plan.

Subsequent to the Enactment activity, the Classification activity involves:

- **Determine if there are any actionable perceptions that need to be created or updated**: The most important task in processing facts is to determine if there is any actionable information. This can be done by classifying incoming facts and associating a risk or opportunity rating. To avoid cognitive overload, it involves human operators only if there is a need and it represents the data in a very compact qualitative language.

MSET and SPRT technologies comprise the core of the KIDS

CARE Loop. In the subsequent sections, we will focus on the capability of MSET to classify the perceptions and the capability of SPRT to generate and assess the hypotheses.

IV. The Multivariate State Estimation Technique

MSET is a practical and versatile classification model for the KIDS CARE loop that provides prognostic system health monitoring of business-critical systems. It comprises a comprehensive methodology for proactively detecting and isolating failures, recommending condition-based maintenance, and estimating in real time the remaining useful life (RUL) of critical components. Oracle has over the last 15 years developed and patented a suite of advanced pattern recognition innovations that leverage MSET prognostics for components, subsystems, and for integrated hardware-software systems in enterprise data centers [8], [27]. The key enabler for achieving Electronic Prognostics capabilities is a continuous system telemetry harness (CSTH) which collects and preprocesses any/all types of time series signals relating to the health of dynamically executing components and subsystems [8]. These time series provide quantitative metrics associated with physical and performance variables. The magnitude of the problem is illustrated by the fact that a typical data center now contains up to one million physical sensors inside the information technology assets. These sensors measure temperatures, voltages, currents, power metrics, fan speeds, vibration and many other variables. Performance variables include processor loads, memory usages, throughputs, queue lengths, and many other metrics.

The CSTH signals are continuously archived to an offline circular file (such as a "Black Box Flight Recorder"), and are also processed in real time using the advanced pattern recognition technique MSET for proactive anomaly detection and for RUL estimation with associated quantitative confidence factors.

The most significant advantages of our approach to supervisory control are the following:

- The ability to proactively catch very subtle incipient disturbances, even when the disturbance signature is a tiny fraction of the inherent variance in the monitored metrics
- Ultra-low probabilities for false alarms and missed alarms
- Separately specifiable probabilities for false and missed alarms[1]
- Real Time signal validation and sensor operability validation[2]
- Low compute cost for large-scale prognostic monitoring applications, i.e., lots of sensors and/or high sampling rates. (In many past "bake off" comparisons between MSET and NNs, MSET achieves an order of magnitude higher sensitivity for catching subtle disturbances in noisy process variables, with an order of magnitude lower compute cost)
- Remaining Useful Life estimation with quantitative confidence factors[3]
- Highly accurate "inferential variable" capability. (i.e., one doesn't have to shut down a million dollar asset because a $2 internal sensor failed. MSET can swap in a highly-accurate inferential variable, so the sensor fix/replacement can be postponed to a scheduled maintenance window).

The benefits listed above can help IoT prognostic system health-monitoring applications achieve higher availability with lower operation and maintenance costs. These benefits can be achieved by extending the prognostic surveillance envelope to include an IoT customer's production assets, programmable logic controllers, power supplies, motor-operated valves, and interconnecting networks.

V. Comparison of MSET with other ML techniques

In this section we justify our choice of MSET compared with other ML techniques such as NNs and SVMs. While these other ML techniques are well-known and popular, we will see that MSET has key advantages for human-in-the-loop supervisory control that the other ML techniques do not provide.

What distinguishes human-in-the-loop supervisory control over other applications of ML techniques is the requirement that the operator be given a rigorous explanation for the advice being presented. In other words, the technique used must be amenable to a rigorous reliability assessment methodology. To achieve this requirement, it is vitally important to be able to conduct a rigorous propagation-of-uncertainty analysis of the "expert system" hypothesis decisions following system upset events. Mission-critical and safety-critical industrial applications can have huge liability impacts if operators make incorrect decisions as a result of faulty prognostic recommendations from the surveillance software. With ML techniques such as NNs and SVMs, it is not possible to conduct a rigorous propagation-of-uncertainty analysis a-priori, because of the stochastic optimization of the weights during setup and training. After a NN or SVM is set up and the weights become fixed, then it is possible to conduct an empirical analysis, e.g., by adding uncertainty to various input signals and examining the impact on uncertainties for output signals. This approach is called a "black box" uncertainty analysis and is adequate for some classes of applications. However, for safety- and mission-critical applications, a black box uncertainty analysis is not acceptable. No matter how cleverly one designs perturbations on various permutations of input signals to the black

[1]Conventional equipment surveillance approaches have a "sea saw" relationship between false and missed alarms.

[2]Most false and missed alarms in prognostic system health management of business-critical and even safety-critical systems are due to sensor degradation events.

[3]RUL capability is a key enabler for "Condition Based Maintenance" of customer IoT assets.

box, it is not possible to prove through analytical propagation-of-uncertainty analyses that there is not some combination of input perturbation signatures that may cause a false alarm or may miss an alarm. False alarms are obviously dangerous for human-in-the-loop applications, and missed alarms can have disastrous consequences.

By contrast, MSET uses a straightforward, deterministic computational algorithm that is readily amenable to rigorous "propagation of uncertainty" reliability analysis. This means that one can apply well established reliability assessment of MSET for any safety-critical applications and compute a quantitative mean time between failure (MTBF) for the expert system software embodying MSET, where a "failure" of the expert system is a false alarm or missed alarm. This is a vital capability and is the reason that in 2000 the US Nuclear Regulatory Commission (NRC) formally accepted MSET for commercial nuclear plant applications. At the same time, the NRC disallowed the use of NNs, unless the utility plant owners could prove in the future that a NN implementation could pass the rigorous propagation-of-uncertainty analysis required for prognostic software applications in nuclear plant applications.

MSET also possesses advantages over Neural Networks for real-time prognostic applications in terms of prognostic accuracy, compute cost, and memory footprint. However, the primary reason that MSET was selected is the requirement of a rigorous reliability assessment methodology.

VI. Inferential Sensor Substitution

We now discuss one of the main features of our approach to supervisory control assistance: the ability to detect failing and failed sensors and to infer their readings. This capability is important for preventing operator cognitive-overload when supervising and controlling complex systems or emergency situations. Our approach can generate highly accurate inferential sensors when instrumentation or individual sensors should degrade intermittently or fail completely in service. Dense sensor IoT applications can contain thousands of physical transducers. Business-critical enterprise and cloud data centers can contain $>$ 1M sensors for a medium sized data center. Even during normal, non-emergency operation of industrial and business assets, it is often the case that sensors have a shorter MTBF than the assets the sensors are deployed to protect. MSET combined with KIDS has the capability to detect the incipience or onset of sensor decalibration bias (sensors drifting gradually out of calibration) and all known modes of sensor degradation in operating assets, and to distinguish sensor degradation from anomalous system behavior. A human operator would be confused by spurious alarms from one or more individual sensors, and would be unaware that a sensor has failed. A particularly insidious failure mode is what is known as a "stuck at" fault, i.e., it retains its last reading but no longer responds to the physical metric it is measuring. [4]

[4]A stuck-at sensor will never trip a high/low threshold alarm.

For degrading or failed sensors, MSET detects the degradation and autonomically swaps in a highly accurate MSET estimate, called an *inferential sensor*, that is computed on the basis of covariance with other correlated sensors.

Figure 2 shows an actual "stuck at" sensor fault in a large business-critical enterprise computer server that possessed over 600 physical sensors. The metric plotted is a real-time temperature signal in one of 16 internal power supplies that supply the processors with power. The blue signal is the real digitized time series measured sensor signal, superimposed on the red inferential signal from MSET. Although it would be obvious to a human looking at the graphic in Figure 2 that the sensor suddenly failed with a "stuck at" fault, the fact that there are over 1M sensors in a typical data center makes it impossible for humans to watch the sensor signals on a 24x7 basis. Moreover, because a "stuck at" fault will never trip a high/low threshold, this failure mode is often undiscovered and can lead to catastrophic system failures, especially when the sensor signal is used in a feedback-control loop. Although in safety-critical industries it is common to provide triple redundancy for sensors, it is cost prohibitive to implement triple redundancy in many IoT industries, including in enterprise computer assets.

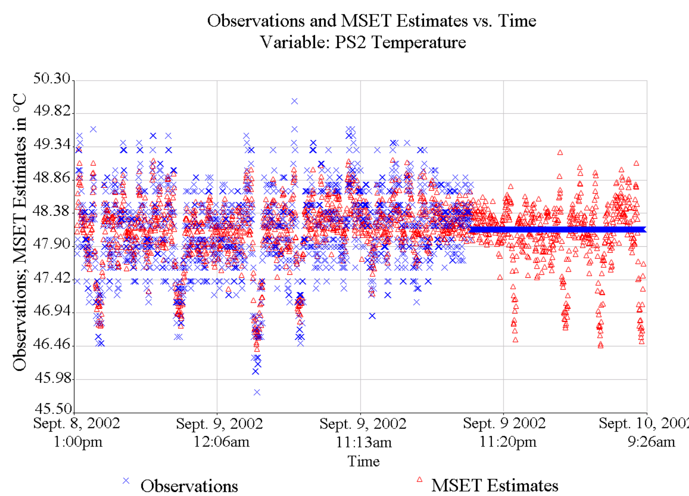

Fig. 2. Illustration of the use of MSET for inferential sensing

In Figure 2, the blue signal is the real digitized time series measured sensor signal, superimposed on the red inferential signal from MSET. When the physical transducer fails in service, the degraded signal is masked out and the MSET inferential signal is substituted into the data acquisition aggregation instrumentation.

For intermittently failing instrumentation that results in missing values in the monitored metrics, MSET performs optimal missing value imputation. It is important to point out here that we are not using conventional "missing value interpolation" algorithms that "fill in" missing values in a univariate time series using conventional interpolation schemes.

Conventional forms of univariate missing value interpolation algorithms suffer from the fact that they are inherently a "lossy" computation. In other words, no matter how cleverly one interpolates to replace a missing value in a univariate time series, the true value could be significantly different. For missing value imputation via MSET, MSET is using correlation patterns with other sensors in a nonlinear, non-parametric algorithm to compute a highly accurate imputed value for signal continuity so that human operators are performing control actions on the basis of fully validated continuous telemetry streams.

VII. The Sequential Probability Ratio Test: Avoidance of False Alarms

The combination of MSET and SPRT is a practical and versatile Assessment model for CARE Loop to unambiguously differentiate between sensor degradation events and degradation in the systems/processes under surveillance. It is an important capability for the root cause analysis.

Many industrial processes have embedded diagnostic systems and online statistical process control techniques that perform real-time analysis of process variables. Most of these systems employ simple tests such as threshold, mean value + three-sigma, statistical process control thresholds, etc. These tests are sensitive only to gross changes in the process mean, or to high step changes or spikes that exceed some threshold-limit test to determine whether or not a failure has occurred or a process is drifting out of control. These conventional methods suffer from either large false alarm rates (if thresholds are set too close) or high missed (or delayed) alarm rates (if the thresholds are set too wide).

For new dense-sensor IoT monitoring applications in industrial manufacturing facilities, utilities, and transportation assets, false alarms are very costly in terms of plant or physical-asset down time. Missed alarms can be even more costly when incipient problems are not identified and expensive assets fail catastrophically.

MSET provides a superior surveillance tool because it is sensitive not only to disturbances in signal mean, but also to very subtle changes in the statistical moments of the monitored signals and the patterns of correlation between/among multiple types of signals. MSET employs SPRT [18], [28]–[30], which provides the basis for detecting very subtle statistical anomalies in noisy process signals at the earliest mathematically possible time, thereby providing actionable warning-alert information on the type and the exact time of onset of the disturbance. Instead of simple threshold limits that trigger faults when a signal increases beyond some threshold 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 alarm. For sudden, gross failures of sensors or system components the SPRT annunciates the disturbance as fast as a conventional threshold limit check. However, for slow degradation that evolves over a long time period the SPRT raises a warning of the incipience or onset of the disturbance long before it would be apparent to any conventional threshold based rules. Slow degradation can occur for a variety of reasons such as a gradual decalibration bias in a sensor or a very subtle voltage drift due to a variety of aging mechanisms that cause resistances to change very slowly with age. Still other reasons for slow degradation include: bearing degradation, lubrication dryout, a buildup of a radial rub in rotating machinery, and the gradual appearance of new vibration spectral components in the presence of noisy background signals.

MSET is a nonlinear, non-parametric regression modeling method that was originally developed by Argonne National Laboratory for high-sensitivity proactive fault monitoring applications in commercial nuclear power applications. It has since been spun off to a variety of other mission-critical and safety-critical industries. Oracle was the first company to develop MSET-based prognostic tools for enterprise computing health-monitoring applications that are truly dense-sensor applications. (A small 4U server now has more than 400 sensors; a large engineered system contains 3400 sensors; and a medium size data center has over 1M sensors). Oracle and our university collaborators are now extending proven MSET prognostics to dense-sensor challenges in IoT [27], [31], [32].

The overall MSET framework consists of a training phase and a monitoring phase (Figure 3 below). 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 (automatically) selects a subset of the data observations (using a similarity operator) 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 observations for all the asset signals are first acquired. These observations are then used in conjunction with the previously trained MSET model to estimate the expected values of the signals. MSET estimates are extremely accurate, with error rates that are usually only 1 to 2 percent of the standard deviation of the input signal. Incidentally, the MSET estimate for a signal originating from any physical transducer is more accurate than the transducer itself. The end-to-end processing steps taken during the MSET surveillance phase are shown in Figure 3.

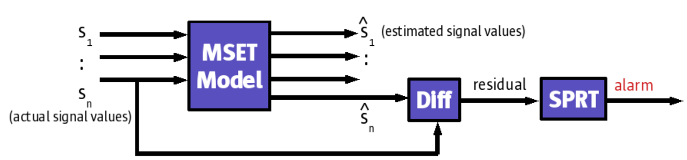

Fig. 3. MSET surveillance phase block diagram

The difference between a signal's real-time MSET estimate and its directly sensed value is termed a *residual*. The residuals for each monitored signal are used as an anomaly indicator for sensor and equipment faults. Instead of using simple thresholds to detect fault indications, MSET's fault detection procedure employs the SPRT to determine whether the residual error value is uncharacteristic of the learned process model and thereby indicative of a sensor or equipment fault. The SPRT algorithm is a significant improvement over conventional threshold detection processes in that it provides more definitive information about signal validity with a quantitative confidence factor through the use of statistical hypothesis testing. This approach allows the user to specify false alarm and missed alarm probabilities, allowing end-customer control over the likelihood of false alarms or missed detection.

VIII. Conclusion

The MSET system comprises a synergistic integration of the SPRT technique with a data-driven modeling method to produce a system with unique surveillance capabilities. Recent large-scale analyses with archived time series signals have shown that the MSET system surpasses conventional approaches, including neural networks, auto-associative kernel regression, and regularized kernel regression using a variety of criteria. The criteria include sensitivity, reliability, robustness to unreliable and possibly degrading sensors, simplicity of training, adaptability when sensor configurations change, and computational efficiency.

References

[1] C. Matheus, M. Kokar, K. Baclawski, J. Letkowski, C. Call, M. Hinman, J. Salerno, and D. Boulware, "SAWA: An assistant for higher-level fusion and situation awareness," in Proc. SPIE Conference on Multisensor, Multisource Information Fusion, vol. 5813, 2005, pp. 75–85.

[2] "The National Information Sharing Consortium," June 2012 [Online], Available: http://www.nisconsortium.org/.

[3] J. Boyd, "Destruction and creation," U.S. Army Command and General Staff College, Tech. Rep., September 3 1976 [Online], Available: http://www.goalsys.com/books/documents/DESTRUCTION_AND_CREATION.pdf.

[4] D. Gawlick, "Mastering situation awareness: The next frontier?" in 7th Biennial Conference on Innovative Data Systems Research (CIDR '15), Asilomar, California, USA, January 4–7, 2015 [Online], Available: http://cidrdb.org/cidr2015/Papers/15_Abstract43GD.pdf.

[5] D. Gawlick, E. Chan, A. Ghoneimy, and Z. Liu, "Mastering situation awareness: The next big challenge?" *SIGMOD Record*, vol. 44, no. 3, pp. 19–24, 2015.

[6] K. Baclawski, E. Chan, D. Gawlick, A. Ghoneimy, K. Gross, Z. Liu, and X. Zhang, "Framework for ontology-driven decision making," *Applied Ontology*, 2017, to appear.

[7] S. Txafestas, Ed., *Knowledge-Based System Diagnosis, Supervision, and Control*. Plenum Publishing Co., 1989.

[8] K. Gross, K. Whisnant, and A. Urmanov, "Prognostics of electronic components: Health monitoring, failure prediction, time to failure," in Proc. New Challenges in Aerospace Technology and Maintenance Conf. 2006, Suntec City, Singapore, Feb 2006.

[9] S. Wegerich, "Similarity based modeling of vibration features for fault detection and identification," *Sensor Review*, vol. 25, no. 2, pp. 114–122, 2005.

[10] D. Kriesel, "A Brief Introduction to Neural Networks," 2005 [Online], Available: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf.

[11] J. Shawe-Taylor and N. Cristianini, *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[12] S. Wegerich, "Condition based monitoring using nonparametric similarity based modeling," in Proc. 3rd Conference of Japan Society of Maintenology, 2006, pp. 308–313.

[13] J. Coble, P. Ramuhalli, L. Bond, J. Hines, and B. Upadhyaya, "Prognostics and health management in nuclear power plants: A review of technologies and applications," US Dept. of Energy, Tech. Rep. Technical Publication PNNL-21515, July 2012.

[14] T. Plume, F. Liard, and D. Therkom, "Similarity based modeling for turbine exit temperature spread monitoring on gas turbines," in Proc. ASME Turbo Expo 2013: Turbine Technical Conference and Exposition, vol. 4, San Antonio, TX, June 2013.

[15] J. Hines, D. Garvey, R. Seibert, and A. Usynin, "Technical review of on-line monitoring techniques for performance assessment, volume 2: Theoretical issues," Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission, Tech. Rep., May 2008.

[16] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory. ACM Press, 1992.

[17] C. Peirce, *Reasoning and the Logic of Things*. Cambridge, MA: Harvard University Press, 1992.

[18] A. Wald, *A Sequential Analysis*. New York: Wiley, 1947.

[19] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988.

[20] R. Fletcher, *Generalized Inverses for Nonlinear Equations and Optimization. Numerical Methods for Non-Linear Algebraic Equations*. London: Gordon and Breach, 1970.

[21] R. Howard and J. Matheson, "Influence diagrams," in Readings on the Principles and Applications of Decision Analysis, R. Howard and J. Matheson, Eds. Menlo Park, CA: Strategic Decisions Group, 1981, vol. 2, pp. 721–762.

[22] A. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society. 1968*, vol. 30, no. 2, pp. 205–247, 1968, series B (Methodological).

[23] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[24] A. Nandi, Y. Yang, O. Kennedy, B. Glavic, R. Fehling, Z. Liu, and D. Gawlick, "Mimir: Bringing CTables into practice," Oracle Corporation, Tech. Rep., 2016.

[25] W. Spoth, B. Arab, E. Chan, D. Gawlich, A. Ghoneimy, B. Glavic, B. Hammerschmidt, O. Kennedy, S. Lee, Z. Liu, X. Niu, and Y. Yan, "Adaptive schema databases," in 8th Biennial Conference on Innovative Data Systems Research (CIDR '17), Chaminade, California USA, January 8–11, 2017.

[26] N. Dalvi, C. Re, and D. Suciu, "Probabilistic databases: Diamonds in the dirt," *Communications of the ACM*, vol. 52, no. 7, pp. 86–94, July 2009.

[27] A. More and K. Gross, "SimML framework: Monte Carlo simulation of statistical machine learning algorithms for IoT prognostic applications," in Proc. Intn'l Symposium on Internet of Things & Internet of Everything (CSCI-ISOT), Las Vegas, NV, Dec 15-17, 2016.

[28] K. Gross and W. Lu, "Early detection of signal and process anomalies in enterprise computing systems," in Proc. 2002 IEEE Int'l Conf. on Machine Learning and Applications (ICMLA), Las Vegas, NV, June 2002.

[29] T. Masoumi and K. Gross, "SimSPRT-II: Monte Carlo simulation of sequential probability ratio test algorithms for optimal prognostic performance," in Proc. 2016 International Symposium on Artificial Intelligence (CSCI-ISAI), Las Vegas, NV, Dec 15-17, 2016.

[30] J. Hines and D. Garvey, "Development and application of fault detectability performance metrics for instrument calibration verification and anomaly detection," *J. of Pattern Recognition Research*, vol. 1, pp. 2–15, 2006.

[31] K. Gross, K. Vaidyanathan, A. Bougaev, and A. Urmanov, "Round-robin staggered-imputation (RRSI) algorithm for enhanced real-time prognostics for dense-sensor IoT applications," Intl. Conf. on Internet Computing and Internet of Things (ICOMP'16), Las Vegas, NV, July 25-28, 2016.

[32] K. Gross, K. Vaidyanathan, , and M. Valiollahzadeh, "Advanced pattern recognition for optimal bandwidth and power utilization for intelligent wireless motes for IoT applications," in 17th International Conference on Wireless Networks (ICWN'16), Las Vegas, NV, July 25-28, 2016.