

Standardizing Research Methods for Prognostics

Serdar Uckun, *Member, IEEE*, Kai Goebel, and Peter J.F. Lucas

Abstract—Prognostics and health management (PHM) is a maturing system engineering discipline. As with most maturing disciplines, PHM does not yet have a universally accepted research methodology. As a result, most component life estimation efforts are based on ad-hoc experimental methods that lack statistical rigor. In this paper, we provide a critical review of current research methods in PHM and contrast these methods with standard research approaches in a more established discipline (medicine). We summarize the developmental steps required for PHM to reach full maturity and to generate actionable results with true business impact.

Index Terms—Prognostics, Prognostics and Health Management, Research Methods, Performance Metrics

I. INTRODUCTION

Over the last decade, the reliable systems community started focusing on the fundamental principles of system failures in an attempt to understand how complex electromechanical systems age and to predict when they might fail. The emerging engineering discipline that links studies of failure mechanisms to system lifecycle management is referred to as Prognostics and Health Management (PHM). In recent years, analysis of vibration and acoustic emissions data from rotorcraft drivetrains have led to breakthroughs in predicting impending failures of these complex mechanical systems, resulting in the development of Health and Usage Monitoring Systems (HUMS) for rotorcraft [1]. Substantial advances were made in life estimation for components ranging from rotating machinery [2] to batteries [3], from printed circuit boards [4] to solid rocket motors [5]. In the U.S. military, two significant weapon platforms were designed with a prognostics capability as an integral element of the overall system architecture: the Joint Strike Fighter Program [6] and the Future Combat Systems Program [7]. Prognostic technology is also finding

its way into future NASA launch vehicles and spacecraft [8]. As the technology matures further, prognostics will play an important role in the design and operation of commercial systems such as passenger aircraft, automobiles, ships, the energy infrastructure, and even consumer electronics.

Despite substantial technical progress over the last decade, PHM does not yet have a universally accepted research methodology. More importantly, the scientific method that underlies all scientific disciplines has not made its way into PHM research. As a result, most component life estimation efforts are based on ad-hoc experimental methods that lack statistical rigor. In this paper, we review current research methods in PHM and contrast these methods with standard research approaches in an established scientific discipline (medicine).

II. METHODS FOR PROGNOSTICS AND REMAINING USEFUL LIFE ESTIMATION

The science of prognostics is predicated on four fundamental notions:

- All electromechanical systems age as a function of use, passage of time, and environmental conditions;
- Component aging and damage accumulation is a monotonic process that manifests itself in the physical and chemical composition of the component;
- Signs of aging (either direct or indirect) are detectable prior to overt failure of the component (i.e., loss of function);
- It is possible to correlate signs of aging with a model of component aging and thereby estimate *remaining useful life* of individual components.

Figure 1 illustrates the states of component lifecycle as a finite state machine. Note that there is no general agreement in the research community with respect to the terminology used for component aging. In this paper, we will use the

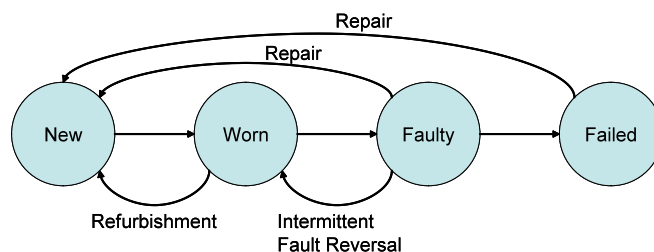


Fig 1. Component lifecycle represented as a finite state machine.

Manuscript received May 23, 2008.

S. Uckun is with the Palo Alto Research Center (PARC) in Palo Alto, CA 94304 USA (phone: 650-812-4329, email: uckun@parc.com).

K. Goebel is with the NASA Ames Research Center in Moffett Field, CA 94035 USA (email: kai.goebel@nasa.gov).

P.J.F. Lucas is with the Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands (email: peterl@cs.ru.nl).

following terminology:

- A **new** component refers to an electromechanical device or component with no discernible damage at the chemical or physical level.
- A **worn** component is a component that exhibits signs of wear and tear due to use or time at the physical or chemical level. These signs may or may not be observable using conventional sensing methods (depending on the size and extent of the damage). The component continues to serve its primary purpose.
- **Remaining Useful Life (RUL)** is an estimation of the remaining life of a component prior to occurrence of a failure. RUL is expressed using units corresponding to the primary measurement of use of the overall system. For example, the primary measurement that correlates with use in commercial aircraft is cycles (typically measured by number of take-offs); in aircraft engines, it is hours of operation; in automobiles, it is miles (or kilometers) driven; in printers or copiers, number of pages printed.
- A **faulty** (or faulted) component is one whose intended function is impaired or has it has ceased to perform its intended function within the system it is integrated in.
- **Failure** is the loss of function of the system.
- An **intermittent fault** is a fault that reverses itself spontaneously, restoring function to the system while transitioning the component back to the “worn” state.
- **Reliability** is the probability of survival of a system, expressed by the probability $R(t)$, which gives the probability that the system functions over time interval $[0, t]$. By definition $R(0) = 1$ and $R(\infty) = 0$, i.e., in the end the system will fail. If $Q(t)$ indicates the probability of unreliability, then it holds that $R(t) + Q(t) = 1$ for each time point t .
- **Refurbishment** of an aged system revert worn components to “new” status through replacement.
- **Repair** of a failed system reverts its faulted components to “new” status through replacement.

To illustrate, wheel-bearing faults in a car may manifest themselves as vibrations or noise emanating from the wheel in question. A failed wheel bearing, in contrast, will manifest itself in failure of the vehicle (the overall system) to move.

Methods for direct measurement of damage accumulation and aging include physical inspection, monitoring of measurands related to the damage (e.g., vibration, acoustic emissions, oil debris), and non-destructive inspection (e.g., x-rays, ultrasound). Methods for indirect assessment of damage involve observation of side effects of damage, such as streaks on printed paper indicating component aging in a printer, or a puddle of oil underneath a car indicating damage to a seal or fitting.

A. Impact of Use on Damage Accumulation

Mean Time Between Failures (MTBF) is a rather crude estimator of when components may fail. Just as life spans of individual persons may vary greatly from the population life expectancy, useful life spans of individual components may vary substantially from the MTBF measured for their population. Component aging is a complex and often non-linear function of time, usage, and environmental conditions.

Accurate RUL estimation often requires monitoring and analysis of environmental conditions the system is operated in (e.g., temperatures, temperature cycles, pressure, humidity, ambient light and vibrations, acceleration, physical loads). In addition, metrics related to use are often correlated with aging. For example, a common life estimation metric quoted for lithium-ion batteries is 500 charge-recharge cycles. Similarly, rotorcraft drivetrain aging correlates well with the number of ground-air cycles [9]. In commercial aircraft, structural fatigue is strongly correlated with the number of pressurization cycles. In fact, the Federal Aviation Administration (FAA) mandates the replacement of life-limited parts (LLP) after a predetermined part-specific number of cycles. Careful observation of use patterns and analysis of usage-related data often reveals interesting (and sometimes unexpected) correlations. For example, incandescent light bulbs almost always fail the moment they are turned on or off. As such, one might postulate that the average number of on-off cycles might be a more accurate indicator of use-induced damage for light bulbs rather than total hours of operation.

B. Data-Driven Methods for Remaining Life Estimation

Most RUL estimation efforts follow a data-driven approach. Unlike traditional reliability engineering approaches where one is primarily interested in the initiation of faults, PHM requires a much more intensive data collection process in order to characterize damage accumulation and progression. In data-driven prognostics, the challenge is to capture and analyze a multidimensional and noisy data stream from a large number of channels (use conditions, environmental conditions, direct and indirect measurements that are potentially related to component damage) from a population of similar components. In many cases, data collection, storage, and analysis requirements are onerous (e.g., vibration or acoustic-emissions data analysis requires sampling rates in the 10 kHz range and above).

The challenge of data-driven RUL estimation does not end with the complexity of the data sets. Unlike diagnosis which aims to isolate the root cause of a fault event that has already occurred, prognosis is the prediction of a future failure

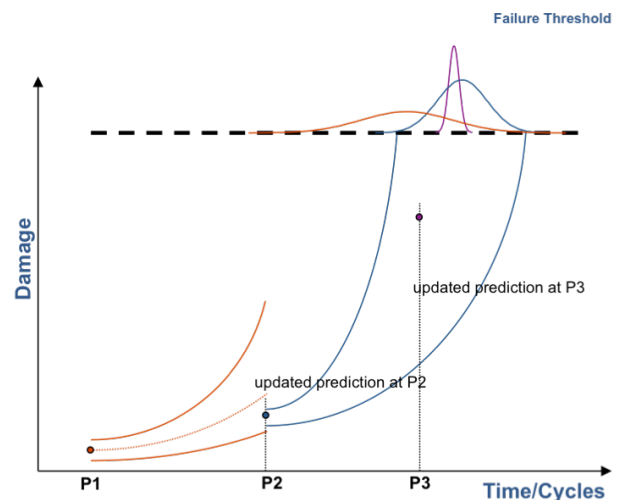


Fig. 2. Uncertainty bands associated with prognostic RUL estimations as time progresses.

condition. Given that the future health of the component depends on future environmental and operational conditions, uncertainty management is an inherent element of RUL estimation. Figure 2 illustrates the uncertainty bands associated with prognostic RUL estimation. While the component is relatively new and accumulated damage is relatively minor, the uncertainty regarding the exact time of failure is high (the wide probability distribution illustrated in red). As the component accumulates more damage and the failure point approaches, there is usually much less uncertainty as to when the component fault may occur (the narrow probability distribution illustrated in blue).

C. Model-Based Methods for Remaining Life Estimation

Another approach to prognostic RUL estimation is the model-based approach. The model-based approach involves development of first-principles models of component use and damage accumulation, and the use of operational data to fine-tune model parameters. Modeling approaches include materials-level models such as crack growth or spall progression models, or macro-level models such as gas-path cycle models for turbine engines. There are fewer examples of model-based approaches in the PHM literature than there are data-driven approaches. This reflects the fact that model-based prognostics is a more involved process than data-driven prognostics. Depending on the fidelity of the model, model-based prognostics typically results in more accurate and precise RUL estimation. Model-based prognostics also has advantages in validation, verification, and certification since the model response can be correlated with laws of nature.

III. PROGNOSTICS IN MEDICINE

Arguably, PHM is a maturing science. The concept of using data on individual components to determine RUL and to optimize maintenance practices using such data (i.e., autonomic logistics [10]) is still in the experimental stage. In contrast, prognostic reasoning, along with diagnosis, is the basis for managing diseases in medicine. As diseases develop in time, one usually speaks of the *natural history* of a disease. The natural history starts before a diagnosis has been made. Further progression of a disease, possibly after clinical intervention, is usually referred to as the ‘clinical course’. Knowledge of the prognosis of a disease is crucial for selecting appropriate treatment. Thus, medicine includes a wealth of prognostic methods [11].

For the prevention of disease, it is important to know what factors determine the occurrence of the disease. A distinction is often made between *risk factors*, which are known to be causally related to the disease, and *risk indicators*, for which such causal relationship is (yet) unknown. For example, blood cholesterol level is a risk factor of myocardial infarction, whereas gender is a risk indicator.

Insight into the natural history and clinical course of disease is obtained by deployment of various study methods. Some of these methods are scientifically rigorous yet expensive and time consuming, while others are less demanding. The purpose of a *follow-up study* is to determine whether there is a difference between two study populations where the so-called index population is exposed to a factor and the reference

population is not exposed. One may then simply compute the relative risk, *RR* for short, of the occurrence of outcome *O* due to exposure or absence of exposure to factor *F*:

$$RR = P(O | F) / P(O | \bar{F}).$$

For simplicity’s sake, the relative risk is often considered time invariant; in practice, however, the relative risk is often time dependent. For example, age is an important factor in determining the effect of smoking on the occurrence of lung cancer: the older the person, the higher the risk.

Data used in a follow-up study is often collected prospectively, although retrospective studies are also possible. A disadvantage of retrospective follow-up studies is that all available patient information has been collected for another purpose than the study at hand. As a consequence, essential information may be missing, the level of detail of information may not be appropriate, and the information often lacks sufficient reliability. Highly relevant follow-up studies, that have had an impact on patient management are, therefore, normally prospective in nature. The study is called ‘observational’ if the outcome is observed without intervention. In an intervention study, the outcome is manipulated by imposed actions such as medical treatment. The latter study design is common in investigating the effectiveness and presence of adverse reactions to drugs, and, in particular, has become the standard in clinical cancer research. The two groups of patients are in that case chosen by randomization, where the control (reference) group is ‘treated’ by a placebo or state-of-the-art treatment, and the treatment group is subjected to the new, experimental treatment. This study design is also known as the randomized clinical trial (RCT). Although the RCT offers one of the most rigorous experimental study designs in clinical medicine, it requires considerable expense, preparation, and staff involvement. Note again that if it is not the purpose to investigate the effects of an intervention, the follow-up study is observational.

In addition to the follow-up study, there exists the *case-control study*, where presence and absence of a particular outcome, rather than the factor, is taken as the starting point. By going back in time, the researcher attempts to find out what factors may have caused the outcome. Case-control studies often suffer from selection bias.

The results of clinical studies are often used to develop prognostic models that help clinicians in estimating outcomes for patients. In particular, very simple prognostic scoring rules are popular in medicine. A prognostic scoring rule is often a simple additive formula, where presence or level of significance of a factor is given a number greater than zero (e.g., 1, 2, etc.), and a zero if absent. The resulting total score conveys prognostic information based on statistics. For example, the International Prognostic Index (IPI) is one of the clinical aids clinicians use to determine the survival of patients with non-Hodgkin lymphoma. Based on age (greater than 60 years), stage of the disease, raised levels of lactate dehydrogenase in blood, clinical performance status and involvement of more than one lymphatic node, a total score is calculated. For example, a total score of 2 points is interpreted as ‘low-to-intermediate risk’ (5 year survival of 51%), whereas a score of 4-5 points correlates with high risk (5 year survival of 26%) [12].

More sophisticated statistical models used in medical prognostics include logistic regression and Cox's regression [13]. Cox's regression, also known as the proportional hazards model, has the advantage that the clinical course of the disease can be explicitly modeled by means of a *probabilistic survival function* $S(t) = P(T > t)$, the probability that the patient survives more than t years. If mortality is the outcome variable, then one speaks of *survival analysis*. If $F(t) = 1 - S(t)$, and $f(t) = F'(t)$ is the first derivative of the distribution function F , then the concept of **hazard**, defined as $h(t) = f(t)/S(t)$, gives the instantaneous risk of demise after time t . Logistic regression and Cox's regression are multivariate statistical regression methods. If these methods are augmented with a cut-off threshold, they can also be used to *classify* individual patients, for example, those likely to survive more than five years and those that do not.

In medical Artificial Intelligence, Bayesian networks have been proposed as alternatives to the traditional multivariate regression models. The advantage of Bayesian networks is that they handle missing data with ease when computing the risk; in addition, input and output variables can be chosen arbitrarily. This freedom of choice renders those models as the most flexible available. Temporal Bayesian networks allow modeling temporal evolution of a disease process, including the effect of interventions, as a function of time. There is now some recent literature describing actual clinical prognostic Bayesian models [14]. However, more work is required to investigate their clinical value.

Although outcomes in medicine are often expressed in terms of morbidity or mortality, many prognostic models yield a range of values corresponding to severity of the disease. These more detailed prognostic models have achieved a crucial role in clinical decision-making and optimal treatment selection. This is especially true for models that have been substantiated with statistical significance. The fact that a prognostic model is based on an RCT makes the conclusions trustworthy irrespective of who conducted the research.

IV. PROGNOSTIC METRICS

In medicine, prognostic methods, such as Cox's regression, usually predict morbidity and mortality, possibly referring to time. There is a clear justification for use of such methods, as they guide the choice of appropriate, effective treatment. Whether or not the choice improves due to use of a disease-specific prognostic model has been established empirically for many prognostic models in medicine. In contrast, there are no universally accepted methods to quantify the benefit of prognostics in engineering. Nonetheless, a number of researchers have investigated economic metrics that attempt to quantify benefits of prognostics, for example, ROI, Total Value, and others [15]. Other metrics include performance based metrics that focus on how well prognostic algorithms are doing. These metrics can be categorized into accuracy-based metrics, precision-based metrics, and robustness-based metrics. Saxena et al [16] provide a comprehensive review of metrics for evaluating performance of prognostic techniques.

Accuracy-based metrics describe the closeness of RUL estimates to the actual remaining life. Estimates that are closer

to the target are considered more accurate. The *error* metric is a representative of the accuracy metric. Error defines the deviation from desired output. Most accuracy-based metrics are derived directly or indirectly from error. It is defined as:

$$\Delta^l(i) = r_*^l(i) - r^l(i),$$

where $r^l(i)$ is the remaining useful life estimation at time index i given that the information (features and conditions) up to time index l and $r_*^l(i)$ is the true remaining useful life.

In contrast, precision-based metrics express the degree of reproducibility. If a large number of estimates are made with different measurements, precision would be the size of the distribution around the target that results from noise in the measurements, uncertainty of the model, future conditions, etc. When all estimates are grouped tightly together, the distribution is considered precise whether or not they are near the target. Figure 3 illustrates the concepts of accuracy and precision. *Standard deviation* is a basic precision measure. It is defined as:

$$S(i) = \sqrt{\frac{\sum_{l=1}^n (\Delta^l(i) - M)^2}{n-1}},$$

where M is the sample mean of the error.

Finally, robustness expresses the quality of being able to withstand changes in input or external disturbances such as environmental conditions. If the algorithm is capable of coping well with those (unpredictable) variations with minimal impact on the accuracy and precision of the RUL estimate, it is called a robust algorithm. *Sensitivity* is a representative robustness-based metric. It is defined as:

$$S(i) = \frac{1}{L} \sum_{l=1}^L \left\{ \frac{\Delta M^l(i)}{\Delta_{input}} \right\},$$

where ΔM is the distance measure between two successive outputs for the value of metric M and Δ_{input} is the distance between two successive inputs.

In addition, other, less common, metrics are used in the PHM community. In order to mature the science of PHM beyond its current state, the PHM community needs to reach an agreement on systematic, objective, and quantitative measures that can be used as prognostic metrics.

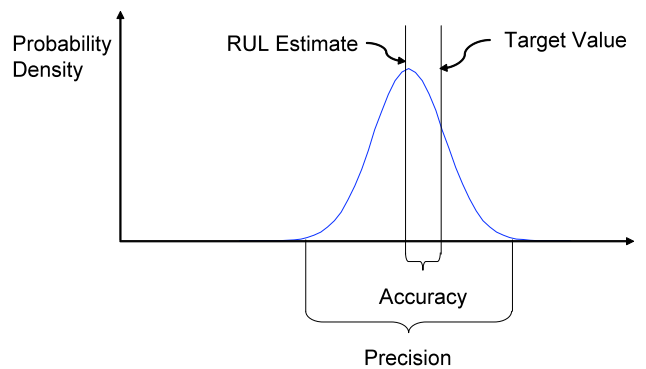


Fig. 3: Illustration of Accuracy and Precision for RUL

V. PHM RESEARCH METHODS

Metrics aside, one of the main issues with PHM today is that there is no robust methodology that governs PHM research. Papers published in the field often utilize home-grown research methodology. While those methodologies may be derived from math, physics, statistics, or other domains, it is often impossible to derive actionable conclusions based on the work presented. To illustrate the point, Table 1 summarizes the research methods used in several PHM-related papers presented at the 2008 IEEE Aerospace Conference. The summary presented in the table is not intended as a criticism of particular authors or research projects. Rather, it is used to illustrate the arbitrary nature of PHM research methods, including the variety of test practices and data analysis methods used and the tentative nature of conclusions. Furthermore, none of these papers reach its conclusion through statistical significance (i.e., z-test or similar methods).

In order to understand the conclusions of a PHM paper, one needs to study the methodology and decide whether the methodology has merit, whether the conclusions are statistically significant, and whether the results are applicable to a broad range of components of similar form or function. This approach is neither scalable nor sustainable, and in fact it impedes progress. In contrast, one only needs to read the abstract of a medical research article in order to understand the research objectives and conclusions. The fact that the research project is based on a standard methodology such as RCT eliminates the guesswork and establishes credibility.

In order to establish a business case around PHM, PHM needs to become less of an art and more of a science, and the community needs to develop a standard research methodology. Arguably, what we are missing in PHM is the equivalent of RCT: a robust, credible research methodology that every prognostic RUL estimation effort is based on.

The purpose of this paper is not to propose a particular research methodology that can be standardized across the PHM community. We believe that a universally-accepted research methodology will evolve through continued dialogue in the community. Rather, we would like to identify the minimum requirements for a universal research methodology for PHM. In the following sections, we will first discuss the elements required for the design of a prognostic system before discussing the research methodology.

A. Sensors

The system-level goals and requirements for PHM need to be flowed down to prognostic needs for components and ultimately to the prognostic method to be employed. Sensors are vital information sources that acquire measurements with some correlation to the damage propagation properties of the fault of interest, at a sampling rate and with a signal-to-noise ratio that allows accurate state assessment and prediction of future behavior. Sensor fusion techniques, advanced feature extraction methods, and virtual sensors are often adopted as data processing methods to improve the quality of prognostic data. In any case, expectations of prognostic performance should be defined with respect to the practical limitations of the sensor and data acquisition systems.

In addition, the availability of sensors is likely to be limited for any legacy system. Even for new systems, there is a strong desire to restrict the number of sensors because of weight concerns and the belief that unreliable sensors may make the overall system less reliable. These concerns have to be weighed against potential benefits of PHM. In the aerospace business, “every sensor needs to earn its way into the system.” PHM sensors are no exception.

B. Tradeoffs During Prognostic System Design

There are fundamental tradeoffs between coverage, accuracy, uncertainty, complexity, time horizon of prediction, and cost. As the complexity and cost of the prognostic system increases (e.g., through addition of sensors or data processing), so does its accuracy -- while uncertainty and coverage decrease.

C. Decision Process

As mentioned earlier, prognostic system designs are driven by system requirements and are constrained by resource availability. To that end, the prognostic decision process should either lead to a choice of prognostic technology that meets the requirements (possibly at different levels of performance) or to the conclusion that high-level requirements cannot be met given overall constraints.

Figure 4 shows the selection flowdown process for prognostic design. The first step is a failure modes analysis such as FMECA to determine the most critical faults. Next, sensor capabilities are assessed to provide information about sampling rates and the trending qualities of their derived features. The availability of models is subsequently assessed and their quality and other characteristics (like computational burden) are analyzed. In a final step, different approaches and their likelihood to meet the requirement metrics are considered. This will yield a degree of fulfillment. If the high-level requirements are not met, the process repeats until all possible approaches have been considered or certain requirements are waived.

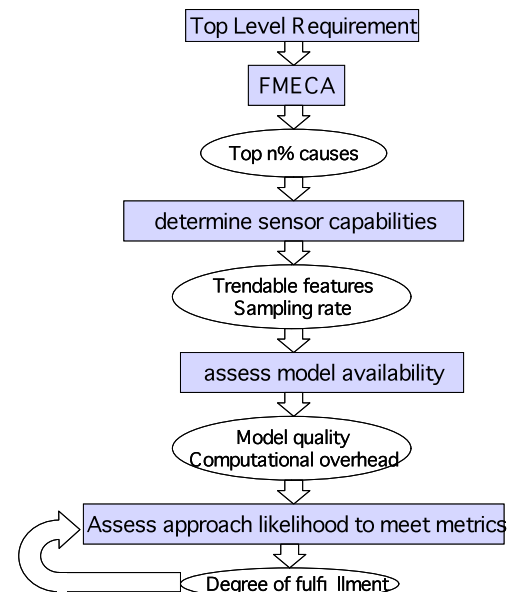


Fig. 4: Selection flowdown process for prognostics [17]

TABLE I

SUMMARY OF PHM RESEARCH PAPERS PUBLISHED AT THE IEEE AEROSPACE CONFERENCE IN 2008

Authors	Target Component	Method Summary	Quantity	Hypothesis Tested	Methods Used	Conclusion
He and Bechhoefer [18]	Bearings	Analysis of bearing and spall progression data collected by Sentient (limited to data collected after fault was seeded)	47	condition indicators do not correlate with bearing life when the damage is small	Multiple linear regression analysis, K-Nearest Neighbor	HUMS condition indicators extracted from vibration data useful in estimating RUL
Chen et al. [19]	Bearings	Custom-designed test rig; run to failure	4 (2 control and 2 test)	AI-based methods are useful in determining RUL	Probabilistic Diagnostic and Prognostic System (ProDAPS)	significantly earlier detection may be achieved using ProDAPS on data from the vibration and electrostatic sensors
Baybutt et al. [20]	Actuator electronic components (FET, capacitor)	rapid thermal cycling, rapid cycling of electromagnetic fields (FET), overcharging (capacitor)	1 set	it is possible to discern between healthy and functionally degraded operation	proprietary data analysis and visualization tools	detection ability was demonstrated for both types of aged transistors and the seeded capacitor fault
Kumar, Sotiris, and Pecht [21]	Laptop computers	Operational settings under multiple environmental conditions (thermal, humidity)	10 (9 for training, 1 for test)	Statistical approaches are useful in predicting the future reliability of complex electronic systems	Mahalanobis Distance and Projection Pursuit Analysis	these two algorithms can be used for fault detection and isolation.
Palazzolo, Scheunemann, and Hartin [22]	Axial-Piston Variable Displacement Pumps	Hydraulic test rig; nominal and seeded fault (piston leakage) conditions	1	Physics-based models are valuable as a prognostic RUL estimation tool for VDPs	high-fidelity physics-based model; time- and frequency-domain data analysis	developed and demonstrated promising fault estimation algorithms for diagnosing axial-piston VDPs
Hofmeister et al. [23]	FPGA solder joints	Special test boards subjected to highly-accelerated life tests (temperature cycling, vibration, drop)	32	SJ-BIST is an effective method for detecting faults in solder joint networks in FPGA I/O ports	canary circuit	All faults of 100 Ω or larger were detected, and there were no false alarms.
Keller et al. [24]	Actuator power converter solder joints	Special test assemblies subjected to deep thermal cycling	20 (5 each of four different substrates)	there is a correlation between thermal cycle solder fatigue and the magnitude of the thermal expansion mismatch	Finite Element Analysis modeling	work in progress
Saha and Goebel [25]	Lithium-ion batteries	Test rig with two different temperature baselines	not specified	It is possible to predict RUL of batteries under environmental and load conditions different from training data sets.	Relevance Vector Machines and Particle Filters	The combined Bayesian regression-estimation approach shows promise in estimating RUL for batteries.

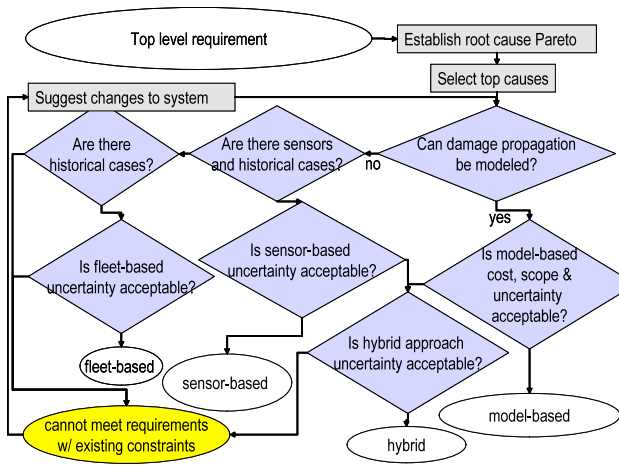


Fig. 5. Prognostic design process [17]

Figure 5 illustrates the prognostic design process in more detail using a decision tree.

Starting again with the high-level requirements, one needs to flow down the requirements and substantiate them at the level that is meaningful for the prognostics technology choices. Specifically, one needs to establish what life-limited module or component to apply prognostics to (or whether prognostics is needed at all, for that matter). To that end, one could analyze the top drivers that impact the high-level requirement, for example, by performing a Pareto analysis of the failure modes for a particular component.

After determining what failure mechanism might benefit from a prognostic assessment, a survey of the existing damage propagation models needs to be made. Here, it is vital to assess whether a fundamental understanding of the damage growth exists or whether it can be established. Damage propagation modeling is often times a very complex undertaking. When performed at the materials level, it can involve finite element modeling which in turn requires precise information about material properties, geometric design information, and load sensing information for both the current time frame as well as future load profiles. A model does not need to be available at the materials level. Simpler models can also be used but they will trade off coverage with accuracy and narrow uncertainty bounds.

If there is (or will be) no suitable model, one still might be able to provide prognostic estimates by engaging in a data-driven approach. Data-driven approaches may be suitable alternatives to model-based approaches when retrospective data exist in sufficient quantity. Sufficient quantity in this context means that run-to-failure has been observed for all fault modes of interest several times. This may be an issue for new systems without service experience. If sufficient historical data are not available, the prognostic requirements cannot be met. In many cases, neither a pure model-based approach nor a pure data-driven approach is able to provide the desired prognostic capability. In these situations, hybrid approaches can help. Hybrid approaches include the use of models that were not specifically designed for damage propagation. Instead, the models can be driven to provide response that is recognized as being consistent with a

particular damage level by comparing that with information from real data.

If neither a model-based approach can be pursued nor a condition-based data-driven approach is feasible, there is still the option to engage in fleet-wide statistics. Here, one would tap into data, typically at a rather coarse level of granularity, gathered at the fleet level from a large number of fielded systems, and adjust reliability curves (e.g., Weibull curves) based on specific conditions.

If the prognostic requirements cannot be met, the design cycle could continue, conceptually, by suggesting system changes. These include requirements modifications or system design modifications. The scope of these changes depends on the design commitments made so far as well as overall system goals. However, the prognostics design can be seen (and should be seen) as integrated into the whole system design processes (as opposed to as an add-on step after the system design is done).

D. Information requirements

Roemer et al. [26] provide an excellent summary of information requirements for various prognostic approaches (reproduced in Table 2). They conclude that as the fidelity of the prognostics approach increases from evidence-based approaches to physics-based models, there is greater need for additional sensor data and detailed failure models. This analysis does not make any statement about the quality of the resulting remaining life estimates.

The constraints posed by the requirements and sensors have a strong correlation with the accuracy of the remaining life prediction as well as the associated uncertainty. Where models are not available, accuracy decreases while uncertainty bounds typically increase. A similar relationship is observed where information from sensors becomes sparser, either because they are poor predictors or because they are simply missing.

E. Uncertainty Management

In most dynamic systems, there is a great deal of uncertainty about the current state of the component and even more about future usage of the components. This means the prediction about the component will typically not be known with certainty [27]. Because of this, prognostic systems may need to have a method for communicating their confidence in each prediction.

Uncertainty management is a fundamental aspect of

TABLE 2
PROGNOSTIC INFORMATION REQUIREMENTS [26]

	Experience-based	Evolutionary	Physics-based
Engineering Model	Not required	Beneficial	Required
Failure History	Required	Not required	Beneficial
Past Operating Conditions	Beneficial	Not required	Required
Current Conditions	Beneficial	Required	Required
Identified Fault Patterns	Not required	Required	Required
Maintenance History	Beneficial	Not required	Beneficial
In General	No Sensors / No Model	Sensors / No Model	Sensors and Model

prognostics both in electromechanical systems and in medicine. Not surprisingly, the basic ideas underlying survival analysis in medicine are similar to reliability analysis of electromechanical systems mentioned earlier (e.g., survival probability $S(t)$ is similar to reliability $R(t)$). However, whereas survival analysis is a very common method in medicine, reliability analysis is not commonly employed in *complex* electromechanical systems. If uncertainty becomes too large, there may be no gain in the resulting action because at a given risk level, the action will always be immediate. Knowing that the average failure is far in the future offers little value when uncertainty is large. Much of the prognostic work is therefore (besides getting an accurate estimate of the remaining time) geared towards managing (i.e., reducing) the uncertainty.

F. Elements of a Universal Research Methodology for PHM

A universal research methodology for prognostics needs to address all of the issues discussed above. To that end, the following elements are required:

- Establish the business case for a prognostic solution that is derived from high-level system requirements.
- Select the set of metrics that support a particular prognostic life estimation requirement.
- Select faults of interest by employing FMECA or similarly rigorous methods.
- For the highest ranked faults (by criticality and frequency), determine the root cause.
- Determine whether existing sensors can support a prognostic solution (or establish a case for needed sensors and data acquisition/processing capability).
- Determine the most appropriate approach, trading off desired performance, resources available for development, acceptable uncertainty, etc.
- Determine the appropriate number of component samples required for testing to reach a statistically significant conclusion at the required confidence level (typically 95% or above).
- Develop test scenarios to operate and age the components under conditions representative of actual use (realistic loads and environmental conditions). If possible, test component samples in the training set all the way to failure. Collect and analyze operational and environmental data as well as direct and indirect indicators of aging from the fleet of components under study.
- If needed, develop a system/subsystem/component model that describes the operation of the system under nominal and – ideally – under abnormal conditions.
- Develop RUL algorithms using data gathered from the training set.
- Develop an uncertainty management model, possibly in conjunction with the RUL algorithms.
- Estimate RUL on components at frequent intervals and compare RUL estimates with actual (eventual) failure times.
- Measure and report accuracy-based metrics, precision-based metrics, and robustness-based metrics for RUL estimation.

- Verify and validate the models and algorithms developed.

VI. CONCLUSIONS AND NEXT STEPS

In this paper, we discussed the traditional approaches to prognostic RUL estimation in engineering and medicine. Despite superficial differences, the two disciplines are based on similar concepts and have similar objectives. The concepts of natural history and disease progression are similar to the concepts of damage accumulation and fault progression in engineering. Risk factors in medicine are similar to prognostic indicators in engineering. In medicine, the purpose of prognostic life estimation is to determine the optimal treatment policy in order to or maximize life expectation for the individual patient while satisfying quality-of-life goals. In engineering, the purpose of prognostic RUL estimation is to determine the optimal maintenance policy in order to minimize the total lifecycle cost while satisfying overall system safety goals.

One might conjecture that disease progression in medicine is subject to substantial variability due to factors such as genetics, environmental factors, etc. Damage accumulation in engineered systems is also subject to variability due to factors such as build quality, operating conditions, etc. The medical approach normalizes such variability through rigorous application of methods such as randomization, elimination of sample bias, properly-selected study population sizes, and careful assessment of statistical significance. It is time to start expecting similar rigor in PHM research studies.

PHM has not yet become standard business practice in the management of complex engineered systems. This is largely due to the lack of scientific rigor in PHM research. The PHM community needs to reach an agreement on systematic, objective, and quantitative measures that can be used as prognostic metrics. Moreover, PHM research studies should be based on clear scientific hypotheses and scientific methods designed to prove or refute the *null hypotheses* based on common prognostic metrics and at actionable levels of statistical significance. As with medicine, these studies may be retrospective (i.e., based on analysis of previously-collected data) or prospective (i.e., based on a data collection plan). The studies may result in the establishment of simple prognostics measures similar to the medical prognostic metrics mentioned earlier. For example, levels of vibration in a rotating component may be classified into a small number of discrete prognostic scores, and these scores may be mapped to appropriate logistics actions.

A. ROLE OF GOVERNMENT AGENCIES

Rigorous medical studies are very expensive to design and conduct. In contrast, typical PHM studies are conducted through Small Business Innovative Research (SBIR) projects in the U.S., which are often too low-budget and short-term to provide meaningful results. Too often, funding priorities and resources change drastically due to programmatic or administrative considerations, seriously impacting or disrupting research projects and making it futile to plan or execute long-term roadmaps.

In the U.S., medical research is funded through two principal sources: the National Institutes of Health (NIH) and pharmaceutical companies (less important funding sources include private research foundations). Unlike government agencies that fund PHM research, NIH research goals are long-term and rather deterministic: for example, NIH will continue to fund cancer research as long as cancer continues to be a significant cause of mortality and morbidity (of course, even NIH research directions are not entirely immune to non-scientific influences, but such disruptions are exceptions rather than the norm).

Unfortunately, given the limited budgets available for PHM research and the fractured nature of the funding sources, there is no current mechanism (in the U.S.) to fund long-term, substantial PHM research projects that can produce actionable results. This is a major problem, since high-quality prognostics research is likely to consume substantial time and resources regardless of whether the domain is medicine or engineering.

One possible solution is for government funding agencies (at a minimum, DARPA, NASA, Air Force, Army, Navy, and the National Science Foundation) to form an alliance to establish a joint research program in prognostics, pool their resources, define a long-term research program, and maintain it for the long term. Despite high-level differences between systems and platforms, underlying failure mechanisms are universal: a ball bearing fails the same way in a submarine as it does on a spacecraft since they are both subject to the same universal laws of physics. There is no defensible reason for all these agencies to define their own PHM programs, conduct similar projects, and reach similar conclusions time after time. Over time, such an alliance could grow to include overseas agencies and funding sources. The Technical Cooperation Program (TTCP) that involves government agencies of the U.S., Great Britain, Canada, Australia, and New Zealand is a great role model for an extended PHM cooperation across national borders [28].

B. ROLE OF INDUSTRY

A final comment is about the role of industry in PHM research. As stated earlier, one of the major sources of funding for medical prognostic research is the pharmaceutical industry. In contrast, manufacturers or integrators of engineered systems provide very little internal funding for PHM research. Rather, PHM research in industry depends almost entirely on government funding. Furthermore, unlike medical prognostic research which is often published in great detail to establish the credibility of a certain drug or treatment, industrial PHM research is often kept proprietary. Critical details (such as detailed RUL estimation models) are not released by companies. Unfortunately, the secrecy around industrial PHM research hurts the credibility of the discipline and prevents the dissemination of fundamental knowledge about how engineered systems fail and how such failures can be predicted. In addition, there is considerable resistance by the OEMs and integrators to adopt innovative and promising solutions from small businesses, of which many were funded through SBIR programs. This is in large part due to reluctance of the industry to embrace technology that was not invented

in-house and the associated unease about licensing 3rd party technology.

We do not have a solution in mind for the lack of proper scientific disclosure and limited adoption of prognostic knowledge in the industry. Rather, we will conclude by posing a question: *what will it take the industry to adopt and share PHM knowledge in a pre-competitive fashion so that we all benefit from safer, more reliable, and more cost-effective engineered systems?*

REFERENCES

- [1] M. Revor and E. Bechhoefer, "Rotor track and balance cost benefit analysis and impact on operational availability," American Helicopter Society 60th Annual Forum, Baltimore, MD, 2004.
- [2] S. Marble and D. Tow, "Bearing Health Monitoring and Life Extension in Satellite Momentum/Reaction Wheels," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2006.
- [3] B. Saha, K. Goebel, S. Poll and J. Christophersen, "An Integrated Approach to Battery Health Monitoring using Bayesian Regression and State Estimation", in Proc. AUTOTESTCON, 2007 IEEE, 646-653, 2007.
- [4] J. Gu, D. Barker, and M. Pecht, "Prognostics Implementation of Electronics under Vibration Loading," Microelectronics Reliability, Vol. 47, Issue 12, pp. 1849-1856, Dec. 2007.
- [5] D. Luchinsky, V. Osipov, V. Smelyanskiy, et al., "Model Based IVHM System for the Solid Rocket Booster," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [6] A. Hess, G. Calvello, and T. Dabney, "PHM a Key Enabler for the JSF Autonomic Logistics Support Concept," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2004.
- [7] P.H. Barton, "Prognostics for Combat Systems of the Future," IEEE Instrumentation & Measurement Magazine, Vol. 10, pp. 10-14, August 2007.
- [8] V.V. Osipov, D.G. Luchinsky, V.N. Smelyanskiy, et al., "In-Flight Failure Decision and Prognostics for the Solid Rocket Booster", AIAA 43rd AIAA/ASME/SAE/ ASEE Joint Propulsion Conference and Exhibit, Cincinnati, OH, July 8-11, 2007.
- [9] R. A. Hess, "From Health and Usage Monitoring to Integrated Fleet Management – Evolving Directions for Rotorcraft," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2005.
- [10] A. Hess, G. Calvello, P. Frith, et al., "Challenges, Issues, and Lessons Learned Chasing the "Big P": Real Predictive Prognostics Part 2," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2006.
- [11] P.J.F. Lucas and A. Abu-Hanna, "Prognostic methods in medicine," Artificial Intelligence in Medicine, 15: 105-119, 1999.
- [12] The International Non-Hodgkin's Lymphoma Prognostic Factors Project, "A predictive model for aggressive non-Hodgkin's lymphoma," N Engl J Med 1993;329(14):987-94.
- [13] D.R. Cox and D. Oakes. Analysis of Survival Data. London: Chapman and Hall, 1984.
- [14] M.A. van Gerven, B.G. Taal, and P.J. Lucas, "Dynamic Bayesian networks as prognostic models for clinical patient management," in J. Biomedical Informatics, Feb 2008.
- [15] B.P. Leão, K.T. Fitzgibbon, et al., "Cost-Benefit Analysis Methodology for PHM Applied to Legacy Commercial Aircraft," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [16] A. Saxena J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for Evaluating Performance of Prognostic Techniques", submitted to PHM'08, Denver, CO, 2008.
- [17] K. Goebel, W. Yan, N. Eklund, H. Qiu, "Technology Selection for Prognostics", Poster Presentation, ISHM conference, Cincinnati, 2006.
- [18] D.He and E. Bechhoefer, "Development and Validation of Bearing Diagnostic and Prognostic Tools using HUMS Condition Indicators," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [19] S.L. Chen, M. Craig, R. Callan, H. Powrie and Robert Wood, "Use of Artificial Intelligence Methods for Advanced Bearing Health Diagnostics and Prognostics," in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [20] M. Baybutt, S. Nanduri, P. W. Kalgren, et al., "Seeded Fault Testing and In-situ Analysis of Critical Electronic Components in EMA Power

- Circuitry,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [21] S. Kumar, V. Sotiris and M. Pecht, “Mahalanobis Distance and Projection Pursuit Analysis for Health Assessment of Electronic Systems,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [22] J. J. Palazzolo, L. D. Scheunemann, and J. R. Hartin, “Leakage Fault Detection Method for Axial-Piston Variable Displacement Pumps,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [23] J.P. Hofmeister, P. Lall, D. Panchagade, et al. “Ball Grid Array (BGA) Solder Joint Intermittency Detection: SJ BIST,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [24] K. Keller, J. Sheahan, J. Roach, et al., “Power Conversion Prognostic Controller Implementation for Aeronautical Motor Drives,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [25] B. Saha and K. Goebel, “Uncertainty Management for Diagnostics and Prognostics of Batteries using Bayesian Techniques,” in Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2008.
- [26] M. Roemer, C. Byington, G. Kacprzynski, and G. Vachtsevanos, “An Overview of Selected Prognostic Technologies with Reference to an Integrated PHM Architecture”, Proceedings of ISHEM Forum, Napa Valley, CA, November 7-10, 2005.
- [27] Barlas, I., Zhang, G., Propes, N., Vachtsevanos, G., Galie, T., Thakker, A., 2005. Confidence Metrics and Uncertainty Management in Prognosis, MARCON, Knoxville, May, 2003.
- [28] The Technical Cooperation Program, <http://www.dtic.mil/ttcp/>