Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation

Abhinav Saxena, *Member IEEE,* Kai Goebel,
Don Simon, *Member, IEEE,* Neil Eklund, *Member IEEE*

Abstract— This paper describes how damage propagation can be modeled within the modules of aircraft gas turbine engines. To that end, response surfaces of all sensors are generated via a thermo-dynamical simulation model for the engine as a function of variations of flow and efficiency of the modules of interest. An exponential rate of change for flow and efficiency loss was imposed for each data set, starting at a randomly chosen initial deterioration set point. The rate of change of the flow and efficiency denotes an otherwise unspecified fault with increasingly worsening effect. The rates of change of the faults were constrained to an upper threshold but were otherwise chosen randomly. Damage propagation was allowed to continue until a failure criterion was reached. A health index was defined as the minimum of several superimposed operational margins at any given time instant and the failure criterion is reached when health index reaches zero. Output of the model was the time series (cycles) of sensed measurements typically available from aircraft gas turbine engines. The data generated were used as challenge data for the Prognostics and Health Management (PHM) data competition at PHM'08.

Index Terms—Damage modeling, Prognostics, C-MAPSS, Turbofan engines, Performance Evaluation

CONTENTS

I. INTRODUCTION

Data-driven prognostics faces the perennial challenge of the lack of run-to-failure data sets. In most cases real-world data contain fault signatures for a growing fault but no or little data capture fault evolution until failure. Procuring actual system fault progression data is typically time consuming and expensive. Fielded systems are, most of the time, not properly instrumented for collection of relevant data. Those fortunate enough to be able to collect long-term data for fleets of systems tend to – understandably – hold the data from public release for proprietary or competitive reasons. Few public data repositories (e.g., [1]) exist that make run-to-failure data available. The lack of common data sets, which researchers can use to compare their approaches, is impeding progress in the field of prognostics. While several forecasting competitions have been held in the past (e.g., [2-7]), none have been conducted with a PHM-centric focus. All this provided the motivation to conduct the first PHM data challenge. The task was to estimate remaining life of an unspecified system using historical data only, irrespective of the underlying physical process.

For most complex systems like aircraft engines, finding a suitable model that allows the injection of health related changes certainly is a challenge in itself. In addition, the question of how the damage propagation should be modeled within a model needed to be addressed. Secondary issues revolved around how this propagation would be manifested in sensor signatures such that users could build meaningful prognostic solutions.

In this paper we first define the prognostics problem to set the context. Then the following sections introduce the simulation model chosen, along with a brief review of health parameter modeling. This is followed by a description of the damage propagation modeling, a description of the competition data, and a discussion on performance evaluation.

II. PROGNOSTICS

To avoid confusion, we define prognostics here exclusively as the estimation of remaining useful component life. The remaining useful life (RUL) estimates are in units of time (e.g., hours or cycles). End-of-life can be subjectively determined as a function of operational

thresholds that can be measured. These thresholds depend on user specifications to determine safe operational limits.

Prognostics is currently at the core of systems health management. Reliably estimating remaining life holds the promise for considerable cost savings (for example by avoiding unscheduled maintenance and by increasing equipment usage) and operational safety improvements. Remaining life estimates provide decision makers with information that allows them to change operational characteristics (such as load) which in turn may prolong the life of the component. It also allows planners to account for upcoming maintenance and set in motion a logistics process that supports a smooth transition from faulty equipment to fully functional. Aircraft engines (both military and commercial), medical equipment, power plants, etc. are some of the common examples of these types of equipment.

Therefore, it is not surprising that finding solutions to the prognostics problem is a very active research area. The fact that most efforts are focusing on data-driven approaches seems to reflect the desire to harvest low-hanging fruit as compared to model-based approaches, irrespective of the difficulties in gaining an access to statistically significant amounts of run-to-failure data and common metrics that allow a comparison between different approaches.

Next we will describe how a system model can be used to generate run-to-failure data that can then be utilized to develop, train, and test prognostic algorithms.

III. SYSTEM MODEL

Tracking and predicting the progression of damage in aircraft engine turbo machinery has some roots in the work of Kurosaki *et al.* [8]. They estimate the efficiency and the flow rate deviation of the compressor and the turbine based on operational data, and utilize this information for fault detection purposes. Further investigations have been done by Chatterjee and Litt on on-line tracking and accommodating engine performance degradation effects represented by flow capacity and efficiency adjustments [9]. In [10], response surfaces for various sensors outputs are generated for a range of flow and efficiency values using a simulation model. These response surfaces are used to identify flow and efficiency health parameters of an actual engine by optimally matching the set of sensor readings with simulated sensor values, resulting in only one possible solution. The process chosen here continues on a similar path and follows closely the one described in [10].

An important requirement for the damage modeling process was the availability of a suitable system model that allows input variations of health related parameters and recording of the resulting output sensor measurements. The recently released C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) [11] meets these requirements and was chosen for this work.

A. C-MAPSS

C-MAPSS is a tool for simulating a realistic large commercial turbofan engine. The software is coded in the MATLAB® and Simulink® environment, and includes a number of editable input parameters that allow the user to enter specific values of his/her own choice regarding operational profile, closed-loop controllers, environmental conditions, etc. C-MAPSS simulates an engine model of the 90,000 lb thrust class and the package includes an atmospheric model capable of simulating operations at (i) altitudes ranging from sea level to 40,000 ft, (ii) Mach numbers from 0 to 0.90, and (iii) sea-level temperatures from –60 to 103 °F. The package also includes a power-management system that allows the engine to be operated over a wide range of thrust levels throughout the full range of flight conditions.

In addition, the built-in control system consists of a fan-speed controller, and a set of regulators and limiters. The latter include three high-limit regulators that prevent the engine from exceeding its design limits for core speed, engine-pressure ratio, and High-Pressure Turbine (HPT) exit temperature; a limit regulator that prevents the static pressure at the High-Pressure Compressor (HPC) exit from going too low; and an acceleration and deceleration limiter for the core speed. A comprehensive logic structure integrates these control-system components in a manner similar to that used in real engine controllers such that integrator-windup problems are avoided. Furthermore, all of the gains for the fan-speed controller and the four limit regulators are scheduled such that the controller and regulators perform as intended over the full range of flight conditions and power levels. The engine diagram in Figure 1 shows the main elements of the engine model and the flow chart in Figure 2 shows how various subroutines are assembled in the simulation.
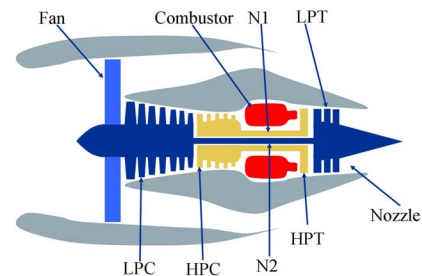

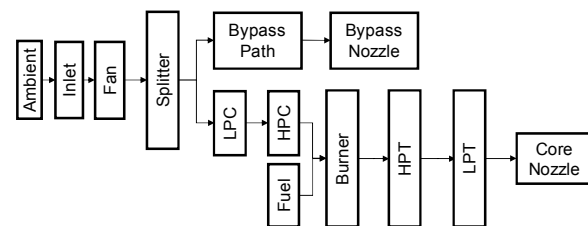Figure 1. Simplified diagram of engine simulated in C-MAPSS [11].

Figure 2. A layout showing various modules and their connections as modeled in the simulation [11].

CMAPSS can be operated either in open-loop (without any controller) or in closed loop (with the engine and its control system) configurations. For the purpose of this

paper, we worked exclusively with the closed-loop configuration. C-MAPSS has 14 inputs (Table 1) and can produce several outputs. Table 2 lists the outputs that were used for the challenge data. The inputs include fuel flow and a set of 13 health-parameter inputs that allow the user to simulate the effects of faults and deterioration in any of the engine's five rotating components (Fan, LPC, HPC, HPT, and LPT). The outputs include various sensor response surfaces and operability margins. A total of 21 variables out of 58 different outputs available from the model were used in this study. C-MAPSS provides a set of Graphical User Interfaces (GUIs) to simplify input and output control for a variety of possible uses, including open-loop analysis, controller-design, and simulation of response of the engine and its control system in a variety of situations. However, for the purpose of this data generation exercise, we ran the model in batch-mode without using the GUIs.

Table 1. C-MAPSS inputs to simulate various degradation scenarios in any of the five rotating components of the simulated engine. For example, to simulate HPC degradation HPC flow and efficiency modifiers were used (highlighted in gray).

Name	Symbol
Fuel flow	W_f
Fan efficiency modifier	fan_eff_mod
Fan flow modifier	fan_flow_mod
Fan pressure-ratio modifier	fan_PR_mod
LPC efficiency modifier	LPC_eff_mod
LPC flow modifier	LPC_flow_mod
LPC pressure-ratio modifier	LPC_PR_mod
HPC efficiency modifier	HPC_eff_mod
HPC flow modifier	HPC_flow_mod
HPC pressure-ratio modifier	HPC_PR_mod
HPT efficiency modifier	HPT_eff_mod
HPT flow modifier	HPT_flow_mod
LPT efficiency modifier	LPT_eff_mod
HPT flow modifier	LPT_flow_mod

Table 2. C-MAPSS outputs to measure system response. Margins were used for health index calculation only and were not available to the participants explicitly.

Symbol	Description	Units
Parameters available to participants as sensor data		
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia
P15	Total pressure in bypass-duct	psia
P30	Total pressure at HPC outlet	psia
Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure ratio (P50/P2)	--
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass Ratio	--
farB	Burner fuel-air ratio	--
htBleed	Bleed Enthalpy	--
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s
Parameters for calculating the Health Index		
T48 (EGT)	Total temperature at HPT outlet	°R
SmFan	Fan stall margin	--
SmLPC	LPC stall margin	--
SmHPC	HPC stall margin	--

B. Response Surfaces

To ensure that the output of the model was producing correct results, we first generated response surfaces for sensed outputs and operability margins from C-MAPSS as a function of flow and efficiency for specific modules. These were compared with those published by Goebel et al. [10]. Although it was not expected that the units would match (in fact, none were revealed in [10]), it was expected that the qualitative response should be similar. For instance, with an increase in flow and efficiency, the response surface behaved in a similar fashion as obtained from the real aircraft engine used in [10]. For each module in the gas path (HPC, HPT, and LPT), the efficiencies and flows were incrementally changed and C-MAPSS was then run under different cruise conditions randomly chosen at each time step. Some resulting HPC module response surfaces for the high pressure compressor stall margin and the exhaust gas temperature (EGT) are shown in Figure 3 and Figure 4, respectively.
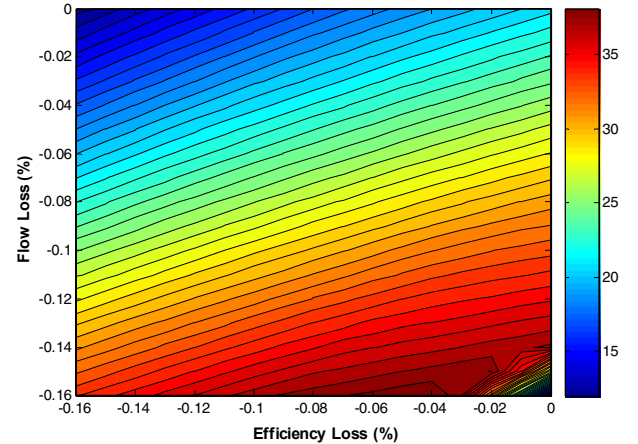

Figure 3. Response surface of HPC stall margin as a function of efficiency and flow losses simulating degradation in HPC module.

The range of the flow and efficiency loss is the same in all figures. Response surfaces for the other modules' sensors and operability margins were also generated using the same process and verified.
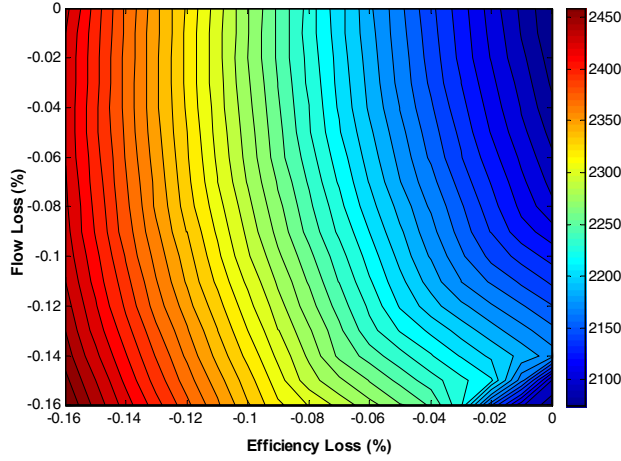
Figure 4. Response surface for Exhaust Gas Temperature (EGT) as a function of efficiency and flow losses.

IV. DAMAGE PROPAGATION MODELING

Having decided on the system model, the next hurdle is to model the propagation of damage. Common models used across different application domains include the Arrhenius model, the Coffin-Manson mechanical crack growth model, and the Eyring model (for more than three stresses or when the above models are not satisfactory) [12]. These models come in numerous variations that will not be discussed here.

A. Arrhenius

The Arrhenius model has been used for a variety of failure mechanisms. Traditionally, it has been applied to those that depend on chemical reactions, diffusion processes or migration processes. While this covers many of the non-mechanical (or non-material fatigue) failure modes that cause electronic equipment failure, lately, variations of the Arrhenius equation have also been employed for mechanical and other non-traditional applications. The operative equation is:

$$ t_f = A e^{\frac{\Delta H}{kT}}, \tag{1} $$

where

t_f is the time to failure,

T is the temperature at the point when the failure process takes place,

k is Boltzmann's constant,

A is a scaling factor, and

ΔH is the activation energy.

B. Coffin-Mason Mechanical Crack Growth Model

A model more typically applied to mechanical failure, material fatigue or material deformation is the (modified) Coffin-Manson model. It has been successfully used to model crack growth in solder and other metals due to repeated temperature cycling as equipment is turned on and off. The operative equation is

$$ N_f = A f^{-\alpha} \Delta T^{-\beta} G(T_{\max}), \tag{2} $$

where

N_f is the number of cycles to failure,

A is a scaling factor,

f is the cycling frequency,

ΔT is the temperature range during a cycle,

$G(T_{max})$ is an Arrhenius term evaluated at the maximum temperature reached in each cycle,

α is the cycling frequency exponent, and

β is the temperature range exponent.

C. Eyring Model

The Eyring Model originates in chemical reaction rate theory and has a theoretical basis in chemistry and quantum mechanics. It describes how time to failure varies with stress. The base model includes temperature and can be expanded to include other relevant stresses. The temperature term by itself is very similar to the Arrhenius empirical model, explaining why that model has been so successful in establishing the connection between the ΔH parameter and the quantum theory concept of "activation energy needed to cross an energy barrier and initiate a reaction".

The model for temperature and additional stress terms takes the general form:

$$ t_f = A T^{\alpha} e^{\left(\frac{\Delta H}{kT} + \left(B + \frac{C}{T}\right)S_1 + \left(D + \frac{E}{T}\right)S_2\right)}, \tag{3} $$

where

t_f is the time to failure,

α, ΔH, A, B, C, D, and E are constants that determine acceleration between stress combinations,

S_1 and S_2 are relevant stresses (e.g., some function of voltage or current),

T is temperature in degrees Kelvin, and

k is the Boltzmann's constant.

The general Eyring model includes terms that have stress and temperature interactions. A disadvantage of the Eyring model is that it has a relatively large number of parameters that need to be determined.

D. Damage Propagation Model for the Challenge Problem

Common to all degradation models is the exponential behavior of the fault evolution. This and the observation of similar degradation trends in practice [10] motivated our use of an exponential term while modeling changes of health parameters in C-MAPSS. For the purpose of a physics-inspired data-generation approach, we assume a generalized equation for wear, $w = Ae^{B(t)}$, which ignores micro-level processes but retains macro-level degradation characteristics. Assuming further an upper wear threshold, th_w, that denotes an operational limit beyond which the component/subsystem cannot be used, the generalized wear equation can be rewritten as a time varying health index, $h(t)$, by subtracting wear from the upper wear threshold and normalizing it with respect to the upper wear threshold as $h(t) = 1 - Ae^{B(t)}/ th_w$. Recasting parameter $A/th_w = e^a$ and expressing $B(t) = t^b$, the health equation can be written as

$$ h(t) = 1 - \exp\{at^b\}. \tag{4} $$

Generally, the system will be observed with some non-

zero initial degradation, d, (allowing the data-generation process to start at an arbitrary point in the wear-space) which will be modeled as an additive term to yield

$$h(t) = 1 - d - \exp\{at^b\}. \qquad (5)$$

The health index can be used to model different phenomena within a subsystem. Specifically, for aircraft engine modules like the compressor and turbine sections, the health is described both by efficiency (e) and flow (f). Trajectories for flow and efficiency vary for different fault modes [4] and are modeled as separate health related indices as shown below.

$$e(t) = 1 - d_e - \exp\{a_e(t)t^{b_e(t)}\}$$
$$f(t) = 1 - d_f - \exp\{a_f(t)t^{b_f(t)}\} \qquad (6)$$

The terms $e(t)$ and $f(t)$ are then aggregated to form the overall health index $H(t)$, the engine simulation response to the given input values.

$$H(t) = g(e(t), f(t)) \qquad (7)$$

where, the function g is the minimum of all operative margins considered (here those for fan, HPC, HPT, and EGT), i.e.

$$g(e(t), f(t)) = \min(m_{Fan}, m_{HPC}, m_{HPT}, m_{EGT}), \qquad (8)$$

where, the margins m in turn are functions of efficiency $e(t)$ and flow $f(t)$. Calculation of the health index is further discussed in section V.D

V. APPLICATION SCENARIO

The scenario developed for the challenge data tracks a number of aircraft engines throughout their usage history. A particular engine unit may be employed under different flight conditions from one flight to another. Depending on various factors the amount and rate of damage accumulation will be different for each engine. It is assumed that the amount of damage accumulated during a particular flight will not be directly quantifiable solely based on flight duration and flight conditions, and hence, one must rely on information extracted from sensor data collected during each flight. This scenario models engine performance degradation due to wear and tear based on the usage pattern of the engines and not necessarily due to any particular fault mode. Therefore, sudden degradation during a flight is rather unlikely. This allows us to take one measurement snapshot per flight to characterize the engine health during or right after that flight. Further, the effects of between-flight maintenance have not been explicitly modeled but have been incorporated as the process noise. This allows the engine performance parameters (flow and efficiency) to improve within allowable limits at any point and hence the loss in efficiency or flow is not locally monotonic (see Figure 5).

In order to simulate the scenario explained above we needed to address several issues in order to make it more realistic. Some of these issues and their resolutions are discussed next.
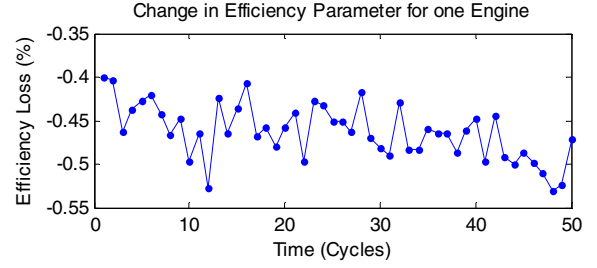

Figure 5. Performance parameters like efficiency and flow may not change monotonically to model process noise that also incorporates between flight maintenance operations, which may lead to improved performance in subsequent flights.

A. Initial Wear

Initial wear can occur due to manufacturing inefficiencies and are commonly observed in real systems. Although it is not considered abnormal, it can make a difference in useful operational life of a component. Initial wear can also be modeled by variations in flow and efficiencies of the various modules, although the magnitude of such variations is relatively low. Chatterjee and Litt [9] give examples for the degree of wear that an engine might experience with progressive usage. These numbers were used as reference values for the challenge data and are recited in Table 3 for reference.

Table 3. Engine wear as manifested in flow and efficiency changes [9].

	Initial Wear (%)	Wear 3000 Cycles (%)	Wear 6000 Cycles (%)
Fan_Efficiency	- 0.18	- 1.5	- 2.85
Fan_Flow	- 0.26	- 2.04	- 3.65
LPC_Efficiency	- 0.62	- 1.46	- 2.61
LPC_Flow	- 1.01	- 2.08	- 4.00
HPT_Efficiency	- 0.48	- 2.63	- 3.81
HPT_Flow	+ 0.08	+ 1.76	+ 2.57
LPT_Efficiency	- 0.10	- 0.54	- 1.08
LPT_Flow	+ 0.08	+ 0.26	+ 0.42

B. Noise

Characterizing noise in a system may be a non-trivial undertaking. Of various sources, the main sources of noise while assessing the true state of system's health are manufacturing and assembly variations, process noise (due to factors not taken in to account while modeling the process), and measurement noise to name a few important ones. These noise sources introduce their respective contributions at different stages of the process and a combined effect is observed in the sensor measurements at the end. A simple approach to model this combined effect is to use approximate models (e.g. random noise models) [13, 14]. In other cases sophisticated noise model identification techniques may be employed [15] if real data are available for such analyses. In both situations a PHM practitioner is faced with characterizing and de-noising tasks before

developing diagnostics or prognostics algorithms.

In this study, since there was no real data available to characterize true noise levels, simplistic normal noise distributions were assumed based on information available from the literature [13, 16-18]. However, to make the signal noise non-trivial, mixture distributions were used and all of these noise sources were combined to present similar challenges in a realistic manner. Since any degradation is modeled by varying (generally decreasing) the efficiency and flow parameters for the engine, the initial wear due to manufacturing and assembly variations was modeled by selecting initial values, e_0 and f_0, for e and f parameters (eq. 6) from a random distribution, such that the maximum initial deterioration is bounded within 1% degradation of the healthy condition as cited in [14]. Therefore, each health index trajectory starts with a number between 1 and 0.99.

To model the process noise, first the degradation trajectory parameters, a_k and b_k, corresponding to a unit under test k were chosen from a normal distribution. Together with e_0 and f_0, these parameters define a deterministic trajectory for degradation for a particular engine. This trajectory was then masked by a mixture of two random distributions with slightly different variances. It has been shown that mixture noise models are more difficult to characterize even if they consist of simple individual components [19]. This contaminated trajectory was fed to the engine model simulation and corresponding sensor outputs listed in Table 2 were collected after the system response reached a steady state. This way, the input process noise gets filtered through system dynamics and overall effect is observed in the output. Lastly, a random measurement noise component was added to all output channels in order to impose sensor noise. This multistage noise contamination resulted in complex noise characteristics often observed in real data and posed a similar challenge in front of competition participants to carry out appropriate de-noising operations.

C. Data Generation

The process for using the model was as follows:

1. Choose initial deterioration (f_0, e_0).
$$e_0 \in [0.99, 1]$$
$$f_0 \in [0.99, 1] \tag{9}$$

2. Impose an exponential rate of change for flow and efficiency loss for each data set, denoting an otherwise unspecified fault with increasingly worsening effect as described in equation (4). This results in the overall health index, $H(t)=g(e(t), f(t))$, varying as a function of time. The randomly chosen direction and evolution of faults is constrained by
$$f_i, e_i \leq 1\%,$$
$$a_k \in [0.001, 0.003], \tag{10}$$
$$b_k \in [1.4, 1.6], \quad k=1,2.$$

3. Stop when health $H= 0$ (this is our failure criterion).
4. Superimpose measurement noise to the output data.

The output is a time series (cycles) of observables (Nf, Nc, w_f, …) at cruise snapshots that were produced by modifying flow and efficiencies of the HPC module from initial settings (indicating normal deterioration) to values corresponding to failure threshold. Degradation of other modules was not included intentionally in the challenge data.

D. Health Index Calculation

The safe operation region for an engine is determined via operability margins - how far the engine is operating from various operational limits like stall and temperature limits. These margins can be calculated by computing the distance between current engine state and pre-defined limits. Among the margins considered, some are directly measurable, such as core speed limits and upper EGT thresholds. Others are "virtual" margins established through simulation. Each of these margins are normalized to the range [0,1], where one signifies a perfectly healthy system and zero denotes a system whose stall margin has reduced by a specified limit. For the challenge data this limit was set at 15% for HPC, LPC and fan stall margins and about 2% for the EGT margin. The underlying premise is that if one engine with certain e and f pairing violates either one of operational margins under any possible operational conditions, such as hot day take off, maximum climb, or cruise, its health index would be zero. Otherwise, whichever normalized margin is lower would be its current health index. As shown in Figure 6, these margins change as a function of operational conditions (e.g. throttle resolver angle (TRA), altitude, ambient temperature, etc.). Therefore, the health index must be adjusted according to operational condition as well.
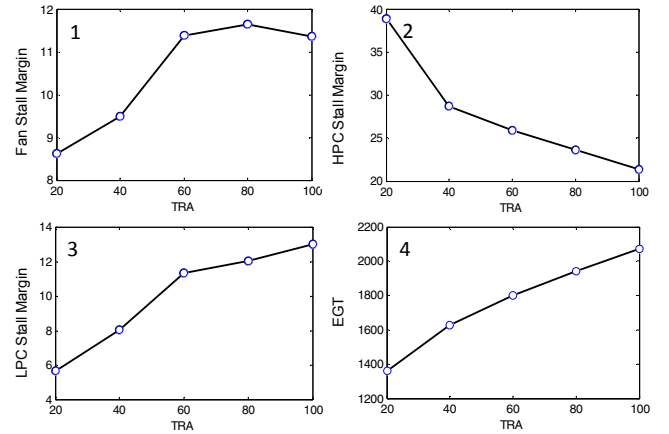

Figure 6. Stall margins vary as a function of operational conditions (TRA in this example).

For the challenge data set six different flight conditions were simulated that comprised of a range of values for three operational conditions: altitude (0-42K ft.), Mach number (0-0.84), and TRA (20-100). Furthermore, these margins change as system degradation takes place. If system degradation is plotted on flow-efficiency axes, various margins indicating the deterioration can be depicted as shown in Figure 7 and Figure 8. A threshold boundary separates the failure region for respective margins.

Depending on the direction of the failure evolution trajectory (simulated by changing *e* and *f* parameters) a threshold may or may not be crossed. Therefore, the overall health index is determined by the margin that approaches the corresponding limit first. For instance, in the following figures, health index is determined by increasing EGT (decreasing EGT margin) as compared to HPC stall margin for all three degradation trajectories. Each degradation trajectory was simulated until the health index reached zero.
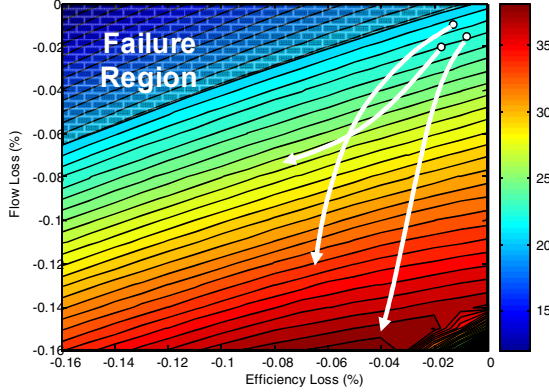

Figure 7. Fault propagation trajectories on HPC stall margin contour map.
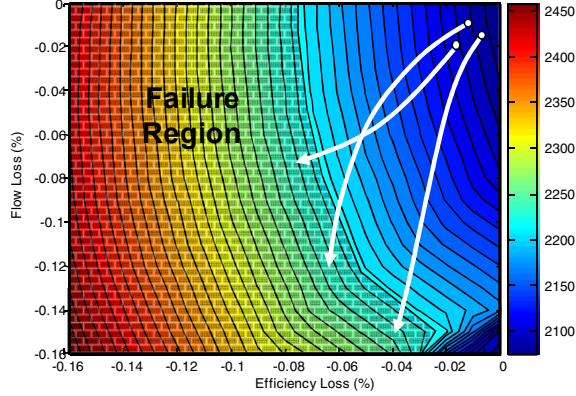

Figure 8. Fault propagation trajectories on EGT contour map with failure threshold.

VI. COMPETITION DATA

The objective was to generate train, test, and validation data sets for development of data-driven prognostics. To that end, a reasonably large number of trajectories were created from C-MAPSS that had the following properties:

1) Simulation of degradation in HPC module under 6 different combinations of Altitude, TRA, and Mach number operational conditions. Sensed margins (fan, HPC, LPC and EGT) were used to compute health index to determine simulation stopping criteria.
2) Time series of observables including operational variables (see Table 1 and Table 2), that change from some undefined initial condition to a failure threshold. Participants were not given access to the health index explicitly and were expected to infer it from the given sensed variables.
3) Division of data into training set, test set, and validation

set. The training set had trajectories that ended at the failure threshold while the test and validation sets were pruned to stop some time prior to the failure threshold.
4) The range of RUL variation was expanded for the validation set to test robustness of the algorithms trained on test data set (a condition that was not announced to the participants). The test data set RULs ranged between 10 and 150 cycles, whereas validation RULs ranged between 6 and 190 cycles. However, all other characteristics like variation in initial wear, noise levels and degradation parameters spread remained changed.

Participants of the challenge were then given details of the scoring function. They could submit their test set results in vector form through a web site where scores were automatically calculated and posted back to the participants, allowing them to improve their algorithms. To avoid over-fitting to the test data, the validation set was withheld and published later, without feedback of the score until after the competition had closed.

VII. PERFORMANCE EVALUATION

Performance evaluation is concerned with employing metrics that help assess if the prognosis meets specifications for the task at hand. In PHM context, since the key aspect is to avoid failures, it is generally desirable to predict early as compared to predicting late. However, in specific situations where failures may not pose life threatening situations and early predictions may instead involve significant economic burden, this equation may change and one may not prefer conservative predictions. Hence, a performance evaluation system should reflect such characteristics to meet specific requirements.

For an engine degradation scenario an early prediction is preferred over late predictions. Therefore, the scoring algorithm for this challenge was asymmetric around the true time of failure such that late predictions were more heavily penalized than early predictions. In either case, the penalty grows exponentially with increasing error. The asymmetric preference is controlled by parameters a_1 and a_2 in the scoring function given below (eq. 11) and Figure 9 shows the score as a function of the error.

$$ s = \begin{cases} \sum_{i=1}^{n} e^{-\left(\frac{d}{a_1}\right)} - 1 \text{ for } d < 0 \\ \sum_{i=1}^{n} e^{\left(\frac{d}{a_2}\right)} - 1 \text{ for } d \geq 0, \end{cases} \tag{11} $$

where

s is the computed score,
n is the number of UUTs,
$d = \hat{t}_{RUL} - t_{RUL}$ (Estimated RUL – True RUL),
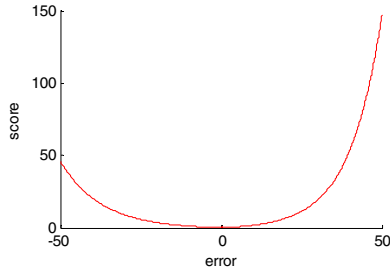$a_1 = 10$, and $a_2 = 13$.

Figure 9. Score as a function of error.

Asymmetric scoring functions like this capture the preference for early prediction pretty well and can be appropriately tuned to quantify the extent of such preference.

While evaluating the results, it was realized that this prognostics metric can be further enhanced in various ways. It must be noted that predicting farther into the future is more difficult than predicting at a time closer to the end of life. Furthermore, it is more important to weigh accuracy of RULs higher when one is closer to the end of life. Keeping these thoughts in mind it may be desirable to assign higher weights for cases with shorter true RULs. Another characteristic of such datasets is that the performance of an algorithm is evaluated from multiple units under test (UUTs), e.g., in fleet applications. Since the metric is a combined aggregate of performance for individual UUTs, an additional correlation metric should be employed to ensure that an algorithm consistently predicts well for all cases as against predicting well for some and poorly for the rest. This idea is illustrated in Figure 10 and Figure 11 shows a simplified asymmetric scoring function used for this illustration.

The various axes shown in Figure 11 show six different cases of prognostics algorithm outputs. Each axis shows estimated RULs and the corresponding true RULs for four UUTs. A simple asymmetric function (see Figure 10) is used to compute scores from RUL predictions for each UUT. As shown, each of these six cases produces an equal aggregated score of six. Clearly one can further differentiate in the output performance based on the suggested correlation metric. In some applications a higher correlation score may be preferred over lower aggregated scores, as a higher correlation may indicate a bias in the algorithm output that can be accordingly adjusted.
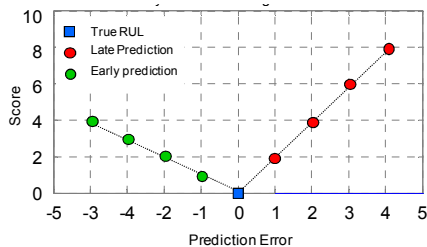

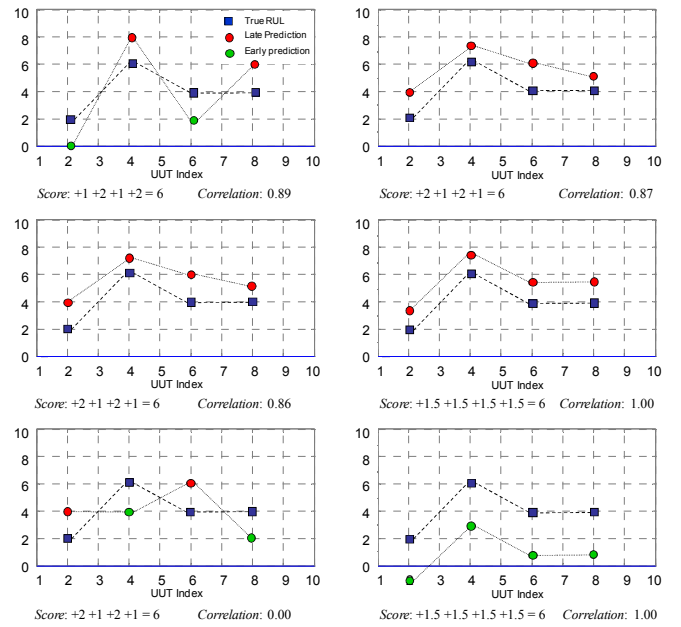Figure 10. A simplified asymmetric scoring function chosen for illustration

Figure 11. Scenarios illustrating various cases where error based aggregated scores may be same but the correlation score distinguishes further between different algorithms.

These suggestions are just to illustrate the point that there may be more modifications possible depending on specific applications and requirements and one must adapt accordingly. A comprehensive review of prognostics metrics is given in [20].

VIII. CONCLUSION

This paper has described how damage propagation can be modeled in various modules of aircraft gas turbine engines for developing and testing prognostics algorithms. A publically available aero-propulsion system simulator, C-MAPSS, was used in this study. Various assumptions and settings have been provided that were used to generate data for the PHM competition at the first international conference on prognostics and health management. Although, the data for the competition consisted of a subset of various possible conditions and settings, an insight into other possibilities can be easily derived. Later, a brief discussion has been provided on the performance evaluation of prognostics algorithms and the aspects of the performance metrics that may be desirable in a PHM application.

REFERENCES

[1]   NASA, "Prognostics Center of Excellence Data Repository", http://ti.arc.nasa.gov/projects/data_prognostics, last accessed on January, 2007.

[2]   H. Madsen, G. Kariniotakis, H. A. Nielsen, T. S. Nielsen, and P. Pinson, "A Protocol for Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models,"

Technical University of Denmark, IMM, Lyngby, Denmark, Deliverable ENK5-CT-2002-00665, 2004.

[3] "NN5: Forecasting Competition for Artificial Neural Networks & Computational Intelligence", http://www.neural-forecasting-competition.com/index.htm, last accessed on July 2008, 2008.

[4] "Time-Series Prediction Competition at 2nd European Symposium on Time Series Prediction (ESTSP'08)", http://www.estsp.org/index.php?pg=welcome, last accessed on July 2008, 2008.

[5] A. Lendasse, E. Oja, O. Simula, and M. Verleysen, "Time Series Prediction Competition: The CATS Benchmark," *Neurocomputing,* vol. 70, pp. 2325-2329, 2007.

[6] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting,* vol. 1, pp. 111-153, 1982.

[7] S. Makridakis and M. Hibon, "The M3-Competition: Results, Conclusions, and Implications," *International Journal of Forecasting,* vol. 16, pp. 451-476, 2000.

[8] M. Kurosaki, T. Morioka, K. Ebina, M. Maruyama, T. Yasuda, and M. Endoh, "Fault Detection and Identification in an IM270 Gas Turbine Using Measurements for Engine Control," *Journal of Engineering for Gas Turbines and Power,* vol. 126, pp. 726-732, 2004.

[9] S. Chatterjee and J. Litt, "Online Model Parameter Estimation of Jet Engine Degradation for Autonomous Propulsion Control," NASA, Technical Manual TM2003-212608, 2003.

[10] K. Goebel, H. Qiu, N. Eklund, and W. Yan, "Modeling Propagation of Gas Path Damage," in *IEEE Aerospace Conference*, Big Sky, MT, 2007.

[11] D. Frederick, J. DeCastro, and J. Litt, "User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)," NASA/ARL, Technical Manual TM2007-215026, 2007.

[12] NIST/SEMATECH, "e-Handbook of Statistical Methods", http://www.itl.nist.gov/div898/handbook/, last accessed on July 2008, 2003.

[13] P.-J. Lu and T.-C. Hsu, "Application of Autoassociative Neural Network on Gas-Path Sensor Data Validation," *Journal of Propulsion and Power,* vol. 18, pp. 879-888, 2002.

[14] R. T. Rausch, K. F. Goebel, N. H. Eklund, and B. J. Brunell, "Integrated In-Flight Fault Detection and Accommodation: A Model-Based Study," *Journal of Engineering for Gas Turbines and Power,* vol. 129, pp. 962-969, 2007.

[15] N. Eklund, "Using Synthetic Data to Train an Accurate Real-World Fault Detection System," in *IMACS Multiconference on Computational Engineering in Systems Applications*, pp. 483-488, 2006.

[16] S. Borguet and O. Leonard, "A Generalised Likelihood Ratio Test for Adaptive Gas Turbine Health Monitoring," in *ASME Turbo Expo 2008: Power for Land, Sea and Air GT2008* Berlin, Germany, 2008.

[17] T. Kobayashi and D. L. Simon, "Evaluation of an Enhanced Bank of Kalman Filters for In-Flight Aircraft Engine Sensor Fault Diagnostics," NASA/ARL, Technical Manual TM2004-213203, 2004.

[18] B. A. Roth, D. L. Doel, and J. J. Cissel, "Probabilistic Matching of Turbofan Engine Performance Models to Test Data," in *ASME Turbo Expo 2005: Land Sea & Air* Reno-Tahoe, Nevada, 2005.

[19] W. E. Yancey, "Working Papers for Mixture Model Additive Noise for Microdata Masking," Statistical Research Division U.S. Bureau of the Census, Washington D.C. May 28 2002.

[20] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for Evaluating Performance of Prognostics Techniques," in *International Conference on Prognostics and Health Management (PHM08)*, Denver CO, 2008.