Enhancing Flight Test Safety with Real-time Early Warning Techniques

Daniel Javorsek II
Air Force Test Center, Edwards AFB, CA

Immanuel Barshi and David Iverson
NASA Ames Research Center, Moffett Field, CA

Contemporary approaches to aerospace vehicle system monitoring rely heavily on thresholds that represent a compromise between providing warning early enough to avert a mishap while simultaneously minimizing false alarms. While this reliance on thresholds has been in place for decades and has permeated both cockpits and control rooms, we often find it insufficient when retrospectively analyzing data from an accident. To modernize and enhance flight test safety, we introduce new methods of monitoring for anomalous patterns of interaction rather than for thresholds exceedance. For systems with a well characterized baseline we show how the Inductive Monitoring System (IMS), utilized by NASA in the aftermath of the Columbia accident, might be implemented in real-time to provide earlier warning than currently employed techniques. For systems without such a baseline, we introduce new developments in statistical methods relating to critical slowing down, first applied in medicine and physics, which show promise for adaptation to flight test. Finally, the familiar resource constrained environment leads to a reliance on increased instrumentation that is challenging the limits of the current "one-sensor, one-indicator" threshold paradigm. Existing methods thus fail to accurately reflect the true complexity of a vehicle rich with interdependent interacting systems. We highlight these concepts in a brief summary of the 2001 Air Transat Flight 236 deadstick landing in the Azores. We then suggest control room and cockpit modifications to better display the information gleaned using these novel monitoring methods.

INTRODUCTION

Contemporary system health management relies heavily on thresholds to alert operators and mission controllers to system failure. In highly federated and well characterized, simple systems this works well. In fact, an isolated laboratory test might be very useful in determining a threshold that indicates the system is operating outside of the normal range. Ultimately the threshold that is selected represents a compromise between providing warning early enough to take action to affect the outcome while also minimizing corrective action when the system is

actually operating within the normal range (often referred to as false alarms). While this "one-sensor, one-indicator" paradigm works well for simple, federated systems (see Figure 1), we repeatedly observe that it is inadequate when examining an aerospace mishap. As many operators are acutely aware, aerospace vehicles often have different limits for a variety of system states.

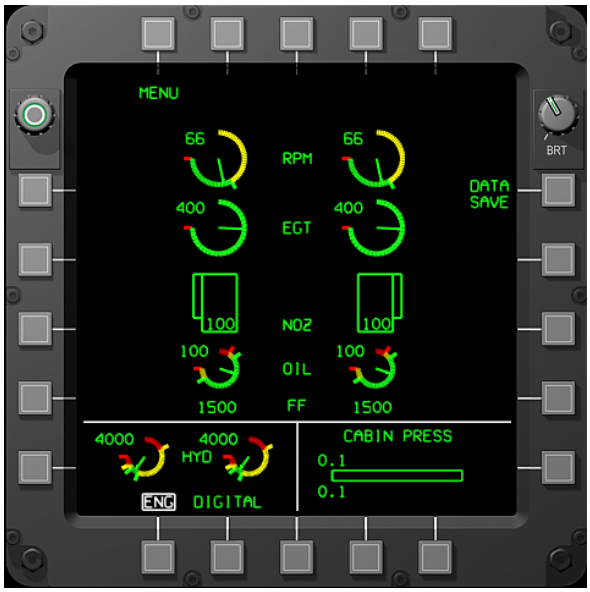

Figure 1. Systems page for the F-22A Raptor representing the contemporary "one-sensor, one-indicator" display paradigm based on thresholds.

For example, the engine oil pressure limits in jet aircraft often have very different tolerances and acceptable limits when the power settings at both idle and afterburner are compared. To the angst of many young pilots, this has meant that separate sets of limits must be memorized and ultimately they are uncertain when intermediate conditions exist. In fact, when presented with potential engine oil problems while at an intermediate setting, most pilots will deliberately modulate the power to one of the two extremes to see if the oil pressure is within the

known parameters. Naturally, this must be done with caution since increasing the power when oil is low may exacerbate the problem leading to premature engine seizure (if the high power setting is selected) or failing to make a landing surface (if the lower power setting is selected).

To complicate things even more, the number of parameters that must be monitored increases considerably in modern vehicles which attempt to reduce the scope of complex vehicle test programs by increasing instrumentation. As a result, flight test pilots and engineering teams rely heavily on the thresholds for each of the thousands of parameters that are monitored on any given mission. However, after a mishap involving a complex system it is often found that anomalous indications were actually present in the real-time data that might have alerted the test team early enough to avert disaster.

In what follows we introduce new methods to improve aerospace vehicle safety by instead monitoring for anomalous patterns of behavior. This departure from simple thresholds represents a significant shift in how system health monitoring in both test and everyday operations would be performed with applications across the entire aviation community. These new methods are also extremely powerful because of their enhanced sensitivities arising from composite parameters which are impossible with thresholds alone.

Finally, we consider how the displays in both the vehicles and the control rooms would need to change to incorporate these new methods in an intuitive way. The introduction of methods sensitive to anomalous behavior and composite parameters is incompatible with the contemporary dials and traces associated with the current displays. Thus we introduce an integrated display that better captures the interrelated nature of modern aerospace vehicles.

EXAMPLE #1: SPACE SHUTTLE COLUMBIA DISASTER

During reentry on 1 February 2003, the space shuttle Columbia (mission STS-107) broke apart due to a compromise of the thermal protection system on the leading edge of the left wing. It was determined that 82 seconds into the flight, a suitcase-sized piece of insulating foam from the external tank (called the Left Bipod Foam Ramp) was liberated at 66,000 feet above Mean Sea Level (MSL) while the vehicle was accelerating through 2.46 Mach. It was later assessed that this 1-meter foam had enough energy to create a 6-10 inch diameter hole in the leading edge of the left wing.[1] Although, this was observed on video it was unknown at the time whether the impact was significant enough to be a cause for concern since "foam shedding" had become a regular occurrence on space shuttle missions. Since no thresholds of the internal wing temperatures were exceeded during the entire eight minutes of the ascent stage, they concluded the spacecraft was safe to continue.

The first time he temperature thresholds were exceeded was during reentry 17 days later, but by then it was too late and an abort was no longer possible. The hot gas of reentry (estimated at over 5,000°F) penetrated the left wing leading edge exposing the aluminum which melts at 1221°F. For over several minutes, dozens of internal temperature sensors began to fail as the heat worked its way through the left wing all the way to the left wheel well (see Figure 2).

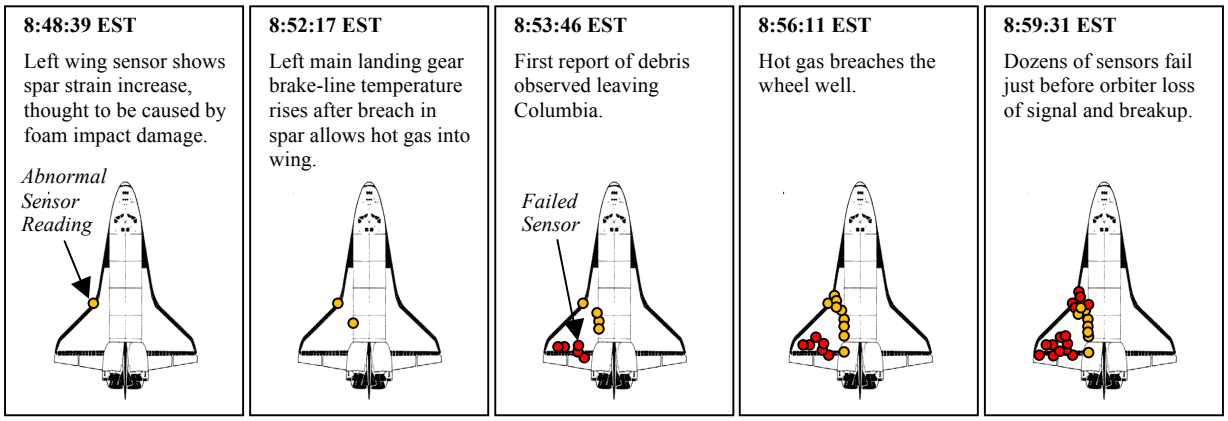

Figure 2. Sensor failures associated with the space shuttle Columbia reentry timeline. Source: Adapted from the Orlando Sentinel (2003, p A7).[2]

Retrospective analysis of the temperature data shows that following the foam impact, the temperatures in the left wing began to behave abnormally when compared to the right wing and when compared to historical data taken on the previous five flights. In fact, within minutes of the foam impact anomalous patterns of behavior were recorded in the left wing temperature data but since these temperatures were "within limits" they were not used to identify if the leading edge had been penetrated. It would seem then that the implementation of a set of algorithms that monitored for anomalous patterns of behavior instead of simple thresholds could provide important early warning not possible with the current system.

INDUCTIVE MONITORING SYSTEM (IMS)

The Columbia accident thus served as one of the first case studies for application of the inductive monitoring system (IMS) to attempt to assess the validity and utility of such algorithms.[3] The IMS takes input either directly from the system under consideration or from high-fidelity simulations.[4] Initially, multi-parameter nominal data samples are processed and

knowledge bases are generated from them. Numerical techniques are used to characterize system behavior by identifying regions of the nominal state space. Specifically, the IMS algorithm employs a hybrid of two clustering techniques to identify these regions: K-means clustering and density-based spatial clustering[5] (see Figure 3 for examples of different clustering techniques[6]). Both of these iterative clustering methods are partitioning methods which focus on different ways to group related data points based on a distance measure defined on the state space (sometimes referred to as the "linkage function").

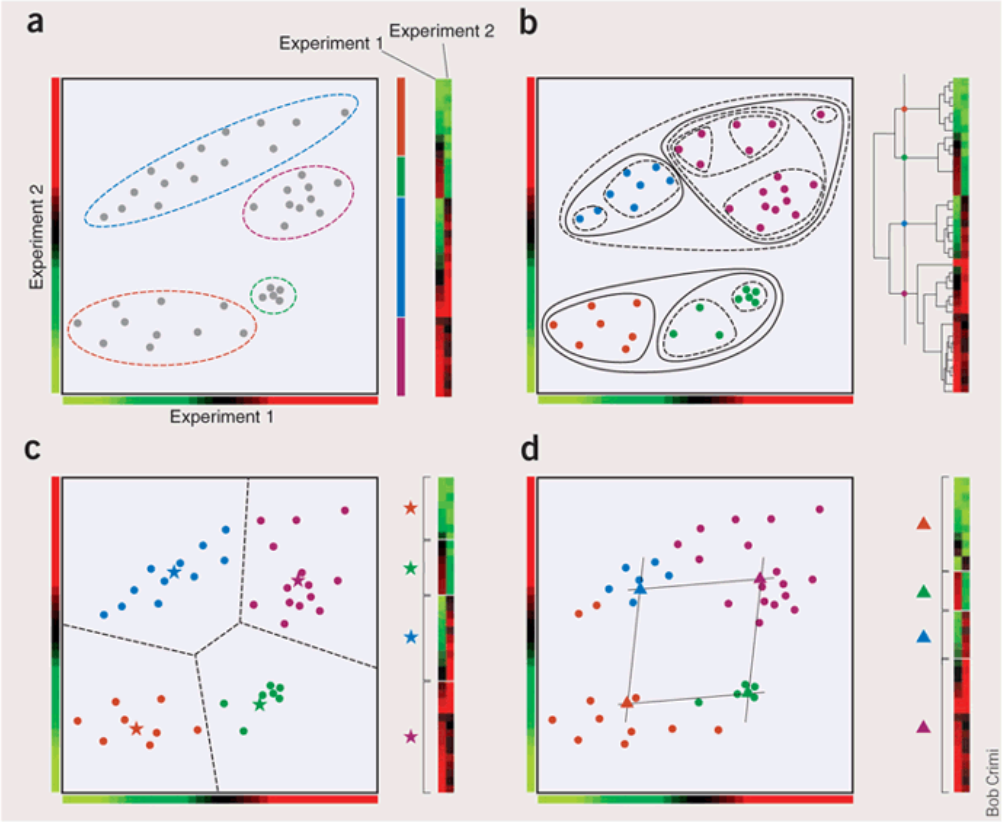

Figure 3. Four different cluster techniques applied to a sample data set. (a) Euclidean distance clustering, (b) Hierarchical clustering where each cluster is subdivided into smaller clusters, (c) K-means (with k = 4) which partitions the space into four subspaces, and (d) Self Organizing Map (SOM) in which the centroids are organized into a grid structure (in this case a simple 2x2). Source: Adapted from D'haeseleer (2005, p. 1499).

The basic data structure of the IMS algorithm is a vector of parameter values (Figure 4). Depending on the sampling rate, a new vector is input into the system at the selected time interval. Theoretically, each such vector defines a point in N-dimensional space. The intention behind the data processing is to create a representation that permits analysis of all parameters of related vectors together.

Pressure A	*Valve 1 Position*	*Pressure B*	*Valve 2 Position*	*Pressure C*	*Temperature 1*	*Temperature 2*
2857.2	86.4%	1218.4	96.2%	1104.1	49.8	37.6

Figure 4: Sample IMS data vector.

The steps of the IMS algorithm are as follows: First, the training data is formatted into a vector and the knowledge base is created. This knowledge base contains clusters that define the ranges of "allowable" value for each vector's parameters for vectors that fall within that cluster. The vector of high values and the vector of low values in a cluster can be thought of as corners defining a minimum bounding rectangle in N-dimensional space (Figure 5). Points that fall inside this rectangle are considered to be within the system's normal operating range. The high and low ranges for each parameter in the cluster are considered as allowable ranges, provided that other parameters are within the ranges specified in that cluster.

Pressure A	*Valve 1 Position*	*Pressure B*	*Valve 2 Position*	*Pressure C*	*Temperature 1*	*Temperature 2*
2860.1	86.9%	1222.4	97.2%	1105.3	51.8	39.6
2855.4	85.7%	1216.3	94.3%	1101.8	48.3	37.4

Figure 5: Sample IMS cluster hyper-box.

Creation of the knowledge base begins with an empty cluster. The first training vector becomes the initial cluster, and each subsequent vector is compared to the existing clusters to find one that is closest to it. The distance between a given vector and a cluster can be measured using a variety of metrics, but the standard Euclidean distance has proven effective for most applications. The point that represents a cluster can be computed in many ways. One option, based on the K-means clustering method, measures the distance from the centroid of the cluster, computed by forming a vector from the averages for each parameter in the cluster.

As in density-based spatial clustering, a threshold value, ε, defines the maximum allowable distance between a cluster and vector to determine if the vector should be incorporated into the cluster. If the vector is within ε, the cluster parameter intervals are expanded to include the new vector. If the distance between the new vector and the closest cluster in the database is greater than ε, a new cluster is formed and the database is updated. This process is repeated until all of the training data has been analyzed and incorporated into clusters. The result is a knowledge base of clusters that characterize system performance in the normal operating envelope.

Once the training knowledge base is established, the next phase concerns the analysis and identification of anomalies. This is accomplished by taking real-time data and comparing the resulting vector to the knowledge base to assess whether a match exists by determining the distance to the nearest cluster. If there is no cluster in the database that would include the new

input vector, the system is behaving in an unexpected manner and a possible anomaly exists.

The objective of the IMS algorithm is to provide the user (pilot, astronaut, mission controller, etc) with an idea of how far the system behavior is deviating from nominal operations. In situations where the input vector is not contained within a known cluster, the distance between the vector and the closest point in the bounding rectangle of the nearest cluster is reported. The implication of the distance between the vector and the bounding rectangle depends on what is defined as an acceptable tolerance range. An input vector close to the acceptable range warrants extra vigilance, while an input vector beyond the tolerance calls for more immediate attention.

In most situations, it is advantageous to normalize the data values before incorporation into the database. For instance, a parameter value can be represented as a percentage of the maximum value or in terms of the standard deviation. Likewise, the output data can be scaled so that the distances between the vectors and clusters represent a percentage of maximum distance in the database. Compared to legacy federated displays such derived variables provide a more meaningful representation of a given parameter value, allow us to compare parameter values of different units (such as pressure and temperature), and provide a more comprehensible measure of how much the system is deviating from nominal operations.

Applying these techniques to the Columbia wing temperature sensors retrospectively provides a good example of how the IMS system might have provided actionable information to the crew and mission control. The knowledge base for the launch and ascent phase were generated from data collected during five previous Columbia flights and then compared to telemetry data from STS-107. The analysis focused on four temperature sensors in each of the wings of the shuttle (see Figure 6). Since ambient temperatures can change between flights, the

data were normalized to the most centrally located sensor (Brake Line Temp B in the figure). This was performed by expressing the other sensor values in relation to the Brake Line Temp B value.

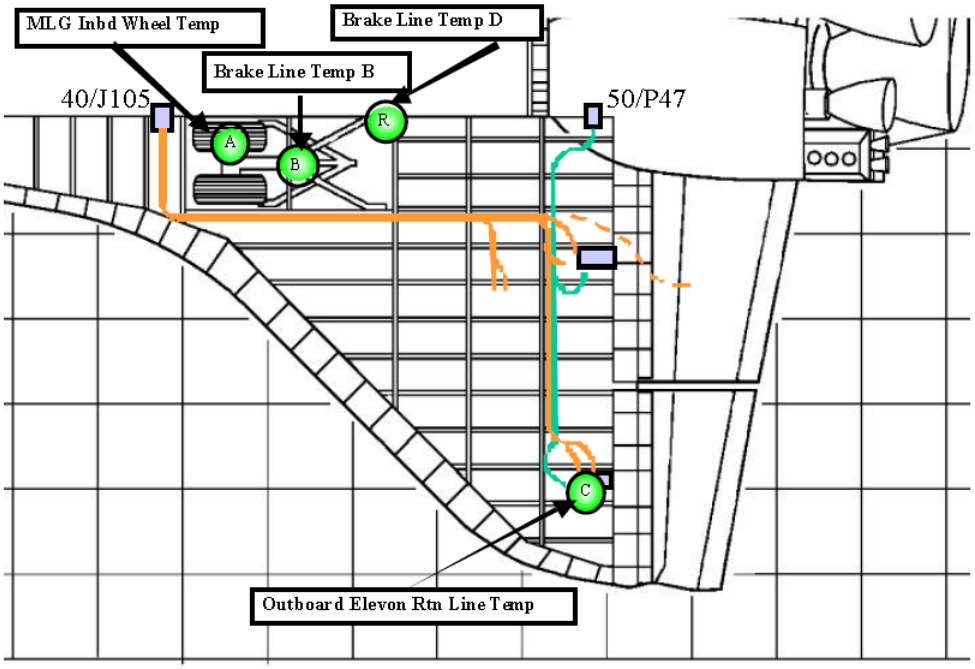

Figure 6. Left wing temperature sensor locations used in the Columbia analysis. Source: Adapted from Iverson (2004, p. 5)

In addition, monitoring was further focused by comparing sets of data in a moving four-second window that typically consisted of 12-16 data points. The data sets used for training and analysis during the ascent phase covered the time period from launch to just before Main Engine Cut Off (MECO) point, which is about eight minutes into the ascent. For example, after the inclusion of the five previous flights the left wing knowledge base contained 502 clusters.

The results of the analysis showed that before the foam impact both the left and right wing temperature sensors were within 5% of the baseline (see Figure 7). However, after the

impact the right wing continues to remain within normal limits whereas the left wing begins a significant departure from the baseline. Note that using this method, the anomalous behavior is noticed immediately with time to consider alternatives and arrive at a backup plan. When compared to the method actually employed at the time, the first indications (corresponding to exceeding a threshold) occurred during the period of maximum heating, at reentry, when it was too late to intervene.

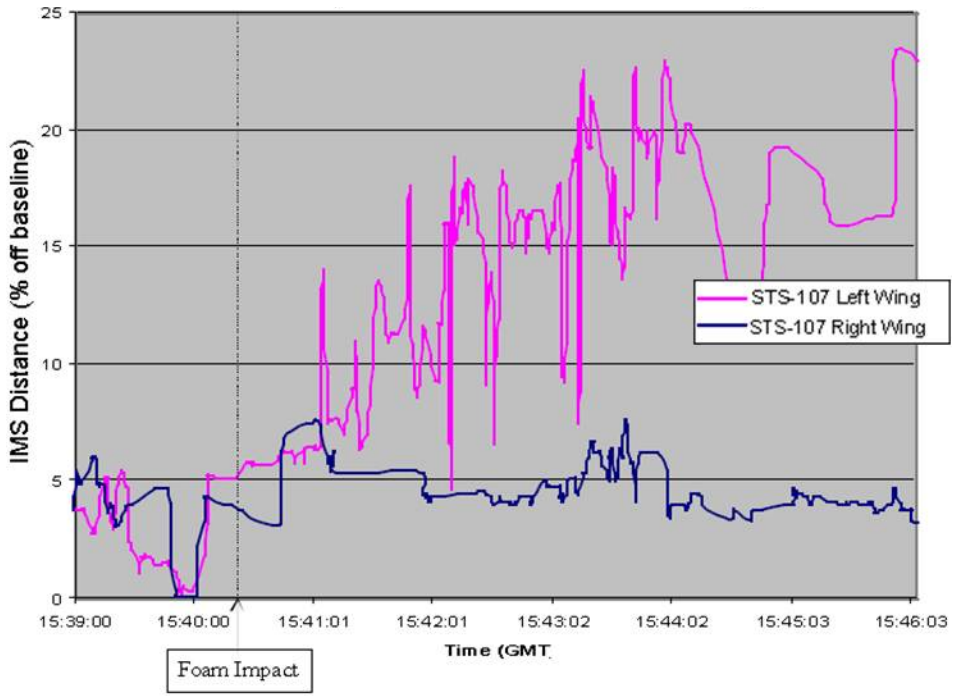

Figure 7. Results from the Columbia IMS temperature analysis. Source: Adapted from Iverson (2004, p. 6)

The retrospective analysis of the Columbia accident data clearly shows the merit of tools that monitor for anomalous patterns of behavior in addition to simple absolute measurements. The results show significant and continuous deviations from expected values for the left wing while the right wing remained consistent with the baseline which represents another indication

that something anomalous is occurring. In effect, algorithms and methods like IMS are able to extract meaningful information that was previously concealed in the data. Additionally, we highlight that although this example is useful in its simplicity and dramatic impact, more complicated composite parameters (those made up from multiple inputs such as might be found in an engine) can be extremely sensitive and even more beneficial.

However, before further developing composite parameters and a display to incorporate them, we first introduce a complementary monitoring method for anomalous patterns of behavior. While the IMS algorithm is extremely useful for systems with an established baseline, it is natural to consider what might be done for systems that are operating for the first time.* During the flight test of new vehicles, a robust baseline is often unavailable or is skeptically approximated from static testing or simulation. As a result, the unfamiliar territory associated with envelope expansion of a new vehicle provides ample motivation for early warning algorithms that do not rely on a well established baseline.

CRITICAL SLOWING DOWN

Phase transitions have been studied for centuries and applications of the general thermodynamic formalism goes beyond that which is taught in introductory science classes. A wide range of systems display behavior governed by highly coupled and non-linear differential equations of state. To name a few, these include the familiar mathematical expressions for fluid dynamics (as in the Navier-Stokes equations), electromagnetics (as in Maxwell's equations and

* We should highlight that the IMS algorithm works most effectively when a well-established baseline is available but could be modified to build the knowledge base in near real-time by dedicating timeline. In the next section we aim to show a slightly different method that exploits the mathematical properties uniquely associated with critical transitions.

the many variations therein), quantum field theory (as in the Yang-Mills, quantum electrodynamics, and quantum chromodynamics equations), and relativity (as in Einstein's field equations). While these state equations can be very successful at dealing with problems that are well defined within the scope of each equation set, many dynamics encountered in a non-laboratory environment experience additional difficulties which the equations do not anticipate. Often such real world systems are very dynamic, with interactions, stochastic influences, and system responses that do not naturally follow from basic principles. Such complex systems have received increased attention since improving computational capabilities have made the field of Complexity Theory more tractable.[7]

In the end, many important complex problems revolve around predicting the state of a system made of many interconnected elements exhibiting a wide variety of complicated feedback relationships. As the internal organization and relationships change, phase transitions can mark particularly important system-level shifts. Identifying such critical thresholds, colloquially called tipping points, could be extremely useful for a variety of fields.[8] The thresholds for such transitions correspond to bifurcations whereby the system stability rapidly undergoes a sudden change in system response or behavior. As an example from medicine, asthma attacks and epileptic seizures provide very little warning of their impending shift in system response and thus represent a bifurcation. A diverse set of such cases was analyzed by Scheffer et al. which led them to propose that "…the dynamics of systems near a critical point have generic properties, regardless of the differences in the details of each system."[9] These generic properties are:

1. A slower recovery from perturbations
2. Increased variance (conditional heteroskedasticity)
3. Increased autocorrelation
4. Fluctuation asymmetries (skewness, kurtosis, and flickering)

All of these are mathematical properties associated with the system residuals (errors) and are thus independent of the actual system specifics. Given this independency, we intend to show that the techniques presented here can also be extremely helpful in mitigating risk in flight test which is inherently performed on systems without a characterized baseline. In fact, flight test is where the baseline is developed so that the IMS system described in the preceding section can be applied both later in the flight test program and once the system is matured into operational status.

Naturally, a wide variety of models can be devised which possess the quality that existing values are related to past ones—either from previous flights with IMS or from more immediate values earlier in the same flight. While the relationship from past to present can theoretically take any form, we use simple linear ones in our development and which can be found in a large number of basic statistics textbooks.[10] In the standard formalism, an autoregressive process is considered first order, AR(1), if the current value is based on the immediately preceding value, second order, AR(2), if based on the previous two values, and so on. In general, as the order increases the estimates on future residuals tends to improve. By plotting the root mean square (RMS) error as a function of order it is sometimes possible to naively determine an optimal order for any given data set.[†] The optimal order and their associated coefficients (typically calculated using a least squares fit) are then reapplied to the model for forecasting. However, to remain consistent with Scheffer et al.'s previous work we will keep the autoregression model simple and utilize a standard AR(1) process.

For a data set represented by $x(t)$ with N elements, the residuals are calculated by simply

[†] More sophisticated methods exist, such as the Akaike Information Criterion, but are beyond the scope of the current discussion.

subtracting the mean, $\bar{x}$, from each individual point to arrive at $y(t_n) = x(t_n) - \bar{x}$. To approximate the recovery speed, λ, we model this behavior such that the system returns to equilibrium over time, Δt, which results in,

$$y(t_{n+1}) = e^{-\lambda \Delta t} y(t_n) + \sigma \varepsilon(t_n),$$

where σ is the standard deviation and ε is a random number from a Gaussian distribution. Since an autoregression model would be represented by

$$y(t_{n+1}) = c + \alpha y(t_n) + \sigma \varepsilon(t_n)$$

where c is a constant and α is the lag-1 (or first order) autoregression coefficient, the model is essentially an autoregression model with $c = 0$. In this nomenclature, for white noise $\alpha = 0$ and for perfectly autocorrelated noise $\alpha = 1$. Assuming λ and Δt independent of $y(t)$, the autoregression coefficient is a direct measure of the system recovery speed and $\alpha = e^{-\lambda \Delta t}$. Thus, as the autoregression coefficient approaches unity, the recovery speed approaches zero, and the variance, $\widehat{\mu_2}$, given by,

$$\widehat{\mu_2} = \frac{\sigma^2}{1 - \alpha^2}$$

increases eventually trending toward infinity (the term conditional heteroskedasticity is often used to describe a changing variance).[11] As a result, by monitoring the lag-1 autoregression coefficient, α, we obtain a direct indication of the system recovery speed, λ. In addition, like α, the variance would be expected to similarly increase in the vicinity of a bifurcation point.

A feature related to the aforementioned autoregression is autocorrelation, which measures the correlation between observations at different times. In an analogous process to the lag-1 autoregression nomenclature, a first order autocorrelation coefficient, β, is the simple correlation coefficient between the subsets of the data including *N-1* observations but shifted by a single observation and is given by,

$$\beta = \frac{\sum_{n=1}^{N-1}[y(t_n)][y(t_{n+1})]}{\sum_{n=1}^{N}[y(t_n)]^2}$$

whereby the autocorrelation function is normalized by mean and variance.

In effect, monitoring the autocorrelation coefficient over time provides a measure of whether the state of the system is becoming more like its past. Like the measures before, it is a reflection of the speed of the system's recovery to stochastic perturbations and was found by Scheffer et al. to be another universal indicator of an impending bifurcation and associated critical transition.

Finally, fluctuation asymmetries such as skewness, kurtosis and flickering also provide amplifying characterization of the variability associated with system residuals that is less related with slowing down and relies instead on the concept that an alternate system state (and hence a

potential critical transition) is nearby (see Figure 6). While skewness and kurtosis represent the third and fourth standardized moments[‡] (and hence indicate that the residuals are no longer Gaussian), flickering relies on the construct that the residuals are actually jumping momentarily to the alternate system state, and indicates that the system could transition at any point. Skewness is essentially a measure of a lack of symmetry and the Fisher-Pearson coefficient, $\widehat{\mu_3}$, is given by

$$\widehat{\mu_3} = f\frac{\sum_{i=1}^{N}(x_i - \bar{x})^3/N}{\sigma^3}$$

assuming large sample sizes (such that the size adjustment coefficient, $f = 1$). For smaller sample sizes, it is customary to introduce an adjusted Fisher-Pearson coefficient where

$$f = \frac{\sqrt{N(N-1)}}{N-1}$$

which approaches 1 as N gets large. For residuals that are Gaussian (from a normal distribution) skewness should be zero. Hence, increasing skewness to the residuals implies an underlying potential change in system dynamics. If we utilize the familiar ball in a bowl example, the introduction of a perturbation results in symmetric residuals if the bowl is symmetric (see Figure 8). However, as conditions change in the vicinity of a bifurcation point the alternate system state

[‡] Variance is the second standardized moment (or more precisely a standardized moment of second degree) and was defined earlier.

causes a skewness to develop in the residuals which grows as the conditions approach those associated with the critical transition.

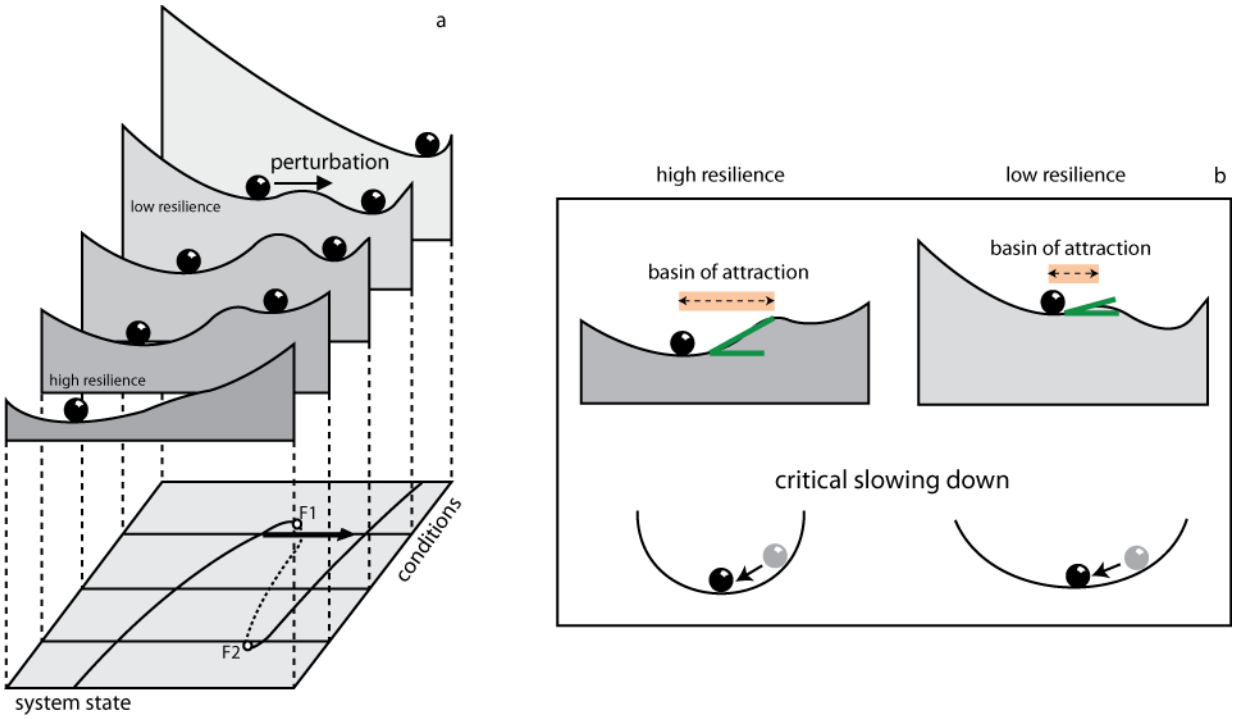

Figure 8. Critical slowing down as a measure of resilience and its relationship to early warning for critical transitions. Source: Adapted from the Early Warning Toolbox.[12]

Kurtosis is another measure of how much the residuals are departing from a Gaussian and is a measure of how heavy-tailed they are, which basically identifies how many extreme data points are expected.[13] Remaining consistent with the previous skewness definition, kurtosis, which represents the fourth standardized moment is given by,

$$\widehat{\mu_4} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4/N}{\sigma^4}$$

which is again the value assuming large *N*. As a result, by monitoring for kurtosis and showing how much it deviates from that associated with a Gaussian (for the normal distribution, $\widehat{\mu_4} = 3$ such that values greater than three are often considered "heavy-tailed" and less than three are "light-tailed") provides evidence of a potential critical transition.

Flickering represents the final statistical early warning indicator and is a term often used to reference bimodal (or even multimodal) behavior in systems with competing system state attractors or alternate regimes. Flickering will manifest itself as large skewness and kurtosis as the system begins the phase transition since stochastic forcing is high enough to begin to realize the new state. Depending on system dynamics, Flickering may not even be recognizable since the resulting transition may no longer be reversible and relying exclusively on this feature would not be recommended.

These critical slowing down indicators (autoregression/autocorrelation, heteroskedasticity, skewness, kurtosis, and flickering) appear to be generic for a variety of reasons, but at a fundamental level, critical slowing down is associated with power-law scaling of the relaxation time (recovery rate). A fast recovery rate implies greater stability and resilience since it is associated with a deeper potential well and provides evidence that the residuals and the system will likely continue to remain well behaved. In contrast, if this recovery rate begins to slow the system may be approaching a transition (see Figure 6) and thus may warn of impending rapid, drastic change. This has been measured for 2D Ising ferromagnetic systems in addition to medical conditions associated with bifurcations (epilepsy, asthma, etc.) such that it may be generalized and, as a result, should be applicable to any system where competing system state

attractors are suspected.

In short, critical slowing down has the potential to be a direct indicator of system resiliency and might be useful in a variety of flight test situations. Consider a test program that has accepted some level of increased risk by forgoing the installation of a Flutter Excitation System (FES). Although considerable modeling and ground vibration testing may be performed, until airborne, the model's prediction accuracy (and any inherent risk) remains unknown (as is frequently proven by flight test, all models are wrong in some ways…some are useful…some may be dangerous). In this particular situation, there is a need to monitor for potential errors by watching accelerometers, but concern would still persist that sufficient warning would be available to avert disaster. Monitoring for the conditions associated with critical slowing down in real-time might provide more advance warning than current techniques, since flutter represents a rather sudden shift in system dynamics. Similar arguments might be made for an aircraft with poor flying or handling qualities where divergent flight control regimes may be suspected. When compared to currently employed methods, critical slowing down compliments IMS in monitoring system dynamics and provides earlier warning to catastrophic events.

In the end, the methods we have introduced to monitor for abnormal patterns of system behavior are enhanced when applied to composite parameters (pairs, triplets, quadrupoles, quintuples, etc). When there are strong interrelationships between primary parameters in a composite, the resulting composite parameter is more sensitive to deviation from expected values than the primary parameters alone and can indicate an anomaly before it manifests in the individual primary parameters. As can be seen in the next section, aircraft engines are a good example where strong interrelationships between individual parameters exist but where the current displays remain federated.

EXAMPLE #2: AIR TRANSAT FLIGHT 236

On 23 August 2001, Air Transat Flight 236 left Toronto, Canada bound for an overnight flight to Lisbon, Portugal. More than four hours into the flight during a routine ops check the crew observed unusual, albeit within limits, oil indications associated with the right engine. In particular, when compared to the left engine parameters, the oil temperature was relatively low, oil pressure was high, and the oil quantity was low. It is worth noting that the asymmetry between the two engines are what really alerted the crew since all the parameters were within pre-established thresholds. Clearly, something was occurring but the cause was not clear or obvious. After consulting with the manuals available on board, the crew performed a "Conference Hotel" by contacting Air Transat's maintenance control center at Montreal-Mirabel International Airport in hopes of gaining additional advice from the technicians and engineers. Unfortunately, the maintenance control center query did not provide any actionable information or amplification beyond what was already known and the captain cited "computer error" as the cause of the unexplained combination of unusual values that were within limits.

Thirty minutes later the onboard Electronic Centralized Aircraft Monitoring system alerted the crew to the presence of a developing fuel imbalance. While fuel imbalances are not routine they are also not uncommon since small differences between engine fuel consumptions over long duration flights can add up. Fortunately, all aircraft are equipped with crossfeed procedures to redistribute fuel before the asymmetry causes controllability concerns. As a result, a fuel imbalance is often not considered overly threatening unless a fuel leak is suspected to be the cause. Since the crew had no reason to suspect that a fuel leak was the cause of the

imbalance, they commenced crossfeed operations.

This action turned out to be a pivotal point in the mishap because the abnormal oil indications and the fuel imbalance were actually related. Just prior to their ops check that discovered the abnormal engine oil indications (at 3 hours and 46 minutes into the flight) the high-pressure fuel line ruptured as a result of hard contact with an adjacent hydraulic line and the right engine developed a significant fuel leak. It was later determined that the hard contact was due to a mismatch between the fuel line and the hydraulic line, each belonging to a different version of the Rolls Royce RB211 Trent engine (see Figure 9).[14]

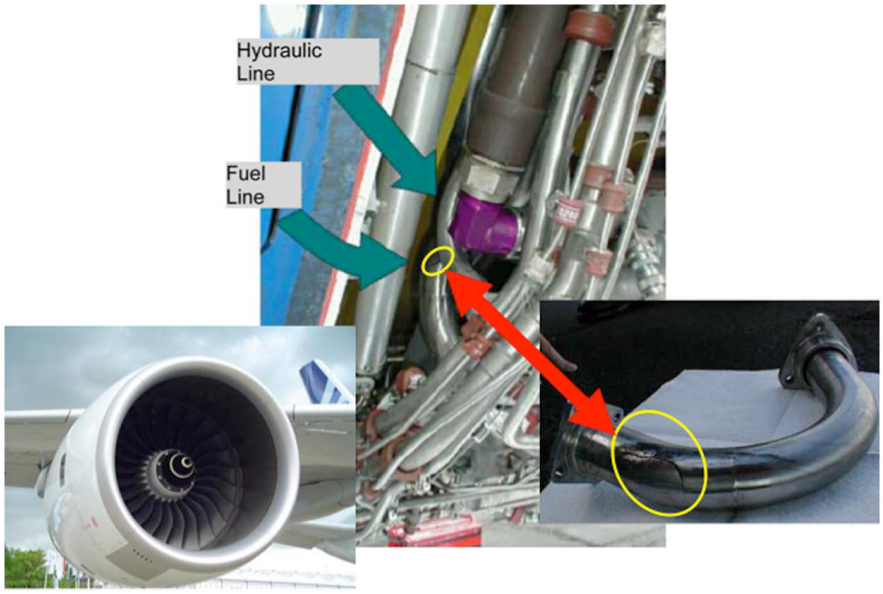

Figure 9. The Rolls-Royce RB211 Trent engine and the juxtaposition of the fuel and hydraulic lines inside the engine. The red arrow points to the actual crack (surrounded with a yellow circle). Source: Adapted from Degani (2013, p. 6).[15]

The leak affected the fuel/oil heat exchanger which is a unit that takes cold fuel from the wing tanks and runs it around lines of hot oil which improves engine efficiency by cooling the

oil and warming the fuel. However, because of the crack in the fuel line no back pressure existed in the line and the fuel was overcooling the engine oil. Once cooled, the oil became more viscous and led to both the high pressure reading and the low quantity because the more viscous oil was slow to return to the reservoir where the quantity sensor was located.

Although the initial crossfeed corrected the fuel imbalance, it also dumped precious fuel overboard during the 13 minutes of operation. Coupled with the earlier leak this action drove their fuel quantity below that required to arrive at Lisbon, the planned destination. As a result, they initiated a divert into Lajes Air Base in the Azores Islands, 1000 miles west of Lisbon. The aircraft continued to lose fuel and both engines flamed out while the aircraft was at 34,000 feet MSL and still 65 nautical miles from Lajes. Fortunately, the glide ratio of the aircraft and skill of the pilots permitted a safe landing and all 306 passengers and crew were saved.

Better system health monitoring could have greatly reduced the risk in this particular scenario. Employing a system that monitors for abnormal patterns of interaction instead of simple threshold exceedance would have informed the flight crew and ground personnel that a serious issue was occurring. While all parties involved were concerned, they were not able to ascertain that the fuel imbalance and oil indications were actually related, and thus exacerbated their condition by performing the crossfeed operation.

Fortunately, the algorithms outlined in the preceding sections have the potential to provide early warning that is also sensitive to interrelated systems through the use of composite parameters. The challenge is then to devise an intuitive way to represent this new information and wean users off of the “one-sensor, one-indicator” paradigm.

COMPOSITE PARAMETER DISPLAY

As aerospace vehicles become more complicated, the number of interrelated systems increases such that automation is required to augment the user for system health monitoring. However, the antiquated display framework that relies on individual indicators places additional reliance on simple thresholds to alert the user to a possible problem. As we have shown in the examples above, these simple thresholds are often insufficient to provide warning early enough. To further complicate the issue, the level of instrumentation on test flights is on the rise in an effort to help reduce the time, scope, and cost of modern flight test programs which increases the number of parameters that must be monitored.

In addition, automation based on thresholds alone can introduce significant and precious delays when early human intervention is required. For example, although the Flight 236 fuel leak started at 04:38 UTC, the fuel imbalance was not highlighted to the crew until 05:33 because the automated system transferred 3.5 tons of fuel from a tank in the tail of the aircraft without alerting the crew. Unfortunately, this automatically transferred fuel was subsequently dumped overboard.

The mathematical techniques introduced in the preceding sections represent an opportunity to incorporate composite parameters in addition to providing a warning earlier than conventional methods. It is important to thoughtfully construct composite parameters with custom weighting factors based on the known interrelationships that arise naturally from the subsystem architecture. In fact, many aerospace vehicles exhibit traits where certain known failures or unique conditions might indicate particularly good composite parameter generation. For example, since the single engine F-16 utilizes engine oil as the hydraulic fluid that controls

nozzle position, anomalous oil temperatures/pressures can cause incorrect nozzle positioning and subsequent compressor stalls. As a result, a combination of oil pressure, oil temperature, and nozzle position would be a good composite parameter more sensitive to anomalous operation and might provide the pilot with sufficient warning to establish a flight path within gliding distance of a suitable airfield.

The health monitoring displays currently employed are not well suited to incorporate composite parameters and, as mentioned previously, are limited to the simple threshold exceedance paradigm. As an example of how such a display might look we include a display concept developed to convey IMS monitoring output from an application to NASA, and studied using the Rotorcraft Aircrew Systems Concepts Airborne Laboratory (RASCAL), which is a modified UH-60 Blackhawk helicopter. During the study, four separate primary parameters were monitored and recorded for the helicopter with an emphasis on the two General Electric T700-701 engines: Power Turbine Speed (Np, in % of rated value), Gas Generator Speed (Ng, in % of rated value), Engine Torque (Tq, in % of rated value), and Fuel Flow rate (FF, in pounds of fuel per minute). Additionally, the combined output of both engines was represented by the helicopter's rotor speed (again in % of rated value).[16]

System dynamics and IMS analysis suggested meaningful composite parameters would include products of rotor speed with Np, Np with Ng, Ng with Tq, and Tq with FF in addition to higher order composite parameters. A display was then created based on medieval girih tilework[17] that includes raw engine parameters in the center with each of the primary (non-composite) parameters as individual petals with their order based on the most meaningful interrelationships (see Figure 10). The left engine parameters were placed on the left side of the

central “flower” and the right engine parameters were placed on the right side with rate of change indicators embedded as anthers (the teardrop looking structures in each petal). Subsequent layers of polygon petals (or garland) embed information on composites of the primary parameters whose subsequent color conveys information about how the system compares to the previously established baseline. The more anomalous the patterns of interaction, the more the color changes to indicate that further inspection is required. Similarly, the color of the line surrounding multiple polygons also changes to represent anomalous patterns for a set of parameters.

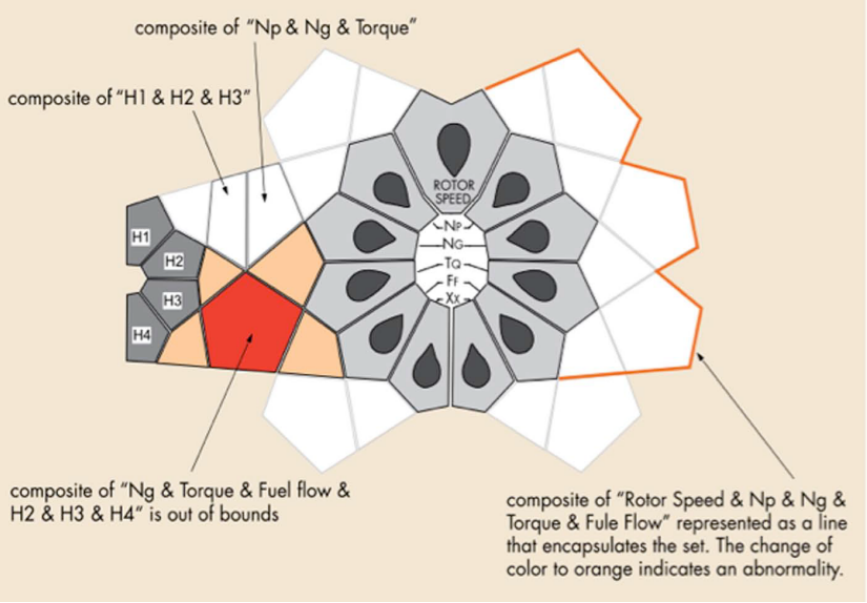

Figure 10. NASA’s Rotorcraft Aircrew Systems Concepts Airborne Laboratory Information Tapestry display which includes the primary engine parameters of Rotor Speed, Power Turbine Speed (Np), Gas Generator Speed (Ng), Engine Torque (Tq), Fuel Flow (FF) and their respective composites. Source: Adapted from Degani (2009, p. 64)

Finally, the display can be modified to accommodate other subsystems and represents their relationship with each other by connecting adjacent flowers. For example, an emergency or

mishap affecting multiple subsystems will cause the polygons between the systems to change color in addition to those just associated with each individual system. Implementing a display that incorporates these elements will enhance safety by assisting operators in monitoring the system's health. This is particularly true for the types of system malfunctions that operators have the least success in identifying, because they involve subtle interactions between subsystems.

CONCLUSIONS

Contemporary system health monitoring utilizes displays that represent inter-related vehicle systems with federated displays based on relatively simple threshold limits not much different from those found in a common automobile. This paper describes ways to modernize system health monitoring and enhance flight test safety in the process. The rapid expansion of computing capabilities, coupled with recent algorithm developments, has enabled real-time monitoring not possible before by actively searching for anomalous patterns of behavior. These techniques are applicable to a wide array of aerospace vehicles ranging from operational platforms, where well established baselines are determined through flight and laboratory testing, to brand new vehicles where increased risk arises from the lack of reference information to define nominal operation limits.

To realize the full benefit of these new early warning techniques, the current "one-sensor, one-indicator" display paradigm must be updated. One of the unique characteristics of algorithms that search for anomalous patterns of interactions is the use of composite parameters which are not reflected in existing displays. When properly chosen, such composite parameters are sensitive to smaller deviations and can thus recognize and alert operators to problems

affecting multiple systems much earlier than the conventional methods. By modernizing system health monitoring displays to include something like the tilework structure introduced earlier (see Figure 11), the operator is quickly informed of the true complexity of the most serious operational problems.

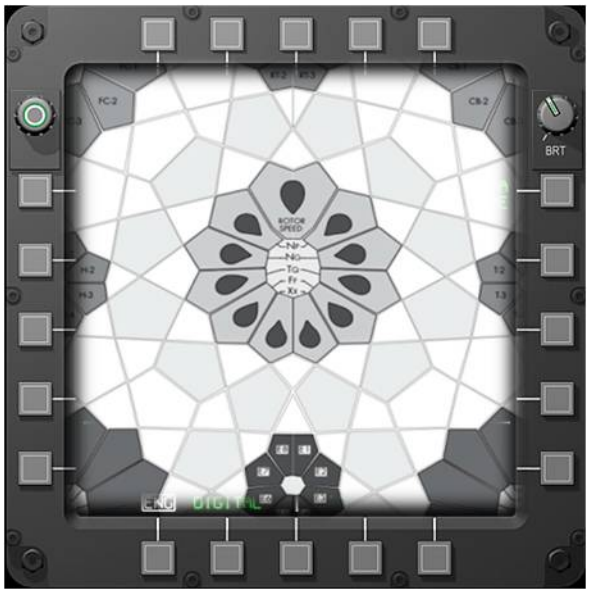

Figure 11. A notional modern system health monitoring display reflecting the interaction of nine interrelated vehicle systems is notably different when compared to the contemporary displays shown in Figure 1. Source: Adapted from Degani (2009, p. 66)

ACKNOWLEDGEMENTS

The views expressed are those of the author and do not reflect the official policy or position of the U.S. Air Force, the Department of Defense, NASA, or the U.S. Government. The material in this paper is UNCLASSIFIED and approved for public release: distribution is unlimited, reference Air Force Test Center Public Affairs reference number 412 TW-PA-#####.

[1] Gehman, H. et al. (2003) *Columbia Accident Investigation Board Report, Vol 1*, NASA and U.S. Government Printing Office, 64-67.

[2] "Columbia Disaster: The Final Report of the Columbia Accident Investigation Board," *Orlando Sentinel* 27 Aug. 2003: A7. Print.

[3] Iverson, D. L. (2004) Inductive System Health Monitoring. *Proceedings of the 2004 International Conference on Artificial Intelligence (IC-AI'04)*.

[4] Degani, A., Jorgensen, C., Iverson, D., Shafto, M., & Olson, L. (2009), On Organization of Information: Approach and Early Work. *NASA/TM-2009-215368.*

[5] Jain, A.K. & Dubes, R.C. (1988). *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall.

[6] D'haeseleer, P. (2005). How Does Gene Expression Clustering Work? *Nature Biotechnology* 23, 1499-1501.

[7] Javorsek, D. (2015). Modernizing the Safety Process: Complexity Theory for Flight Test Professionals. *Society of Experimental Test Pilots 59th Annual Banquet and Symposium*.

[8] Javorsek, D. (2016) Cultivating the Assertive Skeptic: A Proposal for Modernizing Flight Test Safety to Address Human Agency. *Cockpit Magazine,* Jan-Jun, 75-95.

[9] Scheffer, M., et al. (2009). Early-warning Signals for Critical Transitions. *Nature* 461, 53-59.

[10] Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Volume II Detection Theory*. Upper Saddle River, NJ: Prentice-Hall PTR.

[11] Kaufman, R.L. (2013) *Heteroskedasticity in Regression*. New York, NY: Sage Publications, Inc.

[12] Early Warning Signals Toolbox. (2016, Aug 24) Retrieved from http://www.early-warning-signals.org

[13] Grinstead, C., and W. Peterson, L. Snell. (2011). *Probability Tales*. New York, NY: American Mathematical Society.

[14] Government of Portugal. (2004). All engines-out landing due to fuel exhaustion, Air Transat, Airbus A330-243, Lajes, Azores, Portugal, 24 August 2001. *Accident Investigation Final Report of the Ministry of Public Works, Transportation, and Communications*, Gatineau: Transportation Safety Board of Canada.

[15] Degani, A, Barshi, I., & Shafto, M.G. (2013). Information Organization in the Airline Cockpit: Lessons From Flight 236," *Journal of Cognitive Engineering and Decision Making* 20, 1-23.

[16] Barshi, I., Degani, A., Iverson, D., & Lu, P.J. (2012). Using Medieval Architecture as Inspiration for Display Design: Parameter Interrelationships and Organizational Structure. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, 1799-1803.

[17] Lu, P.J. and Steinhardt, P.J. (2007). Decagonal and Quasi-Crystalline Tilings in Medieval Islamic Architecture. *Science* 315, 1106.