FAULT DIAGNOSIS IN GAS TURBINES USING A MODEL-BASED TECHNIQUE

Graeme L. Merrington
DSTO
Aeronautical Research Laboratory
Victoria
Australia

ABSTRACT
Reliable methods for diagnosing faults and detecting degraded performance in gas turbine engines are continually being sought. In this paper, a model-based technique is applied to the problem of detecting degraded performance in a military turbofan engine from take-off acceleration type transients. In the past, difficulty has been experienced in isolating effects of some of the physical processes involved. One such effect is the influence of the bulk metal temperature on the measured engine parameters during large power excursions. It will be shown that the model-based technique provides a simple and convenient way of separating this effect from the faster dynamic components. The important conclusion from this work is that good fault coverage can be gleaned from the resultant pseudo steady-state gain estimates derived in this way.

NOMENCLATURE

\( T \)  
Total temperature

\( u \)  
Measured inputs

\( WF \)  
Fuel flow

\( WF_{ss} \)  
Steady-state fuel flow

\( y \)  
Measured outputs

\( \Delta \)  
Difference operator

\( \varepsilon \)  
Error vector

\( \varphi \)  
Measurement vector

\( \kappa \)  
Acceleration starting point

\( \theta \)  
Parameter vector

\( \sigma \)  
Standard deviation

SUBSCRIPT

\( t \)  
Sample time

\( n \)  
Number of samples in a record

\( p,q \)  
Number of parameters in Eq. 1

\( ss \)  
Steady-state

SUPERSCRIPT

\( T \)  
Vector transpose

Presented at the International Gas Turbine and Aeroengine Congress and Exposition
Cincinnati, Ohio May 24–27, 1993
This paper has been accepted for publication in the Transactions of the ASME
Discussion of it will be accepted at ASME Headquarters until September 30, 1993
INTRODUCTION

The significant cost-benefits associated with maintaining gas turbine engines on-condition are now widely recognised. On-condition maintenance is best performed on modular engines and therefore it is desirable to be able to diagnose performance related problems to module level if possible. It follows then that reliable engine health monitoring and condition assessment techniques are important components of an effective engine maintenance program. Early monitoring techniques mainly involved trending manually recorded data corrected for variations in ambient conditions. Thus important effects, such as measurement noise and sensor biases, were often wrongly interpreted as degraded gas path components. The realisation that the problem was basically stochastic in nature rather than deterministic quickly lead to the adoption of more advanced signal processing techniques.

Until fairly recently, most of the effort in the engine condition assessment area has been devoted to developing improved techniques for extracting the required information from steady-state data. For example, commercial airline operators have largely based their engine health assessment on cruise data. The major engine companies have proprietary computer based techniques available for operators to use as part of their engine management programs. For example, General Electric have TEMPER (Doel, 1992) for use with data recorded on the wing and another version for evaluating engines under test-cell conditions. Similarly, Rolls Royce have COMPASS (Provost, 1988) which is capable of providing engine condition information from on-wing engine data (cruise, take-off, ground-run and start) as well as from test-cell data. These techniques address the stochastic nature of the real problem and use optimal estimation theory to take account of measurement noise. Thus their performance is far superior to many of the earlier methods which tried to treat the problem as a deterministic one.

In many situations, good quality steady-state data are simply not available. However, some current generation aircraft are equipped with Engine Monitoring Systems (EMS) which also have a capability to record transient data. For example, the EMS in the F/A-18 records a range of engine parameters during take-off, in-flight when a parameter exceedance is detected and at other times when manually selected by the crew. The potential to use the transient take-off record for engine condition purposes has been recognised by several groups. The techniques that have emerged as likely candidates include correlative methods (Cue and Muir, 1990; Eustace et. al., 92; Henry, 1988; Muir et. al., 1988) and model based methods (Merrington, 1988; 1989; Merrington et. al., 1991; Villaneuva et. al., 1991). Moreover, it has been shown that the fault coverage capability of these transient techniques is similar to the steady-state methods for a range of operational type faults (Eustace et. al., 1992). For some problems, the transient approach gives additional insight that may not be extracted easily from steady-state tests. For example, some fuel control related problems are clearly visible from transient test data. Many engines utilise a WF/Ps3 versus corrected compressor speed fuel scheduling loop in order to provide some degree of surge protection during transients. Sensor type problems, such as a leaky Ps3 line, can lead to reduced over-fuelling resulting in slow accelerations. Thus by monitoring the WF/Ps3 ratio parameter, this type of fault can be readily isolated from other gas path problems.

Whilst the initial impetus for much of this work has stemmed from the military arena, the potential to use transient engine data analysis in other areas is beginning to emerge. There is scope for civil operators to use the technology to estimate take-off power under varying ambient conditions by monitoring previous take-off records. Currently, in most cases, this sort of information is extracted from steady-state cruise data. In addition, there is increasing interest in trying to apply the technology to land-based applications (Meher-Homji and Bhargava, 1992).

In this paper, model based fault diagnosis in a military turbofan is examined using transient data. A simple technique is outlined for separating the slower 'bulk-temperature' component of an acceleration-type transient from the faster 'spooling-up' component. The method is applied to fault implant test data obtained from an F404 engine under Sea-Level-Static (SLS) conditions.

ACCELERATION TRANSIENTS

When a gas turbine engine accelerates from one steady-state operating point to another, the process involves two
important physical processes. The first is associated with spooling up the engine to the required speed demanded by the power lever movement. The engine controller will limit the degree of over-fuelling based on prior knowledge of the available surge margin from a specification engine. As the actual spool speed approaches the desired value, the controller enters a governing phase to ensure that the final speed is achieved without excessive overshoot. This process in a current generation military gas turbine takes place in a relatively short time. Typically an acceleration from idle power to maximum dry power takes less than five seconds with a characteristic time constant of the order of 0.5s. This is accompanied by a much slower thermal process involving heat addition to the metal components surrounding the gas stream. The time taken to achieve thermal equilibrium conditions can be of the order of five minutes for large power excursions. This time is not only strongly dependent on the magnitude of the power lever movement but also on the immediate prior operating history of the engine. This second physical process is often termed the bulk metal temperature effect.

Take-off transient data records from operating aircraft, such as the F/A-18, can vary in duration for a number of reasons. In the F/A-18, the take-off record is terminated when the power lever is moved out of the Intermediate Rated Power (IRP) position or when the weight comes off the wheels. Thus aircraft weight, temperature of the day and differences in individual crew operating techniques can give rise to considerable scatter in the length of these records. For example, results from RAAF operational aircraft indicate that the length of the take-off record can vary between zero and fifteen seconds, as shown in Figure 1 (taken from Frith, 1992). These times are certainly well short of the time required to achieve true steady-state conditions. It is therefore important to define a consistent pseudo steady-state end-point criterion for an acceleration transient if a model-based fault detection technique of the type described in the next section is to be utilised.

In general, as take-off accelerations can be initiated from any part-power position (in the case of the F/A-18, an upper limit of approximately 80% power is invoked because of power to weight ratio considerations), an end-point criterion based on a fixed record length from some designated starting point or one based on a fixed time interval after the acceleration peak, is usually not suitable. This is because neither criterion addresses the real problem of how to account for the variability in the bulk metal temperature effect when different size transients are involved. One approach would be to mathematically model the bulk metal temperature effect and incorporate this into the data analysis. An alternative approach is to develop a procedure which is capable of isolating the bulk metal temperature effect from the 'fast transient' component. This latter approach is the one that has been adopted here.

**FIGURE 1. DISTRIBUTION OF TAKE-OFF RECORD LENGTHS FROM OPERATIONAL F/A-18 AIRCRAFT (TAKEN FROM FRITH, 1992)**

**ANALYSIS PROCEDURE**

A model based technique based on the method outlined by Merrington, 1989 was used. Briefly this technique is given in the form:

$$y_t = \theta^T \varphi_t + \epsilon_t$$  \(1\)

where

$$\varphi_t^T = (-y_{t-1},-\ldots,-y_{t-p},u_{t-1},\ldots,u_{t-q})$$

$$\theta^T = (a_1,\ldots,a_p,b_1,\ldots,b_q)$$

The least squares estimate (LSE) is the one that minimises the cost function,

$$J(\theta) = \sum_{t=1}^{n} (y_t - \varphi_t^T \theta)^2$$  \(2\)

where n is the number of sample points and the estimate of the model parameters is given by:
Equation (1) can be re-arranged to enable the steady-state gain, \( K \), corresponding to the slope of the steady-state curve, to be extracted. A steady-state condition exists when \( y_t = Y_{t-1} \) and \( u_{t-1} = u_{t-1} \) giving

\[
K = \frac{\Delta y_{ss}}{\Delta u_{ss}} = \frac{\sum_{i=1}^{q} \hat{a}_i}{1 + \sum_{i=1}^{p} \hat{a}_i}
\]  

(4)

The primary input for a gas turbine is fuel and the effect of the significant non-linearity in the steady-state relations over the operating speed range is reduced by reformulating the problem (Merrington, 1989) in terms of an overfuelling parameter defined by

\[
\Delta WF = WF - WF_{ss}
\]

(5)

where \( WF_{ss} \) is the steady-state fuel at the given speed.

Thus if \( WF \) is the only input (\( q = 1 \)), then \( u_t \) becomes

\[
u_t = WF_t - WF_{ss}
\]

(6)

The start of an acceleration was determined by monitoring the rate of change of the relatively fast response compressor outlet static pressure signal (\( \Delta y_3 \)). The actual sample point \( k \) corresponding to this starting point was obtained from

\[
Ps_{3k+1} - Ps_{3k-1} > 4\sigma
\]

(7)

where \( \sigma \) corresponds to the standard deviation of the measurement noise on the \( \Delta y_3 \) signal. A forty point sample of the steady-state segment of the record immediately preceding the acceleration transient was included to identify the initial steady-state condition. The LSE analysis was then performed to determine how the minimum error variance, \( J(\theta)_{\text{min}} \), varies with record length. Commencing with a record of 80 samples (to ensure a reasonable length of the ‘fast-response’ part of the transient was included) the LSE was then applied to progressively longer records by adding five additional sample points at a time until all the data were included.

Since most of the fault information is contained in the steady-state gain term rather than in the dynamics (Villaneuva, 1991), it is important to define a method for isolating that component of the gain attributable to the acceleration from that associated with the bulk metal temperature effect. The bulk metal temperature effect becomes more important as the scale of the transient increases in the absence of re-slam (when a snap acceleration immediately follows a rapid deceleration) effects due to the larger differential between gas and metal temperatures. As re-slam effects are usually absent during pre-takeoff procedures, the distribution of \( J(\theta)_{\text{min}} \) with record length can be used to separate these two important effects.

![Figure 2. F404 Engine and Station Locations](http://proceedings.asmedigitalcollection.asme.org/proceedings.asmedigitalcollection.asme.org/10/17/2018/Terms_of_Use/http://www.asme.org/about-asme/terms-of-use)
The criterion found most suitable for separating the bulk temperature effect was to locate the minimum of the cost function minima (Eq. 2) for record lengths in the range \( (n_0 < n < N) \) where \( n_0 \) is the initial record length (80 sample points) and \( N \) is the total record length. Intuition would suggest that for large transients, such as a flight-idle to IRP acceleration, the desired pseudo steady-state end point should be in close proximity to the peak of the transient and move progressively downstream from the peak as the size of the transient diminishes.

For convenience a first order model \((p=q=1)\) was adopted for this investigation. These models are known to give reasonable accuracy in the vicinity of an operating point, that is, for small transients. However, they can also be applied to larger transients to extract useful comparative information. Whilst the errors in the absolute values will tend to increase as the magnitude of the transient increases (when the non-linear effects become significant), test results indicate that the relative errors between the fault and no-fault cases are still small. Thus the absolute accuracy is less important in the present application than the relative consistency between the fault and no-fault estimates.

**TEST PROGRAM**

The test program has been described in detail in Eustace et. al., 1992, and therefore it will suffice, for present purposes, to summarise the important points. The aim of the test program was to implant a range of operational type faults in a military turbofan engine under test-cell conditions and acquire data from a number of simulated take-off transients.

**Engine**

The engine used in this investigation was a GE-F404 taken from a fleet of engines operated by the RAAF in the F/A-18. The F404 is a low by-pass, twin spool afterburning turbofan. A schematic of the engine showing the major station locations is given in Figure 2. The engine is designed to be maintained on-condition and has six main modules, comprising the fan, compressor, combustor, high-pressure turbine, low-pressure turbine and afterburner section. Accelerations to maximum dry of Intermediate Rated Power (IRP) are scheduled as a function of corrected compressor speed and then limited near IRP via fan-speed (NL) and low pressure turbine outlet temperature (T3) limiter loops in the controller. The engine has variable geometry in the fan, compressor and in the final nozzle.

**Test Data**

Data from a number of engine mounted sensors were captured using a transient data acquisition system. These parameters included those normally acquired by the aircraft’s EMS, namely, inlet temperature (T1), inlet pressure (P1), fan speed (NL), compressor speed (NH), compressor outlet static pressure (Ps3), fuel flow (WF) turbine outlet pressure (P56), low pressure turbine outlet temperature (T5), final nozzle area (A8) and power lever angle (PLA). Additional parameters were measured in the test-cell including compressor outlet temperature (T3) and the variable vane settings in the compression system. Snap accelerations, with and without implanted faults, were performed from several part power positions (PLA = 30, 60, 80 and 90 degrees) to IRP (PLA = 102 Degrees) to simulate typical take-off transients. The data were acquired at a sampling rate of 32 Hz and measured parameters were then corrected to standard day conditions but no pre-filtering was performed. All the data have been normalised to facilitate the presentation of the results.

**Implanted Faults**

The implanted faults comprised misscheduled fan variable geometry (FVG), misscheduled compressor variable geometry (CVG), final nozzle biased open 5% (A8+5%), biased TS sensor via a faulty TS harness (Biased TS) and a variation in...
Separation of the Bulk Metal Temperature Effect

A number of accelerations to IRP from various part power positions was performed on the non-faulty engine. Corrected data for a snap acceleration from flight-idle (PLA = 30 deg.) are shown in Figure 3. The bulk temperature effect is readily apparent from the decay in WF, Ps3 and EPR during the latter part of the transient when the engine is under the control of the NL and T5 limiters. Moreover, this trend is reflected in the rising A8 signal required to maintain the scheduled T5 value. The J(θ) distribution derived from the recursive least squares method is shown in Figure 4. The minimum J(θ)

---

RESULTS

Data obtained from an F404 engine in a SLS test-cell were used to examine the method outlined above as a way of isolating the fast acceleration component from the much slower bulk metal temperature effect. This technique was then applied to the problem of detecting a range of implanted faults.
point is located soon after the transient peak, as expected. Furthermore, the minimum variance point closely corresponds to the minimum gain position for this transient (Figure 4).

By comparison, results for a small transient commencing just below the IRP setting (PLA = 90 deg.) are shown in Figure 5. The bulk temperature effect is less readily discernible in this case because the changes in the metal temperatures will be small. Moreover, the minimum \( J(\theta) \) point is located approximately eight seconds after the peak of the transient (Figure 6).

The trend in the minimum \( J(\theta) \) position for snap accelerations commencing at PLA'S in the range 30 to 90 degrees is shown in Figure 7. The position of the minimum \( J(\theta) \) point for each acceleration is given as a ratio of \( \Delta t_{\theta}/\Delta t_{P3} \), where \( \Delta t_{\theta} \) represents the time interval between the minimum \( J(\theta) \) point and the initial starting point for the transient (based on the P3 criterion defined in Eq. 7) and \( \Delta t_{P3} \) is the corresponding P3 rise-time. A suitable definition of P3 rise-time for this purpose is the time taken to reach the peak P3 value during the transient. This proved to be a convenient and consistent way of scaling the individual transients. The P3 rise-time is shown to vary almost linearly with initial PLA (Figure 8). The results in Figure 7 indicate that the scatter increases as the scale of the transient decreases, that is, as the influence of the bulk metal temperature effect becomes smaller. The increased level of scatter associated with the small scale transients is still acceptable because it only has a small effect on the resultant gain estimates (Figure 6). Conversely, the scatter for the large flight-idle to IRP acceleration is small indicating that the boundary is clearly delineated under these conditions, as shown by the 'peaky' distribution in Figure 4.

It is immediately apparent from the above results that the boundary between the 'fast transient' and the slower 'bulk metal temperature' component moves toward the transient peak as the scale of the transient increases. Moreover, the criterion developed here is capable of separating these two effects without reference to other any parameters, such as final nozzle area.
The next step is to use the above criterion to enhance the detection process for a range of faults using transient data. The fault detection procedure consisted of monitoring the estimated pseudo steady-state gains from a number of take-off type acceleration transients, with and without faults. Normalised fault deltas were then formed by taking the difference between the fault and no-fault gain estimates as a percentage of the no-fault value.

Results for the three levels of misscheduled CVG fault are given in Figure 9. It is important to note the different scales used in Figure 9 in order to make sensible comparisons between the results for the various accelerations. The 6 degrees closed fault condition is clearly evident from all the accelerations. The sensitivity was found to increase as the magnitude of the acceleration decreased, that is, as the initial part-power starting point approached the IRP position. This can be more easily seen if the data for one fault level are plotted for the range of transients as in Figure 10. Insufficient no-fault data were available to reliably ascertain the variability in the no-fault estimates. Preliminary indications are that for parameter gains based on fuel flow, the no-fault uncertainty appears to fall in the range of 3-5% meaning that parameter deltas that fall below this will be undetectable. The smaller 2 degrees open fault condition is detectable from most of the parameters. However, the 2 degrees open fault is not easily detected from any of the parameters. One possible explanation for the reduced sensitivity to this latter fault condition is that the adjustable linkage was found to have a residual error.

Results for the other fault conditions are given in Figure 11. In the case of the FVG fault, in which the blade angles were biased 3.5 degrees closed from the nominal no-fault setting, it is detectable from accelerations commencing at or above PLA of 80 degrees. However, the parameter set is not adequate for the larger accelerations. The final nozzle fault, comprising a 5% bias in the open direction, is readily detectable across the full range of the acceleration transients but the sensitivity improves as the part-power starting point increases toward the higher power settings.

Again, the biased T5 indicator (approximately 60°C) arising from a faulty T5 harness is detectable from NH, Pa3, T3, T5 and EPR parameters from the accelerations with initial PLA's in the range 60 to 90 degrees. The results from flight-
DISCUSSION

The model based technique examined here is shown to be capable of detecting a range of operational type faults. The procedure developed to identify the boundary between the faster spooling-up part of the transient and the slower bulk temperature effect proved effective in separating these two quite distinct physical processes. Failure to incorporate this step in the analysis of transient engine data can lead to an increased level of uncertainty in the resultant gain estimates. This aspect is likely to be important when applying the technology to a field based system, such as that for an F/A-18 where the record length can vary significantly.

The model based method provides additional information to that obtained with a snapshot technique (Eustace et al., 1992) because the gain terms are given by parameter rates of change or slopes rather than absolute values at a particular instant, as is the case with a snapshot method. As a result, the model-based technique will be less sensitive to bias effects in the measured parameters. However, bias effects will still be important in the parameters that are used for some controller functions, such as T5, which is used in the F404 for limiting gas temperatures and therefore metal temperatures near IRP.

Some fault conditions were more easily detected than others. For example, the CVG relatively large 6 degree bias fault was easily detected from all part-power accelerations. The remaining faults were best detected from those accelerations commencing at the higher part-power positions. Generally, all fault conditions were less easily detected from the flight-idle to IRP accelerations. The addition of compressor outlet temperature to the instrumentation set increased the resultant fault coverage. This was especially true in the case of the compressor bleed fault.

Further information can be gleaned from the data set by substituting other parameters in place of fuel for the input. In addition, other techniques are being examined to further improve the fault coverage capability. For example, neural networks are being applied to the problem at ARL with encouraging results. However, these methods do require a comprehensive training data-base in order to perform well.

CONCLUSIONS

A model based system for detecting a range of faults in a turbofan engine has been examined. A new procedure has been defined for separating the slower bulk temperature effect from the faster spooling-up effect. This is an important consideration in some field based engine monitoring systems where the length of the captured record can vary significantly. It was shown that the records derived from the F/A-18 during take-off do not achieve true steady-state conditions and therefore a procedure is required to determine a consistent
The cut-off point for any transient. The minimum variance condition developed in this paper yielded excellent results when applied to test-cell data. However, the system has yet to be proven on field-based data.

The model-based method provides additional information to that obtained with a snapshot technique. Furthermore, the measurement noise is accounted for in a more meaningful way using the properties of a least squares estimator.

REFERENCES


