Gas turbine aero-engines real time on-board modelling: A review, research challenges, and exploring the future

Zhiyuan Wei a, Shuguang Zhang a, Soheil Jafari b, Theoklis Nikolaidis b*

a. School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
b. Centre for Propulsion Engineering, School of Aerospace Transport and Manufacturing, Cranfield University, Bedfordshire MK43 0AL, UK

Abstract

On-board real time modelling for gas turbine aero-engines has been extensively used for engine performance improvement and reliability. This has been achieved by the utilization of on-board model for the engine’s control and health management. This paper offers a historical review of on-board modelling applied on gas turbine engines and it also establishes its limitations, and consequently the challenges, which should be addressed to apply the on-board real time model to new and the next generation gas turbine aero-engines. For both applications, i.e. engine control and health management, claims and limitations are analysed via numerical simulation and publicly available data. Regarding the former, the methods for modelling clean and degraded engines are comprehensively covered. For the latter, the techniques for the component performance tracking and sensor/actuator diagnosis are critically reviewed. As an outcome of this systematic examination, two remaining research challenges have been identified: firstly, the requirement of a high-fidelity on-board modelling over the engine life cycle, especially for safety-critical control parameters during rapid transients; secondly, the dependability and reliability of on-board model, which is critical for the engine protection in case of
on-board model failure. Multiple model-based on-board modelling and runtime assurance are proposed as potential solutions for the identified challenges and their potential and effectiveness are discussed in detail.

**Keywords:** On-board modelling, Gas turbine aero-engines, Control, Health management, Degradation, Multiple model based on-board modelling, Runtime assurance, Research challenges

**Nomenclature**

- **ACARE**  Advisory Council for Aviation Research and Innovation in Europe
- **ADIA**  Advanced Detection, Isolation and Accommodation
- **ARREM**  Adaptive Real-time Reference Engine Model
- **ARTERI**  Analytical Redundancy Technology for Engine Reliability Improvement
- **ANFIS**  Adaptive Neuro-Fuzzy Inference System
- **CGEKF**  Constant Gain Extended Kalman Filter
- **CLM**  Component-Level Model
- **E3**  Energy Efficient Engine
- **EKF**  Extended Kalman Filter
- **EPDMC**  Engine Performance Degradation Mitigation Control
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPR</td>
<td>Engine Pressure Ratio</td>
</tr>
<tr>
<td>eSTORM</td>
<td>enhanced Self Tuning On-board Real-time Model</td>
</tr>
<tr>
<td>FADEC</td>
<td>Full Authority Digital Electronic Control</td>
</tr>
<tr>
<td>FICA</td>
<td>Fault Identification and Corrective Action</td>
</tr>
<tr>
<td>FOD</td>
<td>Foreign Object Damage</td>
</tr>
<tr>
<td>GTEs</td>
<td>Gas Turbine Engines</td>
</tr>
<tr>
<td>HEP</td>
<td>Hybrid Electric Propulsion</td>
</tr>
<tr>
<td>HKF</td>
<td>Hybrid Kalman Filter</td>
</tr>
<tr>
<td>KF</td>
<td>Kalman Filter</td>
</tr>
<tr>
<td>IHKF</td>
<td>Improved Hybrid Kalman Filter</td>
</tr>
<tr>
<td>ISA</td>
<td>International Standard Atmosphere</td>
</tr>
<tr>
<td>MBEC</td>
<td>Model-based Engine Control</td>
</tr>
<tr>
<td>MHKF</td>
<td>Multiple model-based Hybrid Kalman Filter</td>
</tr>
<tr>
<td>N1</td>
<td>Low-pressure spool speed</td>
</tr>
<tr>
<td>N2</td>
<td>High-pressure spool speed</td>
</tr>
<tr>
<td>NARMAX</td>
<td>Nonlinear Autoregressive Moving Average with eXogenous inputs</td>
</tr>
<tr>
<td>NARX</td>
<td>Nonlinear Autoregressive network with eXogenous input</td>
</tr>
<tr>
<td>NGDF</td>
<td>Novel Generalized Describing Function</td>
</tr>
<tr>
<td>NOBEM</td>
<td>Nonlinear On-Board Engine Model</td>
</tr>
<tr>
<td>OTEKFW</td>
<td>Optimal Tuner Extended Kalman Filter</td>
</tr>
<tr>
<td>OTKF</td>
<td>Optimal Tuner Kalman Filter</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RTA</td>
<td>Runtime Assurance</td>
</tr>
<tr>
<td>PF</td>
<td>Particle Filter</td>
</tr>
<tr>
<td>PLA</td>
<td>Power Lever Angle</td>
</tr>
<tr>
<td>PWL</td>
<td>Piecewise Linear</td>
</tr>
<tr>
<td>SCLM</td>
<td>Simplified Component-Level Model</td>
</tr>
<tr>
<td>SLS</td>
<td>Sea Level Static</td>
</tr>
<tr>
<td>SM</td>
<td>Surge Margin</td>
</tr>
<tr>
<td>SN</td>
<td>Smoke Number</td>
</tr>
<tr>
<td>STF</td>
<td>Strong Tracking Filter</td>
</tr>
<tr>
<td>STORM</td>
<td>Self Tuning On-board Real-time Model</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>TET</td>
<td>Turbine Entry Temperature</td>
</tr>
<tr>
<td>UHBR</td>
<td>Ultra High Bypass Ratio</td>
</tr>
<tr>
<td>Δ</td>
<td>Delta</td>
</tr>
<tr>
<td>Γ</td>
<td>Flow capacity</td>
</tr>
<tr>
<td>η</td>
<td>Efficiency</td>
</tr>
<tr>
<td>π</td>
<td>Pressure ratio</td>
</tr>
</tbody>
</table>

**Subscript**

C  Compressor  
F  Fan
1 Introduction

The mechanisms that cause gas turbine aero-engines degradation over the life cycle inevitably lower the safety margin [1, 2], overall efficiency [3], and reliability of the engine [4]. These mechanisms include blade surface changes (due to erosion, corrosion, or fouling) that influence blade aerodynamics, seal clearances changes that affect parasitic flows, combustion system changes, and abrupt fault from foreign object damage [5]. Degradation effects deteriorate the gas path component health status, which is still difficult to accurately obtain [6, 7]. As a consequence, a significantly conservative safety margin that is usually set to end-of-life engines must be considered under the existing industrial engine control system using sensor-based strategies [8-10]. In the meantime, advanced condition-based maintenance approach is hard to achieve without having accurate component health status information of the monitored engine [3]. Moreover, next generation of aero-propulsion systems (e.g. Ultra High Bypass Ratio (UHBR) engines and Hybrid Electric Propulsion (HEP)) should be designed to deal with challenging targets of ultra-high efficiency and ultra-low emissions [11-13]. Advisory Council for Aviation Research and Innovation in Europe (ACARE) Flightpath 2050 has established very challenging targets for emissions and noise, including reductions
of 75% in CO$_2$, 90% in NO$_x$, and 65% in noise by 2050 that are compared to the capabilities of the typical new aircraft in 2000 [14]. These ambitious targets could not be achieved only by means of marginal improvement in the engine component design and development. On-board modelling of gas turbine aero-engines is a promising solution for the engine performance improvement, maintenance decisions optimization, health management, and model-based controllers design as well as future aero-propulsion requirements [7, 15].

On-board modelling of gas turbine aero-engines concentrates on the in-flight and real-time simulation for some pre-defined critical parameters, to accommodate the engine-to-engine variations [1, 6, 16]. On-board models could provide estimated measurable values (e.g. engine shaft rotational speed, and exhaust gas temperature) and unmeasurable performance parameters of interest (e.g. surge margin, thrust, and health parameters) with the maximal fidelity and affordable computation burden. Therefore, on-board models of gas turbine engines (GTEs) could be applied to advanced model-based controllers and health management systems [3, 17-19].

Since 1977, when the piecewise linear modelling approach was utilized as the initial application of on-board model by Teren for GTEs [20], two different fields have been covered by on-board modelling approaches:

- On-board modelling for GTEs control purposes
- On-board modelling for GTEs health management purposes

Different applications of on-board models utilize different modelling features. Control oriented on-board models are usually applied in the control loops based on their estimated values. These models are capable of giving the estimation for thrust, turbine exit temperature (TET) and surge margin (SM)
with high accuracy so that they provide the health awareness of the controlled engine which is totally different from current industrial sensor-based control. Therefore, personalized engine control would be expected to accommodate gradual degradation, abrupt fault, and even manufacturing tolerance or weather conditions [1]. In contrast, on-board models for health management are usually served as indicators for maintenance recommendations and control system integrity assurance [5, 6]. Outputs from these on-board models are usually not fed back to the control loops. Component performance tracking, also known as gas path analysis, is dedicated to the accurate perception of component health status. Estimated outputs of component health parameters enable the condition-based maintenance and optimize the shop visit scope [15]. Sensor and actuator diagnosis are focusing on the estimation of their measurable values that could be regarded as virtual measurements. Only if sensor fault reconstruction is achieved, the estimated value would be transmitted back to the control loop in current sensor-based control strategies [21].

To cover all the above-mentioned aspects, a comprehensive historical review of on-board models for gas turbine aero-engines is firstly conducted in section two in which all claims and constraints are confirmed with references and simulation results. Subsequently, the existing challenges of GTEs on-board models are identified and analysed in section three. Moreover, future exploration and corresponding potential solutions for the next generation of aero-propulsion systems using on-board models are proposed in section four.


2 A history of on-board modelling of gas turbine aero-engines

On-board modelling for gas turbine aero-engines has covered the scope of control purposes and health management, as depicted in Fig. 1. In one hand, on-board models for control purposes is mainly emphasized on nominal engines (specifically, ideal or new engine performance) and degraded engines (under abrupt fault, gradual degradation and a special problem for large turbofan engines). The on-board model for control purposes, capable of estimating unmeasurable thrust, surge margin and TET, could contribute to model-based control strategies, under which the conservatism in current sensor-based controller would be significantly decreased [10]. A more ultimate control structure based on on-board model could adapt itself to different weather conditions or fault scenarios, thus intelligent engine control with the capability of health state awareness of the controlled engine would be expected [1]. On the other hand, the on-board model for health management gives detailed information on engine component performance tracking and sensor/actuator diagnosis [3]. Component health status tracking, including abrupt fault and gradual degradation, could contribute to the cost-efficient condition-based maintenance [5]. Sensor/actuator diagnosis, which historically focused on nominal and degraded engines, could potentially assure the control system integrity [6]. The next section focuses on a comprehensive critical review on the history of on-board modelling approaches in gas turbine aero-engines for both control and health management purposes.
2.1 On-board modelling of gas turbine aero-engines for control purposes

For control oriented on-board modelling, safety-critical unmeasurable parameters including surge margin (SM), turbine exit temperature (TET) and thrust, are highly concerned because these parameters must be controlled or limited to ensure the safe operation of GTEs. Different approaches have been utilized on the on-board modelling for nominal and degraded GTEs. Nominal engine on-board modelling approaches for control purposes are mainly piecewise linear models, nonlinear identification models, block structure models, and Novel Generalized Describing Function (NGDF). Degraded engines on-board modelling techniques for control aims are dominated by estimation algorithms using Kalman filter framework (e.g. Linear Kalman filter, Extended Kalman filter, Hybrid Kalman filter, et al.). These control-oriented on-board modelling methods are systematically discussed.
with their advantages, disadvantages, and limitations via analysis and simulation cases in following sub-sections.

2.1.1 Nominal engines on-board modelling

Several efforts have been spared on the on-board modelling for nominal gas turbine aero-engines. The term, nominal engines modelling, aims to describe the ideal/new engine performance without accounting for the un-deterministic engine-to-engine variation from degradation effects through the engine life cycle [22].

Using piecewise linear (PWL) model for on-board applications is dating back to 1977 when Teren investigated the minimum acceleration response for F100 turbofan engine [20]. The PWL model contains a set of steady-state points and the corresponding linear state-space models along the steady operating line of the engine. Thus, the engine transient response could be predicted via the interpolation between the steady points and the linear state-space models using a scheduling parameter (e.g. engine spool speed). The pseudo-code of PWL modelling approach is presented in Table 1. The simple structure of PWL model enables the on-board and real-time simulation capability. Steady-state points could be retrieved from nonlinear aero-thermal engine model or engine test data. Linear state-space models are usually obtained via partial derivative methods, fitting methods, and identification approaches [23, 24]. In 2004, Reberga and Breikin compared different interpolation methods for PWL model, including linear interpolation and polynomial interpolation, to avoid the discontinuity of this approach [25]. They confirmed that different interpolation approaches have very limited influence on the accuracy of PWL model. The next step was the research by Tian on the comparison of different
scheduling parameters of PWL model with the application to a two-spool turbofan engine [26]. Different scheduling parameters were compared, including low-pressure spool speed (N1), high-pressure spool speed (N2) and Euclidean norm of N1 and N2. It was demonstrated that Euclidean norm of N1 and N2 as the scheduling parameter for PWL model surpasses the other two over the flight envelope. However, PWL model only shows acceptable accuracy in mild transients due to its inherent piecewise linear nature.

Table 1 PWL on-board modelling algorithm adapted from [27]

<table>
<thead>
<tr>
<th>Algorithm: PWL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td>State-space model, static points, initial conditions, scheduling parameter, control inputs, time step, final simulation time</td>
</tr>
<tr>
<td><strong>Process</strong></td>
</tr>
<tr>
<td>1. repeat</td>
</tr>
<tr>
<td>2. Search current static points using the relationship between static points model and scheduling parameter at the current time step</td>
</tr>
<tr>
<td>3. Find current linear model coefficients by the interpolation or polynomial fitting between scheduling parameter and state-space model</td>
</tr>
<tr>
<td>4. Calculate the incremental of control input and state variables</td>
</tr>
<tr>
<td>5. Get the derivative for state variables</td>
</tr>
<tr>
<td>6. Generate real value for state variables and output variables at the current time step</td>
</tr>
<tr>
<td>7. until Final simulation time is reached</td>
</tr>
<tr>
<td><strong>Output</strong></td>
</tr>
<tr>
<td>Parameter outputs from PWL</td>
</tr>
</tbody>
</table>

From 2001 to 2002, Chiras, Evans, and Rees investigated the nonlinear global identification modelling based on Rolls Royce Spey MK202 turbofan engine experiment data [28-30]. Previous studies on the local linear model identification clearly confirmed that the engine dynamics varies with the engine operating points so that a global nonlinear engine model was motivated [31, 32]. They firstly applied NARMAX (Nonlinear autoregressive moving average with exogenous inputs)
method to identify the relationship between fuel flow and engine spool speed [28, 29]. NARX (Nonlinear autoregressive network with exogenous input), a NARMAX model with the noise terms excluded, is taken as an example here. As a time-series modelling approach, NARX relates current output $y(t)$ to past outputs $y(t-m_y)$, past inputs $u(t-n_u)$, and current inputs $u(t)$, as Eq.(1),

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-m_y), u(t), u(t-1), \ldots, u(t-n_u))$$

(1)

where $y$ and $u$ denote system outputs and inputs; $m_y$ and $n_u$ are lags for outputs and inputs with corresponding model orders; $f$ is an identified nonlinear function; $t$ is the time step. The nonlinear function and the model order must be carefully selected to represent the engine response based on a set of engine operating data, which is usually excited by a well-defined excitation input to show sufficient engine dynamics. Then, they used a feed-forward neural network to identify the same relationship [30]. Small signal tests and large signal tests demonstrated that the identified neural network model almost eliminates the steady error that exists in the linear transfer function model.

In 2003, Ruano utilized a multi-objective genetic algorithm (MOGA) to determine the model structure for NARX and neural network in Spey turbofan engine identification [33]. It was concluded that a quadratic NARX model is appropriate for the relationship between fuel flow and engine spool speed. However, only the measurable parameter, shaft speed, was focused by this study. At the same time, Maggiore et al. applied neural network to nonlinear identification modelling for XTE46 variable cycle engine [34]. Based on the discussion of the minimal sets of estimator inputs and estimation feasibility for unmeasurable parameters, they concentrated on the unmeasurable thrust, surge margin and turbine entry temperature. Only steady-state identification was carried by this research. The next step was the study by Tavakolpour-Saleh et al. who proceeded with NARX model for turbojet engine
identification [35]. Parametric identification (known model structure) and non-parametric identification (unknown model structure) were compared. It showed the mean square error between actual outputs and predicted outputs from non-parametric identification is decreased by at least 50% than that of parametric identification. However, it is difficult to find a physical explanation for the non-parametric identification model. Recently, Jafari et al. combined different identification models for different engine parameters (engine spool speed, compressor discharge pressure, and emission values) of CFM56-5A turbofan engine [36]. They managed to find the best candidate models for different parameters and to develop the most accurate combination for control purposes. A reliable validated model for a turbofan engine is firstly developed using Adaptive Neuro-Fuzzy Inference System (ANFIS) and nonlinear autoregressive network with exogenous inputs (NARX) techniques (Polynomial & Hammerstein-Wiener). The control structure is then designed using the Min-Max control strategy which is the most widely used control algorithm for gas turbine aero-engines.

Another on-board modelling technique for nominal engines is Novel Generalized Describing Function (NGDF) proposed by Lichtsinder and Levy in 2006 [37]. The main idea behind NGDF is to obtain a high-fidelity on-board model, especially under rapid fuel flow variations. This significant fuel flow change is highly concerned in GTEs control system because it could drive the engine closer to the safety limits (e.g. surge limit, TET limit, and blown-out limit) during transient states. The accuracy of NGDF is guaranteed through an error minimization optimization approach, as shown in Fig. 2.
In 2019, Jafari and Nikolaidis gave a comprehensive review of the block structure models with application to GTEs modelling [8]. These models are Wiener model, Hammerstein model, Wiener-Hammerstein model, and Hammerstein-Wiener model. They contain different cascade connections of linear dynamic elements and nonlinear static elements. Their structures are illustrated in Fig. 3. Variables of these models could be tuned via engine experimental data or verified aero-thermal engine model [38, 39].

- Hammerstein models are usually regarded as the candidate for systems under which nonlinearity is only affected by the direct current gain [27]. The input amplitude does not influence the dynamic behaviour of the model. Thus, it seems this model is not appropriate for GTEs modelling since GTEs dynamics vary significantly with different input amplitudes.

- Wiener models represent the nonlinearities for different input amplitudes, conversely [27]. This dynamics from Wiener model is consistent with the dynamic characteristic of gas turbine engines whose gains and response time vary with input magnitude and operating points [40]. It also has fewer parameters to be tuned than Wiener-Hammerstein model and Hammerstein-
Wiener model.

- Wiener-Hammerstein models and Hammerstein-Wiener models have relatively complex structures so that these models are hard to produce and validate.

![Block structure models](image)

Fig. 3 Block structure model (a) Hammerstein (b) Wiener (c) Wiener-Hammerstein (d) Hammerstein-Wiener adapted from [8]

Therefore, in block structure models, Wiener model seems to be the superior choice for GTEs onboard modelling due to its dynamics and simple structure.

In order to identify the most high-fidelity onboard model of nominal engines for control purposes, PWL model described in [27], NGDF model described in [37], and Wiener model described in [8] are tested on a small turbofan engine. The examined engine is a dual-spool, separate exhaust turbofan engine with intended application to general aircraft. The specification of the engine at take-off condition, i.e. international standard atmosphere (ISA) and sea-level static (SLS), is shown in Table 2. All the above-mentioned onboard modelling approaches are compared with the turbofan engine aero-thermal model that has been validated against experimental data.
Table 2 Take-off specification of the examined turbofan engine

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient condition</td>
<td>ISA SLS</td>
</tr>
<tr>
<td>Intake mass flow (kg/s)</td>
<td>13.68</td>
</tr>
<tr>
<td>Low-pressure spool speed (rpm)</td>
<td>44233</td>
</tr>
<tr>
<td>High-pressure spool speed (rpm)</td>
<td>50990</td>
</tr>
<tr>
<td>Fan pressure ratio</td>
<td>1.20</td>
</tr>
<tr>
<td>High-pressure compressor pressure ratio</td>
<td>4.38</td>
</tr>
<tr>
<td>Fuel flow (kg/s)</td>
<td>0.036</td>
</tr>
<tr>
<td>Maximum thrust (kg)</td>
<td>246.48</td>
</tr>
</tbody>
</table>

As shown in Fig. 4, the examined PWL, NGDF, and Wiener on-board models were tested with the turbofan engine under a series of rapid acceleration and deceleration transients between idle and take-off thrust at sea level static condition. The engine was represented by its validated aero-thermal model. The unmeasurable safety-critical parameters, surge margin (SM) of high-pressure compressor (HPC), and thrust were selected to demonstrate the performance of different on-board models. From the zooming plot in Fig. 4(a), Wiener model shows the best tracking performance for surge margin in acceleration states. However, the PWL model shows an excessive undershot, which will be too conservative for engine transient performance, if the estimates from PWL were feedback to the control system. Meanwhile, NGDF model is subject to an unfavourable accumulation of steady error. Although the transient accuracy from Wiener model for surge margin at the end of deceleration states (e.g. at 60 s) is not that high, it is still acceptable. This is based on the fact that the engine at deceleration states operates far away from its HPC surge line. From Fig. 4(b), it is also concluded that Wiener model is the best on-board modelling approach in terms of thrust tracking in both acceleration and deceleration transients.
Therefore, Wiener model could be the best candidate on-board model of nominal turbofan engines for control purposes, especially when unmeasurable parameters are important. However, due to the
normal aging of GTEs from erosion, corrosion and fouling, as well as the unexpected incidents from foreign object damage, the engine performance deviates from its nominal state. Nominal engines on-board model could not provide accurate information for a particular engine. Therefore, on-board modelling for GTEs over the full life cycle is reviewed in the following sub-section.

2.1.2 Degraded engines on-board modelling

On-board modeling for degraded GTEs has incorporated the scope of abrupt fault and gradual degradation since gas turbine aero-engine degradation is usually categorized into gradual degradation and abrupt fault [41]. Abrupt fault is usually caused by foreign object damage (FOD, e.g. ice ingestion and bird ingestion), thus only limited components (one, at most two) are affected very rapidly at a time [42]. However, gradual degradation typically triggered by erosion, corrosion, fouling, wear and tip clearance change is a slow aging process with many flights while all components deviate slowly and simultaneously from their nominal/new states with usage [6]. A comparison of the physical cause and the impact on components of abrupt fault and gradual degradation in GTEs is summarized in Table 3.

Table 3 Comparison between abrupt fault and gradual degradation in GTEs

<table>
<thead>
<tr>
<th></th>
<th>Abrupt fault</th>
<th>Gradual degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical physical cause</td>
<td>Foreign object damage (ice, bird ingestion)</td>
<td>Erosion, corrosion, fouling, wear, tip clearance change</td>
</tr>
<tr>
<td>Number of affected components at a time</td>
<td>Limited (one, at most two)</td>
<td>All</td>
</tr>
<tr>
<td>Influence on variation rate (time scale) of component performance</td>
<td>Rapid, usually within several seconds</td>
<td>Slow, usually over several flight cycles</td>
</tr>
<tr>
<td>Evidence</td>
<td>Luppold et al. [43]</td>
<td>Sallee et al. [46, 47]</td>
</tr>
<tr>
<td></td>
<td>Volponi [44]</td>
<td>Kobayashi et al. [48]</td>
</tr>
<tr>
<td></td>
<td>Borguet et al. [42, 45]</td>
<td>Simon et al. [49, 50]</td>
</tr>
</tbody>
</table>
The primary concern is how to realistically model the degradation effects in the nonlinear engine aero-thermal model. Once the nonlinear engine aero-thermal model is validated against engine test data, it could be treated as a baseline model for on-board models validation. Degradation on GTEs performance is usually modelled by introducing health parameters (e.g. degradation coefficients for flow capacity, isentropic efficiency, and pressure ratio) for compressors and turbines, which represent the health status/deviation of major components in the gas path [1, 3]. Therefore, degraded component maps could be obtained via the movement of their original maps based on these health parameters. This is based on the assumption that the shape of degraded component maps remains almost the same as their original maps since component geometries do not hugely change due to the degradation effects [51]. The health parameters are defined in Eq.(2)-(3) for compressors and turbines,

$$\Delta \Gamma = \frac{\Gamma_{\text{real}} - \Gamma_{\text{clean}}}{\Gamma_{\text{clean}}} = \frac{\Gamma_{\text{real}}}{\Gamma_{\text{clean}}} - 1$$ (2)

$$\Delta \eta = \eta_{\text{real}} - \eta_{\text{clean}}$$ (3)

where $\Gamma_{\text{real}}$ and $\Gamma_{\text{clean}}$ are corrected flow capacity for degraded maps and clean maps of compressors or turbines, respectively; $\eta_{\text{real}}$ and $\eta_{\text{clean}}$ denote isentropic efficiency for degraded maps and clean maps of compressors or turbines; $\Delta \Gamma$ and $\Delta \eta$ are the health parameters for flow capacity and isentropic efficiency, respectively.

Particularly, for a realistic representation of the degradation effects for the compressors including fan, booster and high-pressure compressor, three health parameters for its flow capacity $\Delta \Gamma_{C}$, pressure ratio $\Delta \pi_{C}$ and isentropic efficiency $\Delta \eta_{C}$ should be introduced [52, 53], rather than only two health parameters (flow capacity and isentropic efficiency). This utilization of three health
parameters in compressor characteristics adjustment due to degradation effects was also confirmed by the experiment on a deteriorated high-pressure compressor of V2500 turbofan engine [54]. The health parameter for pressure ratio $\Delta \pi_c$ of compressor is usually assumed to be equal to that of flow capacity $\Delta \Gamma_c$ since they have same effects on engine performance, and this avoids multiple solutions in health parameter estimation, as shown in Eq. (4) [51]. Additionally, turbine degraded maps are usually obtained through two health parameters for flow capacity and isentropic efficiency, as shown in Eq. (2)-(3) [55].

$$\Delta \pi_c = \Delta \Gamma_c$$ (4)

The difference in modelling approaches between abrupt fault and gradual degradation in GTEs lies in the variation rate and numbers of health parameters. For abrupt fault modelling, step changes to some health parameters are usually imposed while other health parameters remain to their nominal value [55, 56]. However, under gradual degradation, all health parameters deviate slowly from their nominal values with increasing flight cycles at the same time [46].

National Aeronautics and Space Administration (NASA) has quantified the gradual degradation trend for JT9D turbofan engine based on in-service data from airliners, airframe manufacturers, and engine test results, as shown in Table 4 [46, 47]. It should be noted that only the health parameters of flow capacity for HPT and LPT are positive due to the expanded turbine throat area resulted from degradation effects. All the other health parameters remain negative over the engine life cycle.
Table 4 Gradual degradation trend of JT9D turbofan engine [46, 47]

<table>
<thead>
<tr>
<th>Flight cycle</th>
<th>$\Delta \eta_F$ (%)</th>
<th>$\Delta \Gamma_F$ (%)</th>
<th>$\Delta \eta_B$ (%)</th>
<th>$\Delta \Gamma_B$ (%)</th>
<th>$\Delta \eta_{HPC}$ (%)</th>
<th>$\Delta \Gamma_{HPC}$ (%)</th>
<th>$\Delta \eta_{HPT}$ (%)</th>
<th>$\Delta \Gamma_{HPT}$ (%)</th>
<th>$\Delta \eta_{LPT}$ (%)</th>
<th>$\Delta \Gamma_{LPT}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3000</td>
<td>-1.50</td>
<td>-2.04</td>
<td>-1.46</td>
<td>-2.08</td>
<td>-2.94</td>
<td>-3.91</td>
<td>-2.63</td>
<td>+1.76</td>
<td>-0.54</td>
<td>+0.25</td>
</tr>
<tr>
<td>6000</td>
<td>-2.85</td>
<td>-3.65</td>
<td>-2.61</td>
<td>-4.00</td>
<td>-9.40</td>
<td>-14.06</td>
<td>-3.81</td>
<td>+2.57</td>
<td>-1.08</td>
<td>+0.42</td>
</tr>
</tbody>
</table>

For clarification, the generation of degraded component maps based on the above-mentioned degradation modelling approach in Eq. (2)-(4) and the gradual degradation trend in Table 4 is presented. Fig. 5 illustrates clean compressor map and degraded compressor maps at 3000 and 6000 flight cycles using the booster degradation data, $\Delta \eta_B$ and $\Delta \Gamma_B$, in Table 4. It can be clearly seen that the degraded compressor map deviates to the left-down side (lower flow capacity, pressure ratio, and isentropic efficiency) from the clean map due to the variation from health parameters $\Delta \Gamma_B$, $\Delta \pi_B$, and $\Delta \eta_B$. Thus, the degraded maps could be loaded into the nonlinear aero-thermal engine model to predict the overall degradation effects on the engine performance.
2.1.2.1 Abrupt fault

On-board engines modelling for abrupt fault was initially proposed by Luppold et al. from Pratt
Their self-tuning on-board real-time model (STORM) for aircraft turbofan engine was based on Kalman filter and piecewise linear (PWL) model to estimate five engine health parameters and unmeasurable thrust and flow capacity under foreign object damage, as illustrated in Fig. 6. Residuals between real engine data and estimated outputs from PWL are sent to Kalman filter that serves as a tuner to estimate engine health status. Then PWL outputs give engine performance parameters due to degradation effects. Experiments showed that this model is able to track the abrupt change to the five health parameters and could estimate mass flow. However, the thrust estimation does not show the undershoot response during the transients when abrupt fault occurs. Followed research by Kerr et al. extended STORM to estimate foreign object damage of PW1128 turbofan engine [57]. Four health parameters for the engine cold section (fan flow capacity, fan efficiency, compressor flow capacity, and compressor efficiency) were selected for ice ingestion as moderate damage and bird ingestion as severe damage in the engine compression part. Engine test data for ice and bird ingestion showed that this model could estimate thrust and flow capacity change under these two abrupt faults. However, the compressor discharge temperature estimation shows an undesirable undershoot opposite to that of the actual engine response (a slight overshoot) under the implanted abrupt faults. In 2002, Turevski and Luppold from Pratt & Whitney built a model-based controller for commercial turbofan engines based on STORM [58]. It consisted of estimated values for thrust, compressor surge margin, combustor blowout margin, emissions, and temperature margin from STORM. The results showed that the model-based controller based on STORM improves the thrust response.
Between 2000 and 2001, Lietzau and Kreiner from MTU aero engine developed an on-board turbofan engine model using a Kalman filter as an observer [18, 59]. Via the estimation of four health parameters (efficiency for fan, HPC, HPT and LPT), surge margin, and TET estimation for the nominal and degraded engine were obtained. The direct control of estimated surge margin and TET, as shown in Fig. 7, demonstrates its superiority over sensor-based control for GTEs with regard to surge margin and TET limitation for the degraded engine. However, the tracking time for abrupt fault in HPC efficiency exceeds an order of 100 seconds.

Fig. 6 STORM model adapted from [44]
In 2011, Guo and Lu et al. further developed an improved hybrid Kalman filter (IHKF) based on-board model [60]. A nonlinear aero-thermal engine model is used to update the health baseline to ensure that the nonlinear on-board model works in the vicinity of the degraded engine. And a piecewise linear Kalman filter serves to tune the aero-thermal engine model. They also proposed different health parameter updating mode for different engine states. Four health parameters, fan efficiency, compressor efficiency, HPT flow capacity, and LPT flow capacity, were focused. Static estimation accuracy for spool speed and thrust is improved than basic HKF [61]. However, thrust estimation from IHKF does not track the abrupt thrust response from FOD in transient states.

It was in 2017 that Lu et al. extended STORM model with an extreme learning machine (a kind of neural network) to form an in-flight adaptive engine model in order to compensate the model-plant mismatch [62]. Bias estimation for health parameters is mitigated. And measurable outputs including engine spool speed, pressure, and temperature were given over the full flight envelope. However, they only focused on measurable outputs estimation.

Nonlinear filter algorithms for GTEs on-board modelling also attracted much attention due to the
increasing computational capacity of microprocessors in modern FADEC (Full Authority Digital Electronic Control) system. Between 1996 and 2000, Sugiyama firstly proposed constant gain extended Kalman filter (CGEKF) for GTEs on-board modelling, in which a real-time nonlinear aero-thermal engine model and extended Kalman filter were exploited [63, 64]. A single gain matrix at ground maximum thrust was computed for the whole envelope to avoid the extensive computational burden from extended Kalman filter (EKF). For abrupt fault in one gas path component, represented by the faulty component’s health parameters for flow capacity and efficiency, simulation results showed that this CGEKF model could estimate health parameters deviation and the unmeasurable parameters (thrust, TET, bypass airflow, and specific fuel consumption) with very little static error. However, turbine entry temperature overshoot (up to 1%) due to a 5% abrupt fault of HPT flow capacity is not tracked by the CGEKF model in transient states. Subsequently, he successfully applied CGEKF to conduct a model-based thrust control experiment for a nominal turbojet engine on altitude test facility [65], but the static error for thrust estimation increases with the increasing Mach number from 0 to 1.6, in which the maximum steady error is up to 7% approximately at Mach number 1.6. In 2005, Kobayashi extended CGEKF to turbofan engines on-board modelling for abrupt fault [48]. It was found that the sensor set with bypass nozzle pressure indicates better performance for thrust estimation since this sensor has a relatively strong coupling with fan health status. However, the thrust estimation during abrupt fault shows undesirable oscillation.

More recently, Zheng et al. applied neural network (NN) for on-board modelling of a dual-spool turbofan engine [66]. Based on the simulation data from nonlinear aero-thermal engine model within a large envelope and five health parameters variation, unmeasurable thrust, surge margin as well as
TET are obtained. However, the training process is very complex and time-consuming due to the limited scope of the engine experimental data under different health status.

In general, on-board modelling approaches for engines under abrupt fault, mainly based on Kalman filter framework, could be summarized as follows:

- Estimate the health parameters of the limited components under abrupt fault by the minimization (fulfilled by Kalman filter) between engine measurements and predicted measurements from the on-board model;
- Predict the unmeasurable safety-critical parameters by tuning the on-board model based on the health parameter estimation.

2.1.2.2 Gradual degradation

In 2005, Kobayashi also extended Sugiyama’s research on CGEKF [63] to gradual degradation for GTEs on-board modelling [48]. Sensor selection strategy was discussed for sensor sets with and without bypass pressure sensor. Simulation cases for gradual degradation showed that with bypass pressure sensor, thrust and fan surge margin estimation accuracy are enhanced due to the strong relationship between bypass pressure and fan health status. However, no transient estimation comparison is shown for unmeasurable control parameters at different engine degradation levels. Recently, DeCastro further associated CGEKF with a neural network to form a hybrid on-board model [67]. It was designed to compensate for the model-plant mismatch. All health parameters were set to -4% to simulate a severe degradation condition. Simulation results demonstrated that this hybrid model could track real thrust despite a relatively obvious steady error (up to 6%).
Shankar et al. developed a neural network (NN) based adaptive observer for GTEs on-board modelling [68]. Taking advantage of radial basis function neural network to handle the nonlinearity, they augmented linear Kalman filter with neural network. This model was trained by two training data sets, new and full degraded engine data. Thus, unmeasurable control parameters could be directly estimated without health parameters estimation. However, the results showed that this model owns a disfavored overshoot, over 10%, for thrust tracking during rapid transient states from idle to take-off condition.

Generally, on-board modelling methods for degraded engines, covering abrupt fault and gradual degradation, with control purposes evolve from the linear to nonlinear approaches. STORM and its extension, inflight adaptive model, both rely on the linear Kalman filter framework. Nonlinear on-board models, including IHKF, CGEKF, neural network observer, and neural network model aim at enhancing the estimation accuracy for unmeasurable safety-critical parameters especially during rapid transient states. However, every modelling technique subjects to advantages and limitations, as summarized in Table 5.
Table 5 Comparison of on-board modelling methods for degraded engines focusing control purposes

<table>
<thead>
<tr>
<th>Model</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>STORM [43]</td>
<td>Pioneer work of on-board modelling for degraded engines with abrupt fault</td>
<td>Limited scope for abrupt fault on some components/health parameters</td>
</tr>
<tr>
<td>IHKF [60]</td>
<td>Proposed different updating modes of health parameters for different engine operating states</td>
<td>Unable to track the rapid thrust response under abrupt fault</td>
</tr>
<tr>
<td>Inflight adaptive model [62]</td>
<td>Combination of extreme learning machine to basic STORM for model-plant mismatch</td>
<td>Only focus on the measurable parameters for control purposes (spool speed and pressure)</td>
</tr>
<tr>
<td>CGEKF [48, 63-65]</td>
<td>Single gain matrix in EKF for whole envelop to mitigate computation burden</td>
<td>Obvious steady error (up to 6%) for thrust tracking</td>
</tr>
<tr>
<td>NN model [66]</td>
<td>Exploitation of the strong nonlinear mapping capacity from neural network for some pre-defined engine degraded levels</td>
<td>Subject to the expanded scope of the training data</td>
</tr>
<tr>
<td>NN observer [68]</td>
<td>No need for health parameter estimation</td>
<td>Disfavored transient overshoot (over 10%) for thrust tracking during rapid transients</td>
</tr>
</tbody>
</table>

Given that STORM (based on Kalman filter) was proposed by Pratt & Whitney as the most successful on-board modelling approach for degraded GTEs, it is taken as a test case for evaluation. Simulation results on different levels of abrupt fault on the compressor efficiency of the same dual-spool turbofan engine (as described in Table 2) are illustrated in Fig. 8. The engine was initially run at a nominal condition at its take-off thrust at sea level static condition. At 1 second, compressor efficiency health parameter was suddenly reduced by -1% to simulate an abrupt fault. During the simulation period, fuel flow to the combustor was maintained constant. It could be clearly seen that the STORM performance is quite favourable for -1% abrupt fault on compressor efficiency, regarding the compressor efficiency health parameter and surge margin of HPC. However, when the health parameter for compressor efficiency was decreased by -3% and -5% to simulate a severe abrupt fault, the steady estimation errors of the health parameter and surge margin increase. The steady estimation
error can be attributed to the linear limitation from Kalman filter. When -1% abrupt fault for compressor efficiency occurred, it could be concluded that the engine still works in the vicinity of its nominal operating point thus the Kalman filter could work to an acceptable level. However, when the abrupt fault level became more severe, Kalman filter is not capable of estimating to a high fidelity. Although the largest steady estimation error for surge margin under -5% compressor efficiency fault is just over 0.01, the gap between the estimation value and true surge margin, namely a potential design space for engine performance improvement, should still be minimized.
Fig. 8 STORM simulations for different abrupt fault levels on HPC efficiency

2.1.2.3 A Special problem for large turbofan engines

The special problem behind control-oriented on-board modelling for large turbofan engines under
degradation effects is the underdetermined estimation. Unmeasurable parameters for control purposes (surge margin, turbine entry temperature, and thrust) have strong coupling effects with the engine health status [69]. However, one existing industrial practice for in-service civil large turbofan engines is that the available on-board sensors are usually less than the to-be-estimated health parameters. For two-spool large turbofan engines with five gas path components (fan, booster, HPC, HPT and LPT), there are 10 health parameters in total. It is assumed that every gas path component has two health parameters (flow capacity and isentropic efficiency), whereas pressure ratio health parameters that equal to flow capacity health parameters for fan, booster and HPC are implicitly embedded, as presented in Eq.(4) [51]. However, available on-board sensors in the gas path are usually six to seven [70, 71]. This inevitably results in an underdetermined problem.

Several efforts have been spared on this special problem. Luppold et al. firstly proposed a subset approach for GTEs on-board modelling [43]. This study selected a subset of all health parameters that influence the overall engine performance mostly. It is not a systematic approach since it is only achieved by trial and error and highly relies on the experience from the designer.

In 2008, Litt et al. further developed a singular value decomposition (SVD) approach for large turbofans on-board modelling [69]. They applied singular value approach to get a low-dimensional health parameter subset appropriate for Kalman filter. The results showed that the relative error for booster surge margin between estimation and actual value based on SVD approach decreases by over 30% than that of the above-mentioned subset method. They further applied SVD on-board modelling to engine performance degradation mitigation control (EPDMC) to reduce thrust response variation under current sensor-based control strategies due to degradation effects [72-74]. In this new control
architecture, as shown in Fig. 9, SVD-based on-boarding model served as an on-board thrust estimator. Thrust estimation concerning degradation effects was calculated by the SVD on-board model. Then the fan speed reference command (N1) was compensated with a delta quantity (\(\Delta N1\)) by the estimated thrust. The results showed that this EPDMC significantly reduces thrust variation from cold-section degradation, hot-section degradation, and various gradual degradation levels in both steady and transient states.

![Fig. 9 EPDMC configuration with an SVD-based on-board model adapted from [74]](image)

Between 2009 and 2012, Simon et al. proposed a systematic on-board modelling approach, optimal tuner Kalman filter (OTKF), for large turbofan to decrease the mean square estimation error from SVD approach [49, 75, 76]. Systematic selection for all health parameters was applied to get a low-dimensional health parameter subset so that the estimation variance is minimum. Simulation results of OTKF showed that it outperforms subset and SVD approach for health parameter estimation and unmeasurable thrust, surge margin, and TET estimation on both linear and nonlinear engine models.

In 2016, Csank and Collony extended OTKF to OTEKF (Optimal Tuner Extended Kalman Filter) to
improve the estimation accuracy for unmeasurable control parameters during rapid transients [77]. OTKF and OTEKF were subsequently applied to form a model-based engine control (MBEC) architecture, as highlighted in Fig. 10 [10, 78, 79]. In MBEC, estimated thrust, surge margin and TET from the on-board model were directly close-loop controlled, as a replacement for engine spool speed loop, acceleration schedule and exhaust gas temperature limit loop in existing sensor-based GTEs control system. Results showed that MBEC outperforms existing sensor-based control for GTEs in view of increasing thrust response and effective TET/surge margin limitation for various degradation levels. However, there is still a noticeable relative estimation error for surge margin, up to 20%, for middle-aged and end-of-life engine in OTKF and OTEKF.

The above-mentioned on-board modelling approaches for large turbofan engines focusing on the underdetermined problem are summarized with their pros and cons in Table 6. The evolution from

Fig. 10 MBEC with OTKF/OTEKF on-board model, adapted from [10, 78, 79]
subset, SVD to OTKF/OTEKF technique witnesses the transition from engineering experience-based method to the more systematic approaches. However, there is still noticeable estimation error for surge margin during rapid transient states even for OTKF/OTEKF method.

Table 6 Comparison of on-board modelling for large turbofan with underdetermined problem

<table>
<thead>
<tr>
<th>Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset [43]</td>
<td>Easy for practical implementation</td>
<td>Non-systematic approach, strong dependence on designer’s experience</td>
</tr>
<tr>
<td>SVD [69, 72-74]</td>
<td>Systematic approach for health parameter selection</td>
<td>Large estimation variance</td>
</tr>
<tr>
<td>OTKF/OTEKF [10, 49, 75-79]</td>
<td>More systematic approach for health parameter selection</td>
<td>Noticeable estimation error for HPC surge margin, up to 20%, during rapid transients</td>
</tr>
</tbody>
</table>

2.2 On-board modelling of gas turbine aero-engines for health management

On-board modelling of GTEs for health management centres on the component performance tracking in conjunction with sensor/actuator fault diagnosis. Component performance tracking, capable of accurately estimating component health status including abrupt fault (e.g. ice and bird damage) and gradual degradation (e.g. erosion, corrosion, fouling, etc.), enables engine performance monitoring and condition-based maintenance decision making [5]. Sensors and actuators diagnosis using on-board models mainly consist of fault detection and reconstruction, which contributes to control system integrity assurance [6].

2.2.1 Component performance tracking

As stated in Section 2.1.2, GTEs component performance in gas path deviates from its nominal
state with the usage or foreign object damage [46, 47, 80]. Component performance deviation is usually represented by the health parameters for compressors and turbines[81], as defined in Eq.(2)-(4). Accurate estimation for these health parameters, known as component performance tracking, has attracted tremendous efforts in the gas turbine community [3, 6, 7]. Reliable component performance tracking could contribute to dependable gas turbine monitoring and prognostics [5, 82].

2.2.1.1 Abrupt fault

Between 2003 and 2008, Brotherton and Volponi et al. from Pratt & Whitney proposed a hybrid on-board model, enhanced Self-Tuning On-board Real-time Model (eSTORM), as further development for STORM [83-86]. eSTORM was mainly designed to deal with the model-plant mismatch. The term, model-plant mismatch, is to describe the phenomenon resulting from the engine hardware modification or engine-to-engine variation. Piecewise linear (PWL) model in STORM cannot solely reflect this mismatch. Thus, a misleading health parameter estimation would probably be generated. eSTORM forms a hybrid model, synthesizing STORM with an empirical model (neural network) to compensate the model-plant mismatch, as shown in Fig. 11. The empirical model was trained to generate the initially estimated residuals due to the model-plant mismatch. Engine parameter residuals were therefore generated through a combination of engine measured outputs, piecewise linear model outputs, and the estimated initial residual from the empirical model. The estimator then calculated a tuner based on engine parameter residuals and fed the tuner back to the PWL model. Therefore, the tuner from the estimator represents component health status as long as the estimated outputs from the hybrid model are closer to the measured engine parameters. Simulations and tests on
PW6000 turbofan engine showed that eSTORM could track the efficiency change in fan, LPC, HPC and HPT to avoid model-plant mismatch issues. However, the training process for the empirical model is much complicated.

In 2018, Daroogheh et al. proposed a dual estimator architecture based on particle filter (PF) for abrupt fault in a single-spool turbojet engine [87]. The dual particle filter strategy was designed for the estimation of both engine states and health parameters, respectively. Simulation results confirmed that this dual PF approach has an average estimation accuracy of 0.3% for compressor fault and 0.5% for turbine fault. However, this dual PF on-board model subjects the extensive computational burden due to the inherent nature behind the particle filter (huge number of particles).
2.2.1.2 Gradual degradation

There are also efforts on gradual degradation tracking with the aid of the prior knowledge on GTEs degradation. Our prior knowledge is that engine health status always degrades, as shown in Table 4, except engine maintenance or overhaul. Simon et al. firstly introduced hard constraints in Kalman filter (KF) for gradual degradation estimation for turbofan engines in 2005 [88]. They augmented quadratic programming as health constraints with standard Kalman filter. It was shown that this Kalman filter with hard constraints has a smaller error variance for health estimation for turbofan than that of unconstrained Kalman filter. Subsequently, Simon further compared unconstrained Kalman filter and Kalman filter with soft/hard constraints for turbofan engine health estimation [89]. Soft constraints were approximated satisfaction of the constraints to smooth the estimation. They found that Kalman filter with hard constraints is superior to soft-constrained KF and unconstrained KF at the cost of increasing computational efforts. The next step was the research by Simon that proposed a constraint switching strategy of KF for component performance tracking of turbofan engines [90]. It was assumed that constrained KF may violate the optimal estimation for standard KF. Therefore, constraint switching logic was applied to find whether the confidence in constrained or unconstrained filters was high.

A more practical consideration in component performance tracking issues is that abrupt fault (e.g. foreign object damage) takes place at a time when the engine is undergoing gradual degradation. Kalman filter shows relatively acceptable accuracy for gradual degradation tracking, but its performance deteriorates for abrupt fault in view of long estimation delay [45]. Besides, there are
smearing effects which means that Kalman filter could probably spread health parameter deviation to other non-degraded components [91]. Therefore, adaptive on-board modelling for abrupt fault is a strong motivation. In 2008, Borguet et al. initially introduced an adaptive extended Kalman filter (EKF) for abrupt fault in turbofan engines [45]. They utilized covariance matching methodology into EKF to address the above-mentioned problem, under which the health parameter variation rate was adaptively released when abrupt fault occurred. It showed that this method takes shorter time for abrupt fault estimation than that of standard EKF. The next footprint was also by Borguet et al. who compared covariance matching method with generalized likelihood ratio in EKF to track abrupt fault plus gradual degradation [42]. They concluded that generalized likelihood ratio is more promising due to its detailed explanation of time and amplitude estimation for abrupt fault. Pu et al. proposed another adaptive on-board model, strong tracking filter (STF), for GTEs abrupt fault [92]. They combined an adaptive fading factor with EKF so that this fading factor automatically increased when abrupt fault occurred to increase the weight of new measurements. During gradual degradation, this fading factor decreased, thus a standard EKF was formed. Results showed that this STF-based on-boarding model could track both gradual degradation and abrupt fault with a particular capability for fast estimation of abrupt fault. However, an evident overshoot for health parameter estimation of turbine flow capacity (two times of the magnitude of implanted abrupt fault) is observed.

In general, the above-mentioned on-board modelling methods in component performance tracking cover the scope of abrupt fault and gradual degradation. eSTORM and dual PF model aim at high-fidelity estimation for abrupt fault. Constrained Kalman filter is based on our prior knowledge to improve the estimation accuracy for gradual degradation. Additionally, adaptive EKF and STF models
are the efforts to shorten the tracking time for abrupt fault that occurs along with gradual degradation.

The advantages and limitations of each modelling approaches are summarized in Table 7.

Table 7 Comparison of on-board modelling approach for component performance tracking

<table>
<thead>
<tr>
<th>Models</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>eSTORM [83-86]</td>
<td>Pioneer research on the compensation for model-plant mismatch</td>
<td>Limited scope to abrupt fault on efficiency health parameters in fan, HPC, HPT, and LPT</td>
</tr>
<tr>
<td>Dual PF [87]</td>
<td>Separate particle filter for engine state and health parameter estimation</td>
<td>Extensive computational burden due to the huge number of particles</td>
</tr>
<tr>
<td>Constrained Kalman filter [88-90]</td>
<td>Using the prior knowledge on health degradation for GTEs to form constraints</td>
<td>Potential violation on the optimal estimation from basic Kalman filter</td>
</tr>
<tr>
<td>Adaptive EKF [42, 45, 91]</td>
<td>Shorter estimation time for abrupt fault compared than standard Kalman filter</td>
<td>Long estimation time, with a time scale of hundreds of seconds</td>
</tr>
<tr>
<td>STF [92]</td>
<td>Combination of a fading factor with EKF for fast estimation for abrupt fault</td>
<td>Large estimation overshoot, i.e. two times of the magnitude of implanted abrupt fault</td>
</tr>
</tbody>
</table>

2.2.2 Sensor/actuator diagnosis

Sensor and actuator may encounter malfunction due to the harsh working environment in GTEs. The diagnosis for sensor and actuator including the detection, isolation, and reconstruction witnesses the transition from nominal engines to degraded engines to enhance the diagnosis effectiveness.

2.2.2.1 Nominal engines

In the 1980s, urgent needs for on-board sensor diagnosis in gas turbine aero-engine also motivated the development of on-board model. Different on-board models for online sensor diagnosis focusing nominal engines were firstly proposed, including parameter synthesis, simplified component-level model and piecewise linear model.

Parameter synthesis (PS) based on-board model was proposed under PW2037 turbofan engine [93]
and digital electronic engine control (DEEC) sensor diagnosis program [94]. The estimation for one measurable output was calculated by an algebraic function from other un-fault sensors. Under DEEC project, if inlet static pressure sensor was failed, its estimation was obtained by the synthesis from burner pressure, high-pressure spool speed, and inlet total temperature. Although this approach is very simple and straightforward, it does not capture the dynamics between these parameters.

Simplified component-level model (SCLM) was developed under the program of fault identification and corrective action (FICA) for F404 turbofan engine [95] and Energy-Efficient Engine (E3) on-board sensor diagnosis [96]. It is a simplified aero-thermal engine model by reducing the intensive computation in the full aero-thermal engine model in the following two parts. One aspect is the substitution for component characteristic maps, a polynomial fitting format with no more than 3rd order. The other side is its permission for single pass in every iteration for mass flow continuity computation, thus no iteration exists in the aero-thermal model to mitigate computational burden. Experiments on F404 engine showed that the steady error of measurable parameters (engine spool speed, pressure and temperature) between SCLM and experiments within the whole envelope is less than 3% despite a transient error within 5%.

Piecewise linear (PWL) model was another on-board model for sensor diagnosis in advanced detection, isolation, and accommodation (ADIA) for sensor diagnosis program [21, 97]. Its detailed principle was described in Section 2.1.1. The shortcoming of PWL model is its poor accuracy in rapid transients.

Overall, the above-mentioned on-board model for sensor diagnosis did not consider component fault and gradual degradation effects, detection threshold has to be set to a very conservative level to
balance the mis-alarm and un-alarm, which may decrease the confidence for sensor diagnosis.

2.2.2.2 Degraded engines

Simplified component-level model (SCLM), combined with a tracking filter was also used in diagnosis purposes for degraded engines. Analytical Redundancy Technology for Engine Reliability Improvement (ARTERI) program pioneered the application of SCLM to sensor and actuator failures in degraded GTEs [98]. To adapt SCLM to the specific engine due to gradual degradation and engine-to-engine variation, a tracking filter was used to update the health parameters for SCLM. The real-time computational ability of SCLM was achieved by the single pass for the iteration loop in aero-thermal equations during each time step. Application to GE23A turbofan engine showed this SCLM updated by the tracking filter could effectively diagnose fan speed sensor fault, fuel flow actuator fault and exhaust nozzle actuator fault. In 2008, Kumar from General Electric (GE) further enhanced SCLM with extended Kalman filter (SCLM+EKF) as an on-board model for engine fault tolerant control [99]. SCLM in their study was achieved by using one Jacobian matrix and single pass technique for the iteration loop in aero-thermal equations. Results showed that their model is helpful for fault detection and accommodation. Kumar also developed an on-board engine model for sensor/actuator/component fault detection and accommodation using real-time SCLM and EKF [100]. They deliberately designed this on-board model to a fixed health baseline, for a half-degraded level within all engine fleet. This setting resulted in a non-zero residual due to the un-match between the on-board models with the engine to be monitored. Therefore, much effort had to be spared on tuning this on-board model to ensure it less sensitive to the engine-model un-match but sensitive to faults.
In 2005, Kobayashi and Simon proposed a hybrid estimator for GTEs component and sensor diagnosis to account for sensor bias in health estimation [101]. The hybrid estimator comprised genetic algorithm and neural network (GA+NN). Neural network is utilized for health parameter estimation, which genetic algorithm is served for sensor detection. However, a bias data set must be designed for sensor bias matching, which highly relies on prior knowledge and requires many efforts. And this study limited its scope to steady engine trending.

From 2003 to 2008, Kobayashi et al. developed a unique Kalman filter, hybrid Kalman filter (HKF), for GTEs on-board modelling focusing on sensor, actuator and component fault diagnosis in conjunction with component gradual degradation [102-105]. This was motivated by the practice that traditional diagnosis system may lose its effectiveness due to the gradual degradation effects since traditional diagnosis system was designed at a fixed baseline for engine health, usually for nominal engines. However, with the usage of GTEs, the health baseline should be updated to the engine to be monitored, so that the diagnosis system could work in the vicinity of the updated health baseline to maintain diagnosis capability. This unique structure consisted of a nonlinear on-board engine model (NOBEM) and a piecewise linear Kalman filter (PWLK). NOBEM was updated to the monitored engine via the estimated health parameters from an off-line trend monitoring system. Simulations on a turbofan engine showed that HKF improves the diagnosis accuracy for sensor, actuator and component fault with the existence of gradual degradation. However, the off-line trending system is extremely hard to implement due to its complexity. Kobayashi subsequently applied HKF to sensor, component and anomaly diagnosis of turbofan engines using dual-channel redundant measurements [106]. This configuration enhances the sensor fault and anomaly detection rate than that of the baseline.
system. The next step on HKF was the study that proposed a multiple model-based HKF (MHKF) on-board model for sensor diagnosis with component degradation in a turbojet engine [107]. They designed a special multiple level detection filters as MHKF. Simulations demonstrated that this MHKF outperforms linear Kalman filter, unscented Kalman filter and extended Kalman filter in view of sensor fault detection time and computational time.

In 2011, Zhang et al. proposed an adaptive real-time reference engine model (ARREM) for sensor, actuator, and component fault diagnosis under gradual degradation [108]. This model consisted of a nominal engine model and neural network element to adapt the engine model to the specific engine due to gradual degradation and engine-to-engine variations. The nominal engine model was updated post-flight by the neural network that was trained on-line with the monitored engine. A unified diagnosis framework based on this model shows its superiority in view of less false alarm and early detection time.

In general, the above-mentioned on-board modelling methods for sensor and actuator diagnosis make the leap from nominal engines (PS, SCLM, and PWL model) to degraded engines (SCLM+EKF, GA+NN, HKF, and ARREM) in order to avoid the false alarm and mis-alarm due to the degradation effects. The pros and cons of each model are summarized in detail in Table 8.
Table 8 Comparison of on-board models for sensor and actuator diagnosis

<table>
<thead>
<tr>
<th>Models</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS [93, 94]</td>
<td>Only algebraic computation, very easy for implementation</td>
<td>No dynamic relationship</td>
</tr>
<tr>
<td>SCLM [95, 96]</td>
<td>Simplified aero-thermal engine model by map modification and no iteration</td>
<td>5% error for measurable spool speed, pressure, and temperature in rapid transients</td>
</tr>
<tr>
<td>PWL [21, 97]</td>
<td>Only the need for interpolation and integration</td>
<td>Poor accuracy in rapid transients</td>
</tr>
<tr>
<td>SCLM+EKF [99, 100]</td>
<td>Nonlinear estimation capacity from EKF</td>
<td>Much efforts required in on-board model tuning due to its fixed health baseline</td>
</tr>
<tr>
<td>GA+NN [101]</td>
<td>Integration from genetic algorithm for health estimation and neural network for sensor diagnosis</td>
<td>Only steady engine trending focused, no consideration for engine transients</td>
</tr>
<tr>
<td>HKF [102, 103, 105-107, 109]</td>
<td>Health baseline update of the on-board model to the monitored engine</td>
<td>Complicated health baseline update from off-line trending system</td>
</tr>
<tr>
<td>ARREM [108]</td>
<td>Neural network for model-plant mismatch compensation</td>
<td>Nominal engine model update highly relies on the degradation trending from neural network</td>
</tr>
</tbody>
</table>

3 The research challenges for on-board modelling of gas turbine aero-engines

Based on the historical review of on-board modelling in GTEs for control and health management purposes, on-board modelling techniques in both applications have evolved from nominal engine to degraded engine, including abrupt fault and gradual degradation, in order to provide the health awareness of the monitored engine. Although there is significant progress in on-board modelling methods for degraded engines with application to control and health management that mainly utilize Kalman filter and its variants, the remaining research questions still need to be addressed to enable this promising approach to be adopted by GTEs designers and manufacturers. So, by systematic examination of the critical areas in the above-discussed sections, the two remaining research
challenges are identified.

3.1 High-fidelity on-board modelling of safety-critical parameters during rapid transients over GTEs life cycle

High fidelity on-board modelling for control-oriented safety-critical parameters in gas turbine aero-engines, especially the unmeasurable thrust, surge margin (SM) and turbine entry temperature (TET), during rapid transient states at different degraded levels is still highly challenging. Recent advancement regarding this issue, OTKF on-board modelling for GTEs from NASA, still showed a relatively undesirable accuracy of surge margin of the high-pressure compressor and TET estimation (up to 20% estimation errors) during idle to take-off rapid transients for new, middle-aged and end-of-life engines [10, 78]. This difficulties in this problem exist in two parts:

- One aspect is that GTEs in rapid transients show strong nonlinearity and large deviation of operating points from the steady operating line;
- The other point is that both the engine steady and transient operating lines vary from their nominal situation when different degradation levels take place.

However, high fidelity on-board modelling for control purposes is very promising for active safety considerations and fuel consumption improvement. One benefit is that it has been demonstrated that using on-board models in the feedback control loops in GTEs could significantly eliminate the thrust response variation for engines undergoing different degraded levels, compared to the thrust response under current industrial sensor-based control strategies [78]. The other aspect is that model-based engine control (MBEC) using on-board models is capable of controlling the surge margin and TET
directly. So, engines could be designed to avoid excessive safety margin. Experiments adopted by NASA on a high-fidelity aero-thermal model of a large turbofan engine with 178 kN (40,000 pounds) thrust, Commercial Modular Aero-Propulsion System Simulation (C-MAPSS40k), showed that a 13% reduction in surge margin threshold contributes to the benefit of a 1% reduction in fuel consumption [79]. This is accomplished by re-designing the engine with a lower surge margin limit so that the compressor operating line runs within a much higher efficiency zone. The only cost behind this benefit is the development of a high-fidelity on-board model, whereas no highly challenging marginal improvement in component design is required.

### 3.2 Dependability and reliability for on-board models of gas turbine aero-engines

Another challenge from on-board modelling of GTEs is its dependability and reliability. Specifically, on-board models generating unmeasurable safety-critical parameters including thrust, surge margin, and TET serve in the control loops. Despite the accuracy of the on-board model could be validated off-line with experiments or the aero-thermal engine model, the on-board model with intelligent and non-deterministic feature is still subject to the difficulty for online application. If estimation values from on-board models went wrong, the whole control system would probably drive the engine to the unpredicted, even unsafe operating region. Moreover, especially for civil aero-engines, the application of on-board models has to pass the strict certification regulations (e.g. Federal Aviation Regulations Part-33 (FAR-33) from Federal Aviation Administration (FAA) and Certification Specifications - Engines (CS-E) from European Union Aviation Safety Agency (EASA)) to guarantee
the safety operation of the engine even if a failure from the on-board model takes place [110, 111]. Therefore, ensuring the dependability and reliability of the on-board models is the prerequisite for its real usage in GTEs.

4 Exploring the future of on-board modelling in gas turbine aero-engines application

For the future exploration of on-board modelling in GTEs application especially for control purposes, the potential solutions with regard to the identified research challenges as well as future aero-propulsion requirements are proposed.

4.1 Multiple model based high-fidelity on-board modelling

As the research challenges identified in Section 3.1, almost all the on-board modelling techniques are under one same model framework. Although on-board models under Kalman filter theory, as the most common practice, shows favourable accuracy for some parameters, estimations of other safety-critical parameters including surge margin and TET are not acceptable. Therefore, high fidelity on-board modelling of safety-critical parameters for all degraded levels are still highly challenging. Moreover, government and organizations have set ambitious targets for emissions and noise reduction. In ACARE Flightpath 2050 plan, the challenging goals are reductions of 75% in CO2, 90% in NOx, and 65% in noise that are compared to the capabilities of the typical new aircraft in 2000 [14]. This could not be achieved only by component performance improvement. If the emission and noise reduction related parameters could be covered in the on-board models, the advanced control system
may help to step forward the challenging goals.

One potential approach for dealing with the high-fidelity on-board modelling is multiple model based on-board modelling (MMOBM). Under this modelling framework, different parameters could be modelled by different modelling techniques to get the maximum accuracy for the whole system. This is based on the fact that different modelling approaches own their advantages and shortcomings. If the best candidate model for each parameter of interest could be found, including safety-critical parameters (thrust, surge margin, and TET) and emission-related values, an integrated on-board model could be obtained in the view of overall optimization. An MMOBM for advanced control system focusing emission reduction for future aero-propulsion system is proposed as Fig. 12 for civil turbofan engines [36]. Within this MMOBM, different on-board models are systematically selected from a family of candidate models by means of minimum Normalized Root Mean Square Error (NRMSE). Parameters of interest, including low-pressure spool speed (N1), high-pressure spool speed (N2) and compressor discharge pressure (Ps3) are modelled by polynomial NARX. Meanwhile, emission-related values including CO, NOx, and smoke number (SN) are modelled by ANFIS, NARX and Hammerstein-Wiener models. The training data for different models was generated by a well-defined excitation input signal, Amplitude Modulated Pseudo Random Signal (APRBS), to excite the full engine dynamics. Simulation results on a verified CFM56-5A turbofan engine aero-thermal model demonstrated that all the best models for different parameters, which are underlined in Table 9, are successfully obtained by the criterion of minimum normalized root mean square error. Moreover, an advanced control system based on this MMOBM demonstrates its effectiveness and superiority in view of overall high-fidelity on-board modelling.
Fig. 12 Multiple model-based high-fidelity on-board modelling for GTEs emission reduction adapted from [36]

Table 9 MMOBM comparisons by the criterion of normalized root mean square error [36]

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>N2</th>
<th>P3</th>
<th>NOx</th>
<th>CO</th>
<th>SN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NARX</td>
<td>0.00014</td>
<td>0.000112</td>
<td>0.000147</td>
<td>0.000153</td>
<td>0.0434</td>
<td>0.0071</td>
</tr>
<tr>
<td>Polynomial NARX</td>
<td><strong>0.00002</strong></td>
<td>0.000011</td>
<td><strong>0.000062</strong></td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Hammerstein-Wiener</td>
<td>0.00026</td>
<td>0.000691</td>
<td>0.000311</td>
<td>0.000302</td>
<td>0.0147</td>
<td><strong>0.000197</strong></td>
</tr>
<tr>
<td>ANFIS</td>
<td>0.00017</td>
<td>0.000129</td>
<td>0.000128</td>
<td>0.000196</td>
<td><strong>0.0005</strong></td>
<td>0.000255</td>
</tr>
</tbody>
</table>

N/A: Not applicable

4.2 Runtime assurance for on-board model dependability and reliability

As stated in Sec 3.2, on-board modelling of GTEs is facing the problem of dependability and reliability. This is a practical obstacle for its application in gas turbine aero-engines. Additionally, future aero-propulsion systems (e.g. Ultra High Bypass Ratio (UHBR) engines, Geared Turbofans (GTF), and Hybrid Electric Propulsion), has regarded model based control that highly relies on on-
board modelling as a potential candidate solution for challenging goals of emission, noise, and fuel consumption reduction [15]. Therefore, a continuous monitoring system for the on-board model of civil aero-engines is required for practical point of view and to pass the strict certification regulations.

One potential solution regarding the challenge of on-board model dependability and reliability is runtime assurance (RTA) monitoring system. RTA monitoring system consists of monitor/switch logic and backup controller in conjunction with the advanced controller using on-board modelling, as shown in Fig. 13. This monitoring system continuously monitors the outputs of the on-board model to ensure that the engine operates within its safety boundaries (e.g. over-speed, over-temperature, and over-pressure). When outputs from on-board model went untrusted beyond a pre-defined threshold that could be defined as the residual of outputs between the advanced controller and the backup controller, then the switch logic would activate the backup controller (usually trusted and certified N1/EPR controller) to limit the engine in a more conservative operating area. Therefore, the dependability and reliability of on-board model in GTEs could be fulfilled.

![Fig. 13 Runtime assurance for on-board modelling in GTEs, adapted from [112]](image)

RTA was originally applied in general software to observe the execution of a running system [113].
Then, it is extended to auto-landing flight control system that comprises adaptive and other advanced controllers [114]. National Aeronautics and Space Administration (NASA) shows great interest in RTA for the dependability and reliability of advanced propulsion control system using on-board models [112, 115]. However, only very simple switch logic between the advanced controller and backup controller was designed. Therefore, RTA for advanced control systems with on-board models for future propulsion needs is still an area of ongoing investigation.

It could be concluded that multiple model based on-board modelling and runtime assurance could be applied as the potential solutions for future research challenges of on-board modelling in gas turbine aero-engines. Table 10 summarizes the research challenges and potential solutions in detail.

Table 10 Future challenges of on-board modelling for GTEs with proposed potential solutions

<table>
<thead>
<tr>
<th>Future challenges</th>
<th>Caused by</th>
<th>Potential solution</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-fidelity on-board modelling of safety-critical parameters during rapid transients over GTEs life cycle</td>
<td>Limitations of each specific modelling approach Challenging emission target by government and organization (e.g. ACARE Flightpath 2050)</td>
<td>Multiple model based on board modelling</td>
<td>Fig. , Table , [36]</td>
</tr>
<tr>
<td>Dependability and reliability of the on-board model</td>
<td>Certification requirement for future propulsion (e.g. UHBR, GTF, and HEP) using on-board modelling in advanced control system</td>
<td>Runtime assurance monitoring system</td>
<td>Fig. , [112, 113, 115]</td>
</tr>
</tbody>
</table>

5 Conclusions

Gas turbine aero-engines real time on-board modelling for both areas, control and health management, was comprehensively reviewed for the past four decades. Since the first on-board
modelling application to GTEs in 1977, on-board modelling techniques in both areas were systematically examined with their pros and cons via analysis and simulation results. The on-board modelling methods in both applications have witnessed the transition from nominal engines to degraded engines (under abrupt fault and gradual degradation) to provide the health status perception of the particular monitored engine. Although there is significant progress in on-board modelling for degraded engines, the remaining research challenges are still successfully identified. These research challenges include:

- High-fidelity on-board modelling of safety-critical parameters during rapid transient states over GTEs life cycle;
- Dependability and reliability for on-board models of GTEs.

If these research challenges could be overcome, advanced personalized control strategy with self-adaptation capability to the particular monitored engine would be expected to enhance the engine performance as well as meet the challenging emission and fuel consumption targets for future aero-propulsion systems. The potential solutions to the identified challenges in view of future challenging aero-propulsion requirements are therefore proposed:

- To use multiple model-based on-board modelling approach to obtain a high-fidelity on-board model;
- To use runtime assurance monitoring system to ensure the dependability and reliability of the on-board model in case that a failure from the on-board model takes place.

The effectiveness of the proposed solutions is confirmed by simulation cases and analysis. This allows researchers, engine designers, and manufacturers to understand how the on-board modelling
approach is significantly promising to the GTEs performance improvement and future aero-propulsion requirements.

Acknowledgement

This research is funded by China Scholarship Council (201906020115).

Reference


Gas turbine aero-engines real time on-board modelling: A review, research challenges, and exploring the future

Wei, Zhiyuan

Elsevier

https://doi.org/10.1016/j.paerosci.2020.100693
Downloaded from Cranfield Library Services E-Repository