

Turbine Engine Performance Estimation Using Particle Filters

Bong-Jun Yang^{*}, Prasenjit Sengupta[†], P. K. Menon[‡]
Optimal Synthesis Inc., Los Altos, CA, 94022

A nonlinear, non-Gaussian Particle Filter is considered for engine health parameter estimation. The algorithm employs a high-fidelity full engine model as its central element to overcome the performance limitations imposed by the Gaussian noise-linear dynamics assumptions required in the Kalman filter formulation of the problem. A central feature of the present estimation problem is that the number of engine health parameters to be estimated is often greater than the number of available sensor measurements. This renders the linearized engine dynamics not fully observable. However, using an analysis of the high-fidelity engine model, it is shown that the system may be fully observable in a nonlinear sense. It is also shown that the system observability can be enhanced by using specific inputs. Ensuing particle filter implementation demonstrates that the number of parameters that need to be estimated by the particle filter can be greater than the number of available measurements, implying that nonlinear filters can overcome the non-determinism imposed by linear Kalman filters.

I. Introduction

The use of an engine model for the estimation of engine performance has been a key concept in next-generation engine control and health management¹⁻⁶. Since thrust cannot be directly measured, an engine model is used to compute the thrust from available measurements. However, for this process to work accurately under varying ambient conditions and deteriorating engine components an estimation process has to be employed to ensure smooth, stall free operation of the engine. Additionally, the estimated parameters can be used for health management, engine performance trend monitoring and gas path fault diagnostics. As a consequence, accurate assessment of the performance deterioration of engine components over engine lifetime and accurately detecting the malfunctioning of engine systems that manifest as a deviation from reference engine responses are expected to play a central role in effective engine diagnostics and maintenance.

The left portion of Figure 1 depicts a general architecture for engine performance estimation.

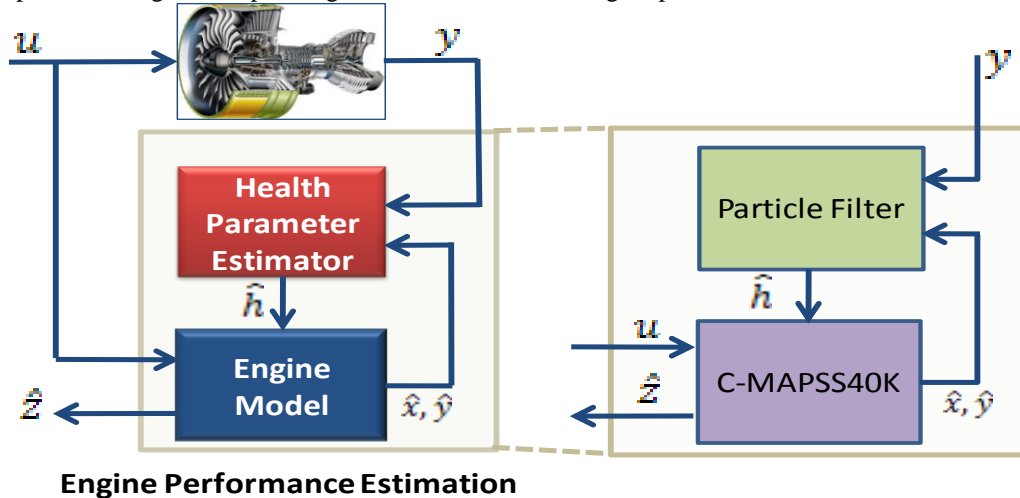


Figure 1. Engine Performance Estimation Architecture for Particle Filtering

^{*} Senior Research Scientist, 95 First Street, jun.yang@optisyn.com, Senior Member AIAA.

[†] Research Scientist, 95 First Street, sengupta@optisyn.com, Senior Member AIAA.

[‡] Chief Scientist, 95 First Street, menon@optisyn.com, Fellow AIAA.

The variable u denotes actuator commands, y denotes the measurements available via engine instrumentation such as pressures, fuel flows, temperatures, and rotor speeds, h denotes an engine health parameter vector associated with efficiencies and capacities of fans, low-pressure compressor (LPC), high-pressure compressor (HPC), low-pressure turbine (LPT), and high-pressure turbine (HPT). With an engine model employed, engine performance estimation is carried out iteratively. The employed engine model generates the expected sensor measurement \hat{y} and the state estimate \hat{x} which, together with y , is used to update the estimate of the engine state and the health parameters. The estimated health parameter vector \hat{h} is again used to update the expected sensor measurement \hat{y} and the expected engine performance \hat{z} that can be obtained as a function of the engine state and the engine health parameters.

The particle filter approach addressed in this paper is motivated by the facts (1) that most existing estimation methods employ piecewise linear engine models and (2) that the use of linearized models are necessitated due to the use of Kalman filters (or their variants) that places restrictions on system models and noise characteristics. Kalman filters (and their variants) require that the process noise and the sensor noise be Gaussian and that the underlying system be linear. While these assumptions lead to an analytical, elegant filter solution for linear systems, a fundamental question that arises in application to nonlinear systems, such as aircraft turbofan engines, is what undesirable effects are introduced because of the requirement of linear system models with Gaussian noise. Following are the technical issues related to estimation approaches based on linear systems for engine health monitoring.

Firstly, in order to approximate a nonlinear system by a series of linear systems, the following steps should be employed.

- A set of trim points must be obtained by trim-solving routines and stored.
- System matrices with respect to the above trim points should be derived and stored.
- A set of schedule parameters, such as Mach number, altitude, engine health parameters, should be selected and tabularized.
- Between trim points, the system matrices should be interpolated. Beyond the trim points, the system matrices should be extrapolated, however, in this case, the system responses obtained by extrapolation can significantly deviate from the true responses of the engine⁶.

Secondly, a linearized system may not be observable while the original nonlinear dynamic system is observable. A fundamental problem in Kalman filter-based health parameter estimation is that the problem is under-determined because the number of sensor measurements is less than the number of parameters that reflect the deterioration of engine performance. This fact may make the linearized system unobservable. In this case, the culprit for this loss of observability may be the linearization process. It is possible that the system is observable only through the nonlinear components of the model, which have been eliminated by the linearization process. The effect of linearization is also present in assessing a nonlinear performance variable. For example, stall margins are inherently nonlinear, and the linearized systems tend to exhibit mismatch with a nonlinear high-fidelity mode⁶.

Thirdly, linear systems cannot model inherently nonlinear phenomenon such as physical bounds on the state and the health parameters, such as limits on maximum fan and core speeds, limits on turbine blade temperatures because of structural strength, limits on pressure to compressor and fans due to stall, and limits on maximum turbine inlet temperature for the duration of engine life. There are approaches that handle state constraints in Kalman filtering⁷, but they tend to be much more complex than conventional Kalman filters.

Finally, Gaussian noise assumption can be restrictive. It has been shown that when there are outliers in the measurements, the fault detection algorithm optimized for Gaussian noise can exhibit significant performance degradation⁸⁻⁹. Moreover, the Gaussian noise model is not appropriate for defining random variables that have hard physical bounds.

As an approach to overcome the limitations posed by Kalman filters, Particle filters (PFs)^{10,11} are employed, in which a high-fidelity nonlinear engine model is incorporated without any modification or linearization. Specifically, it is proposed to use the NASA (National Aeronautics and Space Administration) C-MAPSS40K (Commercial Modular Aero-Propulsion System Simulation for 40,000 lb class thrust engine) in the estimation scheme as shown Figure 1. The key idea of the PF is to represent the system uncertainty distributions using a cloud of particles instead of the state estimate and error covariance matrix employed in a Kalman filter and its variants, such as an Extended Kalman Filter (EKF) and an Unscented Kalman Filter (UKF). The PF formulation has its basis in the Monte Carlo simulation technique. However the PF incorporates a specific process called resampling, in order to choose more important high-weighted particles and to discard lower weighted particles. Although the approach taken by the PF does not lead to an analytical solution, it imposes no restrictive assumptions on the system dynamics or statistical distributions of the uncertainties. Consequently, it can provide excellent results in highly nonlinear dynamic systems

with non-Gaussian noise sources. In particular, the PF can be much more effective for a system whose uncertainty distribution is multi-modal, i.e., the probability distribution has more than one peak, or in heavy-tailed distributions¹². More traditional filters like the EKF and the UKF¹³ which approximate all uncertainties as a unimodal Gaussian distribution functions will perform poorly in such cases.

This paper presents the observability analysis of C-MAPSS40K engine model with a set of sensor measurements and show that the system is indeed observable in the nonlinear sense while the linearization around an engine trim condition leads to an unobservable system. Moreover, it is shown that PF-based nonlinear estimation also allows for the estimation of more health parameters the number of which is larger than that of available measurements. The paper is organized as follows. The engine performance estimation problem is formulated as the estimation problem for engine health parameters in Section II. An observability analysis both in linear and nonlinear senses is presented in Section III. The nonlinear observability forms the basis for PF-based nonlinear filtering described in Section IV that also includes an example that illustrates the overall concept. The performance of particle filtering with the NASA C-MAPSS40K engine model is presented in Section V. Conclusions are given in Section VI.

II. Problem Formulation

Figure 2 shows the engine schematic with flow labels in Reference 14.

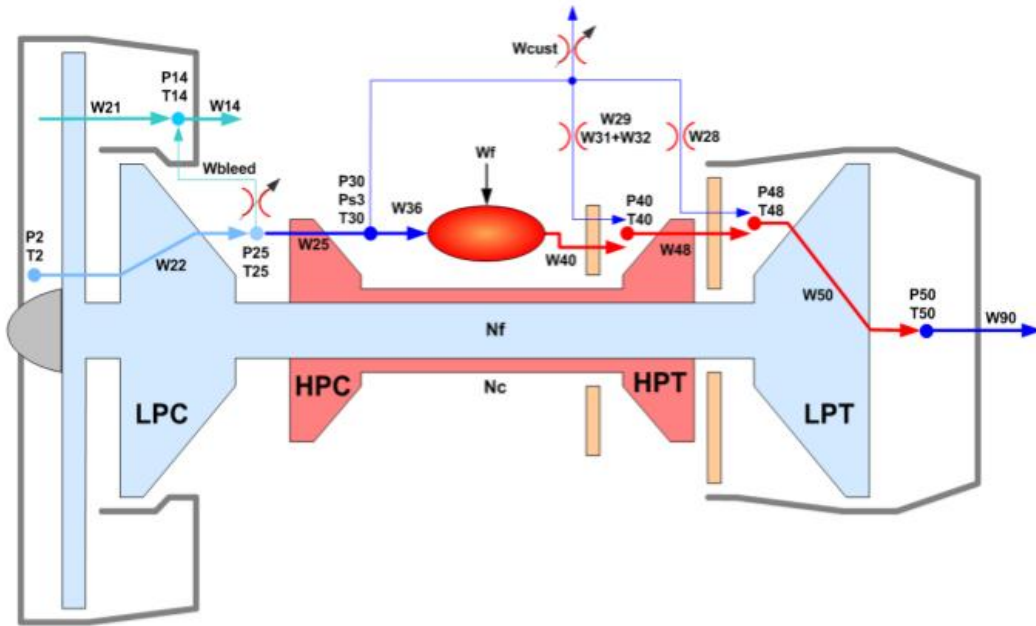


Figure 2. Twin Spool Engine Schematic with Flow Labels¹⁴

The C-MAPSS40K engine model employs 7 states (with the option for including or excluding temperature variables) and 13 health parameters. However, following the problem formulation in Reference 4, the estimation problem for 2 states and 10 health parameters are considered as a baseline problem, which exclude the pressure ratios for Fan, LPC, and HPC out of 13 health parameters. The system states, health parameters to be estimated, and the available sensor measurements considered in the estimation problem are given in Table 1.

Given the state variable and the health parameters in Table 1, the engine dynamics are governed by general equations of the form:

$$\begin{aligned}
 \dot{x} &= f(x, u, h, \mu) \\
 y &= g(x, u, h, \mu) \\
 z &= g_z(x, u, h, \mu)
 \end{aligned} \tag{1}$$

where $x = [N_c, N_f]^T \in \mathbb{R}^2$ represents the engine state vector. The vector $y = [T_{25}, T_{50}, P_{50}, N_f, N_c, P_{s3}]^T \in \mathbb{R}^6$ is the vector of measurements, and z is the vector of unmeasured performance variables such as stall margins and the thrust. The vector $h = [\eta_{HPT}, \gamma_{HPT}, \eta_{LPT}, \gamma_{LPT}, \eta_{FAN}, \gamma_{FAN}, \eta_{HPC}, \gamma_{HPC}, \eta_{LPC}, \gamma_{LPC}]^T \in \mathbb{R}^{10}$ denotes the engine health

parameters shown in Table 1. The functions $f(\cdot)$ and $g(\cdot)$ are characterized by C-MAPSS40K engine model, and μ denotes the additional parameters that are required to run C-MAPSS40K engine model. More specifically, they denote engine operating condition variables such as ambient temperature, Mach number, the altitude, power options, target Engine Pressure Ratio (EPR). Finally, the vector u denotes the control variables of fuel flow, Variable Stator Vane (VSV) position, and Variable Bleed Valve (VBV) angle.

In the engine performance estimation problem, the input parameter vector h is treated as a system state, which is constant vector, and the model employed in the estimator development has the form:

$$\begin{aligned}\dot{x}_a &= f_a(x_a, u, \mu) \\ y &= g(x_a, u, \mu)\end{aligned}\quad (2)$$

where $x_a = [x, h]^T \in \mathbb{R}^{12}$ and $f_a(x_a, u, \mu) = [f(x_a, u, \mu), 0]^T$. The control signals u and the additional function parameters μ are treated as known variables. Then the questions addressed in the paper are posed as follows. In linear Kalman filtering, estimating the health parameters whose number is larger than the dimension of y is infeasible because linearization leads to unobservable system. Is C-MAPSS40K engine model is observable in a nonlinear sense? If the engine model is nonlinearly observable, can a PF, which is a nonlinear estimator, improve the estimation performance?

Table 1. States, Health Parameters, and Measurements of the CMAPSS-40K.

States (x)	Health Parameters (h)	Measurements (y)
N_f , fan speed	η_{HPT} , HPT Efficiency	T_{25} , LPC outlet temperature
N_c , compressor speed	γ_{HPT} , HPT Capacity	T_{50} , exhaust temperature
	η_{LPT} , LPT Efficiency	P_{50} , turbine outlet pressure
	γ_{LPT} , LPT Capacity	N_f , fan speed
	η_{FAN} , Fan Flow Efficiency	N_c , compressor speed
	γ_{FAN} , Fan Flow Capacity	P_{53} , burner inlet static pressure
	η_{HPC} , HPC Efficiency	
	γ_{HPC} , HPC Capacity	
	η_{LPC} , LPC Efficiency	
	γ_{LPC} , LPC Capacity	

III. Observability Analysis

Observability analysis can help determine if the estimates of the system states and parameters can be derived from the given the initial conditions, inputs, system dynamics and the output equations, and the measurements. This section examines the observability of the engine states and parameters from a given set of measurements. In the following, the symbol μ in Equation (1) is dropped for the simplicity of presentation. As it constitutes the known variables, it does not affect any analysis presented below.

A. Observability Rank Test

When the system in Equation (1) is linearized with respect to a trim condition at (x_0, u_0) , the resulting linearized system is given by:

$$\begin{aligned}\dot{x} &= A x + B u + L h \\ \dot{h} &= 0 \\ y &= C x + D u + M h\end{aligned}\quad (3)$$

where $A = \partial f / \partial x |_{x_0, u_0}$, $B = \partial f / \partial u |_{x_0, u_0}$, $L = \partial f / \partial h |_{x_0, u_0}$, $C = \partial g / \partial x |_{x_0, u_0}$, $D = \partial g / \partial u |_{x_0, u_0}$, and $M = \partial g / \partial h |_{x_0, u_0}$, which further leads to the following augmented linearized system:

$$\begin{aligned}\dot{x}_a &= A_a x_a + B u \\ y &= C_a x_a + D u\end{aligned}\quad (4)$$

where

$$A_a = \begin{bmatrix} A & L \\ 0 & 0 \end{bmatrix}, C_a = [C \quad M].$$

Then, the linearized system is observable if and only if the following observability matrix¹⁵ :

$$\mathcal{O}(C_a, A_a) = \begin{bmatrix} C_a^\top & A_a^\top C_a^\top & \dots & (A_a^{(n-1)})^\top \end{bmatrix}^\top$$

has the full rank, where n denotes the dimension of the augmented state x_a .

A standard result in Kalman filtering is that when the system in Equation (2) is linearized, the number of health parameters that can be estimated is limited to the number of sensors, the dimension of y ¹⁶, 4. This renders the engine performance estimation problem unobservable because the dimension of the health parameter vector is usually larger than that of the measurement.

In nonlinear systems, there have been two similar concepts that are related to the estimation of parametric constants and system states. They are observability and identifiability¹⁷. Whereas the observability denotes the feasibility of deducing the states of the system from given input-output behavior, the identifiability represents the feasibility of estimating the parameters of the system uniquely from inputs, outputs, and their time derivatives. Necessary and sufficient conditions for the identifiability for nonlinear systems have been established in References 18 and 19. The relationship between the identifiability and observability is addressed in Reference 20. For observability tests, there have been differential geometric approaches²² and algebraic approaches²⁰, and their equivalence has been established under certain regularity assumptions²¹.

One distinguishing feature of nonlinear observability analysis is that, unlike the observability test for linearized dynamic system, the control signals play a significant role by providing the necessary persistency of excitation. Another difference is that whereas the linear observability considers a linear system in the neighborhood of a trim condition, and therefore the linear observability holds in the neighborhood of a trim condition, the nonlinear observability is considered in the neighborhood of any point. The local observability of the system in Equation (2) is guaranteed if

$$\text{rank} \left\{ dg_i(x_a, u), \dots, d \left(\mathcal{L}_{f_a}^{k_i-1} g_i \right) (x_a, u) \mid 1 \leq i \leq s \right\} = n \quad (5)$$

for a given state and control (x_a, u) , where s is the dimension of y , i.e., $s = 6$ in the estimation problem in Section II. The operator d denotes the differential, \mathcal{L} denotes Lie derivative, and k_i is the observability index for i th measurement²². In other words, if the observability rank condition in Equation (5) holds, the system state is observable in the neighborhood of x_a with the given control u . If the rank condition holds for any $x_a \in \mathbb{R}^n$, then the system is globally observable with the given control u . It is well known that observability of the linearized system is not guaranteed by the observability of the original nonlinear system, since the linearized systems only consider the slope of the outputs in relation to the system state.

B. Observability Test with C-MAPSS40K Engine Model

1. Computation of Differentials

Suppose that $y = g(x_a)$ denotes a scalar output function. Then, the differential dg is numerically obtained by:

$$dg_i = \frac{\partial g}{\partial x_{a_i}} \approx \frac{g(\dots, x_{a_i} + \varepsilon, \dots) - g(\dots, x_{a_i} - \varepsilon, \dots)}{2\varepsilon}, \quad i = 1, \dots, n. \quad (6)$$

The next step in Equation (5) is to compute $d(\dot{y}) = d(\mathcal{L}_{f_a}^1 g)$. It can be seen that

$$\mathcal{L}_{f_a}^1 g(x_a^0) = \frac{\partial g}{\partial x_a} f_a |_{x_a=x_a^0} = \frac{\partial g}{\partial x} f |_{x_a=x_a^0} + \frac{\partial g}{\partial h} \cdot 0 = \frac{\partial g}{\partial x}(x_a^0) f(x_a^0) \quad (7)$$

That is, for any $x_a^0 = (x^0, h^0)$, $\dot{y}(x_a^0)$ can be computed using Equation (7), which can be numerically computed by Equation (6). Subsequently, $d(\mathcal{L}_{f_a}^1 g(x_a^0))$ is again obtained by applying the procedure in Equation (6) with g_i be replaced by $(\mathcal{L}_{f_a}^1 g(x_a^0))_i$. This process can be repeated for higher order differentials.

2. Observability Test

The observability test is performed for the condition with the altitude of 791 ft, Mach number of 0.4, and the power level angle of 78⁰. The engine is in 30% deteriorated condition from new engine condition. For the linearized dynamics, a trim solution is first obtained, and then system matrices A_a and C_a are derived. The observability test for

(A_a, C_a) leads to the rank 8 for the observability matrix, which confirms that for the linearized system, only the same number of health parameters as the measured outputs can be estimated, as noted in Reference 4. The rank 8 of the observability matrix implies that the 2-dimensional state vector and 6 health parameters can be estimated. The nonlinear observability is tested by investigating the rank of $\mathcal{O}_N = [d(y), d(\dot{y})]^T$. As the nonlinear observability is the local property of a point (x_a, u) , it is tested for a few points.

Figure 3 shows the results of the observability tests both in the linear (with respect to the trim condition) and in the nonlinear sense (in the neighborhood of a certain point). It may be noted that the system is observable in the nonlinear sense, implying that the linearization may be contributing to decreased observability. The nonlinear test result also shows the condition number of \mathcal{O}_N .

```
>> checkLinearObservability
rank of observability

ans =

     8

>> checkObservability
rank of gradients of y and ydot

ans =

    12

condition number

ans =

    197637.92336274
```

Figure 3. Linear and Nonlinear Observability Tests

IV. Particle Filter Description

Particle filters belong to the class of filters known as nonparametric filters. They do not rely on a fixed functional form of the posterior[§] such as Gaussians. Instead, they approximate posteriors by a finite number of values (cloud of particles), each roughly corresponding to a region in the state space. As the number of particles goes to infinity, the particles tend to converge uniformly to the correct posterior under certain smoothness assumptions. Particle filters do not make strong parametric assumptions on the posterior density and thus are well suited for representing complex, multimodal probability density functions. However, the representational power of these techniques comes at the cost of added computational complexity. Since the PF is a realization of Bayes filtering using a cloud of particles¹⁰, the following subsection will briefly review general Bayes filtering approach. Note the classical Kalman filter is a special case of the Bayes Filter when the system dynamics is linear and the noise components are Gaussian.

A. Bayes Filter

Consider the following discrete dynamic system:

$$\begin{aligned} x_k &= f(x_{k-1}, v_{k-1}) \\ y_k &= g(x_k, w_k) \end{aligned} \quad (8)$$

where x_k denotes the discrete system state. With slight abuse of notation, f and g still represent the discrete system function and output function in this section. The process noise v_{k-1} and w_k are also added in the filter formulation. It is assumed that their probability distributions are known. Let the accumulation of measurements up to the current time step k be denoted by Y_k , i.e., $Y_k := \{y_i | i = 1, 2, \dots, k\}$. Then, the objective of filtering is to derive $p(x_k | Y_k)$, the probability density function of the state x_k conditioned on the accumulated measurements up to the k^{th} time step. Three following major steps constitute the Bayes filter.

1. The initial probability distribution $p(x_0) = p(x_0 | Y_0)$ is assumed to be known.

[§] Posterior denotes a representation of the distribution of the estimate after including the measurement.

2. Given the known initial distribution for the state vector, subsequent probability distributions are obtained recursively. Given a previous probability distribution $p(x_{k-1}|Y_{k-1})$, the posterior distribution $p(x_k|Y_k)$ is obtained by the following step.
 - a. Prediction

$$p(x_k|Y_{k-1}) = \int_{x_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1})dx_{k-1} \quad (9)$$

where $p(x_k|x_{k-1})$ is the transitional probability density that is governed by the time marching function $x_k = f(x_{k-1}, v_{k-1})$ in Equation(8).

- b. Update

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(y_k|Y_k)} \quad (10)$$

where $p(y_k|x_k)$ denotes the probability distribution of the output y_k conditioned on x_k and is governed by the measurement function $y_k = g(x_k, w_k)$ in Equation(8).

B. Description of the Particle Filter Algorithm

As noted before, the PF is a numerical implementation of the Bayes filter for nonlinear systems with non-Gaussian noises using the idea of Monte Carlo simulation to create a large number of samples for the state variable (particles) from the specified distributions. That is, $p(x_k|Y_k) \approx \sum_{i=1}^{N_p} w_k^i \delta(x_k - x^i)$, where N_p is the total number of particles, and $\delta(\cdot)$ is the Dirac-Delta function. The first step in the PF algorithm is to initialize a set of particles according to the a priori distribution of the state at the first time step.

$$\{ \hat{x}_0^i \}_{i=1}^{N_p} \quad (11)$$

Here \hat{x}_0^i denotes the state of the i th particle at the first time step. The weight of each particle is denoted by ω_k^i and the initial weights for all particles are assumed to be equal:

$$\omega_0^i = \frac{1}{N_p}, \quad i = 1, \dots, N_p \quad (12)$$

Then the following loop, consisting of Equations (13) through (25), is recursively executed from the initial time to the final time step. Once the measurement y_k is available at the current time, the weight ω_k^i for the i th particle can be calculated as follows:

$$\{ \omega_k^i = p(y_k|\hat{x}_k^i)\omega_{k-1}^i \}_{i=1}^{N_p} \quad (13)$$

Here, the particle weight denotes the probability that the particle generates the observed measurement. Note that the measurement noise can be added to the particle measurement if a parametric form of the measurement uncertainty is not available. For example, this may be the case when the statistics of the measurement uncertainty depend on the state, or the measurement model is multiplicative. For example, when an uncertainty histogram is provided, the weight of the particle is proportional to the bin height of the histogram location of the measurement. The particle weights are then normalized by dividing the weight of each particle by the sum of the weight of all particles as shown in Equations (14) and (15).

$$W = \sum_{i=1}^{N_p} \omega_k^i \quad (14)$$

$$\{ \omega_k^i := \frac{\omega_k^i}{W} \}_{i=1}^{N_p} \quad (15)$$

Once the normalized particle weights are available, the state estimate and its statistics are calculated from the particle values and weights. For example, when the expected value of the particles is used as the state estimate, the following formula is used:

$$\hat{x}_k = \sum_{i=1}^{N_p} \omega_k^i \hat{x}_k^i \quad (16)$$

This is followed by the resampling step which draws N_p replacement particles from the current set of N_p particles, such that the probability of drawing each particle depends on its importance weight. The resampling step is a crucial part of the PF. In the absence of resampling, many of the particles end up in regions of low posterior probability. This is because of the state uncertainty at the previous time step and the perturbation due to the process noise while transitioning from the previous time step to the current. The resampling step refocuses the set of particles to the regions in the state space with high posterior probability. By doing so, the PF focuses the computational resources of the filter algorithm to the regions in the state space where they matter the most. Thus, the performance of the filter is maintained using a smaller number of particles. Equations (17) through (23) describe the resampling process.

Resampling need not be performed at every time step. It is only performed when the number of particles in regions of high posterior probability falls below a predefined threshold. The metric used to determine the number of particles in regions of high posterior probability is denoted by N_{eff} and is calculated as follows.

$$N_{\text{eff}} = 1 / \left[\sum_{i=1}^{N_p} (\omega_k^i)^2 \right] \quad (17)$$

Typically, resampling is performed when the effective number of particles falls below 2/3 of the total number of particles. The resampling process generates particles with replacement. Note that the resampled particles are denoted by an under bar.

$$\left[\{ \underline{\hat{x}}_k^j, \underline{\omega}_k^j \}_{j=1}^{N_p} \right] = \text{RESAMPLE} \left[\{ \hat{x}_k^i, \omega_k^i \}_{i=1}^{N_p} \right] \quad (18)$$

There exist a few resampling methods in the literature²³. For example, the simple random resampling is carried out as follows. First a cumulative distribution function of the particle weights is calculated as shown by Equations (19) and (20).

$$c_i = 0, \quad i = 1 \quad (19)$$

$$c_i = c_{i-1} + \omega_k^i, \quad i = 2, \dots, N_p + 1 \quad (20)$$

Then, a uniform random number is generated between 0 and 1.

$$\{ u_j \sim \mathcal{U}[0, 1] \}_{j=1}^{N_p} \quad (21)$$

The index of the particle to be included in the resampled set is obtained from the following equation:

$$\{ \text{find } i, \text{ such that } c_i \leq u_j < c_{i+1} \}_{j=1}^{N_p} \quad (22)$$

The state of the i^{th} particle in the prior particle set is assigned to the j^{th} particle in the resampled set, and the importance weight is set uniformly.

$$\{ \underline{\hat{x}}_k^j = \hat{x}_k^i, \quad \underline{\omega}_k^j = \frac{1}{N_p} \}_{j=1}^{N_p} \quad (23)$$

Then the particles are propagated to the next time instant as shown by Equations (24) and (25). Since the individual particles are propagated numerically, any complex nonlinear function can be used in this particle filter framework. Equation (25) shows how the sampling in Equation (24) is performed; the process noise is sampled from a given noise distribution for the particle \hat{x}_k^i , then the state at the next time step \hat{x}_{k+1}^i is obtained by propagating the state with the sampled process noise following the system dynamics. Note that the noise process need not be additive as in Kalman filters. This is because the function f can be any form or any algorithm that can compute the next state given the state and the process noise.

$$\{ \text{sample } \hat{x}_{k+1}^i \sim p(x_{k+1} | \hat{x}_k^i) \}_{i=1}^{N_p} \quad (24)$$

$$\{ \hat{x}_{k+1}^i = f(x_k^i, v_k^i) \}_{i=1}^{N_p} \quad (25)$$

This sequence of steps is continued till the final time step.

C. An Example of a Particle Filter for a Linearly Unobservable but Nonlinearly Observable System

Before delving into PF results with the C-MAPSS40K Engine model, the overall approach is illustrated by considering a simple example that leads to an unobservable linearized system but is nonlinearly observable. Consider a two-state system whose dynamics are given by the following equations:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) = 0 \\ \dot{x}_2 &= f_2(x_1, x_2) = -x_1 x_2. \end{aligned} \quad (26)$$

The measurement model for this system is given as follows:

$$y = g(x_1, x_2) = x_1 x_2 \quad (27)$$

The true trajectory is obtained by integrating Equation (26) with initial conditions $x_1(0) = 1.0$ and $x_2(0) = 10.0$, over the interval $t = [0, 2]$, with time step $\Delta t = 0.04$. In the present example, a variable step-size 4th-order Runge Kutta method with 5th-order error correction is used. The evolution of the true trajectory is shown in Figure 4.

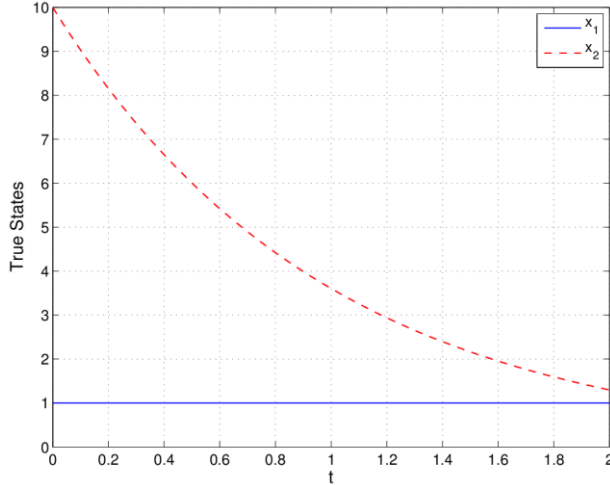


Figure 4. True States for System Given by Equation (26).

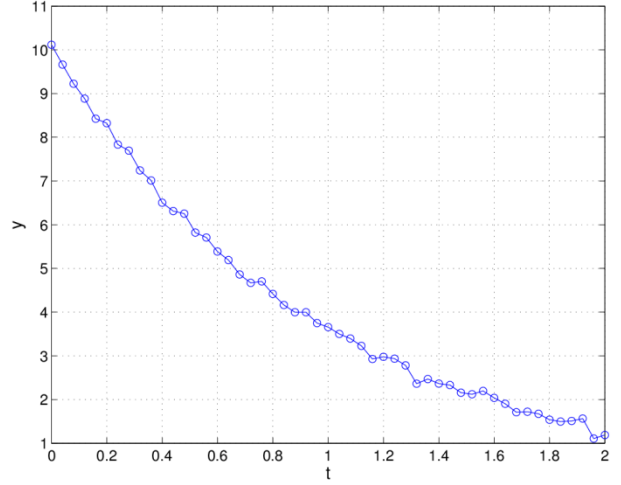


Figure 5. Measurements for System Obtained Using Equation (27).

Measurements for this system are defined by Equation (27) with additive zero-mean Gaussian measurement uncertainty whose variance is given by $\sigma_R^2 = 0.01$. The measurements are shown in Figure 5. Both the EKF and PF are initialized with initial state estimates given by $\hat{x}_1(0) = 3.0$ and $\hat{x}_2(0) = -1.0$. The initial state covariance is given by $P(0) = \text{diag}(10, 25)$.

1. EKF Formulation

A continuous-discrete EKF formulation²⁴ is first implemented. In this approach, the system estimates are propagated in continuous time using the nonlinear model given by Equation (26), whereas measurements are made available at discrete time steps.

The EKF utilizes a linearization of the measurement model when calculating the Kalman gain, and a linearization of the system dynamics when propagating the estimation error covariance matrix. At the k th time step, the linearization of the measurement model is denoted by C_k and the linearization of the system dynamics is denoted by A_k . These matrices can be calculated by taking the derivative of Equations (26) and (27) with respect to states x_1 and x_2 , and are as follows:

$$C_k = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} \end{bmatrix} = [x_2 \quad x_1] \quad (28)$$

$$A_k = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -x_2 & -x_1 \end{bmatrix} \quad (29)$$

2. Linear and Nonlinear Observability

The convergence properties of the EKF depend on the observability of the linearized system. The observability matrix for this system is denoted by \mathcal{O}_k and is given by:

$$\mathcal{O}_k = \begin{bmatrix} C_k \\ C_k A_k \end{bmatrix} = \begin{bmatrix} x_2 & x_1 \\ -x_1 x_2 & -x_1^2 \end{bmatrix} \quad (30)$$

Where C_k and A_k are obtained from Equations (28) and (29). Since the second row of \mathcal{O}_k is equal to the first row scaled by $-x_1$, the rank of the observability matrix is at most 1, and the system is not fully observable when linearized. The EKF is therefore not expected to converge.

It can be shown, however, that the system is nonlinearly observable. Nonlinear observability can be demonstrated by first taking the time derivative of the measurement equation:

$$\dot{y} = \dot{g} = \frac{\partial g}{\partial x_1} \dot{x}_1 + \frac{\partial g}{\partial x_2} \dot{x}_2 = \frac{\partial g}{\partial x_1} f_1 + \frac{\partial g}{\partial x_2} f_2 = -x_1^2 x_2 \quad (31)$$

Linearizing the measurement and its time derivative without first linearizing the system dynamics preserves some of the nonlinear structure of the system. The local nonlinear observability is then governed by the rank of the matrix \mathcal{O}_N , which can be obtained using Equations (27) and (31):

$$\mathcal{O}_N = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} \\ \frac{\partial \dot{y}}{\partial x_1} & \frac{\partial \dot{y}}{\partial x_2} \end{bmatrix} = \begin{bmatrix} x_2 & x_1 \\ -2x_1 x_2 & -x_1^2 \end{bmatrix} \quad (32)$$

The foregoing matrix has rank 2 as long as $x_2 \neq 0$.

3. PF Formulation and Comparison with EKF

A PF is implemented for the foregoing problem. For the example, $N_p = 10000$ particles are initialized by sampling from a bivariate Gaussian distribution with mean given by $\{\hat{x}_1(0) \quad \hat{x}_2(0)\}^T$ and covariance given by $P(0)$. At each iteration, residuals are calculated for each particle with respect to the measurement, using the measurement model given by Equation (27), and a Gaussian uncertainty model is used to update the weights.

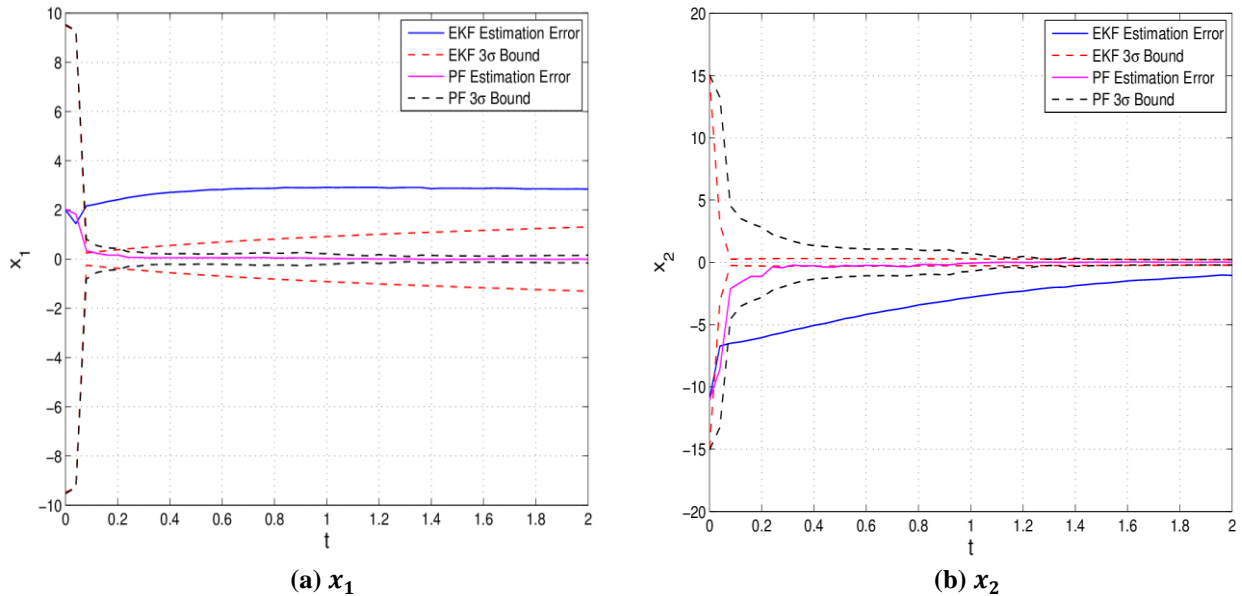


Figure 6. Comparison between Estimation Errors and Confidence Bounds Obtained Using PF and EKF

The posterior mean and variance are calculated for both states, and compared with the EKF estimate and (co)variance. The comparison between PF and EKF for both states is shown in the foregoing Figure 6. The blue solid line and magenta solid line depict the estimation errors in the states, as obtained from the EKF and PF, respectively. The red broken line and the black broken lines depict the 3σ confidence bounds, as obtained from the EKF and PF, respectively. For the EKF, these bounds are obtained by propagating the Riccati equation for the state covariance matrix, whereas in the case of the PF, variances are calculated from the particle weights.

It is immediately obvious that the EKF is not able to successfully estimate the states of the problem. Whereas the estimation error in the second state converges to zero very slowly, the estimation error in the first state appears to converge to a non-zero bias. The estimation errors are well outside the 3σ confidence bounds. It is also evident that the PF can successfully track the states, and the variance of the errors converges to steady-state values relatively quickly. The foregoing example is a relatively simple demonstration of the PF's ability to leverage nonlinear observability in cases where the KF or its derivatives fail.

V. Particle Filter Performance with the C-MAPSS40K Engine Model

A. Simulation Scenario

The performance assessment of the PF with the C-MAPSS40K engine model is approached in two steps. Firstly, under complete linear observability conditions, a KF and a PF are designed, and it is shown that the PF leads to the same level of performance as that of the KF. Secondly, it is shown that estimating more parameters than the number of measurements is not an infeasible problem if the PF is employed for the engine in transient state. For this purpose, the control variable, the fuel flow rate, is altered for different scenarios. In the first scenario, the fuel flow rate is kept constant at 1.6. In the second scenario, the fuel flow rate is varied as shown by the blue line in Figure 7. The second scenario represents an engine undergoing transient operation, wherein the fuel flow rate is varied. Such transient operation has been found to improve observability of the nonlinear engine dynamics.

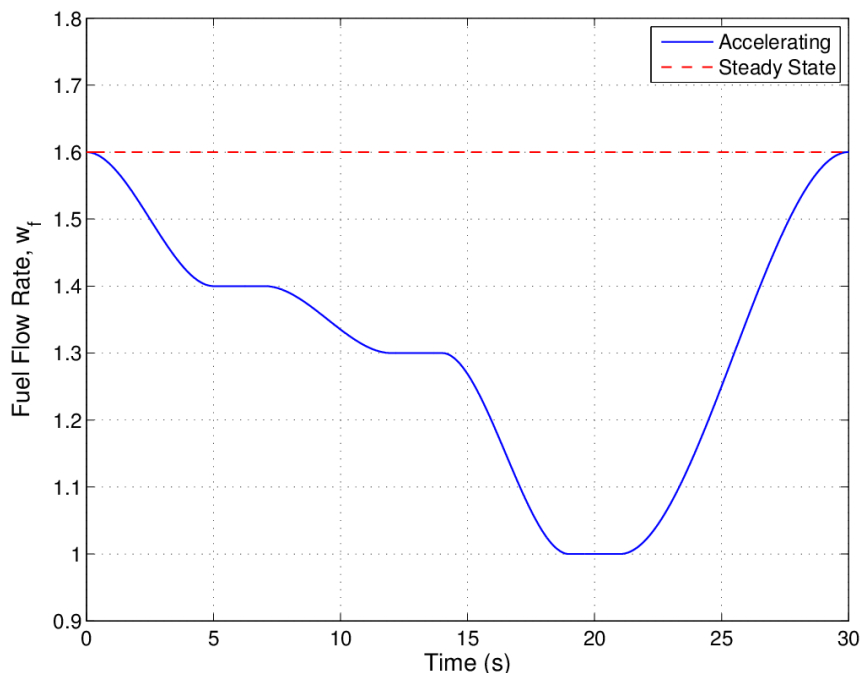


Figure 7. Fuel Rate Variation used in the Present Study

The system consists of 6 measurements, denoted by measurement vector $y \in \mathbb{R}^6$, and listed in Table 1. The measurement error is assumed to follow a Gaussian distribution, with standard deviation of the error equal to 0.25% of the steady state value for N_c and N_f , 0.75% of the steady state value for T_{25} and T_{50} , and 0.50% of the steady state

value for P_{50} and P_{s3} . The standard deviation for fan speed measurement error, temperature measurement error, and pressure measurement error are denoted by σ_N , σ_T , and σ_P , respectively.

B. Baseline Kalman Filter with 6 Parameters Estimated

A Kalman filter is implemented for the engine health parameters for the linearized system in Equation (4) at a trim condition (x_0, u_0) . The linearized system is further discretized with respect to time using an Euler integration scheme. Numerical experiments show that a time step of 0.015s, is sufficiently small for numerical stability of the Euler integration, with errors less than 0.1% in system states when compared with a 4th-order Runge Kutta integrator. The resulting linear, discrete, autonomous system is then given by the following:

$$\begin{aligned}\Delta x_{k+1} &= F_d \Delta x_k + w \\ \Delta y_k &= G \Delta x_k + v\end{aligned}\quad (33)$$

where $F_d = I + A_d \Delta t$, and $w \sim \mathcal{N}(0, Q)$ and $v \sim \mathcal{N}(0, R)$ denote the process and measurement uncertainty, respectively. Both uncertainties are assumed to be zero-mean Gaussian in nature, with covariance matrices Q and R respectively. Whereas $R = \text{diag}([\sigma_T^2 \ \sigma_T^2 \ \sigma_P^2 \ \sigma_N^2 \ \sigma_N^2 \ \sigma_P^2])$ is given by sensor noise characteristics, Q is a tuning parameter chosen to adjust filter performance.

Since the rank of the observability matrix $\mathcal{O}(C_a, A_a)$ is 8, only 8 states are observable. It is assumed therefore, that parameters η_{HPT} , γ_{HPT} , η_{LPT} , and γ_{LPT} are known, and the remaining 8 states of the augmented state vector are to be estimated. The initial estimates of states N_f and N_c are set equal to their observed steady-state measurements, and the initial estimates of the health parameters are set to random values such that the standard deviation of the errors is 5% of the nominal (true) values. Initial uncertainty covariance matrix is set equal to a diagonal matrix; i.e. standard deviation of initial uncertainty is equal to 100 rpm for the fan and compressor speeds, and 1e-03 for the health parameters. Process uncertainty covariance matrix Q is assumed diagonal with standard deviations of 1.2, 1.2, 0.0001, 4e-6, 2.2e-5, 2.2e-5, 2.2e-4, and 2.2e-5 for the eight dynamical equations. However, for the first 1.5 seconds of the simulation process uncertainty standard deviation is increased by ten times in magnitude to improve convergence properties of the filter. The estimation errors in the state are shown in Figure 8, with the estimation error being depicted by a blue solid line, and the 3σ confidence bound as a red dashed line. Figure 9 shows the estimated health parameter as a blue solid line, superimposed on their true values, depicted by a black line. The feasible range of health parameter values is shown as a red, dashed line, whereas the 3σ confidence bounds are indicated in magenta.

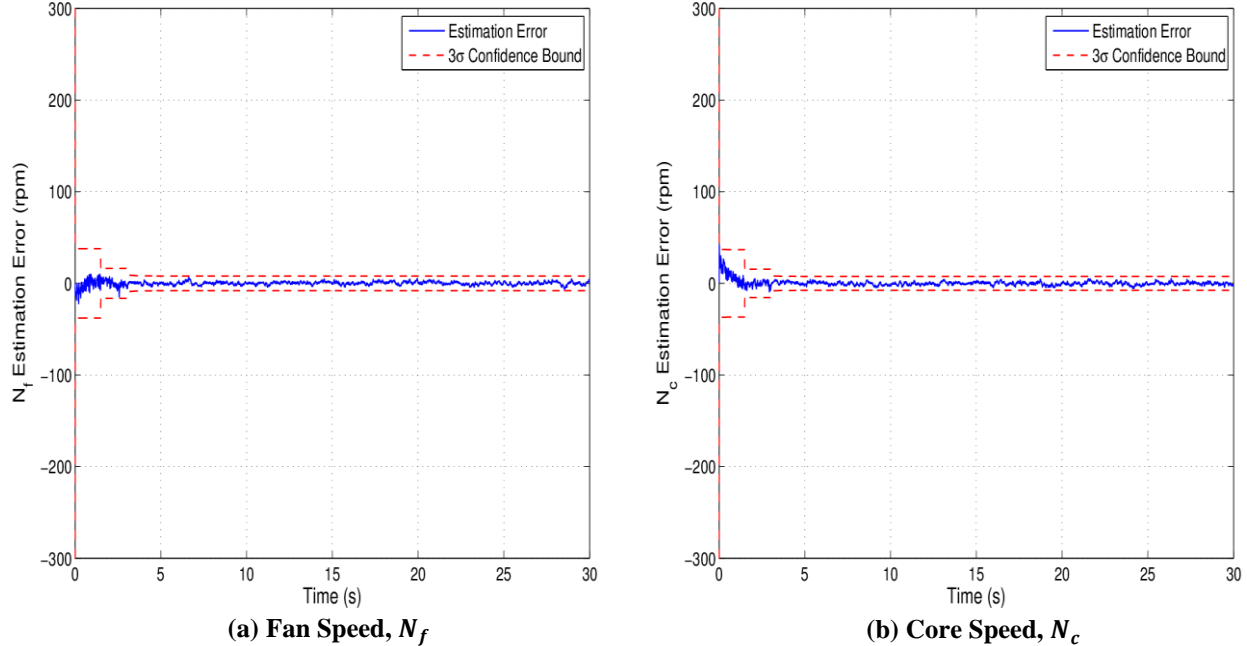
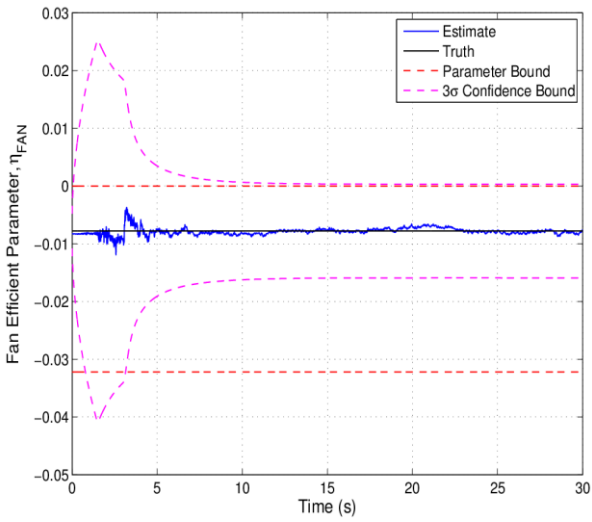
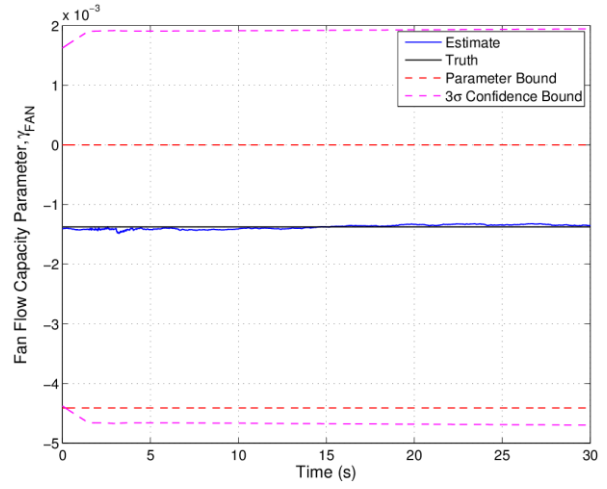


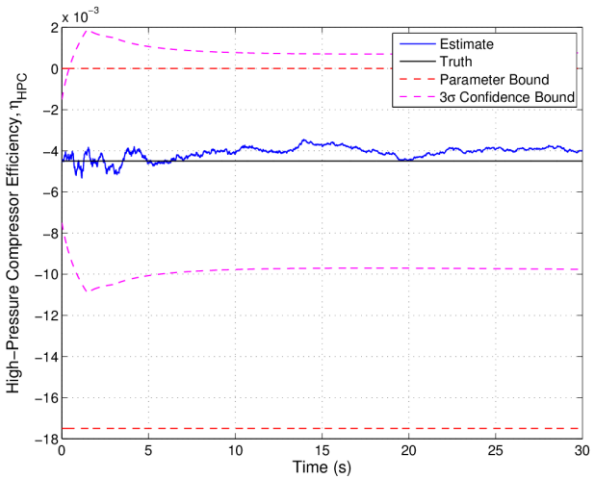
Figure 8. KF Estimation Error in the State



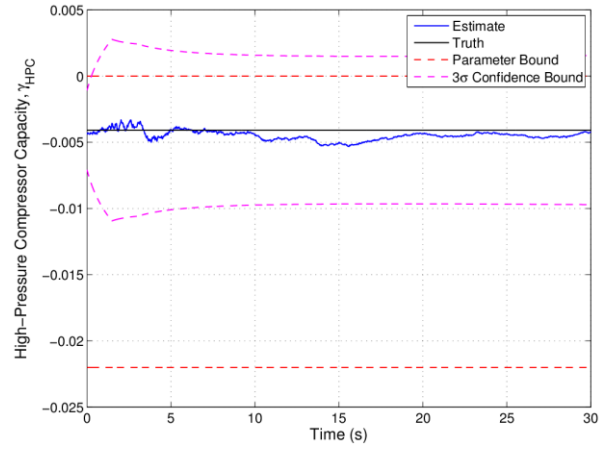
(a) η_{FAN}



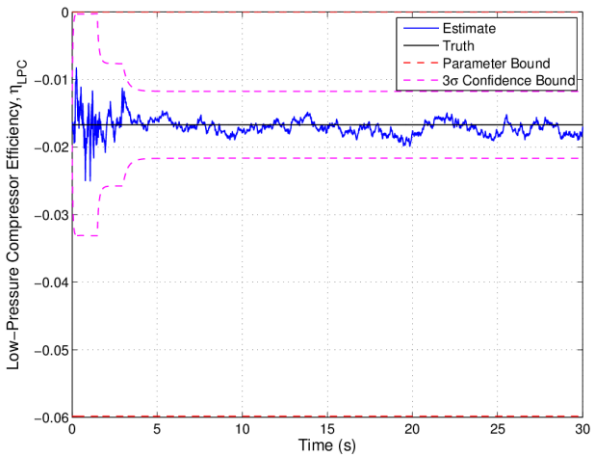
(b) γ_{FAN}



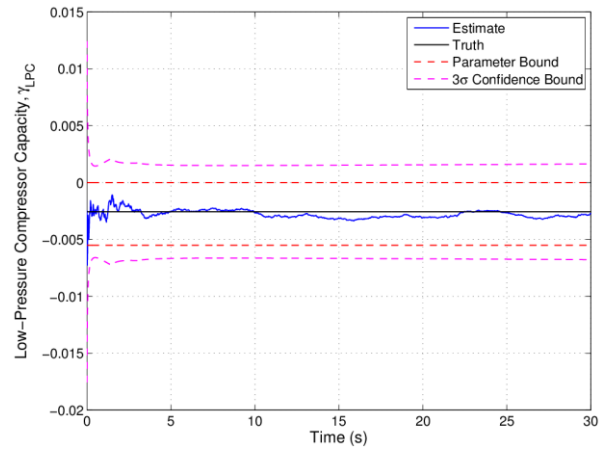
(c) η_{HPC}



(d) γ_{HPC}



(e) η_{LPC}



(f) γ_{LPC}

Figure 9. KF Estimates for Health Parameters

C. Particle Filter with 6 Parameters Estimated

A Particle filter was implemented for the estimation of 2 states and 6 parameters, with identical initial values to the KF estimator described in Section V.B. The PF implementation differs from the KF implementation in the following aspects. Firstly, although Euler integration is used to integrate particles through the system dynamics, the dynamical equations and the measurement model are not linearized. Instead, the system dynamics and measurement model process the particles in parallel using the nonlinear engine model. Secondly, the PF implementation utilizes histograms to represent the process and measurement uncertainty in the general case. However, arbitrary histograms are not explored in this research. Instead, discrete near-Gaussian histograms with finite support up to $\pm 4\sigma$ limit were used to generate random numbers for the process uncertainty and for the measurement uncertainty. Thirdly, a key feature of the present problem is that the values of the health parameters are bounded in a given range. For instance, Figure 10 shows bounds for 10 health parameters, which have values for deterioration levels of 0, 0.5 and 1 (end-of-life), and utilize linear interpolation for intermediate values of the deterioration level.

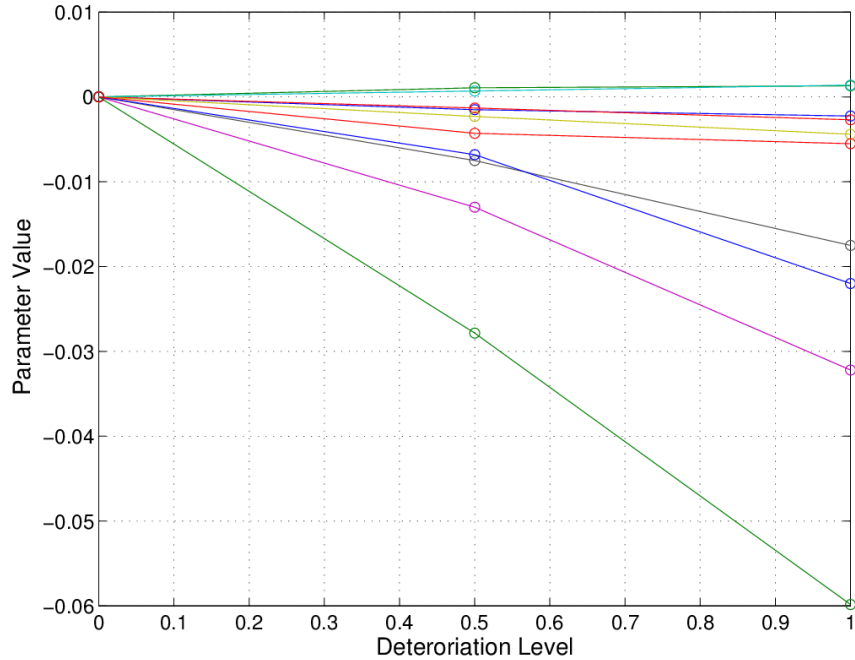


Figure 10. Health Parameter Bounds

Since values of the health parameter outside this range can be considered infeasible, at every step of the PF, the value of the health parameters in the particles representing the augmented state can be evaluated, and weights of those particles that have values outside the bounds can be set to zero. This ensures a more accurate domain of validity of the PF in comparison with the KF. Finally, histograms of the augmented state can be constructed from the particles. For example, a histogram for parameter γ_{LPC} is shown in Figure 11 (Left). Using the state Probability Density Function (PDF), the Cumulative Density Function (CDF) can be constructed numerically, as shown in Figure 11 (Right). The median of the distribution can be obtained by interpolation of the CDF and obtaining the value of the state where the CDF is equal to 0.5. At this point, the probability mass on either side is equal. Furthermore, the limits enclosing the 0.997 probability mass of the distribution can also be calculated. It should be noted that these limits can be asymmetric with respect to the median. When the distribution is Gaussian, this is equivalent to the expected value and the 3σ bound of the distribution, respectively. However, when histograms with finite support are considered, this equivalence fails. Therefore, results with the PF will be depicted using the median, 0.997 probability mass limits, and the support limits of the histogram.

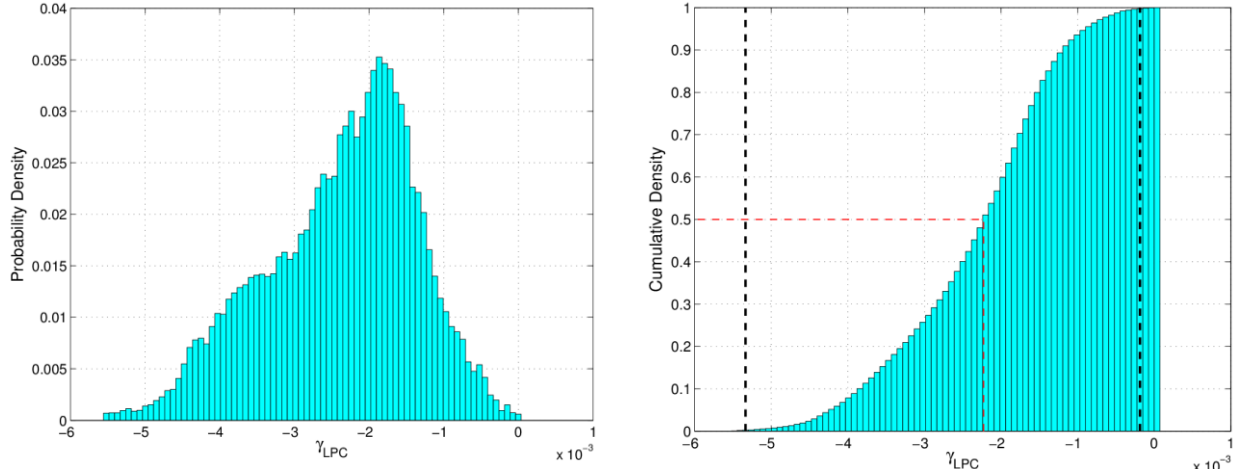
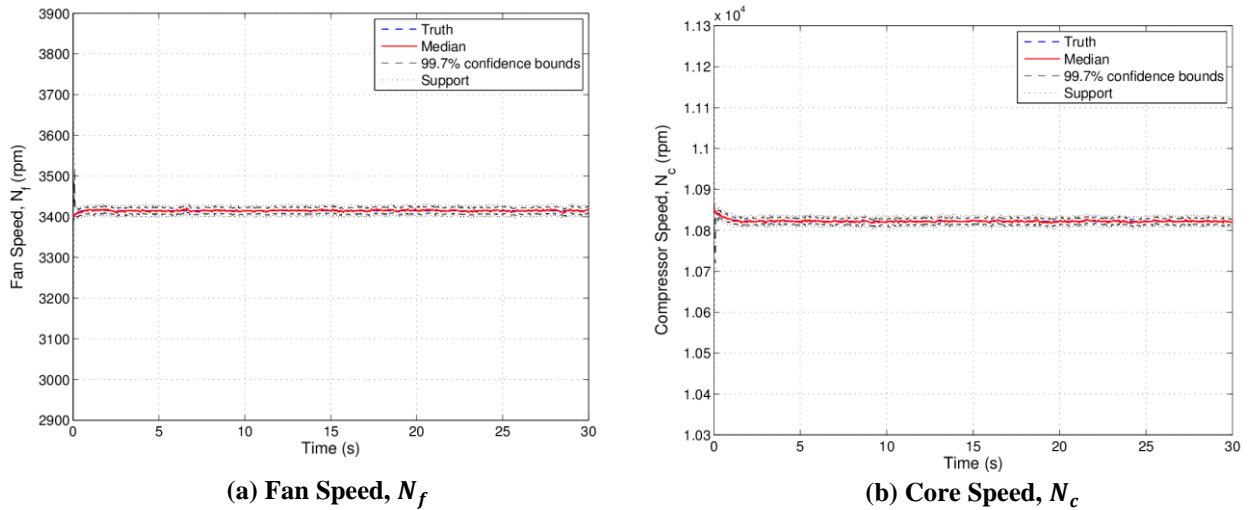


Figure 11. Probability Density of γ_{LPC} at $t=30s$ (Left), and Cumulative Density with Median and 0.997 Probability Mass Limits (Right)

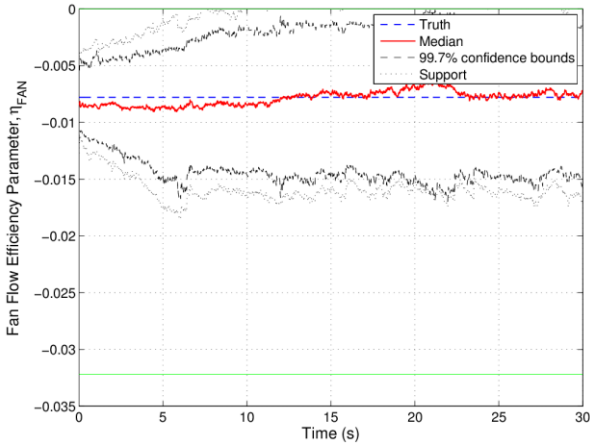
The PF was implemented with 250,000 particles. Results of the PF implementation for an 8-state estimator are shown in Figure 12 and Figure 13. Upon comparing with the results from the KF, it is observed that PF performs as well as the KF and in some cases exceeds KF performance under identical conditions. For example, the 99.7% / 3σ bounds are noticeably smaller in the PF.



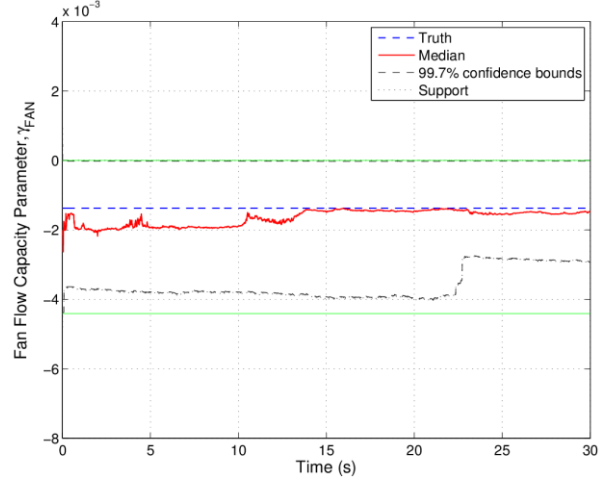
(a) Fan Speed, N_f

(b) Core Speed, N_c

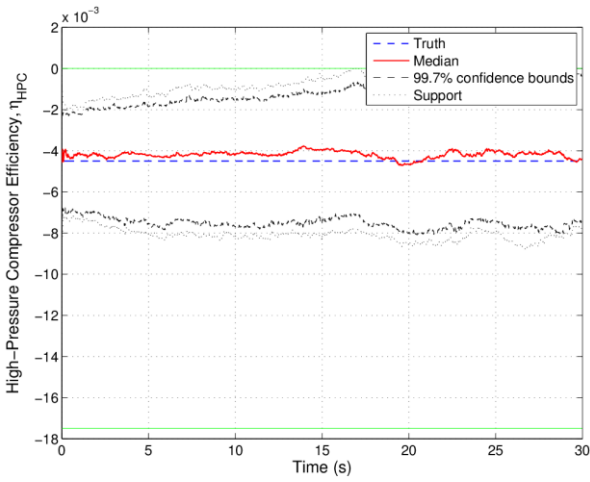
Figure 12. PF Median Estimates and Dispersions for the State



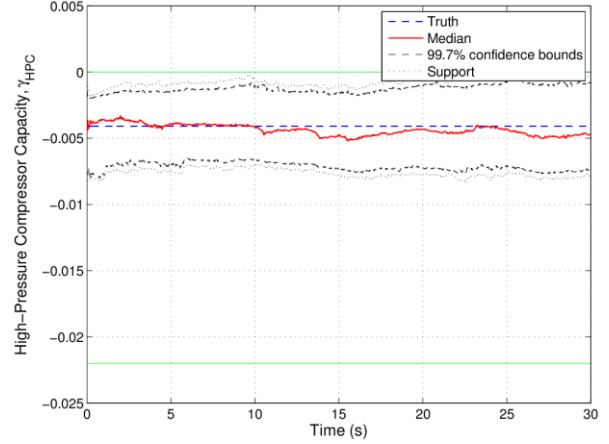
(a) η_{FAN}



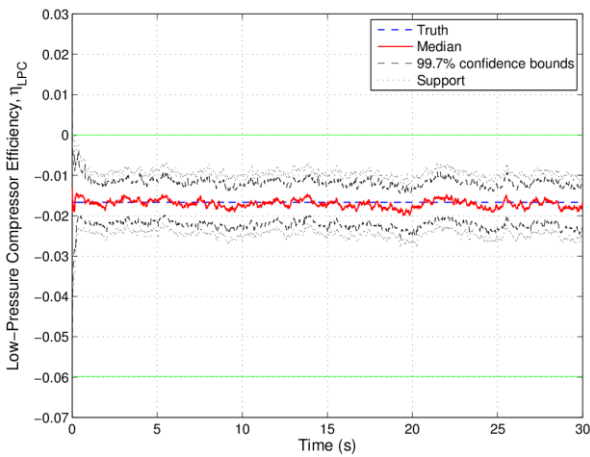
(b) γ_{FAN}



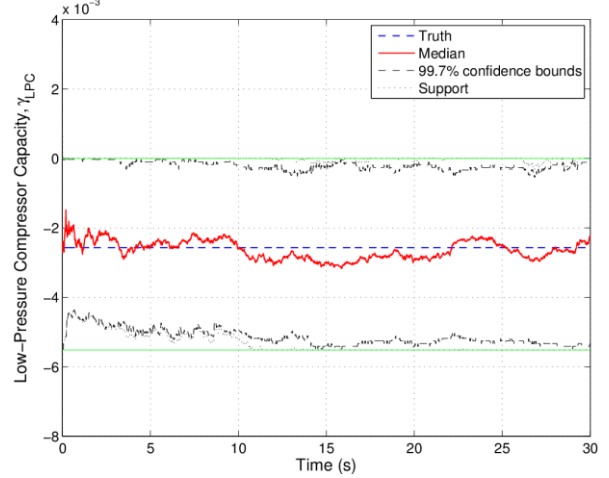
(c) η_{HPC}



(d) γ_{HPC}



(e) η_{LPC}



(f) γ_{LPC}

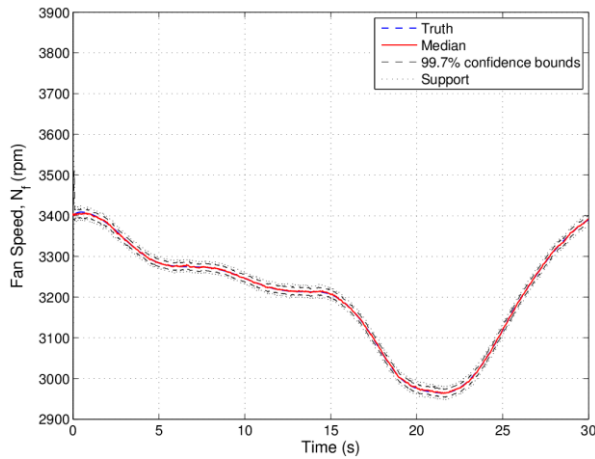
Figure 13. PF Median-Estimates and Dispersions for Health Parameters

D. Particle Filter with 7 Parameters Estimated

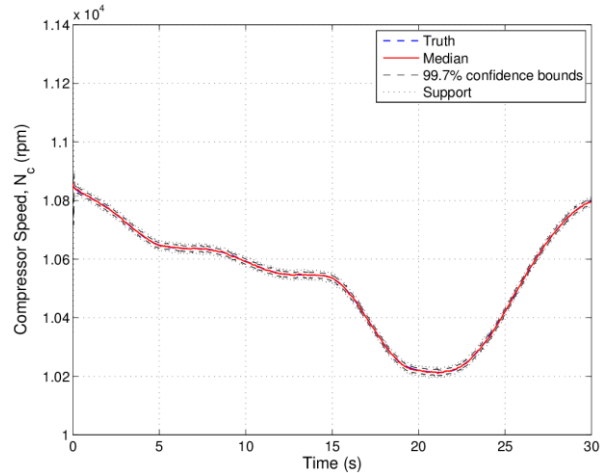
1. Particle Filtering Results

This corresponds to the case where the number of health parameters to be estimated exceeds the number of measurements. It has been noted in Section III.B that this case is unobservable in a linear sense, and a conventional KF will fail to converge. The linear observability analysis shows that two states (fan and compressor speeds) and six parameters can be estimated using six measurements. In this section, the PF is used to estimate an additional parameter, η_{HPT} by exploiting the nonlinear observability. Towards that end, the transient engine performance scenario is used. This is achieved by varying the fuel flow rate as shown in Figure 7. This ensures that the time derivatives of the fan and core speed, denoted by \dot{N}_f and \dot{N}_c are never zero and the measurement themselves exhibit variation with time.

The estimates for the nine augmented states are shown in Figure 14 and Figure 15. In addition to the eight augmented states listed in Section V.C, the high pressure turbine efficiency parameter η_{HPT} is also estimated. It is evident from the results that under transient engine operating conditions, the PF can enable the estimation of more than the six health parameters. Additional parameters can possibly be estimated by a combination of proper filter tuning and selection of a sequence of engine transient operating conditions.

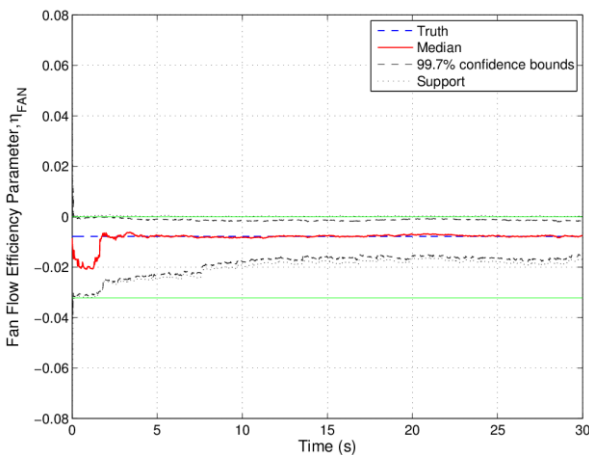


(a) Fan Speed, N_f

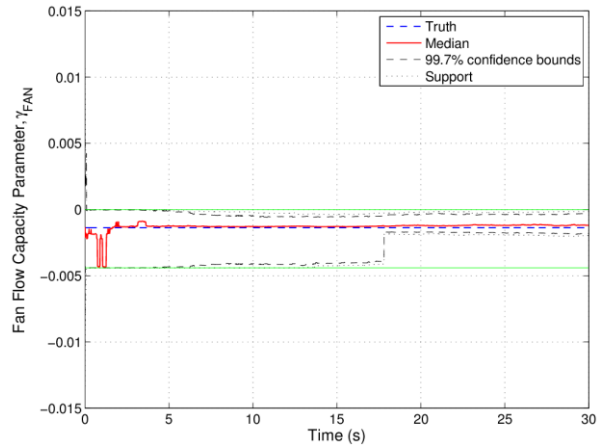


(b) Core Speed, N_c

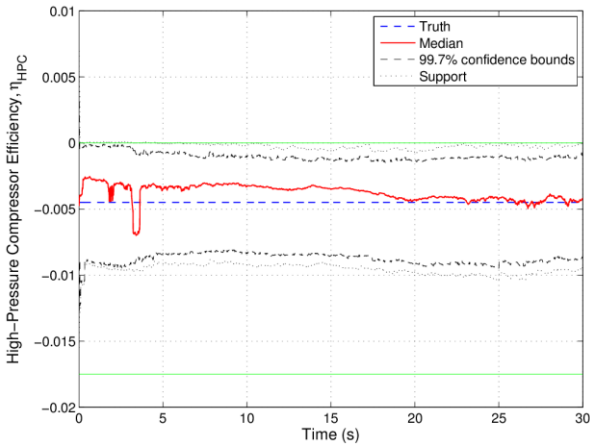
Figure 14. PF Median Estimates and Dispersions for the State in Transient Operation



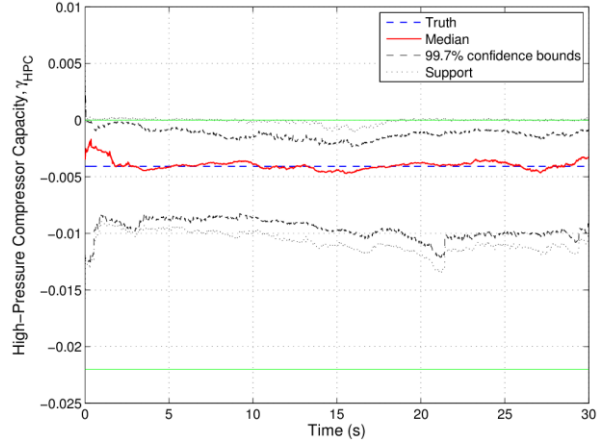
(a) η_{FAN}



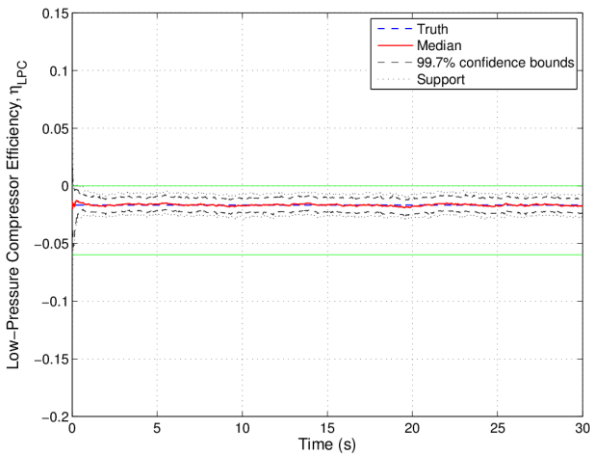
(b) γ_{FAN}



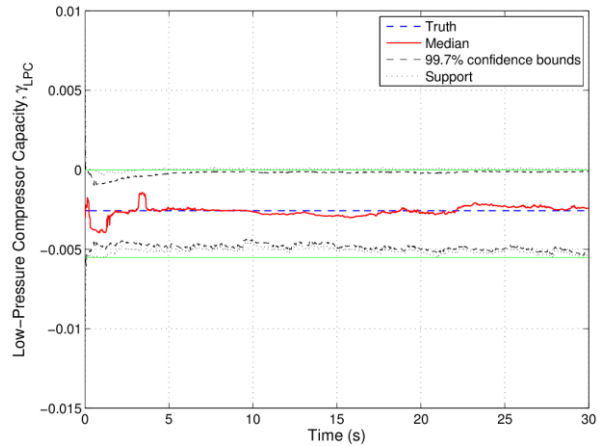
(c) η_{HPC}



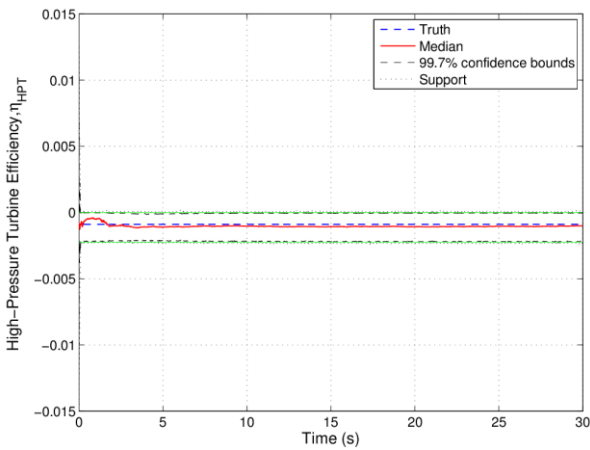
(d) γ_{HPC}



(e) η_{LPC}



(f) γ_{LPC}



(g) η_{HPT}

Figure 15. PF Median-Estimates and Dispersions for Health Parameters

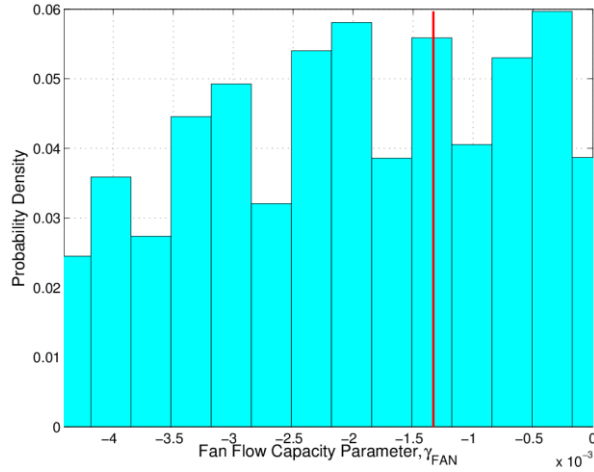
2. Discussions on Posterior Distributions

A closer examination of Figure 15 (b) depicting the estimate of the fan flow capacity parameter γ_{FAN} shows that the median of the posterior state PDF has a small bias with respect to the true value. Further analysis shows that this

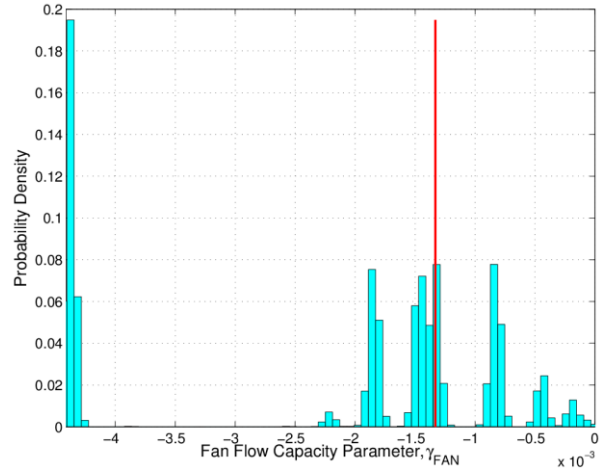
is a consequence of a multimodal posterior PDF, in which neither the median nor mean is necessarily the best statistic to compare with the true value. This is because the median or the mean of a bimodal distribution can lie in a region of very low probability of occurrence, between two or more regions with high probability of occurrence.

The time evolution of the posterior PDF of γ_{FAN} is shown as a series of ‘snapshots’ in Figure 16. The initial distribution is shown in Figure 16 (a). Although the distribution was assumed unimodal (the underlying distribution from which the histogram or discrete PDF is constructed is Gaussian, multiple peaks appear in the initial part of the simulation Figure 16 (b). This may be a consequence of the nonlinear C-MAPSS40K engine model.

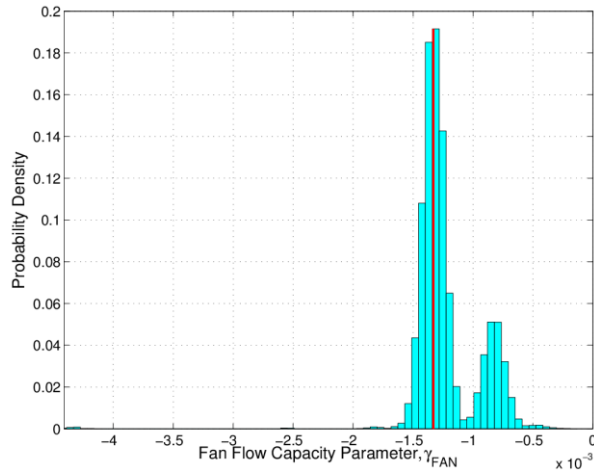
As time progresses, with the availability of measurements, the multiple modes resolve into two strong modes. The true value, shown as a solid red line, corresponds to one of the peaks of the distribution. Addition of larger process uncertainty can possibly accelerate the convergence of the two modes into one, at the cost of increasing the support of the distribution and larger confidence bounds. Due to the presence of two modes, the median of the distribution is shifted from the correct mode, and this results in the bias observed in Figure 15 (b).



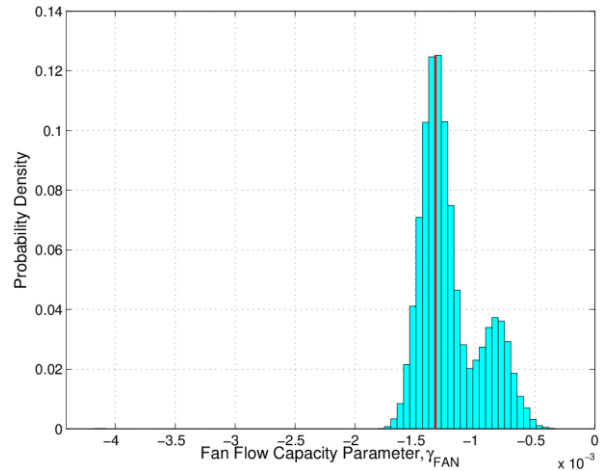
(a) $t = 0s$



(b) $t = 1.5s$



(c) $t = 7.5s$



(d) $t = 15s$

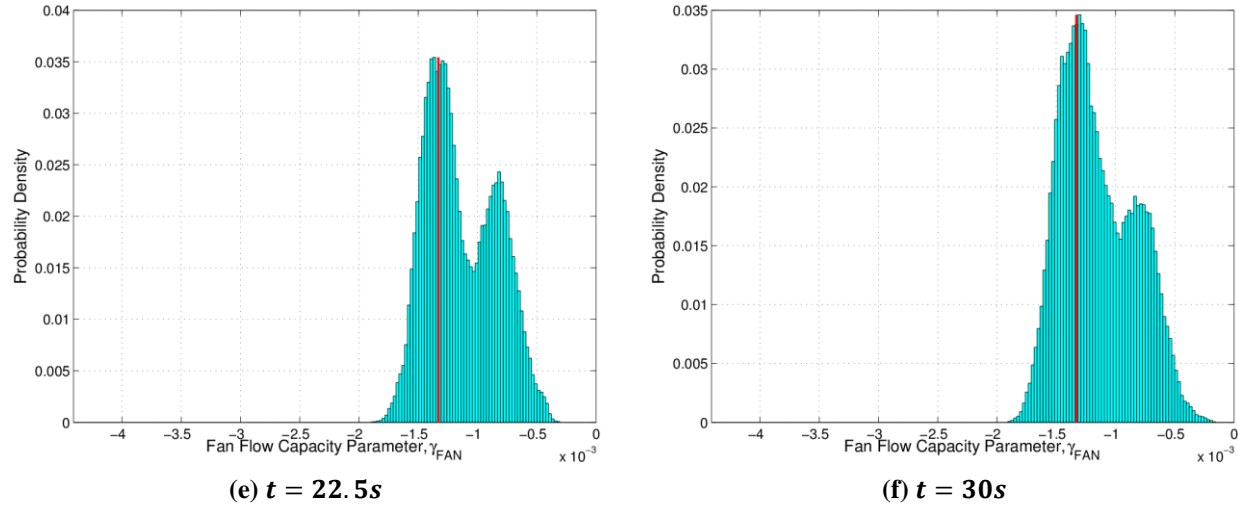


Figure 16. Time Evolution of Posterior Probability Density Function of Fan Flow Capacity Parameter

VI. Conclusions

This paper addresses the feasibility of employing particle filters for the estimation of engine health parameters. The investigation of nonlinear observability for C-MAPSS40K engine model forms the keystone for the nonlinear particle filter approach. Nonlinear observability test shows that the number of engine health parameters that can be estimated is not limited by the number of measurements. It is shown that a nonlinear estimation approach such as particle filter can exploit this fact and can estimate more engine health parameters than the number of available measurements. Investigating an example posterior probability density function for a engine health parameter reveals additional nonlinear nature of the engine health parameter problem: multimodal probability distributions.

While this paper demonstrates the feasibility of nonlinear estimation for engine health parameter estimations, the intricate relations between the control signals and the number of engine health parameters are not fully investigated. An in-depth investigation for the interplay between the control signals and estimation performance of nonlinear estimation methods with a high-fidelity engine model is left for future research.

Acknowledgments

This research was supported under NASA Contract No. NNX13CC21P, with Donald L. Simon at NASA Glenn Research Center as the Technical Monitor. The authors appreciate the guidance and interesting discussions with the Technical Monitor and the engine controls group at NASA Glenn Research Center.

References

- ¹Luppold, R.H., Roman, J.R., Gallops, G.W., and Kerr, L.J., "Estimating In-Flight Engine Performance Variations Using Kalman Filter Concepts," *AIAA Joint Propulsion Conference*, Monterey, CA, 1989.
- ²Brotherton, T., Volponi, A., Luppold, R., and Simon, D., "eSTORM: Enhanced Self Tuning Onboard Real-time Engine Model," *IEEE Aerospace Conference*, Big Sky, MT, 2003.
- ³Volponi, A., "Enhanced Self Tuning On-Board Real-Time Model (eSTORM) for Aircraft Engine Performance Health Tracking," NASA/CR-2008-215272, NASA 2008.
- ⁴Simon, D. and Garg, S., "Optimal Tuner Selection for Kalman Filter-Based Aircraft Engine Performance Estimation," *Journal of Engineering for Gas Turbines and Power*, Vol. 132, No. 3, 2010.
- ⁵Simon, D., Armstrong, J., and Garg, S., "Application of an Optimal Tuner Selection Approach for On-Board Self-Tuning Models," *Proceedings of the ASME Turbo Expo: Turbine Technical Conference*, Vancouver, Canada, 2011.
- ⁶Armstrong, J. and Simon, D., "Constructing an Efficient Self-Tuning Aircraft Engine Model for Control and Health Management Applications," Available at <http://www.phmsociety.org/node/864>, 2012.
- ⁷Simon, D., "Kalman Filtering with State Constraints: A Survey of Linear and Nonlinear Algorithms," *IET Control Theory and Applications*, Vol. 4, pp.1303-1318, 2010.
- ⁸Simon, D., "A Comparison of Filtering Approaches for Aircraft Engine Health Estimation," *Aerospace Science and Technology*, Vol. 12, pp.276-284, 2008.

- ⁹Verma, R. and Ganguli, R., “Denoising Jet Engine Gas Path Measurements Using Nonlinear Filters,” IEEE/ASME Transactions on Mechatronics, Vol. 10, pp.461-464, 2005.
- ¹⁰Thrun, S., Burgard, W. and Fox, D., Probabilistic Robotics, The MIT Press, Cambridge, Massachusetts, 2005.
- ¹¹Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T., “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking,” IEEE Transactions on Signal Processing, Vol. 50, No. 2, pp. 174-188, 2002.
- ¹²Menon, P.K., Kim, J., and Tandale, M.D., “Next-Generation Target State Estimation Algorithm for the Interception of Maneuvering Ballistic Missiles,” Final Report Prepared Under MDA Phase I SBIR Contract No. HQ0147-09-C-7111, 2009.
- ¹³Wan, E. A. and Merwe, R. van der, “The Unscented Kalman Filter for Nonlinear Estimation,” Proceedings of IEEE Symposium 2000, Alberta, Canada , 2000.
- ¹⁴May, R.D., Csank, J., Lavelle, T.M., Litt, J.S., and Guo, T.-H., “A High-Fidelity Simulation of a Generic Commercial Aircraft Engine and Controller,” NASA/TM-2010-216810, 2010.
- ¹⁵Kailath, T., *Linear Systems*, Prentice Hall, Englewood, NJ, 1980.
- ¹⁶Espna, M.D., “Sensor Biases Effect on the Estimation Algorithm for Performance-Seeking Controllers,” Journal of Propulsion and Power, Vol.10, pp.527-532,1994.
- ¹⁷Angelova, M., “Nonlinear Observability and Identifiability: General Theory and a Case Study of a Kinetic Model for *S. cerevisiae*,” Thesis for the Degree of Licentiate of Engineering, Department of Mathematics, Chalmers University of Technology and Göteborg, 2004.
- ¹⁸Tunali, E. T. and Tarn, T.-J., “New Results for Identifiability of Nonlinear Systems,” IEEE Transactions on Automatic Control, Vol. 32, No. 2, pp.146-154, 1987.
- ¹⁹Xia, X. and Moog, H., “Identifiability of Nonlinear Systems With Application to HIV/AIDS Model,” IEEE Transactions on Automatic Control, Vol. 48, No. 2, pp.330-336, 2003.
- ²⁰Diop, S. and Fliess, M., “Nonlinear Observability, Identifiability and Persistent Trajectories,” *IEEE Conference on Decision and Control*, Brighton, U.K., pp.714-719, 1991.
- ²¹Diop, S. and Wang, Y., “Equivalence between algebraic observability and local generic observability,” *IEEE Conference on Decision and Control*, San Antonio, TX, pp.2864-2865, 1993.
- ²²Marino, R. and Tomei, P., *Nonlinear Control Design: Geometric, Adaptive, and Robust* , Prentice Hall International (UK), 1995.
- ²³Hol, J.D., Schon, T.B., and Gustafsson, F., “On Resampling Algorithms for Particle Filters,” Proceedings of IEEE Nonlinear Statistical Signal Processing workshop, pp.79-82, 2006.
- ²⁴Gelb, A., *Applied Optimal Estimation*, The MIT Press, Cambridge, Massachusetts, 1974.