# Remaining Useful Life Prediction for Aero-Engines Combining Sate Space Model and KF Algorithm

*Cai Jing* *, *Zhang Li*, *Dong Ping*

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China

(Received 4 May 2016; revised 24 January 2017; accepted 25 February 2017)

**Abstract**: The key to failure prevention for aero-engine lies in performance prediction and the exhaust gas temperature margin (EGTM) is used as the most important degradation parameter to obtain the operating performance of the aero-engine. Because of the complex environment interference, EGTM always has strong randomness, and the state space based degradation model can identify the noisy observation from the true degradation state, which is more close to the actual situations. Therefore, a state space model based on EGTM is established to describe the degradation path and predict the remaining useful life (RUL). As one of the most effective methods for both linear state estimation and parameter estimation, Kalman filter (KF) is applied. Firstly, with EGTM degradation data, state space model approach is used to set up a state space model for aero-engine. Secondly, RUL of aero-engine is analyzed, and expected RUL and distribution of RUL are determined. Finally, the sate space model and KF algorithm are applied to an example of CFM-56 aero-engine. The expected RUL is predicted, and corresponding probability density distribution (PDF) and cumulative distribution function (CDF) are given. The result indicates that the accuracy of RUL prediction reaches 7.76% ahead 580 flight cycles (FC), which is more accurate than linear regression, and therefore shows the validity and rationality of the proposed method.

**Key words**: remaining useful life; exhaust gas temperature margin (EGTM); Kalman filter; Sate space model

**CLC number**: TB301 **Document code**: A **Article ID**: 1005-1120(2017)03-0265-07

## 0 Introduction

According to the statistics data, the maintenance cost of the aero-engine accounts for more than 30% of the total operating cost of airlines, which reveals much controllable margin[1]. It is widely proven that prognostics and health management (PHM) can reduce the risk of catastrophic system failure as well as the maintenance cost, compared with traditional maintenance techniques. The remaining useful life (RUL) prediction is a key technique of PHM, because it provides the base for aero-engine maintenance plans through predicting RUL of aero-engine accurately, reducing maintenance costs and lowering the probability of risk[2-3]. As the most important measure of aero-engine health, the exhaust gas temperature margin (EGTM) is an effective approach to identify deterioration of aero-engine. EGTM can not only indicate an abnormal operation but also predict RUL without the need to stop the aero-engine and perform expensive full inspection.

There are several studies on the subject of RUL prediction of the aero-engine. García et al.[4] described a hybrid PSO-SVM-based model for the prediction of RUL of aircraft engines. Malinowski[2] presented an approach, based on shapelet extraction, to estimate RUL, and evaluated it by a case study turbofan engines. Li[5] described a prognostic approach to estimate RUL of gas turbine engines based on gas path analysis. Zaidan et al.[6] selected a Bayesian hierarchical model to utilize fleet data from multiple assets to perform probabilistic estimation of RUL. EGTM was taken as a measure of the engine health to

* Corresponding author, E-mail address: caijing@nuaa.edu.cn.

**How to cite this article**: Cai Jing, Zhang Li, Dong Ping. Remaining useful life prediction for aero-engines combining sate space model and KF algorithm[J]. Trans. Nanjing Univ. Aero. Astro., 2017,34(3):265-271.
http://dx.doi.org/10.16356/j.1005-1120.2017.03.265

predict RUL of the aero-engine in Refs. [7,8].

Aero-engine has the character of complexity and high reliability, so it is difficult to obtain sufficient failure data within a short time and set up physical model. However, the data-driven method only depends on applicable historical data and statistical models, which makes the method more applicable to predict the RUL of aero-engine. So based on the existing research, a life prediction method for aero- engine is proposed in this work. Firstly, the state space model of aero-engine will be established. Secondly, RUL of aero-engine is analyzed, and expected RUL and RUL distribution are determined. Finally, the sate space model and KF algorithm are applied to an example of CFM-56 aero-engine, the expected RUL is predicted, and corresponding probability density function (PDF) and cumulative distribution function (CDF) are given.

## 1 Sate Space Model Approach for Aero-Engines

EGTM is time series data to measure the degree of performance deteriorations for aero-engine, and the sate space model approach (SSMA) offers a very general and powerful framework to operate with time series data. SSMA infers best historical estimates and forecast performance trends based on historical data. So RUL prediction for aero-engine will be studied with SSMA based on EGTM.

### 1.1 State space model

State space model differentiates state variables and observed variables to establish models, without loss of generality. Considering the following state space model

$$\begin{cases} y_t = f_\theta(x_t, v_t) \\ x_t = g_\theta(x_{t-1}, w_t) \end{cases} \tag{1}$$

where $f_\theta(\cdot)$ is the observation equation, which describes the mapping relationship between system state variables $x_t$ and observed variables $y_t$; $g(\cdot)$ the state equation, describing the evolution of system states with time; $\theta$ the model parameter; and $v_t$ and $w_t$ are the observation noise and the process noise, respectively

For aero-engine, in this work, the observed variable $y_t$ is EGTM, which is defined as

$$\mathrm{EGTM} = \mathrm{EGT}_{\mathrm{red}} - \mathrm{EGT}_{\mathrm{e}} \tag{2}$$

where $\mathrm{EGT}_{\mathrm{red}}$ is the red line temperature of aero-engine provided by manufacturers and $\mathrm{EGT}_{\mathrm{e}}$ the exhaust gas temperature of the engine with full power take-off in standard conditions.

It is provided that actual EGTM observed at in-service time $t$ is a function of the aero-engine unobservable degradation state $x_t$

$$y_t = \mathrm{EGTM}_t = x_t + v_t \tag{3}$$

where $v_t$ is the observation noise, including errors caused by model error, sensor noise and other factors. Let us suppose that observation noise is subject to Gaussian distribution, which means $v_t \sim N(0, V)$, where variance $V$ need to be estimated.

With regard to a state space model, its state sequence $x_{1:T}$ is often supposed to be first-order Markov chain, which means that the current state only depends on the system state at the previous moment and the current observation $y_t$ only depends on the current system state $x_t$. The structure diagram of state space model is shown as Fig. 1.

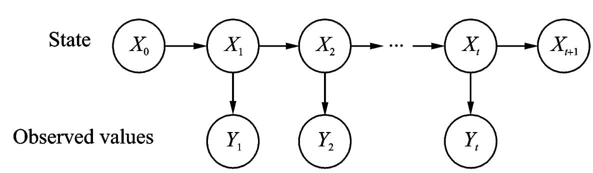


Fig. 1 State space model

### 1.2 Bayesian state estimation

Bayesian state estimation can obtain posterior probability distribution, including all the statistical information of state variables through two-step iteration of prediction and update. Some other statistical information like mean and variance can be obtained through posterior probability distribution. Given aero-engine observation sequence $y_{1:t}$, Bayesian state estimation utilizes all the current observation information to obtain the posterior distribution $\pi(x_t \mid y_{1:t})$ of unknown state. Suppose the initial aero-engine state at time $t=0$

is $x_0 \sim \pi(x_0)$, then for any in-service time $t \geqslant 1$[9]:

(1) Given the aero-engine state posterior distribution $\pi(x_{t-1} | y_{1:t-1})$ at time $t-1$, the one-step prediction of aero-engine state variable is

$$\pi(x_t \mid y_{1:t-1}) = \int \pi(x_t \mid x_{t-1}) \pi(x_{t-1} \mid y_{1:t-1}) \mathrm{d}x_{t-1} \tag{4}$$

The one-step prediction of aero-engine observation variable is

$$\pi(y_t \mid y_{1:t-1}) = \int \pi(y_t \mid x_t) \pi(x_t \mid y_{1:t-1}) \mathrm{d}x_t \tag{5}$$

Given the actual observations at time $t$, we can obtain posterior distribution of aero-engine state at time $t$ in accordance with Bayesian theorem

$$\pi(x_t \mid y_{1:t}) = \frac{\pi(y_t \mid x_t) \pi(x_t \mid y_{1:t-1})}{\pi(y_t \mid y_{1:t-1})} \tag{6}$$

Eqs. (4)—(6) provide the optimal state estimation under Bayesian theory framework.

### 1.3 State prediction

Given that state posterior distribution at in-service time $t$ is $\pi(x_t | y_{1:t})$, the probability distribution of aero-engine state $x_{t+k}$ after $k$-step prediction is[10]

$$\pi(x_{t+k} \mid y_{1:t}) = \int \pi(x_{t+k} \mid x_{t+k-1}) \pi(x_{t+k-1} \mid y_{1:t}) \mathrm{d}x_{t+k-1} \, (k \geqslant 1) \tag{7}$$

Given $x_t \sim N(a_t(0), R_t(0))$, so at time $(t+k)$, the state $x_{t+k}$ is

$$x_{t+k} \sim N(a_t(k), R_t(k)) \tag{8}$$

where $a_t(k) = G_{t+k} a_t(k-1)$, $R_t(k) = G_{t+k} R_t(k-1) G'_{t+k} + W_{t+k}$.

Similar to Bayesian state estimation, suppose the joint posterior distribution $\pi(x_{0:t}, \theta | y_{1:t})$ of state and parameter at time $t$ is approximated by a set of particles $\{(x_{0:t}, \theta_t)^{(i)}\}_{i=1}^N$, then the set of particles $\{x_{t+k}^{(i)}\}_{i=1}^N$ which approximates the state prediction distribution $\pi(x_{t+k} | y_{1:t})$ at time $t+k$ can be obtained by the following algorithm:

$k$-step state prediction

For $j = 1:k$

For $i = 1, \cdots, N$, sample $x_{t+j}^{(i)} \sim \pi(x_{t+j} | x_{t+j-1}^{(i)}, \theta_t^{(i)})$, make $x_{t:t+j}^{(i)} \triangleq (x_{t:t+j-1}^{(i)}, x_{t+j}^{(i)})$.

The state prediction distribution at time $t+k$ can be approximated by

$$\pi(x_{t+k} \mid y_{1:t}) \approx \hat{\pi}(x_{t+k} \mid y_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{t+k}^{(i)}}(x_{t+k}) \tag{9}$$

After obtaining new observation information, Bayesian state estimation would firstly update the posterior distribution of state and parameter by absorbing new observation information, and then update state prediction distribution. The prediction can be described and quantified by probability distribution.

## 2 RUL Prediction Model for Aero-engines

Through the analysis above, the $k$-step predictive state $x_{t+k}$ can be predicted based on the historical data $y_{1:t}$, but it is obvious that $x_{t+k}$ is uncertainty, and the value and trends of $x_{t+k}$ determines RUL, so in fact RUL is also not a definite value. In aero-engine engineering, RUL has two kinds of expressions: (1) Expected RUL the mean of RUL of failure; (2) RUL distribution. Since aero-engine has high reliability requirements and failures need to be avoided, it is difficult to manage aero-engine in practice merely based on expected RUL, and the failure probability at any time in the future is useful, so RUL distribution is a stronger guidance to management of aero-engine.

### 2.1 Expected RUL

As mentioned above, generally, $x_{t+k}$ decreases with in-service time. If $x_{t+k}$ reaches 0, it means that the performance degradation of the aero-engine is very serious, and the aero-engine will always be removed. So the expected RUL is equal to $k$ from current time $t$ to removal time $t+k$, and we can get the RUL at time $t$

$$\mathrm{RUL}_t = \{k \mid x_{t+k} = 0\} \tag{10}$$

### 2.2 RUL distribution

The aero-engine maybe fail at any time, and the probabilities of corresponding failure are different. In order to avoid the failure, the failure probability of the aero-engine at time $t+k$ should be estimated, and make $F_t(k)$ represent the failure probability. It is assumed that the failure time is $T$, and without loss of generality, for a monotone increasing degradation process with upper bound, we can get

$$F_t(k)=\Pr[T\leqslant t+k]=\Pr[x_{t+k}\leqslant 0]=\int_{-\infty}^{0}\pi(x_{t+k})\mathrm{d}x_{t+k} \tag{11}$$

where $\pi(x_{t+k})$ represents aero-engine degradation state distribution at time $t+k$.

According to Eq. (11), we can get CDF and PDF as follows

$$F_t(k)=\int_{-\infty}^{0}\pi(x_{t+k})\mathrm{d}x_{t+k}=\phi\left(\frac{-a_t(k)}{\sqrt{R_t(k)}}\right) \tag{12}$$

$$f_t(k)=\frac{\partial(F_t(k))}{\partial(k)}=\frac{1}{\sqrt{2\pi}\sqrt{R_t(k)}}\exp\left(-\frac{(a_t(k))^2}{2R_t(k)}\right) \tag{13}$$

## 3 Case Study

Fig. 2 shows the observed EGTM data collected during the whole in-service life of one CFM56-5B engine of certain airlines. The engine was removed after 2080 FC because of performance degradation.

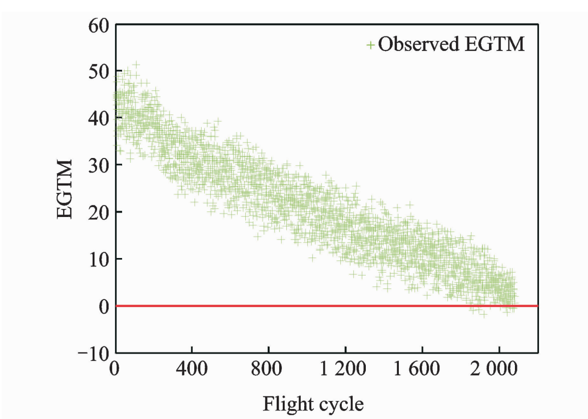


Fig. 2 Observed EGTM data of the CFM56-5B engine

### 3.1 Establishment of state space model

Set observation time as $T$ and observed EGMT as $X$, the correlation coefficient of $T$ and $X$ is[11]

$$\rho_{TX}=\frac{\sigma_{TX}}{\sigma_T\sigma_X} \tag{14}$$

where the covariance of two random variables $T$ and $X$ means $\mu_T$ and $\mu_X$, respectively, is given by

$$\sigma_{TX}=E(TX)-\mu_T\mu_X,\sigma_T^2=E(T^2)-\mu_T^2,\sigma_X^2=E(X^2)-\mu_X^2$$

The correlation coefficient $\rho_{TX}$ satisfies the inequality $-1\leqslant\rho_{TX}\leqslant 1$. It assumes a value of zero when $\sigma_{TX}=0$, where there is an exact linear dependency, say $\rho_{TX}=\pm 1$.

Based on the observed EGTM data, we can get $\rho_{TX}=-0.941\,6$, so the performance degradation process can approximately be regarded as linear degradation shown in Fig. 3. Therefore, we adopt a linear growth model as state equations to describe the aero-engine degradation path [9]

$$x_t=x_{t-1}+\mu_{t-1}+w_{x,t}\quad w_{x,t}\sim N(0,W_x) \tag{15}$$

$$\mu_t=\mu_{t-1}+w_{\mu,t}\quad w_{\mu,t}\sim N(0,W_\mu) \tag{16}$$

where $x_t$ is the actual degradation state, $\mu_t$ the rate of change for degradation state; $w_{x,t}$ and $w_{\mu,t}$ are the process noises subject to Gaussian distribution; and $W_x$ and $W_\mu$ are unknown and to be estimated. So we obtained a Gaussian linear state space model to describe aero-engine performance degradation based on EGTM observation sequence as follows

$$\begin{cases}Y_t=\boldsymbol{F}_t\boldsymbol{X}_t+v_t & v_t\sim N(0,V)\\ \boldsymbol{X}_t=\boldsymbol{G}_t\boldsymbol{X}_{t-1}+w_t & w_t\sim N(0,\boldsymbol{W})\end{cases} \tag{17}$$

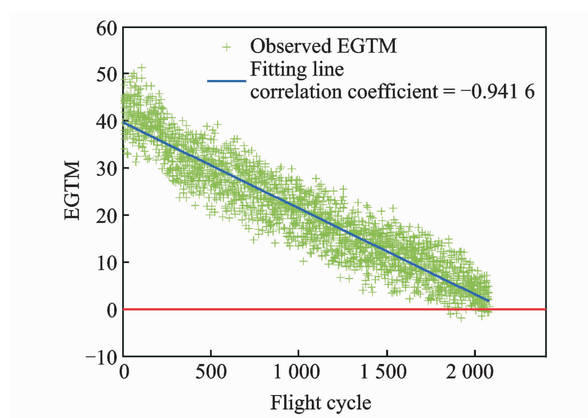


Fig. 3 Fitting line of observed EGTM data

where $\boldsymbol{X}_t=\begin{bmatrix}x_t\\ \mu_t\end{bmatrix}$, $Y_t=\mathrm{EGTM}_t$, $\boldsymbol{W}=\begin{bmatrix}W_x & 0\\ 0 & W_\mu\end{bmatrix}$, $\boldsymbol{G}_t=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$, $\boldsymbol{F}_t=[1\quad 0]$.

### 3.2 Estimation of parameters

Given the EGTM observation sequence of aero-engine and the state space mode above, the performance degradation state can be estimated and its trend can be predicted in accordance with

Bayesian reasoning. The statistical analysis software $R$ and its software package dlm[12] are applied to estimate the aero-engine performance degradation state and the unknown model parameters based on the historical EGTM data.

For example, based on EGTM data of historical 1 000 FC, firstly, the variance $V$, $W_x$ and $W_\mu$ can be estimated with function dlmMLE as follows

$$V=3.90, W_x=0.116, W_\mu=1.165\times10^{-7}$$

Then, with function dlmFilter, the corresponding parameters can be obtained:

$$\pi(x_{1\,000} \mid Y_{1:1\,000})=N(20.55, 0.0612^2), \pi(\mu_{1\,000} \mid Y_{1:1\,000})=N(-0.022, 0.825^2)$$

### 3.3 RUL prediction

After estimating the model parameters, with the function dlmForecast, the predictive distribution of degradation state after the period of $k$ can be obtained. For example, at $k=400$ and 800, the corresponding parameter of aero-engine, as shown in Fig. 4, are as follows

$$a_{1\,000}(400)=11.81, R_{1\,000}(400)=2.93^2$$

$$x_{1\,000+400}\sim N(11.81, 2.93^2)$$

$$F_{1\,000}(400)=\varphi\left(\frac{-11.81}{2.93}\right)=2.78\times10^{-5}$$

$$a_{1\,000}(800)=3.07, R_{1\,000}(800)=4.62^2$$

$$x_{1\,000+800}\sim N(3.07, 4.62^2)$$

$$F_{1\,000}(800)=\phi\left(\frac{-3.07}{4.62}\right)=0.253\,2$$

The blue curves in Fig. 4 are PDF at $k=400$ and 800, respectively, and the sizes of red shadows indicates CDF, such as $F_{1\,000}(400)$ and $F_{1\,000}(800)$. From Fig. 4, the state $x_{1\,000+940}$ is equal to zero, so $\mathrm{RUL}_{1\,000}=940$ FC, and the actual $\mathrm{RUL}_{1\,000}$ is 2 080−1 000=1 080 FC, so the prediction error is 12.96%. And if the traditional linear regression is adopted to predict the RUL as shown in Fig. 5, the expected $\mathrm{RUL}_{1\,000}=1\,899-1\,000=899$ FC, so the corresponding prediction error is 16.76%.

With the increase of aero-engine flight cycles, EGTM is more and more close to zero, so the accuracy of prediction will be getting higher and higher. For example, when the aero-engine reaches 1 500 FC, the predicted state after 300 and 500 FC will be shown in Fig. 6. And the expected $\mathrm{RUL}_{1\,500}$ is 625 FC, and the actual $\mathrm{RUL}_{1\,500}$ is 580 FC, so the prediction error is 7.76%. So the accuracy of RUL prediction increases with more observed EGTM data. And if the traditional linear regression is adopted to predict the RUL as shown in Fig. 7, the expected $\mathrm{RUL}_{1\,500}=2\,028-1\,500=528$ FC, so the corresponding prediction error is 9.00%.

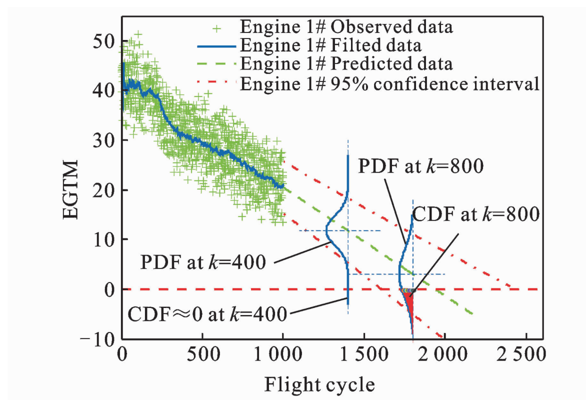


Fig. 4 State estimation and prediction (1 000 FC) with SSMA

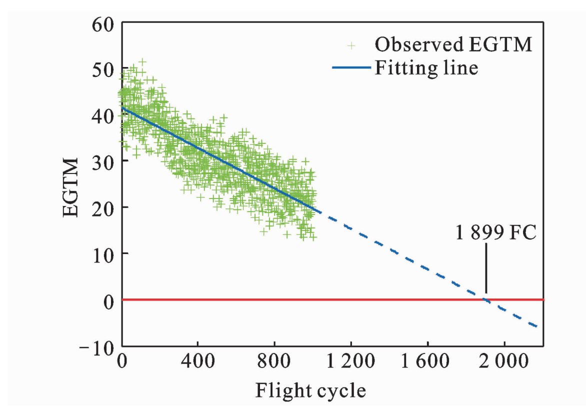


Fig. 5 State estimation and prediction (1 000 FC) with linear regression

Fig. 8 describes the distributions of RUL=200 FC, 400 FC, and 800 FC at in-service time $t=1\,000$ FC and 1 500 FC, respectively. And it can be seen that the variances of distribution are more and more obvious as RUL increases, which is consistent with the actual situation that the more steps of prediction is, the more uncertainty of prediction will be. Similarly, it can be seen that the prediction is more accurate as close to the time when the engine is removed.

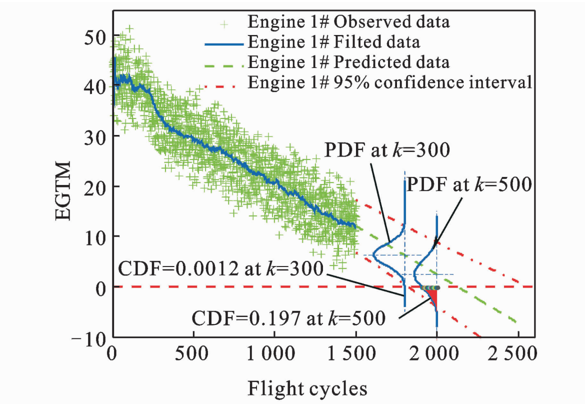


Fig. 6 State estimation and prediction (1 500 FC) with SSMA

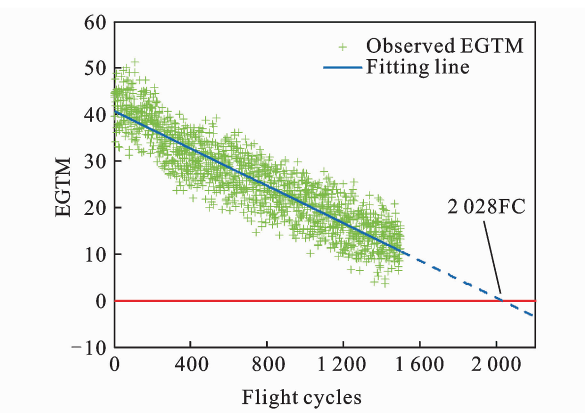


Fig. 7 State estimation and prediction (1 500 FC) with linear regression

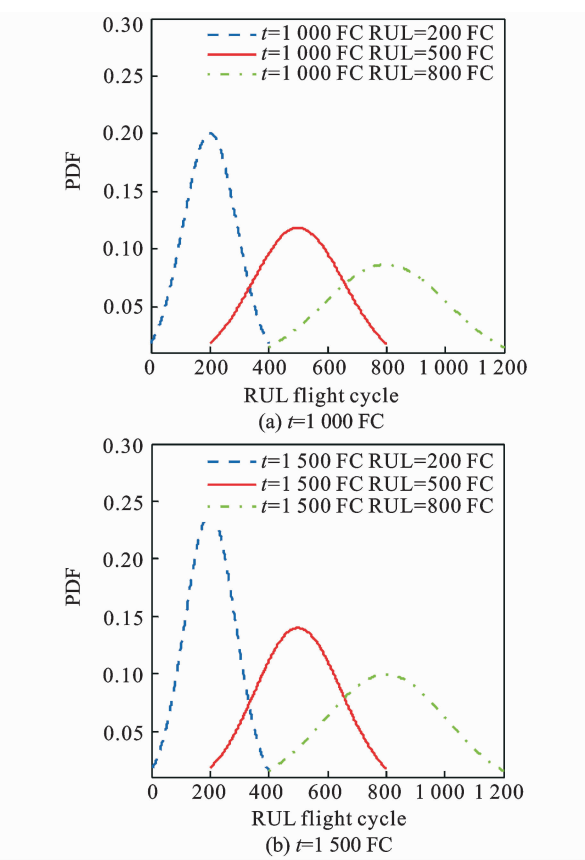


Fig. 8 PDF of aero-engine RUL at time $t=1\ 000$ FC and $t=1\ 500$ FC

Through the analysis above, we can find that CDF will increase with aero-engine flight cycles, and the CDF of aero-engine from the time $t=1\ 000$ FC and $t=1\ 500$ FC are described in Fig. 9, and the right part of Fig. 9 is a zoom of the left from $k=0$ to $k=420$. The CDF from $t=1\ 500$ FC is more than the CDF from $t=1\ 000$ FC, which is consistent with the actual situation that when the steps of prediction is the same, the more the in-service life is, the higher the prediction risk will be.

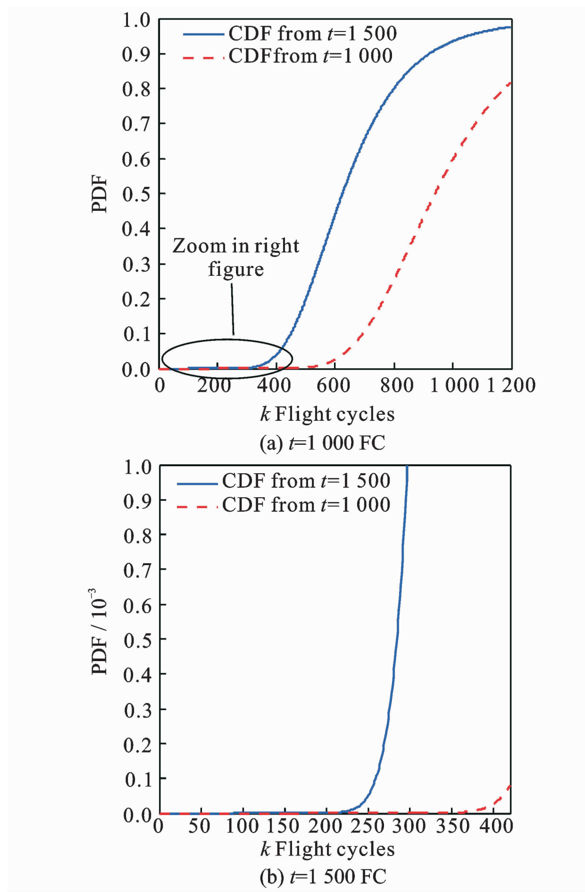


Fig. 9 CDF of aero-engine from the time $t=1\ 000$ FC and $t=1\ 500$ FC

## 4 Conclusions

We present our studies on RUL prediction method for aero-engine, and the conclusions are as follows:

(1) The state space model approach is established, including parameters such as EGTM and degradation rate, etc, which can better reflect the performance degradation process of aero-engine.

(2) The accuracy of RUL prediction increa-

ses with more observed EGTM data through KF, and it is more accurate than traditional linear regression.

(3) Real-time update of performance degradation model parameters can be achieved through Bayesian method. The more steps the prediction conducts, the more uncertainty and risk it will face.

**References:**

[1] ZHANG Haijun. Research on methods of performance synthetic estimation and CBM decision-making models for aero-engine[D]. Nanjing University of Aeronautics and Astronautics, 2007. (in Chinese)

[2] MALINOWSKI S, CHEBEL M B, ZERHOUNI N. Remaining useful life estimation based on discriminating shapelet extraction[J]. Reliability Engineering & System Safety. 2015,142(10):279-288.

[3] TAN X, QIU J, LIU G, et al. A novel approach of testability modeling and analysis for PHM systems based on failure evolution mechanism[J]. Chinese Journal of Aeronautics. 2013(3):766-776.

[4] GARCÍA N P J, GARCIA G E, Sánchez L F, et al. Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability[J]. Reliability Engineering & System Safety. 2015,138(7):219-231.

[5] Li Y G, NILKITSARANONT P. Gas turbine performance prognostic for condition-based maintenance [J]. Applied Energy. 2009,86(10):2152-2161.

[6] ZAIDAN M A, MILLS A R, HARRISON R F, et al. Gas turbine engine prognostics using Bayesian hierarchical models: A variational approach[J]. Mechanical Systems and Signal Processing. 2016,70/71 (S):120-140.

[7] LIU J, ZHANG M, ZUO H, et al. Remaining useful life prognostics for aeroengine based on superstatistics and information fusion[J]. Chinese Journal of Aeronautics. 2014,27(5):1086-1096.

[8] REN S H, ZUO H F. A combined prediction method for the residual life of civil aviation engines based on performance degradation [J]. Mech Sci Technol Aerosp Eng. 2011;30(1):23-29.

[9] WEST M, HARRISON J. Bayesian forecasting and dynamic models[M]. USA: Springer, 1997.

[10] DIACONIS P. Sequential Monte Carlo Methods in Practice[J]. Journal of the American Statistical Association. 2003,98(462):496-497.

[11] WALPOLE R E, MYERS R H, MYERS S L, et al. Probability & Statistics for Engineers &; Scientists [M]. Pearson Education International, 2007.

[12] G. P, S. P, P. C. Robert Gentleman Kurt Hornik Giovanni Parmigiani[M]. Springer- Verlag: Springer-Verlag , 2009.

Dr. **Cai Jing** was born in 1976. He is currently an associate professor in Nanjing University of Aeronautics and Astronautics (NUAA). His research interests include reliability statistics, maintenance theory, prognostic and health management (PHM). From 1995. 09 to 1999. 06, he studied for B. A. in NUAA. From 1999. 08 to 2001. 08, he worked as a designer in Nanjing Mechanics Electronics Hydraulics Engineering Research Center. From 2001. 08 to 2007. 06, he studied for PH. D. in NUAA. From 2007. 07, he works as a teacher in NUAA. He worked on temporary post in Shanghai Aircraft Customer Service Co. , Ltd. From 2012. 09 to 2012. 08. He was studying on condition based maintenance in University of Toronto as a visiting scholar from 2015. 11 to 2016. 11.

Ms. **Zhang Li** was born in 1990, She is currently a postgraduate in Nanjing University of Aeronautics and Astronautics (NUAA).

Mr. **Dong Ping** was born in 1991, He is currently a postgraduate in Nanjing University of Aeronautics and Astronautics (NUAA).

(Executive Editor: Zhang Bei)