

Process fault prognosis using a fuzzy-adaptive unscented Kalman predictor

Xuemin Tian¹, Yu Ping Cao¹ and Sheng Chen^{2,*},[†]

¹*College of Information and Control Engineering, China University of Petroleum (East China), Shandong 257061, People's Republic of China*

²*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.*

SUMMARY

By monitoring the future process status via information prediction, process fault prognosis is able to give an early alarm and therefore prevent faults, when the faults are still in their early stages. A fuzzy-adaptive unscented Kalman filter (FAUKF)-based predictor is proposed to improve the tracking and forecasting capability for process fault prognosis. The predictor combines the strong tracking concept and fuzzy logic idea. Similar to the standard adaptive unscented Kalman filter (AUKF) that employs an adaptive parameter to correct the estimation error covariance, a Takagi–Sugeno fuzzy logic system is designed to provide a better adaptive parameter for smoothing this regulation. Compared with the standard AUKF, the proposed FAUKF has the same strong tracking ability but does not suffer from the drawback of serious tracking fluctuation. Two simulation examples demonstrate the effectiveness of the proposed predictor. Copyright © 2011 John Wiley & Sons, Ltd.

Received 23 November 2010; Revised 21 January 2011; Accepted 20 February 2011

KEY WORDS: process fault prognosis; unscented Kalman predictor; fuzzy logic; error covariance

1. INTRODUCTION

Process safety and reliability are critically important for large-scale and highly integrated modern industry. Faults can cause long-term disturbances, influence normal operation, and may even lead to large economical loss and ecological crisis. The development of fault detection and diagnosis techniques began in 1970s [1]. In the early stage, fault detection and diagnosis is designed to monitor process status, to detect fault, and to isolate the causes of faults. A fault alarm is given when the current measurements or statistics lie outside the normal-operation region. By this time, however, the fault may have already developed into a serious state. In order to minimize the loss caused by faults and to guarantee that the process operates in a safe, stable, and optimal status, it is important to be able to predict the future process status. Future prediction can help to detect the faults as early as possible and to deal with the faults when the current measurements are still within the normal-operation region. Moreover, predictive maintenance based on fault prognosis can reduce maintenance and production costs. Hence, process fault prognosis has become an active research area [2, 3]. Lu and Saeks [4] investigated failure prediction for an online maintenance system, which may be considered as the earliest example of fault prognosis. However, the development of fault prognosis techniques have been slow, owing to the difficulty in predicting the faults that have small amplitude and inexplicit characteristics.

*Correspondence to: Sheng Chen, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

[†]E-mail: sqc@ecs.soton.ac.uk

Recently, fault prognosis has attracted renewed interests from both academic and industrial communities with the development of various fault detection and diagnosis techniques, which may be divided into three main categories: model-based approaches [5–9], signal processing-based approaches [10–17], and intelligent approaches [18, 19]. Model-based approaches may be considered as the most effective by using system kernel information such as the inherent structure and first-principles. In particular, Yang [5, 6] used Kalman filter (KF) to predict one-step-ahead as well as two-step-ahead states for prognosing ageing failure, and provided a Monte–Carlo simulation as well as an actual experiment on a DC motor for predictive maintenance. Juricek *et al.* [7] applied KF to forecast the future outputs and to signal the impending emergency limit violations. Both these works employed KF and, therefore, they require the assumption that the process involved can be accurately described by a linear stochastic state space model. For nonlinear fault prognosis, we used the support vector machine to forecast the future Kalman innovation, and applied the innovation to predict multi-step-ahead outputs [8]. Chen *et al.* [9] proposed a new particle predictor for nonlinear time-varying systems, in which the process fault model was simplified into a parameter vector.

Information forecasting is the key issue for fault prognosis. By estimating the states and fault parameters, KF can be used to track the process running trend and to predict the future outputs. At the initial stages of a fault process, however, the fault amplitude is small and, therefore, it is difficult to estimate the states and fault parameters accurately. For filtering problems, the innovation can be used as feedback to correct the estimation. But for forecasting problems, the innovation is unavailable, and such a correction cannot take place. Thus, predictors with accurate estimation ability and strong tracking ability are essential for fault prognosis. Extended KF (EKF) is the most widely used filtering method for nonlinear dynamic system. EKF however has poor robustness against model mismatch as well as has poor tracking ability to sudden changes in the steady state. For this reason, Zhou *et al.* [20] proposed a concept of strong tracking filter (STF), and presented a suboptimal fading EKF (SFEKF), which employs a suboptimal fading factor to adjust the error covariance using the innovation. Jwo and Wang [21] proposed an adaptive fuzzy algorithm to calculate the smoothing factor in the SFEKF for the GPS navigation application. However, first-order linearization in EKF can introduce large error and high instability as a consequence of propagating the covariance information through a linear transformation. Specifically, the EKF-based approach for nonlinear process fault prognosis has the following three drawbacks: (i) the linearized transformation is reliable only if the error propagation can be well approximated by a linear function, which is generally not the case for a nonlinear process; (ii) this linearization can be applied only if the Jacobian matrix exists; and (iii) calculating the Jacobian matrix is a very difficult and costly as well as error-prone process.

Recently, unscented KF (UKF), proposed by Julier *et al.* [22, 23], has attracted wide interests [24, 25]. UKF uses a minimal set of sigma points to approximate the state's mean and covariance. These sigma points are propagated through a nonlinear transformation and, therefore, it is not necessary to compute Jacobian matrices in UKF. In other words, the mean and covariance information are propagated through nonlinear transformations in an UKF, which is more accurate and easier to implement. Consequently, the UKF-based approach does not suffer from the above-mentioned drawbacks of the EKF method. In particular, the UKF matches the mean correctly up to the second order in Taylor series and predicts the covariance correctly up to the third order, while the EKF can only approximate the mean up to the first order [22, 23]. UKF has been used for nonlinear estimation and process monitoring [26–30]. We also combined the UKF with the concept of STF to improve its tracking ability to states and fault parameters, so as to improve fault prognosis performance [31]. Although the stability of the UKF was not addressed in the original work of Julier *et al.* [22], this important issue has been discussed in [24, 32, 33]. The estimation error behavior of the UKF was analyzed in [24] for nonlinear stochastic processes with linear measurement equations. It was proved that by adding an appropriately chosen positive-definite matrix to the covariance matrix, the stability of the UKF can be ensured and the estimation error of the UKF remains bounded under certain conditions. This result was then extended to the more general case involving nonlinear measurement equations in [32]. Based on a similar modification to the covariance matrix, the authors of [33] also proved that the estimation error of the UKF

remains bounded under certain conditions. Stability analysis of the UKF remains an active research area and is beyond the scope of this study.

Adaptive UKF (AUKF) based on the concept of STF [20, 22, 23] can capture process changes quickly and offers much superior tracking performance over UKF, while possessing all the advantages of UKF. This makes it a powerful technique for application to nonlinear process fault prognosis. However, we note that the introduction of the suboptimal fading factor in AUKF, while enhancing the tracking ability, can increase the tracking fluctuation of the process state and fault parameters, especially when the process is in steady state. The fluctuation of the state and fault parameter estimation in turn may degrade the prediction performance. Our motivation is to prevent this overregulation to the error covariance by the AUKF, while maintaining its fast tracking ability. In this contribution, a fuzzy-adaptive UKF (FAUKF) is proposed to alleviate the unnecessary regulation of the error covariance, leading to an enhanced prediction accuracy without any sacrifice in the tracking performance. Specifically, a Takagi–Sugeno fuzzy logical system [34] is designed to calculate a fuzzy-adaptive parameter, which depends on the membership functions of the normal and failure states. A predictor based on the proposed FAUKF is applied for fault prognosis. We demonstrate that the proposed FAUKF has the same strong tracking ability as the standard AUKF while avoiding the unnecessary regulation drawback of the latter, resulting in an enhanced performance in fault prognosis. The simulation results obtained confirm the effectiveness of this FAUKF for process fault prognosis application.

The remainder of this contribution is organized as follows. Section 2 briefly describes the problem of nonlinear process fault prognosis. After introducing the AUKF-based approach, the proposed FAUKF is presented in Section 3 for the application to nonlinear process fault prognosis. Monte–Carlo simulations involving two nonlinear stochastic systems are given in Section 4 to compare the performance of the proposed FAUKF with that of the standard AUKF. The concluding remarks are summarized in Section 5.

2. PROBLEM FORMULATION

Consider the following nonlinear dynamic system:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) + \mathbf{d}(k) + \boldsymbol{\varepsilon}(k), \quad (1)$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \quad (2)$$

where $\mathbf{u}(k) \in \mathbf{R}^{l \times 1}$, $\mathbf{x}(k) \in \mathbf{R}^{n \times 1}$, and $\mathbf{y}(k) \in \mathbf{R}^{m \times 1}$ are the vectors of measured inputs, state variables, and measured outputs, respectively, $\mathbf{d}(k) \in \mathbf{R}^{n \times 1}$ denotes the vector of unmeasurable fault variables, while $\boldsymbol{\varepsilon}(k) \in \mathbf{R}^{n \times 1}$ and $\mathbf{v}(k) \in \mathbf{R}^{m \times 1}$ are the process and measurement noise vectors, which are uncorrelated with one another and obey the zero-mean normal distributions with the covariance matrices $\mathbf{Q}(k) = E[\boldsymbol{\varepsilon}(k)\boldsymbol{\varepsilon}^T(k)]$ and $\mathbf{R}(k) = E[\mathbf{v}(k)\mathbf{v}^T(k)]$, respectively. The nonlinear vector function $\mathbf{f}(\bullet): \mathbf{R}^{(n+l) \times 1} \rightarrow \mathbf{R}^{n \times 1}$ in the state equation (1) is assumed to be continuously differentiable with respect to $\mathbf{x}(k)$, while the matrix $\mathbf{H}(k) \in \mathbf{R}^{m \times n}$ specifies the linear measurement Equation (2).

In general, process fault prognosis is suitable for the slowly changing behavior caused by fault. The root cause of fault may be step or ramp change in the unmeasured disturbance caused by $\mathbf{d}(k)$. Yang used the exponential model to describe ageing failure [5, 6], while Juricek *et al.* adopted a pseudo-disturbance term to describe step or ramp changes [7]. In the work [9], a time-varying equivalent parameter vector was used to represent the difference between the fault process and the known normal model. We will also adopt the approach of Juricek *et al.* [7] by considering the pseudo-fault model

$$\mathbf{d}(k) = \boldsymbol{\Gamma}(k, k_0)\boldsymbol{\theta}, \quad (3)$$

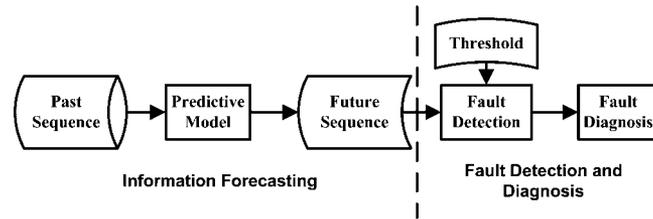


Figure 1. Schematic diagram of process fault prognosis program.

where $\theta \in \mathbb{R}^{q \times 1}$ denotes the vector of fault parameters, k_0 is the beginning time of the fault, and $\Gamma(k, k_0)$ is a matrix function of time. A pseudo-fault model for second-order system is given by Juricek *et al.* [7]

$$\mathbf{d}(k) = \Gamma(k, k_0)\theta = \begin{bmatrix} \theta_{1,0} + \theta_{1,s}T_s(k - k_0) \\ \theta_{2,0} + \theta_{2,s}T_s(k - k_0) \end{bmatrix}$$

with

$$\Gamma(k, k_0) \triangleq \begin{bmatrix} 1 & 0 & T_s(k - k_0) & 0 \\ 0 & 1 & 0 & T_s(k - k_0) \end{bmatrix}, \quad \theta \triangleq [\theta_{1,0} \ \theta_{2,0} \ \theta_{1,s} \ \theta_{2,s}]^T,$$

where T_s is the sampling period. The pseudo-fault model (3) has no physical interpretation, and it simply describes the process changes caused by faults.

Process fault prognosis techniques use the collected data and the information prediction model to forecast the future status and trend, and then apply fault detection and diagnosis methods to determine whether the monitored variables or statistics will be outside the normal-operating region in the future. Thus, the prognosis program can be summarized in two steps: (1) information forecasting and (2) fault detection and diagnosis, as illustrated in Figure 1. The ability to accurately predict future outputs is very important, since process fault prognosis techniques use the future prediction values to judge the process operating status. Define the measurement innovation or the predicted output error as $\tilde{\mathbf{y}}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k)$, where $\hat{\mathbf{y}}(k)$ is the prediction for $\mathbf{y}(k)$. The autocorrelation function of the predicted output error, defined as

$$\mathbf{R}_{\tilde{\mathbf{y}}}(k + j, k) = E[\tilde{\mathbf{y}}(k + j)\tilde{\mathbf{y}}^T(k)], \quad 1 \leq j \leq N \quad (4)$$

can be used as a forecast measure. If $\mathbf{R}_{\tilde{\mathbf{y}}}(k + j, k) = \mathbf{0}_{m \times m}$, $1 \leq j \leq N$, the predicted output error sequence $\{\tilde{\mathbf{y}}(k), \tilde{\mathbf{y}}(k + 1), \dots, \tilde{\mathbf{y}}(k + N)\}$ is uncorrelated and has a zero-mean Gaussian distribution. Then the error is only caused by the Gaussian measurement noise. If, on the other hand, $\mathbf{R}_{\tilde{\mathbf{y}}}(k + j, k) \neq \mathbf{0}_{m \times m}$, the error sequence is correlated. Then the errors fluctuate around a nonzero constant or increase gradually. A key problem of process fault prognosis is to find the future output predictions which ensure $\mathbf{R}_{\tilde{\mathbf{y}}}(k + j, k) = \mathbf{0}_{m \times m}$.

3. A FUZZY-ADAPTIVE UNSCENTED KALMAN PREDICTOR

In the area of adaptive KF and UKF, most studies focused on the process noise covariance estimation [35–37]. However, Zhou *et al.* [20] proposed the SFEFK, in which the error covariance was regulated with the innovation. The adaptive algorithm of Zhou *et al.* [20] was then integrated into the UKF by the work [38]. We consider the AUKF, which introduces an adaptive parameter to correct the error covariance based on innovations [31]. In this AUKF, the filter gain is regulated so that the filter output can better track the actual process state. While enhancing the tracking ability, the introduction of this adaptive parameter may cause the estimated values of the process state and fault parameters to fluctuate. The fluctuation of the fault parameter estimation in turn may

affect the prediction performance. In order to prevent overregulation and to smooth estimation, we combine a fuzzy logic method with the AUKF to form a FAUKF-based predictor.

3.1. Adaptive unscented Kalman filter

3.1.1. *Unscented Kalman filter.* The UKF [22, 23] uses the unscented transform (UT) to propagate the state’s mean and covariance. Unlike EKF, linearization and computation of Jacobian matrices are unnecessary in UKF. In particular, linearization is done at a single point in EKF, while UKF samples nonlinear behavior at multiple sigma points. Furthermore, UKF matches the mean correctly up to the second order in Taylor series, while EKF only approximates it up to the first order [22, 23]. Consider the nonlinear system given in (1) and (2), in which $\mathbf{d}(k) = \mathbf{0}$ is assumed. The standard UKF algorithm can be summarized as follows.

(a) *Prediction:* Form a set of sigma points $\tilde{\mathbf{X}}_i(k-1) \in \mathbb{R}^{n_s \times 1}$ and their weights w_i , where $0 \leq i \leq 2n_s$ and $n_s = 2n + m$, for the enlarged state variable $\tilde{\mathbf{x}}_s(k-1) = [\mathbf{x}^T(k-1) \boldsymbol{\varepsilon}^T(k-1) \mathbf{v}^T(k-1)]^T$. Specifically,

$$\tilde{\mathbf{X}}_i(k-1) = \begin{cases} \tilde{\mathbf{x}}_s(k-1), & i = 0, \\ \tilde{\mathbf{x}}_s(k-1) + \mathbf{A}_i(k-1), & i = 1, 2, \dots, n_s, \\ \tilde{\mathbf{x}}_s(k-1) - \mathbf{A}_{i-n_s}(k-1), & i = n_s + 1, n_s + 2, \dots, 2n_s, \end{cases} \quad (5)$$

$$w_i = \begin{cases} \frac{\eta}{n_s + \eta}, & i = 0, \\ \frac{1}{2(n_s + \eta)}, & i = 1, 2, \dots, 2n_s. \end{cases} \quad (6)$$

Here $\eta \in \mathbb{R}$ is determined by the distribution of the state $\mathbf{x}(k)$, and $\eta = 3 - n_s$ for the Gaussian distribution. Moreover, $\tilde{\mathbf{x}}_s(k-1) = [\hat{\mathbf{x}}^T(k-1) \mathbf{0}_{1 \times n} \mathbf{0}_{1 \times m}]^T$, $\hat{\mathbf{x}}(k-1)$ is the estimate of $\mathbf{x}(k-1)$, and \mathbf{A}_i denotes the i th column of the matrix $\mathbf{A} \in \mathbb{R}^{n_s \times n_s}$ which satisfies

$$\mathbf{A}(k-1)\mathbf{A}^T(k-1) = (n_s + \eta)\hat{\mathbf{P}}_s(k-1), \quad (7)$$

where $\hat{\mathbf{P}}_s(k-1)$ is the enlarged error covariance defined by

$$\hat{\mathbf{P}}_s(k-1) = \begin{bmatrix} \hat{\mathbf{P}}(k-1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}(k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}(k-1) \end{bmatrix} \in \mathbb{R}^{n_s \times n_s} \quad (8)$$

with $\hat{\mathbf{P}}(k-1) \in \mathbb{R}^{n \times n}$ denoting the state error covariance. Then, the predicted state $\hat{\mathbf{x}}(k|k-1) \in \mathbb{R}^{n \times 1}$ and the related error covariance $\hat{\mathbf{P}}(k|k-1) \in \mathbb{R}^{n \times n}$ are computed according to

$$\hat{\mathbf{X}}_i(k) = \mathbf{f}(\tilde{\mathbf{X}}_i^x(k-1), \mathbf{u}(k-1)) + \tilde{\mathbf{X}}_i^\varepsilon(k-1), \quad (9)$$

$$\hat{\mathbf{x}}(k|k-1) = \sum_{i=0}^{2n_s} w_i \hat{\mathbf{X}}_i(k), \quad (10)$$

$$\hat{\mathbf{P}}(k|k-1) = \sum_{i=0}^{2n_s} w_i (\hat{\mathbf{X}}_i(k) - \hat{\mathbf{x}}(k|k-1))(\hat{\mathbf{X}}_i(k) - \hat{\mathbf{x}}(k|k-1))^T, \quad (11)$$

where $\tilde{\mathbf{X}}_i^x(k-1) \in \mathbb{R}^{n \times 1}$ and $\tilde{\mathbf{X}}_i^\varepsilon(k-1) \in \mathbb{R}^{n \times 1}$ are the sigma-point components which correspond to the process state and noise variables, respectively. More specifically, $\tilde{\mathbf{X}}_i^x(k-1)$ consists of the first n elements of $\tilde{\mathbf{X}}_i(k-1)$, while $\tilde{\mathbf{X}}_i^\varepsilon(k-1)$ contains the $(n+1)$ th to $(2n)$ th elements of $\tilde{\mathbf{X}}_i(k-1)$.

(b) *Update*: The predicted output $\hat{\mathbf{y}}(k) \in \mathbf{R}^{m \times 1}$, the corrected state estimate $\hat{\mathbf{x}}(k) \in \mathbf{R}^{n \times 1}$ and the corresponding error covariance $\hat{\mathbf{P}}(k)$ are updated according to

$$\hat{\mathbf{y}}(k) = \mathbf{H}(k)\hat{\mathbf{x}}(k|k-1), \quad (12)$$

$$\mathbf{K}(k) = \hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)(\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}, \quad (13)$$

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)(\mathbf{y}(k) - \hat{\mathbf{y}}(k)), \quad (14)$$

$$\hat{\mathbf{P}}(k) = \hat{\mathbf{P}}(k|k-1) - \mathbf{K}(k)(\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))\mathbf{K}^T(k). \quad (15)$$

3.1.2. *Adaptive unscented Kalman filter*. The error covariance $\hat{\mathbf{P}}(k|k-1)$ for the standard UKF is computed in open loop. The predicted output error $\tilde{\mathbf{y}}(k)$ can be used as feedback to correct the error covariance adaptively, and AUKF introduces a regulation parameter $\lambda(k) \geq 1$ to regulate the error covariance, yielding the actual error covariance used

$$\hat{\mathbf{P}}(k|k-1) = \lambda(k)\hat{\mathbf{P}}(k|k-1). \quad (16)$$

Thus, for the AUKF, equations (13) and (15) are replaced by

$$\mathbf{K}(k) = \hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)(\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}, \quad (17)$$

$$\hat{\mathbf{P}}(k) = \hat{\mathbf{P}}(k|k-1) - \mathbf{K}(k)(\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))\mathbf{K}^T(k). \quad (18)$$

The adaptive regulation parameter $\lambda(k)$ is designed to fulfill $\mathbf{R}_{\tilde{\mathbf{y}}}(k+j, k) = \mathbf{0}_{m \times m}$, $1 \leq j \leq N$. In the study [20], $\mathbf{R}_{\tilde{\mathbf{y}}}(k+j, k) = \mathbf{0}_{m \times m}$ is used as the performance measure for STF. When this measure is met, the error sequence is orthogonal. This measure is referred to as the orthogonality principle, and it ensures that the predictor is insensitive to the initial states and the system model mismatch as well as has the strong tracking ability to suddenly change the system status even in the steady state.

Define $\mathbf{P}_{\tilde{\mathbf{y}}}(k) = E[\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}^T(k)] \in \mathbf{R}^{m \times m}$. From the system model and the UKF algorithm described above, the autocorrelation $\mathbf{R}_{\tilde{\mathbf{y}}}(k+j, k)$ given in (4) becomes [31]

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{y}}}(k+j, k) &= \mathbf{H}(k+j)\boldsymbol{\beta}(k+j-1)\mathbf{F}(k+j-1) \left(\prod_{i=k+j-2}^k (\mathbf{I} - \mathbf{K}(i+1)\mathbf{H}(i+1))\boldsymbol{\beta}(i)\mathbf{F}(i) \right) \\ &\quad \times \lambda(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)(\mathbf{I} - (\lambda(k)\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}\mathbf{P}_{\tilde{\mathbf{y}}}(k)), \end{aligned} \quad (19)$$

where \mathbf{I} denotes the identity matrix of appropriate dimension, while $\mathbf{F}(k+j-1)$ denotes the Jacobian matrix of the nonlinear function $\mathbf{f}(\bullet)$ and $\boldsymbol{\beta}(k) = \text{diag}\{\beta_1(k), \dots, \beta_n(k)\}$ is a diagonal matrix, which have no effect on the following derivation of $\lambda(k)$. Setting $\mathbf{R}_{\tilde{\mathbf{y}}}(k+j, k) = \mathbf{0}_{m \times m}$ leads to

$$\mathbf{I} - (\lambda(k)\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k))^{-1}\mathbf{P}_{\tilde{\mathbf{y}}}(k) = \mathbf{0}_{m \times m}. \quad (20)$$

From (20), a suboptimal $\lambda(k)$ can be found as follows [20]:

$$\lambda(k) = \max \left\{ \frac{\text{tr}[\mathbf{P}_{\tilde{\mathbf{y}}}(k) - \zeta\mathbf{R}(k)]}{\text{tr}[\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k)]}, 1 \right\}, \quad (21)$$

where $\text{tr}[\bullet]$ denotes the matrix trace operator and $\zeta \geq 1$ is a smoothing factor. Define $\mathbf{V}(k) = \tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}^T(k)$ and give $0 \leq \rho \leq 1$. Then the matrix $\mathbf{P}_{\tilde{\mathbf{y}}}(k)$ can be approximated using

$$\mathbf{P}_{\tilde{\mathbf{y}}}(k) = \begin{cases} \mathbf{V}(k), & k=1, \\ \frac{\rho\mathbf{V}(k-1) + \mathbf{V}(k)}{1+\rho}, & k \geq 2. \end{cases} \quad (22)$$

3.2. Takagi–Sugeno fuzzy logic system

For the above AUKF, the smoothing factor ζ is used to soften the regulation of $\lambda(k)$. However, ζ is a time-invariant constant, which cannot adapt to the process operating status. In order to adapt the regulation according to the actual process operating status, a Takagi–Sugeno fuzzy logic system [34] is adopted to produce a fuzzy-adaptive regulation parameter $\lambda^*(k)$ for replacing $\lambda(k)$. This fuzzy logic system is designed to fulfill the following requirement: when the process $\mathbf{y}(k)$ is in a normal-operating status, correcting the error covariance $\hat{\mathbf{P}}(k|k-1)$ is undesired; while when the status of $\mathbf{y}(k)$ is failure, it is necessary to regulate $\hat{\mathbf{P}}(k|k-1)$. Specifically, this Takagi–Sugeno fuzzy logic system takes the measured variable vector $\mathbf{y}(k)$ as its input and produces the fuzzy-adaptive parameter $\lambda^*(k)$ as its output. Let the fuzzy sets ‘Normal’ and ‘Failure’ denote the normal- and failure-operating states, respectively. Furthermore, let $\mu_N(\mathbf{y}(k))$ be the membership function of the fuzzy set ‘Normal’ and $\mu_F(\mathbf{y}(k))$ be the membership function of the fuzzy set ‘Failure’. Design the fuzzy inference rules as follows:

- If $\mathbf{y}(k)$ is ‘Normal’, then $\lambda_N^*(k) = 1$,
- If $\mathbf{y}(k)$ is ‘Failure’, then $\lambda_F^*(k) = \lambda(k)$.

Then, the output $\lambda^*(k)$ of the fuzzy logic system is the weighted average

$$\lambda^*(k) = \mu_N(\mathbf{y}(k))\lambda_N^*(k) + \mu_F(\mathbf{y}(k))\lambda_F^*(k). \tag{23}$$

To explain the selection of membership function, consider that $\mathbf{y}(k) = y(k)$ is univariate. When the process is in the ‘Normal’ status, the variable $y(k)$ can be described by the normal distribution $N(\mu, \sigma^2)$ with the mean μ and the standard deviation σ , according to the central limit theorem. The collected data will be centralized around μ . The further away from μ , the fewer data are. The cumulative probability that the data lie within $(\mu - 3\sigma, \mu + 3\sigma)$ is 99.7%, and the event that the data lie outside the region $(\mu - 3\sigma, \mu + 3\sigma)$ is a very small probability event. In the statistical process control (SPC) [39], it is considered that such a small probability event will not occur normally. If it occurs, a fault may exist and the process may be in a fault status. This is the so-called 3σ theory. In the work [40], control limit for the SPC Shewhart charts is designed based on the 3σ theory to signal faults. Since the measurement noise and the process noise will make the data noisy, the boundary between the normal status and the failure status is fuzzy. The closer the data are to μ , the more possibly that the process belongs to ‘Normal’. On the other hand, the further the data are away from μ , the more possibly that the process belongs to ‘Failure’. Thus, the cumulative probability distribution function of $y(k)$ can be selected as the membership function.

In general, assume that the probability distribution of $\mathbf{y}(k) = [y_1(k) \ y_2(k) \ \dots \ y_m(k)]^T$ is the normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. Since $\mu_N(\mathbf{y}(k)) + \mu_F(\mathbf{y}(k)) = 1$, we only need to calculate one membership function. Based on the above discussion on the ‘ 3σ theory’, the membership function of the fuzzy set ‘Failure’ can be defined as

$$\begin{aligned} \mu_F(\mathbf{y}(k)) &= P(-|y_1(k)| \leq \psi_1 < |y_1(k)|, \dots, -|y_m(k)| \leq \psi_m < |y_m(k)|) \\ &= \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}} \int_{-|y_1(k)|}^{|y_1(k)|} \dots \int_{-|y_m(k)|}^{|y_m(k)|} e^{-\frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\psi} - \boldsymbol{\mu})} d\boldsymbol{\psi}, \end{aligned} \tag{24}$$

where $\boldsymbol{\psi} = [\psi_1 \ \psi_2 \ \dots \ \psi_m]^T$. Equation (24) can be evaluated using Monte–Carlo methods or subregion adaptive methods through a series of transformation [41]. Process fault prognosis often deals with low-dimensional $\mathbf{y}(k)$. Taking the bivariate $\mathbf{y}(k) = [y_1(k) \ y_2(k)]^T$ for instance, define the probability distribution function $g(c, d) = P(\psi_1 < c, \psi_2 < d)$. Then, Equation (24) can be evaluated by several values of this probability distribution function as

$$\begin{aligned} \mu_F(\mathbf{y}(k)) &= P(-|y_1(k)| \leq \psi_1 < |y_1(k)|, -|y_2(k)| \leq \psi_2 < |y_2(k)|) = g(|y_1(k)|, |y_2(k)|) \\ &\quad - g(|y_1(k)|, -|y_2(k)|) - g(-|y_1(k)|, |y_2(k)|) + g(-|y_1(k)|, -|y_2(k)|). \end{aligned} \tag{25}$$

The function $g(c, d) = P(\psi_1 < c, \psi_2 < d)$ can be computed using the Gauss–Legendre integration algorithm [42].

If $1 \leq \lambda(k) \leq a$, from $0 \leq \mu_N(\mathbf{y}(k)), \mu_F(\mathbf{y}(k)) \leq 1$ and $\mu_N(\mathbf{y}(k)) + \mu_F(\mathbf{y}(k)) = 1$, it can be obtained that $1 \leq \lambda^*(k) \leq \lambda(k) \leq a$. This means that the fuzzy-adaptive parameter $\lambda^*(k)$ has a ‘weaker’ regulation ability in comparison to $\lambda(k)$, particularly when the process is in the ‘Normal’ state, which is useful for preventing overregulation and large fluctuation.

3.3. Fuzzy-adaptive predictor

3.3.1. *FAUKF for state estimation.* The proposed FAUKF can now be summarized.

- (1) The predicted state $\hat{\mathbf{x}}(k|k-1)$, the unregulated error covariance $\hat{\mathbf{P}}(k|k-1)$ and the predicted output $\hat{\mathbf{y}}(k)$ are calculated according to (9)–(12).
- (2) The adaptive parameter $\lambda(k)$ is determined by (21), and the fuzzy-adaptive parameter $\lambda^*(k)$ is computed according to (23). Then the regulated error covariance $\hat{\mathbf{P}}(k|k-1)$ is obtained by substituting $\lambda(k)$ with $\lambda^*(k)$ into (16).
- (3) The Kalman gain $\mathbf{K}(k)$, the updated state $\hat{\mathbf{x}}(k)$ and the error covariance $\hat{\mathbf{P}}(k)$ are calculated according to (17), (14) and (18), respectively.

Remark

Since the FAUKF requires a numerical calculation of the fuzzy membership function $\mu_F(\mathbf{y})$, it imposes a slightly higher complexity than the AUKF.

3.3.2. *FAUKF for process fault prognosis.* Assume that the fault parameter vector can be represented by the random-walk model given by

$$\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}(k) + \boldsymbol{\xi}(k), \tag{26}$$

where the noise vector $\boldsymbol{\xi}(k) \in \mathbb{R}^{q \times 1}$ has a zero-mean normal distribution. Augmenting the state vector $\mathbf{x}(k)$ with the fault parameter vector $\boldsymbol{\theta}(k)$ yields the augmented state vector $\mathbf{x}_a(k) = [\mathbf{x}^T(k) \boldsymbol{\theta}^T(k)]^T \in \mathbb{R}^{(n+q) \times 1}$, with the augmented nonlinear dynamic system

$$\mathbf{x}_a(k+1) = \mathbf{f}_a(\mathbf{x}_a(k), \mathbf{u}(k)) + \boldsymbol{\varepsilon}_a(k), \tag{27}$$

$$\mathbf{y}(k) = \mathbf{H}_a(k)\mathbf{x}_a(k) + v(k), \tag{28}$$

where the augmented process noise $\boldsymbol{\varepsilon}_a(k) = [\boldsymbol{\varepsilon}^T(k) \boldsymbol{\xi}^T(k)]^T \in \mathbb{R}^{(n+q) \times 1}$ has a covariance $\mathbf{Q}_a(k) \in \mathbb{R}^{(n+q) \times (n+q)}$, and the nonlinear vector function $\mathbf{f}_a(\bullet): \mathbb{R}^{(n+q+l) \times 1} \rightarrow \mathbb{R}^{(n+q) \times 1}$ is defined by

$$\mathbf{f}_a(\mathbf{x}_a(k), \mathbf{u}(k)) = \begin{bmatrix} \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) \\ \boldsymbol{\theta}(k) \end{bmatrix}, \tag{29}$$

while the ‘augmented’ linear measurement equation is specified by the matrix $\mathbf{H}_a(k) = [\mathbf{H}(k) \mathbf{0}_{m \times q}] \in \mathbb{R}^{m \times (n+q)}$. Note that the measurement equation (28) is identical to (2). The process state and fault parameters, namely, $\mathbf{x}_a(k)$, can be estimated via the FAUKF.

Specifically, the set of sigma points $\tilde{\mathbf{X}}_i(k-1) \in \mathbb{R}^{n_s \times 1}$, where $0 \leq i \leq 2n_s$, is now defined for the enlarged state variable $\tilde{\mathbf{x}}_s(k-1) = [\mathbf{x}_a^T(k-1) \boldsymbol{\varepsilon}_a^T(k-1) v^T(k-1)]^T$, which has the dimension $n_s = 2(n+q) + m$. In calculating these sigma points of (5), we now have $\tilde{\mathbf{x}}_s(k-1) = [\hat{\mathbf{x}}_a^T(k-1) \mathbf{0}_{1 \times (n+q)} \mathbf{0}_{1 \times m}]^T$ with $\hat{\mathbf{x}}_a(k-1)$ being the estimate of $\mathbf{x}_a(k-1)$. Moreover, the enlarged error covariance $\hat{\mathbf{P}}_s(k-1)$ of (8) becomes

$$\hat{\mathbf{P}}_s(k-1) = \begin{bmatrix} \hat{\mathbf{P}}_a(k-1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_a(k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}(k-1) \end{bmatrix} \in \mathbb{R}^{n_s \times n_s} \tag{30}$$

with $\widehat{\mathbf{P}}_a(k-1) \in \mathbf{R}^{(n+q) \times (n+q)}$ denoting the augmented state error covariance. Note that, for the notational simplicity, we have kept the same notations $n_s, \widetilde{\mathbf{X}}_i(k-1), \widetilde{\mathbf{x}}_s(k-1), \widetilde{\mathbf{x}}_s(k-1)$ and $\widehat{\mathbf{P}}_s(k-1)$, but their exact dimensions, as defined here, are clearly different from those given in Section 3.1. The FAUKF for estimating the augmented state can now be summarized as follows.

- (1) The predicted augmented state $\widehat{\mathbf{x}}_a(k|k-1) \in \mathbf{R}^{(n+q) \times 1}$, the unregulated augmented error covariance $\widehat{\mathbf{P}}_a(k|k-1) \in \mathbf{R}^{(n+q) \times (n+q)}$ and the predicted output $\widehat{\mathbf{y}}(k)$ are given by

$$\widehat{\mathbf{X}}_{ai}(k) = \mathbf{f}_a(\widetilde{\mathbf{X}}_i^x(k-1), \mathbf{u}(k-1)) + \widetilde{\mathbf{X}}_i^\varepsilon(k-1), \tag{31}$$

$$\widehat{\mathbf{x}}_a(k|k-1) = \sum_{i=0}^{2n_s} w_i \widehat{\mathbf{X}}_{ai}(k), \tag{32}$$

$$\widehat{\mathbf{P}}_a(k|k-1) = \sum_{i=0}^{2n_s} w_i (\widehat{\mathbf{X}}_{ai}(k) - \widehat{\mathbf{x}}_a(k|k-1)) (\widehat{\mathbf{X}}_{ai}(k) - \widehat{\mathbf{x}}_a(k|k-1))^T, \tag{33}$$

$$\widehat{\mathbf{y}}(k) = \mathbf{H}_a(k) \widehat{\mathbf{x}}_a(k|k-1), \tag{34}$$

where $\widetilde{\mathbf{X}}_i^x(k-1) \in \mathbf{R}^{(n+q) \times 1}$ consists of the first $n+q$ elements of $\widetilde{\mathbf{X}}_i(k-1)$ and $\widetilde{\mathbf{X}}_i^\varepsilon(k-1) \in \mathbf{R}^{(n+q) \times 1}$ contains the $(n+q+1)$ th to $(2(n+q))$ th elements of $\widetilde{\mathbf{X}}_i(k-1)$, which should not be confused with the notations defined after (11).

- (2) The adaptive parameter $\lambda(k)$ is determined by

$$\lambda(k) = \max \left\{ \frac{\text{tr}[\mathbf{P}_{\widehat{\mathbf{y}}}(k) - \zeta \mathbf{R}(k)]}{\text{tr}[\mathbf{H}_a(k) \widehat{\mathbf{P}}_a(k|k-1) \mathbf{H}_a^T(k)]}, 1 \right\}, \tag{35}$$

which is identical to (21), and the fuzzy-adaptive parameter $\lambda^*(k)$ is computed according to (23). Then, the regulated augmented error covariance $\widehat{\mathbf{P}}_a(k|k-1) \in \mathbf{R}^{(n+q) \times (n+q)}$ is obtained by

$$\widehat{\mathbf{P}}_a(k|k-1) = \lambda^*(k) \widehat{\mathbf{P}}_a(k|k-1). \tag{36}$$

- (3) The augmented Kalman gain $\mathbf{K}_a(k) \in \mathbf{R}^{(n+q) \times m}$, the updated augmented state $\widehat{\mathbf{x}}_a(k) \in \mathbf{R}^{(n+q) \times 1}$ and the augmented error covariance $\widehat{\mathbf{P}}_a(k) \in \mathbf{R}^{(n+q) \times (n+q)}$ are given by

$$\mathbf{K}_a(k) = \widehat{\mathbf{P}}_a(k|k-1) \mathbf{H}_a^T(k) (\mathbf{H}_a(k) \widehat{\mathbf{P}}_a(k|k-1) \mathbf{H}_a^T(k) + \mathbf{R}(k))^{-1}, \tag{37}$$

$$\widehat{\mathbf{x}}_a(k) = \widehat{\mathbf{x}}_a(k|k-1) + \mathbf{K}_a(k) (\mathbf{y}(k) - \widehat{\mathbf{y}}(k)), \tag{38}$$

$$\widehat{\mathbf{P}}_a(k) = \widehat{\mathbf{P}}_a(k|k-1) - \mathbf{K}_a(k) (\mathbf{H}_a(k) \widehat{\mathbf{P}}_a(k|k-1) \mathbf{H}_a^T(k) + \mathbf{R}(k)) \mathbf{K}_a^T(k). \tag{39}$$

Given the estimate $\widehat{\mathbf{x}}_a(k_p) = [\widehat{\mathbf{x}}^T(k_p) \widehat{\boldsymbol{\theta}}^T(k_p)]^T$ obtained by the FAUKF, the future forecasted outputs starting from k_p can be made by assuming that the fault parameters remain constant for the prediction horizon N , resulting in the prediction equation for the state $\mathbf{x}(k_p+i)$

$$\begin{aligned} \widehat{\mathbf{x}}(k_p+i) &= \mathbf{f}(\widehat{\mathbf{x}}(k_p+i-1), \mathbf{u}(k_p+i-1)) + \widehat{\mathbf{d}}(k_p+i-1), \\ \widehat{\mathbf{y}}(k_p+i) &= \mathbf{H}(k_p+i) \widehat{\mathbf{x}}(k_p+i), \end{aligned} \tag{40}$$

where $1 \leq i \leq N$, $\widehat{\mathbf{d}}(k_p+i-1) = \boldsymbol{\Gamma}(k_p+i-1, k_0) \widehat{\boldsymbol{\theta}}(k_p)$, and $\widehat{\boldsymbol{\theta}}(k_p)$ is an estimate of $\boldsymbol{\theta}(k_p)$.

Assume that a control limit for the output variable is set by considering the process demands or based on the knowledge of the plant operation. If a prediction value violates this control limit, a fault may exist, which indicates that, even if the current measurement lies inside the normal-operating region, the measurement may violate the normal operation constraint at some time in the future. Thus, warning can be made, and corrective actions can be taken at the present to prevent the faults when they are still small. To explain this in a simpler and clearer way, take the univariate case $\mathbf{y}(k) = y(k)$, for example, and assume that the control limit is C_y which has been appropriately

chosen and is larger than the steady-state value of $y(t)$. If the current measurement is $y(k_p) > C_y$, then the control limit is violated and a fault is detected. If the measurement is $y(k_p) < C_y$ but the future predictions, forecasted starting from k_p , violates the control limit, i.e. $\hat{y}(k_p+i) > C_y$ for some $i \geq 1$, then a fault may exist at its initial stage and the future measurement at some point $y(k_p+k_f)$ may violate the control limit. In this case, a fault warning should be signaled, and corrective actions should be taken at the present. Through this process fault prognosis, faults may be prevented when they are still small.

4. SIMULATION RESULTS

Two simulation examples, a second-order nonlinear system and a continuous stirred tank reactor (CSTR), were used to compare the performance of the proposed FAUKF-based fault prognosis method with that of the standard AUKF-based method.

4.1. State estimation for the second-order nonlinear system

Consider the nonlinear stochastic system given by Xiong *et al.* [24]

$$\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} x_1(k-1) + \tau x_2(k-1) \\ x_2(k-1) + \tau(-x_1(k-1) + (x_1^2(k-1) + x_2^2(k-1) - 1)x_2(k-1)) \end{bmatrix} + \boldsymbol{\varepsilon}(k), \quad (41)$$

$$y(k) = x_2(k) + v(k), \quad (42)$$

where $\tau = 0.001$, the covariance matrix of $\boldsymbol{\varepsilon}(k)$ and the variance of $v(k)$ are $\mathbf{Q}(k) = 0.003^2 \mathbf{I}_2$ and $R(k) = 0.001^2$, respectively. The initial conditions for the system were $x_1(0) = 0.8$ and $x_2(0) = 0.2$. Monte-Carlo simulations were carried out for 50 times, yielding the 50 realizations of the state and measurement trajectories. The standard UKF, the AUKF, and the FAUKF were used to estimate the state. The initial conditions for the three filters were $\hat{x}_1(0) = 2.3$, $\hat{x}_2(0) = 2.2$ and $\hat{\mathbf{P}}(0) = \mathbf{Q}(0)$. The regulation rate was restricted by setting $1 \leq \lambda(k)$, $\lambda^*(k) \leq 1.1$, in order to guarantee the stability of the AUKF algorithm. For the FAUKF, the probability distribution parameters were $\mu = 0$ and $3\sigma = 0.1$.

Define the root mean-squared error (RMSE) for the estimation of the state $x_1(k)$ as

$$\text{RMSE}(k) = \sqrt{\frac{1}{50} \sum_{l=1}^{50} (x_1^{(l)}(k) - \hat{x}_1^{(l)}(k))^2}, \quad (43)$$

where $x_1^{(l)}(k)$ and $\hat{x}_1^{(l)}(k)$ denote the actual and estimated states for the l th Monte-Carlo simulation, respectively. Figure 2 plots the RMSE values of the first state estimation obtained by the three algorithms. It can be seen from Figure 2 that both the AUKF and FAUKF achieved a similar performance at the initial-convergence stage which was better than that for the UKF. However, the RMSE(k) values from the FAUKF after the initial convergence were smaller than those from

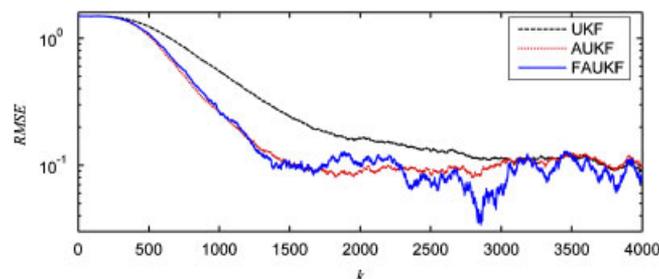


Figure 2. RMSE comparison of the UKF, AUKF and FAUKF for the simulated second-order nonlinear system.

the AUKF, as can be seen clearly in Figure 2. This confirms that the FAUKF, while preserving the strong tracking ability of the AUKF, reduced the tracking fluctuations in the regulated error covariance matrix to a certain extent, leading to a better performance in the state estimation.

4.2. Process fault prognosis for the CSTR

The nonisothermal CSTR with a first-order irreversible reaction $A_r \rightarrow B$ and a cooling jacket to remove heat was described by Juricek *et al.* [7]

$$\begin{aligned} \frac{dC_A}{dt} &= -K_0 e^{-\frac{E}{RT}} C_A + \frac{F(C_F - C_A)}{V}, \\ \frac{dT}{dt} &= \frac{(-\Delta H)K_0 e^{-\frac{E}{RT}} C_A}{\rho C_P} + \frac{F(T_F - T)}{V} + \frac{UA_a(T_J - T)}{V\rho C_P}, \end{aligned} \quad (44)$$

where A_a is the heat transfer area; C_A is the concentration of reactant A_r ; C_F is the feed concentration, 1 (mol/L); C_P is the mass specific heat, 239 (J/(kgK)); E is the activation energy; F is the feed flow rate, 100 (L/min); K_0 is the frequency coefficient, 7.2×10^{10} (min^{-1}); R is the gas constant, and E/R is 8750 (K); T is the reactor temperature; T_F is the feed temperature, 350 (K); T_J is the coolant temperature, 309.9 (K); U is the overall heat transfer coefficient, and UA_a is 5×10^4 (J/(minK)); V is the volume, 100 (L); ρ is the density, 1 (kg/L); ΔH is the reaction enthalpy, -5×10^4 (J/mol).

The state vector and the output vector were both defined to be $[C_A \ T]^T$, while the input vector was given by $[T_J \ F]^T$. A list of process faults are given in Table I. The normal process and the failure process were simulated with the sampling period of 0.2 min. All the faults occurred at $k = 101$. The augmented state vector was given by $[C_A \ T \ \theta_{1,0} \ \theta_{2,0} \ \theta_{1,s} \ \theta_{2,s}]^T$. The output T was monitored with the control limit 393 K. The filter parameters are listed as follows:

- Augmented process noise covariance: $\mathbf{Q}_a(k) = \text{diag}\{10^{-7}, 10^{-3}, 10^{-7}, 10^{-3}, 10^{-7}, 10^{-3}\}$;
- Measurement noise covariance: $\mathbf{R}(k) = \text{diag}\{10^{-6}, 0.04\}$;
- Initial covariance for filters: $\hat{\mathbf{P}}_a(0) = \mathbf{Q}_a(0)$; Regulation rate: $1 \leq \lambda(k), \lambda^*(k) \leq 2$;
- Probability distribution parameters for the FAUKF: $\boldsymbol{\mu} = [0.1 \ 383.78]^T$, $\boldsymbol{\Sigma}^{\frac{1}{2}} = \text{diag}\{0.002, 0.46\}$.

A normalized squared prediction error was defined as

$$e^2(k) = \|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}(k) - \hat{\mathbf{y}}(k))\|_2^2 \quad (45)$$

to compare the filters' performance. Tracking errors $e^2(k)$ under the normal-operating condition are shown in Figure 3 for $1 \leq k \leq 300$, where it can be seen that the squared errors of the FAUKF have smaller magnitudes than those of the AUKF. Regulation parameters $\lambda(k)$ and $\lambda^*(k)$ are plotted in Figure 4. It can be seen from Figure 4 that, at most of the sampling times, $\lambda^*(k)$ was less than or equal to $\lambda(k)$, and this had the desired effect of smoothing the regulation for the error covariance under the normal status. The averaged normalized squared prediction error over $K_1 \leq k \leq K_2$ was defined as

$$\text{ASE} [K_1, K_2] = \frac{1}{K_2 - K_1 + 1} \sum_{k=K_1}^{K_2} e^2(k) \quad (46)$$

Table I. The fault settings for the CSTR.

No.	Operating condition	Description and size
Fault 1	Catalyst deactivation	The ramp rate for E/R is +3 (K/min).
Fault 2	Ramp change in feed temperature	The ramp rate for T_F is +2 (K/min).
Fault 3	Ramp change in feed concentration	The ramp rate for C_F is +0.01 ((mol/L)/min).
Fault 4	Heat exchanger fouling	The ramp rate for UA_a is -125 ((J/(minK))/min).

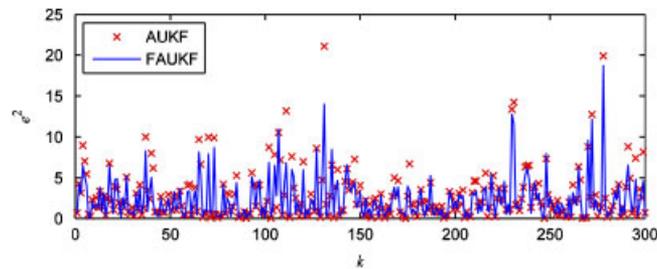


Figure 3. Tracking errors under the normal-operating condition for the CSTR.

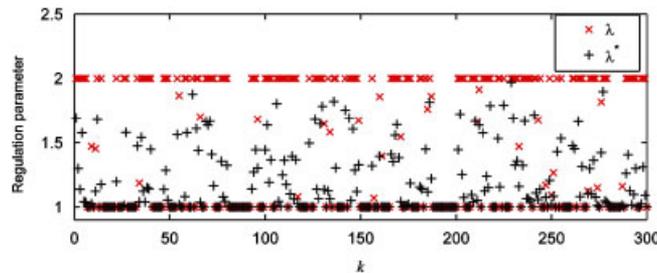


Figure 4. Regulation parameters under the normal-operating condition for the CSTR.

Table II. Comparison of the averaged normalized square prediction errors for the AUKF and FAUKF.

Operation status	Operation mode	ASE	AUKF	FAUKF
Normal	Tracking	ASE [1, 300]	2.7232	2.3531
	Forecasting	ASE [301, 350]	2.4403	2.2357
	Forecasting	ASE [302, 351]	2.1163	2.0633
	Forecasting	ASE [303, 352]	2.0779	2.0157
Fault 1	Tracking	ASE [1, 100]	2.1878	1.7999
	Tracking	ASE [101, 200]	4.8968	4.0981
Fault 2	Tracking	ASE [1, 122]	2.8625	2.4133
	Forecasting	ASE [123, 172]	5.3088	5.7559
Fault 3	Tracking	ASE [1, 122]	2.9277	2.6013
	Forecasting	ASE [123, 172]	3.5814	2.3728
Fault 4	Tracking	ASE [1, 300]	2.9730	2.7833
	Forecasting	ASE [301, 350]	3.3502	3.3502

The values of ASE [1, 300] for the AUKF and FAUKF in tracking are compared in Table II. The next 50 measurements were then forecasted starting at $k_p = 300, 301,$ and $302,$ respectively, and the associated ASE values for the AUKF and FAUKF are also listed in Table II, where it can be seen that the ASE values produced by the AUKF were larger than those produced by the FAUKF.

Fault 1 was the catalyst deactivation simulated as a ramp change in the activation energy, with a ramp rate 3 K/min for E/R . The results obtained under Fault 1 operating condition are shown in Figure 5. The outputs were tracked during the period of $k = 101$ to 401 , and the forecasts of the next fifty measurements were made starting at $k_p = 401$. As the fault grew up, the outputs increased slowly and then jumped to the neighborhood of a new steady-state point near $k = 300$. The estimation from the UKF had the overshoot when tracking the large-amplitude change at $k = 300$ to 320 , as can be seen from Figure 5. Although the estimation of the concentration C_A by the UKF at $k = 401$ was accurate, its future predictions had large errors. Moreover, for the neighborhood of the new steady-state point, the temperature T was tracked by the UKF with large

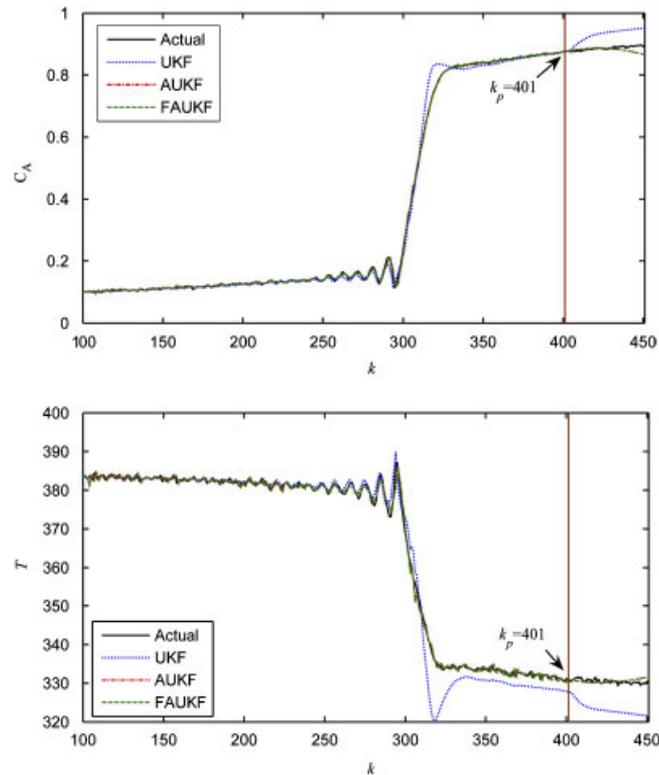


Figure 5. Tracking and forecasting results under Fault 1 operating condition for the CSTR.

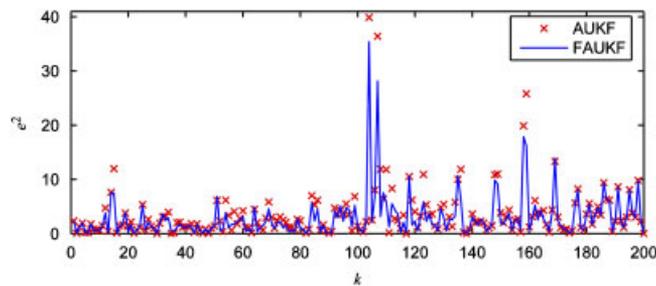


Figure 6. Tracking errors under the normal and initial Fault 1 operating conditions for the CSTR.

errors, and the future predictions by the UKF starting at $k_p=401$ were very poor. The AUKF and FAUKF could track changes more quickly and accurately, as clearly seen in Figure 5. The tracking errors obtained by the AUKF and FAUKF for the normal-operating stage ($k=1$ to 100) and the initial failure stage ($k=101$ to 200) are depicted in Figure 6. After the fault occurred at $k=101$, the squared prediction error increased sharply and then decreased quickly again, which indicated that both the AUKF and FAUKF could capture the process change from the normal state to a failure state quickly. The values of ASE [1, 100] and ASE [101, 200] obtained by the AUKF and FAUKF are also listed in Table II. It can be seen that the FAUKF produced smaller errors than the AUKF. Figure 7 depicts the regulating parameters $\lambda(k)$ and $\lambda^*(k)$. It is interesting to see that, under the normal-operating condition, $\lambda^*(k)$ was smaller than $\lambda(k)$, while at the initial failure stage, $\lambda^*(k)$ started to catch up $\lambda(k)$ quickly. As the fault grew up, $\lambda^*(k)$ was almost equal to $\lambda(k)$. This confirms that the FAUKF only softens the regulation under the normal-operating state.

Fault 2 was a 2 K/min ramp change in the feed temperature T_F , and the results obtained under this failure status are shown in Figures 8 and 9. Specifically, the outputs were tracked during the

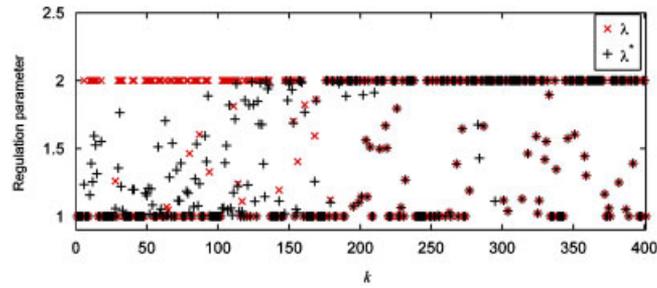


Figure 7. Regulation parameters under the normal and Fault 1 operating conditions for the CSTR.

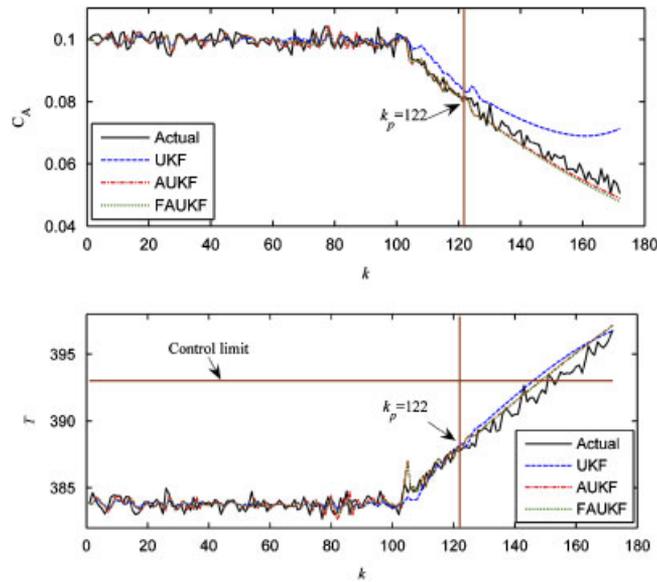


Figure 8. Tracking and future forecasting results under Fault 2 operating condition for the CSTR, where the period of $k=1$ to 100 is the normal-operating period.

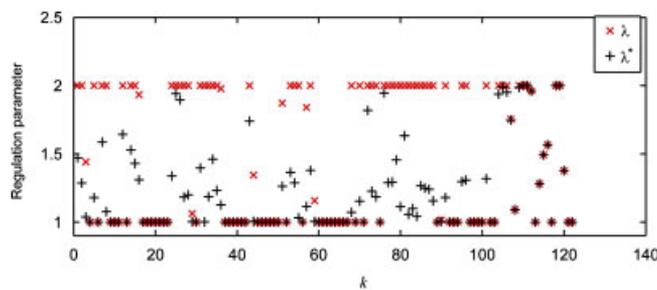


Figure 9. Regulation parameters under the normal and Fault 2 operating conditions for the CSTR.

period of $k=1$ to 122, and the forecasts of the next 50 measurements were made starting from $k_p=122$. Figure 8 shows that the forecasted temperature T values increased continuously and consequentially violated the control limit. Note that the actually observed process data violated the control limit at $k=151$. Since the forecasts generated at $k_p=122$ signaled the violation of the control limit, the predictor was able to signal this future violation $\delta k=151-122=29$ sample-time ahead. From Figure 9, it is again seen that under the normal-operating condition, $\lambda^*(k)$ was smaller than $\lambda(k)$, while when the fault occurred at $k=101$, $\lambda^*(k)$ caught up with $\lambda(k)$ quickly. The

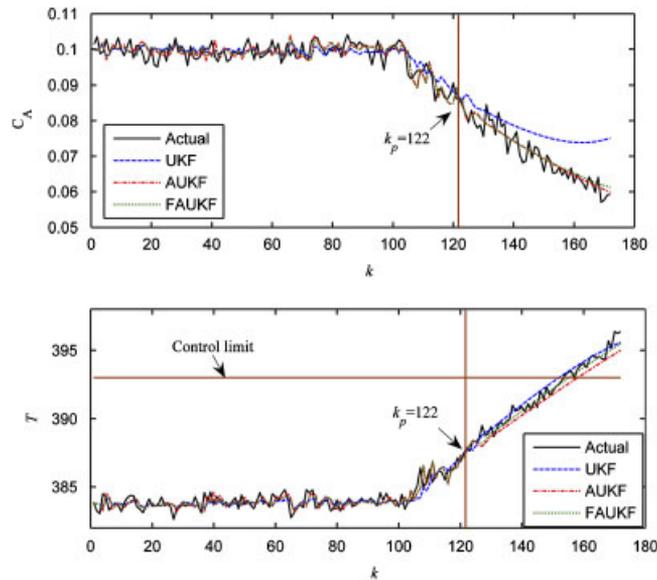


Figure 10. Tracking and future forecasting results under Fault 3 operating condition for the CSTR, where the period of $k=1$ to 100 is the normal-operating period.

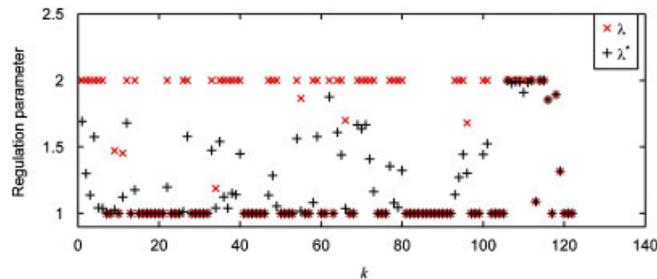


Figure 11. Regulation parameters under the normal and Fault 3 operating conditions for the CSTR.

tracking ASE [1, 122] and the forecasting ASE [123, 172] obtained by the AUKF and FAUKF are compared in Table II. This was the only case that the forecasting error of the FAUKF was larger than that of the AUKF.

The results obtained under Fault 3 condition, which was a 0.01 (mol/L)/min ramp change in the feed concentration C_F , are plotted in Figures 10 and 11. In this case, the outputs were tracked during the $k=1$ to 122 samples, and the forecasts of the next 50 measurements were made starting from $k_p=122$. Figure 10 shows that the forecasted temperature T values violated the control limit. Since the actual process data violated the control limit at $k=154$, the predictor signaled the future violation in the $\delta k=32$ samples ahead. The same relationship between $\lambda(k)$ and $\lambda^*(k)$ during the normal- and failure-operating conditions can be observed from Figure 11. The tracking error ASE [1, 122] and the forecasting error ASE [123, 172] obtained by the AUKF and FAUKF are listed in Table II.

Fault 4 was the heat exchanger fouling simulated as a ramp change in the heat transfer coefficient. The ramp rate for $U A_a$ was -125 (J/(min K))/min. The tracking and future forecasting results are illustrated in Figures 12 and 13. In this case, the outputs were tracked for the duration of $k=201$ to 300, and the forecasts of the next 50 measurements were made starting from $k_p=300$. Figure 12 shows that forecasted temperature T values violated the control limit. Since the actual process data violated this limit at $k=325$, the predictor signaled the future violation $\delta k=25$ samples ahead. Again the result of Figure 13 confirms that the FAUKF only softens the regulation under the

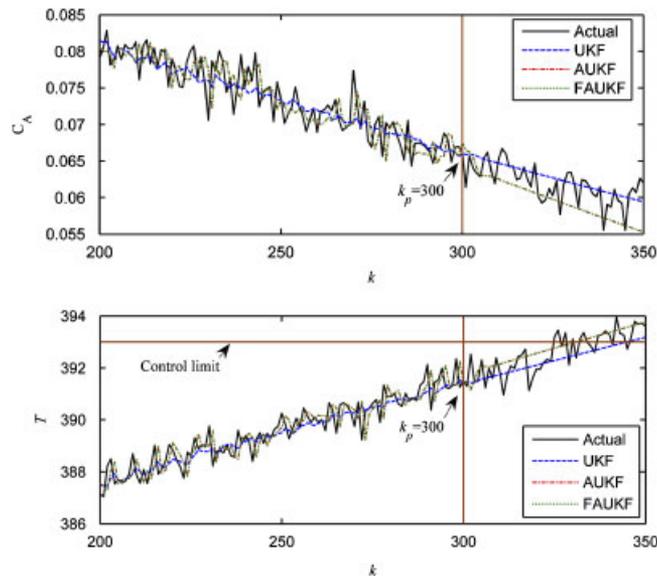


Figure 12. Tracking and future forecasting results under Fault 4 operating condition for the CSTR.

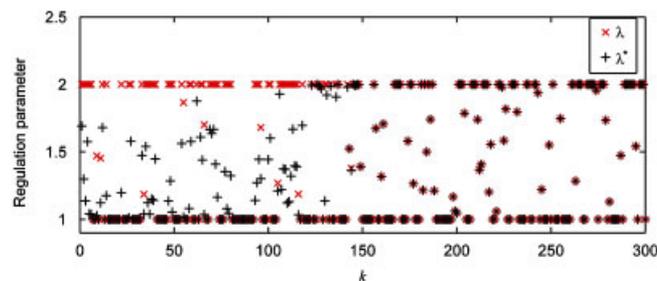


Figure 13. Regulation parameters under the normal and Fault 4 operating conditions for the CSTR.

normal-operating state. The tracking error ASE [1, 300] and the forecasting error ASE [301, 350] obtained by the AUKF and FAUKF are also compared in Table II.

5. CONCLUSIONS

A fuzzy-adaptive unscented Kalman predictor has been proposed for nonlinear process fault prognosis. Our motivation for this FAUKF is based on the observation that the standard UKF has a weak tracking ability, which can influence its future forecasting performance, while the adaptive UKF with the property of STF can capture process changes quickly but may introduce unnecessary regulation. It has been shown that the FAUKF guarantees the strong tracking ability, while softening unnecessary regulation. The FAUKF-based predictor can forecast the future process output accurately and provides early fault alarm. Two examples have been used to demonstrate the effectiveness of the proposed FAUKF predictor for process fault prognosis.

ACKNOWLEDGEMENTS

This work was supported by China National 863 program (2007AA04Z193) and Shandong province Nature Science Foundation (Y2007G49).

REFERENCES

1. Willsky AS. A survey of design methods for failure detection in dynamic systems. *Automatica* 1976; **12**(6):601–611.
2. Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 2006; **20**(7):1483–1510.
3. Tian XM, Cao YP. Situation and perspectives of statistical process control. *Journal of China University of Petroleum* (Edition of Natural Science) 2008; **32**(5):175–180.
4. Lu KS, Saeks R. Failure prediction for an on-line maintenance system in a Poisson shock environment. *IEEE Transactions on Systems, Man, and Cybernetics* 1979; **9**(6):356–362.
5. Yang SK, Liu TS. State estimation for predictive maintenance using Kalman filter. *Reliability Engineering and System Safety* 1999; **66**(1):29–39.
6. Yang SK. An experiment of state estimation for predictive maintenance using Kalman filter on a DC motor. *Reliability Engineering and System Safety* 2002; **75**(1):103–111.
7. Juricek BC, Seborg DE, Larimore WE. Predictive monitoring for abnormal situation management. *Journal of Process Control* 2001; **11**(2):111–128.
8. Cao YP, Tian XM. Nonlinear system fault prognosis based on SVM and Kalman predictor. *Control and Decision* 2009; **24**(3):477–480.
9. Chen MZ, Zhou DH, Liu GP. A new particle predictor for fault prediction of nonlinear time-varying systems. *Developments in Chemical Engineering and Mineral Processing* 2005; **13**(3–4):379–388.
10. Ho SL, Xie M. The use of ARIMA models for reliability forecasting and analysis. *Computers and Industrial Engineering* 1998; **35**(1–2):213–216.
11. Weber P, Medina-Oliva G, Simon C, Lung B. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence* 2010; DOI: 10.1016/j.engappai.2010.06.002.
12. Wang WQ, Golnaraghi MF, Ismail F. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing* 2004; **18**(4):813–831.
13. Wang WQ. An adaptive predictor for dynamic system forecasting. *Mechanical Systems and Signal Processing* 2007; **21**(2):809–823.
14. Wang MH, Hung CP. Novel grey model for the prediction of trend of dissolved gases in oil-filled power apparatus. *Electric Power System Research* 2003; **67**(1):53–58.
15. Zhao Z, Wang FL, Jia MX, Wang S. Predictive maintenance policy based on process data. *Chemometrics and Intelligent Laboratory Systems* 2010; **104**(1):137–143.
16. Watson M, Bingtong C, Edwards D, Amin S. Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems. *Tribology Transactions* 2005; **48**(2):208–217.
17. Li Y, Kurfess TR, Liang SY. Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing* 2000; **14**(5):747–762.
18. Yang SK. A condition-based preventive maintenance arrangement for thermal power plants. *Electric Power Systems Research* 2004; **72**(1):49–62.
19. Tran VT, Yang BS, Oh MS, Tran ACC. Machine condition prognosis based on regression trees and one-step-ahead prediction. *Mechanical Systems and Signal Processing* 2008; **22**(5):1179–1193.
20. Zhou DH, Su YX, Xi YG, Zhang ZJ. Extension of Friedland's separate-bias estimation to randomly time-varying bias of nonlinear systems. *IEEE Transactions on Automatic Control* 1993; **38**(8):1270–1273.
21. Jwo DJ, Wang SH. Adaptive fuzzy strong tracking extended Kalman filtering for GPS navigation. *IEEE Sensors Journal* 2007; **7**(5):778–789.
22. Julier S, Uhlmann J, Durrant-Whyte HF. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control* 2000; **45**(3):477–482.
23. Julier SJ, Uhlmann JK. Unscented filtering and nonlinear estimation. *Proceeding of IEEE* 2004; **92**(3):401–422.
24. Xiong K, Zhang HY, Chan CW. Performance evaluation of UKF-based nonlinear filtering. *Automatica* 2006; **42**(2):261–270.
25. Xiong K, Wei CL, Liu LD. Robust unscented Kalman filtering for nonlinear uncertain systems. *Asian Journal of Control* 2010; **12**(3):426–433.
26. Rambabu K, Bjarne F, Lars I. Applying the unscented Kalman filter for nonlinear state estimation. *Journal of Process Control* 2008; **18**(7–8):753–768.
27. Romanenko A, Castro JAAM. The unscented filter as an alternative to the EKF for nonlinear state estimation, a simulation case study. *Computers and Chemical Engineering* 2004; **28**(3):347–355.
28. Qu CC, Hahn J. Process monitoring and parameter estimation via unscented Kalman filtering. *Journal of Loss Prevention in the Process Industries* 2009; **22**(6):703–709.
29. Cao YP, Tian XM. Fault detection method based on multi-variable sequential probability ratio test of unscented Kalman filter innovation. *Journal of China University of Petroleum* (Edition of Natural Science) 2010; **34**(3):165–169.
30. Cao YP, Tian XM. Information divergence based process fault detection and diagnosis. *Journal of Zhejiang University* (Engineering Science) 2010; **44**(7):1315–1320.
31. Cao YP, Tian XM. An adaptive UKF algorithm for process fault prognostics. *Proceeding of 2nd International Conference on Intelligent Computation Technology and Automation*, Changsha, China, 10–11 October 2009; 487–490.

32. Xiong K, Zhang HY, Chan CW. Author's reply to 'Comments on Performance evaluation of UKF-based nonlinear filtering'. *Automatica* 2007; **43**(3):569–570.
33. Xu JH, Dimirovski GM, Jiang YW, Shen C. UKF design and stability for nonlinear stochastic systems with correlated noises. *Proceeding of 46th IEEE CDC*, New Orleans, U.S.A., 12–14 December 2007; 6226–6231.
34. Sugeno M. *Industrial Applications of Fuzzy Control*. Elsevier Science Inc.: New York, 1985.
35. Maybeck PS. *Stochastic Models, Estimation and Control*. Academic Press: New York, 1979.
36. Lee DJ. *Nonlinear Bayesian Filtering with Applications to Estimation and Navigation*. Ph.D. Dissertation, Texas A&M University, 2005.
37. Song Q, Han JD. An adaptive UKF algorithm for the state and parameter estimations of a mobile robot. *Acta Automatica Sinica* 2008; **34**(1):72–79.
38. Li J. A strong tracking unscented Kalman filter and its application in passive target tracking. *Telecommunication Engineering* 2005; **45**(1):160–164.
39. Stoumbos ZG, Reynolds MR, Ryan TP, Woodall WH. The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association* 2000; **95**(451):992–998.
40. Chiang LH, Russell EL, Braatz RD. *Fault Detection and Diagnosis in Industrial Systems*. Springer: London, 2001.
41. Genz A. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* 1992; **1**(2):141–149.
42. Genz A. Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing* 2004; **14**(3):251–260.