

Article

# An Adaptive Early Fault Detection Model of Induced Draft Fans Based on Multivariate State Estimation Technique

Ruijun Guo <sup>1</sup>, Guobin Zhang <sup>1</sup>, Qian Zhang <sup>1</sup>, Lei Zhou <sup>1</sup>, Haicun Yu <sup>1</sup>, Meng Lei <sup>2</sup>, and You Lv <sup>2,3,\*</sup>

<sup>1</sup> Inner Mongolia Power Research Institute, Hohhot 010020, China; zrui\_75@sina.com (Ruijun Guo)

<sup>2</sup> School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, China; leimeng\_0908@ncepu.edu.cn (Meng Lei)

<sup>3</sup> Key Laboratory of Power Station Energy Transfer Conversion and System, North China Electric Power University, Beijing 102206, China; you.lv@ncepu.edu.cn (You Lv)

\* Correspondence: you.lv@ncepu.edu.cn, you.lv@hotmail.com

**Abstract:** The induced draft (ID) fan is important auxiliary equipment in the thermal power plant. It is of great significance to monitor the operation of the ID fan for safe and efficient production. In this paper, an adaptive warning model is proposed to detect early faults of ID fans. First, a non-parametric monitoring model is constructed to describe the normal operation states with the multivariate state estimation technique (MSET). Then, an early warning approach is presented to identify abnormal behaviors based on the results of the MSET model. As the performance of the MSET model is heavily influenced by the normal operation data in the historic memory matrix, an adaptive strategy is proposed by using the samples with a high data quality index (DQI) to manage the memory matrix and update the model. The proposed method is applied to a 300 MW coal-fired power plant for early fault detection, and it is compared with the model without an update. Results show that the proposed method can detect the fault earlier and more accurately.

**Keywords:** fault detection; induced draft fan; multivariate state estimation technique (MSET); model update; power plant

---

## 1. Introduction

The ID fan is a mechanical device that exhausts the gas from other equipment with a wheel rotor to achieve gas compression and gas delivery. ID fans are widely used in ventilation plants including mine, tunnel, and cooling tower. The ID fan is one of the crucial auxiliary equipment in the thermal power plant, where it induces the flue gas generated in the boiler to maintain a certain negative pressure in the furnace. ID fans are usually exposed to high-temperature flue gas and fly ash, which can make the impeller and blades easily eroded; moreover, the bearings are prone to wear and tear because of the mechanical vibration. Thus, faults can occur to ID fans in daily operations. Even worse, ID fan faults may cause an unscheduled downtime of the power plant, seriously affecting the secure operation and economic generation [1, 2].

Early fault detection can identify current status based on historical operating information of process or equipment and give a warning on the abnormal operation, thereby providing enough time to take mitigation measures. Early fault detection strategy has been widely applied in different industrial fields [3-5]. For example, a mechanism model was constructed to access the performance of airport landing lightening system in [6]; Odgaard et al. [7] developed a fault detection model of the coal mill based on energy balance. These models, which are developed according to the physical mechanism, have a high acceptance; however, it is usually difficult to obtain adequate knowledge in practice, thus affecting the prediction accuracy in condition monitoring. Some literature tried to solve the fault detection problem by using historical operating data and they are also known as data-driven models. Shen et al. [8] employed an optimized deep belief network (DBN) technique to diagnose the

bearing faults of the rotating machinery. Dai et al. [9] presented a fault detection strategy of the rolling bearing by combining sparse autoencoder (SAE) and denoising autoencoder (DAE). When developing fault-detection models based on operating data, model parameters need to be determined, which is known as model training. Usually, the model parameters are searched by using training data. In [10], the 10-fold cross-validation support vector machine (SVM) is adopted in the classification of wind turbine faults with data sampled from Supervisory Control and Data Acquisition (SCADA) at an increased frequency. Guo et al. [11] introduced an improved random forest approach in fault recognition of direct current (DC) motors. In general, those state-classification methods divide the operating states into two types: fault states and fault-free states. To construct a precise model with state classification methods, it is important to get the operating data of typical faults of the equipment or process. However, it is difficult to describe the dynamic characteristics of the ID fan with physical mechanism knowledge. Besides, the fault data of the ID fan is rarely seen in practice and not enough to train the model. Considering practical application, the model with excessive adjustable parameters is also not suitable for fault detection.

The multivariate state estimation technique (MSET) is one of the most used data-driven fault detection methods. The MSET model is developed with normal operating data, but unlike ANN, complicated parameter searching is not necessary during the MSET model construction, making it much easier and more feasible to be implemented in practical use. By using MSET, faults can be detected with no need for fault records (which are difficult to obtain), physical mechanism knowledge, and complicated parameter determination [12-14]. MSET can learn the process features with historical normal operations and produce a prediction of the actual measured data sample, which is named the observation. By evaluating the similarity between the predicted value and the observation, operating faults can be identified. Thus, MSET can be used in constructing the early fault detection system of ID fans. MSET models have been widely applied in the industrial field. For example, Zhang et al. [15] introduced the MSET model combined with similarity theory for early fault warning of auxiliary equipment; Cui et al. [16] integrated MSET with principal component analysis (PCA) to monitor the operating condition of an air compressor. Unlike model training of the ANN, the parameters of MSET can be directly solved by using the memory matrix.

MSET is a widely used data-driven method and it mines the features of healthy states from the history database [13-15]. However, like all the data-driven models, the accuracy of the MSET model has an important dependence on the training data, which are called as a memory matrix and represent the relationship among feature variables. If the current operating conditions have a large change and deviate from the historical conditions, the prediction accuracy will be affected. Thus, it is necessary to absorb operating information from newly acquired data and achieve the model update. Model update strategies have been presented to guarantee prediction accuracy. Russell et al. [17] proposed a recursive technique to monitor the batch process online. Liu et al. [18] introduced a moving window kernel principal component analysis method to achieve adaptive detection. Kaneko et al. [19] presented an adaptive prediction model based on a database monitoring index. The index is used to measure information contained in newly acquired data, which determines whether the new data are added to the database. The monitoring index can also be used in the MSET model to update the memory matrix.

The accuracy performance of the MSET model has an important dependence on the historic memory matrix, which consists of a set of operating data to represent normal working conditions. Thus, how to construct an adequate memory matrix is crucial when applying MSET for early fault detection. Like preparing the training data of ANN models, the memory matrix needs to be elaborately designed, which has not received enough attention. In some research, several segments of data were manually selected from the historical operating database and then considered as the memory matrix. Such a method is implemented without any specific indicators; thus, all possible operating conditions cannot be fully covered. Another type of method tried to design the memory matrix with a uniform distribution by evenly selecting samples from historical operating data. Nevertheless, to the best of our knowledge, the memory matrix is usually set to be invariable in previous research. If the observed data are beyond the operating condition described by the memory

matrix, the MSET model will exhibit a decline in accuracy, which is called model degradation. Model degradation is a crucial problem of the practical use of most data-driven models [17,18]. To maintain a high prediction accuracy, adaptive update strategies are usually applied [19]. One commonly used framework for the adaptive update is the moving window strategy, in which the model is constructed with the most recent data of the window with a certain width that moves over time. For example, in [20], the moving window is utilized to optimize the monitoring model based on the dynamic principal component analysis; Sheriff et al. [21] proposed a moving window generalized likelihood ratio test approach to reduce the missed fault detection rate. The just-in-time (JIT) learning is another widely used update method. JIT learning method selects data samples that are similar to the observation from the historical operating database to construct the model; in addition, the model can be updated automatically by storing the newly acquired data into the database. Chen et al. [22] proposed a JIT-support vector regression model to achieve better silicon content prediction. Ahmad et al. [23] applied the JIT strategy to develop an online model. Both moving window and JIT learning strategies need to reconstruct the model with completely reselected data, which has an influence on the computation time and the sensitivity of the model.

To solve the above problems, the update of the MSET model can be implemented incrementally. In the incremental update, the unnecessary sample is replaced and the update is executed once a new measured sample is acquired. The incremental update can reduce the computation cost compared with model reconstruction. In [24], the incremental update was adopted to modify the parameters of the sequential assimilation model. Kaneko et al. [25] defined a database monitoring index to realize incremental updates and solve the model degradation problem. The index is used to measure information contained in newly acquired data, which determines whether the new data are added to the database. The monitoring index can also be used in the MSET model to update the memory matrix.

The main contribution of this study is to introduce an adaptive early fault detection model for ID fans in power plants. The MSET model is adopted to monitor the operating state of the ID fan; furthermore, to solve the problem of the model performance degradation, the memory matrix used for model training is updated incrementally. Based on the prediction errors of the MSET model, potential faults of ID fans are detected by catching the anomalies in operating data, thereby alarming the operators in advance. The remainder of this paper is organized as follows. Section 2 gives a problem description of the early fault detection of ID fans with MSET. Section 3 describes the framework of the proposed adaptive MSET model, including the MSET theory, the adaptive update method, and the early warning approach. Section 4 presents a case study, and the results and discussion are also provided. Section 5 gives the conclusions.

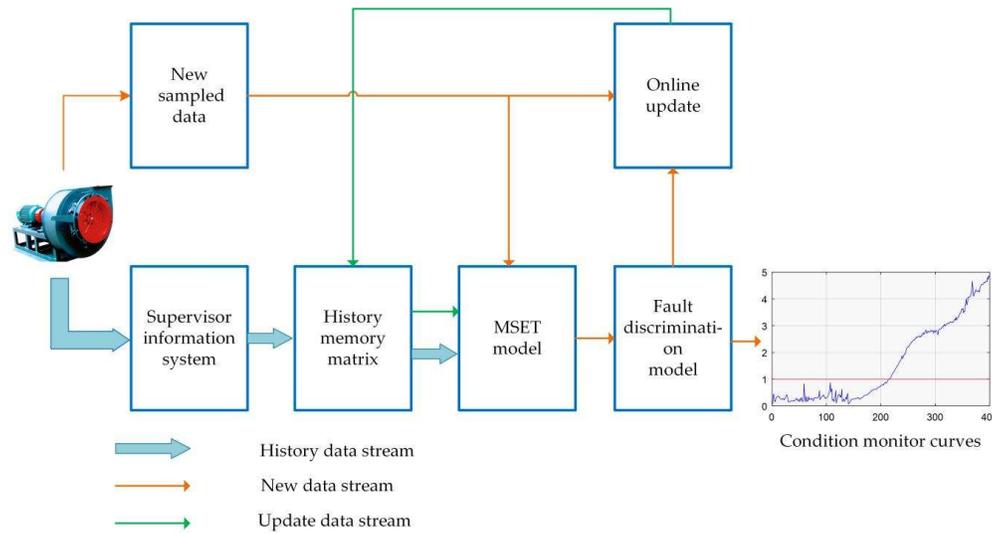
## 2. Problem Description

The MSET method has been widely applied to the fault detection of the equipment including wind turbine gearbox, aircraft, etc., and it is utilized to detect early faults of ID fans in power plants in this study. The historic memory matrix used for the MSET model training consists of data samples that represent normal operating conditions. When current observation deviates from the data in the memory matrix, the MSET model will produce an inaccurate estimation, indicating a failure may occur. However, the operating conditions of the ID fan in power plants are inconstant because of the environmental change and boiler load regulation, and it is difficult to guarantee that all possible normal operating conditions are contained in the memory matrix when selecting training data from the historical database. If the normal operation runs beyond the coverage of the initially obtained memory matrix, the inaccurate prediction will be produced, thus leading to a false alarm. Therefore, the samples in the memory matrix are supposed to be adaptive by covering newly operating information. Furthermore, considering the computation complexity, it is infeasible to construct a memory matrix with all the operating data. It is highly desirable to propose an adaptive model to ensure prediction accuracy and improve fault detection performance.

### 3. Adaptive Early Fault Detection Model Framework

#### 3.1 Model structure

The framework of the adaptive early fault detection model is illustrated in Figure 1. It consists of three parts: offline MSET model, online update, and fault detection model. The offline MSET model calculates the estimation of the current observation based on the initial memory matrix; the online update is used to achieve the update of the memory matrix; and the fault discrimination model performs anomaly detection and produces warnings of early faults. The specific implementation of the proposed model for the fault detection of ID fans is described below. First, with the parameters related to the ID fan as feature variables, normal operating data are collected from the supervisor information system (SIS) to construct the memory matrix. Then, the MSET model is established by using the memory matrix, and the estimation of the current observation is calculated. According to the residual between the observation and its estimation, the fault discrimination model gives the evaluation result to detect early faults. Finally, the newly obtained data are analyzed in the online update part. If they are identified to be in normal operating conditions according to the fault evaluation results, the memory matrix is updated to cover the new operating information from these data, and furthermore, the MSET model is updated accordingly based on the new memory matrix.



**Figure 1.** Framework of the adaptive early fault detection model

#### 3.2. MSET Theory

MSET was first proposed by Argonne National Laboratory [26], and it is widely applied in condition monitoring, fault detection, and residual life prediction. The MSET model uses normal operating data, which are stored in the memory matrix, to characterize the monitored process or equipment. Considering that the memory matrix is constructed by a dataset of  $n$  variables at  $m$  time stamps, which are collected from the plant historical operating database, the memory matrix can be denoted as follows:

$$\mathbf{D} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_m)] = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \cdots & x_2(t_m) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_m) \end{bmatrix} \quad (1)$$

where  $x_i(t_j)$  ( $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ ) is the value of the variable  $x_i$  at time stamp  $t_j$ . Each column of  $\mathbf{D}$  represents a normal operation state at a certain time stamp, and each row represents the values of a certain variable at different time stamps. Thus,  $\mathbf{D}$  consists of  $m$  vectors  $\mathbf{x}(t_j)$  that represent different normal operating states of a process.

For a given operating state, denoted as the observation vector  $\mathbf{x}_o$ , the MSET model will produce its estimation  $\hat{\mathbf{x}}_o$  by taking dot products between a weight vector  $\mathbf{w}$  and the memory matrix  $\mathbf{D}$ ; that is:

$$\begin{aligned}\hat{\mathbf{x}}_o &= \mathbf{D}\mathbf{w} \\ &= \mathbf{D} \cdot [w_1, w_2, \dots, w_m]^T \\ &= w_1\mathbf{x}(t_1) + w_2\mathbf{x}(t_2) + \dots + w_m\mathbf{x}(t_m)\end{aligned}\quad (2)$$

As shown in Equation (2), the estimation vector  $\hat{\mathbf{x}}_o$  can be considered as a linear combination of  $m$  normal operating states with the weight vector  $\mathbf{w}$ . The estimation vector  $\hat{\mathbf{x}}_o$  is supposed to be similar to the observation vector  $\mathbf{x}_o$  if the equipment is in normal operations; that is, the residual  $\boldsymbol{\varepsilon}$  between  $\hat{\mathbf{x}}_o$  and  $\mathbf{x}_o$  should be minimized. The residual  $\boldsymbol{\varepsilon}$  can be written as follows:

$$\boldsymbol{\varepsilon} = \mathbf{x}_o - \hat{\mathbf{x}}_o \quad (3)$$

Then the quadratic sum of  $\boldsymbol{\varepsilon}$  is written with respect to the weight vector  $\mathbf{w}$ ; that is:

$$\begin{aligned}J(\mathbf{w}) &= \|\boldsymbol{\varepsilon}\|^2 = (\mathbf{x}_o - \hat{\mathbf{x}}_o)^T(\mathbf{x}_o - \hat{\mathbf{x}}_o) \\ &= (\mathbf{x}_o - \mathbf{D}\mathbf{w})^T(\mathbf{x}_o - \mathbf{D}\mathbf{w}) = \sum_{i=1}^n \left( \mathbf{x}_o(t_i) - \sum_{j=1}^m w_j \mathbf{D}_{ij} \right)^2\end{aligned}\quad (4)$$

To get the minimum of the residual, we need to take the partial derivative of the  $J(\mathbf{w})$  and make it equal to 0; that is:

$$\frac{\partial J(\mathbf{w})}{\partial w_j} = -2 \sum_{i=1}^n \left( \mathbf{x}_o(t_i) - \sum_{j=1}^m w_j \mathbf{D}_{ij} \right) \mathbf{D}_{ik} = 0 \quad (5)$$

Namely,

$$\sum_{i=1}^n \mathbf{x}(t_i) \mathbf{D}_{ik} = \sum_{i=1}^n \sum_{j=1}^m w_j \mathbf{D}_{ij} \mathbf{D}_{ik} = \sum_{j=1}^m \left( \sum_{i=1}^n \mathbf{D}_{ij} \mathbf{D}_{ik} \right) w_j \quad (6)$$

Equation (6) can be written in a matrix form:

$$\mathbf{D}^T \mathbf{D} \mathbf{w} = \mathbf{D}^T \mathbf{x}_o \quad (7)$$

Then, the weight vector  $\mathbf{w}$  can be calculated accordingly:

$$\mathbf{w} = (\mathbf{D}^T \mathbf{D})^{-1} (\mathbf{D}^T \mathbf{x}_o) \quad (8)$$

As shown in Equation (8), the inversion of  $\mathbf{D}^T \mathbf{D}$  is calculated to obtain the weight vector. The dot product calculation is replaced with a nonlinear operation, such as the Euclidean distance operation, to guarantee the existence of the inversion. The Euclidean distance is expressed as follows:

$$\varphi(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (9)$$

Then Equation (9) can be expressed as:

$$\mathbf{w} = \mathbf{K}^{-1} \cdot \mathbf{f}(\mathbf{x}_o) \quad (10)$$

where  $\mathbf{K}$  is an  $m$ -by- $m$  matrix, and  $\mathbf{K}_{ij} = \varphi(\mathbf{D}_i, \mathbf{D}_j)$ ,  $\mathbf{f}$  is an  $m$ -by-1 function vector, and  $\mathbf{f}_i = \varphi(\mathbf{D}_i, \mathbf{x}_o)$ ,  $\mathbf{D}_i = \mathbf{x}(t_i)$ .

Thus, the weight vector can represent the similarity of  $\mathbf{x}_o$  and  $\mathbf{D}$ . Assuming that the observation vector  $\mathbf{x}_o$  is very similar to the vector  $\mathbf{x}(t_i)$  in  $\mathbf{D}$ , the weight  $w_i$  can have a large value. After submitting Equation (10) into Equation (2), the estimation vector  $\hat{\mathbf{x}}_o$  can be written as:

$$\hat{\mathbf{x}}_o = \mathbf{D} \mathbf{K}^{-1} \mathbf{f}(\mathbf{x}_o) \quad (11)$$

If the observation  $\mathbf{x}_o$  is consistent with most vectors in memory matrix  $\mathbf{D}$ , the estimation vector  $\hat{\mathbf{x}}_o$  will be close to  $\mathbf{x}_o$ . On the contrary, if a potential fault occurs, the observation  $\mathbf{x}_o$  will diverge from normal operating states, and an estimation vector with large deviations is generated.

### 3.3. Fault Detection

The estimation error, which is produced by the MSET model, can indicate the deviation between the current state from norm operations. Thus, in this study, the norm value of the residual  $\boldsymbol{\varepsilon}$ , which is denoted as the deviation  $\delta$ , is utilized to identify early faults of ID fans. The deviation  $\delta$  is calculated as follows:

$$\delta(\boldsymbol{x}) = \sqrt{\sum_{i=1}^n (\boldsymbol{x}(i) - \hat{\boldsymbol{x}}(i))^2} \quad (12)$$

where  $\hat{\boldsymbol{x}}$  is the estimation of  $\boldsymbol{x}$  and can be calculated by Equation (11).

When the deviation  $\delta$  is larger, the dissimilarity between the observation vector  $\boldsymbol{x}$  and the estimation vector  $\hat{\boldsymbol{x}}$  is higher. Based on the deviation, we definite the similarity index  $s$ , which is limited to the range of  $(0, 1)$ , for better detection of ID fan faults.

$$s(\boldsymbol{x}) = 1 - \frac{\sqrt{\sum_{i=1}^n \mu_i (\boldsymbol{x}(i) - \hat{\boldsymbol{x}}(i))^2}}{\theta} \quad (13)$$

where  $\theta$  is a given threshold,  $\mu_i$  is the weight of the  $i$ th variable. If the equipment runs in normal operations, the estimation  $\hat{\boldsymbol{x}}$  will have a high similarity to the observation  $\boldsymbol{x}$ , and thus  $s$  is close to 1. If the equipment runs in abnormal operations, the similarity will be beyond the normal range and  $s$  is close to 0. Therefore, the similarity index can be taken as a visual indicator for fault judgment.

Moreover, the similarity value is easily influenced by the random disturbance. The moving window method is applied to optimize the similarity value. Considering a period of  $l$  data samples, the similarity sequences  $[\kappa_1, \kappa_2, \dots, \kappa_T, \dots, \kappa_l]$  can be calculated by Equation (13). Assuming that the moving window width is  $T$ , the mean similarity value of each sub-window is obtained. That is,

$$\bar{\kappa}_j = \frac{\sum_{i=j}^{T+j-1} \kappa_i}{T} \quad (14)$$

where  $j = 1, 2, \dots, l - T + 1$ .

The warning threshold  $\kappa_s$  can be set to the minimum of the similarity sequences multiplied by a coefficient; that is:

$$\kappa_s = \gamma \cdot \min \{\bar{\kappa}_1, \bar{\kappa}_2, \dots, \bar{\kappa}_{l-T+1}\}, \quad (15)$$

where  $\gamma$  is a warning threshold coefficient and is generally less than 1. A warning signal will be produced when the similarity of the current observation falls below the threshold  $\kappa_s$ .

### 3.4. MSET Update Strategy

The ID fan usually runs in variant conditions because of the environmental change and boiler load regulation. The MSET model trained with the memory matrix selected from certain operations fails to represent the whole characteristics of the ID fan. When it runs in a different operating condition, the observed value may quite differ from the data in the memory matrix, leading to inaccurate estimations. Thus, it is necessary to implement an update strategy to make the memory matrix cover new information to update the MSET model.

However, if all the newly acquired data are added into the memory matrix, the size of  $\boldsymbol{D}$  will be constantly increased, leading to a heavy computation. Thus, the update strategy can be implemented incrementally. Once a new observation is obtained, it is analyzed to check whether it satisfies the updating criteria. If yes, then it is added into  $\boldsymbol{D}$ ; and meanwhile, an old sample is selected and deleted from  $\boldsymbol{D}$  to maintain the matrix size unchanged. After a new  $\boldsymbol{D}$  is obtained, the MSET model is calculated accordingly.

The flowchart of the MSET update is shown in Figure 2. The first step is to check whether the newly acquired data sample belongs to normal operations. To achieve this, the similarity index  $s$  is calculated according to Equations (8)-(13). If the similarity index  $s$  exceeds the warning threshold  $\kappa_s$ , this sample is considered to reflect an abnormal behavior and the next data sample is focused on; otherwise, its estimation deviation  $\delta$  is analyzed next. If the deviation  $\delta$  keeps in an instant growth, this data sample is considered to in the fault-development stage and then we move to the next data sample. After a new data sample that is identified to a normal operation, the next step is to check

whether this sample contains new operating information that is not covered by the memory matrix  $\mathbf{D}$ . The data quality index (DQI), which builds the relationship between the new data sample and the memory matrix, is defined to select the data sample that represents new operating information. DQI is expressed as follows:

$$r(\mathbf{x}_o) = \frac{\prod_{i=1}^n (x_o(i) - \bar{x}(i))}{\sum_{i=1}^n (x_o(i) - \bar{x}(i))^2} \quad (16)$$

where  $r(\mathbf{x}_o)$  is the DQI value of the data sample  $\mathbf{x}_o$ ,  $x_o(i)$  is the  $i$ th variable value in the new data sample,  $\bar{x}(i)$  is the average value of the  $i$ th variable in  $\mathbf{D}$ ,  $n$  is the number of variables. The data sample with a large DQI is considered to help increase the operating information in  $\mathbf{D}$ . A threshold  $T_r$  is set to determine whether the data sample is added into  $\mathbf{D}$ :

$$T_r = \max_{1 \leq j \leq m} \left( \frac{\prod_{i=1}^n (x_j(i) - \bar{x}(i))}{\sum_{i=1}^n (x_j(i) - \bar{x}(i))^2} \right) \quad (17)$$

where  $m$  is the number of data samples in  $\mathbf{D}$ . If the DQI value of the new sample has a larger value than  $T_r$ , it is added into  $\mathbf{D}$ . Meanwhile, to keep the matrix size unchanged, the oldest data sample, which may have become invalid, is deleted from  $\mathbf{D}$ . The update procedure is shown as follows:

1. Construct the memory matrix  $\mathbf{D}$  with the initial data and develop the MSET model by using Equations (1)-(11);
2. Acquire a new data sample and obtain the estimation by using the MSET model;
3. Calculate the similarity  $s$ , and if it exceeds the warning threshold  $\kappa_s$ , go to 2, else go to 4;
4. Calculate the deviation  $\delta$ , and if it exceeds the given deviation threshold  $T_\delta$ , go to 2, else go to 5;
5. Calculate the DQI value  $r$ , and if it exceeds the threshold  $T_r$ , go to 6, else go to 2;
6. Update the memory matrix  $\mathbf{D}$  with this new data sample;
7. Delete the oldest data sample in  $\mathbf{D}$  and go to 2.

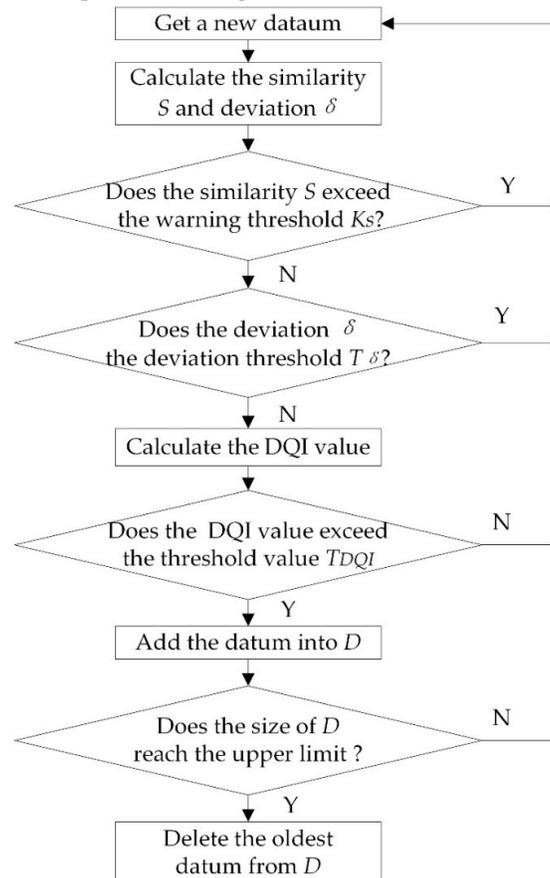


Figure 2. The update procedure

## 4 Industrial Case Study

### 4.1. Data Preparation

The operating data of the ID fan are acquired from the SIS in a 300 MW coal-fired power plant. More than 20 parameters related to the ID fan operation are recorded in the database, but only 7 fault-sensitive parameters are selected as the feature variables to develop the MSET model. These parameters include power, pressure, temperature, and vibrations. The selected parameters are illustrated in Table.1.

Table.1 Parameters related to the fault detection of ID fans.

Parameter	Unit
Motor power	<i>kW</i>
Outlet pressure	<i>MPa</i>
Front bearing temperature	<i>°C</i>
Horizontal vibration of waist bearing	<i>mm</i>
Vertical vibration of waist bearing	<i>mm</i>
Horizontal vibration of side bearing	<i>mm</i>
Vertical vibration of side bearing	<i>mm</i>

These parameters are sampled with a period of 10 minutes, with total 1800 samples obtained. At the 1560<sup>th</sup> data sample, a cumulative deviation with a step length of 0.06 is artificially added to the front bearing temperature to simulate the abnormal temperature state caused by the fault of the induced draft fan. The first 1400 samples are considered as the available training dataset to construct the memory matrix  $\mathbf{D}$  and the next 140 samples as the test dataset to observe the prediction performance of the model and determine the warning threshold  $\kappa_s$ . Both the training and test data are obtained during normal operation. The remaining 260 samples cover the span from normal operation to fault occurring are considered as the validation dataset to verify the fault-detection performance of the model. The partition of the training dataset, test dataset, and validation dataset is shown in Figure 3.

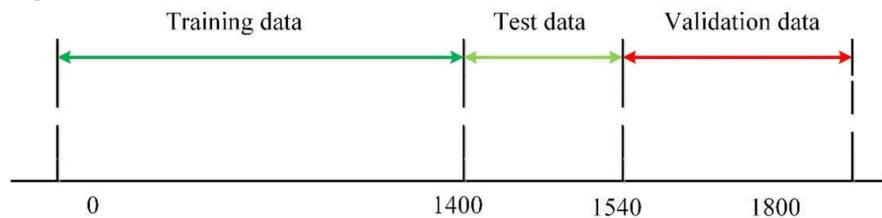


Figure 3. The division of operating data

Note that the feature variables have different ranges; thus, the acquired data need to be normalized to the same range. The z-score standardization is utilized here, and it is defined as follows:

$$x' = \frac{x - \bar{x}}{\sigma} \quad (18)$$

where  $x$  is the acquired value,  $\bar{x}$  is the average of  $x$ ,  $\sigma$  is the standard deviation, and  $x'$  is the standardized value.

### 4.2. Model Construction

The memory matrix  $\mathbf{D}$  is firstly constructed when developing the MSET model. The data samples in  $\mathbf{D}$  should cover all possible operating conditions to characterize the ID fan system. However, the historical operating data recorded in SIS is large in volume, and if the memory matrix  $\mathbf{D}$  is constructed with all operating data, the computational complexity will increase greatly, leading

to a decline in real-time warning performance. Thus, the memory matrix  $\mathbf{D}$  is designed to consist of 200 samples, which are selected according to a well-designed criterion. In this way, the samples in the memory matrix  $\mathbf{D}$  can be representative to guarantee good prediction accuracy of the developed MSET model. The detailed selection criterion can be found in [27]. There are many sample-selection methods, such as uniform sampling method, random sampling method, etc. In this paper, the memory matrix  $\mathbf{D}$  was constructed based on an evaluation index. The evaluation index  $P$  can reflect the data distribution and richness of the operating conditions in  $\mathbf{D}$ , as defined as follow:

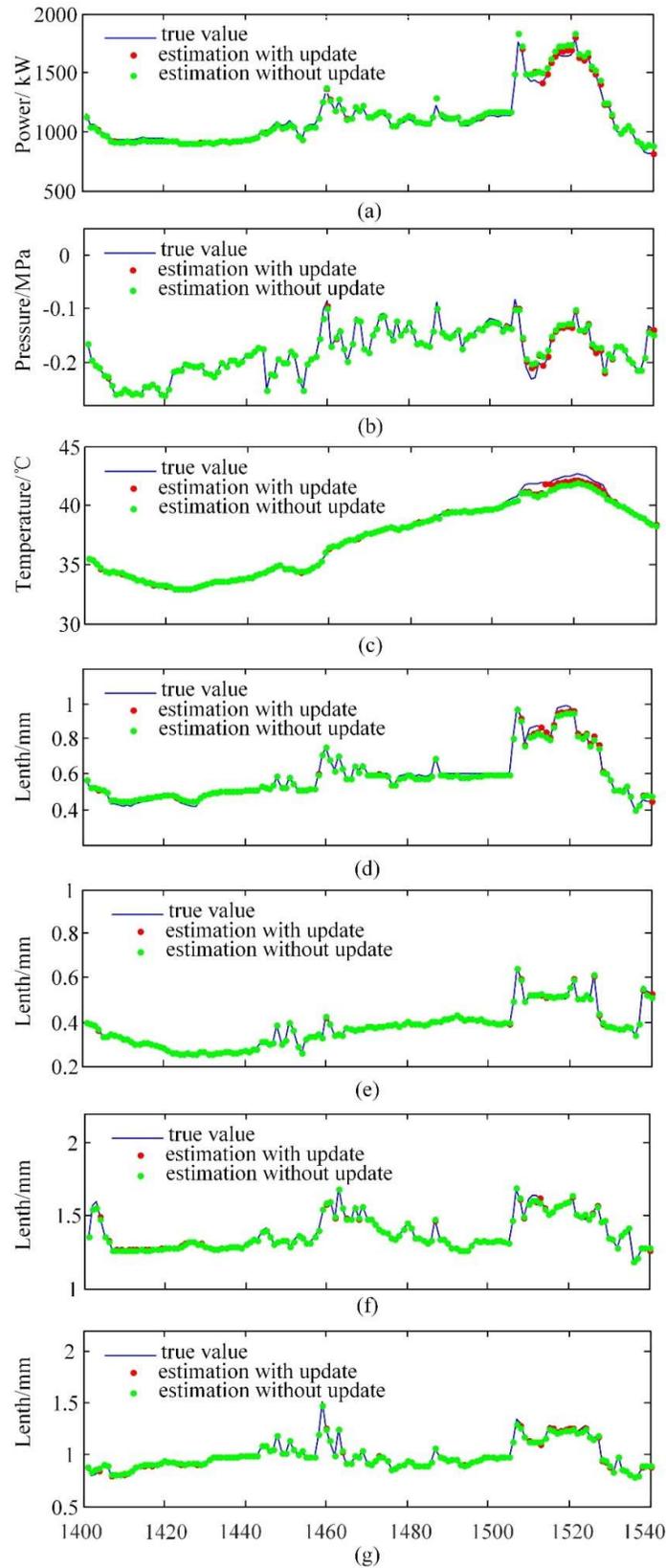
$$P(\mathbf{X}) = \left( \frac{1}{d_{max}} \cdot \max \left( \min_{1 \leq i, j \leq s} d_{ij} \right) \right)^{\gamma_1} \cdot \frac{1}{(1 + L_2(\mathbf{X}))^{\gamma_2}} \quad (19)$$

where  $\gamma_1$  and  $\gamma_2$  are the weight coefficients,  $d_{ij}$  is the euclidian distance of two data samples in  $\mathbf{D}$ ,  $d_{max}$  is the maximum euclidian distance of data samples,  $L_2(\mathbf{X})$  is the centered deviation of data samples,  $s$  is the number of data samples. When the evaluation index  $P$  is larger, it is believed that the data distribution is more uniform and the richness of the operating conditions covered in  $\mathbf{D}$  is higher. The discrete particle swarm optimization algorithm (DPSSO) was adopted to search the optimum value of the evaluation index  $P$  and the optimum dataset to construct the memory matrix.

#### 4.3. Modeling Results and Discussion

The MSET model was developed based on Equations (1)-(11). To validate the proposed update strategy two models were constructed: the MSET with adaptive update strategy (denoted as M1) and the MSET model without update (denoted as M2). The prediction results of the test dataset are calculated by using models M1 and M2. During the prediction of model M1, the estimation result of the new data sample in the test dataset was analyzed and the new data sample was decided to add into the memory matrix  $\mathbf{D}$  based on the MSET update strategy. The data samples in  $\mathbf{D}$  were updated when the new operating condition occurred. Whereas, during the prediction of the model M2, the estimation results were calculated without such an update procedure. The prediction results of the test dataset are illustrated in Figure 4.

Figure 4 (a) gives the estimations and actual measured values of the motor power by using models M1 and M2. The estimations are basically coincident with actual measured values of the motor power from the 1401<sup>st</sup> to the 1512<sup>th</sup> data samples. Furthermore, the estimations produced by the model M1 are much closer to the measured values of the motor power from the 1513<sup>th</sup> to the 1522<sup>nd</sup> data samples. The prediction errors are within 0.027, which indicates that the model M1 has higher prediction accuracy. Similar results can be obtained from Figure 4 (b) to (g). In Figure 4 (b), the model M1 and M2 produce almost the same estimations of outlet pressure from the 1401<sup>st</sup> to the 1509<sup>th</sup> data sample. At the 1510<sup>th</sup> data sample, a small deviation between the estimation and actual measured values occurs for both two models, with a prediction error of 0.094 and 0.1175 respectively. Furthermore, the estimations of the model M1 gradually approximate actual measured values from the 1511<sup>th</sup> to the 1514<sup>th</sup> data sample. The prediction error of this data segment drops to 0.0072. However, the estimations of the model M2 still maintain a small deviation from the actual measurements, with a prediction error of 0.0672. It indicates that the memory matrix of model M1 covered new operating conditions after the update, and thus the prediction accuracy is improved. Figure 4 (c) shows the prediction results of front bearing temperature. The estimations of the two models are basically the same except for the 1515<sup>th</sup> to the 1525<sup>th</sup> data sample, where the estimations produced by model M1 are more approaching the actual measured values.



**Figure 4.** Estimation results of test dataset by using the model M1 and M2: (a) motor power, (b) outlet pressure, (c) front bearing temperature, (d) heretical vibration of waist bearing, (e) vertical vibration of waist bearing, (f) heretical vibration of side bearing, (g) vertical vibration of side bearing.

The estimation results of the vibration parameters are shown in Figure 4 (d) to (g). The estimations of the two models M1 and M2 are generally close to actual measured values, but the estimations of some samples produced by the model M1 are more accurate than those produced by the model M2. For horizontal vibration of waist bearing (see Figure 4(d)), the prediction errors of model M1 are smaller than those of model M2 around the 1515<sup>th</sup> and the 1520<sup>th</sup> data sample; for vertical vibration of waist bearing (see Figure 4(e)), model M1 and model M2 produce almost identical estimations; for horizontal vibration of side bearing (see Figure 4(f)), model M1 has lower prediction errors than model M2 when predicting the 1515<sup>th</sup> data sample; for vertical vibration of side bearing (see Figure 4(g)), the estimations of model M1 are closer to actual measurements around the 1520<sup>th</sup> data sample.

In addition, three kinds of estimation errors of the model M1 and M2 were calculated as follows:

$$SSE_r = \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (19)$$

$$MSE_r = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n} \quad (20)$$

$$RSME_r = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \quad (21)$$

where  $x_i$  is the true value at time stamp  $i$ ,  $\hat{x}_i$  is the estimation at time stamp  $i$ ,  $n$  is the number of data samples. Table 2 shows the average relative estimation errors of the two models.

Tables 1 to 7 illustrates detailed prediction errors of the feature variables. It is shown that model M1 and model M2 produce almost the same-level prediction errors from the 1401<sup>st</sup> to the 1510<sup>th</sup> data samples. For example, for the variable of outlet pressure, the RMSE values of the two models are 0.0881 and 0.0975, respectively. However, from the 1511<sup>th</sup> to the 1540<sup>th</sup> data sample, the estimation errors of model M1 are much lower than those of model M2. The RMSE of the outlet pressure produced by M2 is 0.3264, but the RMSE produced by model M1 is only 0.1938, about 40.9% reduced. Thus, one can find that the prediction accuracy is greatly improved after the proposed update strategy is adopted. We only analyzed the variable of outlet pressure, and the accuracy improvement of model M1 is still valid to other variables.

Table 1. The estimation errors of the motor power.

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.8796	0.0080	0.0894	0.4287	0.0143	0.1195	1.3083	0.0093	0.0967
M2	0.9200	0.0084	0.0915	1.0353	0.0345	0.1858	1.9533	0.0104	0.1182

Table.2 The estimation errors of the outlet pressure

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.9124	0.0083	0.0911	0.3779	0.0126	0.1122	1.2903	0.0092	0.0960
M2	1.1761	0.0107	0.1034	0.9722	0.0324	0.1800	2.1483	0.0153	0.1239

Table.3 The estimation errors of the front bearing temperature

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.8535	0.0078	0.0881	1.1265	0.0375	0.1938	1.9799	0.0141	0.1189
M2	1.0455	0.0095	0.0975	3.1960	0.1065	0.3264	4.2416	0.0303	0.1741

Table.4 The estimation errors of the horizontal vibration of waist bearing

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.9841	0.0089	0.0946	0.4417	0.0147	0.1213	1.4258	0.0102	0.1009
M2	1.1162	0.0101	0.1007	0.8621	0.0287	0.1695	1.9784	0.0141	0.1189

Table.5 The estimation errors of the vertical vibration of waist bearing

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.1463	0.0013	0.0365	0.2068	0.0069	0.0830	0.3532	0.0025	0.0502
M2	0.1511	0.0014	0.0371	0.2382	0.0079	0.0891	0.3893	0.0028	0.0527

Table.6 The estimation errors of the horizontal vibration of side bearing

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.3093	0.0028	0.0530	0.2952	0.0098	0.0992	0.6045	0.0043	0.0657
M2	0.3335	0.0030	0.0551	0.3597	0.0120	0.1095	0.6932	0.0050	0.0704

Table.7 The estimation errors of the vertical vibration of side bearing

Models	From 1401 <sup>th</sup> to 1510 <sup>th</sup> data samples			From 1511 <sup>th</sup> to 1540 <sup>th</sup> data samples			Entire test dataset		
	SSE	MSE	RMSE	SSE	MSE	RMSE	SSE	MSE	RMSE
M1	0.0558	0.0051	0.0711	0.1806	0.0060	0.0776	0.7363	0.0053	0.0725
M2	0.6681	0.0061	0.0779	0.3379	0.0113	0.1061	1.0060	0.0072	0.0848

Considering the estimations of all the feature variables, the similarity index of the test dataset can be calculated by Equation (13), and the results are shown in Figure 5. It can be found that the similarity index values calculated by model M1 and model M2 generally coincide from the 1401<sup>st</sup> to the 1497<sup>th</sup> data sample. The similarity index of model M1 is much closer to 1.0 than that of model M2 since the 1498<sup>th</sup> data sample. If we set the warning threshold coefficient  $\gamma$  to 0.83, the fault warning threshold of model M1 and model M2 are respectively 0.7884 and 0.7628. The simulation results of model M1 and M2 about the test dataset show that model M1 can achieve better estimation than model M2 during normal operation.

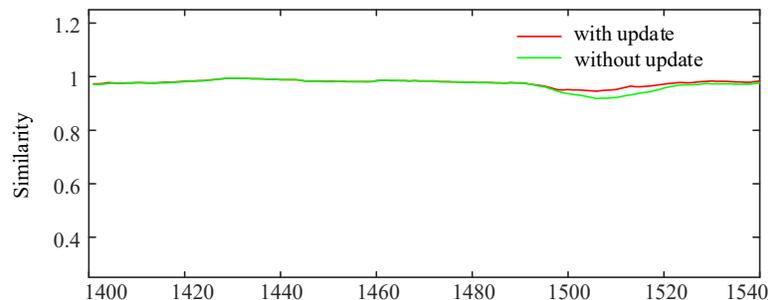


Figure 5. Similarity index values of the test dataset

#### 4.4. Fault Detection Validation

The validation dataset, which consisted of 260 data samples (from the 1541<sup>st</sup> data sample to the 1800<sup>th</sup> data sample), was used to evaluate the fault-detection performance of the proposed model. To

simulate the abnormal temperature variation caused by a fault, a cumulative deviation with a step length of 0.06 is artificially added to the front bearing temperature since the 1560<sup>th</sup> data sample. The similarity index values of the validation dataset are calculated by using models M1 and M2, and the results are shown in Figure 6. The similarity index values calculated by the two models show a general downward trend. The similarity index of model M1 drops from 0.9832 to 0, and the similarity index of model M2 drops from 0.9832 to 0.2715. Thus, both two models can capture the abnormal increasing trend of the front bearing temperature. In addition, the similarity index calculated by model M1 decreases at a faster rate compared with model M2. The similarity index of model M1 decreases to the warning threshold 0.7884 after the 1602<sup>nd</sup> data sample, indicating that abnormal operation is identified. The similarity index of model M2 reaches the warning threshold of 0.7628 after the 1630<sup>th</sup> data sample. Thus, we can find that the model M1 can detect the abnormality of the ID fan 28 data points (i.e. 4 hours and 40 minutes) earlier than M2, which indicates that the proposed adaptive update strategy is beneficial to advance the fault warning time.

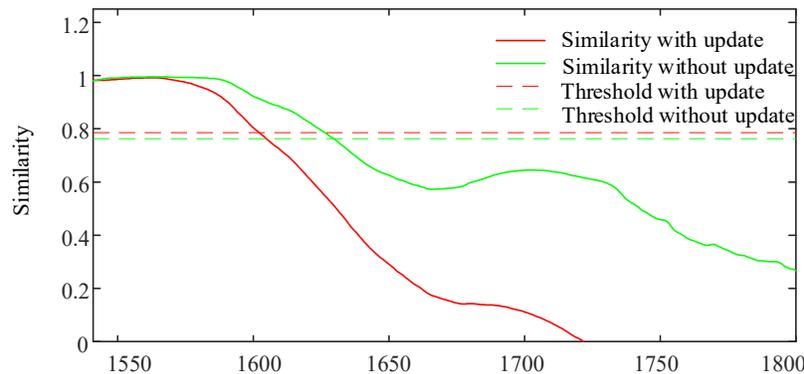


Figure 6. Similarity of the validation dataset

## 5. Conclusion

To achieve the high safety and efficiency of the ID fan operation in the thermal power plant, an adaptive early fault detection method is proposed in this study. The experiments on real-time data of the ID fan in a 300 MW coal-fired power plant show that the adaptive early fault detection model can detect the fault before it occurs and provide time for taking emergency measures. Firstly, the operating data was divided into three parts: training dataset, test dataset, and validation dataset. The training dataset was used to construct the offline MSET model. Every single datum in the test dataset and validation dataset was sent into the current MSET model to calculate its estimation at first. Then, the fault discrimination model analyzed the difference between the datum and its estimation and identified the fault that might be occurred. Based on the values provided by the fault discrimination model, the online update module calculated the DQI value of the data sample and decided whether it should be added into the historical memory matrix  $D$  to update the MSET model. Besides, the early fault detection model without updating  $D$  is constructed in comparison. Finally, the modeling results on the test dataset showed that the adaptive early fault detection model has a higher prediction accuracy than the model without updates, and the modeling results on the validation dataset showed that the adaptive fault early warning model is more sensitive to the fault of the ID fan and can achieve an earlier warning than the model without updates.

**Funding:** This research was funded by the NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA, grant number 52061635102 and the FUNDAMENTAL RESEARCH FUNDS FOR THE CENTRAL UNIVERSITIES, grant number 2020MS076.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bulloch, J.H.; Callagy A.G. An in situ wear-corrosion study on a series of protective coatings in large induced draft fans. *Wear* **1999**, *233*, 284-292.
2. Das K.K.; Acharya D. Optimal replacement policy for induced draft fans. *Microelectronics Reliability* **1988**, *28*, 519-523.
3. Jalil, B.; Leone, G.R.; Ouhrouche, M.; Martinelli, M.; Moroni, D.; Pascali, M.A.; Berton, A. Fault Detection in Power Equipment via an Unmanned Aerial System Using Multi Modal Data. *Sensors* **2019**, *19*, 3014.
4. Gao, L.; Li, D.H.; Li, D.; Yao, L.L.; Liang, L.M.; Gao, Y.N. A Novel Chiller Sensors Fault Diagnosis Method Based on Virtual Sensors. *Sensors* **2019**, *19*, 3013.
5. Zhang, M.; Wang, T.Z.; Tang, T.T.; Liu, Z.; Claramunt, C. A Synchronous Sampling Based Harmonic Analysis Strategy for Marine Current Turbine Monitoring System under Strong Interference Conditions. *Energies* **2019**, *12*, 2117.
6. Niblock, J.; Peng, J.X.; McMenemy, K.; Irwin, G.W. Fast model-based feature matching technique applied to airport lighting *IET Science, Measurement & Technology* **2008**, *2*, 160-176.
7. Odgaard, P. F.; Mataji, B. Fault Detection in Power Equipment via an Unmanned Aerial System Using Multi Modal Data. *Control Engineering Practice* **2008**, *16*, 909-921.
8. Shen, C.; Xie, J.; Wang, D.; Jiang, X.; Shi, J.; Zhu, Z. Improved Hierarchical Adaptive Deep Belief Network for Bearing Fault Diagnosis. *Appl. Sci.* **2019**, *9*, 3374.
9. Dai, J.; Tang, J.; Shao, F.; Huang, S.; Wang, Y. Fault Diagnosis of Rolling Bearing Based on Multiscale Intrinsic Mode Function Permutation Entropy and a Stacked Sparse Denoising Autoencoder. *Appl. Sci.* **2019**, *9*, 2743.
10. Vidal, Y.; Tutivén C.; Pozo, F. Wind Turbine Multi-Fault Detection and Classification Based on SCADA Data. *Energies* **2018**, *11*, 3018.
11. Guo, Z.; Liu, M.; Qin, H.; Li, B. Mechanical Fault Diagnosis of a DC Motor Utilizing United Variational Mode Decomposition, SampEn, and Random Forest-SPRINT Algorithm Classifiers. *Entropy* **2019**, *21*, 470.
12. Suh, S.; Lee, H.; Jo, J.; Lukowicz, P.; Lee, Y.O. Generative Oversampling Method for Imbalanced Data on Bearing Fault Detection and Diagnosis. *Appl. Sci.* **2019**, *9*, 746.
13. Caesarendra, W.; Lee, J.M.; Ha, J.M.; Choi, B.K. Slew Bearing Early Damage Detection based on Multivariate State Estimation Technique and Sequential Probability Ratio Test, IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Busan, SOUTH KOREA, JUL 07-11, 2015; IEEE: NEW YORK, USA, 2015; 1161-1166.
14. Guo, P.; Bai, N. Wind Turbine Gearbox Condition Monitoring with AAKR and Moving Window Statistic Methods. *Energies* **2011**, *4*, 2077-2093.
15. Zhang, W.; Liu, J.; Gao, M.; Pan, C.; Huusom, J.K. A fault early warning method for auxiliary equipment based on multivariate state estimation technique and sliding window similarity. *Computers in Industry* **2019**, *107*, 67-80.
16. Cui, C.; Li, W.; Yang, Y.; Kuang, X.; Xiao, Y. A novel fault measure and early warning system for air compressor. *Measurement* **2019**, *135*, 593-605.
17. Liu, Y.; Gao, Z.; Li, P.; Wang, H. Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. *Industrial & Engineering Chemistry Research* **2012**, *51*, 4313-4327.
18. Kadlec, P.; Grbić, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Computers & Chemical Engineering* **2011**, *35*, 1-24.
19. Shao, W.; Tian, X.; Wang, P.; Deng, X.; Chen, X. Online soft sensor design using local partial least squares models with adaptive process state partition. *Chemometrics and Intelligent Laboratory Systems* **2015**, *144*, 108-121.
20. Ammiche, M.; Kouadri, A.; Bensmail, A. A modified moving window dynamic PCA with Fuzzy Logic Filter and application to fault detection. *Chemometrics and Intelligent Laboratory Systems* **2018**, *117*, 100-113.
21. Sheriff, M. Z.; Mansouri, M.; Karim, M. N.; Nounou, H.; Nounou, M. Fault detection using multiscale PCA-based moving window GLRT. *Journal of Process Control* **2017**, *54*, 47-64.
22. Chen, K.; Liang, Y.; Gao, Z.; Liu, Y. Just-in-Time Correntropy Soft Sensor with Noisy Data for Industrial Silicon Content Prediction. *Sensors* **2017**, *17*, 1830.
23. Ahmad, I.; Ayub, A.; Mohammad, N.; Kano, M. Data-Based Prediction and Stochastic Analysis of Entrained Flow Coal Gasification under Uncertainty. *Sensors* **2019**, *19*, 1626.

24. Jiang, J.; Xiao, Z.; Wang, J.; Song, J. Sequential Method with Incremental Analysis Update to Retrieve Leaf Area Index from Time Series MODIS Reflectance Data. *Remote Sens* **2014**, *6*, 9194-9212.
25. Kaneko, H.; Funatsu, K. Adaptive database management based on the database monitoring index for long-term use of adaptive soft sensors. *Chemometrics and Intelligent Laboratory Systems* **2015**, *146*, 179-185.
26. Gross, K.C.; Singer, R.M.; Wegerich, S.W.; Herzog, J.P. Application of a model-based fault detection system to nuclear plant signals. In *Proceedings of 9th International Conference on Intelligent Systems Application to Power System*, Seoul, Korea, 6-10 July 1997; pp. 212-218.
27. Lv, Y., Fang, F., Yang, T. T., Liu J.Z. An early fault detection method for induced draft fans based on MSET with informative memory matrix selection. *ISA Transactions* **2020**, *102*, 325-334.