Energy 211 (2020) 118555

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A data-based approach for benchmark interval determination with varying operating conditions in the coal-fired power unit

Jing Xu [a, *], Dapeng Bi [b, **], Suxia Ma [a], Jin Bai [b]

[a] College of Electrical and Power Engineering, Taiyuan University of Technology, No.79 of Yingzexi Street, Taiyuan, Shanxi, 030024, PR China
[b] State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, No. 27 of Taoyuan South Road, Taiyuan, Shanxi, 030001, PR China

ARTICLE INFO

Article history:
Received 10 April 2020
Received in revised form
13 July 2020
Accepted 28 July 2020
Available online 15 August 2020

Keywords:
Coal-fired power plant
Performance degradation
Benchmark
K-means
Gaussian mixture model

ABSTRACT

The modern coal-fired power units in China are mostly operated in a flexible manner. However, flexible operation results in performance degradation, energy-efficiency penalties, and increased energy consumption, which necessitates the detection of performance degradation to save energy. This paper presents a model for detecting the performance degradation of coal-fired power units by determining the benchmark intervals of variables under varying operating conditions using data-mining methods. The K-means clustering method is employed to categorize the operating conditions according to the similarity of historical operational data. Gaussian mixture model is adopted to determine the benchmark interval with respect to the varying operating conditions by estimating the probability of historical runtime data. The methodology is validated using a feedwater heating system of an on-duty coal-fired power unit. The results indicate that in comparison with the design-based method, the proposed method can provide benchmark intervals for 225 operating conditions. In addition, the determined benchmark interval can detect performance degradation earlier than design-based values, thereby providing opportunities for energy-efficiency enhancement.

1. Introduction

Currently, one of the most urgent issues facing China is saving energy in the power industry [1]. The renewable energy sources have increased their shares of power generation installed capacity recently [2]. Owing to the uncertainties related to renewable energy sources, coal-fired power units need to adjust their output flexibly in order to maintain the real-time power balance of the power grid [3]. However, the flexible operation may degrade the performance and increase the energy consumption of coal-fired power units [4]. The performance of a coal-fired power unit is often evaluated in terms of the coal-consumption rate, which varies with operation boundaries, running time, and component malfunctions [5]. An essential part of detecting performance degradation is determining the benchmark of the variables related to performance (coal-consumption rate).

There are mainly two ways to determine the benchmark of the variables to detect the performance degradation: advanced exergy-based method and data-based method. Since George Tsatsaronis [6,7] proposed the advanced exergy analysis method to detect the performance degradation of the energy conversion process in 1984, the terms themselves and the methodologies of advanced exergy-based analysis have been developed by many researchers in the past twenty years [8]. For instance, Wang et al. [9] employed exergy-based methods to determine the energy-consumption benchmark state with varying operation boundaries. Fu et al. [10] used the advanced exergy-based method to locate the components having degraded performance effectively. Recently, Wang et al. [11] generalized an effective advanced exergy analysis-based method to accurately locate the malfunction component and quantify the effects of anomalies due to multiple malfunctions. However, in reality, the exergy-based calculation and detection can be difficult because the mathematical formula is based on physics that is not always well understood and the benchmark calculations utilize the designed parameters provided by the manufacturer without considering the benchmark variation. Since the performance of the coal-fired power unit is dynamically time-dependent, especially due to the flexible operation and varying environment [12], the

* Corresponding author.;
** Corresponding author.
E-mail addresses: xujing@tyut.edu.cn (J. Xu), bidp@sxicc.an.cn (D. Bi).

https://doi.org/10.1016/j.energy.2020.118555
0360-5442/

benchmark of a variable can deviate from the designed one due to the wide existence of varying operation boundaries, component aging or upgrading [13].

Thanks to the development of information and communication technology, a large amount of sensor data on the actual status of the coal-fired power unit is available. Massive amounts of operational data have enabled the development of the data-based approach to detect performance degradation [14]. Recently, several data-mining methods have been applied to optimize the performance of coal-fired power units, and these methods have made satisfactory results [15]. For example, Xu et al. [1] proposed a data-based methodology to provide the reference values of independent variables for the operator to minimize the heat consumption rate with respect to the variations in load and ambient temperature. This was successfully implemented as an online optimization system for an on-duty steam-turbine system. Zagorowska et al. [16] proposed a data-driven algorithm, which combines the moving window approach with the adaptive regression analysis of operating data, to predict the expected value of the performance degradation indicator. Wang et al. [17] proposed the operation optimization benchmark based on dynamic data-mining technology for a direct air-cooled combined heat and power plants. Liu et al. [18] utilized big data mining techniques based on the Hadoop platform, including the fuzzy rough set attribute reduction method and the Canopy-*K*-means algorithm, to excavate the reference values of controllable operating parameters of the net coal consumption under typical load conditions. The studies mentioned above set the benchmark as a fixed value to optimize the performance of the unit. However, when taking account into the performance degradation, the benchmark should be an interval that varies with the operating conditions rather than a fixed value in order to avoid false or missing alarms. The benchmark interval refers to the upper and lower thresholds of the variables when the performance is normal.

This paper proposes a methodology based on the data mining method to determine the benchmark interval of variables related to the performance of a coal-fired power unit, in order to detect its performance degradation. The benchmark interval is determined using the Gaussian Mixture Model (GMM), which is a widely used data-based method, by estimating the historical runtime data with respect to varying operating conditions (load and ambient temperature).

The main contributions of this work are as follows: (1) proposing a data-based methodology to detect performance degradation by integrating both data mining and data estimation techniques; this method can detect performance degradation earlier than the designed-based method; (2) determining the benchmark interval of variables instead of a fixed value considering their normal fluctuations; (3) providing the benchmark interval for a series of operating conditions rather than a few typical ones, which is useful for the flexible operation of power units.

This paper is organized as follows: The proposed novel methodology for performance degradation detection is described in Section 2. A feedwater heating system of a coal-fired power unit is considered for the case study, which is described in Section 3. Results and discussions are provided in Section 4. Some conclusions are drawn in Section 5.

2. Proposed methodology

2.1. Description of the proposed methodology

The basic idea of the proposed methodology is to determine the benchmark interval for detecting the performance degradation of a coal-fired power unit with respect to the varying operating conditions. The flowsheet of the proposed approach is illustrated in Fig. 1.

The proposed methodology is hierarchical, consisting of the following four successive steps:

(1) Steady-state detection and operating conditions classification. The flexible operation of a coal-fired power unit causes the variables to fluctuate. Therefore, the steady-state detection is a prerequisite. In addition, variables are also subjected to both load changes and ambient temperature variations [19]. Therefore, K-means clustering is employed to categorize the operating conditions according to the similarity of historical data.
(2) Key performance indexes (KPIs) selection. The coal-fired power unit process involves hundreds of interconnected variables and infrastructures. Some of these variables are directly related to the performance of the unit, while the others have minimal influence [20]. Selecting the KPIs is vital for performing the calculations.
(3) Benchmark interval determination of KPIs. GMM is employed to estimate the probability density distributions of the KPIs based on the stored historical runtime data. The benchmark interval is determined by setting the confidence level at 95%.
(4) Performance degradation detection. The sliding window technique is taken to detect whether the mean value of the window data is outside the benchmark interval, considering the uncertainty. When it is outside the interval, the performance can be identified as degradation.

Every step is described in detail in the following sections, sequentially.

2.2. Steady-state detection and operating conditions classification

2.2.1. Steady-state detection

The output of a coal-fired power unit varies with the requirements of the power gird [21]. The variables are unstable because of the transition operating conditions. However, since the steady-state is one of the most important and common assumptions [22], it is necessary to detect whether the process is steady. Table 1 lists the five key variables that are identified based on the rules of ASME PTC6 [23]. The process has been described in detail in the literature [1,20].

2.2.2. K-means clustering based operating conditions classification

The criterion of the operating conditions classification depends on the actual operational scenarios. Therefore, K-means clustering, a data-mining method, is adopted to categorize the operating conditions according to the similarity of the historical operational data.

K-means clustering method, proposed by MacQueen in 1967 [24], is one of the most widely used clustering methods. Given a data set $\boldsymbol{X} = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \in \boldsymbol{R}^p$, where n is the number of samples and p is the variable dimension, K-means clustering starts by selecting the initial cluster center randomly. It keeps updating the centroid of each cluster until it finds the minimum sum-of-squares based on the coordinate descend [25]. That is,

$$E = \sum_{i=1}^{k} \sum_{x_j \in S_j} \|x_j - m_i\|^2 \tag{1}$$

where E is the sum-of-squares of the data, k is the number of clusters, x_j is a sample of the i th cluster, and m_i is the clustering center of the i th cluster.

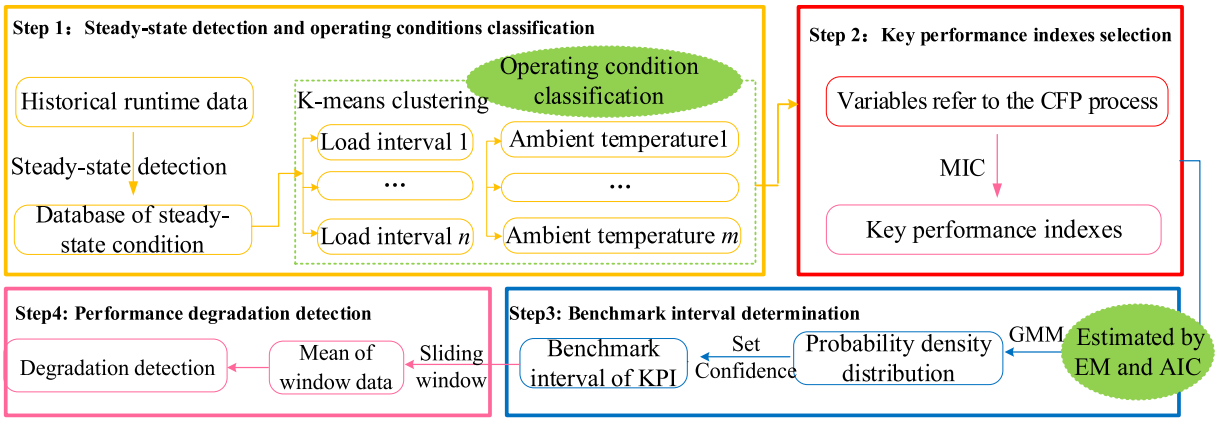

Fig. 1. Flowsheet of the proposed methodology.

Table 1
Steady-state detection criterion.

Parameter	Code	Unit	Threshold
Load	L	MW	30
Main steam pressure	P_1	MPa	1
Main steam temperature	T_1	°C	10
Reheat steam temperature	T_2	°C	10
Feed-water flow	G_{fw}	t/h	60

2.3. Key performance indexes selection by the maximum information coefficient

The process of the coal-fired power unit involves a great number of variables including independent variables, boundary parameters, and dependent variables. While some of them are highly related to the coal-consumption rate, the others are not [26]. The maximum information coefficient (MIC), proposed by Reshef [27] in 2011, is employed to quantify the association between the variable and the coal-consumption rate without any bias. Variables with larger MIC are selected as the KPIs to reduce the complexity of the model.

Given a dataset $\boldsymbol{S}(x, y)$ of two variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, we can partition the x-values and y-values of $\boldsymbol{S}$ into i and j bins, respectively. A pair of the partition is called an x-by-y grid $\boldsymbol{M}$ [28]. *MIC* of the dataset $\boldsymbol{S}$ is given by

$$MIC = \max_{i,j} \frac{\max(M(i,j))}{\log(\min M(X,Y))} \tag{2}$$

where max $\boldsymbol{M}(i,j)$ and min $\boldsymbol{M}(X,Y)$ are the maximum and minimum mutual information over grid $\boldsymbol{M}$ with i columns and j rows, respectively. The larger *MIC* is, the more powerful the $\boldsymbol{X}$ associated with $\boldsymbol{Y}$ would be.

2.4. Benchmark interval determination using GMM

The variables fluctuate dynamically during daily operation. One of the most important issues related to the performance degradation detection is distinguishing between normal fluctuation intervals (benchmark intervals) and degradation. The runtime data of the KPIs is collected and stored in a database, which can reflect the daily operating characteristics and provide the operational data for analysis. GMM, a data-based method, is adopted to estimate the distribution of the KPIs. Then, the benchmark interval can be determined by setting a certain confidence level.

2.4.1. Gaussian Mixture Model

GMM, a semi-parametric density distribution estimation method, has been widely used for the parameter estimation [29]. It is not subjected to the probability density function, which makes it suitable for engineering applications. The probability density of the GMM is described as follows.

$$p\left(X|\theta\right) = \sum_{k=1}^{K} \omega_k \varphi_k(X|\theta_k) \tag{3}$$

where $X = [x_1, x_2, \ldots, x_L]^T$ is a vector with L-dimensional columns, K is the number of sub-models, ω_k is the weight coefficient, $\omega_{k \geq 0}$, $\sum_{k=1}^{K} \omega_k = 1$, $\varphi_k(X|\theta_k)$ represents the Gaussian probability density function of the k th sub-model. The probability density of the k th sub-model is shown in Eq. (4).

$$\varphi_k\left(X|\theta_k\right) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}\left(X-\mu\right)\right) \tag{4}$$

where μ represents the mean of the density function, and $\sum$ is the covariance matrix of the density function. Both μ and $\sum$ need to be estimated.

It is worth noting that the accuracy of GMM depends on the number of sub-models K [30]. The larger K is, the more accurate the model would be. However, the complexity of the model increases with K. Therefore, Akaike Information Criterion (AIC), proposed by Akachi Hiroshi in 1987 [31], is used to determine the optimal number of sub-models, and is given by

$$\mathrm{AIC} = 2\mathrm{K} - 2\ln(\mathrm{L}) \tag{5}$$

where K represents the number of sub-models, and $\ln(L)$ is the log likelihood function of the model.

When AIC is minimum, the trade-off between the accuracy and complexity of the model is balanced. The corresponding K is the optimal number of GMM sub-models.

2.4.2. EM -based estimation of μ and $\sum$

The expectation maximization (EM) algorithm, a type of the maximum likelihood estimation method, was proposed by Dempster and his co-workers in 1996 [32]. The objective function can be described as follows:

$$\min L(\theta) = \log\left(\Pi_{i=1}^{N} p(X|\theta)\right) = \sum_{i=1}^{N} \log p(X|\theta) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} \omega_k \varphi_k(X|\theta_k) \quad (6)$$

The EM algorithm consists of the following three steps [30]:

Step 1(Selection): Initialize parameter u and $\sum$.
Step 2 (E-step): Calculate the response of the kth sub-model:

$$\widehat{\gamma}_{ik} = \frac{\omega_k \varphi(X_i|\theta_k)}{\sum_{k=1}^{K} \omega_k \varphi(X_i|\theta_k)}, i = 1,2,\ldots N; k = 1,2,\ldots K \quad (7)$$

Step 3 (M-step): Iterate u and $\sum$ according to equation (8).

$$\widehat{\mu}_k = \frac{\sum_{i=1}^{N} \widehat{\gamma}_{ik} X_i}{\sum_{i=1}^{N} \widehat{\gamma}_{ik}}, k = 1,2,\ldots,K$$

$$\widehat{\Sigma}_k^2 = \frac{\sum_{i=1}^{N} \widehat{\gamma}_{ik}(X_i - \widehat{\mu}_k)^{\mathrm{T}}(X_i - \widehat{\mu}_k)}{\sum_{i=1}^{N} \widehat{\gamma}_{ik}}, k = 1,2,\ldots,K \quad (8)$$

$$\widehat{\omega}_k = \frac{1}{N} \sum_{i=1}^{N} \widehat{\gamma}_{ik}, k = 1,2,\ldots,K$$

Terminate the iteration (from Step 1 to Step 3) when the following termination criterion is met.

$$|L(\theta) - \widehat{L}(\theta)| x < \varepsilon \quad (9)$$

where ε is usually taken as 10^{-5}.

2.4.3. Benchmark interval determination

After estimating the variable distribution $\widehat{f}(x)$, we can set a certain confidence level α as follows:

$$\int_{x_l}^{x_u} \widehat{f}(x)\, \mathrm{d}x = \alpha \quad (10)$$

where x_u and x_l represent the upper and lower thresholds of the variable x at the confidence level α, respectively. The benchmark interval of variable x is $[x_l, x_u]$.

2.5. Sliding window-based degradation detection

Considering the uncertainty related to the random interference of sensors, we check whether the mean value of the variable in a fixed window n is outside the determined benchmark interval. If it is within the interval, the performance can be identified as normal and vice versa. The sliding window moves forward to the next sample data to iterate the check as shown in Fig. 2. The window length n and step m are taken as 5 and 1, respectively.

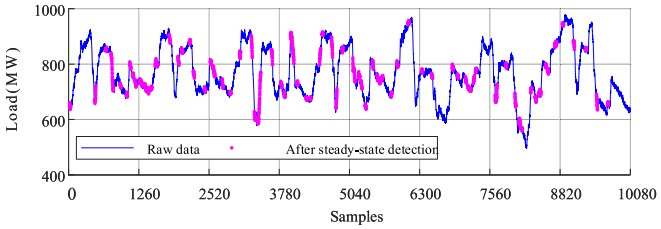

Fig. 2. Sliding window-based performance degradation detection.

3. Case study

3.1. Case feedwater heating system and its working process

In this paper, a 1000-MW ultra-supercritical coal-fired power unit was considered for the case study. The steam-turbine is of N1000−26.25/600/600 type, manufactured by Siemens and Shanghai Steam-Turbine Co., Ltd. It is a single reheat unit with four cylinders, eight steam extractors, and two open-loop condensers. Its feedwater heating system (FHS), which is a typical interconnected sub-system of the coal-fired power system, is taken as the case study.

Fig. 3 shows a schematic diagram of a regenerative feed heating system (RFHS). FHS is a part of the RFHS. It mainly consists of three high-pressure heaters (HPHs), pipes and valves. Water leaving the deaerator goes to the feed water pump suction, and is then pumped into the next HPH. Physically, a proportion of the steam is bled from the turbine, and then condensed in the HPH to heat the feed water on its return to the boiler, sequentially [33,34].

3.2. Steady-state detection and operating condition classification

3.2.1. Data description

The studied case unit is configured with a PI (Plant Information) database, which records the operational data over many years. We sampled the variables historical runtime data of March, August, and December 2018 from the PI to ensure that the data are representative of a range of operating scenarios. The historical data were sampled over periods of 1 min, with 1440 samples per day and a total of 133,920 samples for the three months.

3.2.2. Steady-state detection and operating condition classification of feedwater heating system

The variation interval of the load and ambient temperature are [500−1010] MW and [9−34]°C, respectively. A total of 27,140 data samples remained after steady-state detection. The steady-state detection data of samples over a one-week period is shown in Fig. 4.

As discussed in Section 2.2.2, the historical data are categorized into 225 operating conditions according to the variables' similarity of historical operational data based on K-means clustering. Here, we take one cluster as an example and discuss it in detail. Its output and ambient temperature ranges are 500−510 MW and 29−34 °C, respectively. The operating condition is labelled as A. The others can be analyzed in the same way. They are not described in detail owing to space constraints.

3.3. KPIs selection of the feedwater heating system

Out of the 123 variables, 28 variables related to the performance of FHS were selected based on MIC. They are listed in Table 2. TTD is short for terminal temperature difference.

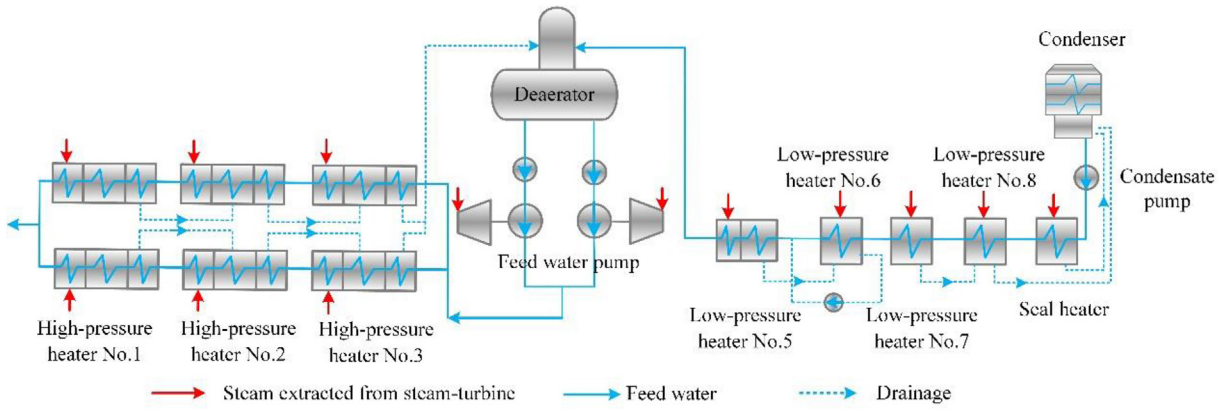

Fig. 3. Schematic diagram of the RFHS of a coal-fired power plant.

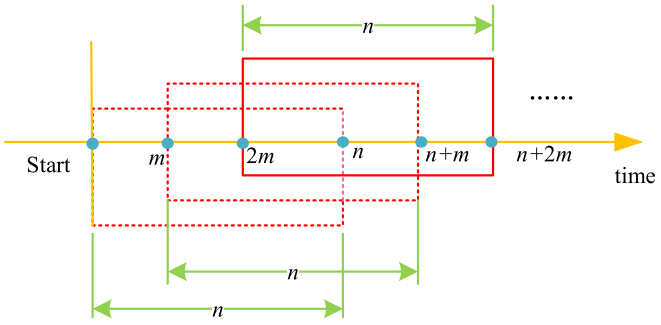

Fig. 4. Steady-state detection data of samples over a one-week period.

3.4. Probability density distributions estimation of KPIs

Fig. 5 illustrates the variation of AIC with the number of sub-models K, which ranges from 1 to 30, for Δt_{S2}. It is shown that when K is 5, the AIC is minimum. Therefore, the optimal number of GMM sub-models of Δt_{S2} is 5. The others can be analyzed in the same way, however, they are not listed herein. The probability density distributions of the KPIs estimated by GMM of operating condition A are shown in Fig. 6 It can be seen from Fig. 6 that KPI fluctuates within a certain range under the same operating condition. The confidence level is set as 95%, that is, the false positive rate is 5%. The benchmark intervals of operating condition A is determined as shown in Table 3.

4. Results and discussion

4.1. Benchmark comparison between the data-based and design-based method

Table 3 compares the benchmarks of 28 KPIs determined by the design-based method and the data-based method in terms of the

Table 2
Codes of KPIs.

KPI	Code	KPI	Code	KPI	Code
Extraction pressure of No.1 HPH	p_{cq1}	Extraction pressure of No.2 HPH	p_{cq2}	Extraction pressure of No.3 HPH	p_{cq3}
Inlet pressure of No.1 HPH	p_{jq1}	Inlet pressure No.2 of HPH	p_{jq2}	Inlet pressure of No.3 HPH	p_{jq3}
Extraction pressure difference of No.1 HPH	Δp_1	Extraction pressure difference of No.2 HPH	Δp_2	Extraction pressure difference of No.3 HPH	Δp_3
Inlet water temperature of No.1 HPH	t_{c1}	Inlet water temperature of No.2 HPH	t_{c2}	Inlet water temperature of No.3 HPH	t_{c3}
Feedwater TTD of No.1 HPH	Δt_{S1}	Feedwater TTD of No.2 HPH	Δt_{S2}	Feedwater TTD of No.3 HPH	Δt_{S3}
Drainage temperature of No.1 HPH	t_{s1}	Drainage temperature of No.2 HPH	t_{s2}	Drainage temperature of No.3 HPH	t_{s3}
Drainage TTD of No.1 HPH	Δt_{x1}	Drainage TTD of No.2 HPH	Δt_{x2}	Drainage TTD of No.3 HPH	Δt_{x3}
Water level of No.1 HPH	L_{j1}	Water level of No.2 HPH	L_{j2}	Water level of No.3 HPH	L_{j3}
Outlet temperature of deaerator	t_{c4}	Extraction pressure of deaerator	p_{cq4}	Water level of deaerator	L_{j4}
Inlet pressure of deaerator	p_{jq4}				

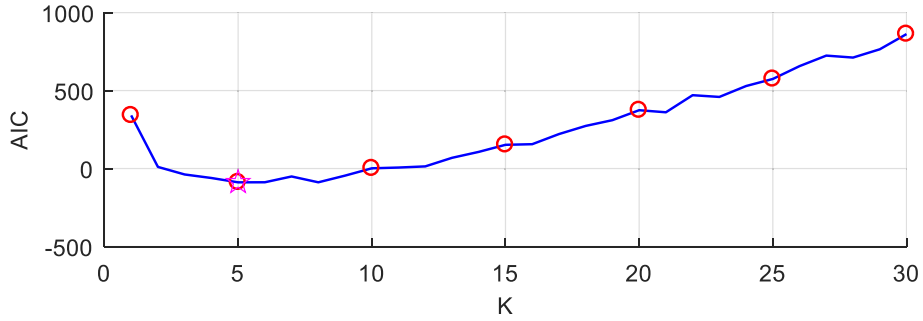

Fig. 5. Variation in AIC with the number of the GMM sub-models K for Δt_{S2}.

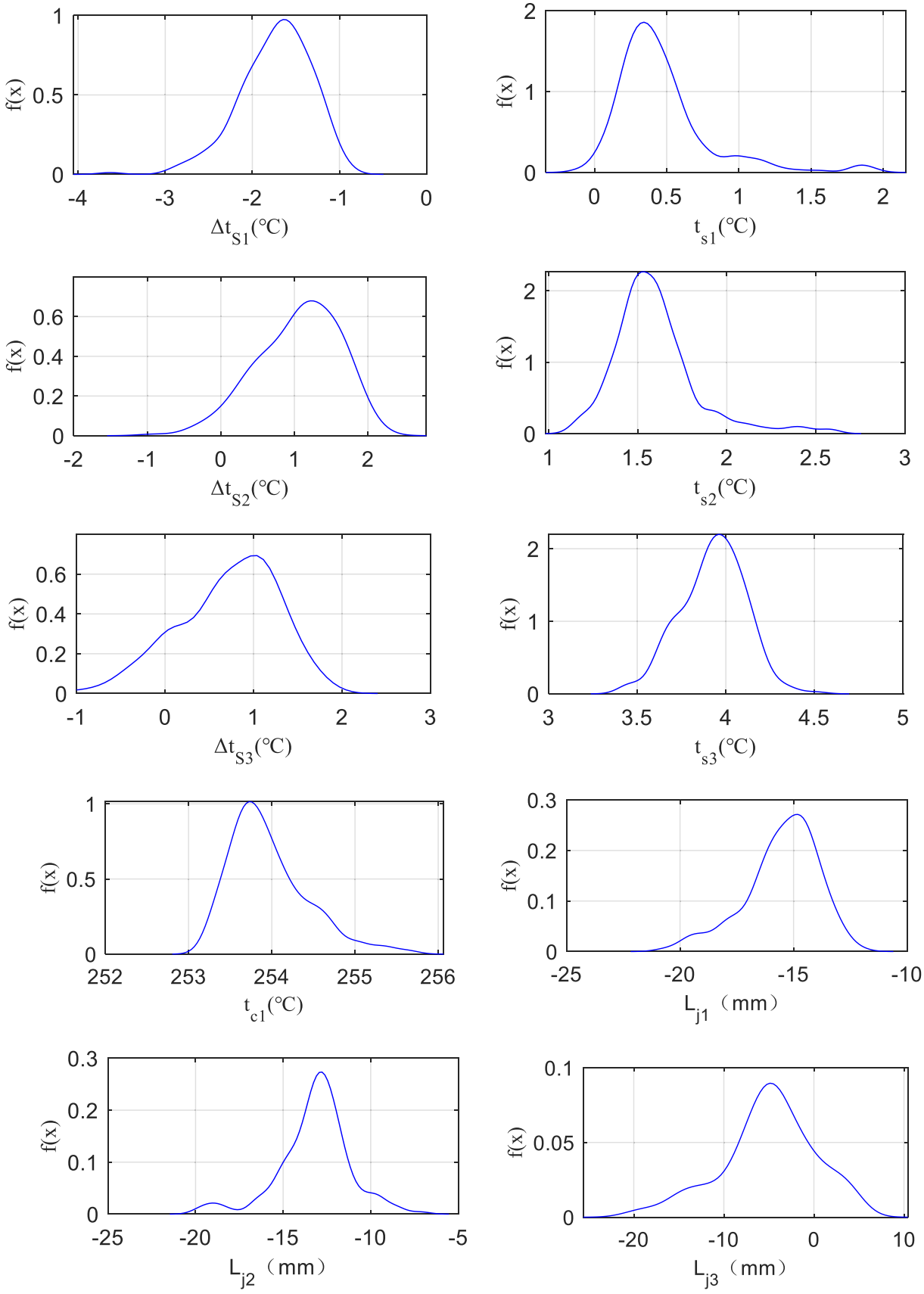

Fig. 6. Probability density distributions of the key performance nodes based on GMM.

Table 3
Benchmark intervals of KPIs.

KPI	Designed	X^+	X^-	Unit	KPI	Designed	X^+	X^-	Unit
p_{cq1}	3.81	4.25	4.04	MPa	p_{cq2}	2.83	3.07	2.94	MPa
p_{jq1}	3.69	4.14	4.00	MPa	p_{jq2}	2.74	3.00	2.86	MPa
Δp_1	0.114	0.079	0.028	MPa	$\boldsymbol{\Delta p_2}$	**0.085**	**0.089**	**0.042**	**MPa**
t_{c1}	247.4	255.2	253.6	°C	t_{c2}	228.9	234.2	232.6	°C
$\boldsymbol{\Delta t_{S1}}$	**−1.7**	**−0.82**	**−2.39**	**°C**	Δt_{S2}	0	2.2	0.3	°C
t_{s1}	234.5	235.1	233.1	°C	t_{s2}	193.2	192.9	191.3	°C
Δt_{x1}	5.6	1.0	0.3	°C	Δt_{x2}	5.6	2.1	1.4	°C
L_{j1}	0	−13.28	−19.28	mm	L_{j2}	0	−9.64	−19.01	mm
p_{cq3}	1.22	1.23	1.10	MPa	p_{jq3}	1.19	1.18	1.06	MPa
$\boldsymbol{\Delta p_3}$	**0.037**	**0.047**	**0.0074**	**MPa**	t_{c3}	187.6	191.6	189.6	°C
Δt_{S3}	0	1.3	−0.8	°C	t_{s3}	167.7	166.9	163.3	°C
Δt_{x3}	5.7	4.3	3.6	°C	$\boldsymbol{L_{j3}}$	**0**	**3.34**	**−15.52**	**mm**
t_{c4}	159.4	163.2	159.4	°C	$\boldsymbol{p_{cq4}}$	**0.65**	**0.72**	**0.54**	**MPa**
$\boldsymbol{p_{jq4}}$	**0.62**	**0.66**	**0.58**	**MPa**	L_{j4}	0	−16.61	−61.54	mm

operating condition A. It is clear that the benchmark determined by GMM is an interval instead of a fixed value, considering the normal operating fluctuation of the KPI. Of the 28 KPIs, only six of the designed values (marked in bold) are within the benchmark interval. It is because after years of operation, the variables deviate from the designed value due to the wide existence of components aging or renewal, and technology upgrading. It can be seen that the data-based method can reflect the actual operating characteristics when determining the benchmark interval.

Compared with the design-based method, the proposed method can provide KPIs benchmark intervals for 225 operating conditions (load and ambient temperature ranging from 500 MW to 1000 MW and 9 °C–34 °C, respectively), rather than only a few typical operating conditions, such as THA, 75% THA, and 50% THA.

Results (not shown herein due to the space constrains) also show that the benchmark intervals of the KPIs vary with the operating conditions, that is, the load and ambient temperature affect the KPIs. Therefore, it is necessary to account for the load and ambient temperature when evaluating the performance of the system. Considering the flexible operation of the unit, the benchmark intervals of 225 operating conditions proposed in this paper can provide accurate anomaly information under various operating conditions, which makes it more suitable for engineering applications than the design-based method.

4.2. Industrial application

An industrial case in point is discussed here. On July 15, 2018, from 1:00 a.m. to 2:20 a.m., the output and ambient temperature of the studied coal-fired power unit were 500 MW and 32 °C, respectively. It was operating under operating condition A.

Fig. 7 and Fig. 8 illustrate the variations of L_{j2} and coal-consumption rate, respectively, from 1:00 a.m. and 2:30 a.m. It can be seen from Fig. 6 that from 1:00 a.m. to 2:10 a.m., L_{j2} fluctuated within the benchmark interval [-9.64 mm, −19.01 mm] listed in Table 3. However, at 2:10 a.m., L_{j2} reached −8.95 mm, which is outside the benchmark interval during sliding window detection. After that, L_{j2} increased dramatically to approximately 20 mm.

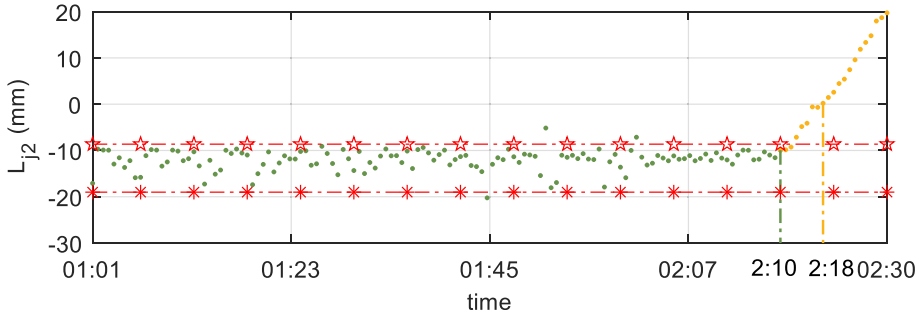

Fig. 7. L_{j2} varies with time.

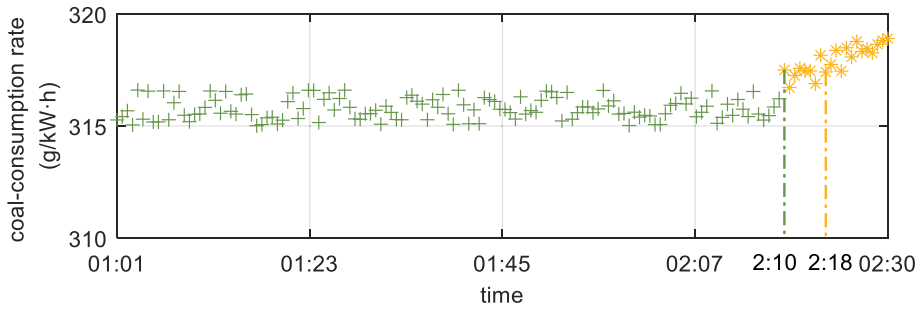

Fig. 8. Variation in Coal-consumption rate with time.

Similarly, the corresponding coal-consumption rate increased from 315.8 g/(kW·h) to 318.9 g/(kW·h), as shown in Fig. 8.

However, the design-based benchmark of L_{j2} is 0 mm, as shown in Table 3. It can be seen from Fig. 7 that L_{j2} did not exceed 0 mm until 2:18 a.m., which indicates that the performance degradation started at 2:18 a.m. Nevertheless, the coal-consumption rate showed a growth trend from 2:10 a.m. as shown in Fig. 8, implicating that the performance begun to deteriorate at 2:10 a.m.

It can be seen that the proposed benchmark interval can detect performance degradation 8 min earlier than the design-based value. It can be estimated that the increase in energy consumption reached up to 23793.2 kg during the 8 min of performance degradation, which is remarkable.

When No. 2 HPH, one of the surface heaters of FHS, was opened, it was found that there was a hole (with a diameter of 20 mm) on one of the tubes. The feed water leaked from the tube to the shell, which increased the water level of No.2 HPH L_{j2}. It can be concluded that the performance degradation was caused by the component malfunction. It also can be concluded that data-based benchmark interval is useful for detecting performance degradation.

5. Conclusions

Due to the increased shares of renewable energy power resources installation, coal-fired power units have been in the flexible operation manner. However, flexible operation deteriorates the performance of the units, which incur energy-efficiency penalty and energy consumption increment. In this paper, a data-based methodology was presented to detect the performance degradation of coal-fired power units. The essential part of the detection is determining the benchmark intervals of KPIs with respect to the varying operating conditions. The proposed methodology utilizes a large amount of the historical runtime data, and it was applied to an on-duty FHS of a coal-fired power unit.

(1) Considering that the benchmark interval varies with the operating conditions, the categorization of operating conditions is a prerequisite for the benchmark determination. K-means clustering method works well in operating conditions classification, according to the similarity of the variables with the historical operational data.
(2) The variables that have the greatest influence on the performance were analyzed using the maximum information coefficient. 28 variables of the FHS having high correlation with the coal-consumption rate were selected as the KPIs.
(3) GMM works well in estimating the probability distributions of KPIs. The benchmark interval was determined by setting the confidence level at 95%. The benchmark interval determined by the proposed methodology could detect performance degradation earlier than the design-based value with respect to varying operating conditions, which is helpful in saving energy.

One limitation of the proposed methodology is that it requires massive amounts of historical operational data. If there were insufficient data for clustering, the results would not have been satisfactory. It would be worthwhile to consider the influence of the quality of the coal on the performance of the unit in the future study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The author Jing Xu is supported by the National Natural Science Foundation of China (grant number 51906171); Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (grant number 2019L0308); Science and Technology Major Project in Shanxi (grant number 20201101013) and the International Clean Energy Talent Program funded by China Scholarship Council (grant number 201904100053). The authors Dapeng Bi and Jin Bai are supported by the National Natural Science Foundation of China (grant number 21761132032).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2020.118555.

References

[1] Xu J, Gu Y, Ma S. Data based online operational performance optimization with varying work conditions for steam-turbine system. Appl Therm Eng 2019;151:344–53.
[2] National Energy Administrator. National electric power industry statistics. Site accessed. http://www.nea.gov.cn/2019-01/18/c_137754977.htm; January, 2019.
[3] Wang J, You Sh, Zong Y, Trætholt Ch, Dong Y, Zhou Y. Flexibility of combined heat and power plants: a review of technologies and operation strategies. Appl Energy 2019;252:113445.
[4] Dong Y, Jiang X, Liang Zh, Yuan J. Coal power flexibility, energy efficiency and pollutant emissions implications in China: a plant-level analysis based on case units. Resour Conserv Recycl 2018;134:184–95.
[5] Gu Y, Xu J, Chen D, Wang Zh, Li Q. Overall review of peak shaving for coal-fired power units in China. Renew Sustain Energy Rev 2016;54:723–31.
[6] Tsatsaronis G. Strengths and limitations of exergy analysis. In: Bejan A, Mamut E, editors. Thermodynamic optimization of complex energy systems. Kluwer Academic Publishers; 1999. p. 93–100.
[7] Tsatsaronis G. Combination of exergetic and economic analysis in energy conversion processes. In: Energy economics and management in industry, proceedings of the European Congress, Algarve, Portugal, April 2-5, vol. 1. England, Oxford: Pergamon Press; 1984. p. 151–7.
[8] Morosuk T, Tsatsaronis G. Advanced exergy-based methods used to understand and improve energy-conversion systems. Energy 2019;169:238–46.
[9] Wang N, Fu P, Xu H, Wu D, Yang Zh, Yang Y. Heat transfer characteristics and energy-consumption benchmark condition with varying operation boundaries for coal-fired power units: an exergy analytics approach. Appl Therm Eng 2015;88:433–43.
[10] Fu P, Wang N, Wang L, Morosuk T, Yang Y, Tsatsaronis G. Performance degradation diagnosis of thermal power plants: a method based on advanced exergy analysis. Energy Convers Manag 2016;130:219–29.
[11] Wang L, Fu P, Wang N, Morosuk T, Yang Y, Tsatsaronis G. Malfunction diagnosis of thermal power plants based on advanced exergy analysis: the case with multiple malfunctions occurring simultaneously. Energy Convers Manag 2017;148:1453–67.
[12] Xiong J, Zhao H, Zhang C, Zheng Ch, Peter BL. Thermoeconomic operation optimization of a coal-fired power plan. Energy 2012;42(1):486–96.
[13] Qiao Z, Wang X, Gu H, Tang Y, Sia F, Romerob C, et al. An investigation on data mining and operating optimization for wet flue gas desulfurization systems. Fuel 2019;258:116–78.
[14] Fan C, Xiao F, Zhao Y, Wang J. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. Appl Energy 2018;211:1123–35.
[15] Fan C, Sun Y, Shan K, Xiao F, Wang J. Discovering gradual patterns in building operations for improving building energy efficiency. Appl Energy 2018;224:116–23.
[16] Marta Z, Frederik S, Arne-Marius D, Lars I, Erling L, Nina F. Adaptive detection and prediction of performance degradation in off-shore turbomachinery. Appl Energy 2020;268:114934.
[17] Wang N, Zhang Y, Fu P, Feng P, Yang Y. Heat transfer and thermal characteristics analysis of direct air-cooled combined heat and power plants under off-design conditions. Appl Therm Eng 2018;129:260–8.
[18] Liu B, Fu Zh, Wang P, Wang Y, Gao X. Big data mining technology application in energy consumption analysis of coal-fired power plant units. Proc. CSEE 2018;38(12):3578–87.
[19] Chen C, Zhou Z, Bollas G. Dynamic modeling, simulation and optimization of a subcritical steam power plant. Part I: plant model and regulatory control. Energy Convers Manag 2017;145:324–34.
[20] Xu J, Gu Y, Chen D, Li Q. Data mining based plant-level load dispatching strategy for the coal-fired power plant coal-saving: a case study. Appl Therm Eng 2017;119:553–9.

[21] Blanco J, Vazquez L, Peña F, Diaz D. New investigation on diagnosing steam production systems from multivariate time series applied to thermal power plants. Appl Energy 2013;101:589–99.
[22] Blanco J, Vazquez L, Peña F. Investigation on a new methodology for thermal power plant assessment through live diagnosis monitoring of selected process parameters; application to a case study. Energy 2012;42(1):170–80.
[23] ASME. Steam-turbines, performance test Codes. 2004. ASME PTC 6-2004.
[24] Hartigan J, Wong M. A K-means clustering algorithm. Appl. Stat. 1979;28:100–8.
[25] Xu J, Gu Y, Wang Zh, Li Q, Yang N. Research on indexes of energy efficiency and its reference-value for coal-fired power units based on data mining. Proc. CSEE 2017;37(7):2009–15.
[26] Tunckaya Y, Koklukaya E. Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural-based models. J Energy Inst 2014;1:11–8.
[27] Reshef D, Reshef Y, Finucane H, Grossman S, McVean G, Turnbaugh P, et al. Detecting novel associations in large data sets. Science 2011;334(6062):1518–24.
[28] Justin B, Gurinder S. Equitability, mutual information, and the maximal information coefficient. Proc. Natl. Acad. 2014;111:3354–9.
[29] Sang W, Jin H, Lee I. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. Comput Chem Eng 2004;28:1377–87.
[30] Wang Z, Gu Y. A steady-state detection method based on Gaussian discriminant analysis for the on-line gas turbine process. Appl Therm Eng 2018;133:1–7.
[31] Akaike H. Factor analysis and AIC. Psychometrika 1987;52(3):317–32.
[32] Tibshirani T. Discriminant analysis by Gaussian mixtures. J. Royal Stat. Soc.Series B (Methodological) 1996;58:155–76.
[33] Wang Y, Cao L, Hu P, Li B, Li Y. Model establishment and performance evaluation of a modified regenerative system for a 660 MW supercritical unit running at the IPT-setting mode. Energy 2019;179:890–915.
[34] Li Y, Wang Y, Cao L, Hu P, Han W. Modeling for the performance evaluation of 600 MW supercritical unit operating No.0 high pressure heater. Energy 2018;149:639–61.