Applied Thermal Engineering 151 (2019) 344–353

Contents lists available at ScienceDirect

Applied Thermal Engineering

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

Research Paper

Data based online operational performance optimization with varying work conditions for steam-turbine system

Jing Xu[a,*], Yujiong Gu[b], Suxia Ma[a]

[a] College of Electrical and Power Engineering, Taiyuan University of Technology, No. 79 of Yingzexi Street, Taiyuan, Shanxi 030024, PR China
[b] National Thermal Power Engineering & Technology Research Center, North China Electric Power University, Beinong Street, Beijing 102206, PR China

ARTICLE INFO

Keywords:
Steam-turbine system
Performance optimization
Heat consumption rate
Fuzzy C-means
Kernel density estimation

ABSTRACT

One of the most urgent problems facing China today is saving energy in the power industry. Moreover, because the steam-turbine system is a main part of the power unit, its operation is of great significance for energy saving. This paper presents an optimization model for steam-turbine systems based on a data-mining method from an operator's perspective. The main aim of the proposed methodology is to provide reference values of the independent variables for the operator to minimize heat consumption rate. These were determined based on fuzzy C-means clustering and statistical methods by mining the historical operation data resources with respect to varying load and ambient temperature. The proposed methodology application was implemented as an online optimization system for an on-duty steam-turbine system. The application results showed that the energy savings reached up to 79,000 GJ, which is remarkable. The reference values of variables were helpful for improving the steam-turbine system's performance.

1. Introduction

Fog and haze across the country has alarmed China's energy utilization and environmental protection since 2013 [1]. As the largest source of carbon emissions in China, the electric power industry consumes approximately 50% of total coal, which constitutes the vast majority of the pollution [2]. The Chinese government has recently made significant endeavors into energy conservation and emission reduction. One of the treatments is replacing coal-fired power plants with renewable energy resources. Unfortunately, the outputs of renewable power generation are random, unstable, and difficult to predicted [3]. Therefore, to maintain the power grid balance, the coal-fired power units must experience load transition and partial outputs situations most of the time [4]. However, when the load decreases, the units' thermal efficiency of the cycle decreases; hence, coal consumption rises eventually [3]. The ever-increasing rate of coal depletion and the severity of environmental damage underline the necessity of higher efficiencies of coal utilization [5].

The energy unitization efficiency is determined by a group of operational variables for an on-duty power unit. Thus, the reference value of operational variables is of great significance for unit operation optimization. The reference value is defined herein as 'the expected operational value of variables that can improve energy efficiency'. The reference values can be broadly categorized as design values, experiment-based values and model-based values [6]. Generally, design values only refer to some typical work conditions, such as THA, 75%THA, and 50%THA. The experiment-based values have high cost when performing experiments. Further, modelling of the unit processes (based on thermodynamic theory) is difficult because of its complexity and nonlinearity [7]. Some assumptions are needed when building the theoretical model, which could sometimes make the model-based reference value unavailable in practice.

Fortunately, sensors installed in the power unit collect and supervise the runtime data of thousands of variables continuously and store them in a database [8]. These data reflect the unit's actual state and operation characteristics. Recently, several data-mining methods have been applied to optimize coal-fired power plant performance and have produced satisfactory results [9]. Wang and Fu [10] proposed the operation optimization benchmark based on dynamic data-mining technology. Niu et al. [11] used support vector regression (SVR) and genetic algorithm (GA) to optimize the turbine optimal initial pressure under off-design operation. Li et al. [12,13] applied the fuzzy association rules mining algorithm to determine a small group of variables' reference values of a fossil-fired power plant. Zhao et al. [14] used the fuzzy C-means (FCM) clustering algorithm to determine reference value of the exhaust-gas oxygen, unburned carbon content, and exhaust-gas temperature in some typical load conditions.

The variables' reference values are dynamically time-dependent,

* Corresponding author.
E-mail address: xujing0810@126.com (J. Xu).

https://doi.org/10.1016/j.applthermaleng.2019.02.032
Received 28 December 2018; Received in revised form 7 February 2019; Accepted 9 February 2019
Available online 10 February 2019
1359-4311/

especially under the off-design work condition and varying operation environments [15]. It has been demonstrated that when the load is decreased, the thermal efficiency of the unit's cycle decreases; hence, the heat consumption rate eventually increases [3]. In addition, the ambient temperature can affect the vacuum of a water-cooling unit. According to the experimental result, a 1-kPa decrease in vacuum can increase the heat consumption rate by 52.8 kJ/(kW·h) [16]. However, most studies focused on the optimization of a small group of variables at some typical load conditions without considering the load variety and ambient temperature. Only few studies have been applied in an on-duty power plant and presented in open literature.

This paper presents a systematic study on the reference value determination of overall independent variables of the steam-turbine system from the operator's perspective. Reference values of independent variables were determined based on FCM clustering and statistical methods by mining the historical operation data resources with respect to varying work condition (load and ambient temperature). In addition, the proposed methodology was applied as an online optimization system for an on-duty unit for the operator to improve the system performance.

The paper is organized as follows: The proposed novel hybrid methodology for operational performance optimization is described in Section 2. Section 3 describes the implementation of the online optimization system. An industrial steam-turbine system was employed to be the case unit and is studied in Section 4. The results and discussion are provided in Section 5. Finally, some conclusions are drawn in Section 6.

2. Proposed methodology

2.1. Description of the proposed methodology

The basic idea of the proposed methodology was to determine the reference values, minimize the heat consumption rate, and improve the system performance with respect to the load and ambient temperature. The flowsheet of the proposed approach is illustrated in Fig. 1.

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

(1) Key performance indexes (KPI) selection. The steam-turbine process consists of hundreds of variables, where some of the variables directly affect the heat consumption rate, and some have minimal influence [17]. Selecting the KPI of the heat consumption rate is a demanding subject.
(2) Steady-state detection and work condition classification. We can obtain a steady-state condition database by steady-state detection from historical runtime data, considering that transient operation of the steam-turbine system is necessary to balance the grid load. The performance of the steam-turbine system is subjected to frequent load changes and ambient temperature variation [18]. Therefore, the work condition was categorized into several types according to the load and ambient temperature.
(3) Decision-making sample selection. Heat consumption rate was set as the decision-making target. The cluster of historical data with the minimum heat consumption rate was selected as the decision-making sample by the FCM clustering method.
(4) Reference value estimation. The probability density distributions of KPI were estimated by the kernel density estimation (KDE) using the decision-making samples. Their expectations were employed as the reference values, considering the accessibility.

Every step is detailed in the following sections, sequentially.

2.2. Key performance indexes selection

Heat consumption rate, one of the most important indicators for the boiler-turbine system, is a comprehensive indicator that can represent the performance of a steam-turbine unit. It refers to the energy required to produce a kilowatt of electrical energy [19]. The lower the heat consumption rate, the better the steam-turbine system performs. It is known that, numerous variables may impact the heat consumption rate, and their relationship is nonlinear [20].

By convention, variables in an on-duty power plant are categorized into three types: independent variables, boundary conditions and dependent variables. Independent variables can be manipulated from the control panel interface directly by operator. Boundary conditions are the load and ambient temperature. Load is determined by the demand of the power grid, and the ambient temperature is the temperature of

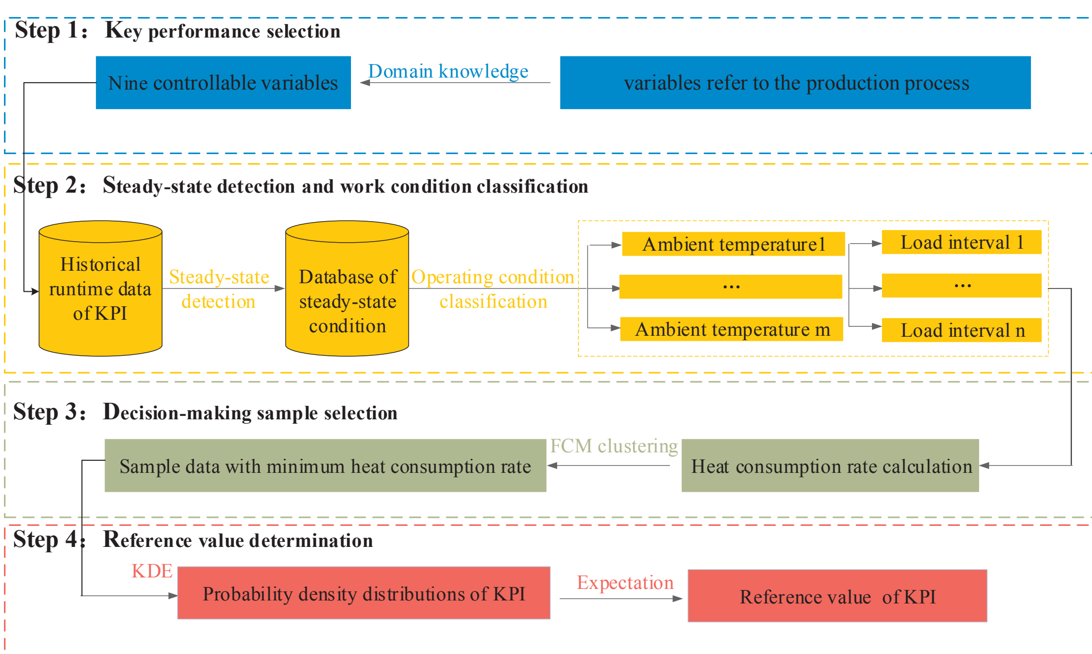

Fig. 1. The flowsheet of the proposed optimization methodology.

Table 1
Information of KPI of the steam-turbine system.

Name	Code	Unit	Design parameter	Sensitivity (%)
Main steam pressure	P_1	MPa	26.25	0.15
Main steam temperature	T_1	°C	600	0.21
Reheat steam temperature	T_2	°C	600	0.08
Condenser vacuum	C	kPa	6.2	0.72
Feed-water temperature	T_{fw}	°C	290	0.15
Feed-water pressure	P_{fw}	MPa	33.0	0.21
Feed-water flow	G_{fw}	t/h	2719.76	0.30
Superheater desuperheating water flow	G_{sd}	t/h	0	0.13
Reheater desuperheating water flow	G_{rd}	t/h	0	0.10
Heat consumption rate	HR	kJ/(kW·h)	7377	1

cooling water, determined by weather. Dependent variables cannot be adjusted directly by the operator, but can be impacted by changing independent variables, such as the reheat steam pressure, and extraction pressure of feed-water heaters.

A 1000-MW ultra-supercritical steam-turbine system was taken as the case subject. The steam-turbine type is N1000-26.25/600/600 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. The turbine design parameters are presented in Table 1. The problem discussed in this paper is how to improve the steam-turbine performance from an operator's perspective. Nine independent variables with great influence on steam-turbine system performance for the case unit were selected as KPI (listed in Table 1) based on domain knowledge.

Sensitivity analysis is employed to confirm that the selected KPI strongly affect the heat consumption. The sensitivity of the variable indicates the heat consumption variation caused by. per-unit variation of the variable, and can be written as follows:

$$\Delta HR|_{\Delta x_i} = HR|_{\Delta x_i} - HR|_{x_i} \tag{1}$$

where x_i denotes the variable; $\Delta HR|_{\Delta x_i}$ is the heat consumption rate variation caused by variation of x_i; $HR|_{\Delta x_i}$ and $HR|_{x_i}$ represent heat consumption rate after and before variation of x_i, respectively.

It also can be expressed in another way as follows,

$$\delta HR|_{\Delta x_i} = \frac{\Delta HR|_{\Delta x_i}}{HR|_{x_i}} \times 100 \tag{2}$$

where $\delta HR|_{\Delta x_i}$ is the relative variation of the heat consumption rate.

The sensitivities of the nine independent variables of the case unit at THA working condition are listed in Table 1. From the table, we can see that the relative variations are all greater than 0.1, which means the per-unit variation of KPI can cause at least 737.7 kJ/(kW·h) variation. It can be concluded that the KPI selection is reasonable.

2.3. Steady-state detection and work condition classification

The working process and the variables of a steam-turbine system are not steady. However, the steady-state is one of the most important and common assumptions [21]. Hence, it is necessary to detect whether the process is steady. Because some of the process-associated variables are transient, the detection needs to be multivariate [22]. Five key variables are identified in Table 2 based on the rules of ASME PTC6 [23].

We check whether the difference between the maximum and minimum of the five variables in a fixed window n is less than a certain threshold (seen in Table 2). If all variables meet the criterion, the window data can be determined as steady-state and vice versa. The sliding window goes backward to the next sample data to check again until the end of the sample data, as shown in Fig. 2. Based on the actual working condition, the window length n was taken as 30.

Table 2
Steady-state criterion.

Parameter Name	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

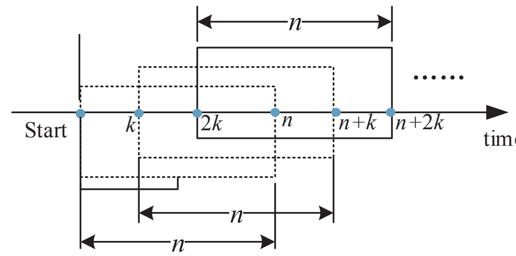

Fig. 2. The steady condition discriminant diagram based on sliding window.

As discussed in Section 2.1, the work condition classification criterion is determined by the actual operation situation and historical operation data. By convention, the KPI of the steam-turbine system are similar when the fluctuation is within 30 MW and 10 °C. Hence, load fluctuations within 30 MW, and ambient temperature fluctuations within 10 °C were categorized as one work condition.

2.4. Decision-making sample selection by FCM clustering

Clustering [24] divides data sample into subsets according to their resemblance. Because of the characteristics of the operation, the training data were partitioned into different subgroups to find the group of KPI data with the minimum heat consumption rate. The FCM clustering algorithm, proposed by Dunn and Bezdek [25], was adopted for its simplicity and efficiency.

Given a data set $X = \{x_1, x_2, ..., x_k, ..., x_n\} \subset \mathbf{R}^p$, where n is the number of sample and p is the variable dimension, FCM partitions the data set $\boldsymbol{X}$ into c clusters by minimizing the errors in terms of the weighted distance of the data to the clusters [26,27]. That is,

$$\min J = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, \|v_i - x_k\|^2 \tag{3}$$

where m is the fuzzifier and generally is taken as 2, v_i denotes the cluster centroid, x_k is the sample datum. $\|\cdot\|$ is the Euclidean distance, and u_{ik} is the membership degree of the kth sample to cluster i. The constraint conditions of u_{ik} are as follows:

$$\begin{cases} \sum_{i=1}^{c} u_{ik} = 1 \\ \sum_{k=1}^{n} u_{ik} > 0 \end{cases} \quad k = 1, \quad 2, \quad ... \quad ,n, \qquad i = 1, \quad 2, \quad ... \quad ,c \tag{4}$$

Combining Eq. (4) and Eq. (3) by the Lagrange multiplier method, we can find the iterative formula of membership degree u_{ik} and cluster centers v_i, as follows:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\| v_i - x_k \|}{\| v_j - x_k \|}\right)^{\frac{2}{m-1}}},\ k = 1,\ 2,\ ...n \tag{5}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{j=1}^{n} u_{ik}^{m}},\ i = 1,\ 2,\ ...c \tag{6}$$

The FCM algorithm solves the classification problem under the condition of satisfying the minimum objective function (Eq. (3)) by iteratively updating Eqs. (5) and (6).

2.5. Reference value estimated by KDE

After clustering, we can obtain the decision-making sample with the minimum heat consumption rate. However, the variables' values are distributed in a wide range in the actual operation. Therefore, it is worthwhile to discuss the reasonable determination of which should be employed as the reference value. Considering the distribution of decision-making sample data, we employ the expectation $E(X)$ as the reference value. $E(X)$ can be described as follows [28]:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{7}$$

As described in formula (7), the probability density function $f(x)$ of the decision-making sample needs to be estimated first. KDE, one of the most widely used estimation methods with a nonparametric model [29], was adopted to estimate $f(x)$. KDE does not depend on the specific form of the probability density function of the observed sample. $f(x)$ can be estimated by the following equation:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{8}$$

where $\hat{f}(x)$ is the estimated probability density value, n is the number of sample data, h is the bandwidth, and $K(\cdot)$ is the kernel function; x_i is the sampling datum of X. It is known that when n is sufficient, $\hat{f}(x)$ converges to $f(x)$.

In fact, it has been proven that the type of $K(\cdot)$ used does not create a significant impact on the accuracy of $\hat{f}(x)$ [30]. Therefore, the Gaussian kernel was employed and its expression is as follows:

$$K(u) = \frac{1}{\sqrt{2x}}\exp(-u^2/2) \tag{9}$$

However, bandwidth h may have a great effect on the estimation accuracy [26]. The minimum squared difference method was used to determine the optimal h in terms of the Gaussian kernel function as follows:

$$h_{\text{opt}} = 1.059\sigma \cdot n^{-\frac{1}{5}} \tag{10}$$

where α and n are the variance and number of the sample data, respectively.

We can obtain $E(X)$ based on the composite integral formula shown in Eq. (11) by assuming that $xf(x) \in C[a, b]$ [31].

$$E(X) = \frac{1}{2}[\frac{b-a}{n}\sum_{k=0}^{n-1} x_k f(x_k) + \frac{b-a}{n}\sum_{k=i}^{n} x_k f(x_k)] \tag{11}$$

3. Implementation of online optimization system

The components of the online optimization system entity are shown in Fig. 3. The system consists of two servers. One server is configured as an interface message processor and integrated server, and the other is configured as a database server (Oracle 10.0 g) and web server. The system was developed in Browser/Server mode, ensuring any computer connected to the local area network can visit it by web browsing. This system is based on the Java programming language and was developed on the Eclipse platform, which is a well-known cross-platform with a freely integrated development environment.

The system of the proposed methodology implementation logic is shown Fig. 4. The interface message processor acquires real-time data from plant information system (PI) by the OPC interface and transmits them to the integrated server and database server based on the UDP protocol. Its sampling interval is 1 min. The integrated server processes data read from the database based on the proposed methodology. Meanwhile, calculation results are also written into the database server. In order to improve the real-time performance of Web-based publishing, the database is divided into two parts: a buffer database (real-time database) and a historical database. Both of them were configured in the database server. The real-time database only stores the latest data, which is relatively small and used for real-time transmission. All buffered data are automatically entered into the historical database for query, analysis, and statistics. The web server is responsible for data issuing. Data are retrieved from the real-time database and sent to the Web server for users to browse. The entire working process takes 0.53 s which can meet the real-time requirement.

4. Industrial case study

4.1. Data description of the case steam-turbine system

The unit is configured with a PI, which records the historical operation data. Considering the influence of the load and ambient temperature, we sampled the variables historical runtime data in January, May, and August 2016 from PI to ensure that the data was representative. The historical data were sampled in 1 min, with 1440 samples per day and a total of 133,923 sample in 3 months.

4.2. Steady-state detection and work condition classification

After data cleaning and steady-state detection, a total of 40,001 sample data remained. The steady-state detection results of one week are shown in Fig. 5.

The ranges of the load and ambient temperature are 450–1020 MW and 5–35 °C, respectively. The data were classified into 57 work conditions based on the work condition classification rule discussed in Section 2.3. They are coded as listed in Table 3.

4.3. Decision making sample selection

The sample data of every work condition were partitioned into four clusters based on FCM. The cluster with the minimum heat consumption rate was selected as the decision-making sample. Take work condition W307 (load interval is [630, 660) MW, temperature interval is [25, 35] °C) as an example herein. The others can be analyzed in the same way. They are not described in detail owing to space limitations. The results are shown in Table 4.

It can be seen from the table that No.4 cluster's center is 7852.2 kJ/(kW·h), which is the minimum among the four clusters. Therefore, this cluster was selected as the decision-making samples. It should be noted that the number of categories is set manually, considering the scale of each cluster. The larger the category number, the fewer samples there are in each category. Considering that the estimation of the probability density distribution needs sufficient samples, the number of categories in this work is taken as 4.

4.4. Reference values estimation of KPI

The probability density distributions of KPI were estimated by KDE, as illustrated in Fig. 6. The expectation of each KPI was calculated and is listed in Table 5.

4.5. Online optimization system for the case unit

Fig. 7 is a screenshot of the online data-mining based optimization system of the case unit. It shows the real-time value, reference value, and deviation of KPI. Operators can adjust the KPI according to the reference value shown in the system interface. In addition, users can query the KPI historical trends by clicking the 'Statistics' button.

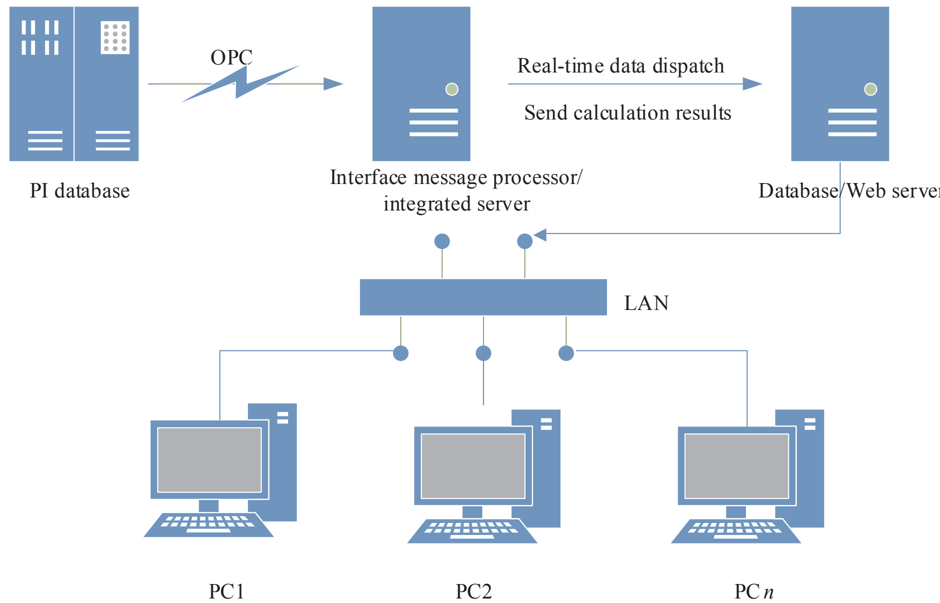

Fig. 3. Components of the optimization system.

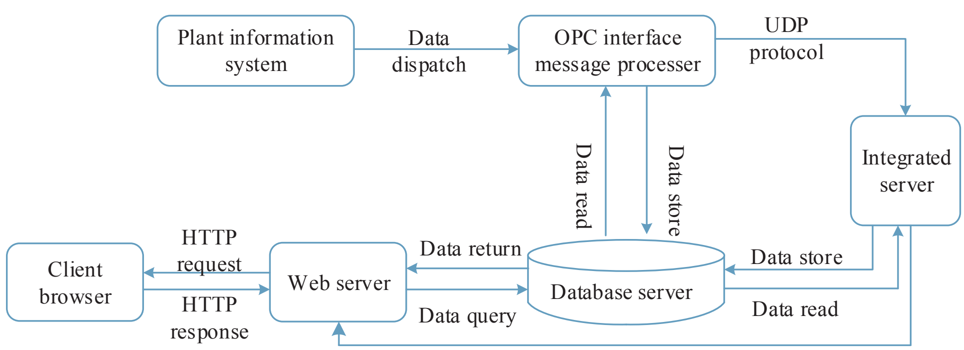

Fig. 4. The implementation logic of the system.

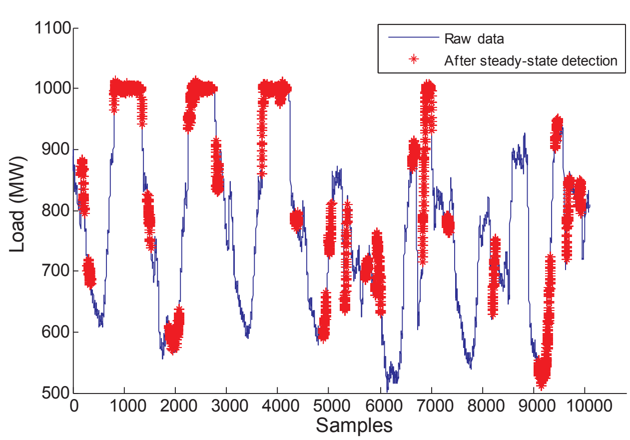

Fig. 5. Steady-state detection results of one-week sample.

5. Results and discussion

5.1. Energy-saving assessment

It is interesting to see how much energy can we save after adjusting the KPI. Compared with the historical data, the energy savings can be obtained according to the following equation:

Table 3
The coding rule of work condition.

Work condition code	Temperature interval (°C)	Load interval (MW)
W101	[5, 15)	[450, 480)
W102	[5, 15)	[480, 510)
……	[5, 15)	……
W119	[5, 15)	[990, 1020]
W201	[15, 25)	[450, 480)
……	[15, 25)	……
W219	[15, 25)	[990, 1020]
W301	[25, 35]	[450, 480)
……	[25, 35]	……
W319	[25, 35]	[990, 1020]

Table 4
Clustering results of work condition W307.

Number	Number of sample	Clustering center (kJ/(kW·h))	Clustering interval (kJ/(kW·h))
1	336	8481.3	[8458.6, 8700.1]
2	311	8321.6	[8259.8, 8458.8]
3	296	8172.6	[8079.1, 8258.0]
4	**216**	**7852.2**	**[7851.9, 8078.2]**

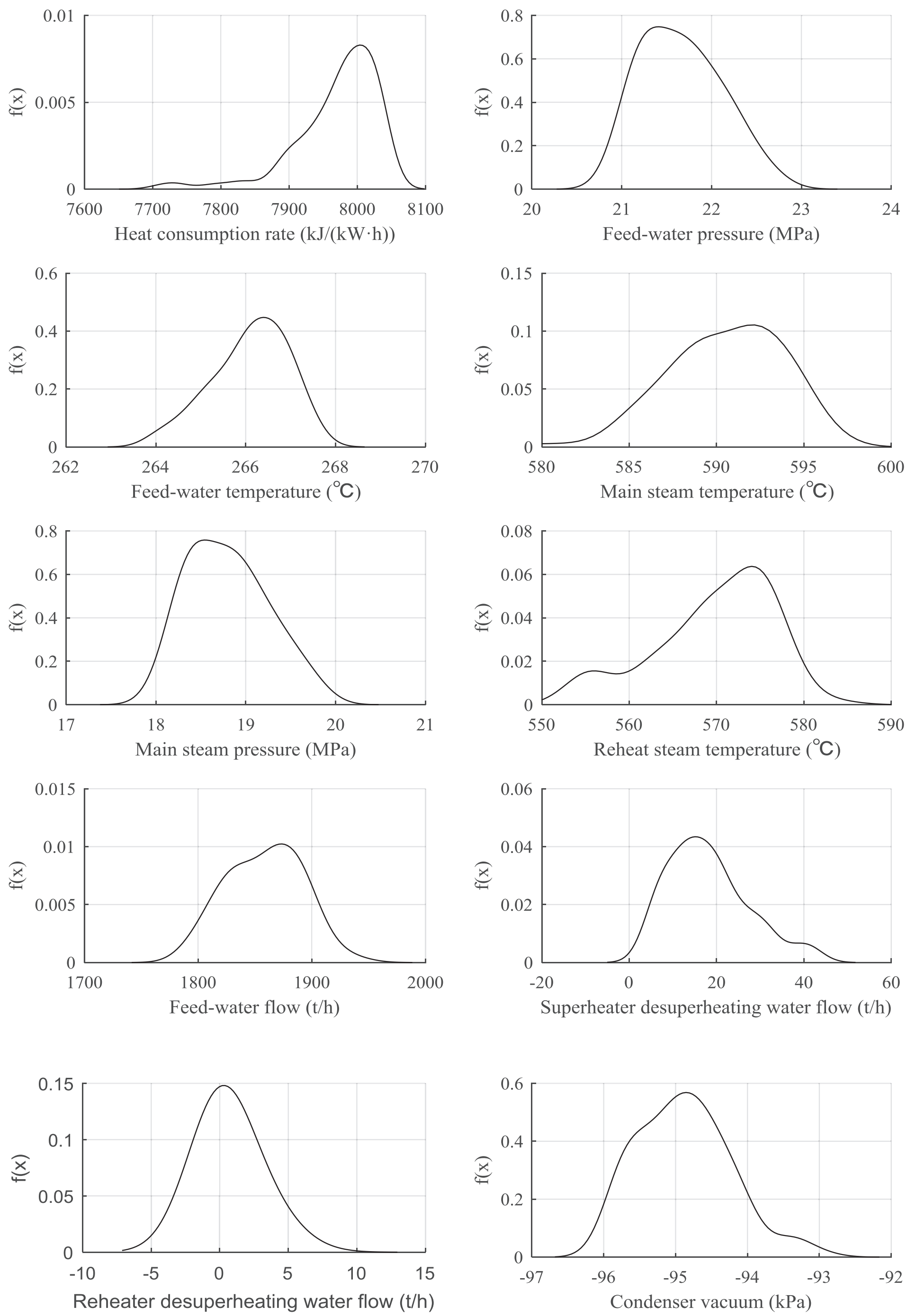

Fig. 6. Probability density distributions of KPI.

Table 5
Reference values of KPI at work condition W307.

Code	P_1	T_1	T_2	C	T_{fw}	P_{fw}	G_{fw}	G_{sd}	G_{rd}
Unit	MPa	°C	°C	kPa	°C	MPa	t/h	t/h	t/h
Reference	18.1	591.1	564.0	95.2	266.1	20.8	1849.4	17.8	3.2

$$\text{Energy saving} = \sum_{i=1}^{n} \frac{(HR_{i,\text{optimized}} - HR_{i,\text{actual}})^{*}L_i}{60} \tag{12}$$

where n is the number of sample data; $HR_{i,\text{actual}}$ and $HR_{i,\text{optimized}}$ are the historical heat consumption rate and optimized heat consumption rate of the i th sample, respectively; L_i is the megawatt load. It should be noted that the optimized state refers to the unit state after the operator adjusts the KPI according to the reference value proposed in this paper.

The heat consumption rate after optimized was calculated based on a particle-swarm optimization and support vector machine (PSO-SVM) regression model, which was discussed in detail in the literature [4]. The remaining 40,001 sample data after data pre-processing in Section 4.2 were divided into two distinct sets: training and test sets. A total of 30,001 pairs of data were taken as the training set. The testing set consists of the remaining 10,000 pairs of sample. The result and prediction error of the regression model based on PSO-SVM are shown in Figs. 8 and 9, respectively. From the figures, we can see that the prediction relative error is within (−0.01, 0.01), which means the model is quite accurate.

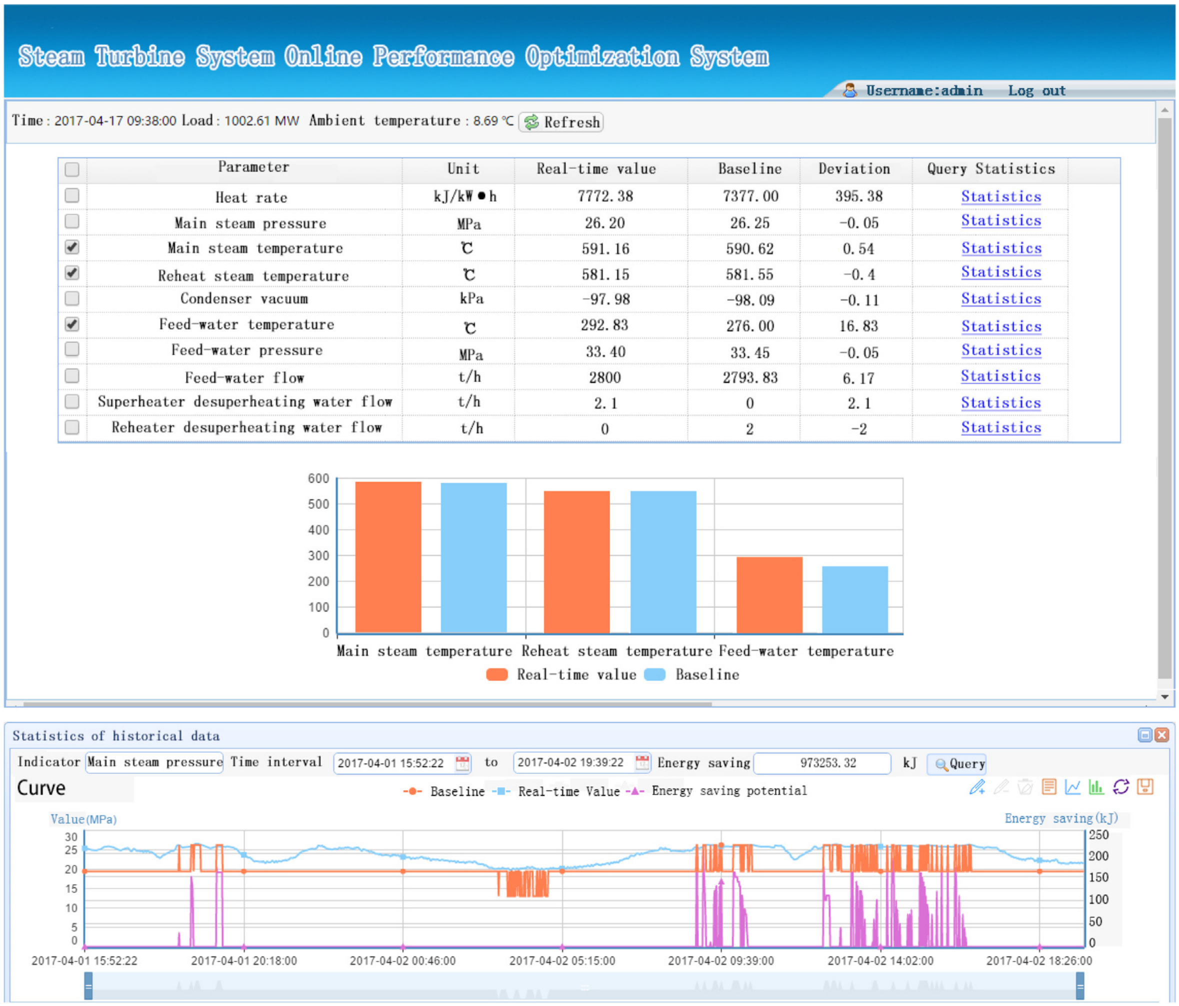

Fig. 7. A screenshot of the online optimization system for the studied steam-turbine system.

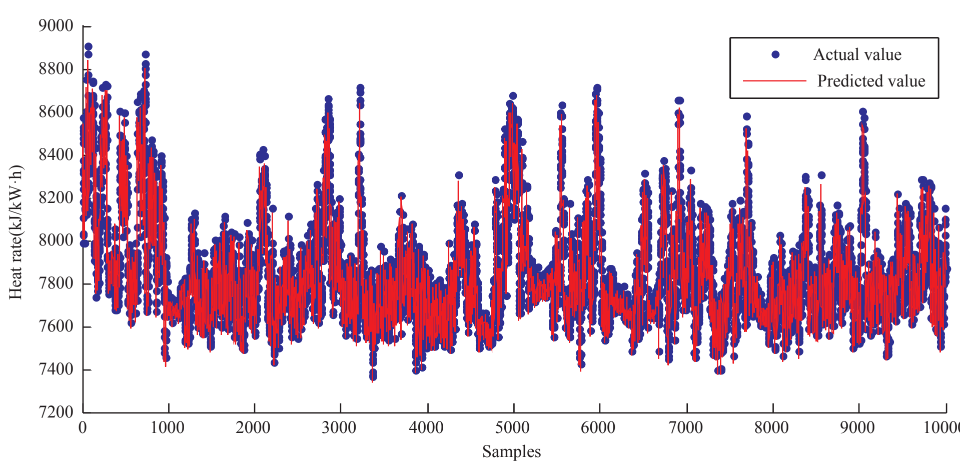

Fig. 8. Regression results of heat consumption rate based on PSO-SVM.

Heat consumption rate of all 57 work conditions (case unit studied) were obtained based on the PSO-SVM regression mode. It can be seen from Fig. 10 that the heat consumption rate (performance) is lower (higher) than most of the actual historical operation levels at all ambient temperature ranges after optimization. The energy saving potential ranges from 57 to 226 kJ/(kW·h) at different loads. In comparison with the historical data, we can calculate that the energy saving reached up to 79,000 GJ during the studied three months, based on Eq. (12). Given that the boiler efficiency is 93%, it can save 2,902.16 t of standard coal, which is remarkable.

5.2. Discussion

The comparison of the reference value, mean value, and design value of the KPI of work condition W307 is shown in Table 6. We can see that all reference values are better than the historical operation

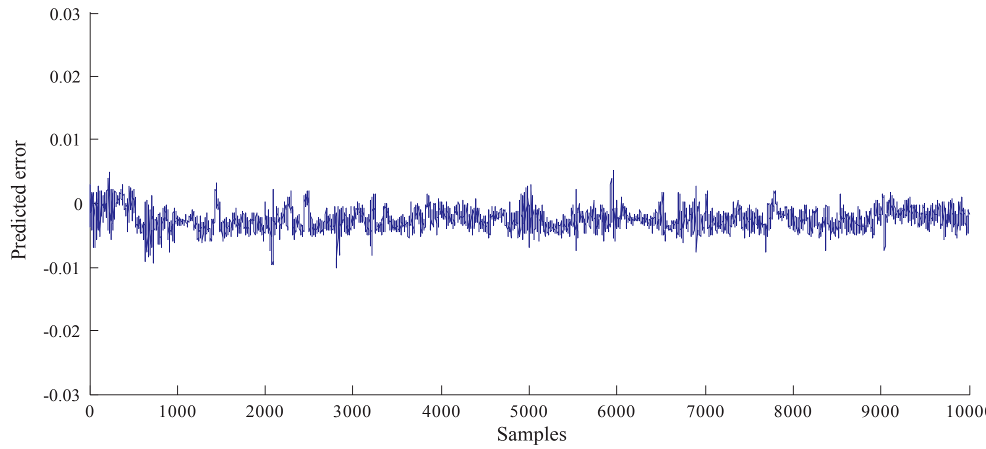

Fig. 9. The relative error of the regression model.

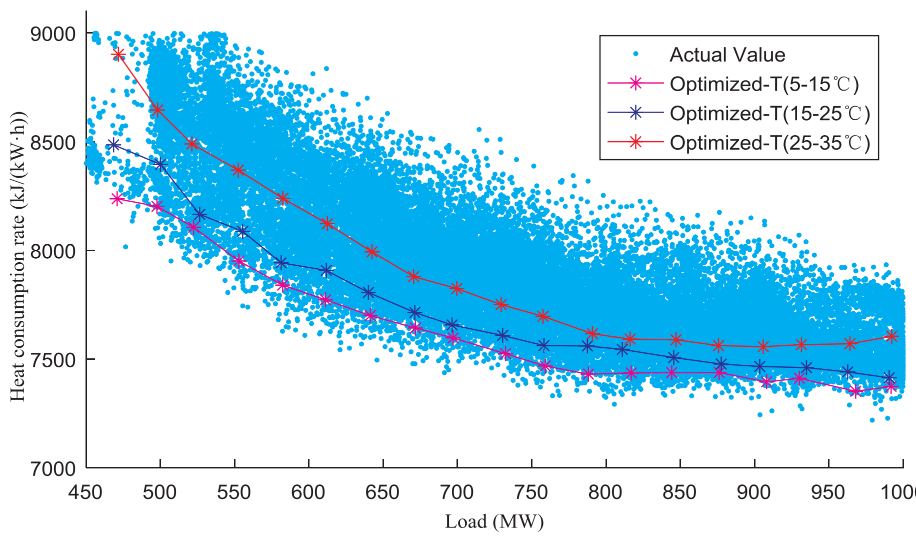

Fig. 10. Comparison between historical and optimized values of heat consumption rate.

Table 6
Comparisons of KPI of work condition W307.

Code	Unit	Reference value	Mean value	Design value
P_1	MPa	18.1	17.8	19.4
T_1	°C	591.1	587.8	600
T_2	°C	564.0	563.0	600
C	kPa	95.2	94.2	95.3
T_{fw}	°C	266.1	265.7	270.9
P_{fw}	°C	20.8	20.3	21.8
G_{fw}	t/h	1849.4	1916.4	1977.1
G_{sd}	t/h	17.8	17.2	0
G_{rd}	t/h	3.2	4.0	0
HR	kJ/kW·h	7852.2	8002.1	7503.0

mean level. That is, the performance of the steam-turbine system can be further enhanced. However, the reference values are worse than the design values. As was discussed in the induction part, the performance of the unit tends to deteriorate along with the running time. Most design values are the ceilings for the old on-duty units which are unavailable. They are unsuitable to act as reference values in actual production.

In contrast with the design or experiment-based method, the proposed method can provide KPI reference values under 57 work conditions, rather than some typical work condition, like 1000 MW, 750 MW, and 500 MW. Considering the load transients and off-design situation, the reference values of the KPI in the 57 work conditions proposed in this paper are more suitable for the realistic operation.

The comparison of the heat consumption rate in the 57 work conditions between the optimized state and the mean level based on the PSO-SVM model is illustrated in Fig. 11.

It can be seen that both the ambient temperature and load can influence the heat consumption rate. The heat consumption rate increases along with the decrease in unit load. In the case of a certain load, the higher the ambient temperature, the greater the heat consumption rate. Therefore, it is necessary to account for the load and ambient temperature when evaluating the system performance. In addition, the figure clearly shows that after optimization, the heat consumption rate (performance) is lower (higher) than the historical average level at any work condition. That is, the heat consumption rate will decrease and the steam-turbine system performance will be improved when the operator adjusts the KPI, referring to the proposed reference value.

6. Conclusions

In this paper, a novel methodology was presented to optimize the steam-turbine system performance from an operator's perspective. The main concept of the optimization is to determine the reference value of KPI. The proposed methodology was developed using a larger amount of historical runtime data and was demonstrated in an actual on-duty unit.

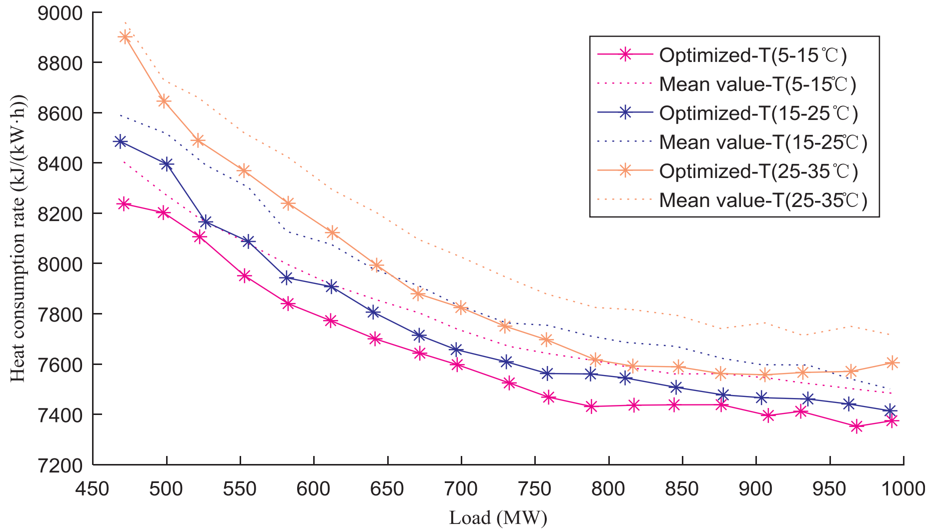

Fig. 11. Comparison between mean and optimized values of heat consumption rate.

Decision samples selection is a key part of the reference-value determination. FCM worked well in selecting the cluster with the minimum heat consumption rate. Expectations were employed as reference values of KPI after achieving the probability density distribution estimated by KDE. The application for the steam-turbine system was discussed in detail. The online optimization system for the steam-turbine system has been configured in the studied on-duty unit.

It is shown that the energy savings reached up to 79,000 GJ, or 2902.16 t of standard coal, during the studied three months, which means that the reference value is helpful for improving the steam-turbine system. Furthermore, the proposed methodology can provide the KPI reference value in a series of work conditions, considering load and ambient temperature variations. This can significantly improve the performance of steam-turbine systems for the operator.

It should be noted that employing expectation as the reference value may not be the best state of the steam-turbine system. However, it is better than the current standard and is an available state for the operator. Additionally, when operators adjust the KPI according to the reference value, the historical states tend to be improved gradually. Therefore, the optimization is a dynamic process. The clustering algorithm learns and generates new knowledge along with the updated sample data, which can be used to update the reference value.

It should also be noted that one limitation of the proposed method is that the study is based on huge amounts of historical operation data. For example, if there were no sufficient data for clustering, the results would not be satisfactory. In addition, the proposed methodology is suitable for steady-state working conditions. Hence, it would be worthwhile to study the transitional working conditions in the future.

Finally, the proposed methodology is also suitable for boiler performance optimization and operational optimization of the whole coal-fired power plant.

Acknowledgements

The data described in this paper were supported by Guohua Power Generation Company.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.applthermaleng.2019.02.032.

References

[1] M. Zeng, S. Wang, J. Duan, J. Sun, P. Zhong, Y. Zhang, Review of nuclear power development in China: environment analysis, historical stages, development status, problems and countermeasures, Renew. Sustain. Energy Rev. 59 (2016) 1369–1383.
[2] Z.Y. Zhao, J. Zuo, P. Wu, H. Yan, G. Zillante, Competitiveness assessment of the biomass power generation industry in China: a five forces model study, Renew. Energy 89 (2016) 144–153.
[3] Y. Gu, J. Xu, D. Chen, Zh. Wang, Q. Li, Overall review of peak shaving for coal-fired power units in China, Renew. Sustain. Energy Rev. 54 (2016) 723–731.
[4] J. Xu, Y. Gu, D. Chen, Q. Li, Data mining based plant-level load dispatching strategy for the coal-fired power plant coal-saving: a case study, Appl. Therm. Eng. 119 (2017) 553–559.
[5] L. Wang, Y. Yang, C. Dong, T. Morosuk, G. Tsatsaronis, Parametric optimization of supercritical coal-fired power plants by MINLP and differential evolution, Energy Convers. Manage. 85 (2014) 828–838.
[6] P. Ran, Research on Operation Optimization of Power Plant Thermal System by Dynamic Data Mining Approaches, North China Elect. Power University, 2012.
[7] A. Kusiak, Z. Song, Combustion efficiency optimization and virtual testing: a data-mining approach, IEEE Trans. Ind. Inform. 2 (3) (2006) 176–184.
[8] J.M. Blanco, L. Vazquez, F. Peña, D. Diaz, New investigation on diagnosing steam production systems from multivariate time series applied to thermal power plants, Appl. Energy 101 (2013) 589–599.
[9] Z. Song, A. Kusiak, Constraint-based control of boiler efficiency: a data-mining approach, IEEE Trans. Ind. Inform. 3 (1) (2007) 73–83.
[10] N. Wang, P. Fu, H. Xu, D. Wu, Zh. Yang, Y. Yang, Heat transfer characteristics and energy-consumption benchmark condition with varying operation boundaries for coal-fired power units: an exergy analytics approach, Appl. Therm. Eng. 88 (2015) 433–443.
[11] P. Niu, W. Zhang, Model of turbine optimal initial pressure under off-design operation based on SVR and GA, Neurocomputing 78 (1) (2012) 64–71.
[12] J. Li, C. Niu, J. Liu, Application of data mining technique in optimizing the operation of power plants, J. Power Eng. 26 (6) (2006) 830–835.
[13] J. Li, J. Liu, L. Zhang, c.h. Niu, The research and application of fuzzy association rule mining in power plant operation optimization, Proc. CSEE 26 (20) (2006) 118–123.
[14] H. Zhao, P. Wang, J. Qian, Z. Su, X. Peng, Modeling for target-value of boiler monitoring parameters based on fuzzy c-means clustering algorithm, Proc. CSEE 31 (32) (2011) 16–31.
[15] J. Xiong, H. Zhao, C. Zhang, Ch. Zheng, B.L. Peter, Thermoeconomic operation optimization of a coal-fired power plan, Energy 42 (1) (2012) 486–496.
[16] Xi'an thermal power research institute. Energy consumption analysis and energy-saving diagnosis technology for coal-fired generating units. China Electr. Power Press, Beijing, 2014.
[17] Yasin Tunckaya, Etem Koklukaya, Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural-based models, J. Energy Inst. (2014) 11–18.
[18] C. Chen, Z. Zhou, G.M. Bollas, Dynamic modeling, simulation and optimization of a subcritical steam power plant. Part I: plant model and regulatory control, Energy Convers. Manage. 145 (2017) 324–334.
[19] W. Zhang, P. Niu, G. Li, P. Li, Forecasting of turbine heat consumption rate with online least squares support vector machine based on gravitational search algorithm, Knowledge-Based Syst. 39 (2013) 34–44.
[20] X. Jiang, P. Liu, Z. Li, Data reconciliation and gross error detection for operational data in power plants, Energy 75 (2014) 14–23.
[21] S. Narasimhan, R. Mah, A. Tamhane, J. Woodward, A composite statistical test for detecting changes of steady states, AIChE J. 32 (9) (1986) 1409–1418.

[22] J.M. Blanco, L. Vazquez, F. Peña, Investigation on a new methodology for thermal power plant assessment through live diagnosis monitoring of selected process parameters; application to a case study, Energy 42 (1) (2012) 170–180.
[23] ASME, Steam-turbines, Performance Test Codes, ASME PTC 6-2004, 2004.
[24] J. Bezdek, R. Ehrlich, W. Full, The fuzzy c-means clustering algorithm, Comput. Geosci. 10 (2–3) (1984) 191–203.
[25] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[26] K. Zhou, S. Yang, Exploring the uniform effect of FCM clustering: a data distribution perspective, Knowledge-Based Syst. 96 (2016) 76–83.
[27] M. Liukkonen, T. Hiltunen, Adaptive monitoring of emissions in energy boilers using self-organizing maps: an application to a biomass-fired CFB (circulating fluidized bed), Energy 73 (2014) 443–452.
[28] Z. Sheng, S. Xie, C. Pan, Probability and Statistics, Higher Education Press, Beijing, 2008.
[29] S. Miao, K. Xie, H. Yang, R. Karki, H. Tai, T. Chen, A mixture kernel density model for wind speed probability distribution estimation, Energy Convers. Manage. 126 (2016) 1066–1083.
[30] Rao BLSPrakasa, Nonparametric Function Estimation, Academic Press, London, 1983.
[31] Q. Li, N. Wang, D. Yi, Numerical Analysis, Tsinghua University Press, Beijing, 2008.