An operating zone model for safety and efficiency monitoring of power generation units in thermal power plants

Mengyao Wei [a], Jiandong Wang [a,*], Song Gao [b], Jie Li [c], Xiangkun Pang [b]

[a] *College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, 266590, Shandong Province, China*
[b] *Shandong Electric Power Research Institute, Jinan, 250003, Shandong Province, China*
[c] *Shandong Huaneng Power Generation Co. Ltd., Jinan, 250013, Shandong Province, China*

ARTICLE INFO

ABSTRACT

Safe and efficient operations of thermal power plants become increasingly important, especially in compensating for electricity supply fluctuations in renewable energy. This paper proposes a method to build an operating zone model for safety and efficiency monitoring of power generation units in thermal power plants. The operating zone is a high-dimensional geometric space formed by all steady-state operating points in safe conditions of the coal flow rate, steam valve position, steam pressure and active power. Those operating points are obtained based on allowable variation ranges of process variables and mechanistic models describing the relationships among process variables. The main technical challenge is how to measure uncertainties of mechanistic models estimated from historical data disturbed by noticeable measurement noise. Bayesian theory and goodness-of-fit tests are exploited to tackle this challenge by yielding probability density functions of model parameters. A performance index is defined based on the operating zone model to assess the safety and efficiency of power generation units. Industrial case studies at a large-scale thermal power plant are provided to illustrate the effectiveness of the proposed method.

1. Introduction

Renewable energy has developed rapidly because of its environmental friendliness. However, the intermittent and random nature of renewable energy poses a threat to the stable operation of power grids (Su, Sun, Xue, & Lee, 2023). Thermal power plants compensate for fluctuations in renewable energy by operating flexibly in a wide load range (Zhang, Gao, Yue and Zhang, 2022). Therefore, it is important to ensure the safe and efficient operations of thermal power plants (Mukherjee, Gupta, Deodhar, & Runkana, 2023; Xie, Liang, Zhang, & Zhu, 2019).

The existing methods for operation monitoring of thermal power plants may be roughly grouped in two categories. The first category is model-based methods. Fu et al. (2016) presented a diagnosis method based on the comparison of exergy destructions between operation and reference states for thermal power plants. Tian, Xu, Yuan, Zhang, and Wang (2017) constructed a monitoring platform with the estimation of key variables and performance evaluation for thermal power plants. Ahmadi et al. (2019) provided thermodynamic and economic analysis on the performance of thermal power plants. The second category is data-driven, which contain three sub-categories. The first is based on the signal processing techniques. Blanco, Vazquez, and Peña (2012) identified quasi-steady states by exploiting noise-filtering theories. Vazquez, Blanco, Ramis, Peña, and Diaz (2015) monitored the operation performance of thermal power plants through deviations between steady states and expected values. Second, machine learning techniques are used to identify abnormal conditions. Moradi, Chaibakhsh, and Ramezani (2018) recognized abnormal conditions of thermal power plants by combining the outlet results of four existing classifiers. Hundi and Shahsavari (2020) demonstrated the efficacy of anomaly detection based on thermal power plant models established by four machine learning methods. Wang, Zhou, Chen, and Wang (2020) provided a mixed hidden naive Bayesian model for anomaly detection in thermal power plants. The third is based on the techniques of multivariate statistical analysis. Principal component analysis (PCA) (Omrani, Beiragh, & Kaleibari, 2015) and multistage PCA (Mahmoudi, Emrouznejad, Khosroshahi, Khashei, & Rajabi, 2019) were taken to evaluate the efficiency of thermal power plants. An autocorrelation feature analysis algorithm (Ma, Wu, Gao, Hou, & Wang, 2023) and a sparse dynamic inner PCA (Zhang, Zhou, Chen, & Hong, 2023) were introduced for process monitoring of thermal power plants. Wang, Zhou, and Chen (2023) developed a dynamic-related component analysis to distinguish anomalies in thermal power plants. To sum up,

---

* Corresponding author.
E-mail address: jiandong@sdust.edu.cn (J. Wang).

model-based methods require detailed process knowledge, so that they possess extrapolation properties, meaning that it is possible to use the laws of known information to obtain unknown information. Data-based methods reply on the information richness of training data. However, training data generally cannot cover all operating conditions.

This paper proposes a method to build an operating zone model for safety and efficiency monitoring of power generation units in thermal power plants. The operating zone is a high-dimensional geometric space formed by all steady-state operating points of power generation units in safe conditions. Those operating points are obtained based on allowable variation ranges of process variables and mechanistic models describing the relationships among process variables. The mechanistic model is formulated based on physical principles and its unknown parameters are estimated from historical data. The main technical challenge is how to measure uncertainties of mechanistic models caused by noticeable measurement noise. This challenge is resolved by exploiting Bayesian theory and goodness-of-fit tests to yield probability density functions (PDFs) of model parameters. A performance index is defined based on the operating zone model to assess the safety and efficiency of thermal power plants.

Operating zone models for process monitoring have been investigated in recent years. Brooks, Thorpe, and Wilson (2004) first proposed multi-variable operating zones derived from existing historical data to manage alarms for multivariate systems; however, no technique details were revealed therein for building mathematical models for operating zones. Afterwards, operating zone models were developed from historical data for multivariate alarm monitoring, using different mathematical models including convex hulls (Yu & Wang, 2019), hyperellipsoids (Yu, Wang, & Ouyang, 2019) and search cones (Wang, Wang, & Hou, 2022). Operating zone models in the above works are solely based on historical data and hardly have the extrapolation properties. Wang, Wang and Liu (2024) established an operating zone model for condensers in thermal power plants based on mechanisms, and considered model uncertainties by multiple sets of model parameter estimates from historical data. This paper attempts to extend the work (Wang, Wang, Liu, 2024) from condensers to power generation units in thermal power plants. There are two major differences: the mechanisms of power generation units are very different from the counterparts of condensers; both safety and efficiency indices are involved in power generation units, instead of only the safety considered in condensers.

The proposed method has three features distinguished from the existing methods. First, the proposed method does not require the data to follow certain statistical distributions such as the Gaussian distribution. Many existing methods based on multivariate statistical analysis essentially assume that data follow a Gaussian distribution. Second, the proposed method can identify safe conditions that do not appear in training data. The data-based method methods in general do not have such extrapolation properties. Third, the proposed method provides real-time monitoring results on the safety and efficiency performance of thermal power plants. By contrast, most existing methods evaluate the efficiency of thermal power plants based on operating data over a period of time rather than in a real-time manner.

The rest of the paper is organized as follows. Section 2 presents the problem description. Section 3 introduces details of the proposed method. Section 4 provides industrial applications of the proposed method in a large-scale thermal power plant. Some concluding remarks are given in Section 5.

2. Problem description

The operating principle of thermal power plants is to convert the heat energy into electrical energy. The heat energy created by coal combustion heats up the water in a boiler, and the associated water steam spins a steam turbine that drives an electrical generator to generate electricity (Fan, Su, Wang, & Lee, 2020). As indicated by some
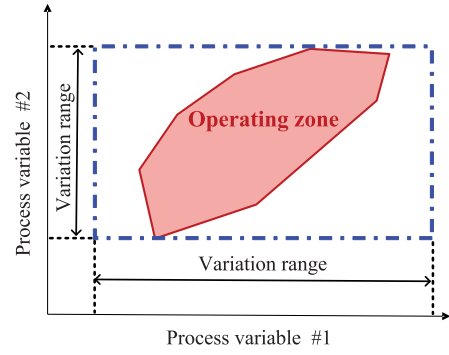

Fig. 1. Schematic diagram of an operating zone for two related process variables. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

existing models for thermal power plants (Gao, Hu, Zeng, Liu, & Niu, 2021; Tan, Fang, Tian, Fu, & Liu, 2008), main process variables in power generation units are the coal flow rate, steam valve position, steam pressure and active power. When thermal power plants are in safe conditions, process variables often do not take fixed values, but are subject to certain variation ranges. The variation ranges are denoted as

$$
\begin{aligned}
E_{min} &\leq E \leq E_{max}, \\
P_{T_{min}} &\leq P_T \leq P_{T_{max}}, \\
\mu_{T_{min}} &\leq \mu_T \leq \mu_{T_{max}}, \\
B_{min} &\leq B \leq B_{max},
\end{aligned}
\tag{1}
$$

where E is the active power, P_T is the steam pressure, μ_T is the steam valve position, and B is the coal flow rate. The subscripts *min* and *max* denote the minimums and maximums of variation ranges, respectively. The allowable variation ranges of four process variables together formulate a rectangular four-dimensional geometric space. For the demonstration purpose, Fig. 1 shows a rectangular space corresponding to the allowable variation ranges of two process variables. However, process variables are related in mechanisms, causing their variation ranges to be mutually constrained. It means that some parts of the rectangular space cannot be reached. Hence, the allowable variation ranges of two related process variables in safe conditions formulate a special space, schematically shown as the red area in Fig. 1, which is referred to as an operating zone in this context.

Let us assume that historical data samples of E, P_T, μ_T and B are given for thermal power plants in safe conditions. Our objective is to build an operating zone model for describing a geometric space composed by operating points of E, P_T, μ_T and B for power generation units. Safety and efficiency indices are defined to monitor the performance of thermal power plants. The operating zone models are developed from historical data disturbed by noticeable measurement noises. Therefore, a technical challenge is how to measure model uncertainties, in order to make sure that the operating zone model is trustworthy in a certain statistical sense.

3. Proposed method

This section proposes a method to establish operating zone models for safety and efficiency monitoring of power generation units, including the modeling of power generation units, the measurement of model uncertainties, and the building of operating zone models with performance indices.

3.1. Modeling of power generation units

This subsection introduces a mechanistic model of power generation units with the identification of unknown model parameters.

Modeling of power generation units in thermal power plants has been extensively studied (Celis, Pinto, Teixeira, & Xavier, 2017; Oko & Wang, 2014; Zhang, Gao, Yu, Fan and Zhang, 2022). Based on these results, a mechanistic model is built to describe the relationships among the process variables E, P_T, μ_T and B. The model consists of a combustion model, an evaporation model, a superheater model, and a turbine model. The details are given as follows.

For the combustion model, the dynamic balance between the feed coal and the heat is regarded as a first-order inertia process (Tan et al., 2008),

$$\tau_f \frac{dQ_w}{dt} = -Q_w + k_f B, \tag{2}$$

where Q_w is the heat from combustion in the furnace, k_f is a gain coefficient, and τ_f is the inertia time of the combustion process. The relationship between the drum pressure and the heat in the evaporation model is represented as Liu, Yan, Zeng, Hu, and Lv (2015)

$$C_d \frac{dP_D}{dt} = Q_w - k_s \sqrt{P_D - P_T}, \tag{3}$$

where P_D is the drum pressure, C_d denotes the thermal inertia, and k_s is the conversion coefficient for pressure and flow. The superheater model in terms of thermodynamic balance is characterized as Gao et al. (2021)

$$C_t \frac{dP_T}{dt} = k_s \sqrt{P_D - P_T} - k_t P_T \mu_T, \tag{4}$$

where C_t is the heat storage coefficient, and k_t is the flow coefficient of the steam valve. The turbine model can be described as another first-order inertia process (Tian, Yuan, Zhang, Kong, & Wang, 2018),

$$\tau_e \frac{dE}{dt} = -E + k_e P_T \mu_T, \tag{5}$$

where τ_e is the inertia time of the turbine generator process, and k_e is the power coefficient.

To sum up, the mechanistic model of power generation units in thermal power plants is formulated by four differential equations in (2)–(5). Unknown model parameters τ_f, k_f, C_d, k_s, C_t, k_t, τ_e and k_e are identified from historical data samples $\left\{ E(t), P_T(t), \mu_T(t), B(t) \right\}_{t=1}^T$. A fitting-error function f measures the errors between the model outputs $\hat{E}(t)$ and $\hat{P}_T(t)$ to the actual output $E(t)$ and $P_T(t)$,

$$f = \sum_{t=1}^T \left(\nu \frac{\left\| \hat{E}(t) - E(t) \right\|}{E(t)} + (1 - \nu) \frac{\left\| \hat{P}_T(t) - P_T(t) \right\|}{P_T(t)} \right), \tag{6}$$

where ν is a weighting coefficient, and $\|\cdot\|$ takes the Euclidean norm of operands. It is ready to estimate these model parameters by the genetic algorithm to solve an optimization problem minimizing the fitting-error function f.

3.2. Measurement of model uncertainties

This subsection measures the uncertainties of the mechanistic model based on Bayesian theory and goodness of fit tests.

Samples of process variables in thermal power plants are often disturbed by measurement noise, as shown in Fig. 4 (to be given later in Section 4), and the coal flow rate B among them is the most severely disturbed. As a result, there are uncertainties in mechanistic models identified from historical data. Operating zone models exploit mechanistic models to describe the relationships of process variables. Therefore, it is necessary to measure the uncertainties of the mechanistic model for thermal power plants.

The main idea is as follows. Since operating zone models are based on static relationships of process variables, only static parameters in $\theta = \{k_f, k_t, k_e\}$ are the ones to be considered in this sequel. Model uncertainties can be measured by the $(1 - \alpha)$ confidence interval of PDFs of $\theta = \{k_f, k_t, k_e\}$. Multiple sets of model parameters are estimated as $\{\hat{\theta}(d)\}_{d=1}^D$ from different groups of data samples. Bayesian theory is

exploited to yield the PDF $\hat{p}_D(\theta)$ of θ based on $\{\hat{\theta}(d)\}_{d=1}^D$. Goodness-of-fit tests are utilized to determine whether the PDF $\hat{p}_D(\theta)$ is convergent for a given sample size D. If the PDF is convergent, then the mechanistic model uncertainties are well captured to be used for establishing operating models.

The main idea is realized in the following steps. First, model uncertainties are described by the $(1 - \alpha)$ confidence interval of PDFs of model parameters. As θ consists of k_f, k_t, and k_e, the model uncertainties are measured by the smallest space Θ_u satisfying the triple integral equality,

$$\iiint_{\Theta_u} \hat{p}(\theta) d\theta = 1 - \alpha. \tag{7}$$

Here Θ_u is a set of model parameters to describe model uncertainties, and $\hat{p}(\theta)$ is the PDF of θ.

Second, the PDF $\hat{p}_D(\theta)$ of θ is estimated based on samples $\{\hat{\theta}(d)\}_{d=1}^D$ by Bayesian theory (Walpole, Myers, Myers, & Ye, 2011). The parameter $\hat{\theta}$ is obtained from a group of historical data, to formulate the model parameter set $\{\hat{\theta}(d)\}_{d=1}^D$. The three-dimensional sample space of $\theta = (k_f, k_t, k_e)$ is divided equally into G small spaces. The gth small space corresponds to the model parameter $\theta_g = (k_{f,g}, k_{t,g}, k_{e,g})$ with $g \in [1, G]$. For the demonstration purpose, the sample space of two model parameters is divided as illustrated in Fig. 2. The probability of the gth small space is

$$f_g = \frac{d_g}{D}, \tag{8}$$

where d_g is the number of samples from the total D samples that fall into the gth sample space. Let f_g be regarded as the realization of a random variable F_g. Since the true value of f_g may take any value within the gth space with the same chance, it is reasonable to assume that the prior PDF $p_{F_g}(f_g)$ is a continuous uniform distribution,

$$p_{F_g}(f_g) = \begin{cases} 1, & \text{if } 0 < f_g < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Given the total number D of realizations, the number d_g under the conditional probability f_g is a realization of a random variable D_g subject to a binomial probability distribution, i.e.,

$$p_{D_g | F_g}(d_g | f_g) = \frac{D!}{d_g! (D - d_g)!} \cdot f_g^{d_g} \cdot (1 - f_g)^{D - d_g}. \tag{10}$$

Using the Bayesian formula, the posterior PDF of f_g is

$$p_{F_g | D_g}(f_g | d_g) = \frac{p_{D_g | F_g}(d_g | f_g) \cdot p_{F_g}(f_g)}{\int p_{D_g | F_g}(d_g | f_g) \cdot p_{F_g}(f_g) df_g}. \tag{11}$$

The Bayesian estimate of f_g is often taken as the conditional mean, i.e.,

$$\hat{f}_g = \int f_g \cdot p_{F_g | D_g}(f_g | d_g) df_g. \tag{12}$$

Therefore, the PDF of θ with the sample size D is estimated as

$$\hat{p}_D(\theta_g) = \hat{f}_g, \quad g = 1, \dots, G. \tag{13}$$

Finally, goodness-of-fit tests (Walpole et al., 2011) are employed to determine whether the PDF $\hat{p}_D(\theta)$ is convergent. The posterior PDF of f_g tends to be constant and close to the true distribution as the sample size D increases (Walpole et al., 2011). A null hypothesis is given by

$$H_0 : \hat{p}_D(\theta) = \hat{p}_{D-1}(\theta). \tag{14}$$

Here $\hat{p}_{D-1}(\theta)$ and $\hat{p}_D(\theta)$ are the estimated PDFs of θ from $\{\hat{\theta}(d)\}_{d=1}^{D-1}$ and $\{\hat{\theta}(d)\}_{d=1}^D$, respectively. A statistic S_D is constructed to characterize the degree of deviation between $\{\hat{\theta}(d)\}_{d=1}^D$ and $\hat{p}_{D-1}(\theta)$,

$$S_D = \sum_{g=1}^G \frac{\left(d_g - (D-1) \cdot \hat{p}_{D-1}(\theta_g) \right)^2}{(D-1) \cdot \hat{p}_{D-1}(\theta_g)} \sim \chi^2(G-1). \tag{15}$$

The statistic S_D approximately follows the χ^2 distribution with the degree of freedom as $(G - 1)$. Accepting the null hypothesis means that
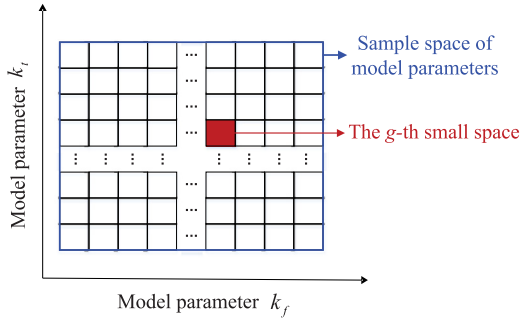
Fig. 2. Schematic diagram of the division of the sample space of model parameters.

the PDF $\hat{p}_D(\theta)$ is convergent. If the null hypothesis is rejected, it is necessary to keep increasing the sample size D until the null hypothesis is accepted. Based on the convergent PDF $\hat{p}_c(\theta) = \hat{p}_D(\theta)$, the model uncertainty is obtained by (7) as

$$\Theta_{u,c} = \{\theta_g | g \in I_{u,c}\} = \{k_{f,g}, k_{t,g}, k_{e,g} | g \in I_{u,c}\}, \tag{16}$$

where $I_{u,c}$ is a set of model parameters belonging to the $(1 - \alpha)$ confidence interval of the convergent PDF $\hat{p}_c(\theta)$.

3.3. Building of operating zone models with performance indices

This subsection presents operating zone models with performance indices for safety and efficiency monitoring in thermal power plants.

First, operating zones are formed by all safe operating points obtained through allowable variation ranges of process variables and mechanistic models with uncertainties. According to the allowable variation ranges of process variables in (1), the model inputs B and μ_T are discretized respectively to take values from their own minimums B_{min} and $\mu_{T_{min}}$ to maximums B_{max} and $\mu_{T_{max}}$ in step sizes B_s and μ_s, i.e.,

$$B[n_B] = B_{min} + B_s n_B, n_B = 1, \dots N_B,$$
$$\mu_T[n_\mu] = \mu_{T_{min}} + \mu_s n_\mu, n_\mu = 1, \dots N_\mu. \tag{17}$$

Here $N_B = \left\lfloor \frac{B_{max} - B_{min}}{B_s} \right\rfloor$ and $N_\mu = \left\lfloor \frac{\mu_{T_{max}} - \mu_{T_{min}}}{\mu_s} \right\rfloor$ are the number of possible values for B and μ_T, respectively. Symbol $\lfloor \cdot \rfloor$ represents the largest integer no larger than the operand. There are $N_B \cdot N_\mu$ combinations between all possible values of B and μ_T, which cover the entire range of input variables. Correspondingly, the output variables P_T and E are generated based on the static relationship in (18) between inputs and outputs derived from (2)–(5),

$$P_T[n; \theta_g] = \frac{k_{f,g}}{k_{t,g}} \frac{B[n_B]}{\mu_T[n_\mu]}, E[n; \theta_g] = \frac{k_{f,g} k_{e,g}}{k_{t,g}} B[n_B]. \tag{18}$$

Here $n \in [1, N_B N_\mu]$ is the sampling index associated with the possible values for B and μ_T, $n_B = n - \left\lfloor \frac{n-1}{N_B} \right\rfloor N_B$ and $n_\mu = \left\lfloor \frac{n-1}{N_B} \right\rfloor - \left\lfloor \left\lfloor \frac{n-1}{N_B} \right\rfloor / N_\mu \right\rfloor N_\mu + 1$ are related to n, and the model parameter vector $\theta_g = (k_{f,g}, k_{t,g}, k_{e,g})$ is a group of parameters from $\Theta_{u,c}$ in (16). According to (17) and (18), steady-state operating points are collected for mechanistic models with model parameters θ_g to formulate a set S_g,

$$S_g = \{x[n; \theta_g]\}_{n=1}^{N_B N_\mu}$$
$$= \{E[n; \theta_g], P_T[n; \theta_g], \mu_T[n_\mu], B[n_B]\}_{n=1}^{N_B N_\mu}. \tag{19}$$

Here the values of B and μ_T are inside their allowable variation ranges. Thus, safe operating points are selected based on the steady-state operating points where the values of E and P_T are within their allowable variation ranges shown in (1), i.e.,

$$S_{s,g} = \Big\{ x[n; \theta_g] | E_{min} \leq E[n; \theta_g] \leq E_{max},$$
$$P_{T_{min}} \leq P_T[n; \theta_g] \leq P_{T_{min}}, \forall x \in S_g \Big\}. \tag{20}$$

Considering the model uncertainty $\Theta_{u,c} = \{\theta_g | g \in I_{u,c}\}$ in (16), the above steps in (18)–(20) are repeated to traverse all model parameter θ_g in $\Theta_{u,c}$. Then, a set S_{all} containing all safe operating points are obtained as

$$S_{all} = \bigcup_{g \in I_{u,c}} S_{s,g}. \tag{21}$$

All safe operating points in S_{all} formulate a four-dimensional geometric space, referred to as an operating zone model. Based on our previous work on alarm monitoring (Wang, Wang and Hou, 2024), a search cone model is used to formulate the operating zone model as

$$f_{ozm} = \left\{ A_q X_q^T - B_q \leq 0 \right\}_{q=1}^{Q}, \tag{22}$$

where

$$\begin{cases} A_q = \left[a_{q,1}, \dots, a_{q,m}, \dots, a_{q,M} \right]^T \\ B_q = \left[b_{q,1}, \dots, b_{q,m}, \dots, b_{q,M} \right]^T \\ 0 = [0, \dots, 0, \dots, 0]^T \end{cases}. \tag{23}$$

Here $a_{q,m}$ is the unit normal vector of the mth facet in the qth search cone, $b_{q,m}$ is the offset distance from the mth facet in the qth search cone to the origin point, Q is the number of the search cone, $M = 2^{N-1}(N - 1) + 1$ is the number of the facet in the search cone, $N = 4$ is the dimension of the operating point $x = (E, P_T, \mu_T, B)$, and X_q is the set of vertices of qth search cones in the operating zone model.

Second, a performance index is defined to assess the safety and efficiency of thermal power plants. The safety index is defined as the shortest nominal distance from the operating point x to the operating zone boundaries,

$$\phi(x) = min\{\phi_1(x), \dots, \phi_q(x), \dots, \phi_Q(x)\}. \tag{24}$$

Here $\phi_q(x)$ with $q \in [1, Q]$ is the nominal distance from the operating point x to the qth operating zone boundary. It is calculated based on the nominal distance from the operating point x to the Mth facet of the qth search cones in the operating zone model,

$$\phi_q(x) = \frac{\bar{a}_{q,M} \cdot (\bar{x} - \bar{x}_q)}{|\bar{a}_{q,M}|}. \tag{25}$$

Here $(\bar{x} - \bar{x}_q)$ denotes the vector consisting of the operating points x and x_q, and x_q is an operating point on the Mth facet of the qth search cone, which satisfies an equality

$$a_{q,M} x_q^T - b_{q,M} = 0. \tag{26}$$

The efficiency of thermal power plants (Ahmadi et al., 2019) is calculated as

$$\eta(x) = \frac{E}{B \cdot H_{coal}}, \tag{27}$$

where H_{coal} is the lower heating value of coal.

Efficiency makes sense only when safety requirements are met. Therefore, a performance index $\varphi(x) \in [0, 1]$ integrating the safety defined by (24) and efficiency by (27) of thermal power plants is

$$\varphi(x) = \begin{cases} \omega \cdot \eta'(x) + (1 - \omega) \cdot \phi'(x), if & \phi'(x) > 0, \\ 0, & if \quad \phi'(x) \leq 0. \end{cases} \tag{28}$$

Here ω is a weighting coefficient, $\eta'(x)$ denotes the normalized efficiency index $\eta(x)$, and $\phi'(x)$ is the normalized safety index $\phi(x)$. In order to visually monitor the performance of thermal power plants, the values of φ for all operating points inside the operating zone are calculated by (28), to yield contour lines in the B-E plane of the operating zone model as shown in Fig. 15 (to be given later in Section 4).
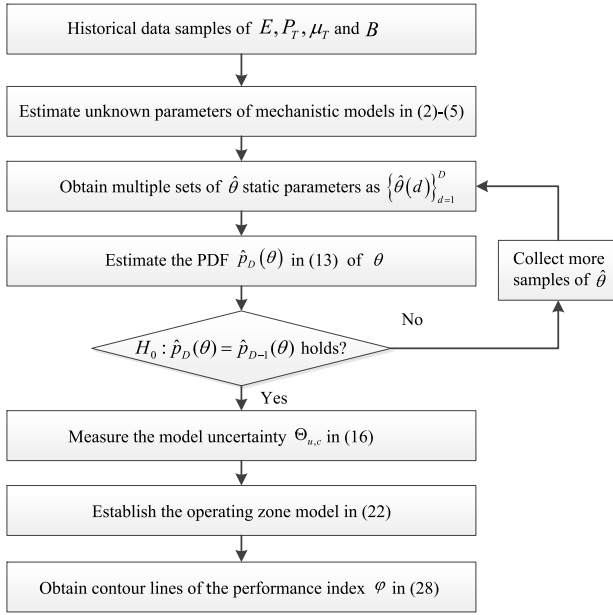
Fig. 3. The flow chart of the proposed method.

Fig. 4. Time sequence plots of (a) E, (b) P_T, (c) μ_T, and (d) B for estimating model parameters.

Fig. 5. Time sequence plots of process variables and estimated outputs from the dynamic model and static model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Summary

The flow chart of the proposed method is presented in Fig. 3. The proposed method consists of six steps:

Step 1: The mechanistic models in (2)–(5) of power generation units in thermal power plants are estimated from historical data.

Step 2: The PDF $\hat{p}_D(\theta)$ in (13) of $\theta = (k_f, k_t, k_e)$ is determined by Bayesian theory based on model parameter samples $\{\hat{\theta}(d)\}_{d=1}^{D}$.

Step 3: Steps 1 and 2 are repeated by using more historical data for gradually adding $\hat{\theta}$ to $\{\hat{\theta}(d)\}_{d=1}^{D}$ until the goodness of fit test in (14) is accepted.

Step 4: The model uncertainty $\Theta_{u,c}$ in (16) is measured as the $(1 - \alpha)$ confidence interval of the convergent PDF $\hat{p}_c(\theta)$ by (7).

Step 5: The operating zone model in (22) is obtained from the allowable variation ranges of process variables and the mechanistic model with uncertainties.

Step 6: The values of the performance index φ for all operating points inside the operating zone are calculated by (28), to yield contour lines of φ.

The operating zone model has a good potential to be used for process monitoring, even though it is static and industrial processes are often dynamic. The operating zone model for process monitoring is analogous to a map for traveling in mountains. Even though the map is static and traveling in mountains is dynamic, the map provides indispensable information for traveling safely and efficiently in mountains. Boundaries of the operating zone model (see e.g., the black solid lines in Fig. 15 given later in Section 4) are like cliff edges in the map. The safety index measures the distance of an operating point away from process operating limits, similar to the distance of a physical location away from cliff edges. If the distance is larger, then the location is safer. The efficiency index (see e.g., the contour lines in Fig. 15) represents the efficiency of an operating point, similar to the altitude of a physical location in mountains. If the altitude is higher, then the location has a better view in general. Hence, the efficiency index denotes the preference in selecting operating points or picking locations in mountains.
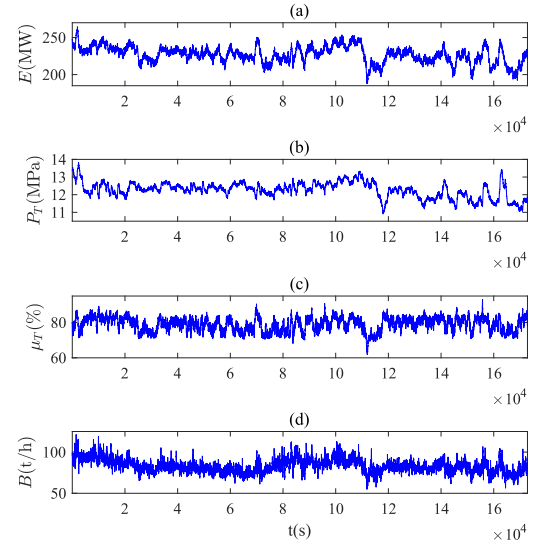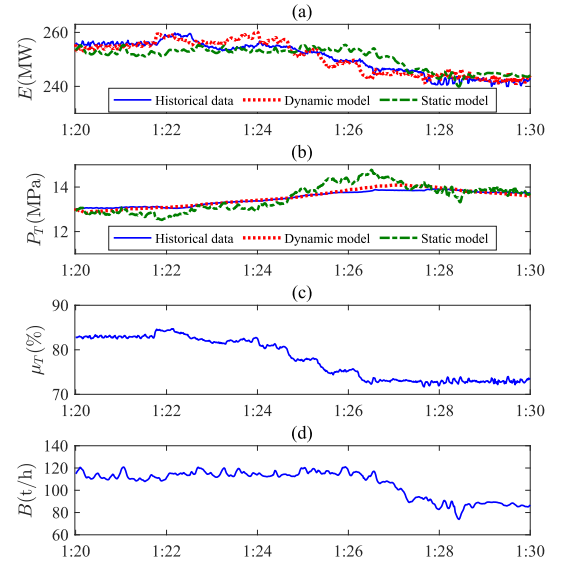
4. Industrial case studies

This section presents industrial case studies of operating zone models for the safety and efficiency monitoring of a large-scale thermal power plant.

4.1. Modeling of operating zone models

This subsection builds an operating zone model for a large-scale thermal power plant.

First, the mechanistic model of the power generation unit is established to describe relationships among the active power E, the steam pressure P_T, the steam valve position μ_T and the coal flow rate B. Fig. 4 presents historical data samples $\left\{ E(t), P_T(t), \mu_T(t), B(t) \right\}_{t=1}^{172800}$ with the sampling period 1 s for two days of the power generation unit. The model parameters in (2)–(5) are estimated from the historical data. The dynamic model is used to yield a more accurate static model than
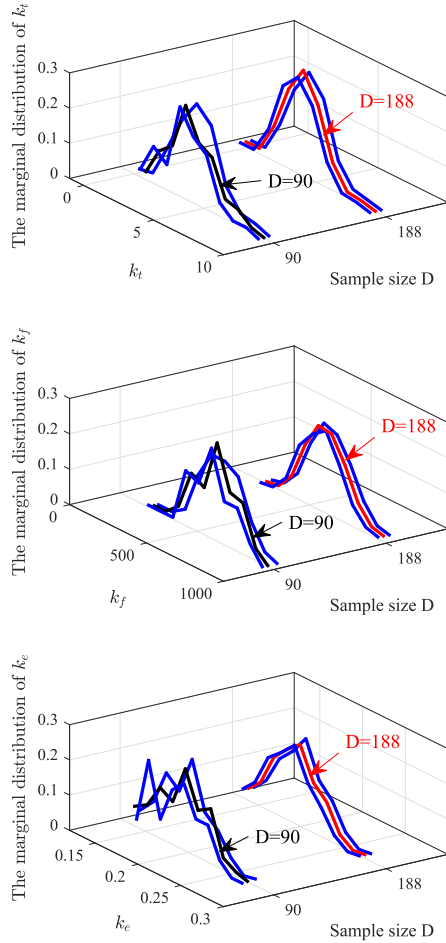
Fig. 6. Marginal distributions of k_f, k_t, and k_e of joint probability distribution $\hat{p}_D(\theta)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Operating zone models (black line), historical data (blue dot) for building the mechanistic model, and operating points from new data on two-dimensional planes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Ten operating points where E takes the value around 250 MW.

	E (MW)	P_T (MPa)	μ_T (%)	B (t/h)
#1	249.6	12.6	84.7	95.1
#2	250.5	12.7	84.0	112.2
#3	250.1	12.6	83.5	106.7
#4	249.9	13.7	76.1	99.3
#5	249.9	12.2	87.1	92.8
#6	250.2	15.1	72.0	105.4
#7	250.0	15.4	70.3	96.4
#8	249.5	14.4	77.8	101.9
#9	249.2	14.1	76.5	90.4
#10	250.6	14.0	77.8	90.8

another static model being estimated directly from data. This is due to two facts: the dynamic model better fits historical data containing both static and dynamic information, and the static model is a special case of the dynamic model. For illustration, Fig. 5 here presents some historical data, from which two models are estimated. They are the dynamic model and the static model being estimated directly from data. In Fig. 5-(a) and (b), the estimated outputs (red dots) from the dynamic model have smaller fitting errors than the counterparts (green dashes) from the static model being estimated directly from data. The fitting errors of the two models are calculated as 3.62 and 9.34 by (6).

Second, the uncertainty in the model parameters is measured. The PDF $\hat{p}_D(\theta)$ in (13) of θ is calculated on the basis of samples of $\{\hat{\theta}(d)\}_{d=1}^{D}$. Here $\theta = (k_f, k_t, k_e)$ is estimated from every four hours of historical data. The null hypothesis in (14) is not accepted until the sample size $D = 188$ by gradually adding the estimated parameter $\hat{\theta}$ to $\{\hat{\theta}(d)\}_{d=1}^{D}$. Thus, the PDF $\hat{p}_{188}(\theta)$ is regarded as being convergent. Fig. 6 presents the marginal distributions of k_f, k_t, and k_e, for a few sample size of D. The marginal distributions for $D = 188$ (red) are similar to the marginal distributions at $D = 187$ and $D = 189$. By contrast, the marginal distributions for $D = 90$ (black) are different from the counterparts for $D = 89$ and $D = 91$. Hence, the model uncertainty is $\Theta_{u,c}$ in (16) based on the 95% confidence interval of $\hat{p}_{188}(\theta)$.

Next, the operating zone model is built based on the allowable variation ranges of process variables and the mechanistic model with uncertainties. According to process knowledge of thermal power plants, the minimums and maximums in (1), (17) and (20) of the allowable variation ranges for four process variables are respectively
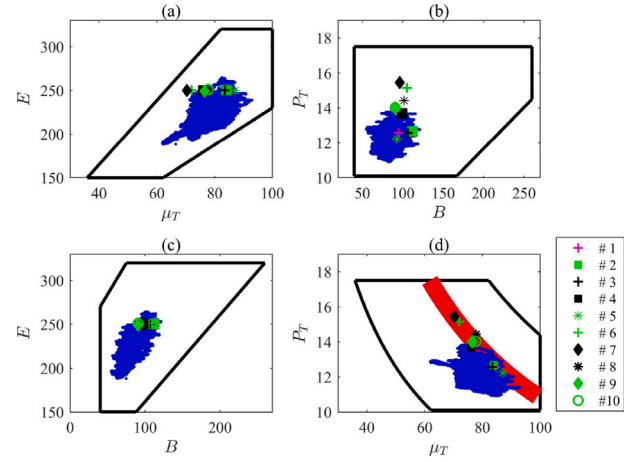
$E_{min} = 150$ MW, $E_{max} = 320$ MW;
$P_{T_{min}} = 10.1$ MPa, $P_{T_{max}} = 17.5$ MPa;
$\mu_{T_{min}} = 0\%, \mu_{T_{max}} = 100\%$;
$B_{min} = 40$ t/h, $B_{max} = 300$ t/h.

The model uncertainties are measured by multiple groups of model parameters in $\Theta_{u,c}$. The safe operating points are collected by (17)–(20) for the model parameters $k_{f,g}$, $k_{t,g}$, and $k_{e,g}$. Then traversing the model parameter θ_g in $\Theta_{u,c}$ yields all safe operating points in S_{all} by (21). These safe operating points in the four-dimensional space together formulate an operating zone model in (22). Fig. 7 shows projections on the two-dimensional space of the operating zone model, since the four-dimensional operating zone cannot be visualized in the Cartesian coordinate.

Finally, the operating zone model is verified to be accurate and possess extrapolation properties. Eq. (5) yields $E = k_e P_T \mu_T$ at steady-state based on mechanism analysis. Taking $E = 250$ MW as an example, the relationship between P_T and μ_T is obtained from the estimated model (5), as shown in the red band in Fig. 7-(d), by considering the uncertainty of the model parameter k_e. As shown in Fig. 7, the blue dots indicate the historical data for building the mechanistic model in Fig. 4, which cannot cover the entire operating zone. Ten data segments of E, P_T, μ_T, and B in steady-state conditions subject to $E = 250$ MW are found from new data in safe conditions. The corresponding ten operating conditions are obtained by calculating the means of these data segments, as given in Table 1. It is clear that similar values of E are associated with multiple different values of P_T and μ_T. As drawn in Fig. 7-(d), ten operating points are all inside the red band. In particular, the first to fifth operating points are inside the region
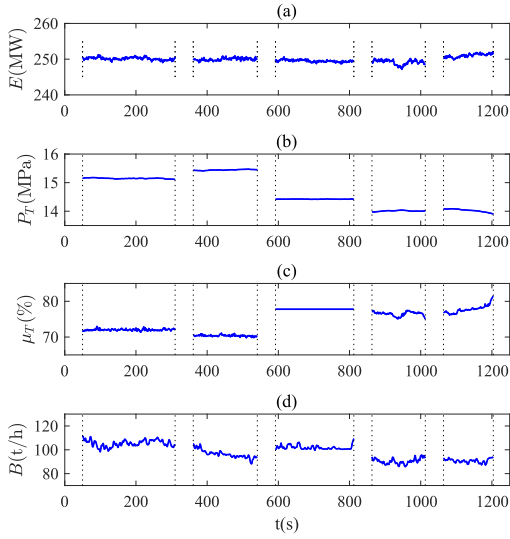
Fig. 8. Time sequence plots of (a) E, (b) P_T, (c) μ_T, and (d) B corresponding to the sixth to tenth operating points in Table 1.
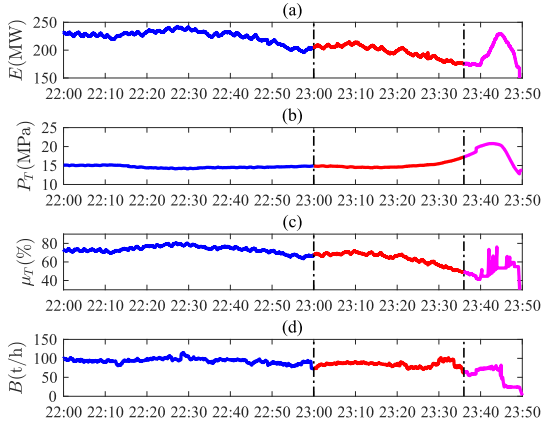

Fig. 9. Time sequence plots of (a) E, (b) P_T, (c) μ_T, and (d) B on January 24. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 10. Monitoring results based on the operating zone model at 23:30, 23:36, and 24:50.

formed by the historical data in Fig. 4 for modeling (blue dots), because similar operating points appear in the historical data. The sixth to tenth operating conditions with $P_T \geq 14$ MPa correspond to time sequence plots shown in Fig. 8. Such operating conditions are not present in the historical data for modeling, where all data samples of P_T in Fig. 4 take the values less than 14 MPa. Therefore, the operating zone model is applicable to operating conditions that do not appear in the historical data for modeling.

4.2. Industrial case study #1

This case study illustrates the good potential of applying the operating zone model in an abnormality detection, and compares with the PCA method as the representative of multivariate statistical analysis methods.

The time sequence plots of the process variables E, P_T, μ_T and B for thermal power plants on January 24 are shown in Fig. 9. The three colored segments indicate the historical data for time periods 22:00–23:30, 23:30–23:36, and 23:36–23:50. The operating zone model is used to monitor the operating condition of thermal power plants. The monitoring results at 23:30, 23:36, and 24:50 are given in Fig. 10. The thermal power plant is in safe operation from 22:00 to 23:30, and all
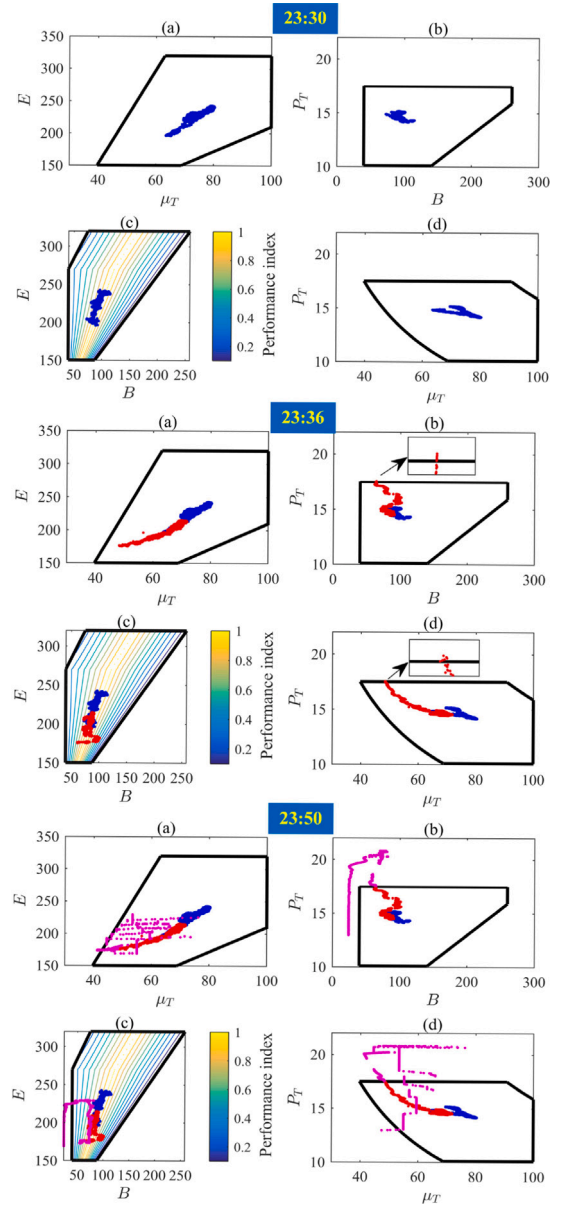
operating points are inside the operating zone model. During the period from 23:30 to 23:36, the trend of data points (red dots) approach the operating zone boundary as shown in Fig. 10. The proposed method detects an abnormality at 23:36 because the operating points are located outside the operating zone model, as illustrated in Fig. 10-(b) and (d) at 23:36. In a few minutes later, the thermal power plant goes out of control and shuts down in emergency. The reason for the abnormality is that the coal flow rate B increases in a wrong direction, while the active power E and the steam valve position μ_T decrease, as shown in Fig. 9, resulting in an excessive increment in the steam pressure P_T.

The proposed method is compared with the PCA method in multivariate statistical analysis. As shown in Fig. 11, a hyperellipsoid corresponding to the PCA method is obtained based on the same historical data of E, P_T, μ_T and B in Fig. 4 for building the operating zone model. Here the hyperellipsoid is the one to enclose 95% historical data samples. Fig. 12 provides the monitoring results based on the hyperellipsoid on January 24, where blue, red, and pink dots indicate the operating data in Fig. 9 for the periods 22:00–23:30, 23:30–23:36,
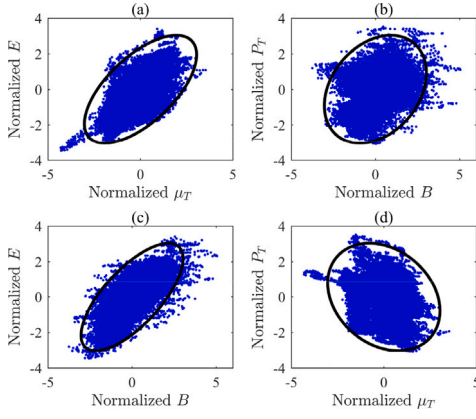
Fig. 11. Normalized data samples (blue dot) and the hyperellipsoid (black solid line) of the PCA method on two-dimensional planes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
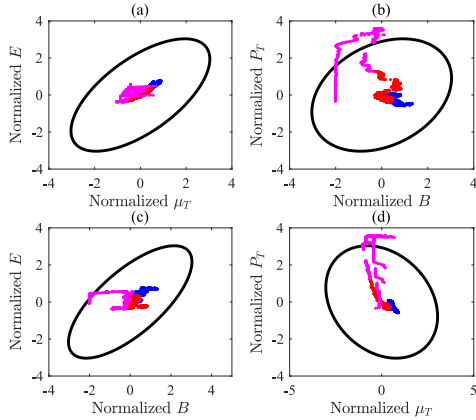

Fig. 12. Monitoring results based on the PCA method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 13. Time sequence plots of (a) E, (b) P_T, (c) μ_T and (d) B on May 13.

Fig. 14. Time sequence plots of (a) B, (b) P_T, (c) μ_T and (d) E on July 7.

and 23:36–23:50, respectively. Operating points in red go outside the hyperellipsoid in Fig. 12. Thus, the PCA method does not detect the abnormality until the last timer period, later than the proposed method. The reason is that historical data samples of the process variables E, P_T, μ_T and B do not satisfy the assumption of having a Gaussian distribution. This is revealed by a fact that data samples in Fig. 11 are not in the exact shape of hyperellipsoids. By contrast, the proposed method does not require the assumption that data samples follow a Gaussian distribution.

4.3. Industrial case study #2

This case study provides an application of the performance index to assess the safety and efficiency based on the operating zone model.

Figs. 13 and 14 give the time sequence plots of four process variables of the coal flow rate, steam valve position, steam pressure and active power on May 13 and July 7, respectively. Clearly, they receive similar power commands: the active power E is requested to increase from 184 MW to 248 MW on May 13; the active power E is requested to increase from 188 MW to 250 MW on July 7. Operators adjust the coal flow rate B and steam valve position μ_T to bring the active power E to the target power. There are multiple paths that allow E to reach the target power. However, different adjustment paths chosen by experienced or immature operators may have distinct operation performances.
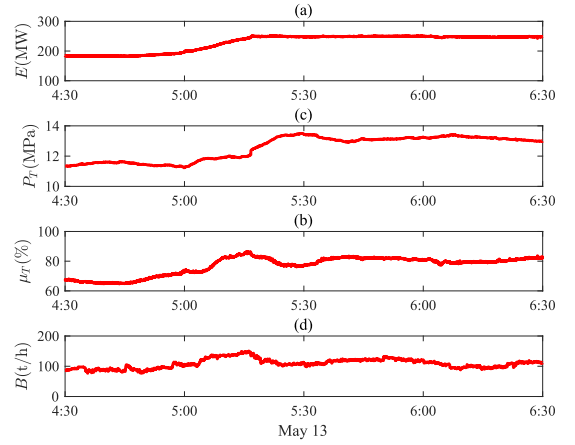
The operating zone model is used for performance monitoring of two power adjustment paths on May 13 and July 7. The monitoring results on May 13 and July 7 are respectively depicted by red dots and blue dots in Fig. 15. Many operating points on May 13 fall in the blue contour area with lower performance indices. In contrast, the operating points on July 7 are largely in the yellow contour area with larger performance indices. The mean values of the performance index φ in (28) for May 13 and July 7 are 0.763 and 0.915, respectively. Therefore, the safety and efficiency performance of the power adjustment path on July 7 is better than that on May 13. This is because an excess of B is provided to increase E on May 13. During this process, the steam pressure P_T has been increasing, while the steam valve position μ_T appears to increase and decrease. According to the operating experience of thermal power plants, it is known that the throttling loss increases due to the decrement of μ_T. By contrast, the two variables B and μ_T are well coordinated to achieve the increment in E on July 7.

As a comparison, the existing method is used to assess the performance of the two power adjustment paths. The efficiency of thermal power plants is defined based on the first law of thermodynamics (Ahmadi et al., 2019),

$$\eta_{unit} = \frac{W \cdot H_E}{C \cdot H_{coal}}.$$

Here $W = \frac{\sum_{t=1}^T E(t)}{3600/h}$ is the power generation capacity, $C = \frac{\sum_{t=1}^T B(t)}{3600/h}$ is the coal consumption, t is the sampling index being associated with a sampling period h, T is the number of samples, $H_E = 3.6$ MJ/kWh
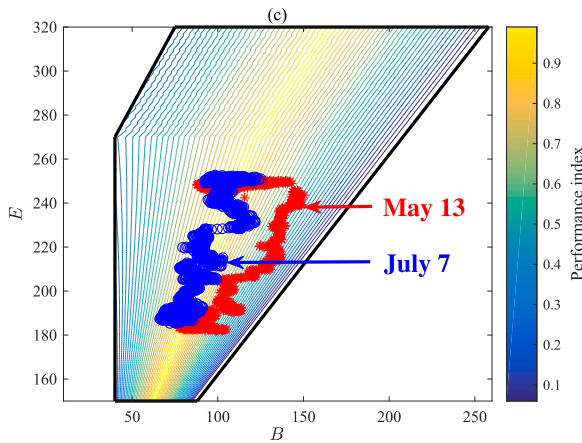
Fig. 15. Comparison of performance monitoring results for May 13 (red points) and July 7 (blue points). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Performance monitoring results for the proposed and existing methods.

Method	Index	May 13	July 7
Existing	W (kWh)	457 803.3	450 796.2
	C (kg)	218 001.3	186 123.6
	η_{unit}	25.83%	29.79%
Proposed	$\bar{\varphi}$	0.76	0.92

denotes the heat per kWh, and H_{coal} = 29.27 MJ/kg is the lower heating value of standard coal. Table 2 gives the power generation capacities and the coal consumptions of the two power adjustment processes with $T = 7200$ and $h = 1$ s. The values of η_{unit} on May 13 and July 7 are calculated as 25.83% and 29.79% respectively. Thus, the efficiency of the power adjustment path on July 7 is better than the counterpart on May 13, which is the same as the result obtained by the proposed method. However, the existing method evaluates the efficiency performance based on historical data from thermal power plants over a time period. It cannot provide real-time information on performance monitoring to assist operators in making adjustments. By contrast, the proposed method can do so, by promptly noticing the poor performance and adjusting the path afterwards.

5. Conclusion

In this paper, an operating zone model was proposed for the safety and efficiency monitoring of power generation units in thermal power plants. The operating zone model was obtained through allowable variation ranges of process variables and mechanistic models with uncertainties. Bayesian theory and goodness-of-fit tests were exploited to resolve a main technical challenge for measuring model uncertainties caused by measurement noises. A performance index was defined based on the operating zone model to assess the safety and efficiency of thermal power plants.

The operating zone model in this paper has a good potential to be used for process monitoring, even though it is static and industrial processes are often dynamic. Industrial case studies in Sections 4.2 and 4.3 have shown such a potential. However, there are quite a few problems to be resolved in the future work for applying the operating zone model for process monitoring. For instance, the process may be falsely identified as being in abnormal conditions when it is in a transition state and close to boundaries of the operating zone model; the resulting false alarms may be removed by delay timers to be designed. In addition, identifying root causes of abnormalities is important in practice; new root-cause analysis approaches may be formulated by looking at the directions of data points going outside of the operating zone model.

CRediT authorship contribution statement

Mengyao Wei: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. **Jiandong Wang:** Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization. **Song Gao:** Visualization, Validation, Formal analysis, Data curation, Conceptualization. **Jie Li:** Writing – review & editing, Validation, Resources, Data curation. **Xiangkun Pang:** Visualization, Resources, Formal analysis, Conceptualization.

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.

Acknowledgments

References

Ahmadi, M. H., Alhuyi Nazari, M., Sadeghzadeh, M., Pourfayaz, F., Ghazvini, M., Ming, T., et al. (2019). Thermodynamic and economic analysis of performance evaluation of all the thermal power plants: A review. *Energy Science Engineering, 7,* 30–65.

Blanco, J. M., Vazquez, L., & Peña, F. (2012). 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,* 170–180.

Brooks, R., Thorpe, R., & Wilson, J. (2004). A new method for defining and managing process alarms and for correcting process operation when an alarm occurs. *Journal of Hazardous Materials, 115,* 169–174.

Celis, C., Pinto, G. R. S., Teixeira, T., & Xavier, E. (2017). A steam turbine dynamic model for full scope power plant simulators. *Applied Thermal Engineering, 120,* 593–602.

Fan, H., Su, Z., Wang, P., & Lee, K. Y. (2020). A dynamic mathematical model for once-through boiler-turbine units with superheated steam temperature. *Applied Thermal Engineering, 170,* Article 114912.

Fu, P., Wang, N., Wang, L., Morosuk, T., Yang, Y., & Tsatsaronis, G. (2016). Performance degradation diagnosis of thermal power plants: A method based on advanced exergy analysis. *Energy Conversion and Management, 130,* 219–229.

Gao, Y., Hu, Y., Zeng, D., Liu, J., & Niu, Y. (2021). Improving the CCS performance of coal-fired drum boiler units base on PEB and DEB strategies. *Control Engineering Practice, 110,* Article 104761.

Hundi, P., & Shahsavari, R. (2020). Comparative studies among machine learning models for performance estimation and health monitoring of thermal power plants. *Applied Energy, 265,* Article 114775.

Liu, J. Z., Yan, S., Zeng, D. L., Hu, Y., & Lv, Y. (2015). A dynamic model used for controller design of a coal fired once-through boiler-turbine unit. *Energy, 93,* 2069–2078.

Ma, X., Wu, D., Gao, S., Hou, T., & Wang, Y. (2023). Autocorrelation feature analysis for dynamic process monitoring of thermal power plants. *IEEE Transactions on Systems, Man and Cybernetics, 53,* 5387–5399.

Mahmoudi, R., Emrouznejad, A., Khosroshahi, H., Khashei, M., & Rajabi, P. (2019). Performance evaluation of thermal power plants considering CO2 emission: A multistage PCA, clustering, game theory and data envelopment analysis. *Journal of Cleaner Production, 223,* 641–650.

Moradi, M., Chaibakhsh, A., & Ramezani, A. (2018). An intelligent hybrid technique for fault detection and condition monitoring of a thermal power plant. *Applied Mathematical Modelling, 60,* 34–47.

Mukherjee, T., Gupta, A., Deodhar, A., & Runkana, V. (2023). Real-time coal classification in thermal power plants. *Control Engineering Practice, 130,* Article 105377.

Oko, E., & Wang, M. (2014). Dynamic modelling, validation and analysis of coal-fired subcritical power plant. *Fuel, 135,* 292–300.

Omrani, H., Beiragh, R. G., & Kaleibari, S. S. (2015). Performance assessment of Iranian electricity distribution companies by an integrated cooperative game data envelopment analysis principal component analysis approach. *International Journal of Electrical Power & Energy Systems, 64,* 617–625.

Su, Z., Sun, L., Xue, W., & Lee, K. Y. (2023). A review on active disturbance rejection control of power generation systems: Fundamentals, tunings and practices. *Control Engineering Practice, 141,* Article 105716.

Tan, W., Fang, F., Tian, L., Fu, C., & Liu, J. (2008). Linear control of a boiler-turbine unit: Analysis and design. *ISA Transactions, 47,* 189–197.

Tian, Z., Xu, L., Yuan, J., Zhang, X., & Wang, J. (2017). Online performance monitoring platform based on the whole process models of subcritical coal-fired power plants. *Applied Thermal Engineering, 124*, 1368–1381.

Tian, Z., Yuan, J., Zhang, X., Kong, L., & Wang, J. (2018). Modeling and sliding mode predictive control of the ultra-supercritical boiler-turbine system with uncertainties and input constraints. *ISA Transactions, 76*, 43–56.

Vazquez, L., Blanco, J. M., Ramis, R., Peña, F., & Diaz, D. (2015). Robust methodology for steady state measurements estimation based framework for a reliable long term thermal power plant operation performance monitoring. *Energy, 93*, 923–944.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability and statistics for engineers and scientists* (9th ed.). Prentice Hall.

Wang, Z., Wang, J., & Hou, J. (2022). Multivariate alarm monitoring for non-convex normal operating zones based on search cones. *IEEE Transactions on Automation Science and Engineering, 21*, 452–462.

Wang, Z., Wang, J., & Hou, J. (2024). Multivariate alarm monitoring for non convex normal operating zones based on search cones. *IEEE Transactions on Automation Science and Engineering, 21*, 452–462.

Wang, Z., Wang, J., & Liu, J. (2024). Multivariate process monitoring for safe operation of condensers in thermal power plants based on normal operating zones. *IEEE Transactions on Control Systems Technology, 32*, 1399–1409.

Wang, Y., Zhou, D., & Chen, M. (2023). Dynamic related component analysis for quality-related process monitoring with applications to thermal power plants. *Control Engineering Practice, 132*, Article 105426.

Wang, M., Zhou, D., Chen, M., & Wang, Y. (2020). Anomaly detection in the fan system of a thermal power plant monitored by continuous and two-valued variables. *Control Engineering Practice, 102*, Article 104522.

Xie, J., Liang, Z., Zhang, X., & Zhu, L. (2019). Efficiency evaluation of thermal power plants in China based on the weighted russell directional distance method. *Journal of Cleaner Production, 222*, 573–583.

Yu, Y., & Wang, J. (2019). Alarm monitoring for multivariate processes based on a convex-hull normal operating zone. *IEEE Transactions on Control Systems Technology, 28*, 2649–2656.

Yu, Y., Wang, J., & Ouyang, Z. (2019). Designing dynamic alarm limits and adjusting manipulated variables for multivariate systems. *IEEE Transactions on Industrial Electronics, 67*, 2314–2325.

Zhang, H., Gao, M., Yu, H., Fan, H., & Zhang, J. (2022). A dynamic nonlinear model used for controller design of a 600 MW supercritical circulating fluidized bed boiler-turbine unit. *Applied Thermal Engineering, 212*, Article 118547.

Zhang, H., Gao, M., Yue, G., & Zhang, J. (2022). Dynamic model for subcritical circulating fluidized bed boiler-turbine units operated in a wide-load range. *Applied Thermal Engineering, 213*, Article 118742.

Zhang, J., Zhou, D., Chen, M., & Hong, X. (2023). Continual learning for multimode dynamic process monitoring with applications to an ultra–supercritical thermal power plant. *IEEE Transactions on Automation Science and Engineering, 20*, 137–150.