IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 31, 2019, accepted November 16, 2019, date of publication November 22, 2019,
date of current version December 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2955249

Abnormal Condition Monitoring and Diagnosis for Coal Mills Based on Support Vector Regression

XINGYUN HONG[1], **ZHENGGUO XU**[1], **(Member, IEEE), AND ZHENWEI ZHANG**[2]
[1]State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, China
[2]Zhejiang Provincial Energy Group Company, Ltd., Hangzhou 311121, China

Corresponding author: Zhengguo Xu (xzg@zju.edu.cn)

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61751307 and Grant 61973269.

ABSTRACT Abnormal condition monitoring is an essential part in ensuring the reliability and safety and guaranteeing the efficiency for industrial processes. This paper proposes a monitoring and diagnosis framework applied to coal mills in thermal power plants. The support vector regression (SVR) method with optimized parameters is utilized to detect the abnormal condition that the operating performances deviate from the normal levels expected. The support vectors in the trained models are considered as the representative operating conditions; then are used for responsible variables diagnosis, measuring how far each performance related variable deviates from its expected value when abnormal events occur. This approach is validated by three real cases from a thermal power plant, and the application results indicate that the proposed approach can detect abnormal conditions in time and accurately. Furthermore, effective diagnosis can be achieved and is validated by the case analysis conclusions from experts, which shows the effectiveness and potential value of the proposed approach.

INDEX TERMS Abnormal condition monitoring, fault diagnosis, support vector regression, coal mill, performance.

I. INTRODUCTION

Thermal power plant is an enormous and complicated factory for generating electricity. It consists of five major parts: fuel system, combustion system, steam-water system, electrical system and control system. Coal mill is an important auxiliary equipment in fuel system, used for pulverizing coal blocks into powder. The well grinded coal powder is then transported by the primary hot wind into the boiler for combustion. Thus, improving combustion efficiency depends heavily on the performance and productivity of the coal mills. According to Maffezzoni [1], incorrect coal fineness or improper drying will lead to burner line plugging, high unburnt carbon loss, slagging, fouling, etc. Inadequate coal powder supply may result in malfunction or shutdown of boilers. The performance of a coal mill can be affected by blockage, shutdown, fire or explosion, and wear of its components. In conclusion, the operating safety, reliability and economic efficiency of the whole plant are highly related to coal mill performance. Therefore, it is necessary to perform abnormal condition detection and diagnosis on coal mills, and optimize their operation, to improve the performance and efficiency of production.

Modernized power plants are usually equipped with large-scale sensors to monitor the real-time operating conditions, so the data generated from sensors can be utilized to ensure better performance or find abnormal conditions in the industrial process. Fault detection and diagnosis (FDD) methods based on sensor data is a major technique in ensuring safety and economy of coal mills, which arouse intensive attention in both industrial and academic fields. Fault detection and diagnosis for coal mills can be divided into mathematical model-based, signal processing based, multivariate statistical analysis based and machine learning based methods. Model-based methods use mathematical models to describe the operating behavior of coal mills, while other approaches only rely on process data analysis.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhigang Liu.

A multi-segment coal mill model that covers the whole milling process was presented in [2], which can be used for online monitoring and fault detection. Odgaard *et al.* [3] developed an optimal unknown input observer from a simplified energy balance model to detect fault, and compared it with motor power regression models that represented normal operating conditions (NOCs) with both static and dynamic principal component analysis and partial least squares (PLS). An on-line model based approach for tube-ball mill condition monitoring and fault detection was proposed in [4], the model parameters were on-line updated/optimized using Genetic Algorithms. A FDD system was developed by Agrawal *et al.* [5] with a dynamic mathematical model to generate residuals. Fuzzy logic was employed for residual evaluation to detect fault type, isolate fault and provide the severity information of the fault, while Bayesian network was used to troubleshoot the root cause. Self-organizing map was trained and used for the reliable isolation of the most frequent mill fault, output fuel-mixture drop due to the coal-stuck in the input bunker, in [6]. Qin *et al.* [7] established an expert database with fuzzy cluster analysis, PCA and tendency analysis, and fitted the multiple regression equations of main operation loops with support vector regression (SVR) to monitor the parameter variation tendency. Grey relation analysis, combined with neural network, was also studied to get accurate identification to early fault, fault degree as well as fault tendency [8]. Besides, signal processing methods such as wavelet analysis [9] and acoustic emission signal analysis [10], [11] were also applied to coal mill performance monitoring.

The accuracy of the mathematical model-based methods [12], [13] depends on how well the models are formulated. However, the coal mill is a highly non-linear system, and some key parameters are difficult to obtain. Data-driven methods don't require complicated process mechanism information or associated expert knowledge. However, signal processing approaches can be sensitive to changing of operating conditions, such as coal volumes, air flow volumes and environmental temperatures. As for statistical analysis approaches and machine learning approaches, Agrawal *et al.* [14] believed that the existing algorithms are unable to find new faults, locate root causes of the faults and interpret the diagnosis results.

As major methods used in industries', data-driven monitoring and fault diagnosis approaches have been applied to other equipment, though there is not much work oriented at coal mills. For example, Hilbert–Huang transform-based method [15], multivariate empirical mode decomposition (MEMD) [16], support vector machine (SVM) [17] and deep neural network (DNN) [18] were used for bearing fault detection or diagnosis. Bayesian network for pump fault diagnosis [19] and remaining useful life estimation [20]. Tree model based methods such as XGBoost [21] was applied to wind turbines, and also other neural network based methods, such as convolutional neural network (CNN) [22], echo state networks (ESN) [23] and deep belief network (DBN) [24]. These methods mainly extracted features that distinguish faulty conditions from normal conditions to detect faults, or used fault samples to classify faults.

The above-mentioned studies mainly focus on fault detection and diagnosis. In practice, however, there are some abnormal conditions, which are not necessarily caused by faults and may affect the equipment performances, to be identified and corrected. As for anomaly detection, data-based methods use the process data collected under the normal operating conditions to build an empirical model which is then used to estimate true values of new measurements and to detect and diagnose anomalies in future data by evaluating the residuals [25].There are statistical-based [26], [27], similarity-based [28] and machine learning based methods [29] aiming at addressing different anomaly detection problems in the industrial field.

As for the coal mill, in practice, the general monitoring scheme is to calculate the indicators under some specific operating conditions named standard conditions. However, there may be a case that the value of an indicator falls in the normal range under the standard conditions but an actual abnormal condition occurs because of the difference between the actual and standard operating conditions. Thus, this study aims to monitor coal mills so as to detect and analyze the conditions with the unexpected performances below the normal levels that may result from incipient faults or different operation settings. As for performance monitoring, Nikula *et al.* [30] provided a data-driven framework to monitor boiler efficiency together with its expected level. The boiler efficiency was represented by the variables that had the strongest correlations with the boiler efficiency according to information-theoretic variable ranking, and calculated using multiple linear regression or by selecting the highest historical efficiency. This information is essential to the operators in the plants, given the fact that a boiler can have hundreds of variables to be monitored and the efficiency is a key factor of the profits. When an abnormal condition occurs, which is indicated by more than one variable may go out of the normal regions, it requires much time to analyze the root causes, even for experienced experts. Therefore, with an automatic detection and diagnosis tool, the working stress of operators can be reduced, thus the operation costs can be cut down. However, multiple linear regression can be insufficient in describing nonlinear relations between the operating variables and the equipment performances such as boiler efficiency. Furthermore, to diagnose the root causes of the abnormal conditions, the detailed information which operating variables give most contributions to the abnormal conditions is also crucial.

This paper proposes an approach based on SVR to realize coal mill performance monitoring and abnormal condition diagnosis. The major tasks to be completed in this paper include: 1) estimate the value for each performance-related variable using the SVR model; 2) the estimated results are compared with the real values, and the residuals are used for anomaly detection; 3) when abnormal conditions occur, the support vectors, which are extracted from the SVR model, are used to diagnose the root causes of performance deviation by assigning ranks to the concerned variables. Three case studies are provided to evaluate and explain the results of the proposed approach and validate its effectiveness.

One of the difficulties to be dealt with in this paper is how to recognize abnormal conditions under different operating conditions. Another is how to accurately locate the root variables resulting in performance degradation. Therefore, the three major contributions of this paper are:

1) An integrated performance monitoring and diagnosis framework for coal mills is proposed. In this framework, SVR is used to build different models for estimation of different performance-related variables, and the monitored performances are compared with their historical recordings in real time to detect performance degeneration of coal mills under multiple operating conditions.

2) The support vectors in the trained SVR models are considered as the representative operating conditions, which indicates the marginal conditions of normal operating. A distance measure is used to find whether there is a similar operating pattern between a real-time observation and the representative support vectors, so as to monitor the deviation of the monitored variables under similar operating conditions and find the responsible variables for performance deviation.

3) The direction and degree of the performance deviation are also given based on distance measures, which is instructive and meaningful for operators in the plant. Based on the monitoring strategy, even without the prior fault information or fault samples, which are usually required in most diagnosis methods, the proposed approach is capable of locating the root causes of the abnormal conditions.

The remaining of the paper is organized as follows. Section II describes the coal mill system and how the performances are defined in this paper. Section III introduces the algorithms used in this paper. Section IV illustrates the case studies which present the implementation and application results of the approach. Finally, the paper ends with a conclusion and some perspectives in Section V.

II. COAL MILL

A. OPERATION OF COAL MILLS

Large-scale thermal power plants usually use medium speed mills which run at a setting speed in the range of 50$\sim$300r/min, because of their economic efficiency. First the coal briquettes fall in the center of the grinder through an inlet pipe, then they are crushed on the rotating grinding table by the grinding roller which is driven by the motor through the reducer. The pulverized coals on the edge of the mill are brought up and dried by the mixed primary air, thus the requirement for different humidity of the product can be satisfied by adjusting hot air temperature. Large and heavy coal particles cannot be carried by the mixed primary air, so they will fall directly onto the grinding table, which is so called the first separation. While passing the rotary separator, the coal particles comes to the second separation. The lower the rotating speed is, the easier the coal particles can pass through, which means larger particles tend to fall back under faster rotating speed. Therefore, the thickness of coal powder can be determined by different separator speeds.

B. MONITORED PERFORMANCES

Coal mill is a complicated nonlinear system, the variables in which are strongly correlated and coupled. Conventional control strategies use several single-loop PID controllers to separate the pulverization process into different phases. The values of some key operating variables, such as rotating speed of the rotary separator, are set by experiments.

Owing to its characteristics of nonlinearity and highly coupling, it is difficult to adjust the variables to ensure the coal mill to run under the optimal conditions. Furthermore, the variation of one single variable may affect other variables and leads to poor performance of the whole system, even resulting in faults like coal leakage, mill choking, wear of mill components and etc. For example, if the outlet temperature of the coal mill declines, there will be a significant deterioration appearing in the quality of coal powder. On the contrary, if the temperature exceeds the normal level, the pulverizer may fire.

The coal powder quality is related to three major factors, i.e., the grinding capability, coal drying capability, and ventilation capability of the coal mills. The grinding capability is determined by the rotary separator. The faster the separator rotates, the better coal powder is. The drying degree of the coal powder is determined by the outlet temperature. The ventilation capability is corresponding to the ventilation pressure that is denoted by the grinding bowl's differential pressure.

While the quality-related variables are all in the normal range, the energy consumption is taken into consideration. One of the major indicators, which evaluates whether the coal mill is running at a satisfactory state, is its power consumption that is calculated by dividing the consumed power per hour by the coal flow during the same period. In general, the lower the power consumption is, the higher the efficiency is. However, there are some special circumstances where this principle may be invalid. For example, when different types of coals are mixed for combustion, the power consumption may increase or remain in the normal range. If only this index is chosen as the sole indicator, a conclusion that pulverizing the coals with high water content will increase the efficiency will be drawn, which is unacceptable. Therefore, energy consumption should be analyzed on the basis of the three quality-related variables.

In conclusion, to estimate these four performances, different operating data sets are chosen for building four specific models according to the combustion mechanism and expert knowledge.

III. PROPOSED APPROACH

A. MONITORING FRAMEWORK

Considering three coal powder quality-related variables and mill power consumption, an integrated coal mill performance

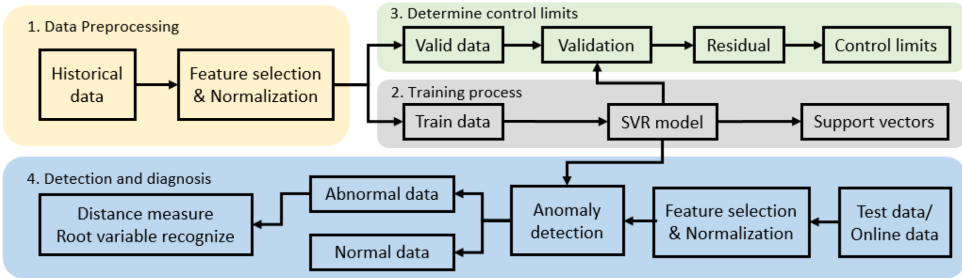


FIGURE 1. The flowchart of the monitoring and diagnosis framework.

monitoring framework based on SVR models is proposed. For each SVR model, the procedure of abnormal condition monitoring and diagnosis is displayed in Fig. 1. The SVR model is used to detect whether the corresponding performance indicator is out of the normal range. The normal range is set between the lower and upper control limits given by well-functioned historical operation data.

When the residuals between the model generated and real values fall out of the normal ranges, indicating the abnormal conditions in practice, a series of abnormal data are used for diagnosis. Then the accuracy of the detection approach is to be validated. Based on the differences between the normal and abnormal data, the top three variables with highest contributions to the deviations are located using the distance measures.

The model is built using SVR, and the kernel method is applied to map the samples into the feature space. The parameters of the models are automatically determined according to the complexity of the data distribution and the grid search strategy, to choose proper number of support vectors and guarantee the models' accuracies and generalization capabilities. The control limits are selected according to the Pauta criterion and the kernel density estimation (KDE). After estimating the performance related indicators, the support vectors obtained from the trained models are utilized to recognize the key variables that lead to performance degradation.

B. PERFORMANCE MONITORING MODELS

Different from the traditional regression methods that calculate the differences between the model outputs and the measured values to obtain the losses, SVR allows ϵ deviations between them.

Given the training samples $D = \{(x_1, y_1), \ldots, (x_m, y_m)\}$, y_i is a one-dimensional output while x_i can be a multi-dimensional vector. The SVR estimates the output $f(x)$ of an unknown input feature vector x. The optimization objective of the model can be written as:

$$f(x) = w^T x + b \tag{1}$$

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} l_\varepsilon(f(x_i) - y_i) \tag{2}$$

where w and b are undetermined parameters of the model, C is the regularization term/penalty coefficient, and m is the number of training samples. $l_\in$ is the ϵ-insensitive loss function:

$$l_\epsilon(z) = \begin{cases} 0, & \text{if } |z| \le \epsilon \\ |z| - \epsilon, & \text{otherwise} \end{cases} \tag{3}$$

The objective function can be solved by introducing the slack variables ζ_i and $\hat{\zeta}_i$ and Lagrange multipliers, then the coefficients a_i, $\hat{a}_i$ are determined, the solution to which is:

$$f(x) = \sum_{i=1}^{m} (\hat{a}_i - a_i) k(x, x_i) + b \tag{4}$$

Taking the form of feature mapping into consideration, the samples are mapped into the feature space, where $k\left(x_i, x_j\right) = \Phi(x_i)^T \Phi(x_j)$ is a kernel function and $\Phi(x)$ is the mapping function. The Gaussian kernel is chosen in this paper:

$$k(x_i, x_j) = exp(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}) \tag{5}$$

The performance of the SVR model with Gaussian kernel largely depends on the selection of its parameters. The penalty coefficient C determines the tolerance of the errors of the samples that fall out of the ϵ interval zone. Larger C will give less tolerance to the error. The support vectors are the training samples that fall out of the ϵ interval zone, making $\left(\hat{a}_i - a_i\right)$ unequal to zero. Thus the parameter ϵ, which determines the width of the interval zone, controls the number of support vectors, which determines the generalization capability of the model. Larger ϵ stands for better generalization capability. Additionally, the parameter σ in the kernel function also affects the model performance. The model with larger σ has weaker nonlinearity and less accuracy. In this paper, therefore, we generalize the method proposed by Yi *et al.* [31] to multi-dimensional inputs to select the appropriate interval zone ϵ, and adopt the grid parameter optimization method to find the best C and σ simultaneously with a given width of the interval zone ϵ. The main idea of the generalized method is based on the fact that the number of the support vectors, which is needed to fully describe the model, is determined by the complexity of the distribution of the training data. Thus, an evaluation index, which is denoted by *CP*, has been put forward to measure the complexity of the data distribution. The original work [31] considered a univariate case and split the data into several regions to evaluate the split data separately. In this paper, the data are considered as a whole and the evaluation index is generalized to the multivariate case.

Given the training sample set $(x_{i,1}, x_{i,2}, \ldots, x_{i,n}, y_i)_{i=1}^{m}$, where n is the dimension of each training sample, the complexity index CP is defined as:

$$CP = \frac{\sum_{i=1}^{m-1} |A_i|/(m-1)}{\sum_{i=1}^{m-2} |cos\theta_i|/(m-2)} \tag{6}$$

$$\begin{aligned} A_i &= (x_{i+1,1}, x_{i+1,2}, \ldots, x_{i+1,n}, y_{i+1}) \\ &\quad - (x_{i,1}, x_{i,2}, \ldots, x_{i,n}, y_i) \end{aligned} \tag{7}$$

$$\cos\theta_i = \frac{|A_i \bullet A_{i+1}|}{|A_i| \bullet |A_{i+1}|} \tag{8}$$

where $|A_i|$ is the Euclidean distance between the *(i+1)-th* sample and the *i-th* sample, θ_i is the angle between the vectors A_i and A_{i+1}. Thus, $\sum |A_i|$ indicates the density of samples, and $\sum |cos\theta_i|$ indicates the samples' variation tendency. Smaller CP implies a less stringent interval zone could be chosen, and fewer support vectors are need for fitting the model. Then ϵ can be modified by the empirical formula as

$$\epsilon = \frac{\max(std^2(x_j)_{j\in 1,2,\ldots,n})}{CP} \tag{9}$$

where *std(x)* calculates the standard deviation of x.

C. CONTROL LIMITS

Due to the variation of operating conditions, the fluctuation of sensor noises and the capabilities of the performance monitoring models, there must be deviations between the real measures and the model-generated values. Thus, the detection is based on the statistical analysis of the residuals.

In this paper, we consider the Pauta criterion and the kernel density estimation as the candidate schemes for determining control limits, and give a comparison between these two schemes. The Pauta criterion assumes that the data to be analyzed only contain random errors. It uses the mean value μ and the standard deviation σ to construct a zone, out of which the data are regarded as abnormal. The probability that the residuals distribute in the range of $(\mu - 3\sigma, \mu + 3\sigma)$ is 0.9973, which means that the normal data are out of the range just with the probability less than 0.3%. This criterion requires that the residuals should follow or approximately follow the normal distribution; otherwise the criterion may be invalid.

For the data without prior distribution information, the KDE is a well-established approach to estimate the probability density function (PDF) for univariate processes [32]. Assume that $\{e_1, e_2, \ldots, e_m\}$ are the independent and identically distributed variables that follow a distribution with the PDF $p(e)$. Then $p(e)$'s kernel estimation function is defined by:

$$\hat{p}(e) = \frac{1}{mh}\sum_{i=1}^{m} K(\frac{e-e_i}{h}) \tag{10}$$

where $K(\cdot)$ is the kernel function, h is the bandwidth. The optimal estimation of the bandwidth was given by Silverman [33] with an empirical formula:

$$h_{opt} = 1.06\sigma m^{-0.2} \tag{11}$$

If the Gaussian kernel is used, the PDF can be rewritten as:

$$\hat{p}(e) = \frac{1}{\sqrt{2}mh}\sum_{i=1}^{m} \exp(-\frac{(e-e_i)^2}{2h}) \tag{12}$$

Then the corresponding upper and lower control limits, $UCL(\alpha)$ and $LCL(\alpha)$, can be obtained from $\hat{p}\,(e)$ by solving the following equations [32]

$$\int_{-\infty}^{UCL(\alpha)} \hat{p}(e)de = \alpha \tag{13}$$

$$\int_{LCL(\alpha)}^{+\infty} \hat{p}(e)de = \alpha \tag{14}$$

where α is the confidential bound.

Similarly, these two methods both use the control limits to detect when residuals exceed the upper or lower bound. In the industrial processes, the anomaly detection algorithms should not only recognize faults accurately, but also avoid reporting faults while under normal circumstances. Thus, the false alarms rate and hit alarms rate are calculated to evaluate the performance of the monitoring scheme.

$$FPR = \frac{number\ of\ false\ alarms}{number\ of\ normal\ data} = \frac{FP}{FP+TN} \tag{15}$$

$$TPR = \frac{number\ of\ hit\ alarms}{number\ of\ abnormal\ data} = \frac{TP}{TP+FN} \tag{16}$$

where FP is the number of the cases that the model claims abnormal when there is no fault occurs, FN is the number of the cases that the model claims normal when faults occur. TP and TN represent the number of the cases that the model can provide right estimation under the abnormal and normal conditions, respectively.

Lower false alarm rate and higher hit alarm rate, which are generally known as recall ratio, show better performance of the monitoring scheme.

D. ROOT CAUSE DIAGNOSE METHOD

After the abnormal condition is detected, the root cause diagnosis method is utilized to find the key variables that contribute most to the performance degradation and clarify the deviations of the variables. The support vectors obtained from the trained model are considered as the representative operating conditions. For each input sample, the Euclidean distance is used to evaluate the similarity between the sample and the support vectors. Thus, by traversing all support vectors, k nearest neighbors in the support vector set of the sample can be found for comparison.

TABLE 1. Common and specialized variables for different models.

No.	Operating variables	Abbreviation	Unit
1	Gross load	GL	MW
2	Coal flow	CF	t/h
3	Motor current	MC	A
4	Motor power	MP	kW
5	Feeder motor speed	FMS	r/min
6	Hot air damper driver position	HADDP	%
7	Cold air damper driver position	CADDP	%
8	Primary air pressure	PAP	kPa
9	Primary air temperature	PAT	DEG C
10	Primary air flow	PAF	t/h
11	Pressure of air/powder mixture	PAPM	kPa
12	Pressure between seal air and grinding bowl	PSAGB	kPa
13	Outlet temperature	OT	DEG C
14	Pressure difference of grinding bowl	PDGB	kPa
15	Separator speed	SS	r/min
16	Separator current	SC	A
17	Power consumption	PC	kWh/t

Given the distance between the sample x and the support vector v:

$$d(x, v) = \sqrt{(x - v)(x - v)^T} \tag{17}$$

The contribution of the$q$-th operating variable to the abnormal condition can be defined as:

$$cont_q = \frac{1}{u}\sum_{i=1}^{u}(x_{i,q} - \frac{\sum_{t=1}^{k} v_{t,q}}{k}) \tag{18}$$

where u is the number of the samples that are estimated to be abnormal by the SVR model, $v_{\mathrm{t},q}$ is the q-th variable in the t-th nearest support vector.

The contribution magnitude is assessed by the absolute value, and the sign is given for judging the deviating direction in a statistical perspective because all input samples that are detected to be abnormal are taken into consideration. The diagnosis process can also be conducted online, which means when the continuous alarms of anomaly accumulate to a certain amount, the abnormal observations of the previous time from the moment will be used for analysis.

IV. CASE STUDY

The proposed approach is applied to the real cases obtained from the coal mills in a large power plant. There are 17 operating variables in total for building the performance monitoring models, and different variables are chosen for different models. All the used variables are listed in Table 1.

We consider first 12 operating variables listed in the Table 1 as common variables since they affect all indicators. Because all the variables have impacts on energy consumption, the power consumption model takes all the variables as inputs. However, this cannot be applied for other three models. As for 'outlet temperature' and 'pressure difference of grinding bowl', their influences are limited to their corresponding quality indicators, so they are removed from other quality-related models. They are the predictors of the temperature model and the ventilation pressure model, respectively. According to the operating principle of the coal mill, the speed of the separator not only reflects coal powder fineness, but also influences the ventilation pressure, so it will be considered as the target of the coal powder fineness model and the input of the ventilation pressure model as well.

A. CASE I: SEPARATOR ANOMALY

The first case is the separator anomaly of a coal mill. At the date of this case, the operator observed a sudden drop of the separator current at 2:15 a.m. From 3:30 a.m. to 6:00 a.m., the separator was stopped twice; and the speed was adjusted slowly in accordance with the current. After the temporal maintenance, the coal mill ran well until 8:27 a.m. Then the senior engineers came to check the causes of the anomaly, concluded that there was something wrong in the reducer of the separator. Therefore, the coal flow and the separator speed were lowered down temporarily. Thus, we choose the data from 2:00 a.m. to 8:30 a.m., except for those in the period from 3:30 a.m. to 6:00 a.m., during which the separator was shut down, as inputs samples. Well-functioned data in five months just before this anomaly case are selected as the modeling data, split into the training and validating sets with the proportion of 7:3.

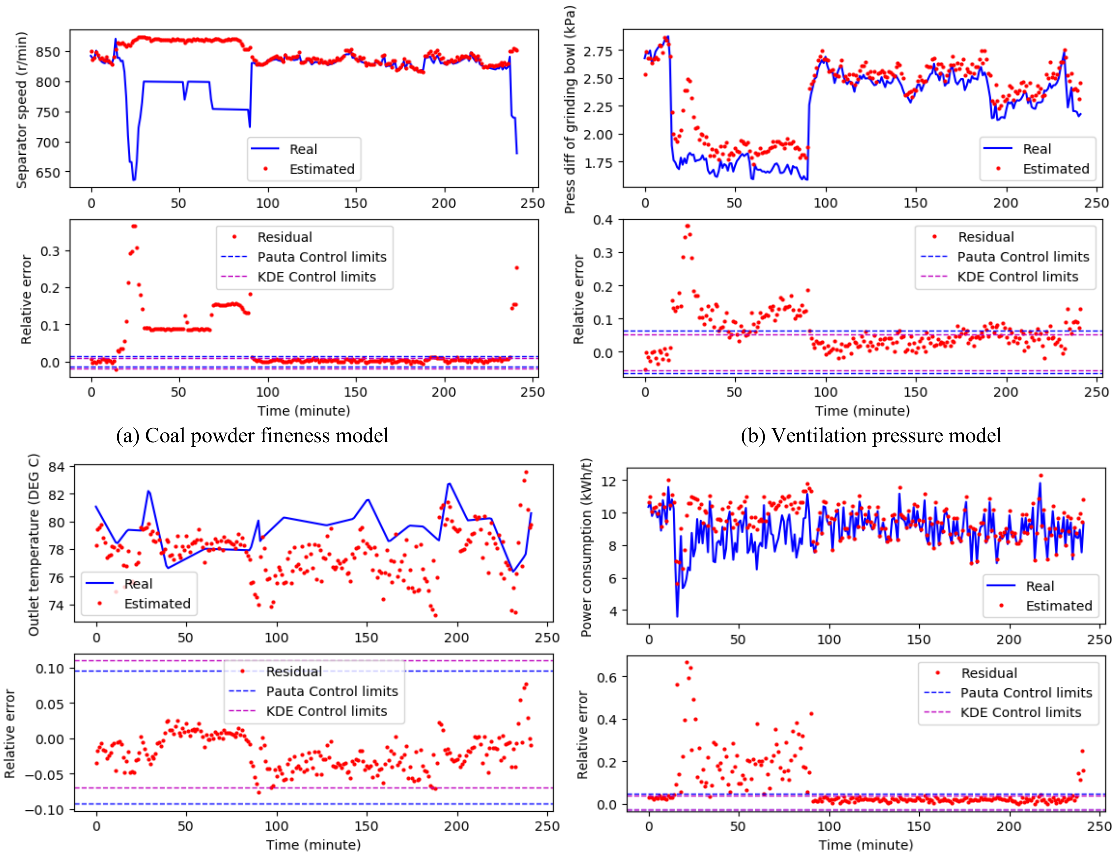
Taking the power consumption model for example, the parameter ϵ is selected as 0.047 concerning the complexity of data distribution. The parameters C and σ are set as 2.0 and 0.5 given by the grid searching strategy. Here both the Pauta criterion and the KDE are used to determine the control limits, with the confidential bound α set as 99%.

From the Fig. 2, the following facts can be observed. During the monitoring period, the residuals between the model-estimated and real values of the separator speed, pressure difference of grinding bowl and power consumption go beyond the control limits for some time, which indicates the occurrence of abnormal conditions. The temperature values of all samples are almost in the normal range.

The abnormal condition is defined by the event that the residuals exceed the control limits. If there are three consecutive estimated indicators falling out of the control limits, we consider the instant corresponding to the first of the three abnormal indicators as the anomaly occurring time. The monitoring results visualized in Fig. 2 show that the abnormal condition happened from the 15th sample to the 91st sample, corresponding to the period from 2:15 a.m. to 3:30 a.m. The abnormal condition occurred again at 8:27 a.m., which is consistent with the operational log time recorded by the operator.

As three models report anomalies based on the input samples, the effect is then compared by calculating the rates of FPR and TPR for each corresponding indicator, shown in Table 2. In general, the TPRs of KDE are higher than those of Pauta criterion but it also misidentifies more normal conditions as abnormal. For the best case, the monitoring scheme can achieve the TPR of 98.8% with a low FPR of 0.6%.

After detecting the abnormal condition, the key variables contribute to the change of operating condition are diagnosed.



(a) Coal powder fineness model

(b) Ventilation pressure model

(c) Temperature model

(d) Power consumption model

FIGURE 2. Plots of three coal powder quality-related models and the power consumption model. Upper: the real and estimated values of performance-related variables; Lower: the relative error of these two values.

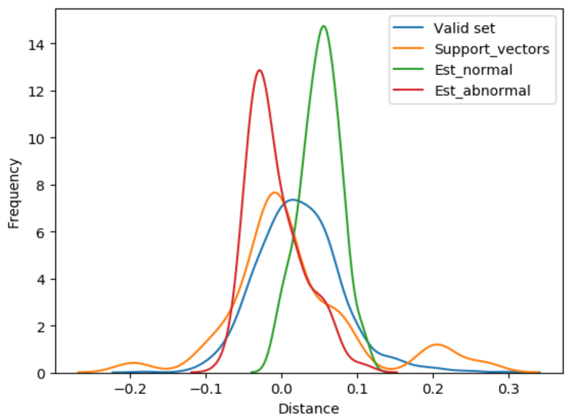
TABLE 2. Comparison between Pauta criterion and KDE estimation for three models.

Model	Method	Upper bound	Lower bound	TPR	FPR	MSE
Coal powder fineness	Pauta	0.014	-0.014	0.988	0.006	3.855
	KDE	0.011	-0.017	1.000	0.050	3.855
Ventilation pressure	Pauta	0.065	-0.064	0.840	0.106	0.059
	KDE	0.053	-0.055	0.938	0.211	0.059
Power consumption	Pauta	0.046	-0.029	0.975	0	0.143
	KDE	0.039	-0.027	0.975	0.019	0.143

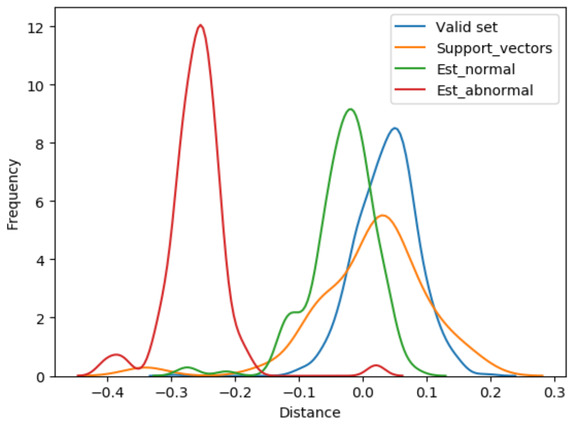
Because the values used in SVR are normalized, the distances between different variables can be compared directly. Here we choose the power consumption model for example. Four data sets are defined, i.e., the validating set extracted from the original modeling data, the support vectors obtained from the trained model, the input samples estimated to be normal, and the input samples estimated to be abnormal. For all the dataset, aiming to the selected variable, the distributions of the distances between the samples and their ten nearest support vectors are displayed together in Fig. 3, from which a statistical conclusion can be drawn that when the abnormal condition occurred, the cause-related variables could be diagnosed through the change of the distances away from their nearest support vectors. In Fig. 4, there are not great differences between the cold air damper driver positions and these four datasets. However, the distribution of the distances between the pressure difference of grinding bowl and the data that are estimated to be abnormal appears to be different from those in the other three cases.

Thus, by comparing the change of the distances between the samples and their ten nearest support vectors, for the ventilation pressure model, we plot Fig. 4 in which all variables' deviations are displayed. The rank of top three key variables and their deviating directions are given in Table 3. The separator speed contributes much more to the ventilation pressure anomaly than other variables.

Similarly, for the coal powder fineness model and the power consumption model which detect anomaly in the monitoring period in this case, we also conduct diagnosis to find the cause-related variables. According to the results of the



(a) Variable: Cold air damper driver position



(b) Variable: Pressure difference of grinding bowl

FIGURE 3. Distributions of the variable distances of input samples in four different datasets from their ten nearest support vectors.

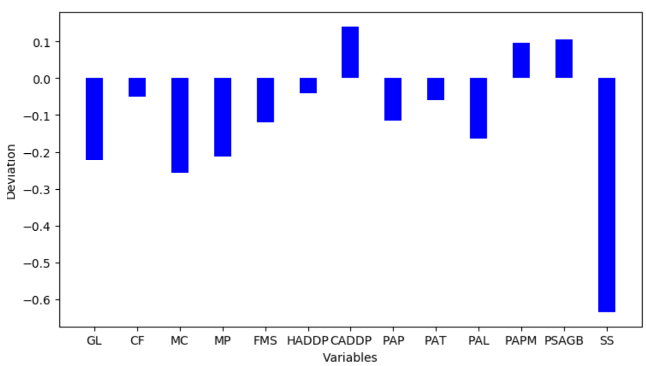


FIGURE 4. The mean distance of each operating variable in the ventilation pressure model between the abnormal input samples and their support vector neighbors.

coal powder fineness model in Table 4, the separator current ranks first, followed by the motor current and the motor power.

Table 5 demonstrates the results from the power consumption model. Among the top three variables, there are two coal powder quality-related variables, which is consistent with detection results in Fig. 2.

Therefore, combining three diagnosis results from three different models, it can be noted that the fluctuation of the pressure difference of grinding bowl was triggered by the variation of the separator speed, and the pressure difference's deviation is less than that of the separator speed. The decrease of the separator speed mainly came from the declined separator current, which is diagnosed as the root cause of the abnormal condition. It is in line with the experts' judgment that the fault was in the separator reducer. Furthermore, this approach is capable of pointing out the variables related to the anomaly explicitly. Besides, although the energy consumption goes down, the quality of the coal powder deteriorates due to the decrease of other two coal powder quality-related variables.

TABLE 3. Top 3 high related variables of the ventilation pressure model in Case I and their deviating directions.

Ranking	Variable	Deviating direction
1	Separator speed	Lower
2	Motor current	Lower
3	Gross load	Lower

TABLE 4. Top 3 high related variables of the coal powder fineness model in Case I and their deviating directions.

Ranking	Variable	Deviating direction
1	Separator current	Lower
2	Motor current	Lower
3	Motor power	Lower

TABLE 5. Top 3 high related variables of the power consumption model in Case I and their deviating directions.

Ranking	Variable	Deviating direction	Quality-related
1	Separator speed	Lower	Yes
2	Pressure difference of grinding bowl	Lower	Yes
3	Motor current	Lower	No

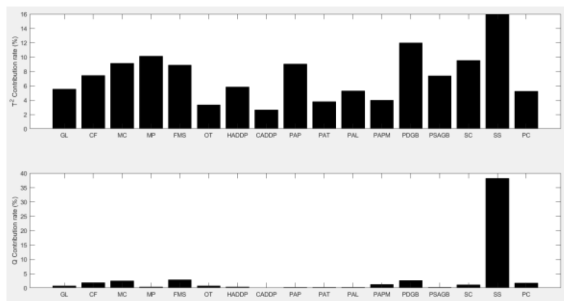
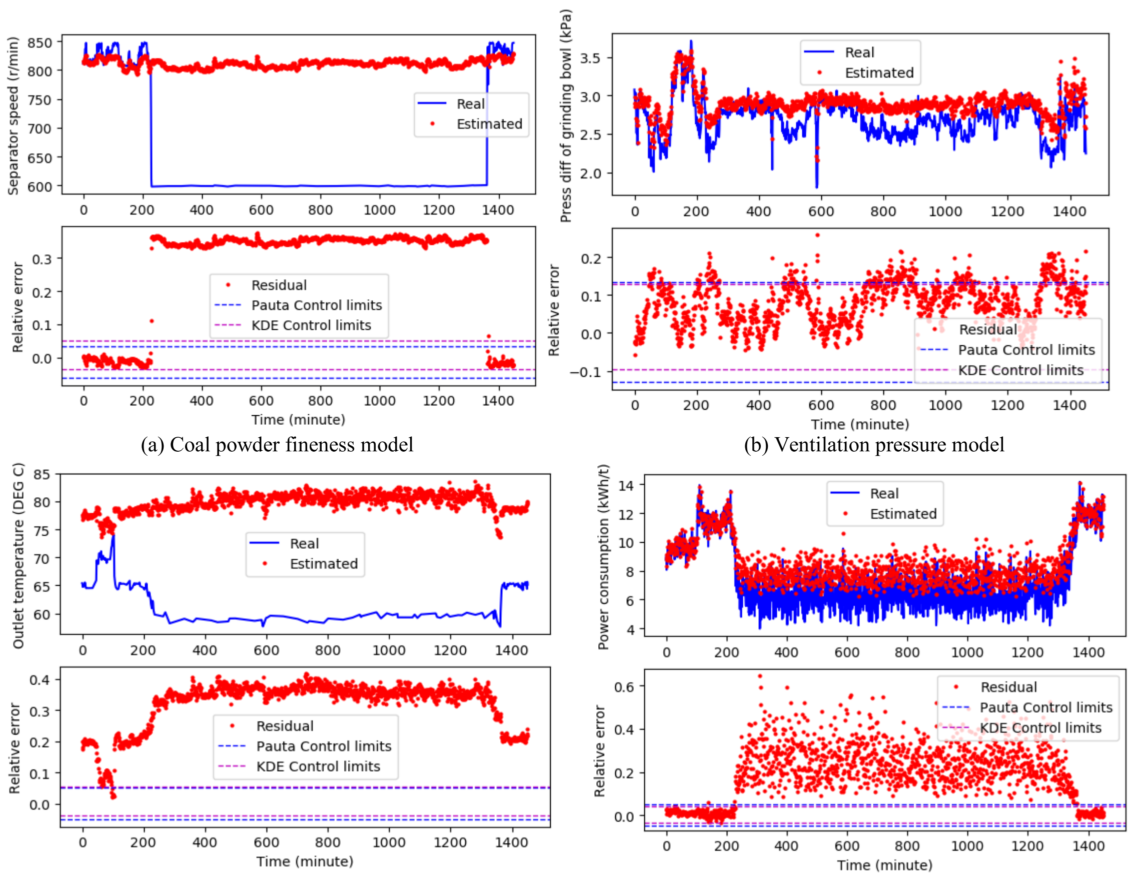


FIGURE 5. The diagnosis results given by PCA.

To validate the detection performance of our method, some different data-driven approaches are used for comparison in Table 6. Both PCA with 95% confidence's statistics and XGBoost with EWMA can achieve quite good results, but they are still inferior to the best monitoring scheme given by our approach (TPR of 98.8% with a low FPR of 0.6%).

As for the diagnosis, since only PCA [34], [35] in the above-mentioned methods can give the contribution of each variable when an anomaly occurs, we calculate the average



(a) Coal powder fineness model
(b) Ventilation pressure model
(c) Temperature model
(d) Power consumption model

FIGURE 6. Plots of three coal powder quality-related models and the power consumption model. Upper: the real and estimated values of performance-related variables; Lower: the relative error of these two values.

TABLE 6. The comparison of different detection approaches.

Method	Control Limits	TPR	FPR
PCA	T^2 with 99% confidence	0.863	0
PCA	T^2 with 95% confidence	0.950	0.025
PCA	Q with 99% confidence	1.000	0.617
KNN	Pauta criterion	0.705	0.085
BPNN	Pauta criterion	0.949	0.055
Random Forest	EWMA	0.987	0.134
XGBoost	EWMA	0.987	0.085

contribution assigned to each variable of the input sample whose statistic exceeds the upper limit. The top 3 variables with the highest contributions turn out to be the separator speed, pressure difference of grinding bowl and the separator current.

The two methods based on PCA and SVR both identify the variables with large deviation, but the PCA based method lacks the deviation information about these variables, which demonstrates the superior of our proposed approach.

According to the results from the proposed performance monitoring approach, it can be concluded that the anomaly in the separator caused its current to be lower than the expected level, which in turn influenced the separator speed, the pressure difference of grinding bowl and the power consumption. In this case, the anomaly of the separator reducer led to a decline of the separator speed, thus more coal blocks without sufficient grinding could pass through the separator, decreasing the quality of reflux. Therefore, the pressure difference of grinding bowl also decreased. The deviation directions of the responsible variables given by our approach agree with the real anomaly causes.

B. CASE II: CO-FIRING

The second case is a co-firing case that the coal with superior quality was mixed with the coal with high moisture and high volatile matter, from April 11th, 2017 to April 18th, 2017. In this case, it is to be checked whether the proposed approach can detect and diagnose the abnormal performances.

The detection and diagnosis procedures are the same as those in the first case. Well-functioned data are used for building the four performance models for the monitoring purpose. Taking the power consumption model for example, the parameter ϵ is set as 0.037. The parameters C and σ are set as 4.0 and 0.25 respectively.

According to the monitoring results given in Fig. 6, all four models report anomaly. It can be seen that the residuals of the temperature model exceed the both of the upper control limits

TABLE 7. Top 3 high related variables of temperature model in Case 2 and their deviating directions.

Ranking	Variable	Deviating direction
1	Primary air flow	Higher
2	Motor current	Lower
3	Motor power	Lower

TABLE 8. Top 3 high related variables of ventilation pressure model in Case 2 and their deviating directions.

Ranking	Variable	Deviating direction
1	Primary air flow	Higher
2	Separator speed	Lower
3	Motor current	Lower

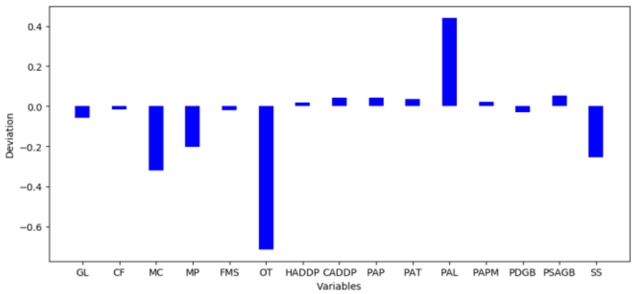


FIGURE 7. The mean distance of each operating variable in the power consumption model between abnormal input samples and their support vector neighbors.

TABLE 9. Top 3 high related variables of power consumption in Case 2 and their deviating directions.

Ranking	Variable	Deviating direction	Quality-related
1	Outlet temperature	Lower	Yes
2	Primary air flow	Higher	No
3	Motor current	Lower	No

of the Pauta and KDE methods from the first time, while the variation of the separator speed and the power consumption begins at the 229th and the 230th samples, corresponding to the real time at 0:10 a.m. and 0:20 a.m. on April 11th, respectively. For the ventilation pressure, it fluctuates and goes beyond the control limits for some time.

As for the separator speed, there is a sharp decrease as it was manually adjusted for proper coal powder fineness, so we mainly focus on the other three models in this case study.

The diagnosis results of the abnormal conditions about the outlet temperature and the pressure difference of grinding bowl are shown in Table 7 and Table 8, where the primary air flow both ranks first and the values are higher than those in the normal conditions.

For the power consumption model, the deviation plot and the top 3 high related variables are given in Fig.7 and Table 9.

From the mechanism point of view, the increase of the primary air flow will result in the decrease of power consumption. Furthermore, both the outlet temperature and the separator speed will also affect the power consumption. During the period in this case, although the coal mill consumed less energy, the decline of the outlet temperature and the fluctuation of the ventilation pressure deteriorated the quality of the coal powder.

TABLE 10. The comparison of different detection approaches.

Method	Control Limits	TPR	FPR
PCA	T^2 with 99% confidence	1.0	0.635
PCA	T^2 with 95% confidence	1.0	1.0
PCA	Q with 99% confidence	0.274	0
KNN	Pauta criterion	0.983	0.029
BPNN	Pauta criterion	1.0	0.196
Random Forest	Pauta criterion	1.0	0.217
XGBoost	EWMA	1.0	0.204

The operator in the plant can locate the root causes of the case with the help of the diagnosis results. As the primary air flow represents the drying capability of the pulverizer, while it is diagnosed as the root cause of the abnormal condition, the reason for the anomaly lies in the increase of the moisture content in the raw coal. The operation logs also proved that the low-grade coal was mixed in during this period.

C. CASE III: FLUCTUATION OF THE OUTLET TEMPERATURE

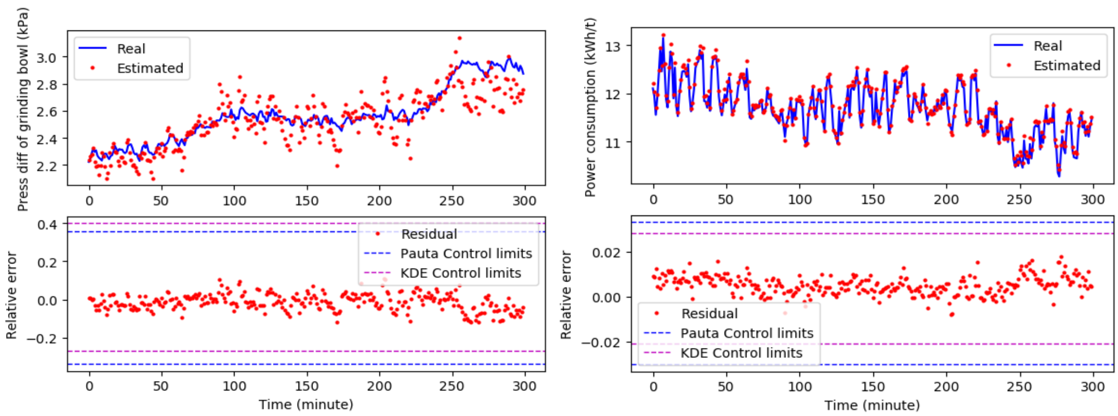
The third case is the fluctuation of outlet temperature happened during a certain period. The data came from a coal mill with the structure different from those in the other two cases. Because the coal powder fineness is simply adjusted using deflector baffle, we use the monitoring variables to build other three models here. As can be seen from the Figure 8 and Figure 9, the ventilation pressure and the powder consumption remained normal while the outlet temperature went beyond the upper control limit at the 235th sample of Pauta criterion and at the 242th sample of KDE, indicating a successful detection of the anomaly in the outlet temperature.

The proposed approach is compared with the other approaches listed in Table 10 which can obtain a high TPR but are with relatively more false alarms mostly. The best one, KNN, is with the TPR of 98.3% and the FPR of 2.9%. However, the modified SVR approach with Pauta criterion as control limits has the TPR of 98.3% and the FPR of 1.7%, and with KDE as control limits has the TPR of 96.7% and no false alarm.

As for the diagnosis results in Table 11, it has been shown that the cold air damper driver position was higher but the primary air temperature was lower. The primary air was mixed from the hot and the cold air, and the seal air also came from the cold air, so the deviations indicated that there might be malfunction in the control loop. The conclusions from the experts also proved that the other parts of the coal mill worked well. Although this case appears to be similar to the Case II since they are both related to outlet temperature anomaly, the diagnosis results provide different explanations, which also validates the effectiveness of our approach.

D. CASE STUDY SUMMARY

Three practical cases from actual thermal power plants are demonstrated in this section. The separator anomaly case



(a) Ventilation pressure model (b) Power consumption model

FIGURE 8. Plots of ventilation pressure model and the power consumption model. Upper: the real and estimated values of performance-related variables; Lower: the relative error of these two values.

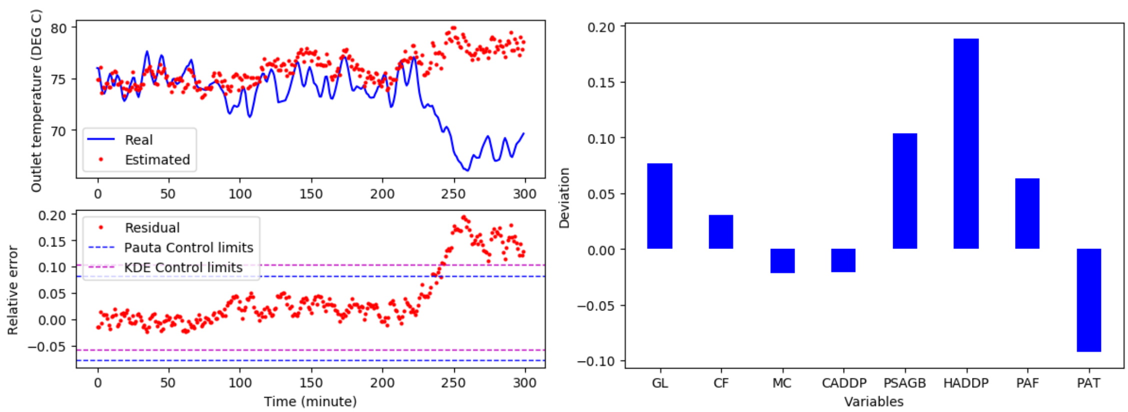


FIGURE 9. Detection and diagnosis results of the temperature model.

TABLE 11. Top 3 high related variables of temperature model in Case 3 and their deviating directions.

Ranking	Variable	Deviating direction
1	Cold air damper driver position	Higher
2	Pressure between seal air and grinding bowl	Higher
3	Primary air temperature	Lower

and the fluctuation of the outlet temperature case that were well recorded are used to conduct the detection accuracy evaluation. The results validate that the modified SVR detection approach achieve better performance than some other detection approaches. As for the diagnosis, in all cases, the diagnosis algorithm by locating the key variables through calculating their deviations is also verified and consistent with the experts' analysis conclusions, which can provide useful anomaly analysis results to operators.

V. CONCLUSION

This paper proposes an abnormal condition detection and diagnosis framework for coal mills in thermal power plants based on support vector regression. The variables of the coal mills are categorized as the operating variables and performance-related variables. The detection scheme includes the SVR model with the parameter selection method, the grid searching strategy, and the control limits computing algorithms based on the Pauta criterion and KDE. Distance based measures are utilized to rank the variables responsible for the abnormal conditions, and both the direction and the magnitude of the deviation are evaluated.

The proposed framework can detect the deterioration of coal powder qualities and the anomaly in the coal mill power consumption in time. Furthermore, the abnormal data are collected for the root causes related variables identification. The diagnosis scheme can also give the deviation directions and magnitudes of the responsible variables, which is useful for the operators in the power plants. Furthermore, both the detection and diagnosis methods are data-driven, and no prior information of the abnormal conditions is required before executing the analysis procedure.

Considering the need of optimizing the operating conditions, the future work may include operating variables optimization by dynamically determining the expected performances under different operating conditions, thus providing adjustment suggestions for the controllable variables.

REFERENCES

[1] C. Maffezzoni, "Concepts, practice and trends in fossil-fired power plant control," in *Proc. IFAC Power Syst. Power Plant Control Conf.*, Beijing, China, 1986, pp. 1–9.

[2] J. L. Wei, J. Wang, and Q. H. Wu, "Development of a multisegment coal mill model using an evolutionary computation technique," *IEEE Trans. Energy Convers.*, vol. 22, no. 3, pp. 718–727, Sep. 2007.

[3] P. F. Odgaard, B. Lin, and S. B. Jorgensen, "Observer and data-driven-model-based fault detection in power plant coal mills," *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 659–668, Jun. 2008.

[4] S. Guo, J. Wang, J. Wei, and P. Zachariades, "A new model-based approach for power plant Tube-ball mill condition monitoring and fault detection," *Energy Convers. Manage.*, vol. 80, pp. 10–19, Apr. 2014.

[5] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "Intelligent decision support system for detection and root cause analysis of faults in coal mills," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 934–944, Aug. 2017.

[6] A. Ž. Rakić, "Early fault detection and isolation in coal mills based on self-organizing maps," in *Proc. 10th Symp. Neural Netw. Appl. Elect. Eng.*, Sep. 2010, pp. 45–48.

[7] W. Qin, W. Yan, and J. Xu, "Application of fault diagnosis expert system in grinding process," in *Proc. IEEE Int. Conf. Autom. Logistics*, Aug. 2010, pp. 290–295.

[8] X. J. Han, J. Xue, "Fault diagnosis of pulverizing system based on fuzzy decision-making fusion method," in *Fuzzy Information and Engineering*, vol. 2. Berlin, Germany: Springer, 2009, pp. 1045–1056.

[9] Z.-G. Su, P.-H. Wang, X.-J. Yu, and Z.-Z. Lv, "Experimental investigation of vibration signal of an industrial tubular ball mill: Monitoring and diagnosing," *Minerals Eng.*, vol. 21, no. 10, pp. 699–710, 2008.

[10] S. Collura, D. Possanzini, M. Gualerci, L. Bonelli, and D. Pestonesi, "Coal mill performances optimization through non-invasive online coal fineness monitoring," in *Proc. Power Gen*, Wien, Jun. 2013, pp. 4–6.

[11] E. Kisić, V. Petrović, S. Vujnović, Z. Đurovic, and M. Ivezić, "Analysis of the condition of coal grinding mills in thermal power plants based on the T^2 multivariate control chart applied on acoustic measurements," *Autom. Control Robot.*, vol. 11, no. 2, pp. 141–151, 2012.

[12] X. Feng and Y. Wang, "Fault estimation based on sliding mode observer for Takagi–Sugeno fuzzy systems with digital communication constraints," *J. Franklin Inst.*, to be published, doi: 10.1016/j.jfranklin.2019.09.020.

[13] Z. Zhou, M. Zhong, and Y. Wang, "Fault diagnosis observer and fault-tolerant control design for unmanned surface vehicles in network environments," *IEEE Access*, to be published, doi: 10.1109/ACCESS.2019.2954352.

[14] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "Review of control and fault diagnosis methods applied to coal mills," *J. Process Control*, vol. 32, pp. 138–153, Aug. 2015.

[15] E. Elbouchikhi, V. Choqueuse, Y. Amirat, M. El Hachemi Benbouzid, and S. Turri, "An efficient Hilbert–Huang transform-based bearing faults detection in induction machines," *IEEE Trans. Energy Convers.*, vol. 32, no. 2, pp. 401–413, Jun. 2017.

[16] S. Liu, Y. Sun, and L. Zhang, "A novel fault diagnosis method based on noise-assisted MEMD and functional neural fuzzy network for rolling element bearings," *IEEE Access*, vol. 6, pp. 27048–27068, 2018.

[17] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050–17066, 2019.

[18] Y. Yang, P. Fu, and Y. He, "Bearing fault automatic classification based on deep learning," *IEEE Access*, vol. 6, pp. 71540–71554, 2018.

[19] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, S. Yu, and R. Ji, "Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network," *Appl. Energy*, vol. 114, pp. 1–9, Feb. 2014.

[20] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2019.2931491.

[21] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018.

[22] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.

[23] N. Li, J. Tuo, Y. Wang, and M. Wang, "Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning," *Neurocomputing*, to be published, doi: 10.1016/j.neucom.2019.10.003.

[24] D. Yu, Z. M. Chen, K. S. Xiahou, M. S. Li, T. Y. Ji, and Q. H. Wu, "A radically data-driven method for fault detection and diagnosis in wind turbines," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 577–584, Jul. 2018.

[25] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 327–346, 2003.

[26] F. Harrou, Y. Sun, and S. Khadraoui, "Amalgamation of anomaly-detection indices for enhanced process monitoring," *J. Loss Prevention Process Ind.*, vol. 40, pp. 365–377, Mar. 2016.

[27] J. Zeng, U. Kruger, J. Geluk, X. Wang, and L. Xie, "Detecting abnormal situations using the Kullback–Leibler divergence," *Automatica*, vol. 50, no. 11, pp. 2777–2786, 2014.

[28] M.-H. Chang, C. Chen, D. Das, and M. Pecht, "Anomaly detection of light-emitting diodes using the similarity-based metric test," *IEEE Trans. Ind. Informat.*, vol. 10, no. 3, pp. 1852–1863, Aug. 2014.

[29] Y. Yan, J. Li, and D. Gao, "Condition parameter modeling for anomaly detection in wind turbines," *Energies*, vol. 7, no. 5, pp. 3104–3120, 2014.

[30] R.-P. Nikula, M. Ruusunen, and K. Leiviskä, "Data-driven framework for boiler performance monitoring," *Appl. Energy*, vol. 183, pp. 1374–1388, Dec. 2016.

[31] H. Yi, X.-F. Song, B. Jiang, Y.-F. Liu, and Z.-H. Zhou, "Flexible support vector regression and its application to fault detection," *Acta Automatica Sinica*, vol. 39, no. 3, pp. 272–284, 2013.

[32] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford, U.K.: Clarendon Press, 1997.

[33] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986, pp. 30–367.

[34] J. Xuan, Z. Xu, and Y. Sun, "Selecting the number of principal components on the basis of performance optimization of fault detection and identification," *Ind. Eng. Chem. Res*, vol. 54, no. 12, pp. 3145–3153, 2015.

[35] J. Xuan, Z. Xu, and Y. Sun, "Incipient sensor fault diagnosis based on average residual-difference reconstruction contribution plot," *Ind. Eng. Chem. Res.*, vol. 53, no. 18, pp. 7706–7713, 2012.

XINGYUN HONG received the B.S. degree in automation from Zhejiang University, Hangzhou, China, in 2017, where she is currently pursuing the M.S. degree in control science and engineering. Her current research interest includes performance monitoring and fault detection and diagnosis for thermal power plant equipment based on machine learning.

ZHENGGUO XU received the B.S. degree in electrical engineering from Changsha Electric Power University, Changsha, China, in 2002, and the M.S. and Ph.D. degrees in control science and engineering from Tsinghua University, Beijing, China, in 2009. He is currently an Associate Professor with the College of Control Science and Engineering, Zhejiang University, China. His research interests include data analysis and artificial intelligence applications in energy systems, fault diagnosis and prediction, and reliability analysis.

ZHENWEI ZHANG received the B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2008. His research interests include condition monitoring and fault diagnosis for steam turbines in power plants.