MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Ibrahim *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 22 (2026) 321–329

RESEARCH ARTICLE

Ash Fouling Prediction of Individual Heating Surface of Boiler for Ultra-Supercritical Coal Power Plant based on Regression Techniques

Maslina Mohd Ibrahim[a,*], Azura Che Soh[a], Asnor Juraiza Ishak[a], Raja Kamil[a], Mohd Amran Mohd Radzi[a], Amir Redzuan Mohd Ibrahim[b]

[a]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Selangor, Malaysia; [b]Sultan Azlan Shah Power Station, Tenaga National Berhad, Seri Manjung, Perak, Malaysia

Abstract The ash fouling inside the boiler has a detrimental effect on its performance, leading to suboptimal efficiency. Soot blower systems are commonly used during power plant operations to mitigate fouling. Most soot blower operations are scheduled and fixed without considering the actual degree of fouling inside the boiler, often resulting in either insufficient or excessive blowing. The former reduces heat transfer efficiency, while the latter leads to the wastage of high-pressure steam and shortens the lifespan of boiler pipes. This study aims to predict the fouling conditions of six individual heating surfaces within the boiler: primary, secondary, and final superheaters; primary and final reheaters; and the economizer, utilizing indirect and data-driven methods. Direct methods, such as sensor installation, are impractical due to the extreme conditions within the boiler. The study initially establishes the relative cleanliness level as an indicator of fouling degree by comparing current heat absorption values with reference values derived from statistical analysis during periods of stable boiler conditions. Data cleaning methods are then applied before employing regression techniques for ash fouling prediction. Gaussian Process Regression (GPR), a nonparametric kernel-based probabilistic model, and Support Vector Machine (SVM) with different kernels are experimented with for comparison. Experimental analysis demonstrates high accuracy, ranging from 91.4% to 98.2% for GPR and 89.1% to 98.1% for SVM on the case study data. The implementation of the prediction model in this study is expected to enhance soot blowing operations, ultimately optimizing boiler performance. This improvement will lead to higher energy efficiency and a reduction in detrimental emissions.

Keywords: Ash Fouling, Regression, Gaussian Process Regression, Support Vector Machine.

Introduction

The combustion of coal in power plants results in the formation of ash deposits, which can cause slagging and fouling on the heating surfaces inside the boiler. This leads to reduced heat transfer efficiency, increased fuel consumption, higher flue gas temperatures, and a decrease in the overall thermal efficiency of the boiler. Additionally, it can result in higher carbon dioxide (CO_2) emissions, incomplete combustion with an increase in carbon monoxide (CO) and nitrogen oxides (NOx) emissions, and corrosion of the boiler's internal components. To mitigate ash fouling inside the boiler, several strategies have been implemented, such as the operation of soot blowers and cleaning systems, where high-pressure steam is periodically used to remove ash deposits. Other methods include optimizing coal combustion control, improving boiler design, and implementing advanced monitoring and predictive maintenance systems.

Ash fouling prediction studies play a crucial role in facilitating the operation of soot blowers and advanced monitoring systems. Predictive models enable plant operators to take preventive or corrective measures in handling ash fouling conditions within the boiler. One of the biggest challenges for operators is determining when and where ash build-up will occur in specific sections of the boiler. The soot-blowing process can be costly, as it uses steam that would otherwise be used for electricity generation. Therefore,

***For correspondence:**
maslina79@yahoo.com

Received: 3 August 2025
Accepted: 11 March 2026

by using ash fouling monitoring and prediction systems, operators can optimize blower intervals and reduce unnecessary soot blowing, leading to significant energy and maintenance cost savings. Moreover, this approach can help extend the lifespan of boiler components.

Recent studies have increasingly focused on indirect methods for predicting slagging and fouling conditions [1] – [9]. Several studies have developed monitoring methodologies using artificial neural networks and other machine learning techniques to evaluate fouling conditions in boiler heating surfaces [1] – [9]. These approaches estimate fouling conditions using different indicators such as system heat loss variations [1], the ratio between actual and theoretical heat transfer coefficients expressed as a cleanliness factor (CF) [2]–[5], [7], [8], dual extended Kalman filtering techniques [6], and thermal resistance network models [9]. Machine learning regression models have demonstrated strong capability in handling complex nonlinear relationships in power plant systems because they can learn directly from operational data without relying solely on predefined physical equations [3], [6], [9]–[11]. Recent studies have further explored machine learning approaches for boiler performance monitoring and fouling prediction. For example, Blackburn *et al*. [12] developed a dynamic machine learning-based optimization framework for coal-fired power plants, while Zeng *et al*. [13] applied deep learning techniques using a multi-modal database for boiler fouling prediction. Yang *et al*. [14] also proposed a machine learning-based dynamic prediction model for evaluating the pollution status of heating surfaces in coal-fired boilers. Despite these developments, fouling prediction based on cleanliness indicators remains sensitive to variations in operating conditions, particularly when using cleanliness factor (CF) [2], [3], [6].

Considering the complexity of fouling inside boilers and the unique design and data availability for each power plant, this research introduces a novel approach to the cleanliness factor, termed the relative cleanliness level (CL). This approach employs a statistical method, using the boiler's optimal condition that is based on the outlet flue gas temperature as an overall assessment of boiler efficiency [15], [16]as a reference value. It is also more dynamic and reliable, considering the gradual deterioration of the boiler's clean state over time. Once the baseline is established, CL can be measured and predicted using raw data from the like Distributed Control System (DCS). However, this data may be compromised due to sensor malfunctions or issues with data acquisition systems. Therefore, data preprocessing is necessary to enhance model quality. Subsequently, model regressors are selected from a set of candidates. Gaussian Process Regression (GPR) and Support Vector Machines (SVM) with different covariance functions were chosen and tested.

Case Study: Coal-Fired Power Plant

This research case study considers a 1000 MW ultra-supercritical (USC) coal-fired boiler at one of the power plant stations in Malaysia. The boiler is a one-through type with a condensate polishing plant, featuring a steam generator capacity of 2989 t/s, a main steam temperature of 600°C, and a main steam pressure of 250 bar. It comprises two main sections: the furnace and the convection section. The convection section includes six individual heat surface areas: secondary superheater (SSH), final superheater (FSH), final reheater (FRH), primary reheater (PRH), primary superheater (PSH), and economizer (ECON). The furnace section (wall) of the boiler is treated as a constant parameter in this study. The general overview of the boiler is presented in Figure 1 below.

The current approach at the power plant for managing ash fouling within the boiler predominantly relies on operator expertise for both combustion control and soot blowing operations. The soot blowing process follows a predetermined schedule and operates on a routine basis. However, adjustments to the schedule are made periodically in response to anomalies, such as elevated CO levels, increased flue gas temperatures, or malfunctioning components like induced fans. For this study, the optimal parameter selection was determined based on the Piping and Instrumentation Diagrams (P&IDs) of the ultra-supercritical boiler at the coal-fired power plant and consultations with the plant operators. The analysis focused on variables such as the combustion state of the boiler, operational conditions, and steam parameters for each specific target surface.

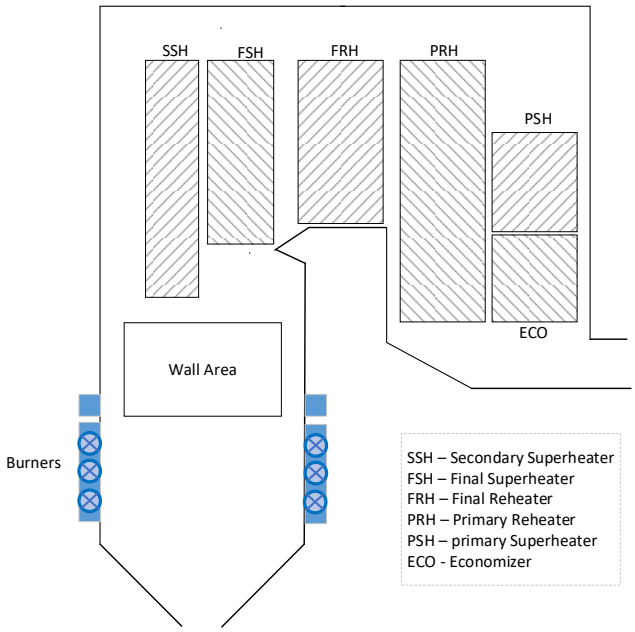


Figure 1. The structure of an ultra-supercritical boiler at a coal-fired power plant

Data Collection and Data Pre-Processing

Data collection was carried out during the power plant's regular and stable operations, focusing on net loads ranging from 740 MW to 1000 MW, with measurements recorded at 5-minute intervals. Over 100 data points from various sensors in the DCS were gathered between 2018 and 2022. This research specifically examines data from January 1, 2022, to January 31, 2022, encompassing more than 12,000 data points for each heat surface. To prepare the data for analysis, we first performed data preprocessing to eliminate outliers caused by measurement errors or anomalies. We employed a moving median due to its robustness and adaptability to non-stationary data. The window size k was determined through experimentation, with k=24 selected as the optimal value for this study.

Statistical Analysis for Q_{ref} Parameters

To establish the target parameter as a basis for CL calculation, a comprehensive statistical analysis was undertaken to identify the optimal period for boiler performance by monitoring the outlet flue gas exit temperature (FGET). The increase in FGET serves as a crucial indicator of combustion efficiency and pollutant formation within the boiler. Therefore, a rise in FGET suggests an escalation in slagging and fouling within the boiler [15] - [18]. The plotted trend of FGET, illustrated in Figure 2, indicates that the period from October 2017 to April 2018 represents the optimal boiler condition. Consequently, this period was selected for analysis to determine the reference value of heat absorption, thereby calculating the relative cleanliness factor as previously discussed.

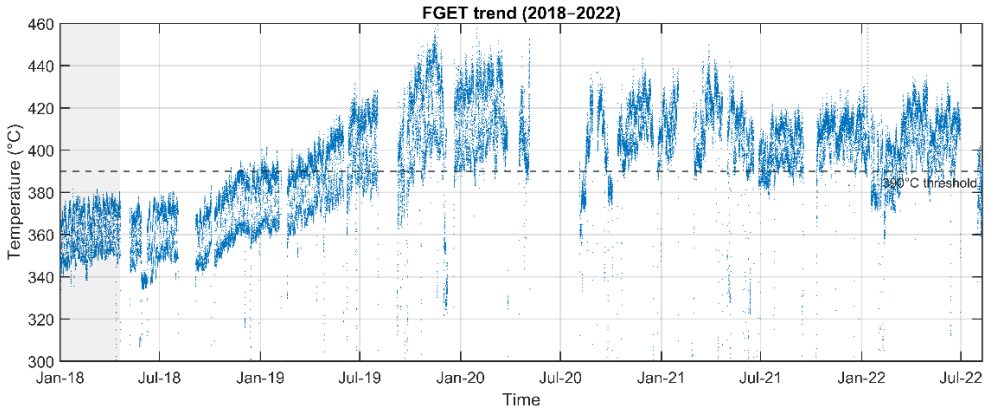


Figure 2. Flue gas exit temperature (FGET) data trend from 2017 to 2022

The statistical analysis involved precise data sampling at 1-minute intervals and categorizing the data into three load categories: 800MW, 900MW, and 1000MW. The mean and standard deviation of these datasets were employed as indicators of the most stable condition. Table 1 and Figure 3 present the outcomes of this analysis on each section heats absorption and the deviation of data respectively to the load of the power plant. From the analysis, the heat absorption data in 2017 are more spread out and less consistent compared to 2018, where the variation is significantly smaller, as clearly shown in Figure 3. Therefore, the period from January to April 2018 was identified as the reference due to its demonstrated stability, consistency, and predictability. The reduced standard deviations across each heat surface indicate fewer fluctuations and less variability, providing a solid foundation for use as a reference value.

Table 1. Mean and Standard Deviation of Heat Absorption Data Classified by Load Categories (800MW, 900MW, 1000MW): (a) Data from October to December 2017, (b) Data from January to April 2018

	800 MW		900 MW		1000 MW	
	Mean	Std-Dev	Mean	Std-Dev	Mean	Std-Dev
ECON	46.63	5.52	61.67	6.41	80.01	8.65
FRH	71.39	5.54	84.41	5.89	94.85	6.78
FSH	87.27	13.16	90.62	5.82	99.57	5.67
PRH	110.26	4.75	123.87	5.22	136.85	5.84
PSH	63.25	26.53	82.63	23.89	104.84	12.69
SSH	98.83	32.30	101.86	13.48	108.24	13.48

(a)

	800 MW		900 MW		1000 MW	
	Mean	Std-Dev	Mean	Std-Dev	Mean	Std-Dev
ECON	52.09	4.86	70.06	6.85	89.41	6.52
FRH	77.69	4.59	91.28	5.39	102.83	4.64
FSH	80.25	6.25	90.20	7.48	97.89	8.04
PRH	103.88	4.32	117.67	5.44	130.29	3.87
PSH	65.96	6.44	89.52	9.47	114.83	9.10
SSH	75.45	8.07	82.27	8.52	87.86	9.35

(b)

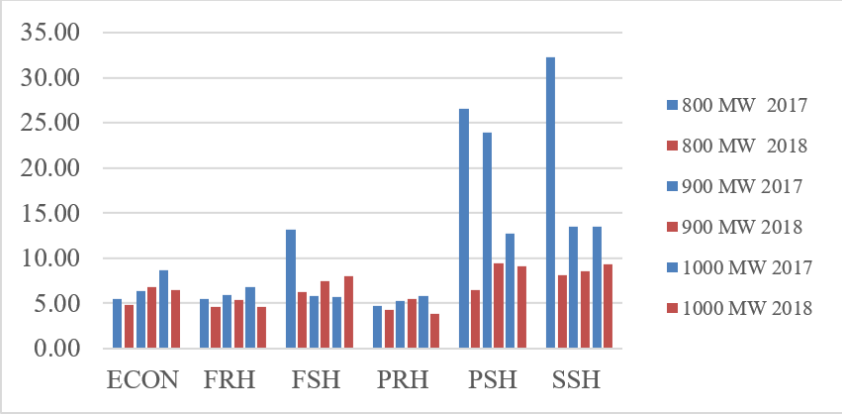


Figure 3. Comparison of Standard Deviation for All Individual Heat Surfaces Between Data from 2017 and 2018

The Thermal Calculation of a Convection Heat Transfer Surfaces

Heat absorption in each section of a boiler refers to the amount of thermal energy transferred to the working fluid in that section. In an ultra-supercritical (USC) boiler, where operating pressures and temperatures exceed the critical point of water, controlling heat absorption is crucial for optimizing performance and efficiency. Balancing and optimizing heat absorption across different components is essential for maximizing efficiency and ensuring safe operation. A discussion on the implementation of

heat absorption for each section is provided in Section 3. In this research, the heat absorption for each section of the boiler is calculated by applying the First Law of Thermodynamics and fundamental principles as shown in equation (1) below:

$$Q = m\,(H_2 - H_1) \tag{1}$$

where Q represents the heat absorption of the working fluid, m represents the mass flow rate of the working fluid (kg/s), and H_1 and H_2 denote the inlet and outlet enthalpy of the working fluid, respectively (kJ/kg). Enthalpy is a thermodynamic property defined as the sum of internal energy and the product of pressure and volume.

In this study, subscripts are used to represent the corresponding boiler heating surface sections, including the economizer (ECON), primary superheater (PSH), secondary superheater (SSH), final superheater (FSH), primary reheater (PRH), and final reheater (FRH). The variables m_e and m_{ms} represent the economizer steam flow and main steam flow, respectively. Additional variables include the spray steam flows for the primary superheater (m_{s-psh}), secondary superheater (m_{s-ssh}), and reheater (m_{s-rh}), as well as the soot blower steam flow (m_{sbr}). The enthalpy terms H_{1-*} and H_{2-*} represent the inlet and outlet enthalpy of the working fluid for each corresponding heating surface section.

The heat absorption Q of all individual heating surfaces is calculated using Equations (2)–(7).

$$Q_{ECON} = m_e(H_{2-ECON} - H_{1-ECON}) \tag{2}$$

$$Q_{PSH} = [m_e - (m_{s-psh} + m_{s-ssh})](H_{2-PSH} - H_{1-PSH}) \tag{3}$$

$$Q_{SSH} = \left[m_{ms} - \left(\frac{\%\, m_{s-ssh}}{\%m_{s-ssh}+\%m_{s-psh}} * m_{s-ssh}\right)\right](H_{2-SSH} - H_{1-SSH}) \tag{4}$$

$$Q_{FSH} = m_{ms}(H_{2-FSH} - H_{1-FSH}) \tag{5}$$

$$Q_{PRH} = [m_{ms} + m_{s-rh} - m_{sbr}](H_{2-PRH} - H_{1-PRH}) \tag{6}$$

$$Q_{FRH} = [m_{ms} + m_{s-rh} - m_{sbr}](H_{2-FRH} - H_{1-FRH}) \tag{7}$$

Determination of Cleanliness Level

The relative cleanliness level or CL is defined as the ratio between the current heat absorption (Q_y) and a reference heat absorption (Q_{ref}), as depicted below (8).

$$Cleanliness\ Level, CL = \frac{Q_y}{Q_{ref}} \tag{8}$$

The Q_{ref} is determined by using a statistical analysis and a point-to-point linear regression technique to ensure that CL remains independent of the power plant's load and adapts according to the current load of the power plant. The data spanning from the boiler's best and ideal condition is taken into consideration and the analysis specifically targets the conditional mean of heat absorption within the load range of 740 MW to 1000 MW. The determination of each load point is based on the frequency of occurrences, which correlates to the stability of the power plant. The mean of each $Q_{filtered}$ is computed using the following equation (9):

$$\bar{Q}_i = \frac{\sum_{i=1}^{n} Q_{filtered}}{n_{filtered}} \tag{9}$$

In deriving Q_{ref}, Q_i is depicted as the y-axis value and formulated with load as the x-axis.

$$Q_{ref} = \beta_o + \beta_i x_i \tag{10}$$

where Q_{ref} is the dependent variable at the i-th observation, β_o is the intercept, β_i is the slope of the line and x is the load of the power plant.

Prediction of CL using Regression Techniques

Regression analysis is a robust technique for predicting a dependent variable (Y) using one or more independent variables (X). Accurate model development requires careful variable selection, model estimation, and validation. Prediction accuracy is evaluated using metrics like R-squared, RMSE, and MAE, with validation performed by splitting the data into training (80%) and testing (20%) sets. This study explores Gaussian Process Regression (GPR) and Support Vector Machine (SVM) regression, utilizing various kernels to identify the most accurate model.

Gaussian Process Regression (GPR)

Gaussian Process Regression (GPR) models are probabilistic, nonparametric models that use kernel functions. In a Bayesian nonparametric context, GPR extends Gaussian distributions to functions, ensuring that any finite collection of random variables within a Gaussian Process (GP) has a joint Gaussian distribution [19]. A GP is specified by its mean function and covariance function, and instead of representing a single function, it defines a distribution over functions. This allows GPR to model a diverse range of potential relationships within the data, offering greater flexibility in capturing underlying patterns. This relationship is expressed by the following equation:

$$f(x) \sim GP\big(m(x), k(x, x')\big) \tag{11}$$

Where $m(x)$ is the mean function and $k(x, x')$ is the covariance function that defines the relationship between data points. The covariance function of the latent variables reflects the smoothness of the response, and the basic functions map the inputs x into a p-dimensional feature space. In supervised learning, it is expected that data points with similar predictor values xi will exhibit closely related response (target) values y_i. In the context of Gaussian processes, this similarity is represented by the covariance function, which defines the covariance between the latent variables $f(xi_i)$ and $f\big(x_j\big)$ with x_i and $x_j$being *d×1* vectors. The covariance implemented for the training data are as listed in equation (12) to (15) below where r is Euclidean distance between x_i and x_j, σ_l is the characteristic length scale, σ_f is the signal standard deviation and α is a positive-valued scale-mixture parameter.

i. Exponential covariance function:

$$k\big(x_i, x_j|\theta\big) = \sigma_f^2 \exp\left(-\frac{r}{\sigma_l}\right) \tag{12}$$

ii. Squared Exponential covariance function:

$$k\big(x_i, x_j|\theta\big) = \sigma_f^2 \exp\left[-\frac{1}{2}\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma_l^2}\right] \tag{13}$$

iii. Matern 5/2 covariance function:

$$k\big(x_i, x_j\big) = \sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)\exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right) \tag{14}$$

iv. Rational Quadratic covariance function:

$$k\big(x_i, x_j|\theta\big) = \sigma_f^2\left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha} \tag{15}$$

Support Vector Machine (SVM)

Support Vector Machine (SVM) analysis is a widely recognized machine learning approach for both classification and regression applications. SVM regression is deemed a nonparametric method due to its reliance on kernel functions. The SVM achieves high performance due to the careful selection of the kernel and the fine-tuning of its parameter settings. A kernel is a function that satisfies specific conditions for any x and y as (16).

$$K(x, y) = \{\emptyset(x), \emptyset(y)\} \tag{16}$$

$\emptyset$ is a mapping from X to F_k that corresponds to an inner product feature space connected with the kernel K.

$$\emptyset: x \in X \rightarrow \emptyset(x) = K(x, .) \in F_k \tag{17}$$

In this study, three kernels have been used which are Quadratic, Cubic and Medium Gaussian due to their suitability for modelling non-linear data, ranging from moderate to complex patterns. Quadratic and Cubic kernel equations are listed in equations (18) and (19). While Medium Gaussian implies that sqrt (the number of predictors) as the kernel scale results in medium distinctions.

$$K(x,y) = (x^t * y + c)^2 \quad (18)$$

$$K(x,y) = (x^t * y + 1)^3 \quad (19)$$

Results and Discussion

In this study, the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were employed as performance indicators for evaluating the prediction models. The R^2 metric indicates the strength of the correlation between observed and predicted values, where values closer to 1 represent stronger correlations. RMSE and MAE measure prediction errors, with values closer to 0 indicating better prediction accuracy. Table 2 presents the performance comparison of the GPR and SVM models with different kernels for the primary superheater (PSH). Similar analyses were conducted for the remaining heating surfaces, including ECON, SSH, FSH, PRH, and FRH.

The results indicate that the GPR model with an exponential kernel provides the best predictive performance. Table 2 presents the results for the primary superheater (PSH), where the highest R^2 value of 0.9822 was achieved during training. Similar analyses were conducted for the remaining heating surfaces. Across all six heating surfaces evaluated in this study, the GPR models produced R^2 values ranging from 0.9144 to 0.9822 during training. The lowest R^2 value was observed in the final reheater (FRH), while the highest value was obtained for the PSH. The testing results are consistent with the training results, demonstrating that the model generalizes well and has strong predictive power. This consistency suggests that the models are well-balanced and do not suffer from overfitting or underfitting.

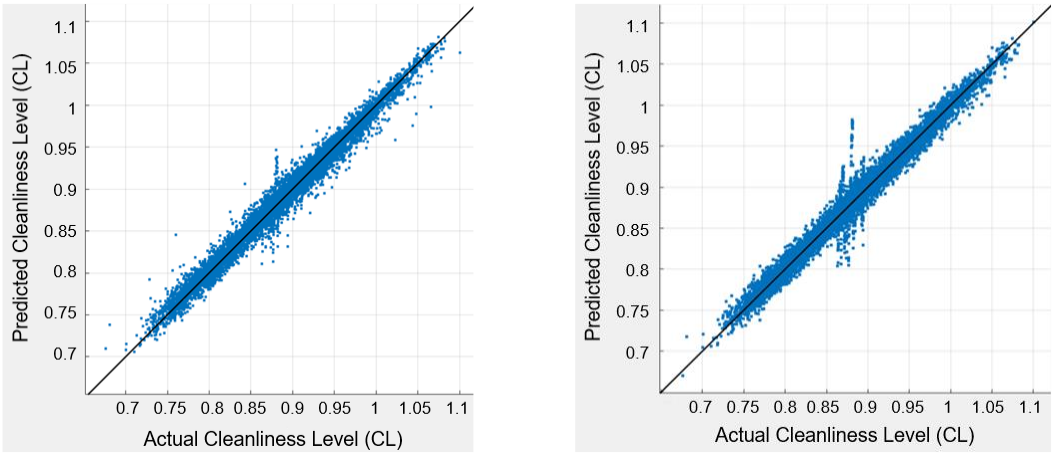
For the SVM model, the best performance was observed with the cubic kernel. The highest R^2 value of 0.9811 was obtained for the PSH, while the lowest value of 0.8576 was observed for the FRH among all heating surfaces evaluated. Overall, the SVM models with cubic and quadratic kernels showed similar performance; however, the quadratic kernel had the lowest MAE value of 0.01032 among the SVM models, indicating better accuracy for individual points compared to the other two kernels. The differences in the results suggest that the data complexity ranges from moderate to high, and the median Gaussian kernel leads to overfitting and unnecessary complexity, as it requires careful tuning of its hyperparameters.

However, overall, the GPR model outperforms the SVM model due to its superior flexibility, better generalization, and ability to handle complex data relationships more effectively, particularly because of its capacity to quantify uncertainty. The absence of uncertainty quantification in SVM reduces its reliability, especially when dealing with noisy data or outliers. Conversely, in this study, the cubic SVM also demonstrates good performance, though not as consistently as GPR.

Table 2. The performance comparison of different model for training and testing data for PSH

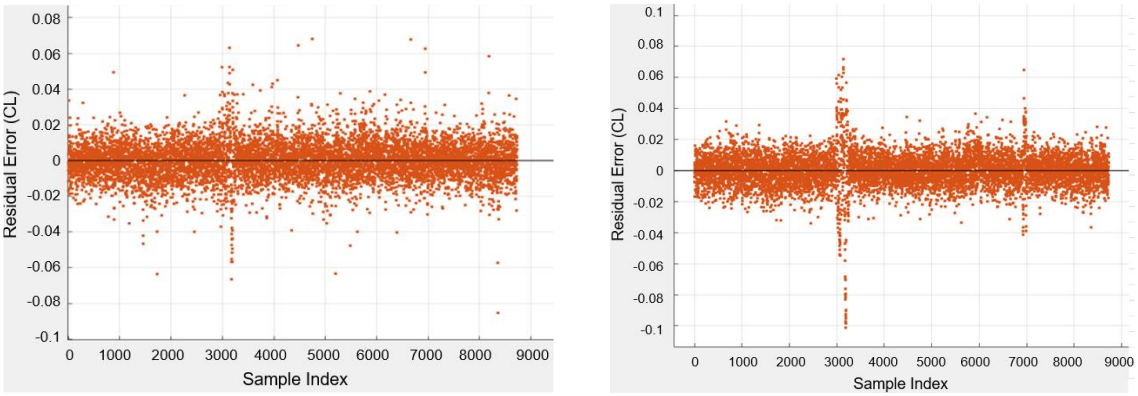
	Training			Testing		
Method	R2	RMSE	MAE	R2	RMSE	MAE
GPR - Exponential	0.9822	0.01413	0.01038	0.9833	0.01375	0.01008
GPR - Matern 5/2	0.9817	0.01430	0.01032	0.9827	0.01400	0.01008
GPR - Rotational Quadratic	0.9817	0.01430	0.01032	0.9828	0.01396	0.01007
GPR - Squared Exponential	0.9817	0.01431	0.01033	0.9826	0.01404	0.01012
SVM - Cubic	0.9811	0.01456	0.01050	0.9820	0.01426	0.01023
SVM - Median Gaussian	0.9724	0.01758	0.01226	0.9742	0.01711	0.01170
SVM - Quadratic	0.9813	0.01446	0.01033	0.9820	0.01426	0.01004

Figure 4 compares the predicted versus actual responses for both the GPR and SVM models. The GPR model shows greater consistency, with points closely clustered around the 45-degree line and fewer outliers compared to the SVM model. Figure 5 illustrates the residual error diagram, where GPR results indicate that most errors are near zero, reflecting accurate predictions for the majority of cases. In contrast, the SVM model displays more noticeable outliers, indicating less uniform prediction accuracy.



(a) (b)

Figure 4. Comparison of prediction and actual response for (a) GPR- Exponential model (b) SVM-Cubic model



(a) (b)

Figure 5. Residual error diagram of prediction results (a) GPR- Exponential model (b) SVM-Cubic model

Conclusions

This study introduced a relative cleanliness level (CL) indicator derived from the ratio between the current heat absorption and the reference heat absorption obtained from statistical analysis of optimal boiler operating conditions. The proposed approach provides a practical data-driven method for evaluating ash fouling conditions in individual heating surfaces of an ultra-supercritical coal-fired boiler. Regression models based on Gaussian Process Regression (GPR) and Support Vector Machine (SVM) were developed to predict the CL values using plant operational data. Among the evaluated models, the GPR model with an exponential covariance function demonstrated the best predictive performance, achieving the highest coefficient of determination and the lowest prediction errors. The results indicate that the proposed regression-based approach is capable of accurately predicting fouling behaviour and can support improved decision making for soot-blowing operations. However, several limitations should be acknowledged. The study is based on operational data from a single power plant unit; therefore, the model performance may vary when applied to boilers with different configurations or operating conditions. In addition, the prediction models rely solely on historical operational data and do not explicitly incorporate physical fouling mechanisms. Future research will focus on integrating the predicted cleanliness level into optimization algorithms for soot-blowing scheduling. The integration of advanced data-driven techniques with operational decision-support systems has the potential to further enhance boiler efficiency and reduce unnecessary steam consumption in coal-fired power plants.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This work is funded by the UPM Putra Grant (GP/2023/9751600: Soot Blowing Sequence Optimization and Prediction to Improve Boiler Heat Absorption for Ultra-Supercritical Coal Boiler). We also acknowledge the Sultan Azlan Shah Power Station, TNB, for providing data and consultation for this project.

References

[1] Xu, L., *et al*. (2021). Improvement of slagging monitoring and soot-blowing of waterwall in a 650 MWe coal-fired utility boiler. *Journal of the Energy Institute, 96*, 106–120. https://doi.org/10.1016/j.joei.2021.02.006.

[2] Shi, Y., Li, M., Cui, F., Wen, J., & Zeng, J. (2021). A neural-network-based method for ash fouling prediction of heat transfer surface in coal-fired power plant boiler. *IEEE Access, 9*, 109584–109604. https://doi.org/10.1109/ACCESS.2021.3100145.

[3] Cui, F., Qin, S., Zhang, J., Li, M., & Shi, Y. (2022). A hybrid method for prediction of ash fouling on heat transfer surfaces. *Energies, 15*(13). https://doi.org/10.3390/en15134658.

[4] Shi, Y., *et al.* (2018). Preventive soot blowing strategy based on state of health prediction for coal-fired power plant boiler.

[5] Shi, Y., *et al*. (2019). Soot blowing optimization for frequency in economizers to improve boiler performance in coal-fired power plant. *Energies, 12*(15). https://doi.org/10.3390/en12152901.

[6] Anitha Kumari, S., & Srinivasan, S. (2019). Ash fouling monitoring and soot-blow optimization for reheater in thermal power plant. *Applied Thermal Engineering, 149*, 62–72. https://doi.org/10.1016/j.applthermaleng.2018.12.031.

[7] Sundaram, T., Basim Ismail, F., Gunnasegaran, P., & Gurusingam, P. (2017). Development and implementation of intelligent soot blowing optimization system for TNB Janamanjung. In *MATEC Web of Conferences*. https://doi.org/10.1051/matecconf/201713101006.

[8] Pattanayak, L., Ayyagari, S. P. K., & Sahu, J. N. (2015). Optimization of sootblowing frequency to improve boiler performance and reduce combustion pollution. *Clean Technologies and Environmental Policy, 17*(7), 1897–1906. https://doi.org/10.1007/s10098-015-0906-0.

[9] Tong, S., Zhang, X., Tong, Z., Wu, Y., Tang, N., & Zhong, W. (2019). Online ash fouling prediction for boiler heating surfaces based on wavelet analysis and support vector regression. *Energies, 13*(1). https://doi.org/10.3390/en13010059.

[10] Band, S. S., *et al.* (2021). Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression. *Engineering Applications of Computational Fluid Mechanics, 15*(1), 1147–1158. https://doi.org/10.1080/19942060.2021.1944913.

[11] Zeng, A., Ho, H., & Yu, Y. (2020). Prediction of building electricity usage using Gaussian process regression. *Journal of Building Engineering, 28*. https://doi.org/10.1016/j.jobe.2019.101054.

[12] Blackburn, L. D., Tuttle, J. F., Andersson, K., Fry, A., & Powell, K. M. (2022). Development of novel dynamic machine learning-based optimization of a coal-fired power plant. *Computers & Chemical Engineering, 163*, 107848.

[13] Zeng, H., Hu, L., Zhang, Q., & Ye, W. (2025). Boiler fouling and slagging prediction based on multi-modal database and deep learning. *Applied Thermal Engineering*, 127878.

[14] Yang, K., Gao, L., Lin, Z., Lian, D., & Lin, Y. (2025). A machine learning based dynamic prediction model for pollution status of heating surface in coal-fired boilers. *Case Studies in Thermal Engineering*, 106953.

[15] Ebrahim, N., Ttlnae001, T., & Jestin, L. (2016). Investigation into methods for the calculation and measurement of pulverised coal boiler flue gas furnace exit temperature.

[16] Beér, J. M. (2000). Combustion technology developments in power generation in response to environmental challenges. Retrieved from http://www.elsevier.com/locate/pecs.

[17] Teler, J., Trojan, M., & Taler, D. (2009). Assessment of ash fouling and slagging in coal fired utility boilers. In *Proceedings of International Conference on Heat Exchanger Fouling and Cleaning VIII*.

[18] Zhang, X., Yuan, J., Chen, Z., Tian, Z., & Wang, J. (2018). A dynamic heat transfer model to estimate the flue gas temperature in the horizontal flue of the coal-fired utility boiler. *Applied Thermal Engineering, 135*, 368–378. https://doi.org/10.1016/j.applthermaleng.2018.02.067

[19] Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.