Data-driven Modeling and Working Condition Prediction in Process Industry Production

Meifang Zhang
College of Computer Science
Beijing University of Technology
Beijing, China
zhangmeifang@emails.bjut.edu.cn

Ziqi Wang
College of Computer Science
Beijing University of Technology
Beijing, China
ziqi_wang@emails.bjut.edu.cn

Jing Bi
College of Computer Science
Beijing University of Technology
Beijing, China
bijing@bjut.edu.cn

Haitao Yuan
School of Automation Science and Electrical Engineering
Beihang University
Beijing, China
yuan@buaa.edu.cn

2024 International Conference on Networking, Sensing and Control (ICNSC) | 979-8-3503-6522-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICNSC62968.2024.10759867

***Abstract*—In the production process of the process industry, precise adjustment of working conditions presents a challenge due to the complexity of processes and unknown disturbances. Central control operators need to adjust setpoints based on deviations in process parameters and monitor target values to maintain system stability. However, many operating procedures excessively rely on human experience, increasing the uncertainty of the production process. In addition, the expert knowledge is not fully embedded in accumulated operations, limiting its potential in decision support. Therefore, data-driven modeling of production processes is essential for developing industrial expert systems to realize intelligent manufacturing. This work proposes a work condition prediction framework based on an Operation Mode Library (OML) to realize Working Condition Prediction (WCP), called for OML-WCP short. Taking the cement rotary kiln adjustment process as an example, a stable OML is constructed using Gaussian mixture clustering technology. Experimental results with real-life operation data of a cement plant reveal that the prediction accuracy of OML-WCP outperforms the existing methods. Moreover, the continuous accumulation of operating mode libraries can improve prediction accuracy in practical applications.**

***Index Terms*—Process industry production, operating mode library, working condition prediction, process modeling.**

I. INTRODUCTION

Process industrial production is influenced by various factors and requires real-time adjustment of main operating parameters to maintain production efficiency [1]–[3]. The ever-changing working conditions, including stemming from equipment wear, variability in raw materials, and environmental fluctuation, necessitate continuous monitoring and fine-tuning by operators. The complex interplay among parameters, *e.g.*, temperature, pressure, and flow rate alterations can significantly impact product quality and throughput. Manual operation relies on subjective judgment and often lacks precision and consistency, making it challenging to recall and update.

This work was supported by the Beijing Natural Science Foundation under Grants L233005 and 4232049, the National Natural Science Foundation of China under Grants 62173013 and 62473014, and in part by Beihang World TOP University Cooperation Program.

To navigate these variable conditions and the intricate coupling of parameters, the process industry must undertake a profound analysis and forecast of production processes to enhance efficiency and quality. Traditional mechanism-based analyses and reliance on expert knowledge have their limitations. It is essential to achieve a harmonious balance between output, fuel, air volume, and kiln speed in cement production. However, these parameters fluctuate and require operator intervention to stabilize the system. The calcination process encompasses complex chemical reactions, complicating the accurate calculation of state parameters. In addition, the calcination state and material fuel composition are uneven, resulting in poor ventilation during feeding and unknown disturbances that affect the stability of systems [4]. These unknown disturbances must be inferred through subtle changes in system parameters, where multiple monitoring variables exhibit intricate relationships with controllable ones.

The prediction of working conditions in industrial processes poses several challenges. Various studies utilize statistical techniques to condense operating principles and predict situations, *e.g.*, the auto-regressive moving average model and its variants [5]. In addition, techniques that rely on multi-layer perceptions, *e.g.*, neural hierarchical interpolation [6], time-series dense encoder [7], and time-series mixer [8] achieve great prediction accuracy. Moreover, to discover abnormalities in the data and prevent equipment breakdowns, researchers employ machine learning methods, including statistical limitations [9], isolation forest [10], and one-class support vector machine [11], [12] to identify aberrant operating circumstances. However, these methods mainly depend on examining individual data points, which makes it challenging to make long-term time series predictions and accurately forecast fluctuation intervals.

Deep learning methods [13] are widely adopted for predicting working conditions. They depend on extensive amounts of past data to train models that predict future operational circumstances. Transformer-based models [14], [15] are widely adopted because of their outstanding performance. However, they have a significant memory demand and suffer from long training times. One of the main difficulties in using machine learning with working condition data is the absence of efficient annotation methods. This makes it harder to forecast changes in condition intervals without labeled data accurately and increases computational complexity. Therefore, developing a collection of operating patterns is an effective method for correctly forecasting working conditions in industrial processes. Wu *et al*. [16] condense arsenic salt cobalt removal working conditions into operating modes using extensive industrial operation data. They derive the initial operating mode and finalize the Operation Mode Library (OML) by analyzing substantial industrial operating data. Zhu *et al*. [17] analyze historical operating status data to identify critical indicators for status judgment. The authors create an OML by clustering the operating status. It can capture the subtle characteristics of changes in working conditions more accurately, resulting in a significant improvement in the accuracy and dependability of predictions. However, some limitations remain, especially in extracting working condition features. Future studies should prioritize critical areas, including optimizing the construction process of the operating mode library and enhancing the accuracy of working condition classification and identification.

Based on the above analysis, this work takes the operation status data of rotary kilns as the basis and constructs an OML containing stable modes of operation status. Moreover, the combination of operation parameters in different stable modes is studied, accumulating control process knowledge by recording its operation experts. Based on the expert control experience in OML, the recognition and prediction of real-time operating status are achieved. It can describe the stable mode of system operation using Gaussian mixture clustering technology. In the modeling process, data-driven methods are adopted while incorporating expert knowledge and experience. The interpretability of data information is considered in parameter selection and feature analysis to improve prediction accuracy. By discerning distinct operational patterns, this OML can furnish vital decision support for the reliable operation of control systems within the process industry.

II. Working Condition Prediction (WCP)

A. Key Parameters Analysis

Correlation analysis and principal component analysis can be employed to eliminate irrelevant and redundant attributes. To mitigate the impact of varying magnitudes among different variables, it is necessary to normalize the data. Several association rule mining algorithms are developed to handle datasets with varying magnitudes. Among these, the Apriori algorithm, FP-Tree algorithm, Eclat algorithm, and grey association analysis stand out as effective solutions [18]. The Apriori algorithm is known for its ability to efficiently discover frequent item sets and generate strong association rules, while the FP-Tree algorithm enhances the performance of the Apriori algorithm by compressing the dataset and reducing the number of database scans required. The Eclat algorithm, on the other hand, uses a different approach called vertical data format to mine frequent item sets, making it suitable for large databases with a high number of transactions. Grey association analysis is useful for handling uncertainty and incomplete information in our datasets.

Expert knowledge gathered from on-site surveys can be integrated with the sample data characteristics to optimize computing resources, thereby preventing the need for complex correlation analysis of numerous parameters. Experiments should discard rules that lack support from mechanistic rules. Once the son and parent sequences have been identified, Grey system theory can be employed to assess the level of correlation between parameters by analyzing the geometric similarity of the parameter change curves. It aligns more closely with experimental data. The grey correlation analysis is shown in Algorithm 1.

Algorithm 1 Grey Correlation Analysis

Input: Sample matrix ($X_{(m,n)}$), reference sequence ($Y_{(1,n)}$), smoothing factor (λ)
Output: Grey relational degrees

1: Calculate the range (r) for each column based on the reference sequence Y
2: **for** each row i in X **do**
3: Calculate the range (r') for each column
4: Calculate the correlation coefficient $c_{(i,j)}=(min\{r_j, r'_{i,j}\}/max\{r_j, r'_{i,j}\})+\lambda$ for column j.
5: Calculate the weight coefficient $w_{(i,j)}=1/((|r_j - r'_{i,j}|/max\{r_j, r'_{i,j}\})+\lambda)$
6: Calculate the Grey Relational Degree ρ_i for the row as the weighted average $\rho_i=(\Sigma w_{(i,j)} \times c_{(i,j)})/n$
7: Get each row's Grey Relational Degree values
8: **end for**
9: Sort the Degree of each row
10: Return the sorted Degree

B. Feature Exaction from Observed Values

This work initially conducts on-site inspections of the production site and incorporates the practical expertise of specialists. It reveals that operators primarily assess on-site working conditions by observing variable changes. By utilizing this information, the specific values of the observed quantities can be chosen to define and distinguish various operating conditions. Meanwhile, the curve characteristics of the observed quantities are indeed influenced by changes in operating conditions. This work does not consider periodic components because they cannot be distinguished from the random and dynamic nature of the changes in operating conditions. The moving average method is adopted to examine values' fluctuating patterns, trajectories, and magnitudes. Specifically, a moving average (MA) is employed to depict the magnitude of the variable. In addition, the statistical indicator coefficient of variation (C.V.) is introduced to evaluate the volatility of variables. It aims to keep the observation volume consistent, which can be mathematically expressed as a preference for lower volatility. Therefore, it is feasible to accurately evaluate operational modifications by monitoring volatility fluctuations. It is obtained as:

$$C_i{=}\frac{S_i}{A_i} \tag{1}$$

where C_i is the ith C.V., S_i denotes the moving standard deviation (SD), and A_i denotes the MA, *i.e.*,

$$A_i{=}\frac{1}{t}\sum_{j=i-t+1}^{i} y_j \tag{2}$$

where t is the window size, and y_j is the data point j in the primary data set. In addition, the moving standard deviation S_i is obtained as:

$$S_i{=}\sqrt{\frac{1}{t}\sum_{j=i-k+1}^{i}\left(y_j-\bar{y}\right)^2} \tag{3}$$

where $\bar{y}$ is the average value of the data within the window.

C. Clustering by Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is adopted to cluster the features of the working conditions. It considers various distribution patterns in the dataset, an unknown covariance matrix, and high-dimensional data.

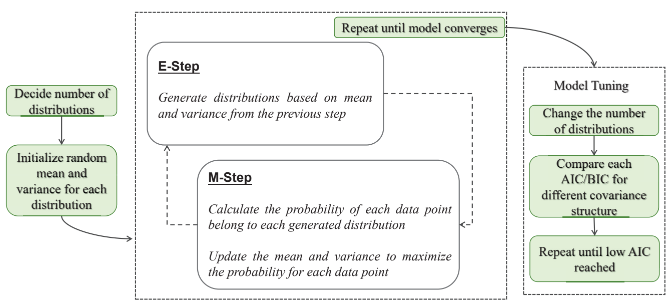


Fig. 1. GMM Clustering and Optimization

GMM and the estimated associated parameters are considered issues involving incomplete data. Therefore, the Expectation Maximization (EM) technique [19] is a viable solution. It is employed to estimate the latent variables and parameters. This process is iterative and may be loosely separated into two steps, including the E-step and the M-step.

In the E-step, the posterior probability is revised. For each data point, the possibility that it belongs to cluster (k) is computed as:

$$h_k^{(j)}(i){=}\frac{\pi_k^{(j)}\mathcal{N}\left(x_i \mid \boldsymbol{\mu}_k^{(j)}, \boldsymbol{\Sigma}_k^{(j)}\right)}{\sum_{k=1}^{K}\pi_k^{(j)}\mathcal{N}\left(x_i \mid \boldsymbol{\mu}_k^{(j)}, \boldsymbol{\Sigma}_k^{(j)}\right)} \tag{4}$$

where k is the count of clusters in the Gaussian distribution, π_k is the mixing coefficient of the Gaussian distribution, which is initialized in the previous stage. The whole process is shown in Fig. 1. $\mathcal{N}\left(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ denotes the probability density function of the Gaussian distribution. It shows the coherence of the mean μ and covariance Σ of data set X. In addition, $\mathcal{N}\left(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ can be obtained as:

$$\mathcal{N}\left(x_i, \mu_k, \Sigma_k\right){=}\frac{1}{(2\pi)^{\frac{1}{2}}\left|\Sigma_k\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left(x_i{-}\mu_k\right)^T\Sigma_k^{-1}\left(x_i{-}\mu_k\right)\right) \tag{5}$$

In the M step, the parameters are updated according to the following rules:

$$\begin{aligned}
\pi_k^{(j+1)}&{=}\frac{1}{N}\sum_{i=1}^{N}h_k^{(j)}(i)\\
\boldsymbol{\mu}_k^{(j+1)}&{=}\frac{\sum_{i=1}^{N}h_k^{(j)}(i)x_i}{\sum_{i=1}^{N}h_k^{(j)}(i)}\\
\boldsymbol{\Sigma}_k^{(j+1)}&{=}\frac{\sum_{i=1}^{N}h_k^{(j)}(t)\left(x_i-\boldsymbol{\mu}_k^{(j+1)}\right)^2}{\sum_{i=1}^{N}h_k^{(j)}(i)}
\end{aligned} \tag{6}$$

where N represents the size of the dataset. This process is repeated until the algorithm converges, *i.e.*, the model parameters do not change significantly from one iteration to the next.

In most applications, k and Σ are unknown. Therefore, comparing information criteria can tune a GMM. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are two popular information criteria. Let k be the desired number of components and Σ be the covariance structure of all components. Algorithm 2 shows the tunning process of the GMM.

Algorithm 2 Fitted GMM

Input: Dataset ($X{=}x_1, x_2, ..., x_N$), the maximum number of Gaussian distributions (K), maximum number of iterations ($\widetilde{n}$), covariance structure (Σ)

. **Output:** Tuned GMM

Initialize GMM parameters: μ_k, Σ_k, π_k
Set $\widetilde{n}$
for $k{=}1$ to K **do**
  for each data point x_i in X **do**
    while $n \leq \widetilde{n}$ **do**
      %Expectation step (E-step)
      Calculate the probability density function $\mathcal{N}\left(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ with (5)
      Calculate the posterior probability $h_k(i)$ with (4)
      %Maximization step (M-step)
      Update μ_k, Σ_k, π_k with (6)
      $n{\leftarrow}n{+}1$
    end while
  end for
end for
Obtain μ_k, Σ_k, π_k for each Gaussian distribution
for each (k, Σ_k) pair **do**
  Estimate the AIC and BIC
end for
Choose the (k, Σ_k) pair with lowest AIC or BIC
Return the fitted GMM

III. Experimental Results and Discussion

A. Combining Expert Knowledge for Correlation Analysis

By examining eight specified settings and four observation tasks, each group accumulates 115,000 data points, representing seven days of operational data. The findings of the correlation study are presented in Table I. The strong and weak association rules can be identified by filtering variables with better scores. After the expert interpretation, the data analysis results dismiss the significant link between the output pressure $Y3$ and the kiln speed $X5$. The wind pressure in a kiln is mainly influenced by factors, including fan speed, ventilation efficiency, and kiln output, rather than the kiln rotation speed. Therefore, this work posits that the cause may be attributed to biased data samples. Hence, utilizing specialized expertise and methodologies can decrease the complexity of data analysis and rectify the results.

TABLE I
Results of Correlation analysis

Series	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$
$Y1$	0.9274	0.6978	0.8249	0.8094	0.7698	0.7347	0.5318
$Y2$	0.8819	0.6919	0.8140	0.8060	0.7408	0.7305	0.8181
$Y3$	0.7683	0.8876	0.8612	0.6852	0.8183	0.6577	0.8441
$Y4$	0.7214	0.6087	0.9152	0.9145	0.7125	0.8218	0.7182

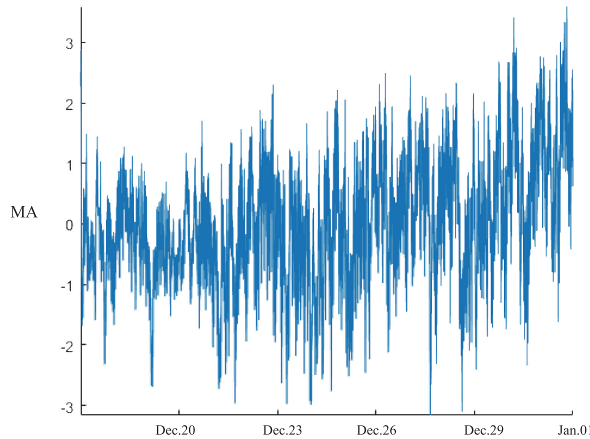
B. Feature Exaction of Working Condition

The dataset's collection period for observed variables is three seconds for each interval, which is not constant. Therefore, the time series is initially resampled at a uniform interval of 30 seconds per time interval for further processing. The data undergoes feature exaction before the unsupervised learning. This work utilizes the two-dimensional properties of MA and C.V. for cluster analysis. The time frame for the MA is configured to five minutes, while the time window for the standard deviation is set to 30 minutes. In addition, the Z-score normalization technique is adopted to preprocess the feature data to address dimensional disparities. It enhances the algorithm's convergence rate and mitigates the influence of feature discrepancies on the model. The results following the extraction of feature values and preprocessing are shown in Fig. 2.

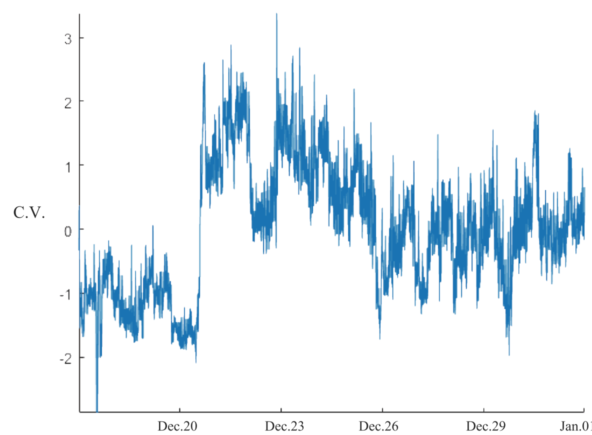
C. Clustering by GMM

Fig. 3 shows the preprocessed dataset of the characteristic variables of $Y1$ kiln current. Initially, it is shown in Fig. 4 that the dataset is represented by the contour of the fitted GMM. After applying the fitted GMM, the data is partitioned into distinct clusters. The outcome of this partitioning is illustrated in Fig. 5. Then, the GMM should be fine-tuned. Specifically, the experiments begin by explicitly stating all the possible options for covariance structures and employing regularization techniques to prevent the occurrence of poorly conditioned covariance matrices. The number of iterations for the EM algorithm is set to 10,000. Next, the GMM is trained using all possible combinations of parameters, determining the values of AIC and BIC for every fitting. The final convergence status of each partner is monitored. According to the AIC and BIC values, the optimal model is found to have five components and a diagonal, non-shared covariance matrix structure.

The best GMM results are shown in Table II. Mean Value (1) refers to the arithmetic mean calculated from the MA values associated with each cluster's set of features. Similarly, Mean Value (2) denotes the average of the C.V. values for the features within each cluster. It is shown that the two feature components chosen in the experiment are independent, and each component has its covariance matrix. An ellipse's major and minor axes are parallel or perpendicular to the x- and y-axes. However, they can vary in terms of their size and direction. This aligns with the current reality, as there is no association between the average and volatility of the observations, and the data size for each state classification is unequal. The most suitable model is ultimately employed to group the training data into clusters. Then, it is shown in Fig. 6 that the clustered data and the plot of component ellipses are generated.



(a) Preprocessing for MA



(b) Preprocessing for C.V.

Fig. 2. Preprocessing feature value

TABLE II
Best GMM results

Component	Mixing Ration	Mean Value (1)	Mean Value (2)
1	0.2834	0.0686	0.3077
2	0.3240	-0.2065	-1.0617
3	0.1568	-0.5962	1.4765
4	0.1791	1.2219	-0.1239
5	0.0566	-1.4033	0.8464

D. Comparative Experiments

Two sets of comparative experiments are conducted to assess the efficiency of the WCP. The experiments are executed using a

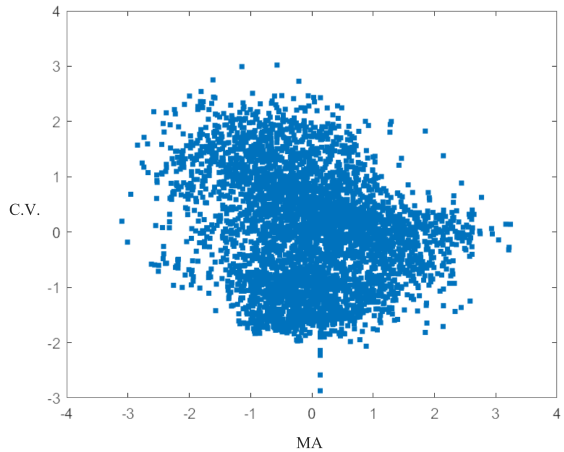


Fig. 3. Scatter plot and Fitted Gaussian Mixture contours

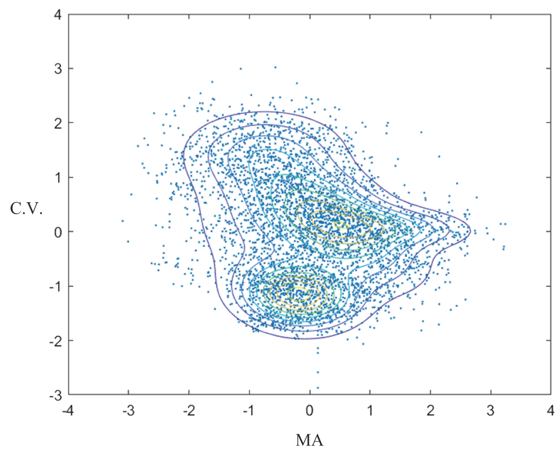


Fig. 4. Data set of Y1 eigenvalue

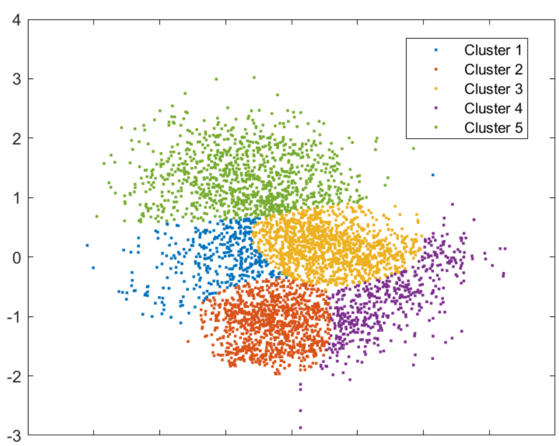


Fig. 5. Clustered data and component structures

consistent dataset, employing a rigorous five-fold cross-validation methodology. The principal aim is to measure the model's accuracy. Thus, the F1-score is adopted to measure the precision. The comparative models are shown as:

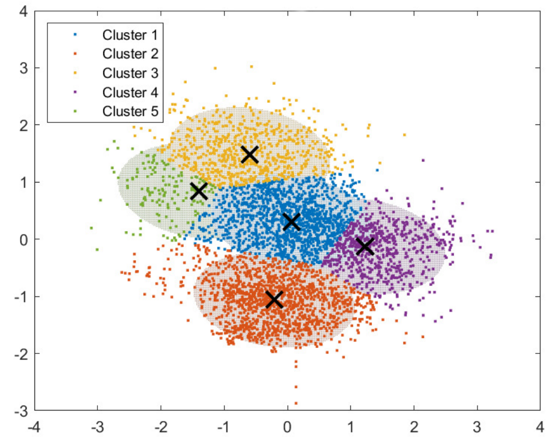


Fig. 6. Scatter plot grouped by cluster after tuned

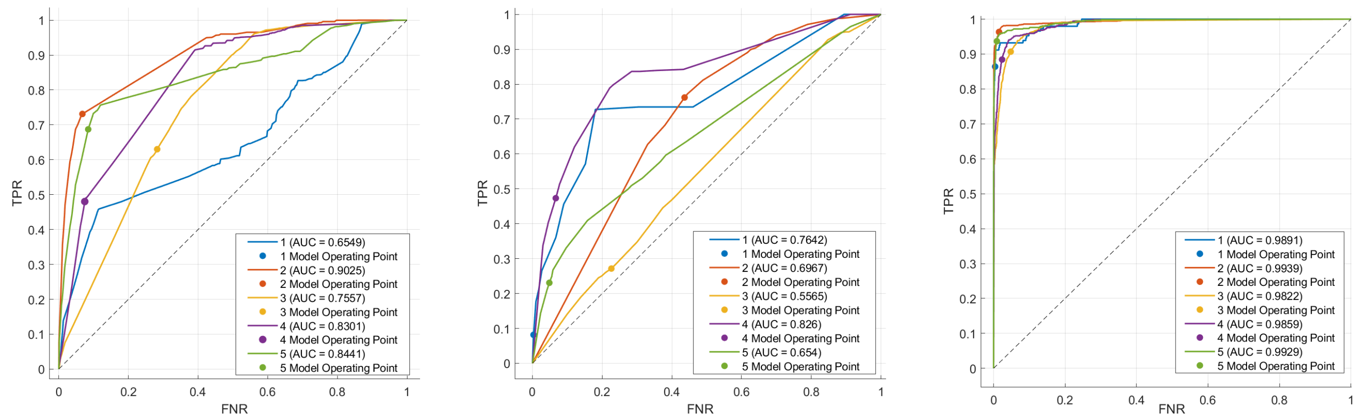
1) Utilize predetermined parameters exclusively for predicting working conditions (Setting-WCP).
2) Utilize observational data for predicting working conditions (Response-WCP).
3) Integrate the predetermined parameters in OML using GMM for predicting working conditions (OML-WCP).

Table III shows the experimental results and Fig. 7 shows the Receiver Operating Characteristic (ROC) curves of each model. They indicate that the Response-WCP achieves an average F1-score of 0.57. It demonstrates that numerical forecasting of a single variable is insufficient in effectively capturing the seasonality, periodicity, and unpredictability present in time series data. Due to numerous intricate factors that can significantly impact observational data, relying exclusively on past data to generate forecasts is incorrect. Correlation parameter analysis can enhance prediction accuracy. However, its effectiveness remains constrained, with a maximum accuracy of 73%. Furthermore, the observed variables exhibit distinct patterns despite the operator configuring the same parameter presets. This has a significant impact on the prediction accuracy. Therefore, the experimental results confirm the efficiency of the feature extraction.

IV. CONCLUSIONS

Real-time parameter adjustment is essential to maintain efficiency and quality in the process industry. Accurate identification and forecasting of working conditions are critical for informed production decisions. However, it is challenging to achieve precise control due to the industry's complex processes and unpredictable disturbances. Traditional approaches often rely on key parameters to develop regression or numerical prediction models. They struggle to capture the full complexity of the production process, resulting in less accurate predictions.

This work proposes a Working Condition Prediction (WCP) framework based on an Operation Mode Library (OML) called OML-WCP. It aims to enhance condition prediction and address the challenges of precise control in complex process industries. By integrating expert knowledge, OML-WCP establishes relationships between highly correlated quantities and observed variables, providing a deeper understanding of the underlying process mechanisms. Observed variables are adopted as proxies for working conditions, and two sets of features are extracted using sliding mean and variance analysis. The Gaussian Mixture model is then applied to construct a robust OML. Experimental results demonstrate that OML-WCP is highly effective for predicting working conditions, achieving a success rate exceeding 92%. Furthermore, the continuous accumulation of OML



(a) ROC (Setting-WCP) (b) ROC (Response-WCP) (c) ROC (OML-WCP)

Fig. 7. ROC of each model

TABLE III
F1-SCORES OF ABLATION STUDIES

Method	Task1	Task2	Task3	Task4	Average
Setting-WCP	0.6451	0.7230	0.7695	0.7837	0.7303
Response-WCP	0.4516	0.4763	0.6388	0.7660	0.5728
OML-WCP	**0.9126**	**0.9393**	**0.9332**	**0.9001**	**0.9213**

can improve prediction accuracy in practical applications. This state-based prediction method is adaptable in various process industries.

REFERENCES

[1] J. Ding, C. Yang, Y. Chen and T. Chai, “Research Progress and Prospects of Intelligent Optimization Decision Making in Complex Industrial Process,” *Acta Automatica Sinica*, vol. 44, no. 11, pp. 1931–1943, Nov. 2018.

[2] J. Bi, Z. Wang, H. Yuan, J. Zhang and M. Zhou, “Cost-Minimized Computation Offloading and User Association in Hybrid Cloud and Edge Computing,” *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 16672–16683, May 2024.

[3] H. Yuan, J. Bi, Z. Wang, J. Yang, and Jia Zhang, “Partial and Cost-minimized Computation Offloading in Hybrid Edge and Cloud Systems,” *Expert Systems with Applications*, vol. 250, pp. 1–13, Sept. 2024.

[4] W. Huang, H. Ding and J. Qiao, “Limestone Calcination Kinetics in Microfluidized Bed Thermogravimetric Analysis (MFB-TGA) for Calcium Looping,” *Catalysts*, vol. 12, no. 12, pp. 1661–1680, Dec. 2022.

[5] M. Valipour, M.E. Banihabib and S. Behbahani, “Comparison of the ARMA, ARIMA, and the Autoregressive Artificial Neural Network Models in Forecasting the Monthly Inflow of Dez Dam Reservoir,” *Journal of Hydrology*, vol. 476, pp. 433–441, Jan. 2013.

[6] C. Challu, K.G. Olivares, B.N. Oreshkin, F.Garza, M.M. Canseco and A. Dubrawski, “NHITS: Neural Hierarchical Interpolation for Time Series Forecasting,” *The 37th AAAI Conference on Artificial Intelligence*, 2023, Washington, DC, USA, pp. 6989–6997.

[7] Y. He, Z. Lu, J. Wang, S. Ying and J. Shi, “A Self-Supervised Learning Based Channel Attention MLP-Mixer Network for Motor Imagery Decoding,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2406–2417, Aug. 2022.

[8] H. Liang, L. Song, J. Du, X. Li and L. Guo, “Consistent Anomaly Detection and Localization of Multivariate Time Series via Cross-Correlation Graph-Based Encoder–Decoder GAN,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, Dec. 2022.

[9] J. Li, C. Shen, L. Kong, D. Wang, M.Xia and Z. Zhu, “A New Adversarial Domain Generalization Network Based on Class Boundary Feature Detection for Bearing Fault Diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, Apr. 2022.

[10] K. Ehdieh, P. Shikhar, K. Pratim and S. Anurag K, “Real-Time Synchrophasor Data Anomaly Detection and Classification Using Isolation Forest, KMeans, and LoOP,” *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2378–2388, May 2021.

[11] R. Luigi, S. Kisan, G. Luigi and A. Antonio, “Fault Detection and Diagnosis in Steel Industry: a One Class-Support Vector Machine Approach,” *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, Melbourne, Australia, pp. 2304–2309.

[12] C. Ma, S. Wei, T. Chen, J. Zhong, Z. Liu and C. Liu, “Integration of Results From Convolutional Neural Network in a Support Vector Machine for the Detection of Atrial Fibrillation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, Dec. 2021.

[13] J. Bi, Z. Wang, H. Yuan, J. Zhang and M. Zhou, “Self-adaptive Teaching-learning-based Optimizer with Improved RBF and Sparse Autoencoder for High-dimensional Problems,” *Information Sciences*, vol. 630, pp. 463–481, Jun. 2023.

[14] Z. Geng, Z. Chen, Q. Meng and Y. Han, “Novel Transformer Based on Gated Convolutional Neural Network for Dynamic Soft Sensor Modeling of Industrial Processes,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1521-–1529, Mar. 2022.

[15] Y. Cao, K. Ngo and D. Dong, “A Scalable Electronic-Embedded Transformer, a New Concept Toward Ultra-High-Frequency High-Power Transformer in DC–DC Converters,” *IEEE Transactions on Power Electronics*, vol. 38, no. 8, pp. 9278–9293, Aug. 2023.

[16] T. Wu, C. Yang, Y. Li, H. Zhu and W. Gui, “Fuzzy Operational-pattern Based Operating Parameters Collaborative Optimization of Cobalt Removal Process with Arsenic Salt,” *Acta Automatica Sinica*, vol. 40, no. 8, pp. 1690–1698, Aug. 2014.

[17] M. Zhu, Y. Ji, Z. Zhang and Y. Sun, “A data-driven decision-making framework for online control of vertical roller mill,” *Computer& Industrial Engineering*, vol. 143, No. 106441, pp. 1–16, May 2020.

[18] N. Ghafoor and M. Ahmad, “Prioritizing Effectiveness of Algorithms of Association Rule Mining,” *Journal of Computational Learning Strategies & Practices*, vol. 1, No.1, pp. 18–30, Nov. 2021.

[19] A. Aubry, A. De Maio, S. Marano and M. Rosamilia, “Structured Covariance Matrix Estimation With Missing-(Complex) Data for Radar Applications via Expectation-Maximization,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 5920–5934, Sept. 2021.