Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Industrial Time Series Forecasting for Fluid Catalytic Cracking Process

Qiming Zhao[a,b], Yaning Zhang[c], and Tong Qiu[a,b]*

[a] Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
[b] State Key Laboratory of Chemical Engineering, Tsinghua University, Beijing 100084, China
[c] PetroChina Planning & Engineering Institute, Beijing 100083, China
* Corresponding Author: qiutong@tsinghua.edu.cn.

ABSTRACT

This study tackles the challenge of accurate yield prediction in fluid catalytic cracking (FCC) units by comparing conventional supervised regression with time series forecasting methods using industrial data collected from the distributed control system (DCS) of an FCC plant. We introduce a shifted forecast paradigm that preserves temporal relationships between predictors and targets. Our preprocessing pipeline, which employs trimmed mean smoothing, addresses common industrial data challenges. Results demonstrate that the forecasting approach significantly outperforms supervised regression, achieving a mean absolute percentage error (MAPE) of 1.56% for 3-hour shifted predictions compared to 6.20% for supervised regression. The model maintains robust performance even with extended shifts during predictions, showing an MAPE of 3.55% for 14-day forecasts. This research provides valuable insights for implementing predictive analytics in industrial FCC operations, demonstrating the superiority of forecasting methods over traditional supervised regression approaches for process yield prediction.

Keywords: Machine Learning, Forecasting, Catalytic Cracking, Predictive Modeling

INTRODUCTION

Fluid catalytic cracking (FCC) is a cornerstone process in modern refineries, playing a crucial role in meeting global energy demands. It enables the efficient conversion of heavy hydrocarbons into lighter and more valuable products, including gasoline and diesel. As crude oil quality continues to decline while global energy demands increase, refineries must navigate intensifying economic and environmental constraints. Under these challenging conditions, process predictive modeling has emerged as an indispensable tool for understanding, controlling, and optimizing the complex behavior of FCC units [1].

Data-driven modeling approaches have gained significant traction in FCC process modeling. These methods excel at capturing complex nonlinearities and interactions and do not require a detailed understanding of the underlying physicochemical phenomena, making them particularly suitable for FCC process modeling [2]. Researchers can leverage the vast amounts of data generated by FCC units to develop empirical models that capture the complex relationships between process variables. By effectively modeling these relationships, operators can make informed adjustments to operating conditions, thereby optimizing performance and enhancing profitability. Machine learning techniques, such as artificial neural networks (ANNs), support vector machines (SVMs), and Gaussian process regression (GPR), have been successfully applied to model various aspects of FCC operation [2]. However, these models often struggle to extrapolate to future operation conditions. Since the FCC process is inherently dynamic and subject to fluctuations over time, effective data-driven modeling requires incorporating time-related information, namely forecasting.

Given these limitations, time series forecasting approaches offer promising alternatives for FCC process modeling. By analyzing temporal dependencies and trends, this modeling approach captures the dynamic nature of the process and provides insights into potential future states. Traditional supervised learning algorithms typically assume independent, identically distributed (IID) inputs [3], and therefore often fail to capture the dynamic nature of industrial processes. In contrast, forecasting

methods emphasize temporal patterns such as seasonality and trends [4], producing more context-aware predictions.

In FCC data, short-term variable relationships remain stable during steady-state operations despite oscillations. However, long-term operations present challenges as operational contexts and system mechanics evolve, complicating the development of consistent input-output mappings. Integrating forecasting techniques with supervised regression methods enhances prediction robustness while accounting for the inherent variability in industrial time series data. Liu *et al.* [5] proposed a hybrid Gaussian Mixture Model (GMM)-Improved Gaussian Process Regression (IGPR) model for time series prediction in industrial processes with varying operating phases and uncertainties. TC-GATN, a novel multivariate time-series forecast model developed by Li *et al.* [6], uses Granger causality-based graph learning and gated recurrent units with self-attention to capture complex nonlinear dependencies in industrial processes, improving prediction accuracy as demonstrated on methanol and chlorosilane datasets. These advancements in industrial time series analysis reflect a broader trend toward more adaptive and context-aware modeling approaches. Nevertheless, their efficacy is predominantly evaluated using moderate-scale systems and datasets, exhibiting more predictable and stable operational characteristics.

In this paper, we develop a shifted forecast paradigm that directly bridges current operational states with future yield predictions of an FCC process. Our approach differs from conventional time series forecasting by explicitly modeling the temporal gap rather than relying on iterative one-step-ahead predictions. The main contributions of this work include: (1) a robust data preprocessing pipeline that addresses common challenges in industrial time series data, including irregular sampling, outliers, and high-frequency fluctuations; (2) a shifted forecast framework that maintains temporal relationships while avoiding the limitations of traditional supervised regression; and (3) comprehensive evaluation of model performance across different prediction shifts, demonstrating its practical utility for industrial applications.

The remainder of this paper is organized as follows. Section 2 describes the data preparation methodology, including variable selection, preprocessing techniques, and dataset organization. Section 3 presents the theoretical framework of our regression and forecasting approaches, with particular emphasis on the shifted forecast paradigm. Section 4 provides detailed experimental results and a comparative analysis of different modeling approaches across various prediction horizons. Finally, Section 5 concludes the paper with a summary of findings and suggestions for future research directions.

METHODS

Figure 1 shows the framework for developing the prediction model, which consists of two modules: data preparation and regression forecasting.

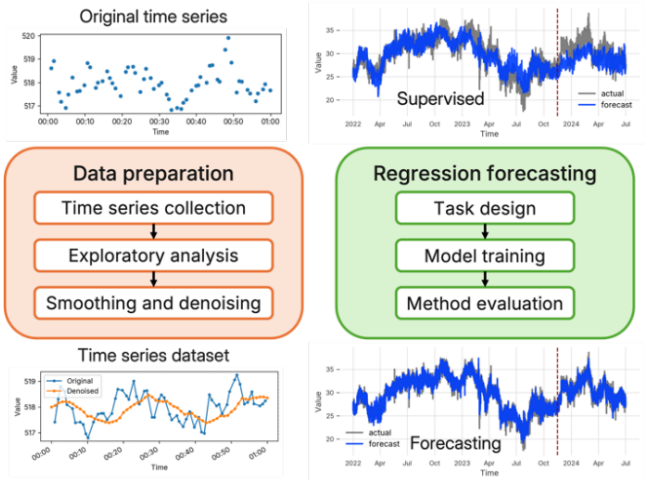

Figure 1. Forecast model development framework.

High-quality datasets are generated and then used to build and validate forecast models. The pipelines for data preprocessing and model training are implemented in Python 3.12, using Darts [7], a time series forecasting library.

Table 1: Variables in the collected dataset.

Symbol	Description	Unit
T_{feed}	Feed preheat temperature	℃
T_{reac}	Reactor temperature	℃
T_{regen}	Regenerator riser temperature	℃
T_{stripper}	Stripper overhead temperature	℃
T_{fractop}	Fractionator overhead temperature	℃
T_{naph}	Heavy naphtha extraction temperature	℃
T_{LCO}	Light cycle oil (LCO) extraction temperature	℃
T_{fracbot}	Fractionator bottoms temperature	℃
F_{stripper}	Stripper overhead flow rate	Nm^3/h
F_{feed}	Flow rate of fresh feed	t/h
F_{reflux}	Flow rate of fractionator reflux	t/h
F_{HCO}	Flow rate of heavy cycle oil (HCO) to riser	t/h
$F_{\text{recyc\_gaso}}$	Flow rate of recycled stabilized gasoline	t/h
COR	Catalyst-to-oil ratio	/
y	Diesel yield	wt%

Industrial dataset preparation

A typical FCC plant is composed of two core modules: the reactor-regenerator and the downstream

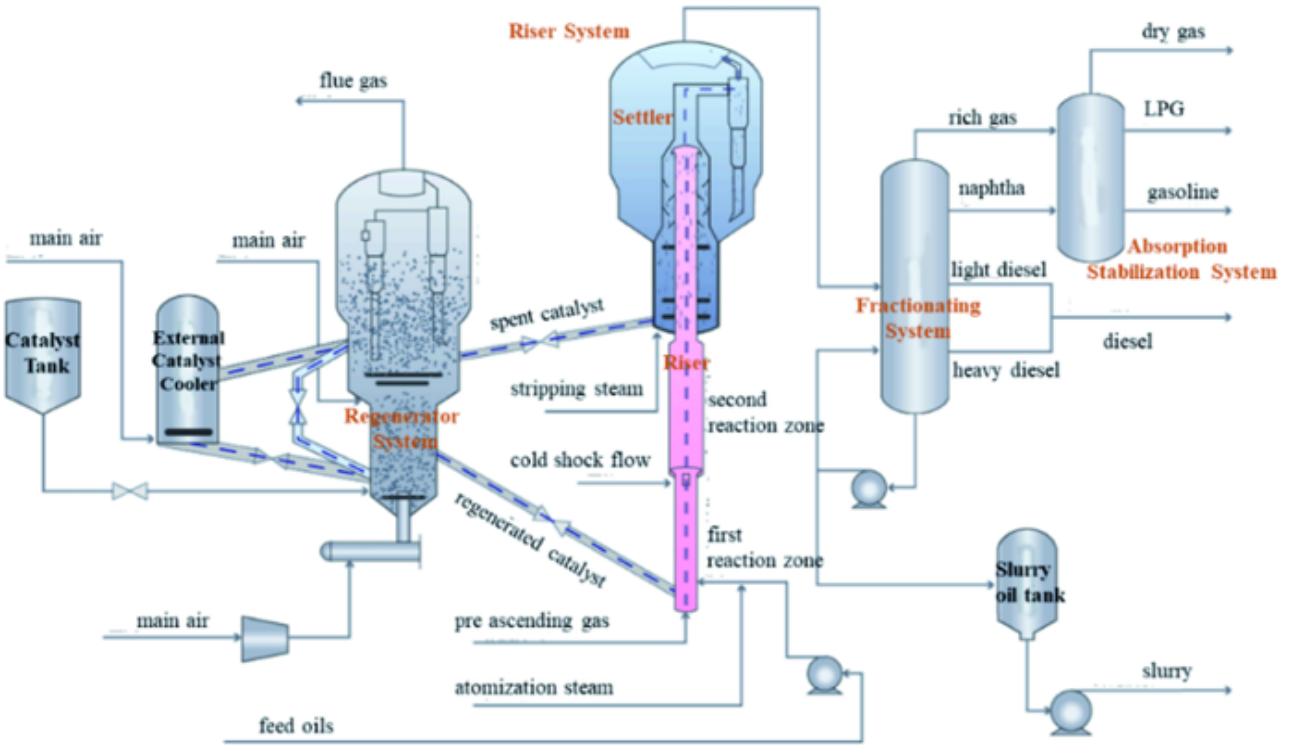

Figure 2. A schematic diagram of the FCC unit [8].

product separation system. Raw feedstock is cracked in the riser reactor upon contact with the regenerated catalyst. Subsequently, the cracked mixture is separated into desired products by one or more separation units. In the regenerator, coke is removed by combustion in a controlled process, typically using air. The detailed substance flows are depicted in Figure 2.

We collected time series data from the DCS system in the FCC plant. We preselected variables of interest using domain knowledge, covering the mass flows of feedstocks and products, major operating parameters of the reactor, regenerator, and downstream separation systems like adsorption and distillation columns. A derived variable, catalyst-to-oil ratio (COR), was added to reflect cracking depth, whose values are calculated using an empirical formula. The target variable, diesel yield, is calculated as the mass flow rate ratio of diesel product to fresh feed. Table 1 lists the details of selected variables.

Effective predictive modeling requires robust dataset preparation to ensure data quality and relevance for model training and testing. Data collected from the distributed control system (DCS) may suffer from outliers, irregular sample intervals, and high-frequency fluctuations, hence the need for specific preprocessing methods. These characteristics stem from sensor failures, inconsistent data collection rates, and external measurement influences. Data quality is critical to forecasting performance, and exploratory data analysis (EDA) is necessary for building an effective preprocessing pipeline. Through inspection and visualization, we found the data to be irregularly sampled, with corruptions, outliers, and local oscillations.

To address these challenges, we employed preprocessing techniques to ensure the dataset adhered to the necessary specifications for time series analysis. Irregularly sampled data are aligned to generate a long-form table, facilitating a consistent modeling approach. In time series analysis, rolling window methods—particularly moving averages (or rolling means)—provide an effective approach to smoothing data, helping distinguish signal from noise by minimizing abrupt fluctuations. While simple averaging considers all data points equally, the trimmed mean provides a more robust approach by removing a specified percentage of the most extreme values that could skew the results. The resulting data are aligned and smoothed, and outliers are implicitly removed in this process. This method is called rolling trimmed mean, which can be described as

$$\mathrm{RTM}(p,t,w) = \frac{1}{w-2\lfloor pw \rfloor} \sum_{i=\lfloor pw \rfloor+1}^{w-\lfloor pw \rfloor} x_{(i)} \tag{1}$$

where t represents the current time point, w is the window size, p is the trimming percentage expressed as a decimal, and $x_{(i)}$ represents the i-th ordered observations within the window $[t-w+1, t]$ that ends at time t.

As a final step, operational parameters are standardized to a similar scale using the z-score method. A temporal split of training and test sets is also carried out,

with the latter taking up 20% of the total data.

Regression and forecasting

With the dataset prepared, the next phase involves applying time series modeling techniques to forecast future outcomes based on historical data. Among various regression techniques, Bayesian Ridge Regression is selected for its ability to handle multicollinearity and incorporate prior distributions, which helps regularize the model [9]. Bayesian Ridge Regression uses prior distributions for the weights and updates them using observed data. The prior distribution for the weights w is chosen as a normal distribution given by

$$w \sim \mathcal{N}(0, \sigma^2/\lambda) \tag{2}$$

where λ is the precision parameter controlling the strength of regularization, and σ^2 represents the noise variance in the observed data. These hyperparameters are automatically tuned during model training using evidence maximization.

The same regression algorithm is used for different prediction tasks, which can be classified into supervised regression and forecasting. They use the same variable set but with fundamentally different settings in temporal structures and prediction objectives. Supervised regression uses features and lagged versions of them to predict targets, while forecasting methods also use historical values of targets. The forecast model uses two types of input data: the target variable (diesel yield) and covariates (process variables like temperatures and flow rates).

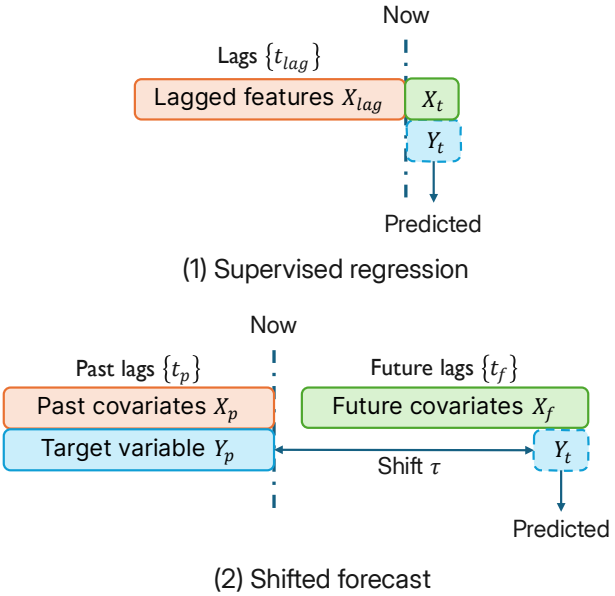

Figure 3. Comparison of temporal prediction frameworks illustrating data availability and prediction targets.

Figure 3 illustrates the details of model inputs and outputs, highlighting the difference in the temporal structure. Solid boxes represent available data at prediction time, while dashed boxes indicate predicted values. Vertical dashed lines indicate prediction time points. Past covariates are historical data known up to now, and future covariates are known in the future, at least up to the last forecasted value.

In supervised regression (Figure 3a), the model learns from features X_t and their lagged values to predict a concurrent target variable Y_t, where all data is available during training. The model predicts each target independently using lagged features, reflecting a global input-output mapping function.

Conversely, external data are introduced in time series forecasting, which are called covariates. Forecasting makes a distinction between past and future time points relative to the prediction moment. Past covariates are historical data known up to now, and future covariates are known at least up to the last forecasted value (often controlled variables). This approach maintains the temporal ordering essential for time series analysis while avoiding data leakage, as only information available at prediction time is utilized. The lags are chosen to provide local gradients with minimum redundancy:

$$\{t_{lag}\} = \{X_{t-60}, X_{t-30}, X_{t-15}, X_{t-10}, X_{t-5}\} \tag{3}$$

where t_{lag} corresponds to past lags t_p or future lags t_f, and t is the timestamp in minutes relative to the end of the corresponding covariate series.

The shifted forecast approach predicts future values by maintaining a fixed time gap between inputs and outputs (Figure 3b). This approach effectively merges lagged inputs (from supervised models) with past covariates (from ordinary forecast models). The temporal difference between past covariates and predicted targets is defined by the prediction shift τ.

A shift of $\tau = 0$ corresponds to one-step-ahead forecasting, where the model leverages data up to the current time point to predict the next value. On industrial datasets with steady states during most periods, the model may degenerate into a naive forecaster that simply repeats the last observation, providing no meaningful information beyond the current state.

Meanwhile, a shifted forecast model with $\tau > 0$ learns to bridge the temporal gap between current information and future outcomes, embedding process information in a way similar to how ordinary supervised regression models learn relationships between input features and output variables. The model implicitly handles slow drifts in process behavior through adaptive corrections of target values with recent historical data.

The models are evaluated with two metrics: mean absolute percentage error (MAPE) and coefficient of variation (CV). In a dataset of n samples, if y_i is the i-th target value and $\hat{y}_i$ is the corresponding prediction, MAPE is defined as

$$\mathrm{MAPE}(\mathbf{y},\hat{\mathbf{y}}) = \frac{1}{n}\sum_{i=0}^{n-1}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \quad (4)$$

as long as the true values are not close to zero.

The coefficient of variation (CV) is also called normalized root mean squared error (NRMSE), which is defined as

$$\mathrm{CV}(\mathbf{y},\hat{\mathbf{y}}) = \frac{\mathrm{RMSE}}{\bar{y}} = \frac{\sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i-\hat{y}_i)^2}}{\frac{1}{n}\sum_{i=0}^{n-1} y_i} \quad (5)$$

In brief, our method integrates data preprocessing (alignment, outlier removal, standardization), a supervised regression baseline, and a shifted forecast paradigm that explicitly models time gaps between observations and predictions. These experiments directly compare traditional supervised learning and the proposed forecasting method with various data requirements.

RESULTS AND DISCUSSION

Data preprocessing

The collected data span two and a half years, from January 2022 to June 2024. Data was sampled approximately once per minute, but the timestamps are not aligned to the minute, and multivariate timestamps are not synchronized. Data points recorded as zero are identified as sensor failures through validation against historical operating ranges. These corrupted readings, comprising approximately 0.12% of the dataset, are removed before further analysis. The dataset undergoes preprocessing using a rolling trimmed mean applied within a 10-minute window. This process involves trimming 5% of extreme values, which effectively smooths the time series and removes outliers, thus preparing it for subsequent modeling. Figure 4 presents examples of original and denoised samples for reaction temperatures and diesel yields.

The denoised temperatures in Figure 4a generally follow the overall trend of the original data but with fewer sporadic deviations. The sharp dip in the original data is significantly reduced in the denoised version, as demonstrated in Figure 4b. The results have smoother and more interpretable trends, providing a good estimate of the system state in the past 10 minutes.

Model training and validation

The models are trained on data from approximately April 2022 to late 2023, and their predictions are generated on the full dataset without retraining. The results for the supervised and forecast models are depicted in Figure 5.

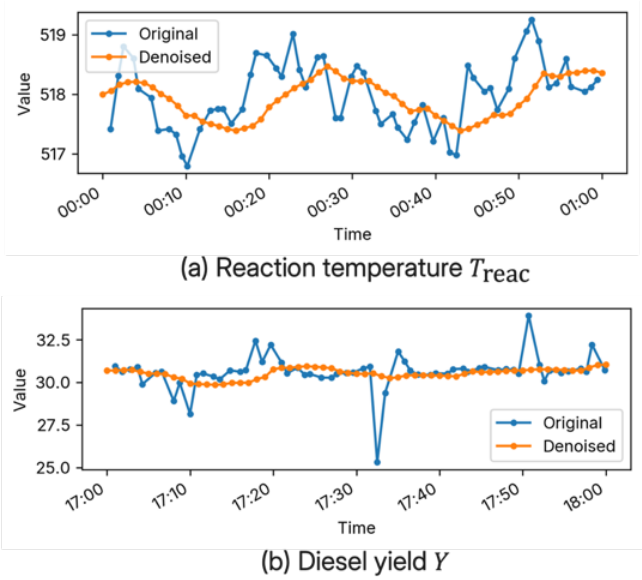

Figure 4. Comparison of original and denoised data.

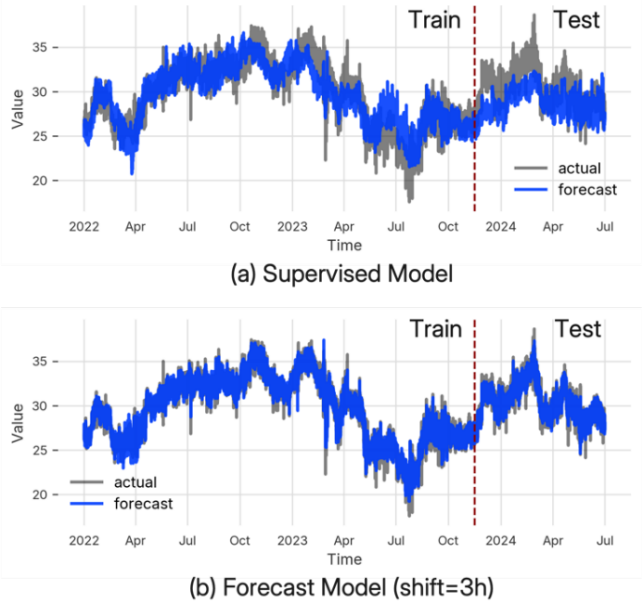

Figure 5. Performance of supervised and forecast models on the whole dataset.

The supervised model captures the main trends within the training dataset, maintaining an MAPE of 4.42% during this period. However, its performance deteriorates significantly in the test set, where the MAPE increases to 6.20%, indicating potential overfitting. In contrast, the forecast model maintains consistent performance across both sets, with MAPE values of 1.55% and 1.56% for training and test sets, respectively.

Furthermore, three forecast models are trained and tested using 3-hour, 1-day, and 14-day prediction shifts to evaluate the accuracy and robustness across various timescales. These shifts are chosen for operational relevance, representing short-term process control, daily

planning, and long-term strategic forecasting. Figure 6 compares the performance metrics of these three forecast models, revealing the gradual change of their predictive capabilities across different temporal resolutions.

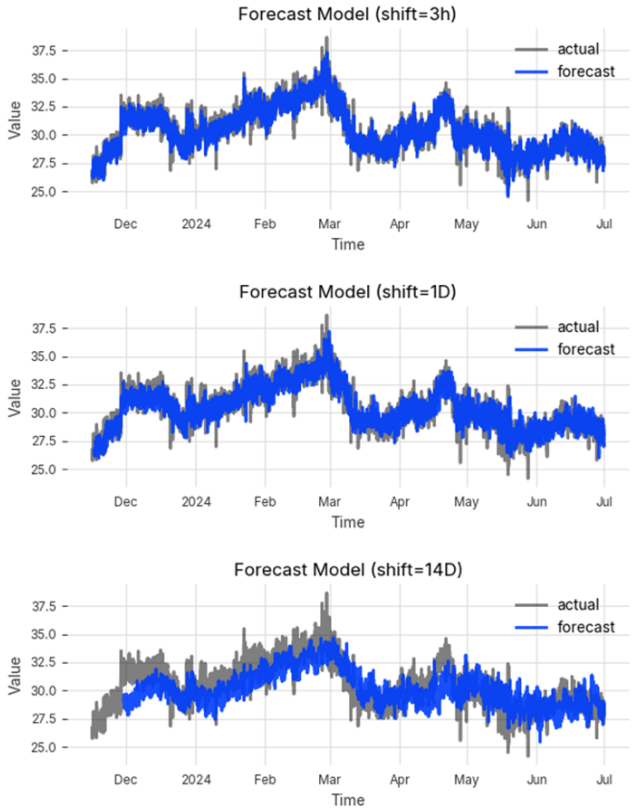

Figure 6. Test performance of forecast models using different shifts.

As the shift is prolonged, the model demonstrates reduced accuracy, generally approaching the predictions from the supervised model. However, the forecast model performs much better on the test set, even when predicting the target value 14 days after the known history of diesel yields. This sustained accuracy suggests the successful capture of underlying process relationships rather than simple trend extrapolation.

Table 2 presents the test metrics of different models.

Table 2: Test metrics of supervised and forecast models.

Model	Shift t_s	MAPE (%)	CV
Supervised	/	6.198	7.567
Forecast	3 hours	1.564	2.120
	1 day	2.067	2.662
	14 days	3.551	4.456

Forecast models perform better than the supervised one on both metrics, which is reasonable since the inputs to the forecast model are the superset of those available to the supervised one. In real industrial applications, the shift can be selected according to the online data collection mechanism and how often the process states are expected to change.

The forecast model's consistent accuracy validates our approach of explicitly modeling temporal relationships in industrial time series. These results indicate that incorporating historical target values and temporal dependencies is crucial for accurate FCC yield prediction.

CONCLUSION

This study demonstrates the effectiveness of time series forecasting approaches over supervised regression for predicting FCC diesel yields in industrial settings. The proposed shifted forecast paradigm achieves superior performance across all evaluation metrics, effectively bridging current process states with future operational plans. Experimental results indicate that incorporating historical target values and temporal relationships significantly improves prediction accuracy, with the 3-hour shifted forecast model achieving a 74.8% reduction in MAPE compared to supervised regression (1.56% vs. 6.20%). The model's ability to maintain reasonable accuracy (MAPE of 3.55%) even for 14-day predictions suggests robust learning of process variable correlations. These findings have important implications for industrial applications, enabling more effective process optimization and planning through accurate long-term predictions. The flexibility in selecting prediction horizons makes this approach adaptable to various operational requirements and data collection mechanisms, providing a practical solution for real-world implementation in FCC units.

It is worth noting that feed quality data, such as density and hydrocarbon distribution profiles, were not incorporated in the current model due to the varied and unstable measurement frequency. Future research could incorporate laboratory analysis data, which typically provides more accurate measurements but at much lower sampling frequencies than DCS data. Specifically, integrating laboratory-measured product qualities with high-frequency sensor data could create multi-resolution models that leverage the strengths of both data sources. Further exploration of deep learning methods, particularly transformer architectures that can handle variable-length sequences and capture long-range dependencies, would also be beneficial. Key implementation challenges include handling mixed sampling rates between DCS and laboratory data, managing computational requirements for real-time prediction, and maintaining model interpretability for process engineers. These advancements could potentially extend the application of our forecasting approach to other complex industrial processes facing similar prediction challenges.

REFERENCES

1. Pinheiro CIC, et al. Fluid Catalytic Cracking (FCC) Process Modeling, Simulation, and Control. *Ind. Eng. Chem. Res.* 51(1):1-29 (2012) https://doi.org/10.1021/ie200743c.
2. Santander O, Kuppuraj V, Harrison CA, Baldea M. An open source fluid catalytic cracker - fractionator model to support the development and benchmarking of process control, machine learning and operation strategies. *Comput. Chem. Eng.* 164:107900 (2022) https://doi.org/10.1016/j.compchemeng.2022.107900.
3. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press (2016).
4. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. OTexts (2021).
5. Liu T, Wei H, Liu S, Zhang K. Industrial time series forecasting based on improved Gaussian process regression. *Soft Comput* 24(20):15853–15869 (2020) https://doi.org/10.1007/s00500-020-04916-6.
6. Li J, Shi Y, Li H, Yang B. TC-GATN: Temporal Causal Graph Attention Networks With Nonlinear Paradigm for Multivariate Time-Series Forecasting in Industrial Processes. *IEEE Trans. Ind. Inform*. 19(6):7592–7601 (2023) https://doi.org/10.1109/TII.2022.3211330.
7. Herzen J, et al. Darts: User-Friendly Modern Machine Learning for Time Series. *J. Mach. Learn. Res.* 23(124):1–6 (2022).
8. Ni, P, Liu, B, He, G. An online optimization strategy for a fluid catalytic cracking process using a case-based reasoning method based on big data technology. *RSC Adv.* 11(46):28557–28564 (2021) https://doi.org/10.1039/D1RA03228C.
9. Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. New York: Springer (2006).