Data-Driven Dynamic Modeling of Renewable CO_2 Emissions in Multimode Industrial Co-processing Processes

Liang Cao[a], Jianping Su[b,*], Yankai Cao[c], Lim C. Siang[d], Gary Lee[d], Jin Li[d] and R. Bhushan Gopaluni[c,**]

[a]Department of Chemical Engineering, Massachusetts Institute of Technology, Boston, 02139, United States
[b]College of Carbon Neutrality Future Technology, China University of Petroleum, Beijing, 102200, China
[c]Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada
[d]Parkland Refining (B.C.) Ltd, Burnaby Refinery, Burnaby, BC V5C 1L7, Canada

ARTICLE INFO

Keywords:
Dynamic Processes Modeling
Multimode Soft Sensors
Co-processing
Industrial Applications
Renewable CO_2

ABSTRACT

Accurate modeling and real-time monitoring of renewable CO_2 emissions in biofeedstock co-processing technologies are critical yet challenging, hindered by limited experimental data and static operational assumptions. This study introduces a novel data-driven dynamic modeling approach using an extensive dataset comprising 43,662 samples from the Parkland refinery. We implement change point detection algorithms to automatically partition the data into segments corresponding to different operating conditions and develop segment-specific robust regression models to predict CO_2 emissions. The proposed framework uniquely integrates change point detection with robust regression, forming a dynamic monitoring system that continuously adapts to multimode industrial processes while balancing numerical accuracy, interpretability, and computational efficiency. These findings reveal that the CO_2 emission ratio per unit of biofeedstock to fossil fuels fluctuates between 51% and 82% under varying operating conditions. The dynamic model exhibits strong agreement with experimental data, providing refineries with a practical, reliable tool for real-time emissions monitoring and regulatory compliance in industrial co-processing applications.

1. Introduction

To mitigate greenhouse gas (GHG) emissions, governments worldwide are enacting legislation to reduce the carbon intensity of transportation fuels(Ebadian et al., 2020; Yeh et al., 2016; Gray et al., 2021; Li et al., 2023). Co-processing has emerged as a promising near-term and cost-effective approach to reduce the carbon intensity of products and decrease process emissions(Bezergianni et al., 2018). Co-processing refers to the simultaneous processing of biofeedstock with fossil fuels in existing refinery infrastructures, allowing for the integration of renewable feedstocks without substantial changes to existing operational systems. Figure 1 shows a diagram of co-processing. The CO_2 released from biofeedstock combustion is part of the renewable carbon cycle, unlike fossil fuel CO_2, which adds to net emissions. By integrating biogenic feedstocks into existing refinery processes, the industry contributes significantly to decarbonizing the oil refining sector without requiring the complete replacement of the current infrastructure, also enabling refineries to meet regulatory mandates for renewable fuel production.

Co-processing represents one of the few methods for traditional oil companies to reduce their scope 3 greenhouse gas emissions, as burning fossil fuels contributes to 70% of the life-cycle emissions of petroleum products. The commercialization of co-processing biogenic feedstocks has been

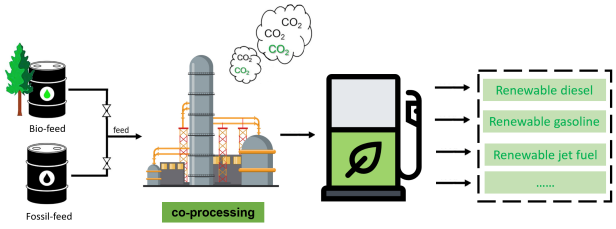

Figure 1: Illustration of the co-processing in refineries

successfully implemented in both hydrotreaters and fluid catalytic crackers (FCCs) (Badoga et al., 2020; Van Dyk et al., 2019; Han et al., 2021). While both technologies offer pathways for reducing fossil carbon emissions, their technical challenges differ. Hydrotreaters require significant modifications to process biogenic feedstocks alongside fossil inputs, posing a technical hurdle (Zacher et al., 2019). In contrast, FCC co-processing is more robust and flexible, enabling greater capacity for biogenic materials and making it more suitable for large-scale use (Stefanidis et al., 2018; Talmadge et al., 2014). The schematic representation of FCC co-processing is shown in Figure 2.

While FCC co-processing demonstrates greater technical feasibility and flexibility, a significant concern remains in accounting the carbon emissions associated with the process. The FCC unit is one of the largest sources of greenhouse gas (GHG) emissions within an oil refinery. This is mainly due to the constant regeneration of the catalyst, where the deposited coke is burned, subsequently releasing CO_2 into

*Corresponding author
**Corresponding author
jianping.su@cup.edu.cn (J. Su); bhushan.gopaluni@ubc.ca (R.B. Gopaluni)
ORCID(s):

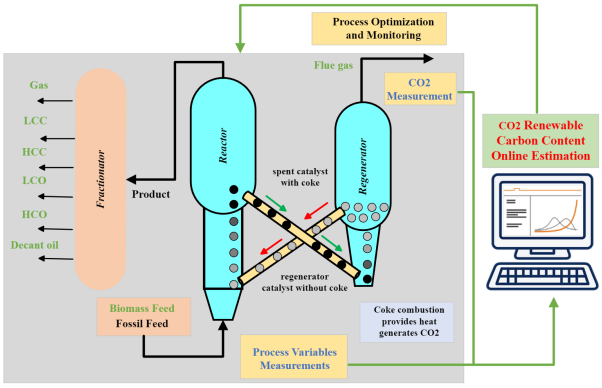

Figure 2: Diagram of a FCC unit with an online renewable CO_2 monitor

the atmosphere. Accurate modeling and monitoring of renewable CO_2 emissions from biofeedstock are crucial for refineries(Lammens, 2022; Su et al., 2022; Cao et al., 2024). However, current refinery systems lack the ability to measure renewable carbon content accurately and promptly. Developing robust methods to track these emissions is essential for refineries to optimize co-processing, confirm decarbonization efforts, and benefit from policy incentives.

This study introduces a novel data-driven dynamic modeling framework that synergistically integrates change point detection (CPD) with robust regression, creating a dynamic monitoring system specifically tailored for multimode industrial co-processing applications. The innovative aspects of our approach are highlighted by addressing significant limitations of traditional modeling methods, including their inability to adapt to continuously evolving operating conditions and their limited interpretability in noisy industrial environments. Our contributions can be summarized in four key points:

- **Synergistic Integration of CPD and Robust Regression:** Our framework uniquely combines change point detection with robust regression to dynamically segment multimode industrial processes and handle noisy, non-stationary data effectively, ensuring both interpretability and computational efficiency.
- **Novel Application and Industrial Validation:** We apply CPD to automatically identify dynamic operational conditions specifically in FCC co-processing, validated extensively on a large-scale industrial dataset of 43,662 samples from Parkland Refining Ltd, with rigorous cross-validation against ^{14}C measurements demonstrating errors predominantly below 5%.
- **Real-Time Monitoring and Process Insights:** Our approach provides accurate real-time tracking of renewable CO_2 emissions, revealing critical dynamic relationships between biofeedstock conversion rates and operational parameters, thus offering actionable insights to optimize co-processing sustainability.
- **Handling Sequential Dynamics via Segmentation:** By developing segment-specific robust regression models, our approach indirectly captures the sequential dynamics of industrial processes, adapting to temporal variations without requiring complex sequential models, while maintaining high interpretability and computational efficiency.

The proposed method continuously adapts to operational shifts and significantly enhances the accuracy, reliability, and practicality of renewable CO_2 emissions monitoring, contributing to the low-carbon transition in the energy industry. Furthermore, by comparing our segment-based approach with advanced sequential modeling techniques, we demonstrate its superior balance of interpretability, robustness, and efficiency, while identifying opportunities for integrating sequential modeling in future work to further enhance dynamic modeling capabilities.

2. Related Work

Existing studies on modeling renewable CO_2 emissions from biofeedstock co-processing have primarily focused on laboratory-scale analysis, using limited datasets. For example, Bezergianni et al. (2018) explored co-processing renewable feedstocks with fossil fuels but noted the need for more accurate emission models. Van Dyk et al. (2019) highlighted the limitations of current emission monitoring approaches, which may not reflect the complexities of industrial processes. Elliott et al. (2012) examined catalytic processes in biofeedstock co-processing, emphasizing the importance of accounting for emissions variations under different operational conditions. Dell'Orco et al. (2021) discussed the limitations of accelerator mass spectrometry (AMS), the current dominant method for measuring renewable CO_2, particularly in terms of cost-effectiveness and real-time capabilities. The AMS method is time-consuming, as the sampling can take several hours or even an entire day. In addition, the costs are significant, and each analysis often amounts to thousands of dollars when sampling costs are considered.

Several studies have explored dynamic modeling and monitoring approaches to handle non-stationary industrial processes. Methods such as Gaussian mixture models (GMM) (Pernkopf and Bouchaffra, 2005) and hidden Markov models (HMM) (Geramifard et al., 2013) have been widely applied for multimode process monitoring, capturing distinct operational states through probabilistic frameworks (Liu and Kadirkamanathan, 2015; Cao et al., 2025; Zhang and Li, 2014). While these methods effectively handle multiple operating modes, they often rely on pre-defined model structures or extensive parameter tuning, limiting their flexibility in highly dynamic and unpredictable industrial settings. Other dynamic techniques like ARIMA and Kalman filters have been extensively used for time-series prediction in industrial applications but are generally limited in capturing abrupt and non-linear shifts due to their linear assumptions (Box et al., 2015; Barbarisi et al., 2006).

In terms of interpretability, substantial progress has been made in developing interpretable machine learning techniques suitable for industrial applications. Linear models and robust regression approaches have been favored due to their inherent transparency and ease of validation in regulatory contexts (Du et al., 2019; Rudin, 2019). Meanwhile, Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have emerged to provide post-hoc explanations for complex models (Lundberg and Lee, 2017; Slack et al., 2020). However, these methods typically deliver explanations after predictions are made, making them less suitable for real-time monitoring and immediate decision-making in industrial scenarios.

Co-processing at the FCC is a challenging industrial process characterized by continuous and dynamic changes in operational conditions over time (Fogassy et al., 2011). This dynamic nature inherently leads to fluctuations in CO_2 emissions. However, previous studies have largely overlooked the impact of changing dynamics on emission modeling. Conventional approaches often fail to account for complex interactions among real-time operational variations, such as changes in feedstock composition, reaction temperatures, and catalyst performance—which can significantly affect the accuracy of emission predictions (Su et al., 2022; Cao et al., 2024). Figure 3 provides an illustrative example of how data and their distribution can evolve over time due to underlying process dynamics. The parameters P_A, P_B, P_C, P_D represent the probability distributions of the process variables within four time segments (A, B, C, D). Each segment corresponds to a distinct operational regime, which might reflect subtle changes in the statistical characteristics of the process data, such as shifts in mean, variance, or underlying distributions. When faced with such dynamic behaviors, a static model trained on data from a single operating regime will likely perform poorly when applied to new conditions with different underlying relationships.

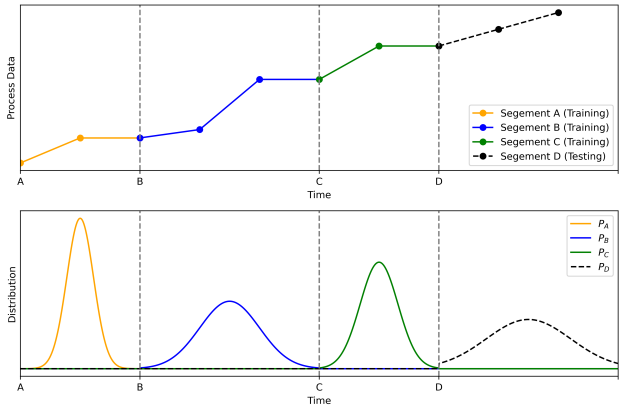

Figure 3: Example of data distribution changes over time due to process dynamics

The limitations of current approaches highlight several critical needs in co-processing modeling: the ability to handle large-scale industrial datasets, account for dynamic operating conditions, provide real-time monitoring capabilities, and adapt to changing process conditions automatically. Traditional models often fail to capture the temporal evolution of process parameters and their complex interactions, leading to reduced accuracy in industrial applications. Furthermore, the lack of comprehensive validation against industrial-scale experimental data has hindered the development of reliable predictive models for renewable CO_2 emissions.

To overcome these limitations, this study proposes an innovative data-driven approach aimed at precise modeling and real-time monitoring of renewable CO_2 emissions during the co-processing process. Our industrial partner, Parkland Refining Ltd in Canada, Canada's largest renewable fuel producer, is actively co-processing oleochemical/lipid feedstocks such as tallow, canola oil, and tall oil, thereby reducing the carbon intensities (CI) of various fuels produced (Parkland, 2022). Parkland Refining Ltd has provided a large amount of valuable data from multiple operational scenarios, laying the crucial foundation for our model development.

3. Methods

3.1. Data-Driven Modeling of Renewable CO_2

The continuous processing of biogenic feedstocks by refineries generates an abundance of data. Notably, we already have the ability to measure Total CO_2 emissions in real time, but the amount of renewable CO_2 can only be calculated through sampling via AMS ^{14}C. We propose the integration of machine learning methodologies to first establish a reliable and accurate real-time model to predict Total CO_2 emissions. Once we have this model in place, we can further analyze it to estimate the renewable CO_2 emissions. For the Total CO_2 emissions, we assumed that Total CO_2 is a linear combination of the input variables, i.e.,

$$\text{Total } CO_2 = \underbrace{\boldsymbol{a} \cdot fossil\ feed + \boldsymbol{b} \cdot bio\ feed}_{CO_2(fossil,\ bio)} + \varepsilon(fossil,\ bio) \tag{1}$$

The predicted Total CO_2 is further decomposed into two components. $CO_2(fossil, bio)$ represents the main contribution of fossil feed and bio feed to the CO_2 production, while $\varepsilon(fossil, bio)$ accounts for the additional contribution or adjustment to the CO_2 production that is not directly explained by the fossil feed and bio feed. $\varepsilon(fossil, bio)$ is defined as follows:

$$\varepsilon(fossil,\ bio) = \boldsymbol{c} \cdot feature\ I + \boldsymbol{d} \cdot feature\ II + \cdots \tag{2}$$

where *fossil feed, bio feed, feature I, feature II, feature III, etc.* represent the selected variables in the co-processing, while ***a,b,c,d etc.*** stand for the coefficients in the linear regression model that are determined through data analysis.

Defining the co-processing ratio, r_{co}, as the proportion of bio feed to the combined amount of bio feed and fossil feed

(bio feed / (bio feed + fossil feed)), it's noteworthy that the renewable CO_2 is only related to the bio feed and the portion of ε(fossil, bio) attributable to the bio feed. The core idea is to proportionally allocate the "additional contribution" term, ε (fossil, bio), to the renewable CO_2 component based on the co-processing ratio (r_{co}). Specifically, we assume that ε(fossil, bio) is influenced by both fossil and bio feeds, and its contribution to renewable CO_2 should be proportional to the bio feed's contribution to the total feed input. This assumption allows us to decompose the total CO_2 emissions into renewable and non-renewable parts in a manner that reflects the underlying process dynamics. Thus, we can further derive the formula for renewable CO_2 as follows:

$$Renewable\ CO_2 = \boldsymbol{b} \cdot bio\ feed + r_{co} \cdot \varepsilon(fossil,\ bio) \quad (3)$$

To clarify, the first term, $\boldsymbol{b} \cdot$bio feed, represents the direct contribution of bio feed to CO_2 emissions, which is entirely renewable. The second term, $r_{co} \cdot \varepsilon$(fossil, bio), allocates a portion of the additional CO_2 captured by ε to the renewable component, proportional to the bio feed's share in the total feed. This proportional allocation is a reasonable approach in the absence of detailed chemical reaction mechanisms that could precisely separate the contributions from fossil and bio feeds.

To facilitate comparison with ^{14}C measurement results, we introduce the biogenic fraction, which represents the renewable CO_2 emissions relative to the total emissions. The biogenic fraction is calculated as the ratio of renewable CO_2 to total CO_2, and can be expressed as follows:

$$Biogenic\ Fraction = \frac{Renewable\ CO_2}{\text{Total } CO_2} \quad (4)$$

This formula not only allows us to compare the biogenic content with ^{14}C measurement results but also enables real-time tracking of the proportion of renewable CO_2 in total emissions, providing more accurate monitoring of biogenic content in the process. To get a simplified, interpretable, and robust online renewable CO_2 monitor, we propose to use robust regression. Robust regression is a form of regression analysis that is designed to be resistant to outliers and can effectively handle noisy data, which is common in industrial processes (Cao et al., 2024; Yu and Yao, 2017). It provides more reliable and accurate estimates compared to ordinary least squares regression when there are outliers or violations of assumptions.

While the robust regression approach provides a static model for renewable CO_2 emissions, co-processing involves various dynamic operating conditions. If a static model is used, it may not accurately capture the variations in emissions under different operating conditions, as each condition may require a different model. To address this limitation, we will develop a data-driven dynamic modeling approach for renewable CO_2 emissions, which can adapt to changing operating conditions and provide more accurate and reliable predictions.

Remark: In this study, we deliberately chose a linear modeling approach for Total CO_2 emissions because of its interpretability and alignment with the physical characteristics of the process. Advanced time-series models like recurrent neural network (RNN) or Transformer might achieve higher predictive accuracy, but they function as black-box models and lack the ability to disentangle Total CO_2 emissions into renewable and non-renewable components (LeCun et al., 2015; Wang et al., 2023). In contrast, our robust linear regression model yields explicit coefficients that clearly represent the contribution of bio feed and fossil feed, thereby facilitating a precise decomposition of total CO_2 into renewable and non-renewable components. Moreover, many regulatory agencies require transparent, auditable methods for carbon accounting, and a linear decomposition approach is more easily justified and validated in these settings (Su et al., 2021).

3.2. Change Point Detection in FCC Co-processing

In FCC co-processing, it is often impossible to determine in advance how many different operating conditions exist. These conditions can change at any time due to various factors, such as variations in reactor feed composition, catalyst activity, operating temperature, pressure, and other parameters. As a result, a data-driven approach is necessary to automatically discover these different operating conditions without relying on prior knowledge. To address this challenge, we use change point detection techniques to automatically identify different operating conditions (Truong et al., 2020; Takeuchi and Yamanishi, 2006). The core idea of change point detection is to detect moments in an observed data sequence where statistical properties change abruptly. These moments, known as change points, allow us to segment the data into relatively stable periods, each representing a potential operating condition.

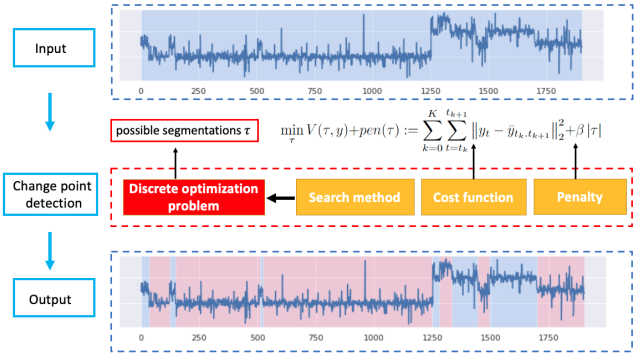

Figure 4: Example of change point detection applied to time-series data

Change point detection offers several key advantages. First, it exhibits strong adaptability, as it can automatically discover changes in statistical properties from the data without requiring manual specification of the number or types of operating conditions. Second, it requires no prior knowledge—only the observed data itself—making it suitable for complex industrial processes. Third, it dynamically captures changes in real-time, ensuring accurate monitoring of operating conditions as they evolve. Finally, being a fully

data-driven method, change point detection can leverage large amounts of industrial operation data to uncover hidden patterns and regularities. Figure 4 gives an example of change point detection. By applying change point detection, we can automatically segment the data, with each segment representing a potential operating condition. This segmentation enables the development of separate models for each operating condition, leading to more accurate predictions and better monitoring of renewable CO_2 emissions.

Let the signal $z = \{z_1, z_2, \dots, z_T\}$ be defined, and assume that it is piecewise stationary, with the process changing abruptly at some unknown time points $s_{1:K} = \{s_1, s_2, \dots, s_K\}$, where $s_1 < s_2 < \cdots < s_K$. The goal of change point detection is to estimate these instants when the signal z is observed. Define $W(\sigma, z)$ as the total cost for choosing segmentation σ, which is given as follows:

$$W(\sigma, z) = \sum_{k=0}^{K} d\left(z_{s_k, s_{k+1}}\right) = d\left(\{z_\sigma\}_1^{s_1}\right) + d\left(\{z_\sigma\}_{s_1+1}^{s_2}\right) + \cdots + d\left(\{z_\sigma\}_{s_{K-1}+1}^{s_K}\right) \tag{5}$$

where $\{s_1, s_2, \dots, s_k, \dots\}$ represents the change point indices, and $d(z_{s_k, s_{k+1}})$ is the cost function for segment $z_{s_k, s_{k+1}} = \{z_{s_k}, z_{s_k+1}, \dots, z_{s_{k+1}}\}$, with K being the number of change points. The segment cost $d(z_{s_k, s_{k+1}})$ tends to be low if the segment is homogeneous (without internal change points) and high if the segment is heterogeneous (with internal change points). Various cost functions have been proposed, such as the L_1-norm cost, L_2-norm cost, Poisson cost, kernel-based cost, etc. In this work, the segment cost function $d(z_{s_k})$ is defined using the L_2-norm:

$$d\left(z_{s_k, s_{k+1}}\right) = \sum_{t=s_k}^{s_{k+1}} \|z_t - \bar{z}_{s_k, s_{k+1}}\|_2^2 \tag{6}$$

where $\bar{z}_{s_k, s_{k+1}}$ is the mean of the segment $z_{s_k, s_{k+1}}$. Since the number of change points K is unknown in FCC co-processing, a regularizer pen$(\cdot)$ on the number of segments is required to avoid overfitting. The choice of penalty is closely related to the magnitude of detected changes. A small penalty will result in detecting many change points, even those caused by noise, while a large penalty will only detect a few significant change points or possibly none. In this work, a linear penalty pen(σ) is chosen, and it is defined as:

$$\text{pen}(\sigma) = \gamma\, |\sigma| \tag{7}$$

where γ represents the regularization coefficient, and a smaller γ results in a weaker penalty. In this work, the Pruned Exact Linear Time (PELT) method with a linear penalty is employed (Wambui et al., 2015). We selected the PELT algorithm for it has $O(T)$ computational efficiency and is effective for large-scale data, which is well-suited for our application involving extensive industrial datasets. Finally, for the signal $z = \{z_1, z_2, \dots, z_T\}$, the problem of detecting change points can be framed as a discrete optimization problem:

$$\min_{\sigma} W(\sigma, z) + \text{pen}(\sigma) = \sum_{k=0}^{K} \sum_{t=s_k}^{s_{k+1}} \|z_t - \bar{z}_{s_k, s_{k+1}}\|_2^2 + \gamma\, |\sigma| \tag{8}$$

While the change point detection is an established method, its application in this context is novel. Previous studies on co-processing have largely overlooked the dynamic nature of industrial operations. Our work addresses this gap by using CPD to automatically identify and adapt to changing process conditions, a critical step for accurate real-time monitoring of renewable CO_2 emissions. We performed careful parameter tuning, combining data-driven optimization with domain expertise from Parkland Refining engineers, to ensure the detected change points were both statistically sound and operationally relevant. Moreover, we integrate the CPD-based segmentation with a robust regression scheme to build dynamic models that capture the non-stationary relationships between feedstocks and CO_2 emissions.

3.3. Data-Driven Dynamic Modeling of Renewable CO_2

Based on the segmentation of operating conditions obtained through change point detection, we develop a data-driven dynamic modeling approach to accurately predict renewable CO_2 emissions in the FCC co-processing unit. The primary objective is to capture the evolving relationships between process variables and CO_2 emissions under varying operational scenarios. Our approach involves constructing segment-specific regression models tailored to the unique characteristics of each identified operating condition.

While our approach primarily relies on change point detection for temporal segmentation rather than explicit sequential modeling, it inherently accounts for the dynamic nature of industrial data through segment-specific local models. Each segment represents a distinct temporal regime with its own dynamic characteristics, effectively capturing different phases of the process over time. This piecewise approach to modeling temporal dynamics offers significant advantages in industrial settings where process behaviors can shift abruptly due to operational changes, feed composition variations, or equipment modifications. By developing separate models for each operational mode, we capture the distinct dynamic relationships that exist within each regime, which might otherwise be obscured in a single global model attempting to represent all possible states.

The dynamic modeling procedure consists of several integrated steps. First, we utilize the change point detection algorithm, such as PELT with a linear penalty, to partition the time-series data into distinct segments. Each segment represents a stable operating condition with consistent statistical properties, allowing for more accurate modeling within homogeneous periods. We then employ robust regression techniques to build predictive models for Total CO_2 emissions within each segment. The local robust regression models trained on each contiguous segment inherently

encode temporal correlations within that regime. In effect, our piecewise framework captures the evolving dynamics of industrial sequences without requiring end-to-end time-series networks. The general form of the regression model for segment k is given by:

$$\begin{aligned}\text{Total CO}_{2t} = a_k \cdot \text{fossil feed}_t + b_k \cdot \text{bio feed}_t \\ + c_k \cdot \text{feature I}_t + d_k \cdot \text{feature II}_t + \dots\end{aligned} \tag{9}$$

where a_k, b_k are the regression coefficients for fossil feed, bio feed, respectively. Subsequently, we decompose the predicted total CO_2 emissions into contributions from fossil fuels and biofeedstock to estimate renewable CO_2 emissions. The renewable CO_2 at each time point t in segment k is estimated using:

$$\textit{Renewable CO}_{2t} = b_k \cdot \text{bio feed}_t + r_{\text{co},t} \cdot \varepsilon_t \tag{10}$$

Building upon the segmentation of operating conditions obtained through change point detection, we develop a data-driven dynamic modeling approach to accurately predict renewable CO_2 emissions in the FCC co-processing unit. The primary objective is to capture the evolving relationships between process variables and CO_2 emissions under varying operational scenarios.

An overview of the dynamic modeling approach is illustrated in Figure 5. This figure demonstrates the workflow from data segmentation to model deployment. The algorithmic implementation is detailed in Algorithm 1, which outlines the sequential steps involved in change point detection, model development, and real-time estimation of renewable CO_2 emissions.

Algorithm 1: Dynamic Renewable CO_2 Monitoring Procedure

Input: A time series $X = [x_0, x_1, \dots, x_T]$, where each x_t includes: Fossil feed rate (fossil feed$_t$), bio feed rate (bio feed$_t$), Additional process feature, Measured Total CO_2 emissions

Hyperparameters: Regularization coefficient γ, Maximum number of change points $N_{\max}$, Minimum segment length $L_{\min}$

Output: Change point list CP; Segment models list M; Renewable CO_2 estimates Renewable CO_2

begin
- Initialize change point list CP ← [ ]
- Initialize segment models list M ← [ ]
- **while** *len(CP)* $< N_{max}$ **do**
  - Set $s \leftarrow$ CP[−1]
  - **if** $s + L_{min} > T$ **then**
    - **break**
  - **end**
  - new_CP ← Detect Next Change Point
  - **if** *new_CP = None* **then**
    - **break**
  - **end**
  - Append new_CP to CP
  - Extract segment data $X_{\text{segment}} \leftarrow X[s : \text{new\_CP}]$
  - Build Total_CO_2 robust regression model m on X_{segment}
  - Append model m to M
- **end**
- **for** *t from* 0 *to* T **do**
  - Identify current segment S_k such that $t \in [\text{CP}[k], \text{CP}[k+1])$
  - Retrieve model m_k from M
  - Compute predicted Total CO_2 emissions: $\text{Total CO}_{2t} = a_k \cdot \text{fossil feed}_t + b_k \cdot \text{bio feed}_t + c_k \cdot \text{feature I}_t + d_k \cdot \text{feature II}_t + \dots$
  - Calculate co-processing ratio: $r_{\text{co},t} = \dfrac{\text{bio feed}_t}{\text{bio feed}_t + \text{fossil feed}_t}$
  - Compute adjustment term: $\varepsilon_t = c_k \cdot \text{feature I}_t + d_k \cdot \text{feature II}_t + \dots$
  - Estimate renewable CO_2 emissions: $\text{Renewable CO}_{2t} = b_k \cdot \text{bio feed}_t + r_{\text{co},t} \cdot \varepsilon_t$
- **end**
- **return** CP, M, Renewable CO_2

end

4. Results

In this section, we present the results of applying our data-driven dynamic modeling approach to the real-time monitoring of renewable CO_2 emissions in the FCC co-processing unit at the Parkland Refinery. We utilized a comprehensive dataset comprising 43,662 samples collected from March 2021 to December 2023 during the refinery's commercial operation. The data captured a wide range of operational conditions, including transitions between different feedstock compositions and process settings. It includes measurements of Total CO_2 emissions, fossil feed rates, bio feed rates, reaction temperatures, catalyst, and other relevant process features. Figure 6 visualizes key variables over time, highlighting fluctuations that reflect the dynamic operating conditions inherent in co-processing.

4.1. Data Preprocessing

To ensure data quality and robustness for subsequent modeling, the dataset should be subjected to a comprehensive preprocessing pipeline (Cao et al., 2025). This pipeline addressed missing data, outliers, and feature scaling, as these issues are common in industrial datasets due to sensor errors, process interruptions, or inconsistent measurements. The preprocessing steps were implemented using Python's scikit-learn library, and their parameters were optimized using a validation subset of the dataset (10% of samples) to minimize bias and ensure generalizability.

Missing values were imputed using the k-nearest neighbors (k-NN) imputation (Pujianto et al., 2019) method with $k = 5$. This approach imputes missing values by computing the weighted average of the five most similar samples based on Euclidean distance in the feature space (excluding the missing feature). The k-NN method was chosen for its robustness to non-linear relationships in industrial data and its ability to preserve local data structures.

Outliers, which can arise from sensor malfunctions or extreme process conditions (e.g., anomalous CO_2 measurements), were detected using the isolation forest algorithm (Thennadil et al., 2018). This method identifies anomalies by constructing random decision trees and isolating samples with shorter average path lengths, which indicate outliers. The parameter of the Isolation Forest was optimized based on visual inspection of key features and domain knowledge from Parkland Refining engineers to avoid removing valid extreme values.

To ensure consistent scales across heterogeneous features, all features were normalized to the range [0, 1] using min-max scaling. Min-max scaling was chosen to preserve the relative relationships between data points while ensuring compatibility with the robust regression models used in subsequent steps.

4.2. Feature Importance Analysis

To systematically analyze the contributions of each feature to Total CO_2 predictions, we conducted a feature importance analysis using various methodologies. The methods include Mutual Information (Jiang et al., 2019), Elastic Net (Sun and Braatz, 2020), Boruta (Kursa and Rudnicki, 2010), Causal Discovery (Su et al., 2022), Random Forest (Biau, 2012), LASSO (Meinshausen and Bühlmann, 2006), XGBoost, LightGBM, CatBoost, and Gradient Boosting (Brownlee, 2021).

Features such as fossil feed, bio feed, catalyst circulation rate and catalyst cooler steam flow consistently demonstrate high importance across all methods. This finding highlights their critical role in predicting Total CO_2. Moreover, features such as reactor temperature and upper regenerator temperature also contribute significantly under certain methodologies, indicating their relevance to specific operating conditions.

Industrial processes may involve unobserved variables such as catalyst aging or feedstock impurities. We mitigate this by focusing on high-impact measured features and by segmenting the process such that each local model remains statistically consistent. Nevertheless, we acknowledge that certain confounding factors outside our measured feature set could still influence CO_2 formation. The residual term ε_t partly accounts for these unmodeled variations, but further investigations on advanced sensor deployment or domain-specific knowledge could help refine the model.

4.3. Change Point Detection

To capture dynamic changes in operating conditions, we applied the PELT algorithm for change point detection on the time series data of Total CO_2 emissions and key process variables. This method dynamically balances the trade-off between model complexity and goodness of fit, ensuring statistically robust segmentations. The regularization coefficient γ was carefully tuned based on operational knowledge and data characteristics to avoid overfitting or underfitting the change points. The algorithm identified 14 significant change points, effectively segmenting the dataset into 15 distinct operational periods, as illustrated in Figure 7. Despite some segments appearing to have less noticeable differences, the segmentation is statistically reliable, as it minimizes the cost function defined in Equation (8).

To objectively evaluate the accuracy of our change point detection approach, we implemented multiple quantitative assessment methods. First, we measured how well the PELT algorithm minimized the statistical cost function in Equation (8). This indicates a strong balance between segment homogeneity and model parsimony. Second, we conducted a sensitivity analysis by varying the regularization coefficient to produce the most stable segmentation when comparing the detected change points against reference points identified from process operation logs.

Additionally, these change points correspond to operational shifts validated by domain knowledge, such as feedstock adjustments, temperature changes, and catalyst variations. This alignment with real-world operational patterns further supports the reliability of the detected change points. The average duration between change points was approximately 1-2 months, underscoring the necessity for a dynamic modeling approach that can adapt to these temporal variations.

4.4. Total CO_2 Emissions Modeling Results

For each identified segment, we developed robust regression models to predict Total CO_2 emissions based on the fossil feed rate, bio feed rate, and additional process features identified through our feature importance analysis. The regression coefficients a_k and b_k for fossil feed and bio feed, respectively, were calculated for each segment, capturing the unique relationships between inputs and emissions under different operating conditions. To ensure that each segment-specific model accurately reflects the relationships within its respective segment, we assess the performance using statistical metrics such as Mean Squared Error (MSE) and the coefficient of determination (R^2).

To comprehensively evaluate the performance of our proposed change point detection-based dynamic modeling approach, we conducted comparisons with alternative modeling methods, including both static models that ignore temporal segmentation and other dynamic approaches that capture multimode behaviors.

Table 1 presents a comprehensive comparison of various modeling approaches evaluated on the Parkland refinery dataset. For static baseline methods, we implemented linear regression (LR), support vector regression (SVR), random forest (RF), and recurrent neural networks (RNN).

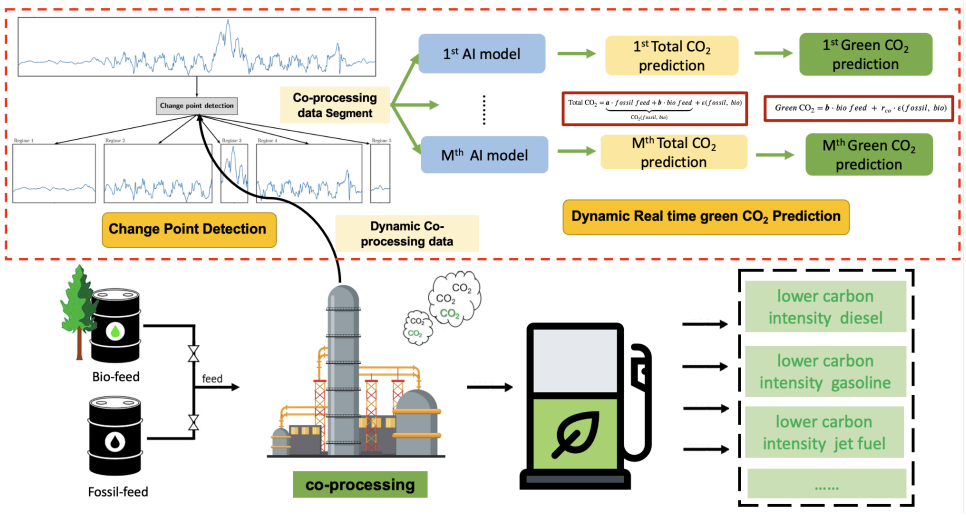

Figure 5: Flowchart of the dynamic modeling for online monitoring of renewable CO_2 emissions

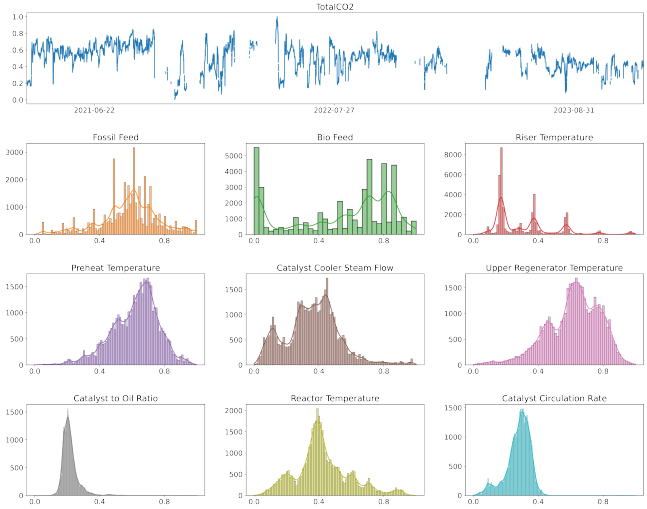

Figure 6: Visualization of process variables over time from the Parkland refinery dataset

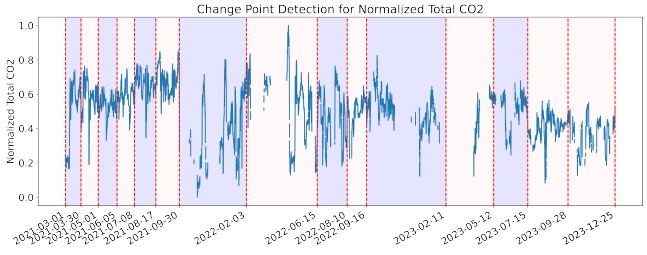

Figure 7: Change point detection results for different operating conditions

For dynamic multimode approaches, we implemented segmented partial least squares (PLS), Gaussian mixture model with linear regression (GMM+LR), hidden Markov model with linear regression (HMM+LR), and a hybrid approach combining Gaussian mixture models with recurrent neural networks (GMM+RNN).

While GMM+RNN achieved the lowest MSE of 0.038 and highest R^2 of 0.937, it came with significant trade-offs in interpretability and computational efficiency. Our proposed approach demonstrated excellent predictive performance with an MSE of 0.041 and R^2 of 0.933, positioning it as the second-best method in terms of pure numerical accuracy. However, our approach substantially outperformed all other methods when considering the full spectrum of desirable model characteristics.

The strong predictive performance of our segment-specific models serves as an indirect validation of the accuracy of our change point detection approach. The relatively low MSE (0.041) and high R^2 (0.933) values suggest that the identified segments represent genuinely distinct operational regimes with consistent statistical properties. Inaccurate segmentation, whether through excessive partitioning of homogeneous data or through failure to identify significant operational transitions, would manifest as diminished predictive performance in subsequent regression models. The observed high performance metrics therefore provide substantive evidence supporting the statistical validity of our segmentation methodology. This performance-based validation complements the direct statistical evaluation of the change point detection algorithm and provides further evidence of the robustness of our approach.

The key advantages of our approach extend beyond numerical performance metrics.

First, the linear structure of our segment-specific models provides high interpretability, which is crucial for industrial applications requiring decomposition of Total CO_2 into renewable and non-renewable components. This transparency is essential for regulatory compliance and process optimization, whereas black-box models like SVR, RNN, and GMM+RNN cannot readily provide such insights.

Second, unlike engineering-based PLS segmentation that relies on domain knowledge, our approach automatically identifies change points through statistical optimization,

Table 1
Performance comparison of various modeling approaches for normalized Total CO_2

Method	MSE	R^2	Interpretability	Noise Robustness	Computational Efficiency
Static Models					
LR	0.059	0.857	High	Low	High
SVR	0.052	0.883	Low	Medium	Medium
RF	0.048	0.902	Medium	High	Medium
RNN	0.046	0.915	Very Low	Medium	Low
Dynamic Models					
PLS (Segmentation)	0.045	0.910	Medium	Medium	Medium
GMM+LR	0.044	0.924	Medium	Medium	Low
HMM+LR	0.043	0.928	Low	Medium	Low
GMM+RNN	**0.038**	**0.937**	Very Low	Medium	Very Low
Our Approach	0.041	0.933	**High**	**High**	**High**

making it adaptable to unforeseen operational shifts. This feature is particularly valuable in dynamic industrial environments where operating conditions change frequently and unpredictably.

Third, the PELT algorithm's computational complexity enables efficient processing of large industrial datasets containing tens of thousands of samples. This efficiency, combined with the robust regression component that effectively handles outliers common in industrial data, makes our approach both practical and reliable for real-time monitoring applications.

4.5. Renewable CO_2 Emissions Modeling Results

Based on the total CO_2 emissions modeling results, we can further model the renewable CO_2 emissions. Table 2 summarizes the proposed model performance and coefficients across different segments. The results indicate that the bio-to-fossil CO_2 emission ratio fluctuates significantly between 51.9% and 82.3%, depending on the operational conditions. This variation underscores the complexity of the co-processing environment, where factors such as feedstock composition, temperature, and catalyst performance interact dynamically. The higher bio-to-fossil CO_2 emission ratios in certain segments suggest that under certain conditions, biofeedstock distribute more to renewable CO_2 emissions (less renewable carbon in the desired liquid fractions). This finding is particularly relevant for optimizing co-processing operations to maximize renewable output, which could lead to higher renewable content in the gasoline and diesel produced.

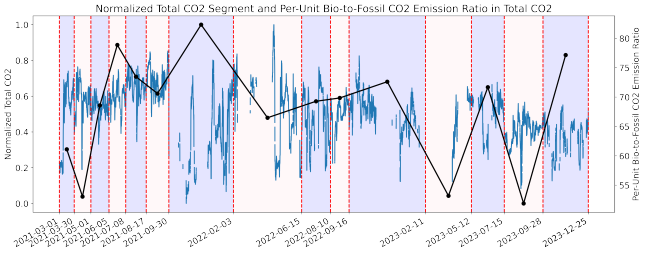

Figure 8: Normalized Total CO_2 segment and per-unit bio-to-fossil CO_2 emission ratio

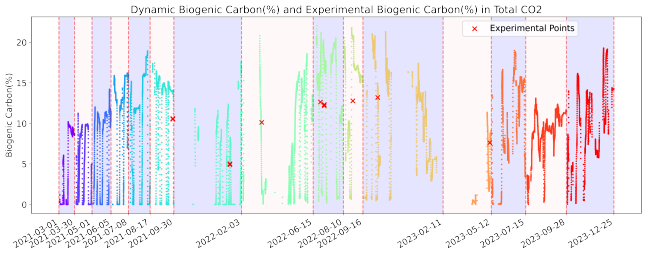

Figure 9: Comparison of dynamic biogenic carbon model predictions with experimental data

The normalized Total CO_2 segments and per-unit bio-to-fossil CO_2 emission ratios are shown in Figure 8, providing a clearer comparison of biofeedstock contributions under different operating conditions. The real-time ratio of renewable CO_2 in Total CO_2, illustrated in Figure 9, exhibits significant fluctuations over time, with different segments represented by different colors. By integrating the real-time ratio curve, the total renewable CO_2 emissions can be obtained, providing crucial support for evaluating emission reduction efforts.

To validate the model's accuracy and flexibility, we compared the predicted renewable CO_2 emissions with experimental measurements obtained from ^{14}C sampling at specific time points. The results in Table 3 demonstrate that the model consistently achieves high accuracy, with prediction errors predominantly below 5%, even during transitional periods with significant changes in operational parameters. This indicates the system's robustness in handling dynamic conditions and its ability to provide reliable real-time monitoring.

Figure 9 illustrates the close alignment between the model predictions and experimental data. To assess the advantages of our dynamic approach, we also developed a static model using the entire dataset without segmentation. Table 3 presents the prediction errors for both models against the experimental data. The dynamic model consistently exhibited lower prediction errors, with most errors below 5%, whereas the static model showed significantly higher errors. The robustness of the proposed model is evidenced by its consistent performance across various operational segments

Table 2
Model performance for normalized Total CO_2 and bio-to-fossil conversion ratios in different segments

Segment Start	Segment End	MSE	R^2	Fossil Feed Coefficient a_k	Bio Feed Coefficient b_k	Bio-to-Fossil Ratio
2021-03-01	2021-03-30	0.015	0.990	0.497	0.303	0.610
2021-03-30	2021-05-01	0.023	0.948	0.345	0.184	0.532
2021-05-01	2021-06-05	0.021	0.834	0.783	0.537	0.686
2021-06-05	2021-07-08	0.020	0.893	0.608	0.481	0.792
2021-07-08	2021-08-17	0.016	0.919	0.642	0.472	0.736
2021-08-17	2021-09-30	0.025	0.897	0.626	0.443	0.707
2021-09-30	2022-02-03	0.033	0.979	0.263	0.217	0.823 (Max)
2022-02-03	2022-06-15	0.158	0.923	0.672	0.447	0.665
2022-06-15	2022-08-10	0.013	0.986	0.612	0.424	0.693
2022-08-10	2022-09-16	0.014	0.923	0.359	0.251	0.699
2022-09-16	2023-02-11	0.042	0.887	0.430	0.313	0.726
2023-02-11	2023-05-12	0.019	0.973	0.351	0.188	0.536
2023-05-12	2023-07-15	0.032	0.910	0.460	0.330	0.718
2023-07-15	2023-09-28	0.018	0.927	0.630	0.327	0.519 (Min)
2023-09-28	2023-12-25	0.019	0.950	0.409	0.315	0.771

Table 3
Comparison of model predictions with experimental data

Test Time	True Value	Proposed Model	Proposed Model Error (%)
2021-09-28	10.59	10.66	0.66
2021-09-28	10.51	10.50	0.10
2022-01-12	4.90	5.17	5.51
2022-01-12	5.01	5.29	5.59
2022-03-23	10.12	10.56	4.35
2022-06-29	12.75	13.15	3.14
2022-06-29	12.63	12.91	2.22
2022-07-06	12.32	12.29	0.24
2022-07-06	12.17	11.99	1.48
2022-08-30	12.78	12.16	4.85
2022-10-27	13.19	13.08	0.83
2023-01-18	8.07	7.84	2.85
2023-01-18	8.62	7.94	7.88
2023-05-10	10.51	10.10	3.90
2023-05-09	7.62	7.32	3.93

and its ability to adapt to changes without manual intervention.

5. Conclusion

This study developed a data-driven dynamic modeling approach to accurately predict renewable CO_2 emissions in biofeedstock co-processing. By integrating change point detection with robust regression techniques, we successfully captured complex relationships between process variables and emissions under varying operational conditions. Compared to other dynamic modeling techniques, our approach provides superior balance of predictive accuracy, interpretability, and computational efficiency—critical factors for industrial implementation. Through extensive modeling with industrial data samples and rigorous validation against multiple ^{14}C measurements, our method achieved prediction errors predominantly below 5%. It offers robust support for refineries to optimize co-processing operations and comply with regulatory mandates for renewable fuel production. Future research should expand data sources to include more refineries and diverse operational conditions to enhance the model's generalizability.

References

M. Ebadian, S. van Dyk, J. D. McMillan, J. Saddler, Biofuels policies that have encouraged their production and use: An international perspective, Energy Policy 147 (2020) 111906.

S. Yeh, J. Witcover, G. E. Lade, D. Sperling, A review of low carbon fuel policies: Principles, program status and future directions, Energy Policy 97 (2016) 220–234.

N. Gray, S. McDonagh, R. O'Shea, B. Smyth, J. D. Murphy, Decarbonising ships, planes and trucks: An analysis of suitable low-carbon fuels for the maritime, aviation and haulage sectors, Advances in Applied Energy 1 (2021) 100008.

J. Li, D. Yu, L. Pan, X. Xu, X. Wang, Y. Wang, Recent advances in plastic waste pyrolysis for liquid fuel production: Critical factors and machine learning applications, Applied Energy 346 (2023) 121350.

S. Bezergianni, A. Dimitriadis, O. Kikhtyanin, D. Kubička, Refinery co-processing of renewable feeds, Progress in Energy and Combustion Science 68 (2018) 29–64.

S. Badoga, A. Alvarez-Majmutov, T. Xing, R. Gieleciak, J. Chen, Co-processing of hydrothermal liquefaction biocrude with vacuum gas oil through hydrotreating and hydrocracking to produce low-carbon fuels, Energy & Fuels 34 (2020) 7160–7169.

S. Van Dyk, J. Su, J. D. Mcmillan, J. J. Saddler, Potential synergies of drop-in biofuel production with further co-processing at oil refineries, Biofuels, Bioproducts and Biorefining 13 (2019) 760–775.

X. Han, H. Wang, Y. Zeng, J. Liu, Advancing the application of bio-oils by co-processing with petroleum intermediates: a review, Energy Conversion and Management: X 10 (2021) 100069.

A. H. Zacher, D. C. Elliott, M. V. Olarte, H. Wang, S. B. Jones, P. A. Meyer, Technology advancements in hydroprocessing of bio-oils, Biomass and Bioenergy 125 (2019) 151–168.

S. D. Stefanidis, K. G. Kalogiannis, A. A. Lappas, Co-processing bio-oil in the refinery for drop-in biofuels via fluid catalytic cracking, Wiley Interdisciplinary Reviews: Energy and Environment 7 (2018) e281.

M. S. Talmadge, R. M. Baldwin, M. J. Biddy, R. L. McCormick, G. T. Beckham, G. A. Ferguson, S. Czernik, K. A. Magrini-Bair, T. D. Foust,

P. D. Metelski, et al., A perspective on oxygenated species in the refinery integration of pyrolysis oil, Green Chemistry 16 (2014) 407–453.

T. M. Lammens, Effect of various green carbon tracking methods on life cycle assessment results for fluid catalytic cracker co-processing of fast pyrolysis bio-oil, Energy & Fuels 36 (2022) 12617–12627.

J. Su, L. Cao, G. Lee, B. Gopaluni, L. C. Siang, Y. Cao, S. van Dyk, R. Pinchuk, J. Saddler, Tracking the green coke production when co-processing lipids at a commercial fluid catalytic cracker (fcc): combining isotope 14 c and causal discovery analysis, Sustainable Energy & Fuels 6 (2022) 5600–5607.

L. Cao, J. Su, J. Saddler, Y. Cao, Y. Wang, G. Lee, L. C. Siang, R. Pinchuk, J. Li, R. B. Gopaluni, Real-time tracking of renewable carbon content with ai-aided approaches during co-processing of biofeedstocks, Applied Energy 360 (2024) 122815.

D. C. Elliott, T. R. Hart, G. G. Neuenschwander, L. J. Rotness, M. V. Olarte, A. H. Zacher, Y. Solantausta, Catalytic hydroprocessing of fast pyrolysis bio-oil from pine sawdust, Energy & Fuels 26 (2012) 3891–3896.

S. Dell'Orco, E. D. Christensen, K. Iisa, A. K. Starace, A. Dutta, M. S. Talmadge, K. A. Magrini, C. Mukarakate, Online biogenic carbon analysis enables refineries to reduce carbon footprint during coprocessing biomass-and petroleum-derived liquids, Analytical Chemistry 93 (2021) 4351–4360.

F. Pernkopf, D. Bouchaffra, Genetic-based em algorithm for learning gaussian mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 1344–1348.

O. Geramifard, J.-X. Xu, J.-H. Zhou, X. Li, Multimodal hidden markov model-based approach for tool wear monitoring, IEEE Transactions on Industrial Electronics 61 (2013) 2900–2911.

J. Liu, V. Kadirkamanathan, Hidden markov model-based fault detection and diagnosis, Annual Reviews in Control 39 (2015) 1–16.

L. Cao, X. Ji, Y. Cao, B. Gopaluni, Adaptive process monitoring for multimode industrial processes through machine learning, IEEE Journal of Emerging and Selected Topics in Industrial Electronics (2025) 1–9.

Y. Zhang, S. Li, Modeling and monitoring of nonlinear multi-mode processes, Control Engineering Practice 22 (2014) 194–204.

G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed., John Wiley & Sons, Hoboken, NJ, 2015.

O. Barbarisi, F. Vasca, L. Glielmo, State of charge kalman filter estimator for automotive batteries, Control engineering practice 14 (2006) 267–275.

M. Du, N. Liu, X. Hu, Techniques for interpretable machine learning, Communications of the ACM 63 (2019) 68–77.

C. Rudin, Stop explaining black box models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019) 206–215.

S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, volume 30, NeurIPS, 2017, pp. 4765–4774.

D. Slack, S. Hilgard, E. Jia, S. Singh, H. Lakkaraju, Fooling lime and shap: Adversarial attacks on post hoc explanation methods, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 180–186.

G. Fogassy, N. Thegarid, Y. Schuurman, C. Mirodatos, From biomass to bio-gasoline by fcc co-processing: effect of feed composition and catalyst structure on product quality, Energy Environ. Sci. 4 (2011) 5068–5076.

Parkland, Parkland announces plans to expand co-processing activities and build british columbia's largest renewable diesel complex, 2022. URL: https://www.parkland.ca/en/investors/news-releases/details.

C. Yu, W. Yao, Robust linear regression: A review and comparison, Communications in Statistics - Simulation and Computation 46 (2017) 6261–6282.

Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

Y. Wang, J. Zhu, L. Cao, B. Gopaluni, Y. Cao, Long short-term memory network with transfer learning for lithium-ion battery capacity fade and cycle life prediction, Applied Energy 350 (2023) 121660.

J. Su, L. Cao, G. Lee, J. Tyler, A. Ringsred, M. Rensing, S. van Dyk, D. O'Connor, R. Pinchuk, J. J. Saddler, Challenges in determining the renewable content of the final fuels after co-processing biogenic feedstocks in the fluid catalytic cracker (fcc) of a commercial oil refinery, Fuel 294 (2021) 120526.

C. Truong, L. Oudre, N. Vayatis, Selective review of offline change point detection methods, Signal Processing 167 (2020) 107299.

J. Takeuchi, K. Yamanishi, A unifying framework for detecting outliers and change points from time series, IEEE Transactions on Knowledge and Data Engineering 18 (2006) 482–492.

G. D. Wambui, G. A. Waititu, A. Wanjoya, The power of the pruned exact linear time(pelt) test in multiple changepoint detection, American Journal of Theoretical and Applied Statistics 4 (2015) 581–586.

L. Cao, J. Su, E. Conde, L. C. Siang, Y. Cao, B. Gopaluni, A novel automated soft sensor design tool for industrial applications based on machine learning, Control Engineering Practice 160 (2025) 106322.

U. Pujianto, A. P. Wibawa, M. I. Akbar, et al., K-nearest neighbor (k-nn) based missing data imputation (2019) 83–88.

S. N. Thennadil, M. Dewar, C. Herdsman, A. Nordon, E. Becker, Automated weighted outlier detection technique for multivariate data, Control Engineering Practice 70 (2018) 40–49.

B. Jiang, Y. Luo, Q. Lu, Maximized mutual information analysis based on stochastic representation for process monitoring, in: IEEE Transactions on Industrial Informatics, volume 15, 2019, pp. 1579–1587. doi:10.1109/TII.2018.2853702.

W. Sun, R. D. Braatz, Alven: Algebraic learning via elastic net for static and dynamic nonlinear model identification, Computers & Chemical Engineering 143 (2020) 107103.

M. B. Kursa, W. R. Rudnicki, Feature selection with the boruta package, Journal of statistical software 36 (2010) 1–13.

G. Biau, Analysis of a random forests model, The Journal of Machine Learning Research 13 (2012) 1063–1095.

N. Meinshausen, P. Bühlmann, High-dimensional Graphs and Variable Selection with the Lasso, The Annals of Statistics 34 (2006) 1436 – 1462.

J. Brownlee, Ensemble Learning Algorithms with Python: Make better Predictions with Bagging, Boosting, and Stacking, Machine Learning Mastery, 2021.