Chemical Engineering Science 301 (2025) 120656

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

# Forecasting methanol-to-olefins product yields based on Relevance Vector Machine with hybrid kernel and rolling-windows

Wenyang Wang [a,b], Nan He [a], Jie Liu [a], Muxin Chen [a], Jibin Zhou [b], Tao Zhang [b], Mao Ye [b,*], Zhongmin Liu [b]

[a] School of Maritime Economics and Management, Dalian Maritime University, Dalian, Liaoning 116026, China
[b] Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, Liaoning 116026, China

ARTICLE INFO

*Keywords:*
Methanol-to-olefins (MTO)
Yields forecasting
Relevance Vector Machine (RVM)
Hybrid kernels
Rolling window

ABSTRACT

Light olefins (ethylene and propylene) have become prominent in chemical industries. Forecasting of the yields of light olefins plays a crucial role in monitoring and optimizing the Methanol-to-olefins (MTO) process. In this work, we introduce an approach for forecasting the yields of ethylene and propylene in the MTO process with the Relevance Vector Machine (RVM) model, which is uniquely enhanced with hybrid kernels and a rolling window methodology. Through an in-depth analysis of 32 independent variables and their pairwise differences, our research pinpoints temperature and pressure as the most critical factors influencing the yields of ethylene and propylene, respectively. The model showcases satisfactory predictive accuracy and reasonable interpretability compared with the traditional statistical and popular machine learning models, marking a step forward in the predictive modeling of chemical engineering processes.

## 1. Introduction

Light olefins, including ethylene and propylene, are pivotal chemical feedstocks that occupy significant positions in the petrochemical and organic chemical industries (Chang, 1984; Monai et al., 2021). The conventional production processes for ethylene and propylene employ naphtha cracking technology, which entails substantial crude oil consumption (Jiao et al., 2016). In contrast, Methanol-to-olefins (MTO) (Tian et al., 2015) represents an essential avenue for the efficient utilization of resources. The Dalian Institute of Chemical Physics (DICP), Chinese Academy of Sciences, has researched MTO technology, DMTO, since the 1980s. In 2010, the first industrial-scale plant using the DMTO technology was completed and operated (Ying et al., 2015). DICP continued the research and developed the DMTO-II and DMTO-III technology. Currently, more than 30 industrial MTO units have been licensed, making MTO one of the primary industrial routines for producing light olefins.

Under the framework of Industry 4.0, DMTO technology is also transforming into intelligence (Moghaddam, 2023). Exploring the deep-level relationships between industrial process inputs and outputs, conducting forecasted research on olefins production has significant practical implications for the future construction of brilliant chemical engineering models, namely the development of self-design and improvement system for DMTO technology (Wang et al., 2021). In addition, for laboratory experiments, pilot scale-up, and factory production at the current stage, mining deep-level relationships between input variables and olefins production, as well as accurate forecasting of olefins output, can provide reliable decision-making references for adjusting equipment and operational conditions (Huang et al., 2023; Abdi et al., 2023).

Classical statistical models have been frequently applied to forecast time series data. Addressing several persistent challenges is essential to construct a durable predictive statistical model. First, the variability and complexity of industrial operations often render these processes non-stationary due to dynamic changes and diverse operational conditions (Cheng et al., 2015). Second, the complexity of prediction is further compounded by the nonlinear relationships and dynamic interplays among the variables involved in these processes. Applying mathematical approaches based on reaction mechanisms and kinetics proves impractical for actual industrial applications. This impracticality stems from the indistinct nature of the physical laws linking inputs to outputs, coupled with the intricate configurations of industrial facilities (Linninger, 2002; Fan et al., 2024). Traditional statistical techniques such as the autoregressive integrated moving average (ARIMA) (Dey et al., 2023) have been widely adopted for forecasting time series data. Nevertheless, these

* Corresponding author.
*E-mail addresses:* wangwenyang@dlmu.edu.cn (W. Wang), maoye@dicp.ac.cn (M. Ye).

https://doi.org/10.1016/j.ces.2024.120656
Received 4 June 2024; Received in revised form 24 July 2024; Accepted 23 August 2024
Available online 28 August 2024

methodologies presuppose a linear association between future outcomes and past observations (De Gooijer and Hyndman, 2006), a presumption that often needing to be revised in industrial processes, leading to suboptimal predictive accuracy.

In recent years, methodologies grounded in machine learning and deep learning have become promising strategies for forecasting time series data. These approaches can autonomously discern complex input and output data relationships without requiring predefined models (Coley et al., 2017). Notably, Recurrent Neural Networks (RNN), along with its derivatives, Long Short-Term Memory (LSTM) (Zhang et al., 2019), have been recognized for their exceptional efficacy (Long et al., 2023). For instance, research by Kumar et al. (2018) highlighted the adeptness of LSTM and Gated-Recurrent-Unit (GRU) models in addressing nonlinearities and seasonal patterns in electricity demand forecasting. Similarly, a study by Wang et al. (2020b) showcased the superiority of LSTM-based frameworks over conventional methodologies in predicting the periodic energy demands of cooling systems. Despite their commendable predictive capabilities, the scalability of RNNs is somewhat constrained due to their limited parallelization potential (Hu and Zheng, 2020). Conversely, Convolutional Neural Networks (CNN) have also been extensively utilized for time series forecasting, demonstrating significant potential. Wang et al. (2020c) introduced two CNN models that leverage symbolic hierarchical clustering for predicting operational trends in methanol production units. Furthermore, attention-based models have gained prominence for their ability to assign variable importance through an attention mechanism, thereby enhancing model focus on critical data segments. For example, studies by Aliabadi et al. (2020) and Li et al. (2021) have validated the superior performance of attention-based RNNs and LSTM models in multi-step forecasting of chemical processes. The integration of attention mechanisms with CNN-LSTM by Yang et al. (2021) facilitated precise predictions of water quality indicators. A Hierarchical Attention-based Recurrent Highway Network (HRHN) proposed by Tao et al. (2018) has demonstrated remarkable accuracy in stock trend forecasting. Additionally, the self-attention mechanism has attracted significant attention for its ability to recognize patterns across extensive historical data, proving advantageous for identifying long-term dependencies. Bi and Zhao (2021) employed dual parallel self-attention layers to capture spatial correlations and temporal dependencies within time series data. Although these methodologies have achieved notable success, the above models' interpretation could be more satisfactory in indicating the significant influential factors. Also, the quest for further improvements in predictive accuracy continues unabated.

In industrial forecasting, vector machines have been widely employed for their ability to meet the accuracy and interpretability evaluation criteria (Wang et al., 2020a; Lim et al., 2021). Tipping (1999, 2001) pointed out the limitations of the Support Vector Machine (SVM) algorithm and proposed a new sparse probabilistic model called Relevance Vector Machine (RVM) based on the Bayesian framework. The main benefits of RVM are as follows (Wang et al., 2023): (1) it is more sparse and suitable for handling large datasets; (2) it offers a broader range of kernel function choices; (3) it is based on a simple linear structure, which can support straightforward model modification and improvements.

RVM regression has demonstrated remarkable achievements in chemical engineering, showcasing excellent forecast capabilities. Fang et al. (2013) proposed an integrated modeling approach based on a unified design, combining RVM with genetic algorithms. This method aimed to enhance the efficiency of microbial fuel cells in converting chemical energy from wastewater into electrical energy, effectively improving battery performance and power density. Xu et al. (2017) presented an effective forecasting interval method based on Bootstrap and RVM. This method forecasts the density of polymers in the high-density polyethylene production process, demonstrating superior accuracy and efficiency. Wang et al. (2021) established a data-driven framework based on RVM for optimizing the operation of industrial MTO processes. The optimal operating mode can be identified by associating the forecasted yield distribution of major products with operating conditions, resulting in improved ethylene yield. Acosta et al. (2021) offered an adaptive differential evolution algorithm combined with RVM for forecasting phosphorus concentration levels in steelmaking processes. The study demonstrated that RVM outperforms random forests (RF), artificial neural networks (ANN), K-nearest neighbor algorithms, and statistical learning techniques. Jiang et al. (2021) proposed a reliable cyclic aging forecasted method based on data-driven models to address the early forecasting of the remaining lifespan in lithium-ion batteries. Experimental results showed that the hybrid kernel function-based RVM accurately forecasts failure cycles and capacity degradation trajectories for different battery types. Reviewing previous literature shows three main application approaches of RVM in chemical engineering: direct utilization of the original RVM model, development of improved RVM models, and utilization of RVM in combination with other methods. This paper improves the RVM by setting the single and hybrid kernel studies for an accurate forecasting result. We also combine the RVM with a rolling window approach for a robust and efficient model, and we provide a more comprehensive research work than previous literature.

The structure of this paper is as follows: Section 2 introduces the data used in this study, Section 3 describes all models employed, Section 4 presents the empirical results, and Section 5 provides conclusions.

## 2. Data description

The data source is from one of the DMTO factories. The simplified flow diagram of the reaction and regeneration unit in a practical DMTO plant is shown in Fig. 1. This system is composed of four components, including wash tower, reactor, stripper, and stripper regenerator. First, gasified methanol enters the reactor and reacts with the catalyst (SAPO-34) to produce product gases. Next, these product gases are then sent to a product analyzer for analysis, while the coker catalyst is circulated back to the regenerator through one of the risers and reacted with air to restore activity. Finally, the regenerated catalyst is returned to the reactor through another riser (Zhou et al., 2024). Since the reaction process will be affected by a variety of environmental factors, we select 32 typical variables in the DMTO flow as the explanation information. The 2 output variables are the yields of ethylene and propylene, the main products of DMTO, shown in Fig. 1, labeled as $AI1603G$ and $AI1603I$, respectively. The data codebook lists all the variables' meanings (Table 1). The collected time is from February 28, 2021, to March 11, 2022, totaling 372 days (missing 4 days' data). The data was measured every 2 hours and produced 12 batches daily so that the dataset has $372 * 12 = 4464$ batches. The whole dataset is $\boldsymbol{S}^{total} = \{\boldsymbol{X}^{total}, \boldsymbol{Y}^{total}\}$, $\boldsymbol{X}^{total}$ is $32 \times 4464$ dimensions and $\boldsymbol{Y}^{total}$ is $2 \times 4464$ dimensions.

The output variables and the corresponding first-order difference are indicated in Fig. 2. The ethylene yield exhibits fluctuations ranging from 45% and 50%, and the propylene yield fluctuates between 28% and 31%. Besides, industrial process data presents more intricate characteristics that pose a considerable challenge for forecasting. For example, anomalies such as data drift occasionally occur. The first-order difference illustrates that the outputs are stationary. In forecasting research, we conduct the normalization for all the dependent and independent variables to eliminate the influence of different units.

### 2.1. Clustering analysis

Considering the high dimensionality, large quantity of variables, and complex characteristics and relationships between data, the clustering analysis is performed. In this paper, 32 independent variables are clustered utilizing the K-shape algorithm, where the optimal number of clusters is determined by the Elbow method. The K-shape clustering process incorporating the Elbow method is described in Appendix A, Section 3.1. K-means is often utilized for cluster analysis, but it performs poorly on data series with significant fluctuations because it employs the Euclidean Distance (ED) to measure similarity (Hartingan and Wong, 1979;

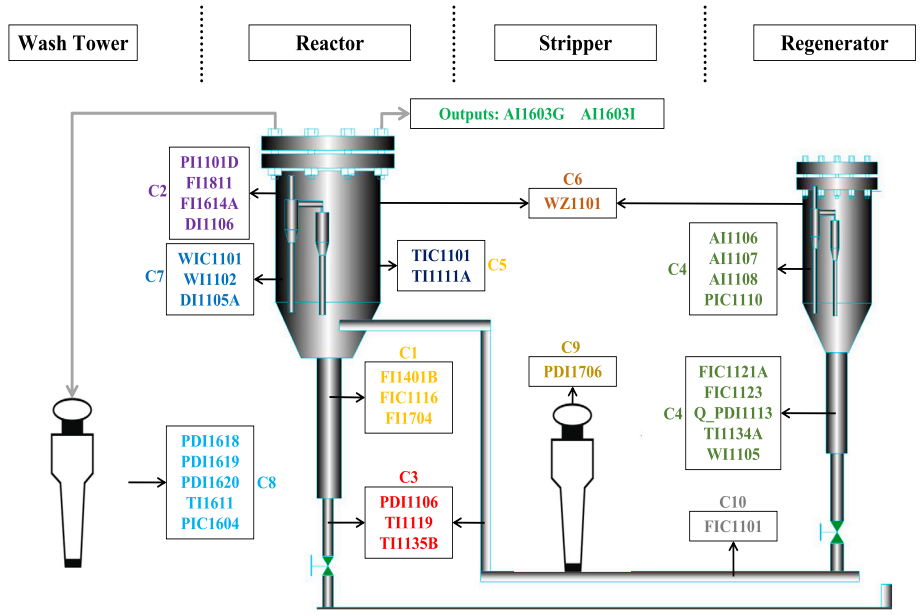


**Fig. 1.** Flow diagram of the reaction-regeneration unit of the MTO process.

**Table 1**
Variables codebook and clustering results.

| Cluster | Code | Name |
|---|---|---|
| Outputs | AI1603G | Ethylene |
| | AI1603I | Propylene |
| C1 | FI1401B | Methanol feed flow rate |
| | FIC1116 | Nitrogen flow rate for fluidization |
| | FI1704 | Non-condensable gas flow rate |
| C2 | FI1811 | Steam production |
| | PI1101D | Reaction pressure |
| | FI1614A | Product gas flow rate |
| | DI1106 | Density of the regenerated catalyst delivery pipe |
| C3 | PDI1106 | Pressure drop of spent catalyst slide valve |
| | TI1119 | The temperature of regenerated catalyst delivery pipe |
| | TI1135B | The temperature of the lower part of the stripping section |
| C4 | AI1106 | Flue gas O2 |
| | AI1107 | Flue gas CO |
| | AI1108 | Flue gas CO2 |
| | PIC1110 | Regeneration pressure |
| | FIC1121A | Main air flow rate |
| C4 | FIC1123 | Supplementary Nitrogen Flow rate to Main Air |
| | Q_PDI1113 | Catalyst circulation flow rate |
| | TI1134A | Regeneration temperature |
| | WI1105 | Total catalyst inventory of the regenerator |
| C5 | TIC1101 | Reaction temperature |
| | TI1111A | Temperature of the reactor dilute phase |
| C6 | WZ1101 | Total catalyst inventory |
| C7 | WIC1101 | Catalyst inventory of the reactor dense phase |
| | WI1102 | Total catalyst inventory of the reactor |
| | DI1105A | Density of the reactor dense phase |
| C8 | PIC1604 | Water washing tower pressure |
| | PDI1618 | Pressure drop between the middle and upper parts of the water washing tower |
| | PDI1619 | Total pressure drop of water washing tower |
| | PDI1620 | Pressure drop at the bottom of the water washing tower |
| | TI1611 | Water washing tower temperature |
| C9 | PDI1706 | Total pressure drop in stripping tower |
| C10 | FIC1001 | Refining C4 flow rate |

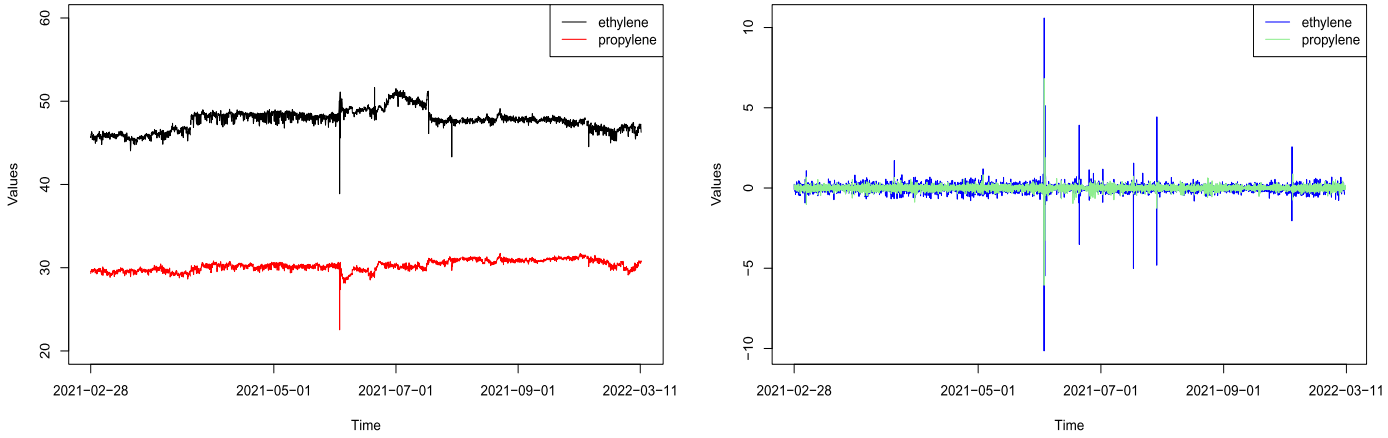


**Fig. 2.** Olefins yields (left) and first-difference (right) plots.

Wu et al., 2022), and cannot capture the dynamic change characteristics such as periodicity and volatility of time-series data. The principle of K-shape is similar to that of K-means, except that it improves the distance calculation method and optimizes the center of mass calculation method. K-shape algorithm uses Shape Based Distance (SBD) to compute the similarity between time series, and thus can efficiently handle long time series. It considers the scale and translation invariance of time series and can reasonably handle data with different amplitudes and phase differences (Yang et al., 2017).

The optimal number of clusters is 10 and the specific clustering results are shown in Table 1 and Fig. 1. Notably, the clustering results based on the K-shape are highly consistent with the actual chemical reaction mechanism, which indirectly confirms the high quality of the data presented in this paper.

### 2.2. Correlation analysis

We apply the Maximal Information Coefficient (MIC) (Kinney and Atwal, 2014) to explore the correlation between output and influencing factors data. Unlike the Pearson Correlation Coefficient (Sedgwick, 2012), which can only capture linear relationships between data, MIC captures both linear and nonlinear relationships. Given the characteristics of the data utilized in this paper, MIC was chosen to calculate correlations. Section 3.2 in Appendix A demonstrates the formula of MIC, and the associated results are shown as Fig. S1. Overall, there are strong correlations between the variables studied in this paper.

Under the MIC correlation test, there is a correlation between the 32 independent variables and light olefins yields. Among those, TI1134A (C4), with the smallest MIC value, is not significantly correlated with both ethylene and propylene. In a word, regeneration temperature is not significantly related to the yield of ethylene and propylene. This result is the same as the empirical results in Section 4. Most of the independent variables correlate with output variables, especially when considering the MIC test, which means the dataset has a time-varying solid correlation. This is the primary motivation for employing the rolling window approach to capture the time-varying correlation information for more acceptable accuracy and efficiency.

## 3. Model specification

### 3.1. Vector machine algorithm

Given a training dataset of the observed values as $\boldsymbol{S}^{tr} = \left\{\boldsymbol{X}^{tr}, \boldsymbol{Y}^{tr}\right\} = \left\{\left(\boldsymbol{x}_1^{tr}, y_1^{tr}\right), \left(\boldsymbol{x}_2^{tr}, y_2^{tr}\right), \ldots, \left(\boldsymbol{x}_n^{tr}, y_n^{tr}\right)\right\}$, where $\boldsymbol{x}_i^{tr} \in \boldsymbol{X} \subseteq \boldsymbol{R}^l$ is the independent variables and $y_i^{tr}$ is the dependent response, $i = 1, 2, \ldots, n$. The vector machine models (including the regression and classification) assume that there is a kernel function $K(\cdot,\cdot)$ such that every arbitrary dependent variable $y_0^{tr}$ in the training dataset can be expressed as a weighted summation of kernel functions

$$f\left(\boldsymbol{x}_0^{tr}; \boldsymbol{w}\right) = \sum_{i=1}^{n} w_i K\left(\boldsymbol{x}_0, \boldsymbol{x}_i^{tr}\right) + w_0, \tag{1}$$

the estimation with an error term is

$$y_0^{tr} = f\left(\boldsymbol{x}_0^{tr}; \boldsymbol{w}\right) + \varepsilon_0 = \sum_{i=1}^{n} w_i K\left(\boldsymbol{x}_0, \boldsymbol{x}_i^{tr}\right) + w_0 + \varepsilon_0. \tag{2}$$

Eq. (1) is often rewritten in a matrix form as

$$f\left(\boldsymbol{X}^{tr}; \boldsymbol{w}\right) = \boldsymbol{\Phi}^{tr}\boldsymbol{w}, \tag{3}$$

where $\boldsymbol{w} = (w_0, w_1, \ldots, w_n)^T$ is the weight vector, $\boldsymbol{\Phi}^{tr} = (\boldsymbol{\phi}_1^{tr}, \boldsymbol{\phi}_2^{tr}, \ldots, \boldsymbol{\phi}_n^{tr})^T$ is the kernel matrix of $\boldsymbol{S}^{tr}$, with $\boldsymbol{\phi}_i^{tr} = \left(1, K(\boldsymbol{x}_1^{tr}, \boldsymbol{x}_i^{tr}), K(\boldsymbol{x}_2^{tr}, \boldsymbol{x}_i^{tr}), \ldots, K(\boldsymbol{x}_n^{tr}, \boldsymbol{x}_i^{tr})\right)$. A test dataset can be defined as $\boldsymbol{S}^{te} = \left\{\boldsymbol{X}^{te}, Y^{te}\right\} = \left\{(\boldsymbol{x}_1^{te}, y_1^{te}), (\boldsymbol{x}_2^{te}, y_2^{te}), \ldots, (\boldsymbol{x}_m^{te}, y_m^{te})\right\}$, where $\boldsymbol{x}_j^{te} \in \boldsymbol{X} \subseteq \boldsymbol{R}^l$, $\boldsymbol{X}$ is in the same vector space as $\boldsymbol{S}^{tr}$, the dependent response is $y_j^{te}$, $j = 1, 2, \ldots, m$. After obtaining the estimation of $\boldsymbol{w}$ as $\hat{\boldsymbol{w}}$ based on $\boldsymbol{S}^{tr}$, the forecasting of the $\boldsymbol{S}^{te}$ is

$$\hat{f}\left(\boldsymbol{X}^{te}; \hat{\boldsymbol{w}}\right) = \boldsymbol{\Phi}^{te}\hat{\boldsymbol{w}}, \tag{4}$$

where $\boldsymbol{\Phi}^{te} = (\boldsymbol{\phi}_1^{te}, \boldsymbol{\phi}_2^{te}, \ldots, \boldsymbol{\phi}_m^{te})^T$ is the kernel matrix of $\boldsymbol{S}^{tr}$ and $\boldsymbol{S}^{te}$, with $\boldsymbol{\phi}_j^{te} = (1, K(\boldsymbol{x}_1^{tr}, \boldsymbol{x}_j^{te}), K(\boldsymbol{x}_2^{tr}, \boldsymbol{x}_j^{te}), \ldots, K(\boldsymbol{x}_n^{tr}, \boldsymbol{x}_j^{te}))$. In Eq. (4), the estimation of parameter $\boldsymbol{w}$ only depends on $\boldsymbol{S}^{tr}$. However, the kernel matrix $\boldsymbol{\Phi}^{te}$ includes both the information of the training and test datasets, the interactive relationship of $\boldsymbol{S}^{tr}$ and $\boldsymbol{S}^{te}$ is expressed by the kernel function.

### 3.2. RVM regression

Reflect on the kernel regression model presented in Eq. (1), (2), where $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$. The likelihood function, which is derived from model (3), follows a Gaussian distribution (Tipping, 2001)

$$\begin{aligned} p\left(\boldsymbol{Y}^{tr} \mid \boldsymbol{w}, \sigma^2\right) &= \mathcal{N}_n\left(\boldsymbol{Y}^{tr} | \boldsymbol{\Phi}^{tr}\boldsymbol{w}, \sigma^2 I_n\right) \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{\left\|\boldsymbol{Y}^{tr} - \boldsymbol{\Phi}^{tr}\boldsymbol{w}\right\|^2}{2\sigma^2}\right\}. \end{aligned} \tag{5}$$

In RVM regression, it is often assumed that most of the coefficients $w_i$ will be zero or of negligible magnitude. This is why a sparse representation of Eq. (1) is sought after. A prior for $\boldsymbol{w}$ is chosen to reflect this belief and obtain the desired sparse representation. To achieve sparsity, a separate independent zero-mean Gaussian prior is specified for each $w_i$

$$\begin{aligned} &\left(w_i \mid \alpha_i\right) \sim \mathcal{N}\left(w_i | 0, \alpha_i^{-1}\right), \\ &p(\boldsymbol{w} \mid \boldsymbol{\alpha}) = \mathcal{N}_{n+1}\left(\boldsymbol{w} | 0, \alpha_i^{-1} I_{n+1}\right). \end{aligned} \tag{6}$$

While it may seem unlikely for a Gaussian prior to result in sparsity, it has been discovered that incorporating a Gamma hyperprior for each $\alpha_i$ leads to a Student-t marginal prior for $w_i$ when $\alpha_i$ is integrated out. This implies that

$$\left(\alpha_i \mid a, b\right) \sim Ga\left(\alpha_i \mid a, b\right), \tag{7}$$

the marginal prior for $w_i$ is

$$\begin{aligned} p\left(w_i\right) &= \int p\left(w_i \mid \alpha_i\right) p\left(\alpha_i\right) d\alpha_i \\ &= \frac{b^a \Gamma(a + 1/2)}{(2\pi)^{1/2}\Gamma(a)} \left(b + w_i^{1/2}\right)^{-(a+1/2)}. \end{aligned} \tag{8}$$

The prior distribution for the vector $\boldsymbol{w}$, consisting of individual components $w_i$, is characterized by a marginal prior that is remarkably sparse. A two-dimensional case $\boldsymbol{w} = \left(w_1, w_2\right)^T$ is employed to demonstrate that this prior (Fig. S2, right) induces even greater sparsity pressure on the joint density of $\left(w_1, w_2\right)$ compared to the sparsity pressure exerted by the Gaussian and Laplace priors (Fig. S2, left and center). Sparse prior forces the most nonsignificant $w_i$ to be 0, making the RVM model efficient and robust (Fokoué et al., 2011). The parameter estimation process of RVM regression is illustrated in Fig. S3 and Section 3.3 in Appendix A. The forecasting process and pseudo-code of the RVM regression model are shown as Section 3.4 and Model 1 in Appendix A.

### 3.3. Kernel functions

One of the critical components of RVM is kernel function $K(\cdot,\cdot)$. By selecting different kernel functions, various vector machine models can be generated (Alvarez et al., 2012).

#### 3.3.1. Fundamentals of kernel functions

For all points $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}^l$ in a certain space, there exists a function $K(\boldsymbol{x}, \boldsymbol{y})$ that satisfies the following expression:

$$K(\boldsymbol{x},\boldsymbol{y}) = \varphi(\boldsymbol{x}) \cdot \varphi(\boldsymbol{y}). \tag{9}$$

If $K(\boldsymbol{x},\boldsymbol{y})$ also satisfies the Mercer condition, it is referred to as a kernel function. Although RVM does not require the Mercer condition (Tipping, 1999), we still use the kernel functions under the Mercer condition for the model comparison convenience. The Mercer condition is defined as:

Assuming $g(\boldsymbol{x}) \in L_2(\boldsymbol{R}^l)$ and $K(\boldsymbol{x},\boldsymbol{y}) \in L_2(\boldsymbol{R}^l \cdot \boldsymbol{R}^l)$. For any $g(\boldsymbol{x}) \neq 0$ with $\int g(\boldsymbol{x})^2 d\boldsymbol{x} < \infty$, the following equation holds:

$$\iint K(\boldsymbol{x},\boldsymbol{y}) g(\boldsymbol{x}) g(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y} \geq 0, \tag{10}$$

i.e., for all training samples, $x_1, x_2, ..., x_n \in \boldsymbol{R}^l$, $K(\boldsymbol{x},\boldsymbol{y})$ is a positive definite matrix.

The positive definite matrix guarantees that the objective function is convex and thus the optimization problem has a unique global optimal solution. The kernel function that satisfies the Mercer condition can map the original data to a high-dimensional space through feature mapping, in which linear operations are performed. By satisfying the Mercer condition, we are able to take advantage of the kernel method to transform a nonlinear problem into a linear problem in a high-dimensional space, thus improving the expressive and predictive performance of the model. It also facilitates the comparison with SVM, which must satisfy the Mercer condition.

The commonly used (also used in this paper) kernel functions include (Lall et al., 1993; Wang et al., 2015) Polynomial kernel, Linear kernel, Gaussian kernel, Sigmoid kernel, and Laplace kernel, defined in Appendix A, Section 3.5.

#### 3.3.2. Choices of kernel functions

Each kernel function type has benefits and drawbacks, which determine its level of non-linearity. As recorded in Appendix A, Section 3.5, based on the characteristics of kernel functions, they can be categorized into two types: local kernel and global kernel (Xu et al., 2015).

Local kernels, including Gaussian, Laplace, and Sigmoid kernels (Fig. S4), emphasize the similarity between neighboring points and are therefore effective in capturing local patterns and variations in the data. Linear kernels and Polynomial kernels are common global kernels, which are less capable of interpolation than local kernels, but are able to model long-distance interactions between the overall structure and features, and therefore capture the global features of the sample more efficiently (Fig. S5). When the input sample values differ and have a wide range of variations, the global kernel function still strongly impacts the samples, indicating its strong generalization ability. Compared to local kernel functions, global kernel functions are characterized by weaker interpolation ability but are more effective in capturing the global characteristics of the samples (Fig. S5). In addition to categorizing them as local and global, each kernel function has unique application scenarios. Table S1 summarizes the application scenarios of the 5 kernel functions considered in this paper (Min and Lee, 2005; Gretton et al., 2012).

#### 3.3.3. Hybrid kernel functions

In high-dimensional feature spaces, samples may exhibit uneven distribution. The results obtained from mapping transformations using a single kernel function may not be particularly satisfactory (Ding et al., 2013). By leveraging the properties of the kernel function and combining them, a hybrid kernel function balancing both aspects of performance and creating a new stable mixed kernel function model can be constructed to improve the stability greatly.

The general properties of kernel functions, which are very important for constructing our hybrid kernel RVM models, are shown in Appendix A, Section 3.6. Based on the five single kernel functions, the hybrid kernel is expressed as:

$$K_{i,j}^{hyb} = \lambda K_i + (1-\lambda) K_j, \tag{11}$$

**Table 2**
Grid search of hyperparameters.

| Hyperparameters | Range |
|---|---|
| Hybrid Kernel Weights | [0.1, 0.9][0.2, 0.8][0.3, 0.7]<br>[0.4, 0.6][0.5, 0.5][0.6, 0.4]<br>[0.7, 0.3][0.8, 0.2][0.9, 0.1] |
| Kernel Function Type | [gaussian][linear][polynomial][sigmoid][laplacian] |
| Gamma | [10][1][0.1][0.01] |
| Window Size | [2][3][4]…[24] |
| Polynomial Kernel Degree | [0.1][1][2][3] |

where $\{K_i, K_j\} \subseteq \{K_{polynomial}, K_{linear}, K_{gaussian}, K_{sigmoid}, K_{laplace}\}$, $\lambda \in (0,1), i \neq j$.

The selection of kernel function-related parameters directly impacts the performance of machine learning methods. Fig. S6 shows the hybrid kernels of Gaussian ($x_j = 0, \sigma = 1$) with Laplace ($x_j = 1, \sigma = 1$), Linear, Sigmoid ($\upsilon = 1, c = 0$), and Polynomial ($d = 3, c = 0$). The hybrid kernel has the edges of two individual kernels and obtains additional freedom to adjust the weights for the most significant performance.

### 3.4. Estimation of hyper-parameters

Except for the primary parameter estimation in Model 1, research conducted by Dioşan et al. (2012) has shown that the performance of vector machines is also significantly determined by the hyper-parameters in the kernel function and the weight coefficients. Therefore, this study also focuses on optimizing hyper-parameter estimation (Claesen and De Moor, 2015). We apply the Grid Search (Lerman, 1980) method for hyper-parameters estimation as Model 2 in Appendix A.

The specific hyperparameters and their optimization search ranges are shown in Table 2.

### 3.5. Rolling-window training and test process

Based on the highly time-varying relationship between variables illustrated in Section 2, the rolling-windows approach is preferred for this paper's training and test process (Inoue et al., 2017). When describing a time-varying state where the value of a time series at a particular moment is highly correlated with data close to the current time point, the input and output data of the model should be continuously updated. This necessitates the introduction of rolling time window technology. Using a fixed range and time-rolling data interval for modeling can significantly improve efficiency. It continuously evolves the base forecasting model by moving forward and including the latest period's new sample in the "time window". In this method, new sample data replaces old data in real time, and changes in the current window's sample data require reconstructing a more optimal forecasting model.

The principle of the rolling window can be described as follows: assume a set of continuous data samples, where $l$ data groups can characterize the state at time $t+l$ from time $t$ to $t+l-1$. Therefore, data from the interval between time $t$ and $t+l-1$ establishes a model and forecasts the state at time $t+l$. When the next moment arrives, while keeping the length of the time window $l$ constant, the data at time $t$ is discarded, and the data at time $t+l$ is added. Then, data from time $t+1$ to $t+l$ is used to establish a model to obtain the forecasted output at time $t+l+1$. The $S$ groups of continuous data intervals characterizing the state as time progresses are dynamically updated. Fig. S7 shows this paper's comprehensive training and test process when $l = 2$.

### 3.6. Criteria for forecasting performance

The loss function must be chosen for statistical tests to compare the forecasting performance of different model specifications. In this paper, two standard loss functions are used: Mean Squared Error (MSE) and Mean Absolute Error (MAE), which are both statistical measures used to evaluate the performance of a forecasted or estimated model. They

**Table 3**
Forecasting results of ethylene with single kernel.

| $C_2H_4$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | Gaussian | | Linear | | Polynomial | | Sigmoid | | Laplace | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | 0.001397 | 0.025564 | 0.001560 | 0.027651 | 0.001366 | 0.025579 | 0.001355 | 0.025425 | **0.001338** | **0.025277** |
| 3 | 0.001666 | 0.028904 | 0.001961 | 0.032018 | 0.001732 | 0.029741 | 0.001716 | 0.029556 | 0.001669 | 0.029143 |
| 4 | 0.001904 | 0.031401 | 0.002308 | 0.035406 | 0.002060 | 0.033117 | 0.002040 | 0.032909 | 0.001970 | 0.032300 |
| 5 | 0.002075 | 0.033243 | 0.002583 | 0.037927 | 0.002320 | 0.035598 | 0.002297 | 0.035383 | 0.002206 | 0.034653 |
| 6 | 0.002176 | 0.034289 | 0.002764 | 0.03960 | 0.002502 | 0.037324 | 0.002478 | 0.037115 | 0.002371 | 0.036247 |
| 7 | 0.002244 | 0.034806 | 0.002861 | 0.040419 | 0.002605 | 0.038221 | 0.002581 | 0.038019 | 0.002462 | 0.037055 |
| 8 | 0.002235 | 0.034638 | 0.002885 | 0.040418 | 0.002629 | 0.038263 | 0.002607 | 0.038072 | 0.002482 | 0.037076 |
| 9 | 0.002174 | 0.033978 | 0.002778 | 0.039539 | 0.00572 | 0.037601 | 0.002552 | 0.037423 | 0.002430 | 0.036484 |
| 10 | 0.002085 | 0.032997 | 0.002637 | 0.038226 | 0.002459 | 0.036486 | 0.002442 | 0.036326 | 0.002319 | 0.035400 |
| 11 | 0.001975 | 0.031812 | 0.002456 | 0.036562 | 0.002322 | 0.035086 | 0.002307 | 0.034946 | 0.002186 | 0.034029 |
| 12 | 0.001875 | 0.030701 | 0.002308 | 0.035109 | 0.002203 | 0.033800 | 0.002191 | 0.033670 | 0.002066 | 0.032780 |
| Average | **0.001982** | **0.032030** | 0.002464 | 0.036625 | 0.002252 | 0.034620 | 0.002233 | 0.034440 | 0.002136 | 0.033677 |
| Optimal window | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | |

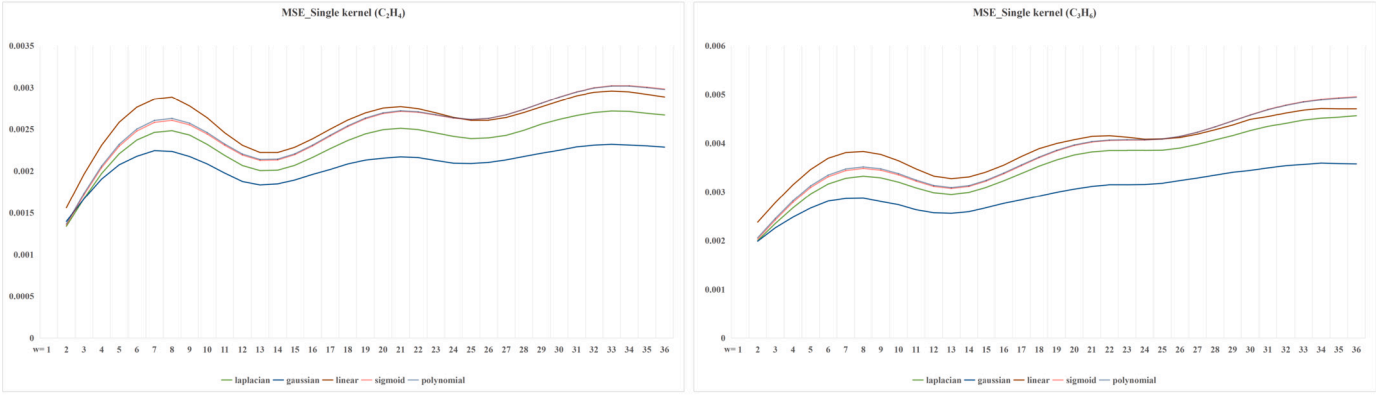


**Fig. 3.** Variation of ethylene (left) and propylene (right) yield forecast error with windows (single kernel).

are both ways to quantify the difference between the actual values and those forecasted by the model.

MSE is calculated as the average squared differences between the forecasted and actual values. Conversely, MAE is calculated as the average absolute differences between the forecasted and actual values. Mathematically, they can be represented as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2,$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \tag{12}$$

where $n$ is the total number of data points, $y_i$ is the actual value, and $\hat{y}_i$ is the forecasted value for the $i$-th observation. A lower MSE or MAE indicates a more satisfactory fit of the model to the data. The key difference between MSE and MAE lies in how they handle outliers. Since MSE squares the differences before averaging, it tends to penalize larger errors more severely than MAE, which could make MSE more sensitive to outliers in the data.

## 4. Results

### 4.1. Olefins yields forecasting study

Apply the model in Section 3 to the data in Section 2 for forecasting the light olefins yields.

#### 4.1.1. Ethylene yields forecasting

(1) Single kernel study

The results of the single kernel study for ethylene yield forecasting are shown in Table 3. From the perspective of average value, the Gaussian kernel is the optimal one, which means the square of the difference between independent variables is most significant in contributing information to forecast the ethylene yield. From the rolling window perspective, 2 window length is most preferred for all the single kernels so that the forecasting can be done efficiently with four hours of data. Notably, as shown in Fig. 3 (left), the error (MSE) varies periodically (roughly 24 hours) as the window increases. Globally, the single Laplace kernel with 2 window length is the best, the corresponding forecasting results are shown in Fig. 4 (left).

(2) Hybrid kernel study

The results of the hybrid kernel study between Gaussian and Laplace for ethylene yield forecasting are shown in Table 4. (0.1·Gaussian+0.9·Laplace) has a better result than the single Laplace kernel concerning MAE with the same window-length 2, and the errors continue to be cyclical. From the perspective of average value, (0.9·Gaussian+0.1·Laplace) is the optimal one, but not more reasonable than the single Laplace. The other hybrid kernel studies are worse than the single one, so we do not show the results.

In conclusion, the Laplace and Gaussian kernels are preferred by ethylene forecasting. The difference between independent variables can explain more about the ethylene yield in the DMTO system. Also, the hybrid of Laplace and Gaussian can obtain the most satisfactory result for forecasting. All the studies for ethylene forecasting show periodic variation in error and return the optimal window length of 2, which provides a desired scheme to conduct the efficient model. This suggests that ethylene yield is characterized by short-term fluctuations, and the rolling strategy can effectively capture this pattern for fast forecasting purposes.

**Table 4**
Forecasting results of ethylene with hybrid kernels of Gaussian and Laplace.

| $C_2H_4$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | 0.1, 0.9[a] | | 0.2, 0.8[a] | | 0.3, 0.7[a] | | 0.4, 0.6[a] | | 0.5, 0.5[a] | | 0.6, 0.4[a] | | 0.7, 0.3[a] | | 0.8, 0.2[a] | | 0.9, 0.1[a] | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | **0.001338** | **0.025260** | 0.001340 | 0.025256 | 0.001343 | 0.025271 | 0.001347 | 0.025296 | 0.001353 | 0.025327 | 0.001359 | 0.025352 | 0.001367 | 0.025398 | 0.001375 | 0.025440 | 0.001385 | 0.025491 |
| 3 | 0.001663 | 0.029087 | 0.001658 | 0.029038 | 0.001657 | 0.029012 | 0.001656 | 0.028991 | 0.001657 | 0.028982 | 0.001652 | 0.028902 | 0.001649 | 0.028862 | 0.001653 | 0.028868 | 0.001659 | 0.028883 |
| 4 | 0.001958 | 0.032187 | 0.001949 | 0.032101 | 0.001942 | 0.032026 | 0.001934 | 0.031932 | 0.001921 | 0.031795 | 0.001912 | 0.031646 | 0.001898 | 0.031508 | 0.001899 | 0.031463 | 0.001902 | 0.031430 |
| 5 | 0.002188 | 0.034481 | 0.002175 | 0.034356 | 0.002160 | 0.034235 | 0.002146 | 0.034074 | 0.002119 | 0.033820 | 0.002093 | 0.033587 | 0.002080 | 0.033395 | 0.002071 | 0.033298 | 0.002073 | 0.033269 |
| 6 | 0.002348 | 0.036042 | 0.002335 | 0.035868 | 0.002312 | 0.035689 | 0.002291 | 0.035494 | 0.002250 | 0.035128 | 0.002221 | 0.034835 | 0.002203 | 0.034602 | 0.002187 | 0.034448 | 0.002178 | 0.034365 |
| 7 | 0.002437 | 0.036821 | 0.002419 | 0.036613 | 0.002388 | 0.036356 | 0.002353 | 0.036087 | 0.002317 | 0.035708 | 0.002293 | 0.035428 | 0.002256 | 0.035123 | 0.002245 | 0.034955 | 0.002240 | 0.034853 |
| 8 | 0.002456 | 0.036846 | 0.002425 | 0.036585 | 0.002393 | 0.036317 | 0.002361 | 0.036022 | 0.002318 | 0.035618 | 0.002284 | 0.035305 | 0.002264 | 0.035104 | 0.002246 | 0.034867 | 0.002234 | 0.034721 |
| 9 | 0.002401 | 0.036230 | 0.002360 | 0.035959 | 0.002336 | 0.035729 | 0.002307 | 0.035438 | 0.002267 | 0.035088 | 0.002232 | 0.034746 | 0.002207 | 0.034448 | 0.002188 | 0.034226 | 0.002180 | 0.034087 |
| 10 | 0.002286 | 0.035118 | 0.002257 | 0.034905 | 0.002235 | 0.034653 | 0.002210 | 0.034383 | 0.002170 | 0.034019 | 0.002132 | 0.033660 | 0.002107 | 0.033417 | 0.002098 | 0.033275 | 0.002090 | 0.033120 |
| 11 | 0.002150 | 0.033733 | 0.002123 | 0.033509 | 0.002103 | 0.033290 | 0.002090 | 0.033078 | 0.002048 | 0.032719 | 0.002012 | 0.032357 | 0.001987 | 0.032091 | 0.001982 | 0.032007 | 0.001979 | 0.031906 |
| 12 | 0.002036 | 0.032513 | 0.002011 | 0.032300 | 0.001992 | 0.032097 | 0.001983 | 0.031904 | 0.001945 | 0.031561 | 0.001912 | 0.031225 | 0.001891 | 0.030997 | 0.001889 | 0.030912 | 0.001883 | 0.030792 |
| Average | 0.002115 | 0.033484 | 0.002096 | 0.033317 | 0.002078 | 0.033152 | 0.002062 | 0.032973 | 0.002033 | 0.032706 | 0.002009 | 0.032459 | 0.001992 | 0.032268 | 0.001985 | 0.032160 | **0.001982** | **0.032083** |
| Optimal window | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | |

[a] Weights of Gaussian and Laplace kernels, respectively.

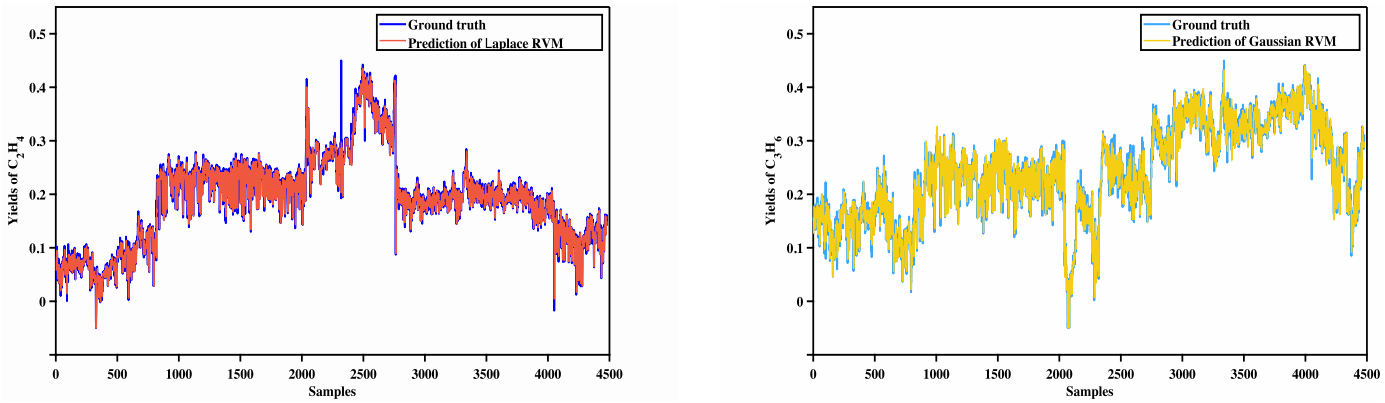


**Fig. 4.** Ethylene (left) and propylene (right) optimal forecasting results with all independent variables.

**Table 5**
Forecasting results of propylene with single kernel.

| $C_3H_6$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | Gaussian | | Linear | | Polynomial | | Sigmoid | | Laplace | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | **0.001988** | **0.033704** | 0.002379 | 0.037005 | 0.002059 | 0.034271 | 0.002031 | 0.034010 | 0.001998 | 0.033780 |
| 3 | 0.002262 | 0.036291 | 0.002778 | 0.040536 | 0.002454 | 0.037799 | 0.002423 | 0.037536 | 0.002350 | 0.037045 |
| 4 | 0.002481 | 0.038283 | 0.003144 | 0.043407 | 0.002816 | 0.040625 | 0.002784 | 0.040347 | 0.002671 | 0.039644 |
| 5 | 0.002668 | 0.039880 | 0.003458 | 0.045719 | 0.003124 | 0.042974 | 0.003090 | 0.042694 | 0.002956 | 0.041839 |
| 6 | 0.002812 | 0.040934 | 0.003690 | 0.047331 | 0.003344 | 0.044569 | 0.003310 | 0.044289 | 0.003162 | 0.043329 |
| 7 | 0.002865 | 0.041336 | 0.003807 | 0.048010 | 0.003471 | 0.045388 | 0.003438 | 0.045111 | 0.003279 | 0.044095 |
| 8 | 0.002871 | 0.041300 | 0.003830 | 0.048036 | 0.003513 | 0.045584 | 0.003482 | 0.045329 | 0.003323 | 0.044327 |
| 9 | 0.002803 | 0.040757 | 0.003768 | 0.047488 | 0.003474 | 0.045174 | 0.003445 | 0.044938 | 0.003289 | 0.043928 |
| 10 | 0.002737 | 0.040134 | 0.003640 | 0.046533 | 0.003376 | 0.044426 | 0.003351 | 0.044206 | 0.003201 | 0.043218 |
| 11 | 0.002633 | 0.039238 | 0.003472 | 0.045357 | 0.003244 | 0.043424 | 0.003222 | 0.043230 | 0.003083 | 0.042295 |
| 12 | 0.002571 | 0.038617 | 0.003325 | 0.044252 | 0.003133 | 0.042550 | 0.003114 | 0.042369 | 0.002984 | 0.041457 |
| Average | **0.002608** | **0.039134** | 0.003390 | 0.044879 | 0.003092 | 0.042435 | 0.003063 | 0.042187 | 0.002936 | 0.041360 |
| Optimal window | $l=2$ | | $l=2$ | | $l=2$ | | $l=2$ | | $l=2$ | |

#### 4.1.2. *Propylene yields forecasting*

(1) Single kernel study

The results of the single kernel study for propylene yield forecasting are shown in Table 5. The Gaussian kernel outperforms in terms of global and average perspectives. The square difference between independent variables contributes the most influential information to forecast the propylene yield. From the rolling-window perspective, as shown in Fig. 3, identical to the ethylene, the MSE for propylene yield forecasting continues to show cyclical variation and 2 window length is most preferred. The results for propylene forecasting with a single Gaussian kernel are shown in Fig. 4 (right).

(2) Hybrid kernel study

The results of the hybrid kernel study between Gaussian and Sigmoid, Gaussian and Laplace for propylene yield forecasting are shown in Tables 6 and 7. (0.8·Gaussian+0.2·Sigmoid) and (0.7·Gaussian+0.3·Laplace) have the most promising results for the hybrid studies, also are more reasonable than the single Gaussian kernel with same window-length of 2, and the errors still continue to be cyclical. (0.7·Gaussian+0.3·Laplace) outperforms all others for propylene forecasting. From the perspective of average value, (0.9·Gaussian+0.1·Sigmoid or Laplace) is the optimal one, which means Gaussian is still the most consequential kernel here. The other hybrid kernel studies are improper, so we do not show the results in the paper.

### 4.2. *Analysis of independent variable clusters*

To identify process variables that have a significant impact on the yields of ethylene and propylene, we conduct single cluster and mixed cluster studies. The single cluster study involves adding each of the 10 clusters to the model and identifying the clusters that have a significant effect on yields forecasts. The mixed cluster study is the process of arranging and combining the different clusters in order to explore the combinatorial effects among them and fully exploit the potential information affecting the light olefins yields.

#### 4.2.1. *Cluster study for ethylene yields*

(1) Single Cluster Studies for Ethylene Yields

Based on the 10 clusters of independent variables in Section 2, we test the forecasting ability of each cluster to locate the influential ones in the DMTO process. The results of single cluster studies for ethylene yield forecasting are shown in Table 8. $C3$ is the most noteworthy one. $C5, C8$, and $C10$ also provide the convincing results. The results indicate that the temperature ($C3$ and $C5$) is the most important factor affecting ethylene yield. The pressure ($C3$ and $C8$) and refining of heavy olefins ($C10$) are also critical for ethylene production. In conclusion, temperature, pressure, and refining of heavy olefins are 3 factors that need to be focused on in the DMTO process to increase the ethylene yields, which is typically identical to the chemical knowledge.

The optimal window length shows that the single cluster needs a larger length than all independent variables studied because it has small dimensions. Hence, a larger data size is needed to enhance the information for forecasting. $C3$ and $C5$ need the smallest 2 of window length, $C8$ and $C10$ require the second smallest 3 of window length, which also illustrate that $C3$ and $C5$ include considerable data information for forecasting ethylene yields. This is an additional proof that $C3, C5, C8$, and $C10$ are necessary factors.

(2) Mixed Clusters Studies for Ethylene Yields

**Table 6**
Forecasting results of propylene with hybrid kernels of Gaussian and Sigmoid.

| $C_3H_6$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | 0.1, 0.9[a] | | 0.2, 0.8[a] | | 0.3, 0.7[a] | | 0.4, 0.6[a] | | 0.5, 0.5[a] | | 0.6, 0.4[a] | | 0.7, 0.3[a] | | 0.8, 0.2[a] | | 0.9, 0.1[a] | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | 0.002019 | 0.033932 | 0.002009 | 0.033866 | 0.002001 | 0.033815 | 0.001994 | 0.033770 | 0.001988 | 0.033733 | 0.001986 | 0.033709 | 0.001985 | 0.033699 | **0.001983** | **0.033682** | 0.001985 | 0.033688 |
| 3 | 0.002401 | 0.037384 | 0.002374 | 0.037215 | 0.002357 | 0.037091 | 0.002336 | 0.036944 | 0.002322 | 0.036843 | 0.002305 | 0.036697 | 0.002286 | 0.036543 | 0.002276 | 0.036437 | 0.002268 | 0.036348 |
| 4 | 0.002748 | 0.040126 | 0.002714 | 0.039916 | 0.002696 | 0.039777 | 0.002666 | 0.039587 | 0.002616 | 0.039328 | 0.002578 | 0.039054 | 0.002542 | 0.038768 | 0.002510 | 0.038525 | 0.002491 | 0.038365 |
| 5 | 0.003042 | 0.042402 | 0.002999 | 0.042131 | 0.002955 | 0.041855 | 0.002911 | 0.041583 | 0.002856 | 0.041277 | 0.002799 | 0.040941 | 0.002746 | 0.040548 | 0.002710 | 0.040269 | 0.002687 | 0.040075 |
| 6 | 0.003260 | 0.043963 | 0.003208 | 0.043638 | 0.003164 | 0.043340 | 0.003112 | 0.043001 | 0.003041 | 0.042620 | 0.002961 | 0.042103 | 0.002898 | 0.041664 | 0.002856 | 0.041318 | 0.002836 | 0.041123 |
| 7 | 0.003383 | 0.044756 | 0.003328 | 0.044403 | 0.003268 | 0.044025 | 0.003199 | 0.043627 | 0.003144 | 0.043251 | 0.003053 | 0.042697 | 0.002966 | 0.042136 | 0.002939 | 0.041851 | 0.002902 | 0.041566 |
| 8 | 0.003424 | 0.044971 | 0.003366 | 0.044610 | 0.003302 | 0.044216 | 0.003231 | 0.043781 | 0.003187 | 0.043462 | 0.003088 | 0.042820 | 0.002997 | 0.042227 | 0.002954 | 0.041898 | 0.002912 | 0.041602 |
| 9 | 0.003388 | 0.044572 | 0.003330 | 0.044214 | 0.003271 | 0.043848 | 0.003206 | 0.043419 | 0.003157 | 0.043061 | 0.003067 | 0.042485 | 0.002989 | 0.041955 | 0.002923 | 0.041506 | 0.002867 | 0.041148 |
| 10 | 0.003297 | 0.043855 | 0.003240 | 0.043499 | 0.003183 | 0.043130 | 0.003111 | 0.042686 | 0.003076 | 0.042404 | 0.002982 | 0.041741 | 0.002898 | 0.041220 | 0.002843 | 0.040833 | 0.002766 | 0.040396 |
| 11 | 0.003168 | 0.042884 | 0.003118 | 0.042551 | 0.003064 | 0.042197 | 0.002994 | 0.041768 | 0.002959 | 0.041513 | 0.002867 | 0.040822 | 0.002783 | 0.040298 | 0.002689 | 0.039727 | 0.002664 | 0.039501 |
| 12 | 0.003064 | 0.042039 | 0.003019 | 0.041729 | 0.002972 | 0.041408 | 0.002903 | 0.040946 | 0.002861 | 0.040628 | 0.002790 | 0.040072 | 0.002696 | 0.039514 | 0.002640 | 0.039129 | 0.002609 | 0.038914 |
| Average | 0.003018 | 0.041899 | 0.002973 | 0.041616 | 0.002930 | 0.041337 | 0.002878 | 0.041010 | 0.002837 | 0.040738 | 0.002770 | 0.040286 | 0.002708 | 0.039870 | 0.002666 | 0.039561 | **0.002635** | **0.039339** |
| Optimal window | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | |

[a] Weights of Gaussian and Sigmoid kernels, respectively.

**Table 7**
Forecasting results of propylene with hybrid kernels of Gaussian and Laplace.

| $C_3H_6$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | 0.1, 0.9[a] | | 0.2, 0.8[a] | | 0.3, 0.7[a] | | 0.4, 0.6[a] | | 0.5, 0.5[a] | | 0.6, 0.4[a] | | 0.7, 0.3[a] | | 0.8, 0.2[a] | | 0.9, 0.1[a] | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | 0.001993 | 0.033750 | 0.001989 | 0.033724 | 0.001985 | 0.033701 | 0.001982 | 0.033686 | 0.001982 | 0.033677 | 0.001982 | 0.033676 | **0.001981** | **0.033668** | 0.001983 | 0.033677 | 0.001985 | 0.033687 |
| 3 | 0.002334 | 0.036931 | 0.002324 | 0.036860 | 0.002313 | 0.036768 | 0.002303 | 0.036697 | 0.002292 | 0.036603 | 0.002281 | 0.036501 | 0.002272 | 0.036421 | 0.002264 | 0.036331 | 0.002263 | 0.036309 |
| 4 | 0.002662 | 0.039573 | 0.002642 | 0.039443 | 0.002602 | 0.039249 | 0.002581 | 0.039095 | 0.002552 | 0.038883 | 0.002523 | 0.038659 | 0.002505 | 0.038496 | 0.002494 | 0.038410 | 0.002483 | 0.038315 |
| 5 | 0.002921 | 0.041627 | 0.002870 | 0.041377 | 0.002853 | 0.041263 | 0.002825 | 0.041085 | 0.002769 | 0.040730 | 0.002728 | 0.040417 | 0.002704 | 0.040223 | 0.002688 | 0.040120 | 0.002679 | 0.040000 |
| 6 | 0.003114 | 0.043053 | 0.003082 | 0.042855 | 0.003051 | 0.042648 | 0.003005 | 0.042393 | 0.002931 | 0.041939 | 0.002877 | 0.041552 | 0.002854 | 0.041317 | 0.002846 | 0.041172 | 0.002827 | 0.041048 |
| 7 | 0.003220 | 0.043780 | 0.003196 | 0.043546 | 0.003137 | 0.043268 | 0.003100 | 0.042991 | 0.003023 | 0.042491 | 0.002952 | 0.042051 | 0.002914 | 0.041786 | 0.002909 | 0.041629 | 0.002885 | 0.041459 |
| 8 | 0.003256 | 0.043951 | 0.003233 | 0.043746 | 0.003185 | 0.043447 | 0.003142 | 0.043152 | 0.003063 | 0.042661 | 0.002991 | 0.042197 | 0.002940 | 0.041858 | 0.002896 | 0.041598 | 0.002900 | 0.041530 |
| 9 | 0.003237 | 0.043624 | 0.003190 | 0.043316 | 0.003156 | 0.043067 | 0.003118 | 0.042818 | 0.003015 | 0.042182 | 0.002963 | 0.041808 | 0.002920 | 0.041530 | 0.002890 | 0.041274 | 0.002848 | 0.041036 |
| 10 | 0.003144 | 0.042872 | 0.003100 | 0.042589 | 0.003068 | 0.042374 | 0.003036 | 0.042135 | 0.002953 | 0.041549 | 0.002883 | 0.041073 | 0.002838 | 0.040767 | 0.002796 | 0.040556 | 0.002753 | 0.040300 |
| 11 | 0.003020 | 0.041913 | 0.002985 | 0.041699 | 0.002961 | 0.041526 | 0.002920 | 0.041225 | 0.002851 | 0.040692 | 0.002763 | 0.040146 | 0.002690 | 0.039765 | 0.002664 | 0.039579 | 0.002650 | 0.039397 |
| 12 | 0.002927 | 0.041103 | 0.002890 | 0.040830 | 0.002867 | 0.040671 | 0.002831 | 0.040415 | 0.002755 | 0.039911 | 0.002677 | 0.039425 | 0.002615 | 0.039033 | 0.002611 | 0.038942 | 0.002596 | 0.038794 |
| Average | 0.002894 | 0.041107 | 0.002864 | 0.040908 | 0.002834 | 0.040726 | 0.002804 | 0.040517 | 0.002744 | 0.040120 | 0.002693 | 0.039773 | 0.002657 | 0.039533 | 0.002640 | 0.039390 | **0.002624** | **0.039261** |
| Optimal window | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | | $l = 2$ | |

[a] Weights of Gaussian and Laplace kernels, respectively.

**Table 8**
Forecasting results of ethylene in single cluster study.

| $C_2H_4$ | | | | | |
|---|---|---|---|---|---|
| Cluster | MSE | MAE | Mean | Kernel | Optimal window length |
| C1 | 0.001353 | 0.025412 | 0.013383 | 0.1·Gaussian+0.9·Laplace | 5 |
| C2 | 0.001352 | 0.025385 | 0.013369 | 0.1·Gaussian+0.9·Laplace | 4 |
| **C3**[a] | **0.001345** | **0.025317** | **0.013334** | 0.5·Gaussian+0.5·Sigmoid | 2 |
| C4 | 0.001351 | 0.025409 | 0.013380 | 0.1·Polynomial+ 0.9·Laplace | 5 |
| **C5**[b] | **0.001346** | **0.025311** | **0.013329** | Gaussian | 2 |
| C6 | 0.001354 | 0.025423 | 0.013389 | 0.1·Polynomial+0.9·Laplace | 4 |
| C7 | 0.001355 | 0.025427 | 0.013391 | Gaussian | 4 |
| **C8**[b] | **0.001347** | **0.025327** | **0.013337** | 0.9·Gaussian+0.1·Laplace | 3 |
| C9 | 0.001354 | 0.025415 | 0.013384 | Gaussian | 5 |
| **C10**[b] | **0.001350** | **0.025344** | **0.013347** | 0.1·Linear+0.9·Laplace | 3 |

[a] Best results are indicated in bold and underlined.
[b] Significant good results are indicated in bold.

**Table 9**
Forecasting Results of ethylene in hybrid clusters study.

| $C_2H_4$ | | | | | |
|---|---|---|---|---|---|
| Cluster | MSE | MAE | Mean | Kernel | Optimal window length |
| **C3+C5** [a] | **0.001337** | **0.025122** | **0.013230** | 0.2·Linear+0.8·Polynomial | 2 |
| **C3+C8**[b] | **0.001339** | **0.025284** | **0.013316** | Gaussian | 2 |
| **C3+C10**[b] | **0.001341** | **0.025238** | **0.013289** | 0.3·Linear+0.7·Laplace | 2 |
| C5+C8 | 0.001354 | 0.025421 | 0.013387 | Gaussian | 2 |
| C5+C10 | 0.001350 | 0.025341 | 0.013346 | 0.1·Linear+0.9·Laplace | 3 |
| C8+C10 | 0.001353 | 0.025399 | 0.013376 | Gaussian | 3 |
| **C3+C5+C8** [a] | **0.001323** | **0.025107** | **0.013215** | Gaussian | 2 |
| **C3+C5+C10**[b] | **0.001323** | **0.025115** | **0.013219** | 0.2·Linear+0.8·Polynomial | 2 |
| C3+C8+C10 | 0.001328 | 0.025207 | 0.013268 | Gaussian | 2 |
| C5+C8+C10 | 0.001333 | 0.025259 | 0.013296 | Gaussian | 2 |
| C3+C5+C8+C10 | 0.001314 | 0.025007 | 0.013161 | Gaussian | 2 |

[a] Best results in each group are indicated in bold and underlined.
[b] Significant good results in each group are indicated in bold.

**Table 10**
Forecasting results of propylene in single cluster study.

| $C_3H_6$ | | | | | |
|---|---|---|---|---|---|
| Cluster | MSE | MAE | Mean | Kernel | Optimal window length |
| C1 | 0.002026 | 0.033992 | 0.018009 | 0.9·Gaussian+0.1·Laplace | 7 |
| C2 | 0.002029 | 0.033972 | 0.018000 | Gaussian | 7 |
| **C3**[b] | **0.002026** | **0.033972** | **0.017999** | Gaussian | 4 |
| C4 | 0.002021 | 0.033976 | 0.017998 | 0.9·Gaussian+0.1·Sigmoid | 6 |
| **C5**[b] | **0.002016** | **0.033899** | **0.017958** | Gaussian | 5 |
| C6 | 0.002034 | 0.034034 | 0.018034 | 0.9·Gaussian+0.1·Sigmoid | 7 |
| C7 | 0.002030 | 0.034002 | 0.018016 | Gaussian | 5 |
| **C8**[b] | **0.002001** | **0.033656** | **0.017829** | 0.9·Gaussian+0.1·Sigmoid | 5 |
| C9 | 0.002030 | 0.034004 | 0.018039 | Gaussian | 7 |
| **C10**[a] | **0.001995** | **0.033610** | **0.017803** | 0.3·Linear+0.7·Laplace | 4 |

[a] Best results are indicated in bold and underlined.
[b] Significant good results are indicated in bold.

Based on the 4 significant clusters of $C3$, $C5$, $C8$, and $C10$, we conducted the 2, 3, and 4 mixed cluster studies. The results are listed in Table 9. $C3+C5+C8$ is the most noteworthy composition, and adding it to the model produces even smaller errors than the single cluster study. This suggests that taking both temperature and pressure into account would be more favorable for ethylene yield forecasting.

### 4.2.2. Cluster study for propylene yields

(1) Single Cluster Studies for Propylene Yields

Same as the single cluster studies for the ethylene yields, we conducted the experiment for propylene yields, listed in Table 10. $C3$, $C5$, $C8$ and $C10$ are still important for propylene. Compared with the ethylene results, $C10$ is the most significant variable. This indicates that the refining of heavy olefins has a powerful impact on propylene yields. $C8$ is more influential than $C3$ and $C5$, which shows that the pressure is more important for propylene yield than ethylene yield.

The optimal window length results show the same information as the ethylene study. $C10$, $C8$, $C3$, and $C5$ have shorter window lengths than other clusters, which means they supply more information than other clusters in explaining the propylene yield.

(2) Mixed Clusters Studies for Propylene Yields

The mixed clusters studies between $C3$, $C5$, $C8$, and $C10$ are shown in Table 11. Unlike the ethylene yield forecasting, $C8+C10$ is the most

**Table 11**
Forecasting results of propylene in hybrid clusters study.

| $C_3H_6$ | | | | | |
|---|---|---|---|---|---|
| Cluster | MSE | MAE | Mean | Kernel | Optimal window length |
| **C3+C5**[b] | **0.002011** | **0.033862** | **0.017936** | Gaussian | 2 |
| C3+C8 | 0.002025 | 0.033964 | 0.017995 | Gaussian | 2 |
| **C3+C10**[b] | **0.002015** | **0.033891** | **0.017953** | Gaussian | 2 |
| C5+C8 | 0.002021 | 0.033789 | 0.017905 | Gaussian | 2 |
| **C5+C10** | **0.002019** | **0.033906** | **0.017963** | Gaussian | 2 |
| **C8+C10** [a] | **0.002005** | **0.033817** | **0.017911** | Gaussian | 2 |
| C3+C5+C8 | 0.002027 | 0.033911 | 0.017969 | Gaussian | 2 |
| **C3+C5+C10** [a] | **0.002009** | **0.033828** | **0.017918** | Gaussian | 2 |
| C3+C8+C10 | 0.002023 | 0.033893 | 0.017958 | Gaussian | 2 |
| **C5+C8+C10**[b] | **0.002012** | **0.033846** | **0.017929** | Gaussian | 2 |
| C3+C5+C8+C10 | 0.002008 | 0.033820 | 0.017914 | Gaussian | 2 |

[a] Best results in each group are indicated in bold and underlined.
[b] Significant good results in each group are indicated in bold.

**Table 12**
Summary of forecasting results based on proposed model.

| Kernel Situation | Single Kernel | | Hybrid Kernel | | Single Cluster | | Hybrid Two Cluster | | Hybrid Three Cluster | |
|---|---|---|---|---|---|---|---|---|---|---|
| Criteria | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| $C_2H_4$[a] | 0.001338 | 0.025277 | 0.001338 | 0.025260 | **0.001265** | 0.024220 | 0.001274 | **0.024159** | 0.001268 | 0.024188 |
| $C_3H_6$[a] | 0.001988 | 0.033704 | **0.001981** | **0.033668** | 0.001995 | 0.033710 | 0.002006 | 0.033807 | 0.002003 | 0.033760 |

[a] Best results regarding MSE are indicated in bold, and best results regarding MAE are indicated in bold and underlined.

**Table 13**
Forecasting Result of comparison test for ethylene.

| $C_2H_4$[a] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Proportions | 50% − 50% | | 60% − 40% | | 70% − 30% | | 80% − 20% | | 90% − 10% | |
| Model | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ANN | **0.002297** | **0.036547** | **0.002042** | **0.033629** | **0.002339** | **0.034124** | **0.002370** | **0.034774** | 0.002070 | 0.032728 |
| CNN | 0.024763 | 0.125945 | 0.015329 | 0.097230 | 0.012234 | 0.087620 | 0.010889 | 0.091608 | 0.016465 | 0.108881 |
| RNN | 0.031805 | 0.157601 | 0.070685 | 0.248327 | 0.007294 | 0.071135 | 0.020908 | 0.116039 | 0.014699 | 0.106425 |
| ARIMA | 0.060204 | 0.212394 | 0.057600 | 0.207858 | 0.010030 | 0.085544 | 0.002458 | 0.039199 | **0.001527** | **0.027147** |
| LSTM | 0.024766 | 0.132009 | 0.009400 | 0.068151 | 0.005690 | 0.056198 | 0.006234 | 0.061259 | 0.006904 | 0.062581 |
| RF | 0.022741 | 0.130386 | 0.023567 | 0.127344 | 0.015858 | 0.091287 | 0.018452 | 0.104594 | 0.029111 | 0.156611 |
| BBP | 0.032244 | 0.161673 | 0.054239 | 0.215113 | 0.006692 | 0.061815 | 0.021797 | 0.124884 | 0.017308 | 0.110983 |
| BDLRF | 0.033500 | 0.141784 | 0.039044 | 0.181855 | 0.006646 | 0.058963 | 0.008934 | 0.071250 | 0.014166 | 0.100207 |

[a] Best results regarding MSE are indicated in bold, and best results regarding MAE are indicated in bold and underlined.

noteworthy composition for propylene yield forecasting. However, it is not as good a forecast as considering only $C10$ or $C8$.

### 4.3. Comparison studies

We conduct two kinds of comparison studies for this work. One compares the forecasting results of the proposed model and the models with the train-test split data approaches; the other compares the RVM and SVM under the same kernel functions and rolling window approach. The first one indicates the advantage of the rolling-window approach in forecasting the yields of MTO, and the second one illustrates the benefit of our proposed model with other vector machines under the same situations. The summary of the best forecasting results of ethylene and propylene in previous subsections are listed in Table 12.

For the train-test split data approach, the prevalent algorithms, including the traditional time-series models (autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM)) and the popular machine learning models (ANN, convolutional neural network (CNN), recurrent neural network (RNN), and RF) are chosen. Nabavi et al. (2009) provided the first forecasting work for the MTO yields with the Basic Backpropagation (BBP) and Backpropagation with Declining Learning-rate Factor (BDLRF) models, which are also considered as the comparative ones.

We set the proportions of the training and test data as $50\% - 50\%, 60\% - 40\%, 70\% - 30\%, 80\% - 20\%$, and $90\% - 10\%$ to conduct the forecasting, and the results are listed in Tables 13 and 14. ANN shows the best results except for the $90\% - 10\%$ proportion situation of ethylene and $50\% - 50\%$ proportion situation of propylene where the ARIMA outperforms others. However, the best results of the train-test split data approach are all worse than the best results of our proposed model shown in Table 12, which support the benefit of the rolling-window RVM compared to the traditional time-series and popular machine learning models.

For the comparison between RVM and SVM under same kernel and rolling-window setting, we list the results of SVM in Table 15, where we only show the Gaussian and Laplace kernel (other kernels or hybrid kernels provide worse results than Gaussian and Laplace kernels) with length of rolling-window setting of $2, 3$, and $4$ (lengths longer than 4 provided the significantly worse results). The results of RVM in Table 12 are slightly better than SVM in Table 15, which supports that RVM outperforms the SVM in forecasting the yields of MTO. Also, the results of SVM based on the rolling-window setting are significantly more profitable than those of Tables 13 and 14, where we can conclude that rolling-window approach is more suitable for the MTO process study than the traditional train-test split data approach concerning forecasting accu-

**Table 14**
Forecasting result of comparison test for propylene.

| $C_3H_6$[a] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Proportions | 50% − 50% | | 60% − 40% | | 70% − 30% | | 80% − 20% | | 90% − 10% | |
| Model | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ANN | 0.002680 | 0.039382 | **0.002587** | **0.039498** | **0.003322** | **0.044740** | **0.003079** | **0.043479** | **0.002693** | **0.039947** |
| CNN | 0.032520 | 0.151288 | 0.028813 | 0.147767 | 0.023069 | 0.127348 | 0.088234 | 0.268584 | 0.014083 | 0.097744 |
| RNN | 0.012660 | 0.083737 | 0.018054 | 0.116784 | 0.007702 | 0.069028 | 0.008684 | 0.067141 | 0.026368 | 0.146983 |
| ARIMA | **0.002283** | **0.035970** | 0.017172 | 0.112159 | 0.017444 | 0.113908 | 0.006532 | 0.066309 | 0.003591 | 0.047113 |
| LSTM | 0.009200 | 0.077333 | 0.007730 | 0.071492 | 0.006107 | 0.063976 | 0.011448 | 0.086046 | 0.028270 | 0.151644 |
| RF | 0.050740 | 0.199272 | 0.153930 | 0.347003 | 0.081269 | 0.250701 | 0.018127 | 0.107941 | 0.051639 | 0.194610 |
| BBP | 0.017303 | 0.110021 | 0.018673 | 0.119221 | 0.007736 | 0.072777 | 0.013665 | 0.089943 | 0.025773 | 0.138350 |
| BDLRF | 0.030981 | 0.153509 | 0.038721 | 0.178011 | 0.007974 | 0.072891 | 0.015786 | 0.093168 | 0.041091 | 0.183767 |

[a] Best results regarding MSE are indicated in bold, and best results regarding MAE are indicated in bold and underlined.

**Table 15**
Forecasting result for SVM with single kernel.

| $l$ | $C_2H_4$[a] | | | | $C_3H_6$[a] | | | |
|---|---|---|---|---|---|---|---|---|
| | Gaussian | | Laplace | | Gaussian | | Laplace | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 2 | **0.001334** | **0.025340** | **0.001355** | **0.025405** | **0.002027** | **0.033981** | **0.002036** | **0.034010** |
| 3 | 0.001670 | 0.029450 | 0.001695 | 0.029530 | 0.002419 | 0.037651 | 0.002434 | 0.037706 |
| 4 | 0.001947 | 0.032589 | 0.001980 | 0.032689 | 0.002743 | 0.040444 | 0.002769 | 0.040542 |

[a] Best results regarding MSE are indicated in bold, and best results regarding MAE are indicated in bold and underlined.

racy. Considering that the rolling-window method is more efficient, the rolling-window approach should be focused and further studied in engineering studies in the future.

## 5. Conclusions

Forecasting the yields of light olefins plays a paramount role in monitoring and optimizing the MTO process. It is not only directly related to the improvement of production efficiency and economic benefits, but also has a far-reaching impact on the enterprise's market share, environmental responsibility, technological innovation, and the stable operation of the whole industrial chain. In this paper, persistent yield data of light olefins are chosen as the dependent, and 32 yield impact indicators are utilized as the independent variables to carry out ethylene and propylene forecasting research. Based on the hybrid kernel RVM and the rolling-window approaches, the following conclusions are obtained:

(1) Regarding the single kernel function, the Laplace is optimal for ethylene yield forecasting, and the Gaussian is optimal for propylene yield forecasting. In terms of the hybrid kernel function, the most reasonable forecasting results are given by the Gaussian+Laplace kernels for both yields, with the Gaussian kernel being the leading one regarding propylene and the Laplace kernel being the leading one for ethylene. The results show that the DMTO yields impact indicators that provide little valuable information for forecasting; the squared difference and absolute value of the difference between the variables predominantly play decisive roles.

(2) The hybrid of local kernel functions has a substantial edge in forecasting the yields of DMTO over the hybrid of local and global kernels, suggesting that the yields of ethylene and propylene have both short volatility trends. Meanwhile, the rolling window study shows that four hours of training information is more optimal in forecasting than a longer window length, also supporting the statements of short volatility trends in DMTO yields. The conclusion is consistent with the traditional DMTO study, which provides a multi-faceted confirmation of the trend study.

(3) The 4-hour time window has the most satisfactory forecasting effect on the yields of DMTO, indicating that the fluctuation cycle is about 4 hours when considering all the variables. In cluster studies, when we consider fewer independent variables and the optimal window lengths become more prominent, the 32 independent variables are all necessary in this study to provide sufficient information for forecasting. Compared with the previous DMTO forecasting studies using train-test split data approaches, this paper's optimal rolling time window can achieve efficient forecasting and increase the practicability of the research content. Moreover, it can provide researchers with data to support the development of more efficient MTO catalysts and processes.

(4) The significant clusters of variables can be effectively screened by the single cluster and mixed cluster studies. In particular, temperature is the most influential factor in ethylene forecasting; pressure and refining of heavy olefins are also consequential. In addition, the combination of temperature and pressure produces better results for ethylene yield forecasting. The refining of heavy olefins is most vital for propylene forecasting, and pressure and temperature are also significant, but the combination of clusters does not have a greater impact on propylene yield forecasts.

It is worth noting that temperature and pressure have the most significant impact on the prediction of ethylene and propylene yields, with ethylene being more sensitive to temperature and propylene being more affected by pressure. In future applications, researchers optimize the MTO process by making timely adjustments to the dosages of these important variables to increase ethylene and propylene yields.

## CRediT authorship contribution statement

**Wenyang Wang:** Writing – original draft, Software, Project administration, Methodology, Funding acquisition, Conceptualization. **Nan He:** Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jie Liu:** Writing – original draft, Visualization, Formal analysis. **Muxin Chen:** Writing – original draft, Visualization, Software, Methodology, Data curation. **Jibin Zhou:** Writing – original draft, Funding acquisition, Formal analysis. **Tao Zhang:** Writing – original draft, Funding acquisition, Formal analysis. **Mao Ye:** Supervision, Formal analysis. **Zhongmin Liu:** Supervision, Formal analysis.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

This work was supported by the Humanities and Social Sciences Youth Foundation, Ministry of Education of the People's Republic of China (22YJC910011), the China Postdoctoral Science Foundation Funded Project (2023M733444), the National Natural Science Foundation of China (22308348), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA29050200).

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ces.2024.120656.

## References

Abdi, J., Mazloom, G., Hadavimoghaddam, F., Hemmati-Sarapardeh, A., Esmaeili-Faraj, S.H., Bolhasani, A., Karamian, S., Hosseini, S., 2023. Estimation of the flow rate of pyrolysis gasoline, ethylene, and propylene in an industrial olefin plant using machine learning approaches. Sci. Rep. 13, 14081.

Acosta, S.M., Amoroso, A.L., Anna, A.M.O.S., Junior, O.C., 2021. Relevance vector machine with tuning based on self-adaptive differential evolution approach for predictive modelling of a chemical process. Appl. Math. Model. 95, 125–142.

Aliabadi, M.M., Emami, H., Dong, M., Huang, Y., 2020. Attention-based recurrent neural network for multistep-ahead prediction of process performance. Comput. Chem. Eng. 140, 106931.

Alvarez, M.A., Rosasco, L., Lawrence, N.D., et al., 2012. Kernels for vector-valued functions: a review. Found. Trends Mach. Learn. 4, 195–266.

Bi, X., Zhao, J., 2021. A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification. Process Saf. Environ. Prot. 156, 581–597.

Chang, C.D., 1984. Methanol conversion to light olefins. Catal. Rev. Sci. Eng. 26, 323–345.

Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., Bukkapatnam, S.T., 2015. Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. IIE Trans. 47, 1053–1071.

Claesen, M., De Moor, B., 2015. Hyperparameter search in machine learning. arXiv preprint. arXiv:1502.02127.

Coley, C.W., Barzilay, R., Jaakkola, T.S., Green, W.H., Jensen, K.F., 2017. Prediction of organic reaction outcomes using machine learning. ACS Cent. Sci. 3, 434–443.

De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. Int. J. Forecast. 22, 443–473.

Dey, B., Roy, B., Datta, S., Ustun, T.S., 2023. Forecasting ethanol demand in India to meet future blending targets: a comparison of arima and various regression models. Energy Rep. 9, 411–418.

Ding, S., Zhang, Y., Xu, X., Bao, L., 2013. A novel extreme learning machine based on hybrid kernel function. J. Comput. 8, 2110–2117.

Dioşan, L., Rogozan, A., Pecuchet, J.P., 2012. Improving classification performance of support vector machine by genetically optimising kernel shape and hyper-parameters. Appl. Intell. 36, 280–294.

Fan, C., Liu, K., Ren, Y., Peng, Q., 2024. Characterization and identification towards dynamic-based electrical modeling of lithium-ion batteries. J. Energy Chem.

Fang, F., Zang, G.L., Sun, M., Yu, H.Q., 2013. Optimizing multi-variables of microbial fuel cell for electricity generation with an integrated modeling and experimental approach. Appl. Energy 110, 98–103.

Fokoué, E., Sun, D., Goel, P., 2011. Fully Bayesian analysis of the relevance vector machine with an extended hierarchical prior structure. Stat. Methodol. 8, 83–96.

Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K., Sriperumbudur, B.K., 2012. Optimal kernel choice for large-scale two-sample tests. Adv. Neural Inf. Process. Syst. 25.

Hartingan, J., Wong, M.K., 1979. Algorithm AS 136: a k-means clustering algorithm. Appl. Stat. 28, 100–108.

Hu, J., Zheng, W., 2020. Multistage attention network for multivariate time series prediction. Neurocomputing 383, 122–137.

Huang, X., Li, X.G., Xiao, W.D., Wei, Z., 2023. Machine learning-assisted multiscale modeling of an autothermal fixed-bed reactor for methanol to propylene process. AIChE J. 69, e17945.

Inoue, A., Jin, L., Rossi, B., 2017. Rolling window selection for out-of-sample forecasting with time-varying parameters. J. Econom. 196, 55–67.

Jiang, B., Dai, H., Wei, X., Jiang, Z., 2021. Multi-kernel relevance vector machine with parameter optimization for cycling aging prediction of lithium-ion batteries. IEEE J. Emerg. Sel. Top. Power Electron.

Jiao, F., Li, J., Pan, X., Xiao, J., Li, H., Ma, H., Wei, M., Pan, Y., Zhou, Z., Li, M., et al., 2016. Selective conversion of syngas to light olefins. Science 351, 1065–1068.

Kinney, J.B., Atwal, G.S., 2014. Equitability, mutual information, and the maximal information coefficient. Proc. Natl. Acad. Sci. 111, 3354–3359.

Kumar, S., Hussain, L., Banarjee, S., Reza, M., 2018. Energy load forecasting using deep learning approach-lstm and gru in spark cluster. In: 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT). IEEE, pp. 1–4.

Lall, U., Moon, Y.i., Bosworth, K., 1993. Kernel flood frequency estimators: bandwidth selection and kernel choice. Water Resour. Res. 29, 1003–1015.

Lerman, P., 1980. Fitting segmented regression models by grid search. J. R. Stat. Soc., Ser. C, Appl. Stat. 29, 77–84.

Li, J., Yang, B., Li, H., Wang, Y., Qi, C., Liu, Y., 2021. Dtdr–alstm: extracting dynamic time-delays to reconstruct multivariate data for improving attention-based lstm industrial time series prediction models. Knowl.-Based Syst. 211, 106508.

Lim, B., Arık, S.Ö., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int. J. Forecast. 37, 1748–1764.

Linninger, A.A., 2002. Metallurgical process design a tribute to Douglas' conceptual design approach. Ind. Eng. Chem. Res. 41, 3797–3805.

Long, J., Chen, Y., Cao, D., Chen, P., Yang, M., 2023. Yield and properties prediction based on the multicondition lstm model for the solvent deasphalting process. ACS Omega 8, 5437–5450.

Min, J.H., Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Syst. Appl. 28, 603–614.

Moghaddam, A.H., 2023. Investigation and optimization of olefin purification in methanol-to-olefin process based on machine learning approach coupled with genetic algorithm. Korean J. Chem. Eng. 40, 1168–1175.

Monai, M., Gambino, M., Wannakao, S., Weckhuysen, B.M., 2021. Propane to olefins tandem catalysis: a selective route towards light olefins production. Chem. Soc. Rev. 50, 11503–11529.

Nabavi, R., Salari, D., Niaei, A., Vakil-Baghmisheh, M.T., 2009. A neural network approach for prediction of main product yields in methanol to olefins process. Int. J. Chem. React. Eng. 7.

Sedgwick, P., 2012. Pearson's correlation coefficient. BMJ 345.

Tao, Y., Ma, L., Zhang, W., Liu, J., Liu, W., Du, Q., 2018. Hierarchical attention-based recurrent highway networks for time series prediction. arXiv:1806.00685.

Tian, P., Wei, Y., Ye, M., Liu, Z., 2015. Methanol to olefins (mto): from fundamentals to commercialization. ACS Catal. 5, 1922–1938.

Tipping, M., 1999. The relevance vector machine. Adv. Neural Inf. Process. Syst. 12.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.

Wang, J., Peng, Z., Wang, X., Li, C., Wu, J., 2020a. Deep fuzzy cognitive maps for interpretable multivariate time series prediction. IEEE Trans. Fuzzy Syst. 29, 2647–2660.

Wang, J.Q., Du, Y., Wang, J., 2020b. Lstm based long-term energy consumption prediction with periodicity. Energy 197, 117197.

Wang, T., Zhao, D., Tian, S., 2015. An overview of kernel alignment and its applications. Artif. Intell. Rev. 43, 179–192.

Wang, W., Sun, Y., Li, K., Wang, J., He, C., Sun, D., 2023. Fully Bayesian analysis of the relevance vector machine classification for imbalanced data problem. CAAI Trans. Intell. Technol. 8, 192–205.

Wang, Y., Zhang, Y., Wu, Z., Li, H., Christofides, P.D., 2020c. Operational trend prediction and classification for chemical processes: a novel convolutional neural network method based on symbolic hierarchical clustering. Chem. Eng. Sci. 225, 115796.

Wang, Z., Wang, L., Yuan, Z., Chen, B., 2021. Data-driven optimal operation of the industrial methanol to olefin process based on relevance vector machine. Chin. J. Chem. Eng. 34, 106–115.

Wu, Z., Mu, Y., Deng, S., Li, Y., 2022. Spatial–temporal short-term load forecasting framework via k-shape time series clustering method and graph convolutional networks. Energy Rep. 8, 8752–8766.

Xu, L., Niu, X., Xie, J., Abel, A., Luo, B., 2015. A local–global mixed kernel with reproducing property. Neurocomputing 168, 190–199.

Xu, Y., Mi, C., Zhu, Q.X., Gao, J.Y., He, Y.L., 2017. An effective high-quality prediction intervals construction method based on parallel bootstrapped rvm for complex chemical processes. Chemom. Intell. Lab. Syst. 171, 161–169.

Yang, J., Ning, C., Deb, C., Zhang, F., Cheong, D., Lee, S.E., Sekhar, C., Tham, K.W., 2017. k-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy Build. 146, 27–37.

Yang, Y., Xiong, Q., Wu, C., Zou, Q., Yu, Y., Yi, H., Gao, M., 2021. A study on water quality prediction by a hybrid cnn-lstm model with attention mechanism. Environ. Sci. Pollut. Res. Int. 28, 55129–55139.

Ying, L., Yuan, X., Ye, M., Cheng, Y., Li, X., Liu, Z., 2015. A seven lumped kinetic model for industrial catalyst in dmto process. Chem. Eng. Res. Des. 100, 179–191.

Zhang, X., Zou, Y., Li, S., Xu, S., 2019. A weighted auto regressive lstm based approach for chemical processes modeling. Neurocomputing 367, 64–74.

Zhou, J., Liu, D., Ye, M., Liu, Z., 2024. Spatial-temporal self-attention network based on Bayesian optimization for light olefins yields prediction in methanol-to-olefins process. Artif. Intell. Chem. 2, 100067.