



HAL
open science

Multi-output machine learning models for kinetic data evaluation: A Fischer–Tropsch synthesis case study

Anoop Chakkingal, Pieter Janssens, Jeroen Poissonnier, Mirella Virginie, Andrei Khodakov, Joris Thybaut

► To cite this version:

Anoop Chakkingal, Pieter Janssens, Jeroen Poissonnier, Mirella Virginie, Andrei Khodakov, et al.. Multi-output machine learning models for kinetic data evaluation: A Fischer–Tropsch synthesis case study. Chemical Engineering Journal, 2022, 446, pp.137186. 10.1016/j.cej.2022.137186 . hal-03863277

HAL Id: hal-03863277

<https://cnrs.hal.science/hal-03863277>

Submitted on 25 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-output machine learning models for evaluating
kinetic data:
A Fischer-Tropsch Synthesis Case Study

Anoop Chakkingal^{a,b}, Pieter Janssens^a, Jeroen Poissonnier^a, Mirella
Virginie^b, Andrei Y. Khodakov^b, Joris W. Thybaut^{a,*}

^a*Laboratory for Chemical Technology (LCT), Department of Materials, Textiles and
Chemical Engineering, Ghent University, Technologiepark 125, 9052 Ghent, Belgium*

^b*CNRS, Centrale Lille, Univ. Lille, ENSCL, Univ. Artois, UMR 8181 – UCCS – Unité
de Catalyse et Chimie du Solide, F-59000 Lille, France*

*Corresponding author

Email address: `joris.thybaut@ugent.be` (Anoop Chakkingal)

Abstract

Predicting the impact of input process variables on chemical processes is key to assess their performance. However, the models explaining this impact through a mechanistic approach are either not readily available or are complex in nature. With increased automation in industries and the availability of high-throughput experimental data, data-driven machine learning models are gaining popularity due to their simplicity and reduced computational effort. In this work, a multi-output Fischer-Tropsch Synthesis process is analyzed with different machine learning models (ML) such as Lasso, K Nearest neighbors (KNN), Support Vector Regression (SVR), and Artificial Neural Network (ANN) regression. The investigated Fischer-Tropsch Synthesis case-study has shown that the temperature and pressure are the dominant input variables. The extent to which multi-response ML models are capable to capture the key trends (i.e., the non-linear conversion and selectivity *w.r.t.* the input variables) of the governing chemical kinetics was evaluated and ANN emerged as the superior performing ML model with respect to the benchmark SEMK results. In addition, the validity of neural network predictions is verified using the so-called Shap-value interpretation technique. The relative impact of input variables obtained using Shap values, on conversion follow the order of temperature (1x) > pressure (0.7x) > space-time (0.5x) > syngas ratio (0.25x). On the other hand, the influence of temperature (1x) and pressure (0.7x) remain the same for light olefin selectivity, but that of space-time (0.25x) and syngas ratio (0.5x) switches. This study provides a reference method for the identification of suitable ML models for multi-output prediction in chemical processes.

Keywords: Fischer-Tropsch Synthesis, Single-Event MicroKinetics, Lasso, K-Nearest Neighbor, Support Vector Regression, Artificial Neural Network, Interpretability, Feature importance, Shap value

Nomenclature

F_{CO} Carbon monoxide molar flow rate at the reactor inlet, mol.s^{-1}

H_2/CO Syngas molar inlet ratio, mol.mol^{-1}

r_p Pearson correlation coefficient

W/F_{CO} space-time, $(\text{kg}_{\text{cat}}\text{s})\text{mol}_{CO}^{-1}$

ANN Artificial Neural Network

FTS Fischer-Tropsch Synthesis

KNN K Nearest Neighbor

ML Machine Learning

MSE Mean squared error, i.e. squared difference between actual and predicted output

RMSE Root of MSE

SEMK Single-Event MicroKinetics

SVR Support Vector Regression

W Catalyst mass, kg

1. Introduction

With the increased volume of data obtained via automation and high-throughput experimentation, data-driven models are becoming increasingly popular and, as a result, machine learning (ML) methods are extensively investigated to assess chemical engineering problems [1–3] where the impact of input variables is highly non-linear in nature [4]. Depending on the process data complexity, the specific type of machine learning method suited to model the phenomena of interest varies. At present, from the modeling perspectives, chemical engineering research challenges are mostly addressed via mechanistic models [5–7] which rely on imposing physico-chemically meaningful relations in the data. However, not all features (mechanistic aspects) are equally contributing and, hence, some particular features are challenging to probe. Also, developing mechanistic models for every chemical engineering problem is not always feasible due to their complexity and the in-depth knowledge required to build such models. This is especially true in the field of applied catalysis where the reaction complexity is more pronounced and the use of simpler kinetic models is not able to capture the detailed chemistry of the process. This presents an opportunity for the development of ML models which are easier [8] to develop as compared to a detailed mechanistic model.

Machine learning methods are typically approached as:

- **unsupervised learning:** ML methods in which patterns are inferred from the unlabeled input data, i.e. without output variables. Unsupervised ML methods identify structure and patterns from the input data by themselves. For e.g. clustering [9], dimensionality reduction [10].

- **supervised learning:** ML methods in which models are trained using labeled data. The supervised ML models determine a mapping function between the input variables and the output variables. For e.g. classification [11], regression [12].

In chemical engineering, supervised learning methods (mainly regression-based methods) are the widely used, in areas such as absorption [13], sludge treatment [14], reactor modelling [15], etc. ML regression methods reported in literature involves the use of linear regression [12], decision tree [16], Support Vector Regressions (SVR) [17], Artificial Neural Network (ANN) [18] etc. The wide application of these techniques in catalysis has recently been reported by Takahashi et al. [19]. As the model complexity increases from Lasso to ANN, the interpretability of the model decreases [20]. To circumvent this, and for a systematic analysis and interpretation of kinetic data, model agnostic interpretation technique Shap could be used [21]. This helps in building confidence on the model’s ability to accurately draw chemical insights. The data required to develop the majority of ML-based models is obtained from experiments [22, 23]. Although the vast majority of the reported ML studies relate to single response scenarios, it should be pointed out that these methodologies can also be applied to multi-response scenarios.

Data-driven modeling in catalysis, also referred to as catalysis informatics is an alternative modeling approach [19, 24] and has led to the re-investigation of many existing reactions, with the demand for developing a sustainable chemical production process as one of the key drivers. Modeling the recycling of plastics is one of the important topics currently investigated, due to the role in environmental pollution [25]. Within the plastic chemical recycling

processes, the Fischer-Tropsch Synthesis (FTS) may play an important role. In FTS [26], the syngas generated via the gasification of plastic waste is catalytically converted into hydrocarbons such as paraffins and light olefins with the aid of a catalyst. An important step in the optimization of the FTS process involves the screening of potential catalysts [27]. There are different types of catalysts, e.g. cobalt, iron, etc. reported in the FTS literature that selectively favor the formation of methane, paraffin or olefins [28]. Several experimental [29–31] and mechanistic/kinetic studies [6, 32] that aim at identifying the properties which steer these selectivities in FTS have been reported. Among the mechanistic models, kinetic model based on Single Event MicroKinetics (SEMK) has been widely applied in the field of oligomerization [33], cracking [34], alkylation [35] etc. and has emerged as a versatile tool to model the FTS [36, 37]. A Single Event MicroKinetic evaluation of the key catalyst properties for enhanced light olefin synthesis via FTS was reported in our previous work [38].

Studies using machine learning methods such as neural network (NN) [39, 40] and support vector regression (SVR) [41] to predict a single response FTS process have also been reported in literature, as an alternative to detailed kinetic models. Compared to that, only a limited number publications on ML methods for a multi-response scenario [42, 43], such as conversion and selectivities in the FTS process, have been reported. Investigating the extent to which these models can match the results obtained with mechanistic models like SEMK, especially in the scenarios with multiple outputs can help us to select a suitable and easy-to-implement ML model. In this work, machine-learning algorithms such as Lasso algorithms [44], K-nearest neigh-

bor (KNN) [45], support vector regression (SVR) [46], and Artificial Neural Network (ANN) [47, 48] regression are evaluated with respect to in-silico SEMK data. This data is obtained at different operating conditions by varying temperature, pressure, space-time (W/F_{CO}), and syngas ratio (H_2/CO) on the promising FTS catalyst exhibiting higher light olefin selectivity as identified in our previous work [38]. This benchmark dataset is then used to investigate the potential of these machine learning models to reproduce and / or predict multiple outputs such as conversion (X_{CO}) and the selectivity towards methane (S_{CH_4}), paraffins ($S_{C_2-C_4}$) and light olefins ($S_{C_2-C_4=}$). A suitable multi-output ML model that matches SEMK results was subsequently identified and the interpretability of the ML model-predictions is confirmed by using Shap values-based interpretation technique [49, 50].

2. Theory

2.1. Single Event MicroKinetic model

To quantify the contributions of competing reaction pathways in a complex mixture process, Single Event MicroKinetic methodology serves as a versatile tool to unravel the complex kinetics of a process. The model-parameters of an SEMK model are classified as kinetic and catalyst descriptors. The kinetic descriptors relate to the kinetic parameters of the chemical reactions such as activation energies and rate coefficients in a catalyst invariant. To quantify the effect of catalyst properties on the chemical reaction, catalyst descriptors such as atomic chemisorption enthalpies are utilized [36]. The parameters governing the microkinetic model are generally obtained after multi-response model regression to the experimental data.

Such detailed calculations on Single Event MicroKinetic studies are reported by, e.g., Lozano-Blanco et al.[36, 51].

2.2. Machine learning models

The increase in the available volume of data and the urge for less complex models has also led to the increased interest in machine learning based models. In the current work, various machine learning methods (Fig.1) are compared to determine the requirements of a model to be suitable for predicting the impact of operating conditions in FTS for multiple outputs rather than a single one. The most commonly used ML regression methods: Lasso, KNN, SVR, and ANN are analyzed here in detail. The predictions made by each of these models are then compared with the microkinetic model predictions and physico-chemical understanding.

2.2.1. Lasso regression

The most general supervised ML method is linear regression. It calculates the dependent (output) variables, (y_i) based on the relationship with the independent (input) variables (x_i) through the parameters also denoted as weights β_i . In the model fitting, linear regression can suffer from model over-fitting where the model fits the training data but does not give a good test data prediction. This is due to the learning of noises in the test data and manifests itself via large parameter estimates associated to independent variables of lesser impact. This issue is addressed by Lasso regression [44], where regularization or shrinkage of the parameters is done to reduce the over-fitting. It penalizes the less important inputs in the dataset and thus creates a simple model by reducing their respective coefficients. Here the

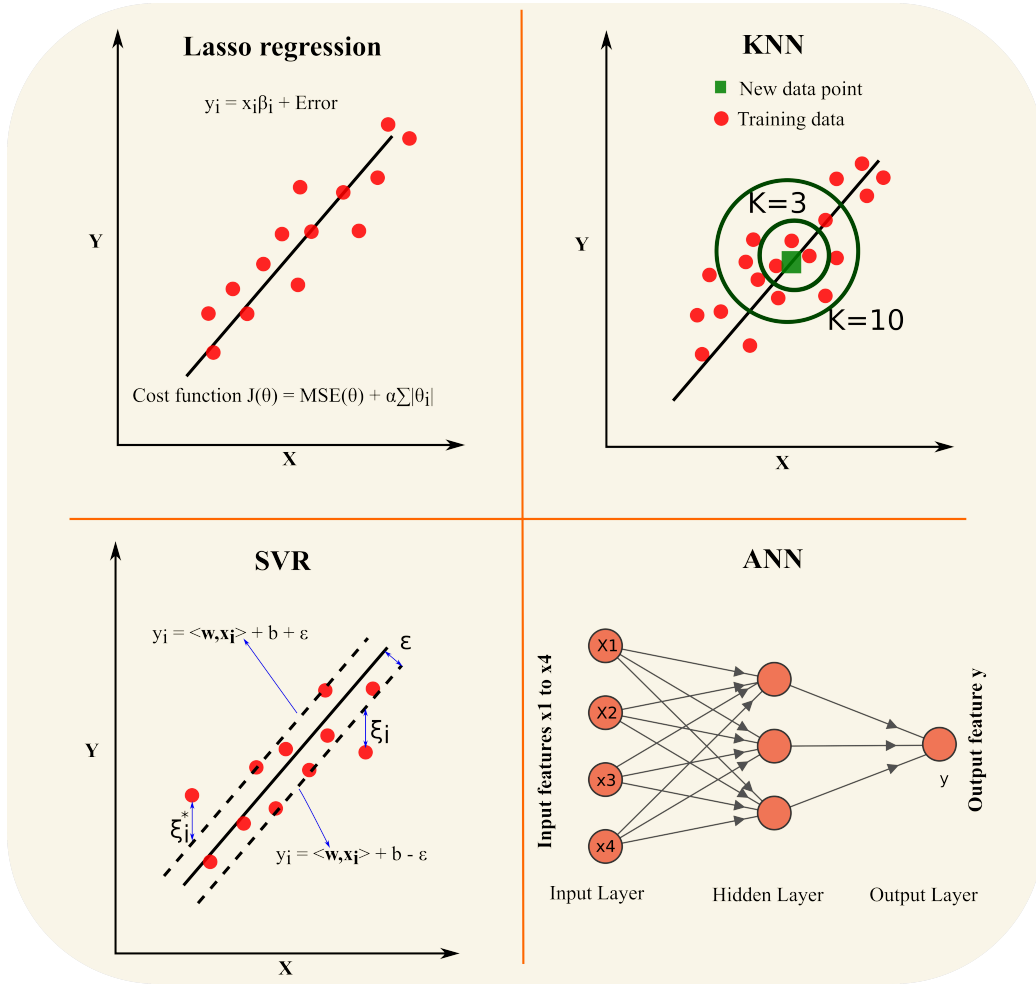


Figure 1: Different machine learning models investigated to analyse multi-output FTS reaction.

L1 norm is used for regularization. The cost function containing the sum of mean square error (MSE) and the penalty as shown in Eq.1, is minimized to find the parameters in the Lasso regression.

$$J(\beta) = MSE(\beta) + \alpha \sum_{i=1}^n |\beta| \quad (1)$$

where α is the hyperparameter that controls the intensity of regularization.

2.2.2. *K Nearest neighbors (KNN) regression*

KNN is a memory-based method [45] that makes predictions for new observations based on its “similarity” to the data used for training. It identifies k observations that are nearest or similar to a newly considered observations, and then assigns the average response of these k observations as the prediction for the new one. The performance of the KNN model is sensitive to the parameter k , which determines the smoothness of the estimation. For low values of k , less neighbors are considered leading to over-fitting, while using large values of k considers more neighbors leading to under-fitting. The similarity between observations (to identify the neighbors) is quantified using the Euclidean distance metric, where the distance between observations x_a and x_b for all input variables is calculated as:

$$d = \sqrt{\sum_{i=1}^n (x_a - x_b)^2} \quad (2)$$

The main parameters associated with the KNN regression include the number of neighbors to be used and the algorithm used to compute the nearest neighbors.

2.2.3. *Support vector regression*

Support Vector Machine (SVM) [46] is an algorithm widely used for solving machine learning problems. Support Vector Regression (SVR) is the most common application form of SVM for regression purposes when the dataset is non-separable. Unlike the least square approach which is sensitive to the outliers due to the use of squared residuals, SVR uses ϵ -insensitive loss as the loss metric which is more robust to potential outliers. The ϵ -insensitive

loss is given as:

$$Le = \max(0, |r(x, y)| - \epsilon) \quad (3)$$

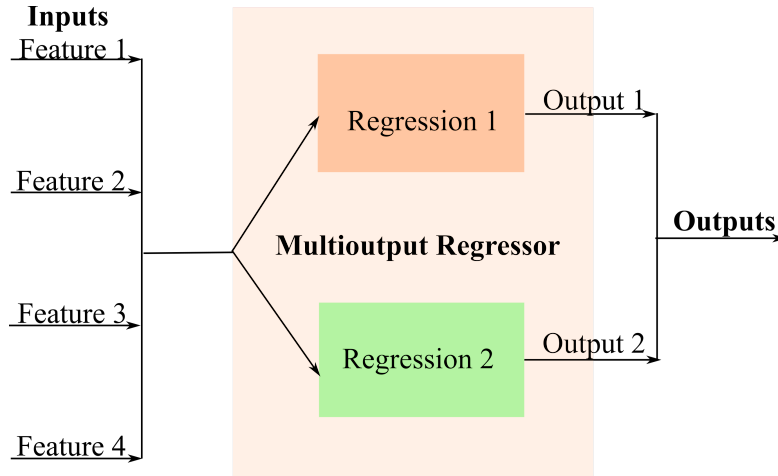


Figure 2: Wrapper function in multi-response support vector regression. Separate regression results for each output is wrapped together by this model.

where $r(x,y)$ is the residual, i.e. the difference between the actual and the predicted output. A margin of width ϵ is created around the regression curve (known as hyperplane in higher dimensional space) within which the exact magnitude of the discrepancy between observation and model prediction does not contribute to the loss function and hence does not influence the regression curve. The margin is defined by the data points that satisfy $r(x, y) \pm \epsilon$. These data points are known as support vectors. When the data points exhibit a non-linear relationship, the data is analyzed in an enlarged feature space. Special functions called kernel functions are used to construct the enlarged feature space. ¹

¹The popular kernel functions used are d^{th} degree polynomial, radial basis function

For multi-response regression, the SVR requires a wrapper function (Fig.2) to combine the separate regression prediction of target outputs, i.e. the interactions among outputs are not considered and thus multi-response prediction is not inherent to this method.

2.2.4. Artificial Neural Network

An artificial neural network (ANN) is a machine learning algorithm [47] inspired by biological neural networks. It is a data-driven model which reveals the hidden patterns or non-linear relationships in the data. The neural network model is composed of its basic units called neurons, which are then stacked to form layers. An ANN network will contain an input layer, hidden layers, and an output layer. Each neuron is connected to the adjacent layer neuron through the weight assigned to it, which does the linear transformation of the input signal in the forward feed propagation step. The activation function assigned to it does the corresponding non-linear transformation. The most commonly used activation functions are Rectified Linear Unit (ReLU), sigmoid, and tanh. In the network training process, recalculation and assigning the new weights happen through the backpropagation step. This process continues until the difference between the prediction and the actual target is within the tolerance limit. If the number of neurons/layers

(RBF), and hyperbolic tangent. The data points that satisfy $r(x,y) \pm \epsilon$ in the kernel induced enlarged feature space form the support vectors. A penalty coefficient C that controls the strength of the penalty term (as defined by the loss function above) is obtained by hyperparameter optimization. Each kernel function mentioned above has a set of hyperparameters (e.g 'spread' of the kernel and therefore the decision region, γ for RBF) that also needs to be optimized [52].

is too limited, it reduces the analysis capability of the network and gives less accuracy in prediction. On the contrary, if the number is high, it results in over-fitting (memorizing) of the data. The optimal number of neurons and layers is found by hyperparameter optimization, or by a trial and error method, which results in a network that yields a perfect prediction with minimization of a loss function.

2.2.5. Interpretability of ML models

With the increase in complexity of ML models (for e.g. ANN model) they become "black box" model, and the interpretation of the models become challenging, when compared to models such as linear regression. For a "simple" model such as linear regression, the weights or coefficients of the independent variables gives an indication of the importance of each variable. As the model complexity increases such as in ANN, though the accuracy of prediction increases, the models become almost uninterpretable [20]. Different interpretation techniques such as permutation importance [53] and Shap values [49, 50, 54] are now being used to overcome this issue. In the current work, the interpretation technique based on Shap values is used. Shap values help in local interpretation by pinpointing the contribution of input variables (features) in each set of operating conditions. This technique builds multiple linear, more easily interpretable models, see Fig.3, which describe the individual data points and thus helps to interpret a complex ML model.

3. Methods and data

In the current work, a dataset is generated using a Single Event MicroKinetic model, validated for Fischer-Tropsch Synthesis experiments. In our

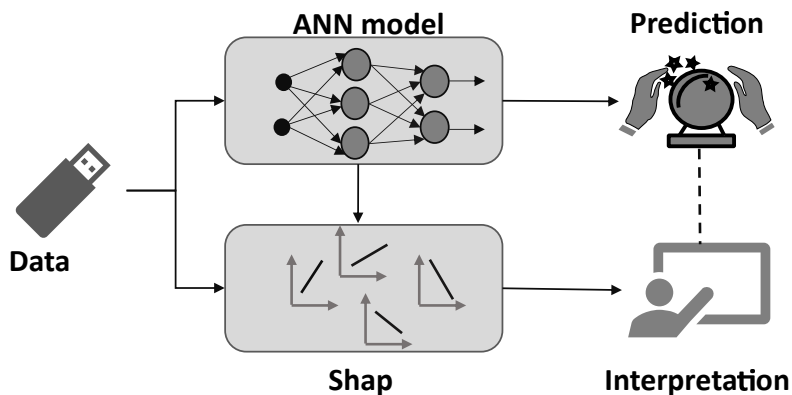


Figure 3: Graphical representation of how a Shap model assists in the interpretation of complex model with the help of linear models. An ANN model is chosen as an example to represent a complex model.

previous work [38], the catalytic descriptors were obtained for an optimally performing virtual catalyst for light olefin production. Our previous work focused on catalyst screening at a single set of operating conditions. Here, we extend our analysis to obtain the multi-response (multi-output response) within a wide range of operating conditions. These are chosen such that the data is uniformly distributed, and there is no concentration of data at a particular condition, which could lead to a biased prediction. The range of operating conditions investigated is as follows:

- temperature: 480-650 K
- total pressure: 1-15 bar(a)
- space-time, W/F_{CO} : 24-68 kg.s/mol
- syngas molar inlet ratio, H_2/CO : 1-5 mol/mol

A total of 2450 synthetic experiment responses were generated and a train

to test split ratio of 80:20 is used in the analysis. Thus, the dataset consists of 4 input variables (temperature, pressure, space-time, syngas ratio) and 4 output variables (conversion, selectivity of methane, paraffins, and light olefins). The generated dataset is scaled and centered using standardization to facilitate the learning by the machine learning models. In the training phase, internal model parameters are learned from the data. The hyperparameter controls the learning process and determines the model parameters the algorithm ends up learning.² The parameters obtained after hyperparameter tuning for the model are as follows:

- Lasso regression: $\alpha = 0.1$
- KNN regression: number of neighbors = 6, weight function = distance
- SVR: $C = 10$, γ for the radial basis function kernel = 0.1
- ANN: number of hidden layers = 1, number of neurons in the hidden layer = 20, activation function = sigmoid (input/output layer) and Relu (hidden layer), loss = MSE

²Hyperparameters are tuned to get the best fit for the ML models. In the current analysis, the simplest strategy based on Grid search which involves forming a grid of the search space for evaluating the hyperparameter is used. K-fold cross-validation is used as the re-sampling procedure to evaluate ML models on the data sample and thus prevent over-fitting [55].

4. Results and discussion

4.1. Analysis of SEMK data

The Pearson correlation coefficient (r_p) between a pair of variables (the input variables and the targeted FTS outputs) calculated as in Eq.4 is shown in Fig.4.

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

where the absolute value of indicates the degree of correlation between variables x and y . If $r_p > 0$, there exists a positive correlation between two variables x and y , and if $r_p < 0$, there exists a negative correlation.

The correlation plot (Fig.4) shows that the temperature is the most important input variable as indicated by the magnitude of the Pearson correlation coefficient (dark red color). The positive correlation indicates that an increase in temperature results in an enhanced formation of FTS products. For this highly active and light olefin selective catalyst, the effect of space-time and syngas ratio on the conversion and selectivities is limited. The impact of total pressure on the FTS reaction varies according to the targeted output variable. There is a slight increase in conversion and paraffin selectivity upon a pressure increase, whereas light olefin selectivity increases upon a decrease in pressure (blue).

The above relation is also visible from the joint plot (Fig.5) drawn for the most sensitive input variable, i.e. temperature. Each joint plot in Fig.5 consists of 3 separate plots, 1 relationship plot, and 2 marginal distribution plots. For e.g. Fig.5(a), the grey-colored contours show the relationship between

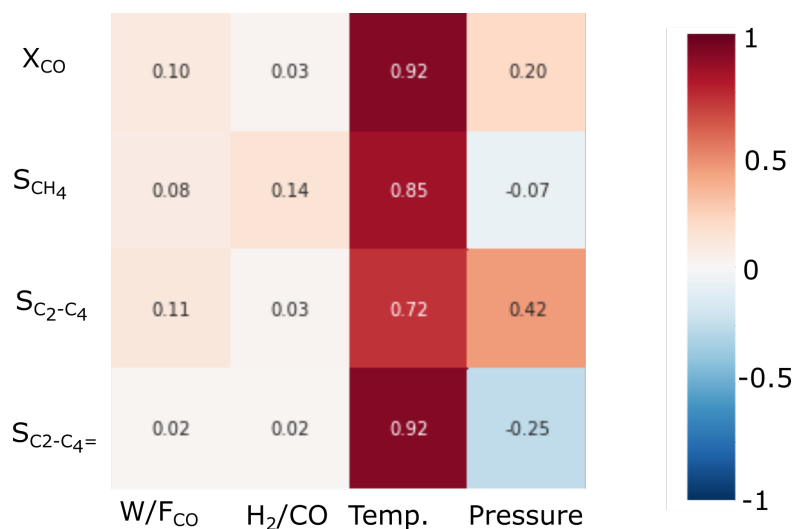


Figure 4: Heatmap of Pearson correlation between input variables: temperature, pressure, space-time and syngas ratio with output variables: conversion and selectivity of methane, paraffins and light olefins.

conversion and temperature (relationship plot). A darker contour indicates a higher concentration of data points in this region. The two distributions at the top and right side of Fig.5(a) give us the marginal distribution of temperature and conversion. The marginal distribution of temperature is obtained by projecting it to a new horizontal axis and then plotting its frequency distribution. Similarly, the marginal distribution of conversion is obtained by projecting the data to a new vertical axis.

In Fig.5(a), darker contours are observed at a lower temperature, indicating that the impact of other input variables on conversion is minimal. As the temperature increases the spread of the contour increases, indicating a higher impact of other input variables on conversion. Similar behavior is also observed for the selectivity of light olefins (Fig.5(b)). Higher conversion and

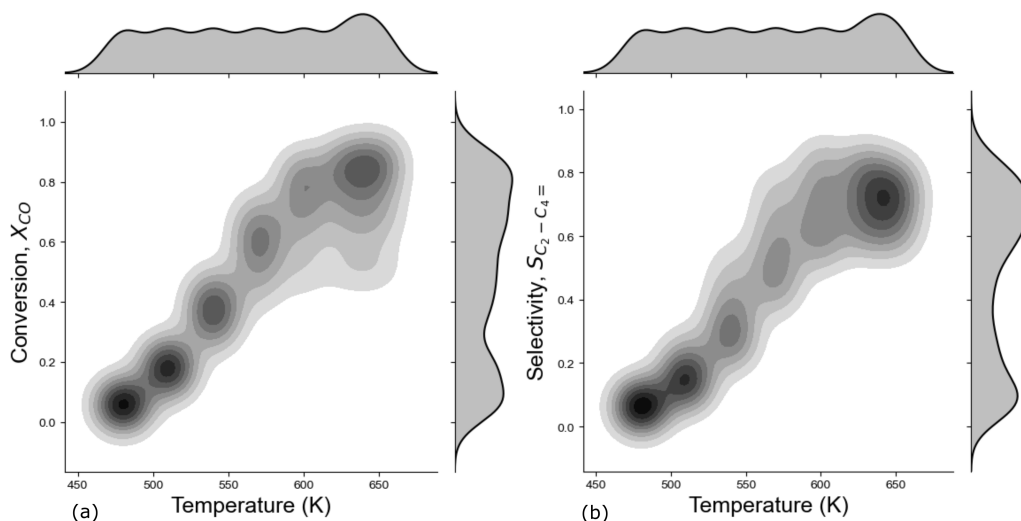


Figure 5: Conversion (a) and selectivity(b) behaviour as a function of temperature for the entire dataset.

light olefin selectivity are obtained in the temperature range 600-650 K. It is also observed that the conversion and light olefin selectivity reach a plateau in this temperature range. This behavior is in line with that observed by Garona et.al [43].

From Fig.5 it is observed that the general trend of change in conversion and light olefin selectivity with temperature is similar. Thus, to understand the combined impact of other input variables, the yield of light olefins is visualized with the help of pairwise box plots (Fig.6). The variation of light olefin yield with the temperature at different syngas ratios (H_2/CO - 2 and 5) is shown in Fig.6(a). It is observed that light olefin yield increases with temperature. The influence of syngas ratio on the mean yield of light olefins is limited, thus indicating a weak influence of the syngas ratio on the total yield of light olefins. The variation of light olefin yield with the temperature

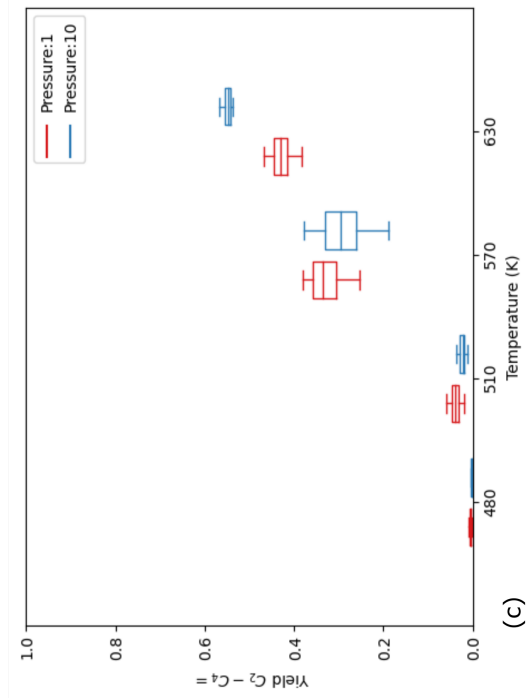
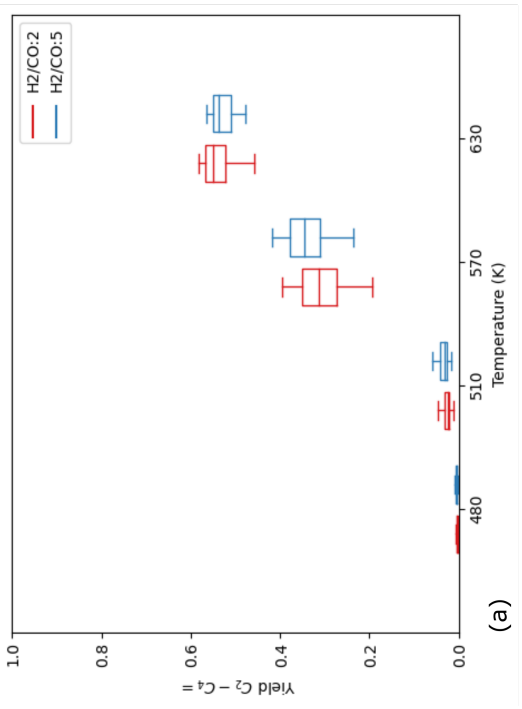
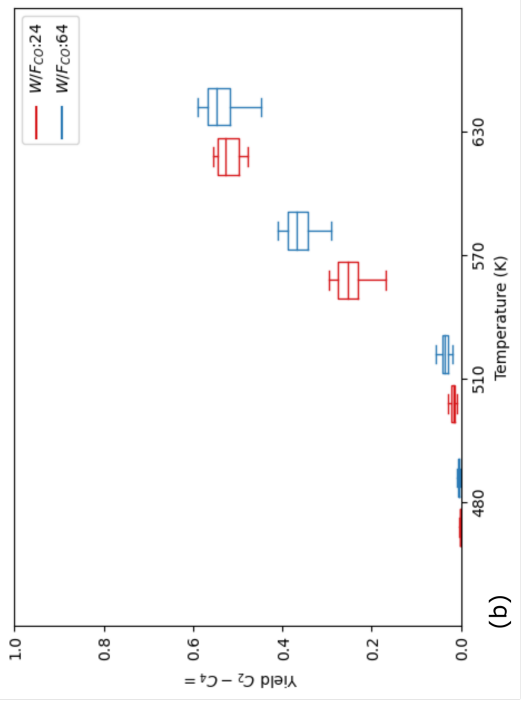


Figure 6: Yield as a function of temperature and syngas ratio (a), temperature and space time (b), temperature and pressure (c) for the entire range of operating conditions in the dataset.

at different space-time (W/F_{CO} - 24 and 64) is shown in Fig.6(b). It is observed that the yield changes with temperature. The influence of space-time is only observed close to a temperature of 570 K, with higher mean yield at a higher space-time. The change in yield of light olefins with temperature and pressure is shown in Fig.6(c). As in the previous cases, the yield increases with an increase in temperature. There is a considerable shift in the mean of box plots (due to influence of change in pressure) at the highest temperature (630K). This clearly indicates that temperature, followed by pressure has an important role in the yield of light olefins and the syngas ratio and space-time have minimal impact.

4.2. Parity plot comparison of the ML models

From the analysis above on SEMK benchmark dataset, the most important input variables and their impact on conversion and selectivity are obtained. This information is compared against the results from different multi-output ML models to find out the extent to which these models can match the benchmark SEMK results. First the output parity diagrams from different ML models, for both train and test data are assessed. The parity diagrams of the conversion and light olefin selectivity behavior for the four machine learning models are presented in Fig.7-8. The SEMK generated dataset is used for training the ML models. The training dataset is indicated by the green colored dots in the parity diagrams whereas the blue dots corresponds to the testing or prediction dataset. The parity diagram with reference to the training dataset show how adequate the model is fitted to the data. The parity diagram for the testing dataset show how well the model can predict for a new dataset.

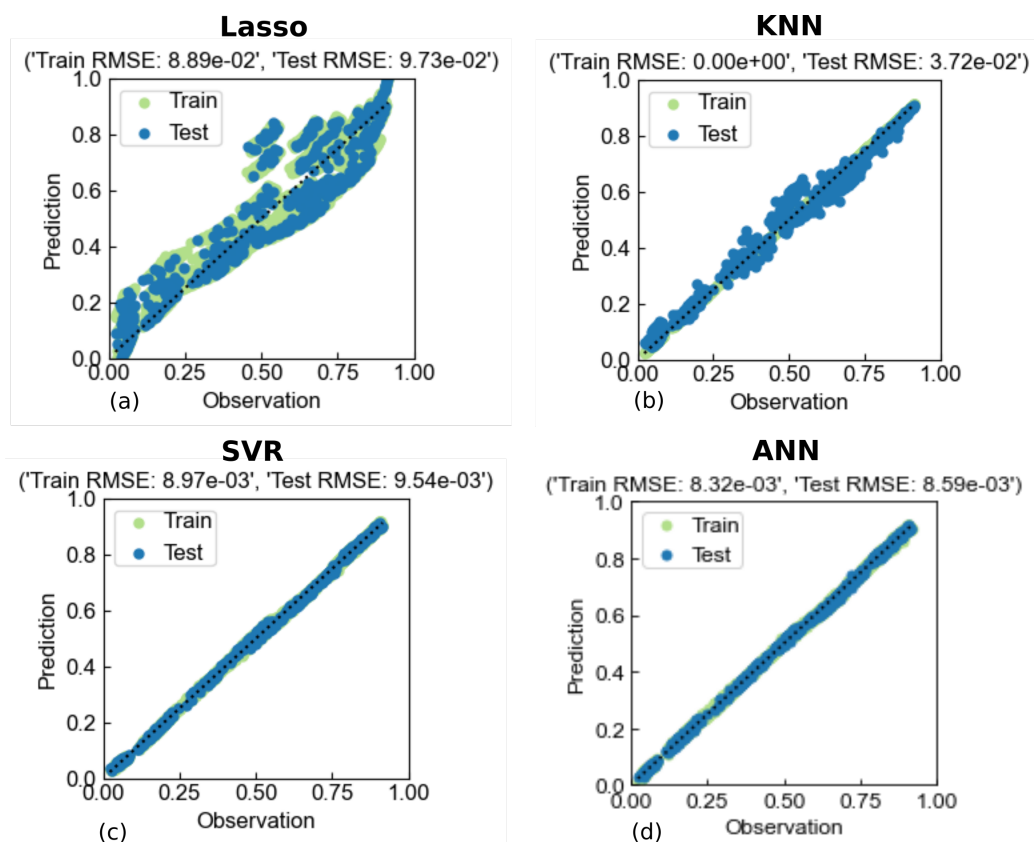


Figure 7: Parity plot of conversion for the machine learning models: Lasso regression (a), KNN regression (b), Support vector regression (c), and ANN regression (d).

It is observed from conversion behavior in Fig.7(a) that the Lasso regression model exhibits a significant spread in the dataset with respect to the parity line and the resulting root mean squared (RMSE) values are higher compared to other ML models (Table 1). The R-square values of all the models (both for training and testing dataset) are in the range of 0.87-0.99. From Fig.7(b-d), it is observed that the KNN, SVR, and ANN better fit the data than the Lasso regression.

It is observed from Fig.8(a-d), that the light olefin selectivity prediction

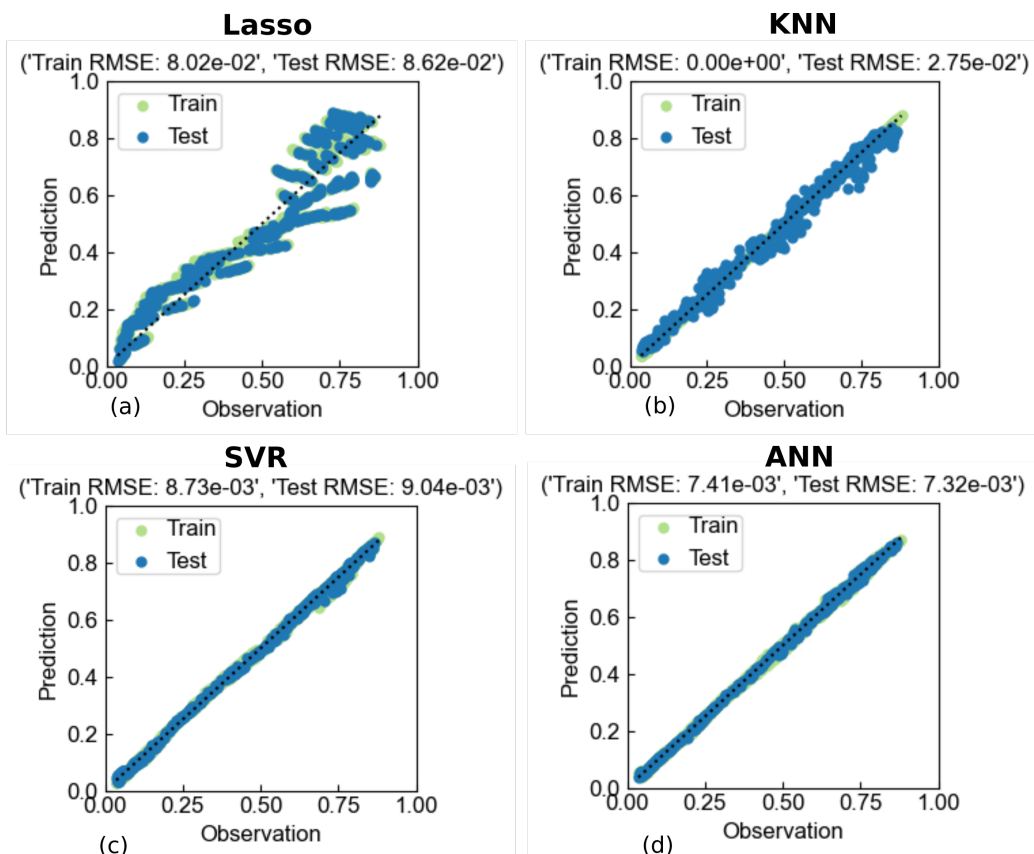


Figure 8: Parity plot of light olefin selectivity obtained for the machine learning models: Lasso regression (a), KNN regression (b), Support vector regression (c), and ANN regression (d).

matches with the parity lines for KNN, SVR, and ANN than for the Lasso regression model. The test root mean squared (RMSE) values of light olefin selectivity of SVR and ANN is slightly lower than that of KNN ³ (Table

³KNN memorizes the data (no function is fitted to the data) during training, and uses this data to make predictions for a new data point depending on its neighbours. Thus, a zero train RMSE is obtained with KNN.

Table 1: Root mean squared error of conversion and light olefin selectivity obtained with different ML models.

	RMSE, X_{CO}		RMSE, $S_{C_2-C_4=}$	
	Train	Test	Train	Test
Lasso	8.89×10^{-2}	9.73×10^{-2}	8.02×10^{-2}	8.62×10^{-2}
KNN	0	3.72×10^{-2}	0	2.75×10^{-2}
SVR	8.97×10^{-3}	9.54×10^{-3}	8.73×10^{-3}	9.04×10^{-3}
ANN	8.32×10^{-3}	8.59×10^{-3}	7.41×10^{-3}	7.32×10^{-3}

1). Further comparison of parity plots of methane and paraffin selectivity, obtained with these ML models are carried out to obtain the best candidate among the four models⁴. The performance of these models are further carried out with the help of contour plot analysis discussed in the section below.

4.3. Analysis of conversion using machine learning models

The feature analysis on the data reveals that the most important input variables impacting the light olefin selective FTS catalyst are temperature and pressure. Thus, a more detailed analysis of variation in conversion with the input variables, temperature and pressure is carried out on the testing dataset. The CO conversion obtained at different temperatures and pressure using the microkinetic model (Fig.9(a)) is compared with the predictions using different machine learning methods Fig.9(b-e) at the same operating

⁴The parity diagrams for these selectivity components are provided in the supplementary material

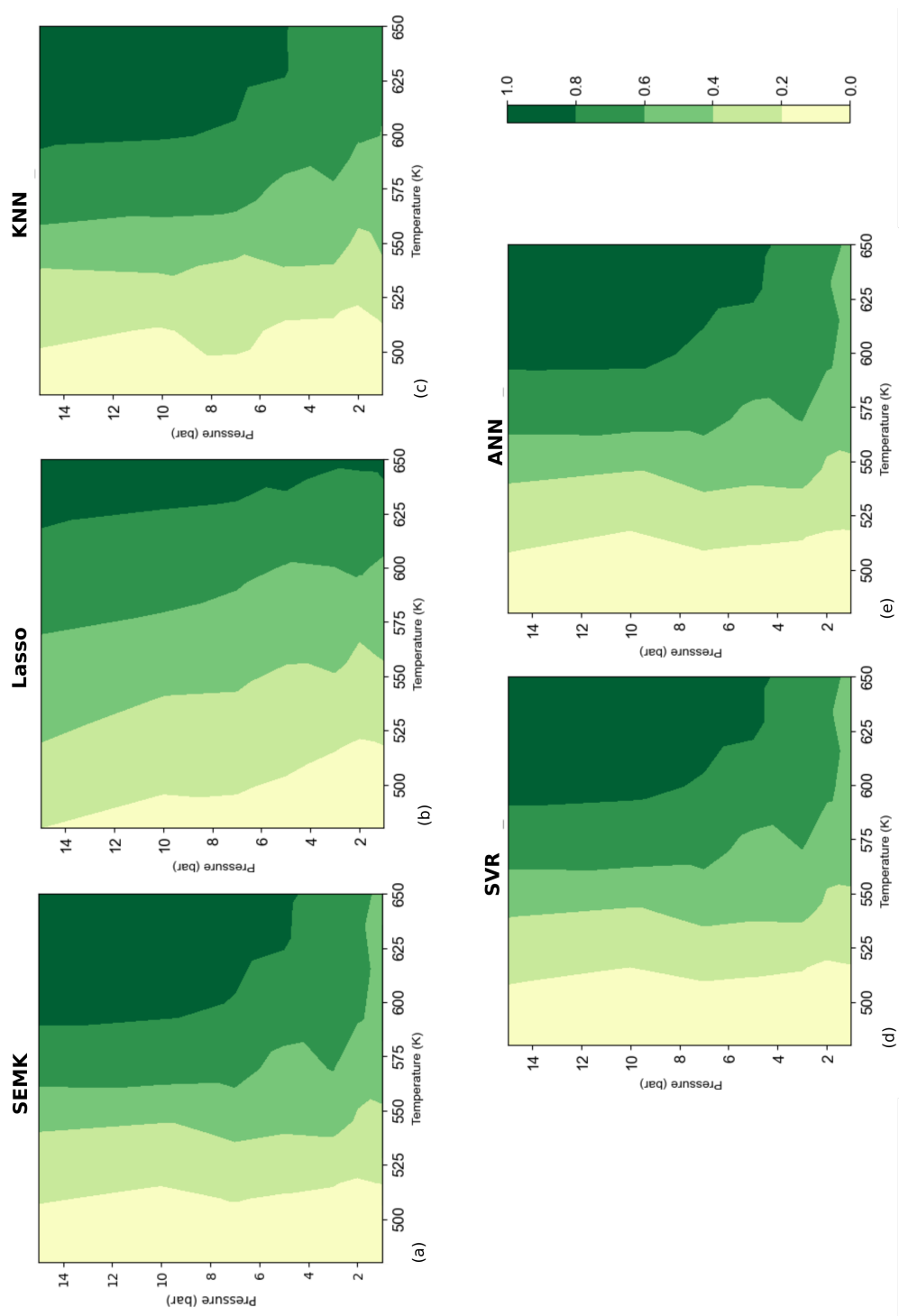


Figure 9: Contours of conversion obtained with SEMK (a) and with the machine learning models: Lasso regression (b), KNN regression (c), Support vector regression (d), and ANN regression (e) for the entire range of operating conditions.

conditions.

From Fig.9(a), it could be observed that the actual variation of conversion with temperature and pressure is non-linear in nature, especially at higher temperatures. There is an increase in the conversion as we traverse horizontally, from the left to the right (increase in temperature) and vertically, from the bottom to the top (increase in pressure). The highest conversion zone (dark green, $X_{CO} > 0.8$) is obtained in the temperature range 600-650K and pressure range varying from 4 to 15 bar. From Fig.9(b), it could be observed that the prediction contours obtained with Lasso regression is linear in nature and the non-linear behavior for conversion (seen by the evolution of curvature of contours) as observed with the microkinetic model is not captured. The KNN model (Fig.9(c)) performs slightly better than Lasso regression, but the evolution of curvature of contours is less pronounced. Conversely, the non-linear trend is captured by SVR (Fig.9(d)), and ANN (Fig.9(e)), with ANN matching the SEMK results very closely.

4.4. Analysis of light olefin Selectivity using machine learning models

Similar to Fig.9, the light olefin selectivity at different temperatures and pressures obtained using the microkinetic model (Fig.10(a)) is compared to the predictions using different ML methods Fig.10(b-e) for the testing dataset. Similar to the conversion plots discussed in the previous section, Fig.10(a) (benchmark SEMK results), shows that the actual variation of light olefin selectivity with temperature and pressure is non-linear in nature. This is evident from the curvature of the selectivity contours. There is an increase in the selectivity values as we traverse horizontally from the left to the right (with the increase in temperature), and vertically from the top to

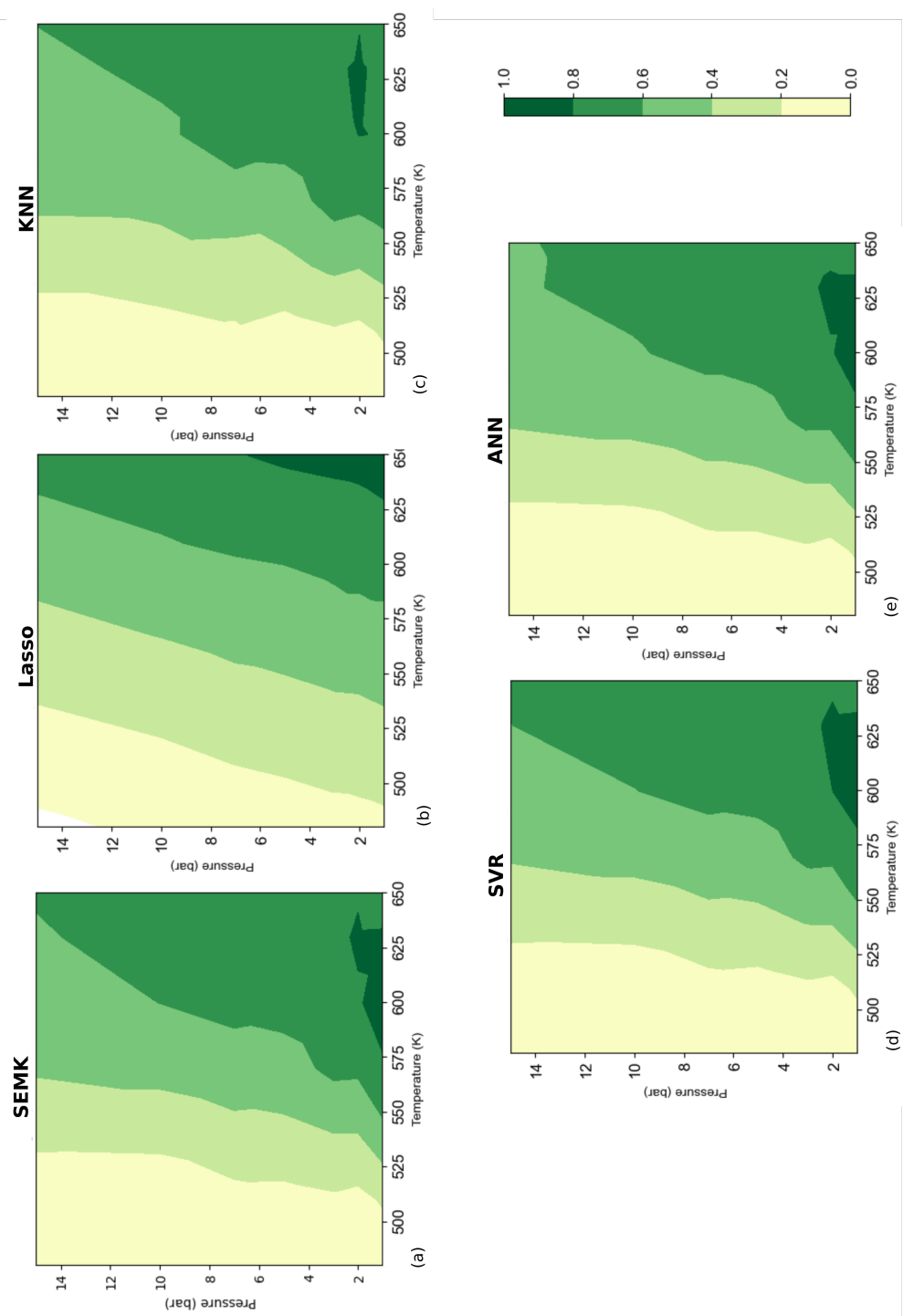


Figure 10: Contours of light olefin selectivity obtained with SEMK (a) and with the machine learning models: Lasso regression (b), KNN regression (c), Support vector regression (d), and ANN regression (e) for the entire range of operating conditions.

the bottom (with the decrease in pressure). Unlike in Fig.9, where conversion increases with pressure, here in Fig.10, the selectivity mostly decreases with an increase in pressure. The highest light olefin selectivity is obtained in a confined zone, in the temperature range 580-635K, and at a pressure below 2 bar.

In Fig.10(b), the light olefin selectivity contour prediction obtained by Lasso regression is linear in nature. The non-linearity as observed from the benchmark SEMK results is however, better captured by KNN (Fig.10(c)), SVR (Fig.10(d)), and ANN (Fig.10(e)) although significant differences between these models are noticeable. For instance, the zone of enhanced light olefin selectivity (temperature: 580-635K and pressure: 1-2 bar) is not captured by KNN, and the SVR and ANN therefore out-perform the Lasso and KNN models. The SVR model requires a work around using a wrapper function, as discussed in Section 2.2.3, to make multi-output predictions. It is therefore concluded that the ANN model is the most preferred model for the considered multi-response system.

4.5. Interpretability of the ANN model

As the model complexity increases from Lasso regression to ANN, the accuracy of prediction increases at the expense of interpretability. Consequently, there is a trade-off between the accuracy and interpretability of machine learning models. To better explain the prediction results of complex models, model agnostic interpretation techniques are useful which perform a systemic analysis and ranking of the important contributing input variables. The Shap values technique contributes to a large extent in the interpretability of the model at a global and local level. Here the global and local in-

interpretability or importance of input variables on the FTS process aimed at the selective light olefin production are discussed for the best performing ML model, i.e ANN by using Shap values.

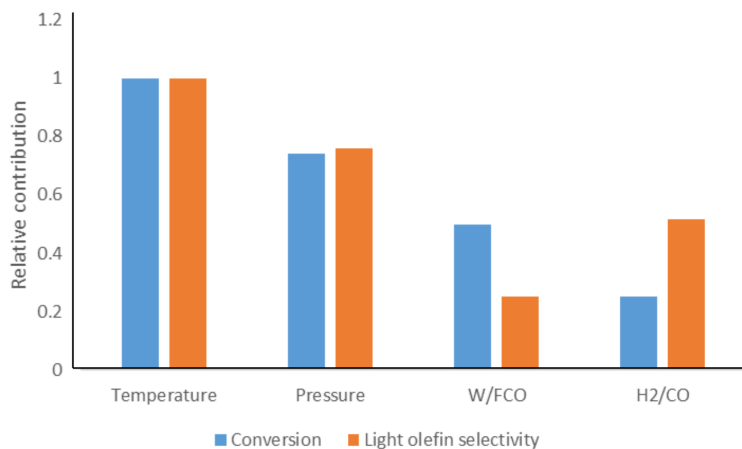
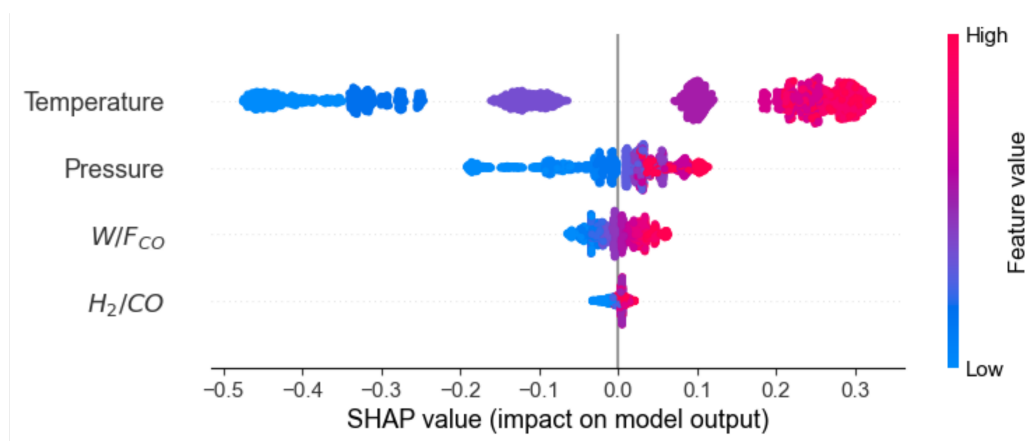


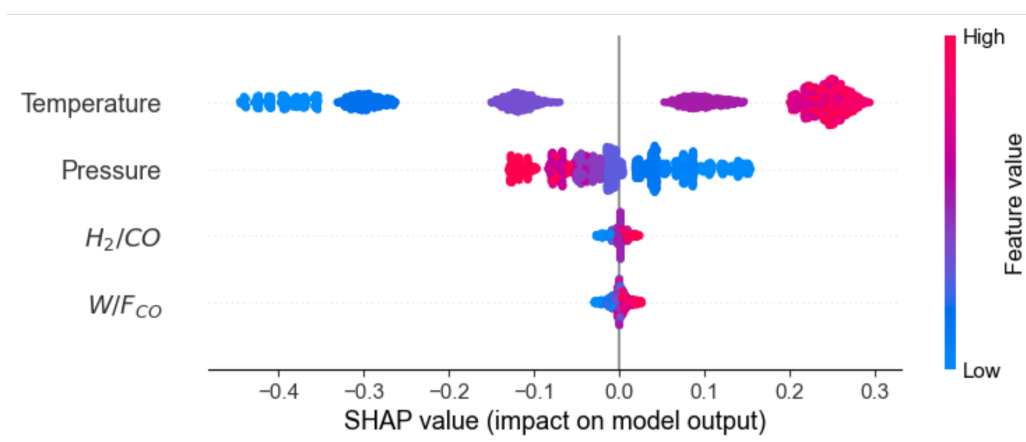
Figure 11: Shap bar plot indicating the global feature importance of the input variables on ANN prediction. The global feature importance is quantified in terms of contribution towards conversion and light olefin selectivity, from the entire training dataset, using the Python package, Shap [56].

From the Shap bar plot (Fig.11), the importance of each input variable on a global level could be observed. The input variables in Fig.11 are ordered according to their relative contribution towards the respective output. For conversion, the most important contributing input variable is the temperature, followed by pressure, space-time, and the syngas ratio. The relative contribution of the input variables towards conversion follow the order: temperature (1x) > pressure (0.7x) > space-time (0.5x) > syngas ratio (0.25x). For the selectivity towards light olefins, the temperature remains the most contributing input variable followed by pressure. However, with respect to conversion, the order of importance of space-time and syngas ratio has switched. In

the case of light olefin selectivity it follows the order: temperature (1x) > pressure (0.7x) > syngas ratio (0.5x) > space time (0.25x).



(a) Conversion



(b) Selectivity of light olefins

Figure 12: Shap tree plot indicating the local feature importance of the input variables on ANN prediction, conversion (a) and light olefin selectivity (b). The local feature importance is quantified, from the entire training dataset, using the Python package, Shap [56].

In addition to the information on the global relative importance of each

input variable towards conversion/selectivity from Fig.11, the relative importance at the local level is shown in Fig.12. The order of input variables in Fig.12, from the top to the bottom follow the order of their relative contribution towards conversion/selectivity, with the most contributing input variable at the top. The Shap values on the horizontal axis represent conversion(Fig.12(a)) and light olefin selectivity (Fig.12(b)). The average conversion/light olefin selectivity in the training dataset, indicated by a Shap value of zero (in Fig.12(a)-(b)), serves as the base value for the analysis. Each colored dot indicates the contribution of the input variable (shown at the left side of the figure) towards the conversion/selectivity. From Fig.12 it could be observed that the input variable with the highest importance has the widest range of Shap values(temperature), while the input variable with the lowest importance has a narrow range of Shap values (space time and syngas ratio). For all the input variables apart from pressure, it could be observed that a low value of the variable results in lower contribution towards conversion and selectivity compared to the average contribution as indicated by negative Shap values. With an increase in the value of each input variable (blue dots to red dots) apart from pressure, their contribution towards conversion/selectivity increases as indicated by the increase in Shap values. This shows that conversion and selectivity are positively correlated to temperature, space-time, and syngas ratio. However, the relative importance of pressure on conversion and selectivity differs. While the increase in pressure favors the conversion (increase in Shap value when pressure changes from blue to red dots), it has an adverse effect on light olefin selectivity. These observations are in line with the results reported in Section 4.1, 4.3 and 4.4.

5. Conclusions and perspectives

Multi-response machine learning models based on Lasso regression, k-nearest neighbors regression, support vector regression, and artificial neural network regression for the Fischer-Tropsch Synthesis were developed using a synthetic dataset generated with a Single Event MicroKinetic model for various operating conditions. The non-linearity in the kinetic data is not captured with the Lasso regression. As the model complexity increases from KNN to SVR and ANN, such non-linear prediction capabilities dramatically improve. In capturing the non-linearity in light olefin selectivity data, SVR and ANN were found to be superior to KNN. The capability of ANN to make multi-output predictions without carrying out separate regressions to predict each output as in SVR, makes it more preferred. The widely used interpretation technique based on Shap values on the (best performing model) ANN prediction allows to rank the considered input variables according to their impact on the process. For the ANN model, the relative impact of input variables, obtained using Shap values, on conversion follow the order of temperature (1x) > pressure (0.7x) > space-time (0.5x) > syngas ratio (0.25x). On the other hand, the influence of temperature (1x) and pressure (0.7x) remain the same for light olefin selectivity, but that of space-time (0.25x) and syngas ratio (0.5x) switches. The obtained results highly agree with what is expected from physico-chemical point of view.

From this work it could be observed that when selecting a machine learning model, the nature of intrinsic data, the number of outputs, and their interactions (such as non-linear behavior with respect to changing operating conditions) have to be taken into account. The methodology employed in

this work could serve as a guideline to identify a suitable, easy to implement ML model for a multi-response chemical process.

Declaration of interests

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.

Acknowledgments

This work has received funding from the European Regional Development Fund (ERDF) via the PSYCHE project (Interreg France-Wallonie-Vlaanderen) with co-financing from the provinces of East-Flanders and West-Flanders.

References

- [1] N. A. Jose, M. Kovalev, E. Bradford, A. M. Schweidtmann, H. C. Zeng, A. A. Lapkin, Pushing nanomaterials up to the kilogram scale – an accelerated approach for synthesizing antimicrobial ZnO with high shear reactors, machine learning and high-throughput analysis, *Chemical Engineering Journal* 426 (2021) 131345. [doi:10.1016/j.cej.2021.131345](https://doi.org/10.1016/j.cej.2021.131345).
- [2] R. Upadhyaya, S. Kosuri, M. Tamasi, T. A. Meyer, S. Atta, M. A. Webb, A. J. Gormley, Automation and data-driven design of polymer therapeutics, *Advanced Drug Delivery Reviews* 171 (2021) 1–28. [doi:10.1016/j.addr.2020.11.009](https://doi.org/10.1016/j.addr.2020.11.009).

- [3] L. C. Xue, D. Dobbs, A. M. Bonvin, V. Honavar, Computational prediction of protein interfaces: A review of data driven methods, *FEBS Letters* 589 (23) (2015) 3516–3526. doi:[10.1016/j.febslet.2015.10.003](https://doi.org/10.1016/j.febslet.2015.10.003).
- [4] A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne, A. A. Lapkin, Machine learning meets continuous flow chemistry: Automated optimization towards the pareto front of multiple objectives, *Chemical Engineering Journal* 352 (2018) 277–282. doi:[10.1016/j.cej.2018.07.031](https://doi.org/10.1016/j.cej.2018.07.031).
- [5] C. Casado, J. Moreno-SanSegundo, I. D. la Odra, B. E. García, J. A. S. Pérez, J. Marugán, Mechanistic modelling of wastewater disinfection by the photo-fenton process at circumneutral pH, *Chemical Engineering Journal* 403 (2021) 126335. doi:[10.1016/j.cej.2020.126335](https://doi.org/10.1016/j.cej.2020.126335).
- [6] C. I. Méndez, J. Ancheyta, Modeling and control of a Fischer-Tropsch synthesis fixed-bed reactor with a novel mechanistic kinetic approach, *Chemical Engineering Journal* 390 (2020) 124489. doi:[10.1016/j.cej.2020.124489](https://doi.org/10.1016/j.cej.2020.124489).
- [7] P. J. Donaubaauer, D. M. Melzer, K. Wanninger, G. Mestl, M. Sanchez-Sanchez, J. A. Lercher, O. Hinrichsen, Intrinsic kinetic model for oxidative dehydrogenation of ethane over MoVTaNb mixed metal oxides: A mechanistic approach, *Chemical Engineering Journal* 383 (2020) 123195. doi:[10.1016/j.cej.2019.123195](https://doi.org/10.1016/j.cej.2019.123195).
- [8] M. R. Dobbelaere, P. P. Plehiers, R. V. de Vijver, C. V. Stevens,

- K. M. V. Geem, Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats, *Engineering* 7 (9) (2021) 1201–1211. [doi:10.1016/j.eng.2021.03.019](https://doi.org/10.1016/j.eng.2021.03.019).
- [9] D. Greene, P. Cunningham, R. Mayer, Unsupervised learning and clustering, in: *Machine Learning Techniques for Multimedia*, Springer Berlin Heidelberg, pp. 51–90. [doi:10.1007/978-3-540-75171-7_3](https://doi.org/10.1007/978-3-540-75171-7_3).
- [10] M. Dash, H. Liu, J. Yao, Dimensionality reduction of unsupervised data, in: *Proceedings ninth IEEE international conference on tools with artificial intelligence*, IEEE, 1997, pp. 532–539.
- [11] M. Castelli, L. Vanneschi, Á. R. Largo, Supervised learning: Classification, in: *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 342–349. [doi:10.1016/b978-0-12-809633-8.20332-4](https://doi.org/10.1016/b978-0-12-809633-8.20332-4).
- [12] X. Zheng, Z. Jiang, Z. Ying, J. Song, W. Chen, B. Wang, Role of feedstock properties and hydrothermal carbonization conditions on fuel properties of sewage sludge-derived hydrochar using multiple linear regression technique, *Fuel* 271 (2020) 117609. [doi:10.1016/j.fuel.2020.117609](https://doi.org/10.1016/j.fuel.2020.117609).
- [13] X. Zhao, X. Li, H. Lu, H. Yue, C. Liu, S. Zhong, K. Ma, S. Tang, B. Liang, Predicting phase-splitting behaviors of an amine-organic solvent–water system for CO₂ absorption: A new model developed by density functional theory and statistical and experimental methods, *Chemistry*

- cal Engineering Journal 422 (2021) 130389. [doi:10.1016/j.cej.2021.130389](https://doi.org/10.1016/j.cej.2021.130389).
- [14] Q. Dai, L. Xie, Z. Guo, L. Ma, J. Yang, N. Ren, G. Tian, P. Ning, Research on the mechanism of synergistic-dehydration/detoxification for sludge under treatment with double-modified potato residue, Chemical Engineering Journal 420 (2021) 127699. [doi:10.1016/j.cej.2020.127699](https://doi.org/10.1016/j.cej.2020.127699).
- [15] S. Bansal, S. Roy, F. Larachi, Support vector regression models for trickle bed reactors, Chemical Engineering Journal 207-208 (2012) 822–831. [doi:10.1016/j.cej.2012.07.081](https://doi.org/10.1016/j.cej.2012.07.081).
- [16] D. S. Palmer, N. M. O'Boyle, R. C. Glen, J. B. O. Mitchell, Random forest models to predict aqueous solubility, Journal of Chemical Information and Modeling 47 (1) (2006) 150–158. [doi:10.1021/ci060164k](https://doi.org/10.1021/ci060164k).
- [17] N. S. Kaveh, F. Mohammadi, S. Ashrafizadeh, Prediction of cell voltage and current efficiency in a lab scale chlor-alkali membrane cell based on support vector machines, Chemical Engineering Journal 147 (2-3) (2009) 161–172. [doi:10.1016/j.cej.2008.06.030](https://doi.org/10.1016/j.cej.2008.06.030).
- [18] F. Kartal, U. Özveren, A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus, Energy [doi:10.1016/j.energy.2020.118457](https://doi.org/10.1016/j.energy.2020.118457).
- [19] K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, K. Hirai, J. Ohyama, T. N. Nguyen, S. Nishimura, T. Taniike, The rise of catalyst informatics: Towards


- catalyst genomics, *ChemCatChem* 11 (4) (2019) 1146–1152. doi:10.1002/cctc.201801956.
- [20] M. E. Morocho-Cayamcela, H. Lee, W. Lim, Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions, *IEEE Access* 7 (2019) 137184–137206. doi:10.1109/access.2019.2942390.
- [21] C. Molnar, *Interpretable machine learning : a guide for making Black Box Models interpretable*, Lulu, Morisville, North Carolina, 2019.
- [22] P. P. Plehiers, S. H. Symoens, I. Amghizar, G. B. Marin, C. V. Stevens, K. M. V. Geem, Artificial intelligence in steam cracking modeling: A deep learning algorithm for detailed effluent prediction, *Engineering* 5 (6) (2019) 1027–1040. doi:10.1016/j.eng.2019.02.013.
- [23] S. Mittal, S. Pathak, H. Dhawan, S. Upadhyayula, A machine learning approach to improve ignition properties of high-ash indian coals by solvent extraction and coal blending, *Chemical Engineering Journal* 413 (2021) 127385. doi:10.1016/j.cej.2020.127385.
- [24] L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike, K. Takahashi, Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts, *Chemical Science* 12 (38) (2021) 12546–12555. doi:10.1039/d1sc04390k.
- [25] J. M. Garcia, M. L. Robertson, The future of plastics recycling, *Science* 358 (6365) (2017) 870–872. doi:10.1126/science.aaq0324.

- [26] M. E. Dry, The fischer-tropsch process: 1950–2000, *Catalysis Today* 71 (3-4) (2002) 227–241. doi:[10.1016/S0920-5861\(01\)00453-9](https://doi.org/10.1016/S0920-5861(01)00453-9).
- [27] M. Dry, Catalytic aspects of industrial fischer-tropsch synthesis, *Journal of Molecular Catalysis* 17 (2-3) (1982) 133–144. doi:[10.1016/0304-5102\(82\)85025-6](https://doi.org/10.1016/0304-5102(82)85025-6).
- [28] Q. Zhang, J. Kang, Y. Wang, Development of novel catalysts for fischer-tropsch synthesis: Tuning the product selectivity, *ChemCatChem* 2 (9) (2010) 1030–1058. doi:[10.1002/cctc.201000071](https://doi.org/10.1002/cctc.201000071).
- [29] E. Iglesia, S. C. Reyes, R. J. Madon, S. L. Soled, Selectivity control and catalyst design in the fischer-tropsch synthesis: Sites, pellets, and reactors, in: *Advances in Catalysis*, Elsevier, 1993, pp. 221–302. doi:[10.1016/S0360-0564\(08\)60579-9](https://doi.org/10.1016/S0360-0564(08)60579-9).
- [30] B. Gu, V. V. Ordonsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, A. Y. Khodakov, Effects of the promotion with bismuth and lead on direct synthesis of light olefins from syngas over carbon nanotube supported iron catalysts, *Applied Catalysis B: Environmental* 234 (2018) 153–166. doi:[10.1016/j.apcatb.2018.04.025](https://doi.org/10.1016/j.apcatb.2018.04.025).
- [31] A. J. Barrios, B. Gu, Y. Luo, D. V. Peron, P. A. Chernavskii, M. Virginie, R. Wojcieszak, J. W. Thybaut, V. V. Ordonsky, A. Y. Khodakov, Identification of efficient promoters and selectivity trends in high temperature fischer-tropsch synthesis over supported iron catalysts, *Applied Catalysis B: Environmental* 273 (2020) 119028. doi:[10.1016/j.apcatb.2020.119028](https://doi.org/10.1016/j.apcatb.2020.119028).

- [32] A. Aguirre, E. Scholman, J. van der Shaaf, M. F. N. d'Angelo, Controlling the selectivity in the fischer-tropsch synthesis using foam catalysts: An integrated experimental and modeling approach, *Chemical Engineering Journal* 409 (2021) 128139. [doi:10.1016/j.cej.2020.128139](https://doi.org/10.1016/j.cej.2020.128139).
- [33] K. Toch, J. Thybaut, G. Marin, Ethene oligomerization on ni-SiO₂-Al₂O₃: Experimental investigation and single-event MicroKinetic modeling, *Applied Catalysis A: General* 489 (2015) 292–304. [doi:10.1016/j.apcata.2014.10.036](https://doi.org/10.1016/j.apcata.2014.10.036).
- [34] T. von Aretin, O. Hinrichsen, Single-event kinetic model for cracking and isomerization of 1-hexene on ZSM-5, *Industrial & Engineering Chemistry Research* 53 (50) (2014) 19460–19470. [doi:10.1021/ie503628p](https://doi.org/10.1021/ie503628p).
- [35] J. M. Martinis, G. F. Froment, Alkylation on solid acids. part 2. single-event kinetic modeling, *Industrial & Engineering Chemistry Research* 45 (3) (2006) 954–967. [doi:10.1021/ie050910v](https://doi.org/10.1021/ie050910v).
- [36] G. Lozano-Blanco, K. Surla, J. Thybaut, G. Marin, Extension of the single-event methodology to metal catalysis: Application to fischer-tropsch synthesis, *Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles* 66 (3) (2010) 423–435. [doi:10.2516/ogst/2009075](https://doi.org/10.2516/ogst/2009075).
- [37] J. V. Belleghem, C. Ledesma, J. Yang, K. Toch, D. Chen, J. W. Thybaut, G. B. Marin, A Single-Event MicroKinetic model for the cobalt catalyzed Fischer-Tropsch Synthesis, *Applied Catalysis A: General* 524 (2016) 149–162. [doi:10.1016/j.apcata.2016.06.028](https://doi.org/10.1016/j.apcata.2016.06.028).

- [38] A. Chakkingal, L. Pirro, A. C. da Cruz, A. J. Barrios, M. Virginie, A. Y. Khodakov, J. W. Thybaut, Unravelling the influence of catalyst properties on light olefin production via fischer–tropsch synthesis: A descriptor space investigation using single-event MicroKinetics, *Chemical Engineering Journal* 419 (2021) 129633. doi:10.1016/j.cej.2021.129633.
- [39] M. Esfandyari, M. Amiri, M. K. Salooki, Neural network prediction of the fischer-tropsch synthesis of natural gas with co (III)/al₂o₃ catalyst, *Chemical Engineering Research Bulletin* 17 (1). doi:10.3329/cerb.v17i1.22915.
- [40] A. Chakkingal, P. Janssens, J. Poissonnier, A. J. Barrios, M. Virginie, A. Y. Khodakov, J. W. Thybaut, Machine learning based interpretation of microkinetic data: a fischer–tropsch synthesis case study, *React. Chem. Eng.* 7 (2022) 101–110. doi:10.1039/D1RE00351H.
- [41] A. Mirzaei, S. Golestan, S.-M. Barakati, Prediction of fe-co-mn/mgo catalytic activity in fischer-tropsch synthesis using nu-support vector regression, *Physical Chemistry Research* 4 (3). doi:10.22036/pcr.2016.14776.
- [42] M.-C. Kang, D.-Y. Yoo, R. Gupta, Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete, *Construction and Building Materials* 266 (2021) 121117. doi:10.1016/j.conbuildmat.2020.121117.
- [43] H. A. Garona, F. M. Cavalcanti, T. F. de Abreu, M. Schmal, R. M. Alves, Evaluation of fischer-tropsch synthesis to light olefins over co-

- and fe-based catalysts using artificial neural network, *Journal of Cleaner Production* 321 (2021) 129003. [doi:10.1016/j.jclepro.2021.129003](https://doi.org/10.1016/j.jclepro.2021.129003).
- [44] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1) (1996) 267–288. [doi:10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x).
- [45] Y. Song, J. Liang, J. Lu, X. Zhao, An efficient instance selection algorithm for k nearest neighbor regression, *Neurocomputing* 251 (2017) 26–34. [doi:10.1016/j.neucom.2017.04.018](https://doi.org/10.1016/j.neucom.2017.04.018).
- [46] G. Smits, E. Jordaan, Improved SVM regression using mixtures of kernels, in: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, IEEE. [doi:10.1109/ijcnn.2002.1007589](https://doi.org/10.1109/ijcnn.2002.1007589).
- [47] W. S. Sarle, *Neural networks and statistical models* (1994).
- [48] M. Mishra, M. Srivastava, A view of artificial neural network, in: *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, IEEE, 2014. [doi:10.1109/icaetr.2014.7012785](https://doi.org/10.1109/icaetr.2014.7012785).
- [49] L. S. Shapley, Stochastic Games, *Proceedings of the National Academy of Sciences* 39 (10) (1953) 1095–1100. [doi:10.1073/pnas.39.10.1095](https://doi.org/10.1073/pnas.39.10.1095).
- [50] S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems 2017-December (Section 2)* (2017) 4766–4775. [arXiv:1705.07874](https://arxiv.org/abs/1705.07874).

- [51] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, G. B. Marin, Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts, *Industrial & Engineering Chemistry Research* 47 (16) (2008) 5879–5891. [doi:10.1021/ie071587u](https://doi.org/10.1021/ie071587u).
- [52] B. Boehmke, B. M. Greenwell, *Hands-on machine learning with R*, CRC Press, 2019.
- [53] L. Breiman, Random forests, *Machine learning* 45 (1) (2001) 5–32.
- [54] S. Lundberg, S.-I. Lee, An unexpected unity among methods for interpreting model predictions, *arXiv (Nips)* (2016) 1–6. [arXiv:1611.07478](https://arxiv.org/abs/1611.07478).
- [55] J. Brownlee, *Machine learning mastery with python*, Machine Learning Mastery Pty Ltd 527 (2016) 100–120.
- [56] [Welcome to the shap documentation](https://shap.readthedocs.io/en/latest/index.html).
URL <https://shap.readthedocs.io/en/latest/index.html>