# Tailoring Fischer Tropsch synthesis product selectivities : insights from microkinetic modelling and machine learning

Anoop Chakkingal

# THESE

Présentée en vue

d'obtenir le grade de

# DOCTEUR

en

***Spécialité : Molécules et Matière Condensée et Génie Chimique***

Par


## Anoop Chakkingal


**DOCTORAT délivré conjointement par CENTRALE LILLE et l'UNIVERSITE de GAND**

Titre de la thèse :

## Réglage de la sélectivité de la synthèse Fischer-Tropsch : aperçu de la modélisation microcinétique et de l'apprentissage automatique

## Tailoring Fischer Tropsch Synthesis Product Selectivities: Insights from Microkinetic Modelling and Machine Learning

Soutenue le 23 septembre 2022 devant le jury d'examen :

**Président :** BRUNET Sylvette, Prof., Université de Poitiers
**Co-président :** TAERWE Luc, Em. Prof., Université de Gand
**Rapporteur :** FONGARLAND Pascal, Prof., Université Claude Bernard Lyon 1
**Rapporteur :** VISCONTI Carlo Giorgio, Prof., Politecnico di Milano
**Membre :** VERMEIRE Florence, Université de Gand
**Membre :** THUINET Ludovic, Prof., Centrale Lille Institut
**Co-encadrante :** VIRGINIE Mirella, Maître de conférence, Centrale Lille
**Co-directeur de thèse :** KHODAKOV Andrei, Directeur de Recherche, CNRS
**Directeur de thèse :** THYBAUT Joris, Prof., Université de Gand


Thèse préparée au laboratoire UCCS, UMR CNRS 8181, École doctorale SMRE 104 et au Laboratoire de Technologie Chimique, Technologiepark 125.


centralelille

GHENT UNIVERSITY

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# NOMENCLATURE

**Acronyms**

$\Delta\tilde{S}_r^0$ Single-event surface reaction entropy, J(Kmol)$^{-1}$

$\Delta$ The difference between experimental and simulation results in percentage

$F_{CO,in}$, $F_{CO}$ Carbon monoxide molar flow rate at the reactor inlet, mol/s

$F_{CO,out}$ Carbon monoxide molar flow rate at the reactor outlet, mol/s

$n_{C,i}$ The number of carbon atoms in component i

$S_i$ The selectivity toward a gas-phase component i

$$S_i = \frac{n_{C,i}F_i}{F_{CO,in} - F_{CO,out}}$$

$X_{CO}$ Carbon monoxide conversion at the reactor outlet

$$X_{CO} = \frac{F_{CO,in} - F_{CO,out}}{F_{CO,in}}$$

$E_a^{for}$ Forward activation energy, kJ/mol

$A_i$ Affinity for elementary reaction $i$, Jmol$^{-1}$

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

$K_{eq}$ Thermodynamic equilibrium coefficient, [−]

$r_p$ Pearson correlation coefficient

$W/F_{CO}$ Space time, $(kg_{cat}s)mol_{CO}^{-1}$

$Y_{CH_4}$ methane yield

$\tilde{A}_j$ Pre-exponential factor, 1/s or 1/MPa s

| | |
|---|---|
| $i$ | Feature of interest |
| ANN | Artificial Neural Network |
| CNT | Carbon nanotube |
| F | All input features except feature $i$ of interest |
| FFF | Fast Flexible Filling design of experiments |
| FTS | Fischer-Tropsch Synthesis |
| GHSV | Gas hourly space velocity, $Lg^{-1}h^{-1}$ |
| KNN | K-Nearest Neighbor |
| ML | Machine Learning |
| MSE | Mean squared error, i.e. squared difference between actual and predicted output |
| n | Number of virtual catalysts |
| P | Pressure, bar |
| R | Universal gas constant, $JK^{-1}mol^{-1}$ |
| RMSE | Root mean square error |
| SEMK | Single-Event MicroKinetic modelling |
| SS | Subset of input features |
| SVR | Support Vector Regression |
| T | Temperature, K |
| W | Catalyst mass, kg |

# SUMMARY

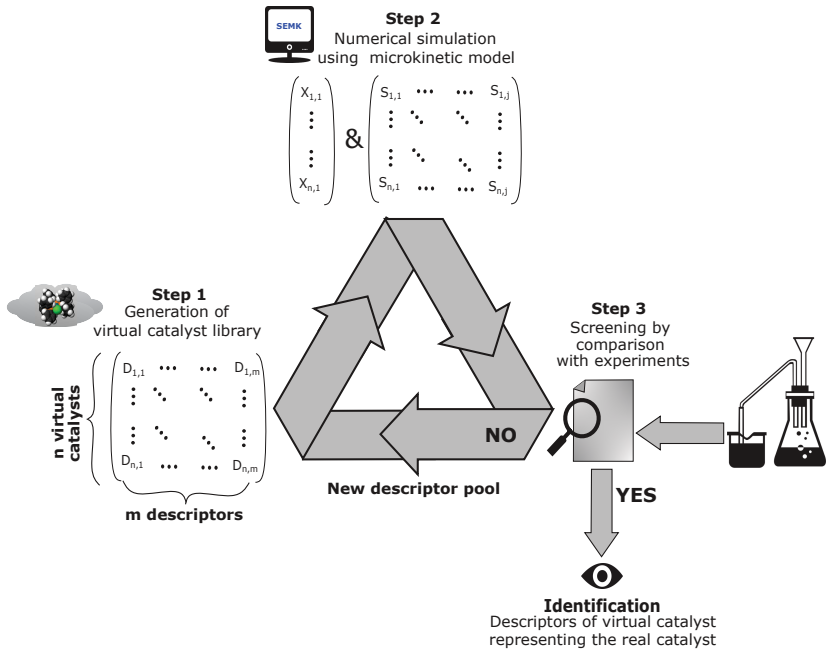
The increased level of plastic usage and disposal rate over the last decades, has recently turned into a major concern. Due to their non-(bio)degradable properties, plastics tend to accumulate as debris in natural habitats worldwide. Their production accounts for around 4% of the usage of non-renewable resources such as oil and gas feedstock and 3 to 4% of the global energy consumption. It is also estimated that the production of plastics will double in the next 20 years. This demands addressing the management of plastic waste on an immediate basis. Among the plastic waste management strategies, landfill inherently induces a long-term contamination risk of soil and groundwater and, hence, is not sustainable. The mechanical recycling of plastic suffers from the inability to fully recover its potential. An important method to fully recover its value is by thermo-chemical conversion. Particularly in the last decade, there has been an increase in research in the area of chemical recycling of plastics.

Within the plastic chemical recycling scheme, Fischer-Tropsch synthesis (FTS) could play a key role as the syngas feedstock that is converted in it, can be generated via the gasification of the considered plastics. This syngas is then chemo-catalytically converted into hydrocarbons such as paraffins and light olefins. Typical FTS catalysts are based on supported cobalt or iron species. Many experimental and mechanistic studies have been reported in the literature to assess their selectivity towards methane, paraffins, olefins, etc. and to determine the key catalyst properties which steer that selectivity. Higher activity, light olefin selectivity, favorable to lower syngas ratio (due to water gas shift reactivity), and lower cost make iron-based catalysts a better candidate in Fischer-Tropsch

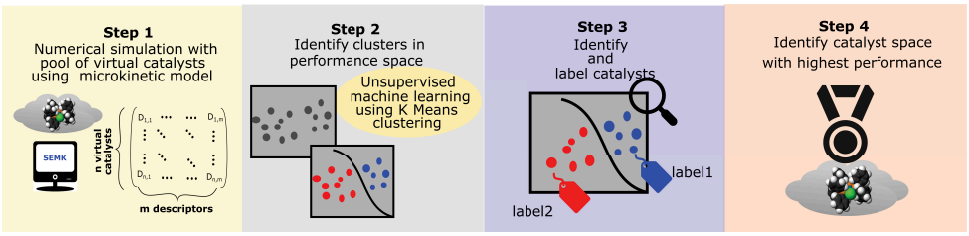
synthesis even though they also lead to proportionally more oxygenates. The improvement in catalyst performance with addition of promoters to iron-based catalysts has attracted recent research interest. The study of the promoter effect is rather complex as it is done with different supports and at different reaction conditions. In order to provide fundamental insights on such catalyst developments, extensive computational studies using different mechanistic models are being carried out.

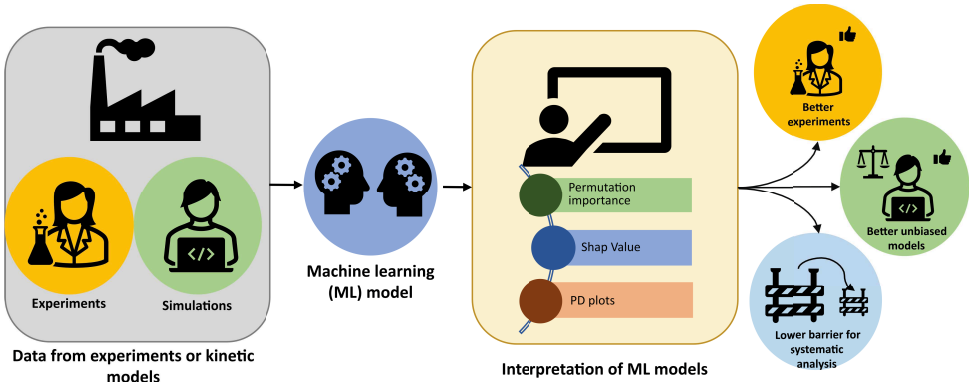
Among the mechanistic kinetic models, the comprehensive variant based on the Single Event MicroKinetics (SEMK) concept has been widely applied in the field of oligomerization, autoxidative curing, etc. and has proven to be a versatile tool to simulate Fischer-Tropsch synthesis. However, developing mechanistic models for every chemical engineering challenge 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 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 machine learning models which are easier to develop as compared to a detailed mechanistic model, which is also investigated in this thesis. The capabilities of both these approaches are used to identify an optimal catalyst with enhanced light olefin selectivity and corresponding suitable range of operating conditions.

In the first instance, this thesis deals with a virtual Single-Event MicroKinetic (SEMK) based screening of potential catalysts targeting enhanced light olefin formation via Fischer-Tropsch synthesis. The catalyst properties were quantified by means of catalyst descriptors such as the atomic chemisorption enthalpy of hydrogen, carbon, oxygen. The descriptor space is explored by simulating the experimental performances of promoted iron-based catalysts and searching for descriptors leading to a high light olefin selectivity. Single-Event MicroKinetic simulations using different combinations of the descriptors are carried to visually (human intervention) identify the catalyst properties of an optimal virtual catalyst with enhanced light olefin selectivity.



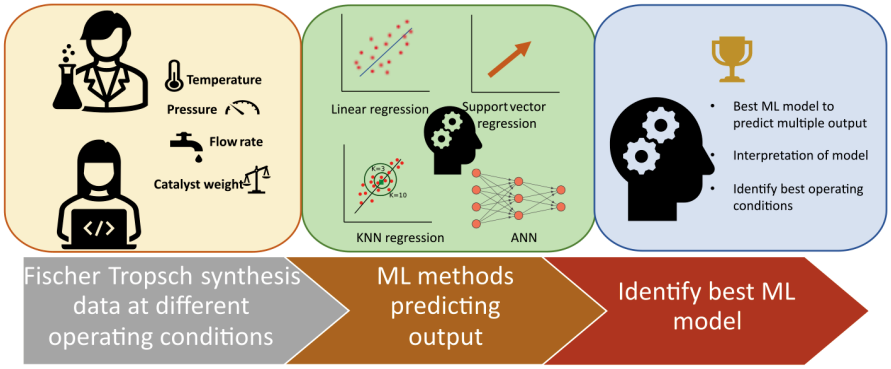

The methodology was subsequently generalized through an unsupervised machine learning approach which allows it to go beyond three dimensions with less human intervention. The descriptor space generated by microkinetic data (virtual catalysts) is explored using a systematic unsupervised clustering (machine learning) and labelling approach. The performance space is grouped into clusters and the minimum number of clusters is identified. Each virtual catalyst (represented by a combination of catalyst descriptors) is then identified in terms of its cluster. The range of catalyst descriptor values






within the cluster with the highest yield of light olefins is then obtained. The descriptor values are observed to be in line with that of the optimal virtual catalyst identified earlier using visual inspection. Thus, a coupled micro-kinetic-machine learning" approach for catalyst design and examining in detail the catalyst properties, reducing the human intervention, has been proposed.

Subsequently, a detailed evaluation on the potential of using machine learning approaches to match the performance of results obtained using the Single-Event MicroKinetic model was carried out. Initially, the focus was on a single dominant output scenario (methane selective catalyst). A machine learning model based on Artificial Neural Network for cobalt catalysed Fischer-Tropsch Synthesis with single dominant output, methane, is developed using a synthetic dataset generated via a Single-Event MicroKinetic model. The relative importance (or ranking) of the process variables (or predictors) in the decision-making process of the employed machine learning model were found using interpretation techniques such as permutation importance, the Shap value and partial dependence plots. The current work thus shows that more widely applied techniques in data science can now be applied for systematic analysis and interpretation of kinetic data. Similar analysis using experimental data can also help experimenters in their preliminary analysis, to detect hidden trends in the data, and thus to identify importance




features. After gaining confidence on the investigated interpretation techniques, for the FTS reaction with single dominant output, a similar investigation on the potential of iron based catalysts with enhanced light olefin selectivity is carried out next.

Finally, the machine learning approach is extended to a multi-output response prediction with primary focus on light olefin production at different operating conditions. 4 alternative machine learning models were employed in this respect, i.e., Lasso regression, k nearest neighbor (KNN) regression, support vector machine regression (SVR), and Artificial Neural Network (ANN) regression. The capabilities of these techniques in reproducing the (pronounced) non-linear behavior of conversion /selectivities as a function of process variables as simulated by the Single-Event MicroKinetic model is assessed. It is found that ANN predictions are best able to match the benchmark Single-Event MicroKinetic results. The best-performing Artificial Neural Network based model is then further analyzed using the Shap interpretation technique for increased explainability.

The overall study yielded key insights such as optimal catalyst descriptors: atomic chemisorption enthalpies of hydrogen ($Q_H \approx 234$kJ/mol), carbon ($Q_C \approx 622$kJ/mol) and oxygen, ($Q_O \approx 575$kJ/mol) for light olefin selective catalyst design using a mechanistic Single-Event MicroKinetic model. Further the range of operating conditions for en-

hanced light olefin selectivity with major process impacting variables being temperature (580-620K), and pressure (1-2 bar) were determined using different modeling strategies (Single-Event MicroKinetics and machine learning). The extent to which different machine learning models can match the Single-Event MicroKinetic results are then investigated. The non-linearity in the kinetic data is best captured with a neural network model with an rmse value of approximately $10^{-2}$. The performed work aimed at unravelling the opportunities of using machine learning models along with mechanistic models, to effectively reduce the computational and manual effort. For instance, for data obtained through high-throughput experimentation, a preliminary analysis could be carried out using a machine learning model and an in-depth chemical understanding could be subsequently obtained using a mechanistic model.

# SAMENVATTING

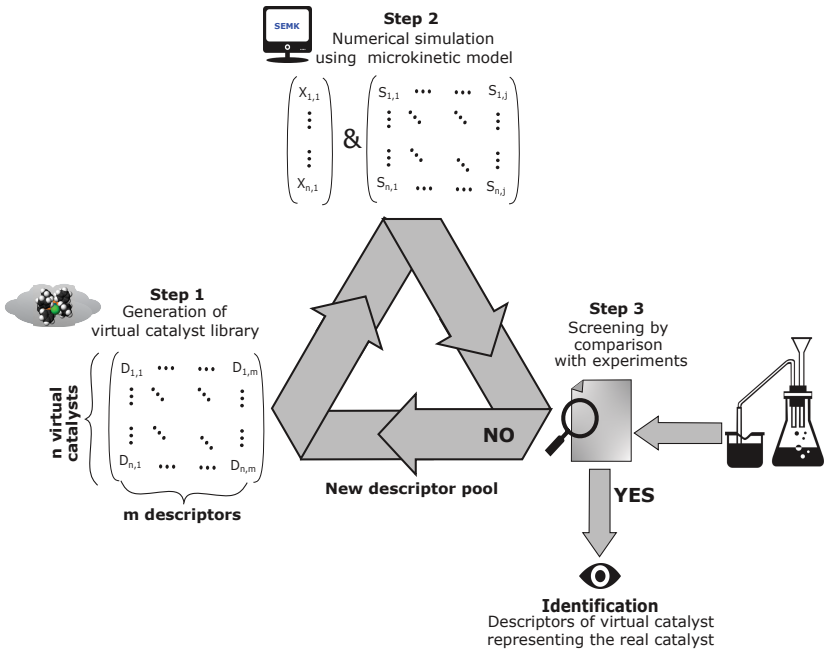
Het streven naar een circulaire economie heeft geleid tot het heronderzoeken van verschillende bestaande chemische processen, met als doel hun duurzaamheid te verhogen. Kunststoffen zijn alomtegenwoordig in onze samenleving en vormen een belangrijk onderdeel van de huidige economie. Door hun corrosieresistentie, laag soortelijk gewicht en interessante prijs kennen deze materialen hun toepassing in een brede waaier van applicaties wat geresulteerd heeft in een steeds toenemende productiecapaciteit tot 367 miljoen ton in 2020.

Het sterk toegenomen verbruik van plastics en in het bijzonder het daarbij horende afvalbeleid vormt een grote uitdaging; de wereldwijde accumulatie van kunststoffen in de natuur is een direct gevolg van hun niet-(bio)degradeerbare eigenschappen. Hun productie is goed voor zowel 4% van de niet-hernieuwbare grondstoffen (zoals olie en gas), als 3 tot 4% van het wereldwijd energieverbruik. Bovendien wordt geschat dat de productie van kunststoffen de komende 20 jaar zal verdubbelen. Dit vraagt om een onmiddellijke aanpak van het afvalbeheer waarin het storten zoveel mogelijk vermeden moet worden omwille van het inherente risico op verontreiniging van zowel bodem als grondwater. Het mechanisch recyclen van plastics is een belangrijke pijler in dit afvalbeheer. Echter, de technologie laat momenteel niet toe het volledige potensteel van de plastics te recupereren. Hiervoor is de thermochemische conversie van plastics een interessant alternatief, getuige de sterk toegenomen interesse vanuit de onderzoekswereld in dit domein.

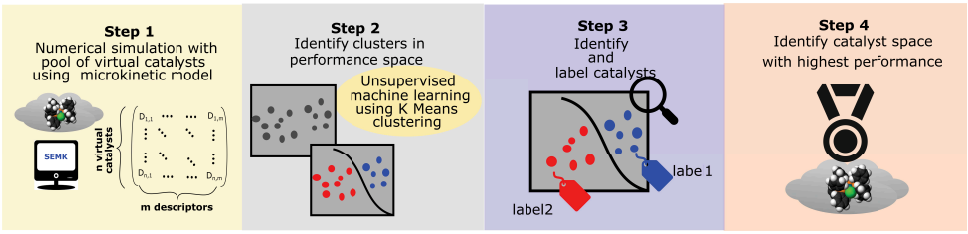
Binnen het gebied van de chemische recyclage kan de Fischer-Tropsch Synthese (FTS)

een sleutelrol spelen aangezien de grondstof voor het proces, i.e., syngas, verkregen kan worden uit de vergassing van plastics. Dit syngas wordt vervolgens chemo-katalytisch omgezet in koolwaterstoffen zoals paraffines en lichte olefinen met behulp van een ijzer of kobalt katalysator. Verschillende experimentele en mechanistische studies werden reeds gerapporteerd waarin onderzocht werd welke katalysatoreigenschappen een rol spelen in de selectiviteit naar methaan, paraffines, olefinen, etc. Naast hun lagere kostprijs en hun activiteit in de water-gas shift reactie vertonen ijzer-gebaseerde katalysatoren een hogere activiteit en selectiviteit naar lichte olefinen. Deze eigenschappen maken dit type katalysatoren uitermate geschikt voor de FTS reactie, ook al zal de verhoogde vorming van zuurstof bevattende species een belangrijke uitdaging zijn bij tijdens ontwikkeling. Hiervoor werd het toevoegen van promotoren recent onderzocht. Echter, het zogenaamde promotor-effect is een complex gegeven aangezien dit afhankelijk is van het dragermateriaal en de onderzochte reactiecondities. Om fundamenteel inzicht te verkrijgen in deze fenomenen worden uitvoerige computationele studies uitgevoerd, gebaseerd op verschillende mechanistische modellen.

Een uitgebreide variant van deze mechanistische modellen is het Single-Event MicroKinetisch (SEMK) model en dit model werd reeds uitvoerig toegepast in het domein van de oligomerisatie, de oxidatieve uitharding en de Fischer-Tropsch synthese. Het ontwikkelen van dit soort modellen voor iedere chemische reactie is echter niet werkbaar omwille van hun complexiteit en, daaruit volgend, de vereiste expertise om zo een model te bouwen. Dit geldt in het bijzonder voor katalytische reacties waarvoor eenvoudige machtswetten niet in staat om de onderliggende chemische fenomenen in detail te beschrijven. Dit creëert een opportuniteit voor de eenvoudiger te ontwikkelen machine learning (ML). Het potentieel van zowel de microkinetische als ML methodes wordt onderzocht in deze thesis met als doelstelling een optimale katalysator te identificeren voor een verhoogde lichte olefine selectiviteit met de daarbij horende werkcondities.

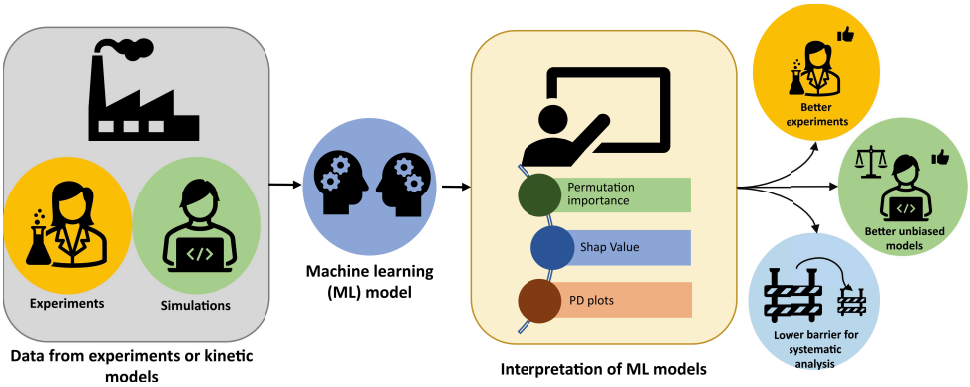



In deze thesis werd in eerste instantie een virtuele katalysator-screening uitgevoerd waarin gebruik werd gemaakt van een SEMK model met als doel een katalysator te identificeren met een verhoogde selectiviteit naar lichte olefinen in de Fischer-Tropsch Synthese. De daarbij horende katalysatoreigenschappen werden gekwantificeerd met behulp van katalysatordescriptoren zoals de atomaire chemisorptie-enthalpie van waterstof, zuurstof en koolstof. De descriptorruimte werd vervolgens verkend door experimentele performanties van gepromoveerde ijzer-gebaseerde katalysatoren te simuleren waaruit optimale descriptorwaarden werden geïdentificeerd. Vervolgens werd gebruik gemaakt van het SEMK model om verschillende descriptor combinaties (i.e., virtuele katalysatoren) te evalueren waaruit de meest geschikte katalysatoreigenschappen visueel werden gedetecteerd ("human intervention").




De methodologie werd vervolgens gegeneraliseerd via een niet-gesuperviseerde ML methode die toelaat om de driedimensionale beperkingen van de visuele detectiemethode te overstijgen. De descriptorruimte die gegenereerd werd met behulp van microkinetische data (virtuele katalysatoren) werd geanalyseerd met behulp van een systematische niet-gesuperviseerde clustering (ML) en labeling methode. De data werd gegroepeerd in clusters en het minimum aantal clusters werd geïdentificeerd. Elke virtuele katalysator (voorgesteld door een combinatie van descriptorwaarden) werd vervolgens geanalyseerd in termen van clusters. Het spanwijdte van de katalysator descriptorwaarden binnen de cluster met het hoogte lichte olefinen rendement wordt dan vervolgens geïdentificeerd. Deze descriptorwaarden waren in overeenstemming met de optimale virtuele katalysator zoals bepaald met de visuele methode. Het succes van deze gekoppelde microkinetische en ML aanpak voor katalysatorontwerp vermindert de manuale werklast aanzienlijk.

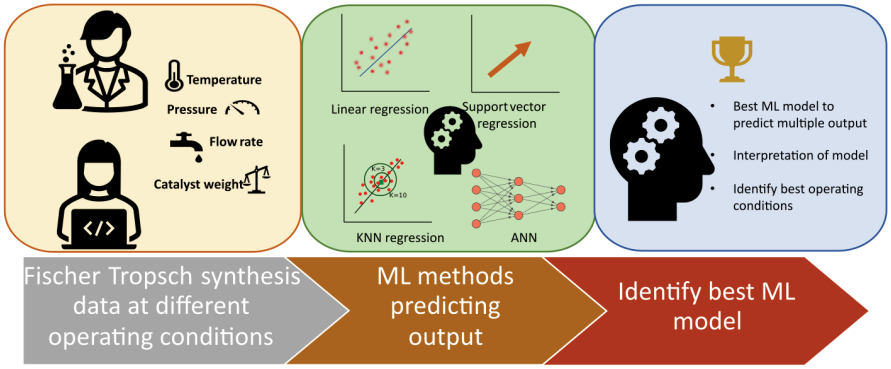
Vervolgens werd onderzocht of ML methoden in staat zijn om de resultaten van het microkinetisch model te evenaren en hierbij werd initieel gefocust op een enkelvoudig dominant output scenario (single dominant output scenario), namelijk de methaan selectiviteit. Een ML model gebaseerd op een artificieel neuraal netwerk (ANN) voor de kobalt-gekatalyseerde FTS met een enkelvoudige dominante output (methaan) werd ontwikkeld met behulp van een synthetische dataset die gegenereerd werd door het SEMK model. Het relatief belang (“ranking”) van de procesvariabelen (“predictors”) in het beslissingsproces van het gebruikte ML model werd onderzocht met behulp van




interpretatietechnieken zoals permutatie belang (“permuation importance”), de Shap waarde en partieel-afhankelijkheidsplots (“partial dependence plots”, PDP). Dit werk toont aan dat data science technieken nu gebruikt kunnen worden voor de systematische analyse en interpretatie van kinetische data. Bovendien kan een analoge (preliminaire) analyse op experimentele data onderzoekers helpen om verborgen trends en belangrijke invloeden te ontdekken in een proces. Op basis van het behaalde succes voor een enkelvoudige dominante output werd de methodologie verder onderzocht in de ijzer-gekatalyseerde FTS van lichte olefinen.

De ML methode werd finaal veralgemeend tot een multi-output respons systeem waarin de primaire focus gericht was op de synthese van lichte olefinen bij verschillende werkcondities. Voor vier ML modellen (i.e., Lasso regressie, k-nearest neighbour (KNN) regressie, support vector machine regressie (SVR) en een artificieel neuraal netwerk (ANN)) werd onderzocht in welke mate het uitgesprokene non-lineair gedrag van conversie en selectiviteit als functie van de procesvariabelen kon gereproduceerd worden, zoals weergegeven door het SEMK model. Hieruit bleek dat een ANN model hiervoor het meest geschikt is. Bijkomende informatie werd vervolgens uit het meest performante ANN model gehaald door middel van de Shap interpretatietechniek.

Met behulp van een SEMK model kon in deze thesis optimale katalysator- descriptor-




waarden voor de synthese van lichte olefinen bepaald worden: de atomaire chemisorptie-enthalpie van waterstof ($Q_H \approx 234$kJ/mol), koolstof ($Q_C \approx 622$kJ/mol) en zuurstof ($Q_O \approx 575$kJ/mol). Met behulp van zowel het SEMK model als de ML modellen kon ook het optimale temperatuursinterval (580-620K) en drukinterval (1-2 bar) bepaald worden. Er werd aangetoond dat de non-lineaire trends in de kinetische data het best beschreven konden worden met behulp van een neuraal netwerk met een root-mean-square-error (rmse) van $10^{-2}$. Het onderzoek dat verricht werd doelde op het gebruik van ML modellen in combinatie met mechanistische modellen om zowel de computationele als de manuele werklast te verminderen. Zo kan, bijvoorbeeld, een preliminaire analyse van high-throughput data uitgevoerd worden met een ML model om potentieel interessante zones van de parameterruimte te identificeren om dan vervolgens een gerichte en diepgaande analyse uit te voeren met een mechanistisch model.

# RÉSUMÉ

En vue de promouvoir l'économie circulaire, de nombreux procédés chimiques sont actuellement réexaminés afin de développer des variantes plus durables. Dans l'économie actuelle et dans la vie quotidienne, les matières plastiques sont des matières importantes et omniprésentes. Elles sont peu coûteuses, durables, résistantes à la corrosion et légères, ce qui les rend propices à l'emploi dans de nombreuses applications. Cela a mené à une forte augmentation de production au cours des 60 dernières années, entraînant une production totale de 367 millions de tonnes en 2020.

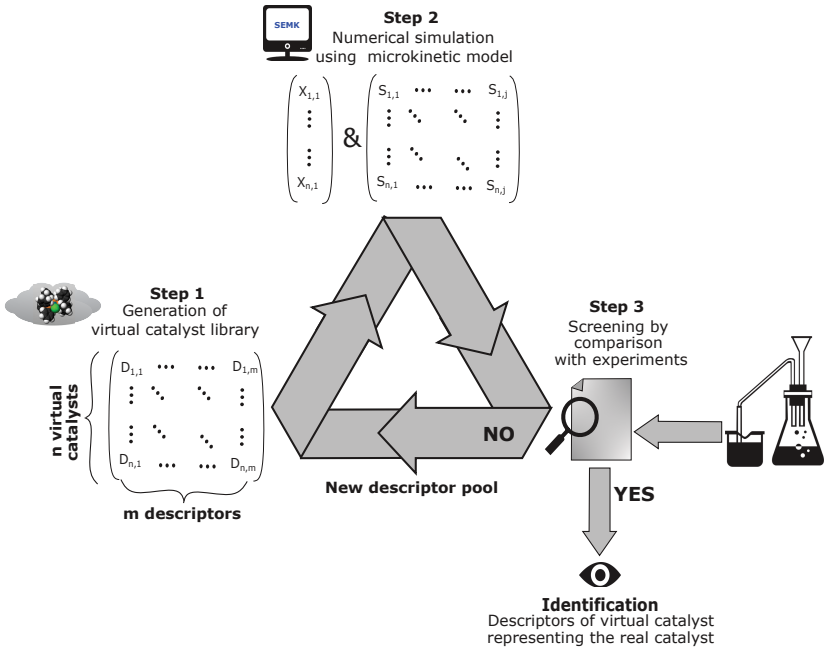
L'utilisation croissante des matières plastiques et la formation de déchets qui en provient sont devenues une source de préoccupation majeure. À cause de leurs propriétés non biodégradables, les matières plastiques ont tendance à s'accumuler dans les habitats naturels à l'échelle mondiale. Leur production compte pour environ 4% de l'emploi des ressources non renouvelables, comme le pétrole et le gaz naturel, et 3 à 4% de la consommation mondiale d'énergie. Il est également prévu que la production des matières plastiques va doubler au cours des 20 prochaines années. Cela exige de remédier rapidement au problème de la gestion des déchets plastiques. Parmi les stratégies de gestion de déchets plastiques, la décharge de déchets entraîne intrinsèquement un risque de contamination du sol et des eaux souterraines à long terme et n'est donc pas durable. D'autre part, le recyclage mécanique n'est pas capable de récupérer entièrement la valeur potentielle des déchets. Une méthode importante pour la récupération complète de la valeur est la conversion thermochimique. Notamment au cours de la dernière décennie, le nombre de recherches sur le recyclage chimique des matières plas-

tiques a fort augmenté.

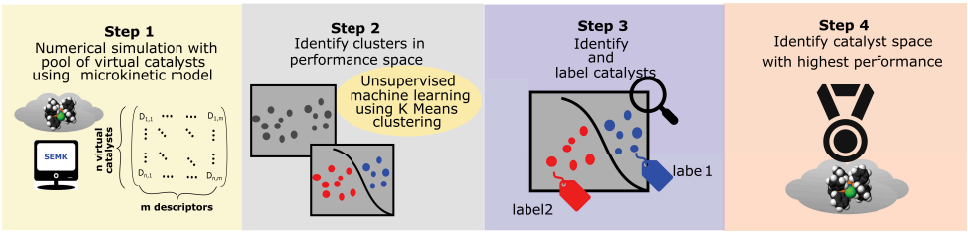
Dans le processus du recyclage chimique des matières plastiques, la synthèse de Fischer-Tropsch (SFT) pourrait jouer un rôle important, puisque le gaz de synthèse qui est converti dans le procédé peut être produit par la gazéification des matières plastiques considérées. Le gaz de synthèse est ensuite converti de façon chimiocatalytique en hydrocarbures comme les alcanes ou les alcènes légers. Les catalyseurs de SFT typiques sont composés de cobalt ou de fer appliqué sur un support. De nombreuses études expérimentales et mécanistes ont été rapportées dans la littérature scientifique dans le but d'évaluer la sélectivité en méthane, en alcanes, en alcènes, etc. et afin de déterminer les principales propriétés du catalyseur qui dirige la sélectivité. L'activité élevée du catalyseur, la sélectivité en alcènes légers qui est favorisée à un rapport H2/CO inférieur (grâce à la réaction du gaz à l'eau) et le coût inférieur font des catalyseurs à base de fer un meilleur choix pour la synthèse de Fischer-Tropsch, malgré le fait qu'ils produisent davantage de composés oxygénés proportionnellement. L'amélioration de la performance du catalyseur en ajoutant des promoteurs aux catalyseurs à base de fer a suscité l'intérêt de beaucoup de chercheurs. L'investigation de l'effet des promoteurs est plutôt complexe car de nombreux supports sont investigués à différentes températures de réaction. Afin d'atteindre une compréhension fondamentale des phénomènes catalytiques de ce genre, beaucoup d'études computationnelles ont été effectuées utilisant divers modèles mécanistes.

Parmi les modèles mécanistes, le modèle cinétique complet à base du concept des Single Event MicroKinetics (SEMK) a été appliqué dans le domaine de l'oligomérisation, de la réticulation par auto-oxydation (autoxidative curing), etc. et s'est avéré utile comme outil polyvalent pour la modélisation de la synthèse de Fischer-Tropsch. Cependant, le développement de modèles mécanistes pour chaque problème d'ingénierie chimique n'est pas toujours faisable vu la complexité des procédés et les connaissances approfondies nécessaires pour la construction d'un modèle de ce genre. Cela est parti-

culièrement le cas dans le domaine de la catalyse, puisque la complexité des procédés est plus importante et les modèles cinétiques plus simples ne sont pas capables de saisir les détails de la chimie du procédé. L'opportunité se présente dans ce cas pour le développement de modèles à base d'apprentissage automatique (machine learning), car ils sont plus faciles à développer par rapport aux modèles mécanistes détaillés, et sont donc également examinés dans cette thèse. Les capacités des deux méthodes sont utilisées pour l'identification du catalyseur optimal menant à une meilleure sélectivité en alcènes légers et pour l'identification des conditions opérationnelles correspondantes appropriées au catalyseur.



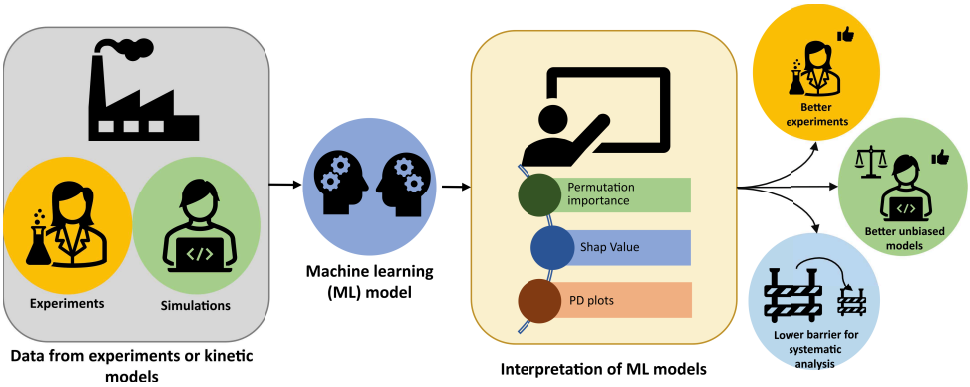

En premier lieu, la thèse concerne un screening virtuel de catalyseurs potentiels à base d'un modèle SEMK, visant une formation d'alcènes légers améliorée dans la synthèse de Fischer-Tropsch. Les propriétés du catalyseur ont été quantifiées à l'aide de descripteurs de catalyseur tels que les enthalpies de chimisorption atomique de l'hydro-



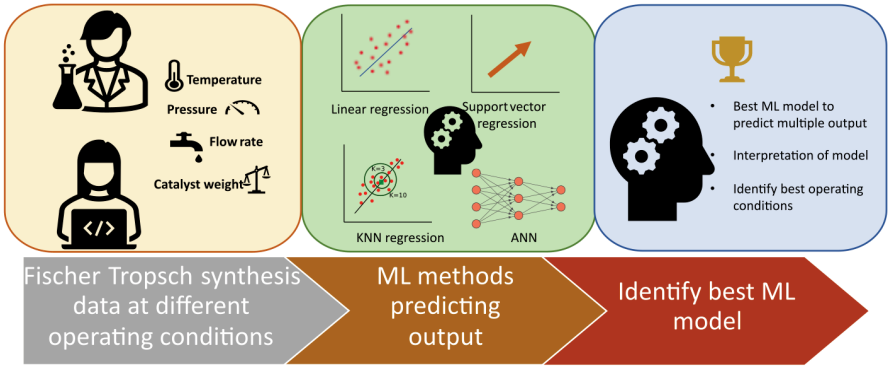

gène, du carbone et de l'oxygène. L'espace des descripteurs est exploré en simulant la performance expérimentale des catalyseurs à base de fer promu, tout en recherchant les descripteurs qui mènent à une sélectivité en alcènes légers supérieure. Les simulations du modèle SEMK ont été réalisées en utilisant plusieurs combinaisons de descripteurs afin d'identifier visuellement (à l'aide d'intervention humaine) les propriétés du catalyseur optimal menant à une meilleure sélectivité en alcènes légers.

La méthodologie a ensuite été généralisée à l'aide d'apprentissage automatique non supervisé, ce qui a permis de dépasser les trois dimensions et de réduire le besoin d'intervention humaine. L'espace des descripteurs généré à partir de données microcinétiques (catalyseurs virtuels) est exploré en utilisant la méthode systématique du regroupement (clustering) et du classement (labelling) non supervisés. L'espace de la performance du catalyseur est regroupé en clusters et le nombre minimale de clusters est identifié. Chaque catalyseur virtuel (représenté par une certaine combinaison de descripteurs) est identifié du point de vue du cluster auquel il appartient. Il est ainsi possible d'obtenir l'étendue des valeurs des descripteurs dans le cluster ayant le meilleur rendement d'alcènes légers. Il est observé que les valeurs obtenues sont conformes à celles du catalyseur virtuel optimal identifié dans l'inspection visuelle précédente. On peut donc conclure qu'une méthode combinant la microcinétique et l'apprentissage automatique a été présentée pour le développement des catalyseurs et pour l'investigation détaillée de leurs propriétés, tout en diminuant le besoin d'intervention humaine.

Ensuite, une évaluation détaillée a été réalisée sur l'apprentissage automatique afin




d'évaluer la possibilité d'obtenir une performance similaire aux modèles SEMK. En première instance, la priorité a été accordée à la recherche d'un catalyseur ayant une sélectivité en méthane supérieure, donc seulement la sélectivité en méthane est considérée dans ce cas-ci. Un modèle d'apprentissage automatique à base d'un réseau de neurones artificiels (Artificial Neural Network) est développé pour la synthèse de Fischer-Tropsch catalysée par le cobalt. Le modèle est basé sur la sélectivité en méthane et est conçu en utilisant un ensemble de données générées par un modèle SEMK. L'importance relative (ou le classement) des variables du procédé (ou les prédicteurs) dans le processus de prise de décisions du modèle d'apprentissage automatique a été trouvée en utilisant des techniques d'interprétation comme l'importance de la permutation (permutation importance), la valeur SHAP (SHAP value) et les graphiques de dépendance partielle (partial dependence plots). Il est donc démontré dans ce travail que les techniques courantes dans le domaine de la science des données peuvent également être utilisées pour l'analyse systématique et l'interprétation des données cinétiques. Une analyse similaire en conjonction avec des données expérimentales pourrait aussi aider les chercheurs dans l'analyse préliminaire des données en trouvant les tendances masquées au sein des données. Une investigation similaire a ensuite été effectuée sur le potentiel des catalyseurs à base de fer, visant une sélectivité en alcènes légers améliorée, donc plusieurs sélectivités




sont désormais considérées.

Finalement, la méthode d'apprentissage automatique a été étendue dans l'intention de pouvoir réaliser des prédictions de plusieurs sélectivités en se concentrant sur la production des alcènes légers à différentes conditions opérationnelles. Afin d'atteindre cet objectif, 4 modèles d'apprentissage automatique alternatifs ont été employés, i.e. la méthode lasso (lasso regression), la méthode des k plus proches voisins (k nearest neighbor regression ou KNN), la méthode de la machine à vecteurs de support (support vector machine regression ou SVR) et le réseau de neurones artificiels (Artificial Neural Network ou ANN). Les capacités de ces techniques sont évaluées par rapport à la reproduction du comportement linéaire de la conversion et la sélectivité en fonction des variables du procédé comme cela a été simulé par le modèle SEMK. Il est constaté que les modèles à base d'un réseau de neurones artificiels correspondent le plus aux résultats de référence du modèle SEMK. Une analyse supplémentaire utilisant la technique d'interprétation de la valeur SHAP a été appliquée aux modèles à base d'un réseau de neurones artificiels ayant la meilleure performance, en vue de mieux expliquer le fonctionnement des modèles.

L'ensemble de l'étude a rapporté des connaissances essentielles, telles que les descripteurs de catalyseur optimales: les enthalpies de chimisorption atomique de l'hydrogène

($Q_H \approx 234$kJ/mol), du carbone ($Q_C \approx 622$kJ/mol) et de l'oxygène ($Q_O \approx 575$kJ/mol), pour la conception de catalyseurs ayant une sélectivité en alcènes légers élevée en utilisant un modèle SEMK mécaniste. De plus, l'étendue des conditions opérationnelles menant à la meilleure sélectivité en alcènes légers a été déterminée en adoptant plusieurs stratégies de modélisation (le concept des SEMK et l'apprentissage automatique). Il a été constaté que les effets de la température (580-620K) et la pression (1-2 bar) étaient les plus importants. Ensuite, une investigation est réalisée dans le but d'évaluer à quel point les résultats des modèles d'apprentissage automatique correspondent à ceux du modèle SEMK. La meilleure reproduction du comportement non linéaire des données cinétiques est produite en utilisant un modèle à base d'un réseau de neurones artificiels ayant une valeur de REQM (RMSE en anglais) d'environ $10^{-2}$. Le travail a été effectué dans l'intention de découvrir les possibilités d'employer des modèles d'apprentissage automatique en conjonction avec des modèles mécanistes afin de réduire véritablement l'effort manuel et computationnel. Par exemple, une analyse préliminaire a pu être réalisée en utilisant un modèle d'apprentissage automatique pour l'analyse des données obtenues à l'aide d'expérimentation à haut débit. Ensuite, le modèle mécaniste a permis d'acquérir une compréhension chimique approfondie du procédé.

# 1

# INTRODUCTION

The circular economy model in the 21st century facilitates overcoming the drawbacks associated with the current system of linear economy, where the take, make, and dispose principle is employed. Circular economy aims at maintaining the value of raw materials, components and material products with minimum waste production and polluting elements. It is based on achieving environmental and social sustainability targets using a closed loop system by employing a system-level thinking. It imposes sustainability at the core of economic activities. Taking into account its relevance, the European Commission has adopted many action plans like circular economy action plan (CEAP-2020), in this direction (Fig.1.1). The action plans aims at reducing pressure on natural resources and creating sustainable growth which will act as a prerequisite for EU's climate neutrality target 2050 [1].

Striving towards a circular economy target has led to the re-investigation of many existing processes, with the demand for developing sustainable chemical production processes as one its key drivers. The life cycle management of plastics is one of the impor-

1

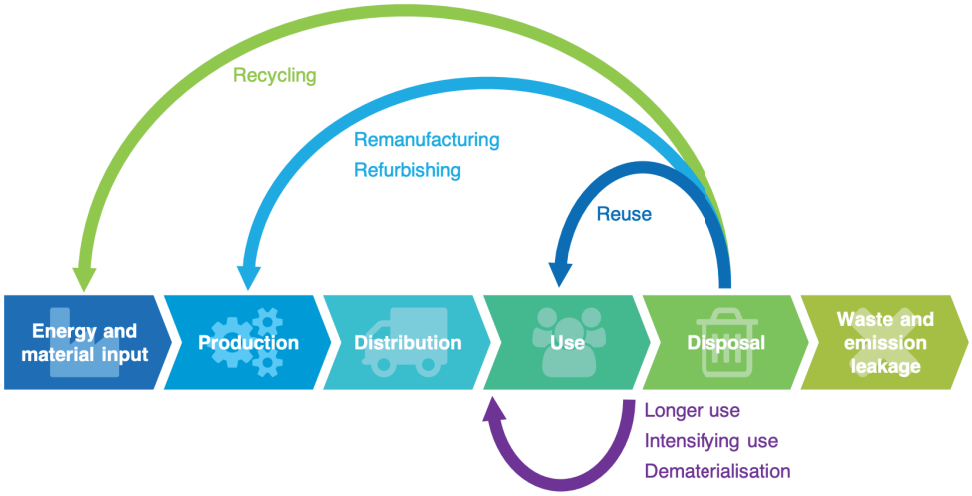



Figure 1.1: Vision for circular economy adapter from [2]

tant topics currently investigated, due to the role of plastics in environmental pollution.

A circular economy for plastics demands employing new and sustainable approaches involving eco-design, reuse, recycle, and use of chemical conversion strategies. Due to the limitation of value reduction associated with mechanical recycling of plastics, technologies based on thermo-chemical conversion, are investigated as an alternative through the EU Interreg project PSYCHE [3].

## 1.1. PLASTIC WASTE AND ITS RECYCLING

In our economy, plastics form an important and omnipresent material affecting our daily lives. They are inexpensive, durable, corrosion resistant, and light weight leading to their use in a wide variety of applications. This has resulted into an enormous increase in the production of plastics in the last 60 years reaching a global production of 367 million tons in 2020 [4]. The increased level of usage and disposal rate has also increased alarmingly, in the last decades. Due to its non-degradable properties, they accumulate as debris in natural habitats worldwide. Its production accounts for around 4 percent usage of non-renewable resource like oil and gas feedstock and 3-4 percent of global energy

Figure 1.2: Plastic accumulation in a landfill. Image obtained from United Nations Development Programme in Europe and CIS photo-stream.

consumption. It is also estimated that the production of plastics will double in the next 20 years [5]. This demands addressing the management of plastic waste at an immediate basis.

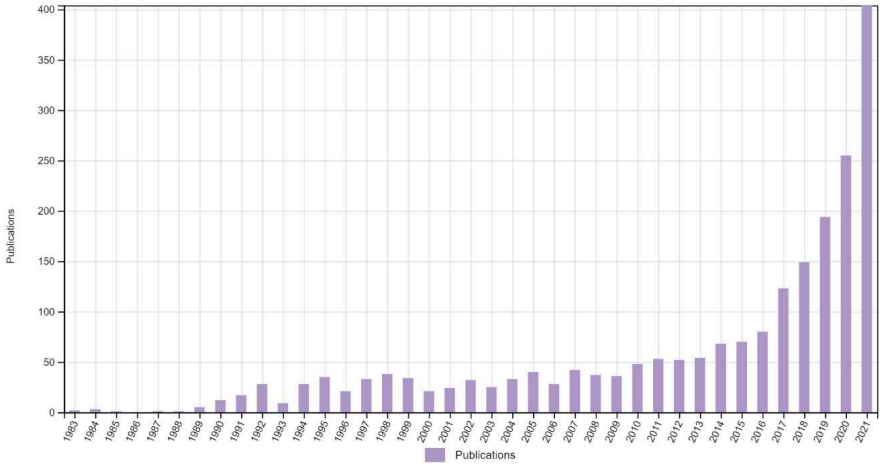



Figure 1.3: Number of publications in the area of plastic chemical recycling [6].

1

Among the plastic waste management the landfill Fig.1.2 strategies possess long-term contamination risk of soil and groundwater and hence makes it non-feasible. The mechanical recycling of plastic suffers from the inability to fully recover its potential. An important method to fully recover its value is by thermo-chemical conversion strategy. Thus, in this decade, there has been an increase in research [7] in this area (Fig.1.3).

## 1.2. GASIFICATION AND SYNGAS PURIFICATION

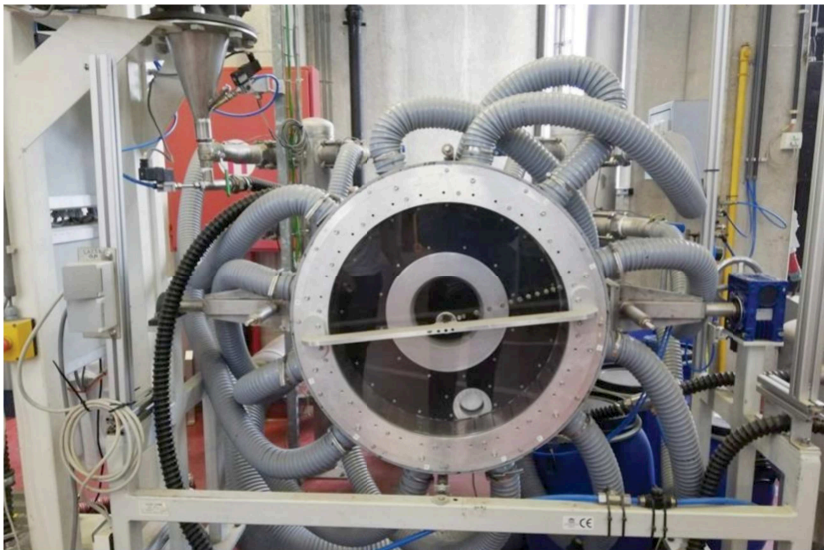


Figure 1.4: Vortex reactor technology developed in Ghent university used for pyrolysis step in gasification [8].

An important step towards chemical recycling of plastics involve a preliminary thermal gasification process at a temperature (800 to 1000° C) where plastics are converted into gases such as CO, $CO_2$, $CH_4$, $H_2$ and halogens [9]. An important advantage of gasification is the flexibility to valorize plastics of varied composition as well as with different feed stocks. The main challenges in the gasification process are the suitability of reactors and the optimal operating conditions for the targeted syngas formation. The different reactors types [9, 10] used are based on fixed bed, fluidized bed and vortex reactor technologies (Fig.1.4).

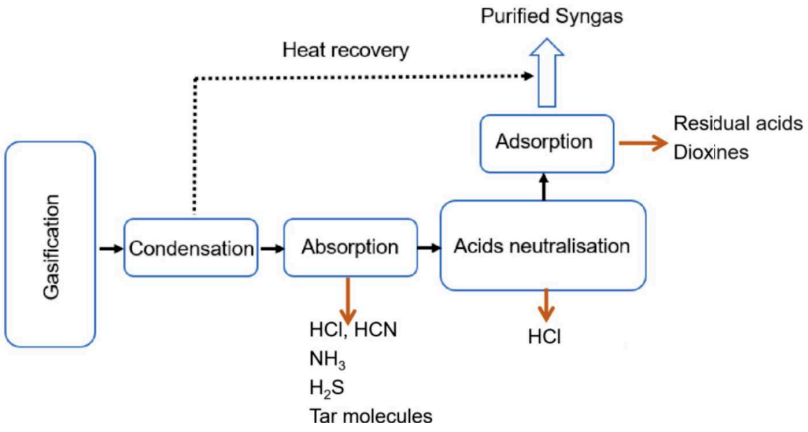



Figure 1.5: Schema: syngas purification after gasification.

The purification methods generally involve using physical methods (for e.g. wet scrubbing, filtration, electrostatic precipitation), thermal and/or catalytic processes [11, 12]. With such purification (Fig.1.5) methods, impurities such as HCl, $NH_3$, $H_2S$, $NO_x$, alkali metals, and condensable tars are reduced to their maximum allowable level in syngas for the Fischer-Tropsch process.

## 1.3. FISCHER–TROPSCH SYNTHESIS, AN OVERVIEW

Fischer-Tropsch synthesis (FTS) is one of the most promising catalytic approaches [14] to convert the syngas (CO and $H_2$) obtained via gasification of coal, biomass, plastics etc. into non-petroleum-based hydrocarbons (Fig.1.6).

$$(2n+1)H_2 + nCO \rightarrow C_nH_{2n+2} + nH_2O \quad \text{Alkane production} \tag{1.1}$$

$$(2n)H_2 + nCO \rightarrow C_nH_{2n} + nH_2O \quad \text{Alkene production} \tag{1.2}$$

$$CO + H_2O \rightleftharpoons CO_2 + H_2 \quad \text{Water-gas shift reaction} \tag{1.3}$$

1

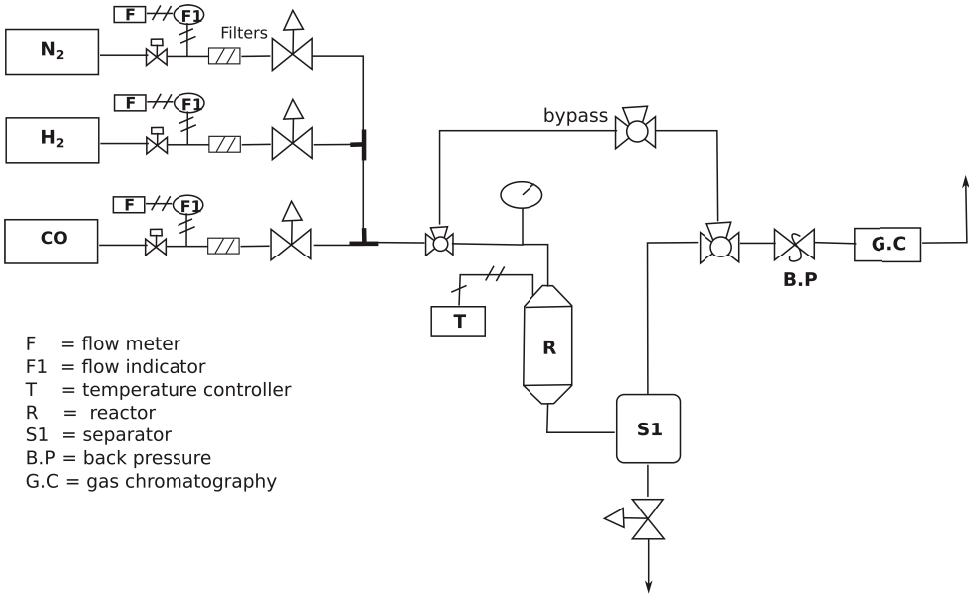



Figure 1.6: Schematics of the Fischer Tropsch catalytic reactor set up in SusCat, Université Lille [13].

This can be subsequently transformed into value added products like transportation fuels and base chemicals. There are two main categories in FTS technologies based on the operating temperature: High Temperature Fischer Tropsch (HTFT) (Fig.1.7 (a-b)) and Low Temperature Fischer-Tropsch (LTFT) (Fig.1.7 (c)) with operating modes between 300° C to 350° C and 180° C to 250° C, respectively [14–16].

The hydrocarbon mixture formed by LTFT mainly consists of long-chain hydrocarbons whereas the HTFT reaction conditions yield shorter hydrocarbons, especially lower olefins. The targeted reaction products and the difference in operating temperature determine the applied reactor technologies such as fixed fluidized bed reactor, tubular trickle bed reactor etc. Apart from these conventional reactor technologies for FTS, other investigated reactor technologies are micro-structured and membrane reactor based technologies.

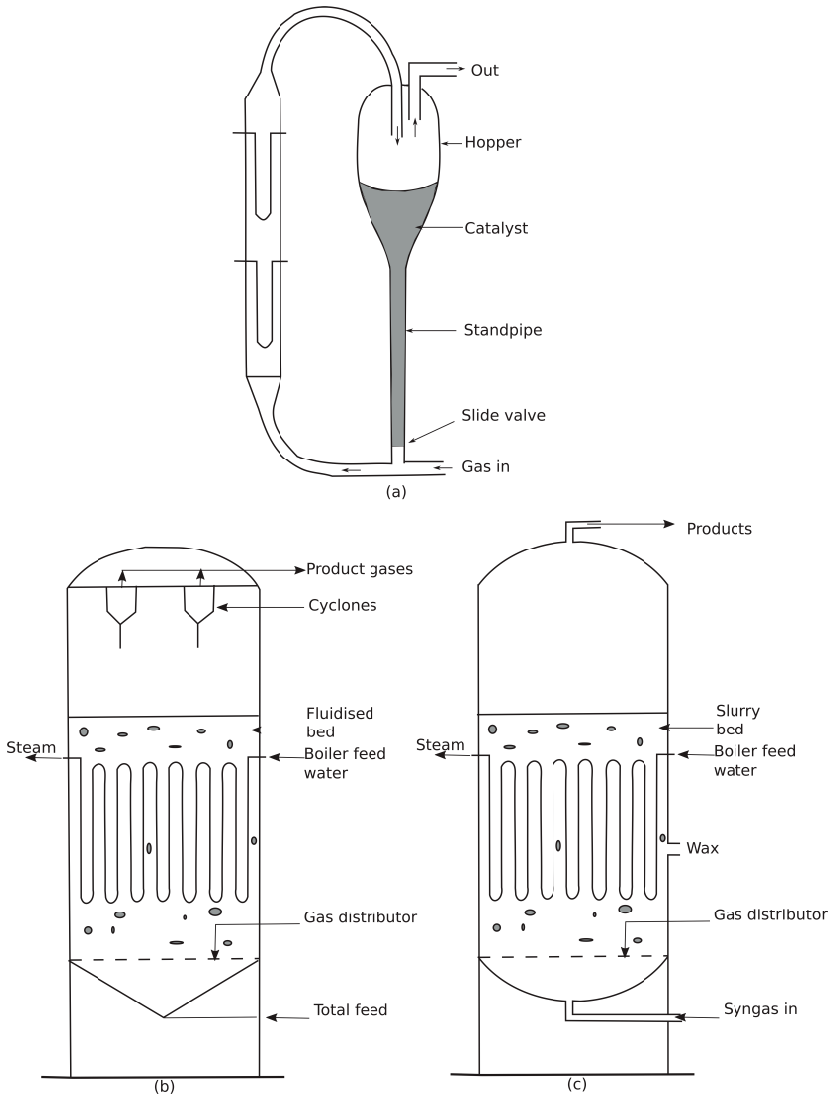



Figure 1.7: Circulating Fluidized Bed Reactor (CFBR) (a), Fixed Fluidized Bed Reactor (FFBR) (b), Slurry Bubble Column Reactor (SBCR) (c). The CFBR and FFBR are High Temperature Fischer-Tropsch (HTFT) reactor technologies. The SBCR is a Low Temperature Fischer-Tropsch (LTFT) reactor technology. The schematics are adopted from [15, 17].

1

## 1.4. FTS CHEMISTRY

In FTS, catalytic conversion of syngas (CO + $H_2$) can synthesize a wide variety of products. This includes paraffins, olefins, and oxygenates (alcohols, ketones, aldehydes). The need to steer the selectivity towards these hydrocarbons has led to increased research interest into the FTS. A variety of mechanisms have been proposed to understand this complex hydrocarbon formation. Some of the widely accepted classes of mechanisms are carbide-based mechanisms, CO-insertion mechanisms, and hydroxy-carbene mechanisms.

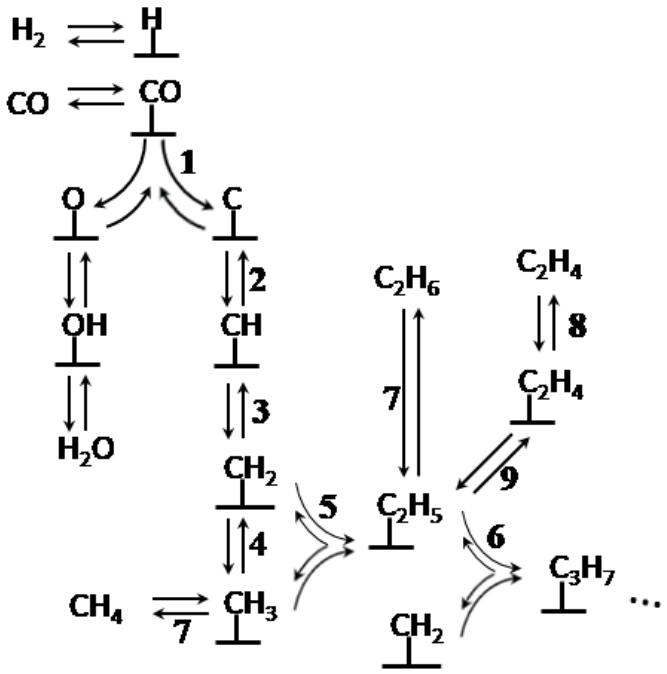



Figure 1.8: The carbene mechanism as proposed by Brady and Pettit [18, 19].

### 1.4.1. CARBIDE-BASED MECHANISM

The mechanistic proposals (originally proposed by Brady and Pettit [19]) in this class has a CO dissociation step to form carbon (C) and oxygen (O). Then, carbon is hydrogenated

to form methylene ($CH_2$) and metal methyl ($CH_3$) surface species which then couples leading to chain growth initiation step. Further chain growth takes place by consecutive insertions of the $CH_2$ species into growing metal alkyl chains. (Fig.1.8). Hydrogenation of the metal- alkyl species leads to alkane formation, and beta-hydride elimination of the corresponding metal alkyl species leads to alkene formation.

### 1.4.2. CO INSERTION MECHANISM

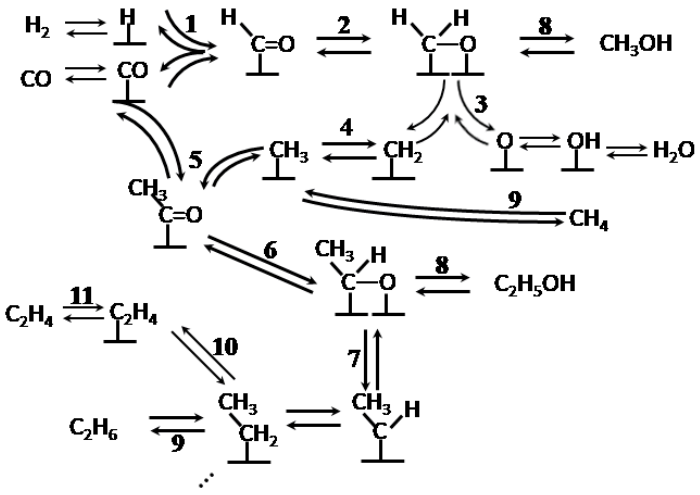



Figure 1.9: The CO insertion mechanism as proposed by Pichler and Schulz [20].

The formation of typical FTS products including oxygenates were explained via the CO insertion mechanism by Pichler and Schulz [20]. Here the surface species undergoes hydrogenation before the C-O bond dissociates. Instead of CH2 coupling (as in case of carbene mechanism), here, chain growth takes place by step-wise insertion of CO in to an $RCH_2$ group and thus inherently accounts the formation of alcohols in the FTS product mixture through hydrogenation of the O atom of an aldehyde surface species (Fig.1.9).

1

### 1.4.3. HYDROXY-CARBENE MECHANISM

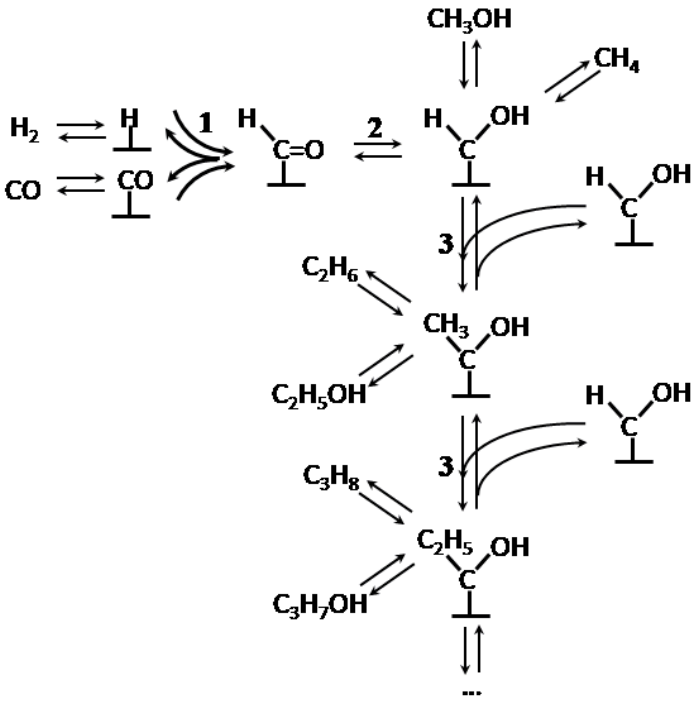



Figure 1.10: The hydroxycarbene mechanism as proposed by Anderson et al. [21].

The mechanism represented in Fig.1.10 was proposed by Anderson et al. [21]. In this mechanism a hydroxy-carbene surface species is formed by H addition to both the C and the O. Here the chain growth occurs by a condensation reaction of the hydroxy-carbene species. Thus it explains the formation of oxygenates in the FTS product mixture. Regardless of the mechanisms, Fischer Tropsch being a polymerization process, the product spectrum follows the Anderson-Schulz-Flory (ASF) distribution [22], and is characterized by means of a single chain growth probability, $\alpha$, defined by:

$$\alpha = \frac{r_p}{r_p + r_t} \tag{1.4}$$

Where $r_p$ the propagation rate and $r_t$ the termination rate. The hydrocarbons mass frac-

tion with n carbon atoms, $w_n$, being a function of the carbon number is related as follows:

$$w_n = n(1-\alpha)^2 \times \alpha^{(n-1)} \tag{1.5}$$

The strategies to steer the reaction to a specific range of hydrocarbons remain one of the key challenges of Fischer Tropsch synthesis where chain growth probability (alpha) can be changed by varying the catalyst composition, reaction temperature, space-time, and $H_2$/CO ratio.

## 1.5. FTS CATALYSTS

Fischer Tropsch synthesis is a catalyzed process which, is typically carried out with Group VIII metals as catalysts. In particular, (Fe), (Co), (Ni), (Ru) are the most important active catalysts for FTS. At present, the FTS process mostly uses iron-based (LT- and HTFT) catalysts due to their high activity and low cost (Table 1.1) compared to other metals. Both cobalt iron catalysts yield higher hydrocarbons. Cobalt catalysts typically have better yields and stability in production of linear saturated hydrocarbons. The high price and high sensitivity to sulfur contamination are disadvantages for their use. Ni, leads to methane synthesis as the main FTS product in practical operating conditions. Ru, a highly active metal for FTS, due to its exceptionally high cost compared to Co and Fe makes it less suitable for practical use [23].

Table 1.1: Relative cost of catalysts [14].

| **Metal** | **Price ratio** |
|---|---|
| Fe | 1 (base) |
| Ni | 250 |
| Co | 1000 |
| Ru | 35000 |

1

Higher activity, light olefin selectivity, and lower costs make iron-based catalysts a better candidate in FTS even though they lead to more oxygenates. Due to their water-gas shift (WGS) activity, syngas with varying $H_2/CO$ ratio, could be used for the process. The improvement in catalyst performance with addition of promoters in iron-based catalysts [24–26] has attracted recent research interests. The study of the promoter effect is rather complex as it is done with different supports and at different reaction conditions. Recent studies investigate the effect of iron-promoted catalysts on performance. This is due to the fact that the catalytic is affected by metallic iron, different iron oxide presence in addition to the iron carbide phase.

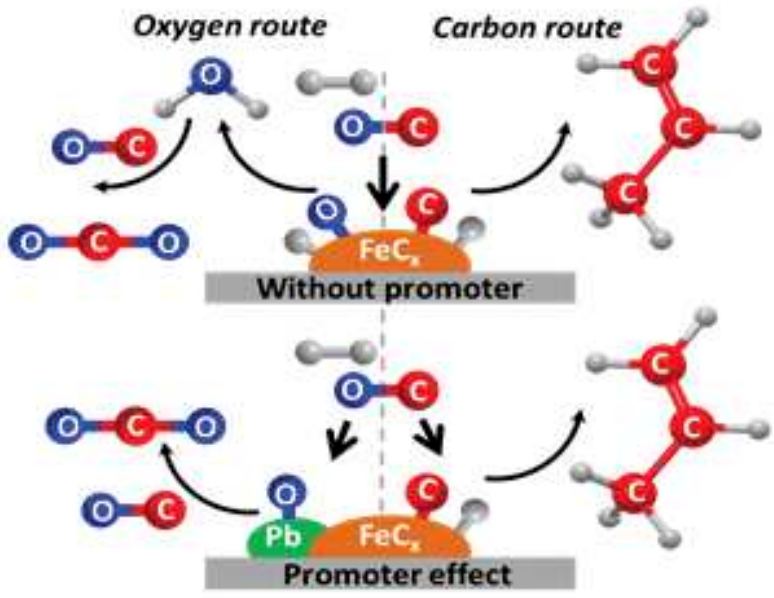



Figure 1.11: The influence of promoter.

The extent of iron carbidization, and increase in iron dispersion, light olefin selectivity and Fischer-Tropsch reaction rates are also reported by using electronic and structural promoters. The active site intrinsic activity is enhanced by the electronic promoters (Fig.1.11). They impact the intrinsic rate of the reaction elementary steps and helps in

shifting the reaction selectivity to the target product. The structural promoters helps in increasing the dispersion of active phase, by improving the mechanical strength, and also by stabilizing the catalyst surface. The overall impact on catalytic performance by promotion is affected by the interplay of structural and electronic promoters, promoter concentration, their surface coverage and the active phase interaction. Gu et al. [27] report a higher light olefin selectivity, specifically on promotion by Bi and Pb on carbon nanotube support, over unpromoted Fe catalyst.

The performance of a Fischer Tropsch process could thus be related to the catalyst design, and finding the operating conditions. Modelling tools based on kinetic (power law, Langmuir Hinshelwood, Microkinetic) and data-driven models can provide fundamental insight in this direction. The modelling approaches used in this thesis are based on Microkinetic and machine learning methodology.

## 1.6. MICROKINETIC MODELING

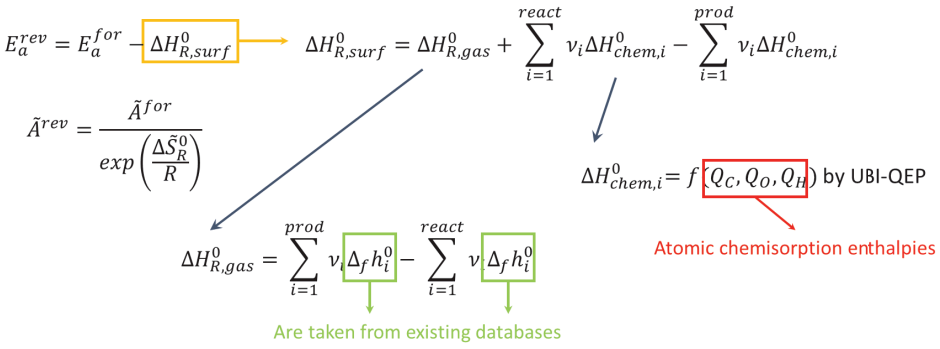



Figure 1.12: The steps in single event Microkinetics for FTS.

In order to provide fundamental insights on catalyst development and relate experimentally observed conversions and selectivities for varied operating conditions, kinetic models are employed. There are kinetic models with different levels of complexity. The commonly encountered simpler kinetic models are based on power law rate expressions.

1

The kinetic models which are comprehensive with detailed reaction mechanism in terms of elementary steps are the Microkinetic models.

Microkinetic models, which represent the occurring chemistry at the elementary step level [28–30] help to study the reaction kinetics of complex reactions such as the Fischer–Tropsch reaction [31], oligomerization [32], autoxidative curing [33], etc. To model reaction systems with a large number of elementary steps using a microkinetic approach, the single event concept can be employed.

In the Single-Event MicroKinetic (SEMK) methodology for FTS, the reactive moieties are considered to determine the reactivity of individual molecules. Accordingly, the elementary reactions are classified into reaction families to reduce the number of parameters [36]. The single-event kinetic coefficient is unique for a reaction family and the adsorption enthalpies of surface species are calculated via the UBI-QEP method [37, 38]. In the SEMK framework, the catalysts properties could be differentiated in terms of model parameters referred to as catalyst descriptors.

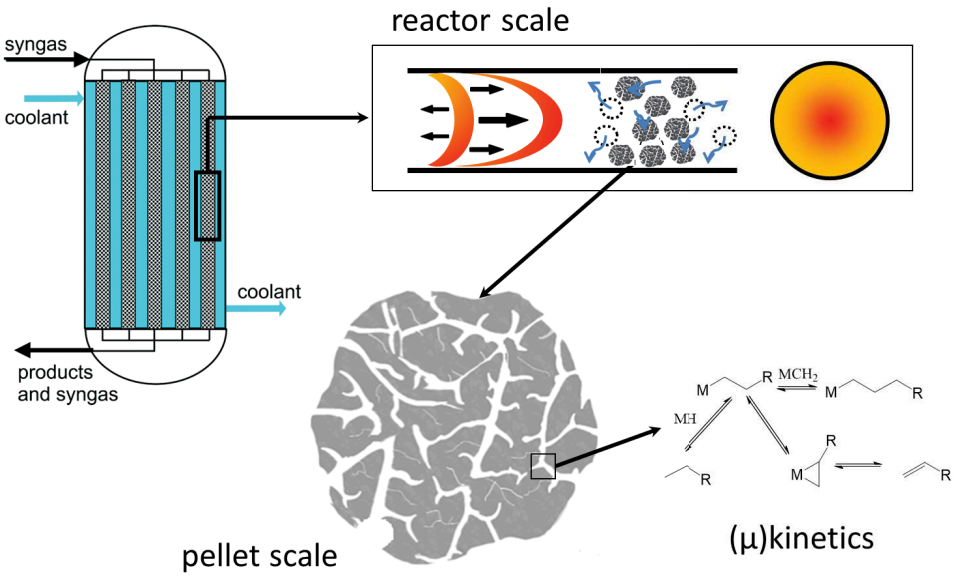



Figure 1.13: Schematics representing multiscale modelling capabilities of the micro kinetic model.

Table 1.2: Elementary reactions and reaction families in the reaction network [34], where $E_a^{for}$ represent the kinetic descriptors, M represents the metal surface, and $\tilde{A}^{for}$, represent the forward pre-exponential factors, and $\Delta\tilde{S}_r^0$ represent single-event surface reaction entropies, respectively. The surface reaction enthalpies is a function of catalyst descriptors. The calculations of enthalpies and entropies are reported by Lozano-Blanco et al. [34, 35].

| ***Reaction family/ elementary reaction*** | $E_a^{for}$ **(kJ/mol)** | $\tilde{A}^{for}$ ($s^{-1}$ **or** $(MPas)^{-1}$) | $\Delta\tilde{S}_r^0$ ($J(Kmol)^{-1}$) |
|---|---|---|---|
| 1. $H_2$+ 2M $\rightleftharpoons$ 2MH | 0 | $3.1 \times 10^8$ | −63.5 |
| 2. CO + 2M $\rightleftharpoons$ MMCO | 0 | $2.2 \times 10^7$ | −165.7 |
| 3. MMCO + 3M $\rightleftharpoons$ MMMC + MMO | 56.81 ± 0.53 | $1.3 \times 10^{13}$ | −17.34 |
| 4. MMMC + MH $\rightleftharpoons$ MMMCH + M | 77.66 ± 0.70 | $8.8 \times 10^{14}$ | +35.08 |
| 5. MMMCH + MH $\rightleftharpoons$ $MMCH_2$+ 2M | 11.94 ± 0.10 | $5.9 \times 10^{11}$ | −25.6 |
| 6. $MMCH_2$ + MH $\rightleftharpoons$ $MCH_3$ + 2M | 61.88 ± 0.50 | $2.2 \times 10^{11}$ | −33.5 |
| 7. $MC_nH_{2n+1}$ + $MMCH_2$ $\rightleftharpoons$ $MC_{n+1}H_{2n+3}$ + 2M | 44.79 ± 0.43 | $8.3 \times 10^9$ | −60.69 |
| 8. $MC_nH_{2n+1}$ + MH $\rightleftharpoons$ $C_nH_{2n+2}$ + 2M | 117.75 ± 0.67 | $3.3 \times 10^{10}$ (n=2)<br>$2.0 \times 10^{10}$ (n=3-10) | +64.3 |
| 9. $MC_nH_{2n+1}$ + M $\rightleftharpoons$ $MC_nH_{2n}$ + MH | 96.27 ± 0.50 | $1.0 \times 10^{10}$ | +12.32 |
| 10. $MC_nH_{2n}$ $\rightleftharpoons$ $C_nH_{2n}$ + M | 63(n=2)<br>61(n=3-10) | $1.3 \times 10^{13}$ | +118.6 |
| 11. O − CHO − M + M − OH +O $\rightleftharpoons$ O − COOH − M + O − H + M | 138.95 ± 1.15 | $1.7 \times 10^{14}$ | +53.2 |
| 12. MMO + MH $\rightleftharpoons$ MOH + 2M | 103.80 ± 0.96 | $1.3 \times 10^{12}$ | +33.92 |
| 13. MOH + MH $\rightleftharpoons$ H2O + 2M | 86.22 ± 0.62 | $2.4 \times 10^{11}$ | +73.7 |

The forward activation energies along with the standard surface reaction enthalpies, are used to calculate the reverse activation energies for the elementary steps by applying the principle of microscopic reversibility (Fig.1.12). The surface reaction enthalpies are calculated starting from those of the corresponding gas phase reactions and by accounting for the enthalpy change induced by chemisorption [39]. Chemisorption enthalpies of various surface species are calculated as a function of the atomic chemisorption enthalpies using the UBI-QEP method. The single-event forward pre-exponential factors as reported in Table 1.2 are determined via statistical thermodynamics calculations. The calculation involves gas phase entropies of the components and the single-event standard entropy change related to chemisorption [40]. The gas phase entropies are obtained either from databases [41] or using group additivity methods [42]. The calculation of the entropy change related to the chemisorption step is based on the loss of translational entropy of a gas phase component, which is inturn calculated with the Sackur-Tetrode equation [43]. The reverse pre-exponential factors are calculated using the forward pre-exponential factors ($\tilde{A}^{for}$) and single event surface reaction entropies ($\Delta\tilde{S}_r^0$) as:

$$\tilde{A_j}^{rev} = \frac{\tilde{A_j}^{for}}{exp\left(\frac{\Delta\tilde{S}^0_{r,j}}{R}\right)} \tag{1.6}$$

The pre-exponential factors are used for determining the rate coefficients of the reaction families in Table 1.2. All the above calculations are explained in detail by Lozano-Blanco et al. [34, 35]. The SEMK embedded in to reactor enables the multiscale modelling capabilities as represented in Fig.1.13.

The major advantage of this Microkinetic model includes extrapolation capabilities of the model on a varied dataset. But this comes at the expense of the effort required to construct the model, and expert-level knowledge required when compared to building a simpler power law or Langmuir kinetic model. Further details on microkinetic model is discussed in Chapter 2.

An alternate approach in predictive modelling is data-driven methods. With developments in increased computational capacity and the ability to handle large volumes of data, high throughput experimental trials, etc., there has been an increased interest in applying data driven, machine learning approaches to chemical engineering problems.

## 1.7. MACHINE LEARNING BASED MODELS

Machine Learning (ML) methods are rapidly becoming popular for modelling and decision support of complex nonlinear process phenomena in the field of chemical engineering [44], owing to faster (cheaper) computational capabilities when compared to full-fledged mechanistic models. Depending on the process data complexity, the specific type of machine learning method suited to model the phenomena of interest varies. The current studies for modeling the chemical engineering research problems are mostly based on mechanistic models [45–47] which rely on imposing the physico-chemically meaningful relations in the data. However, developing mechanistic models for processes such as FTS 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 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 machine learning based models which are easier [48] to develop as compared to a detailed mechanistic model. The machine learning methods are typically subcategorized as either unsupervised (e.g. clustering, dimensionality reduction) or supervised (classification, regression) learning.

### 1.7.1. UNSUPERVISED MACHINE LEARNING

Unsupervised learning is the subcategory of machine using data without any label or target attribute. It involves exploring the unlabeled data to find intrinsic structures or hidden pattern with minimum of human supervision [49]. The major unsupervised ML

1

methods used are:

- Dimensionality reduction, where high dimensional dataset is reduced in dimension
- Cluster analysis, where data points that show similarity are made in to groups.

In this thesis, the investigation based on unsupervised methods is limited to the use of clustering.

**Clustering:**

Clustering is a technique for identifying similar groups in data, called clusters. i.e., a grouping of data instances which are similar into one cluster and instances that are very different into different clusters. It works on the principle of maximization of inter-clusters distance and minimization of intra-clusters distance to ensure the quality of clustering. The clustering technique used in the thesis is K-means clusters based on partition based clustering [50].

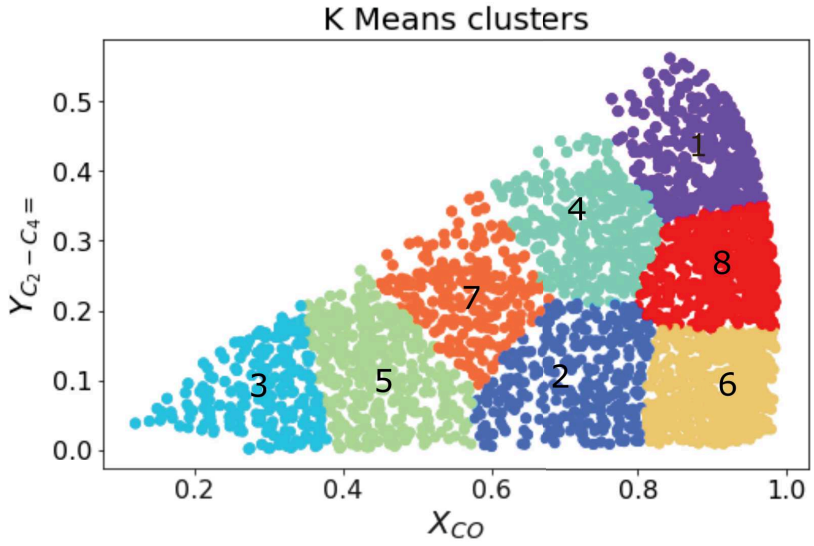



Figure 1.14: llustration of cluster obtained via K- means clustering. The data is separated into k [k=8] clusters in the feature space (yield of light olefins and carbon monoxide conversion).

In the algorithm k data points (see Fig.1.14) are chosen to be the initial centroids or

cluster centers and each data point is assigned to the closest centroid. The is followed by reassigning the centroid of each group formed with the corresponding mean of the data and iteratively repeated until it converges. With the advances in computational capacity, the investigations of different chemical engineering problems such as automated Phase Identification [51], identifying key features in reactive flows [52], fault diagnosis in chemical process [53], electrocatalysis [54], etc. have recently grown.

### 1.7.2. SUPERVISED MACHINE LEARNING

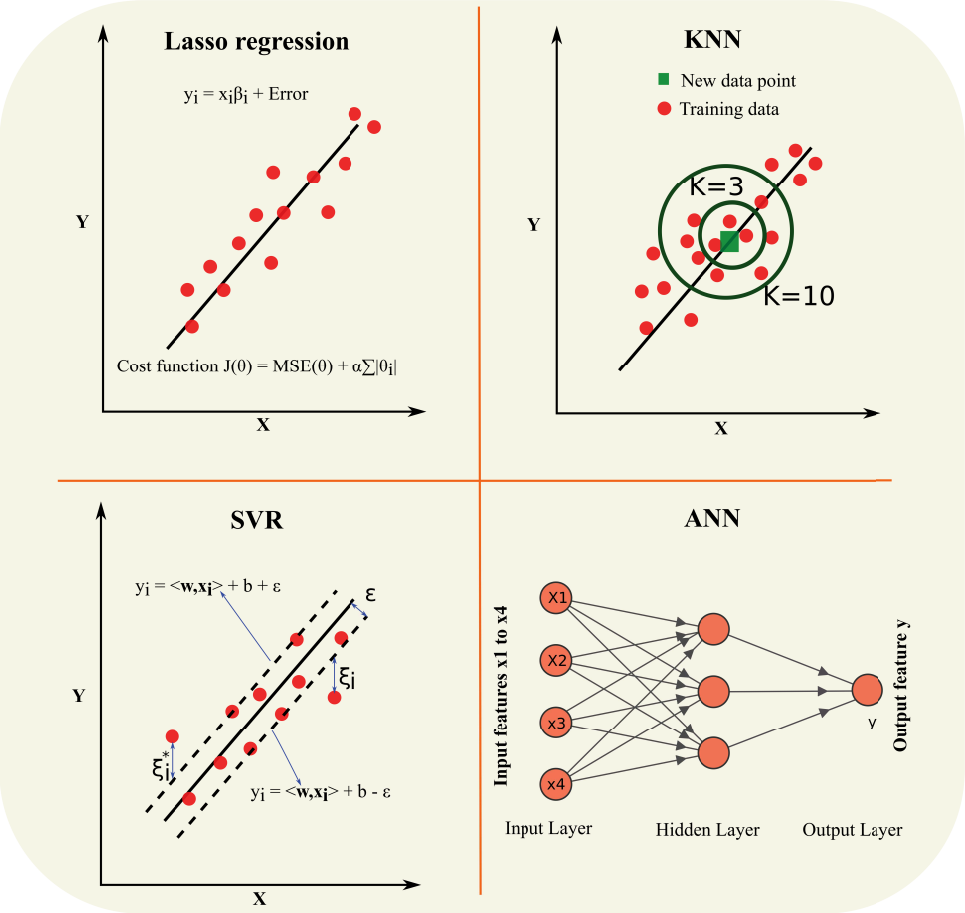



Figure 1.15: Different machine learning models investigated to analyse multi-output FTS reaction.

1

Supervised learning is the training of a machine using data with a label or target attribute. The main type of supervised machine learning for continuous value prediction is the regression method. The aim is to approximate the mapping function (f) such that when a new input data (x) is provided the output variables (Y) for that data can be predicted.

In chemical engineering, supervised learning methods (mainly regression-based methods) are the widely used, for applications such as absorption [55], sludge treatment [56], reactor modelling [57], etc. ML regression methods reported in literature include the use of Lasso regression [58], decision tree [59], K-Nearest Neighbor (KNN) regression [60], Support Vector Regression (SVR) [61], Artificial Neural Network (ANN) [62] etc. The wide application of these techniques in catalysis has recently been reported by Takahashi et al. [63]. In this thesis the investigation based on supervised regression methods are limited to use of Lasso, KNN, SVM and ANN regression.

### LASSO REGRESSION

The most general supervised ML model 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 $\beta_i$. 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 [64], where regularization or shrinkage of the parameters is done to reduce over-fitting. It penalizes the less important inputs in the dataset and thus creates a simple model by reducing their respective weights. Here the L1 norm [1] is used for regularization. The cost function containing the sum of mean square error (MSE) of the actual and predicted output, and the

[1] L1 norm is defined as the sum of the absolute values of the weights, $\beta_i$

penalty as shown in Eq.5.1, is minimized to find the parameters in the Lasso regression.

$$J = MSE + \alpha \sum_{i=1}^{n} |\beta_i| \tag{1.7}$$

where $\alpha$ is the hyperparameter that controls the intensity of regularization.

### K NEAREST NEIGHBORS (KNN) REGRESSION

KNN is a memory-based model [65] 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 observation, 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 potentially leading to over-fitting, while using large values of k considers more neighbors may result in not capturing sufficient variation. 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:

$$d = \sqrt{\sum_{i=1}^{n} (x_a - x_b)^2} \tag{1.8}$$

Here $x_a$, denotes a new point, $x_b$, an existing point and n, the number of data points.The main parameters associated with the KNN regression include the number of neighbors to be used and the metric used to compute the nearest neighbors.

### SUPPORT VECTOR REGRESSION

Support Vector Machine (SVM) [66] 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. SVR uses e-

1

insensitive loss as the loss metric. The e-insensitive loss is given as:

$$Le = max(0, |r(x, y)| - \epsilon) \tag{1.9}$$

where r(x,y) is the residual, i.e. the difference between the actual and the predicted output. A margin of width, $\epsilon$, as set by the modeller, 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. Thus, in essence, the aim is to contain maximum data pints within the margin, with a minimal number of violations. 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. [2]

#### ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is a machine learning algorithm [68] inspired by biological neural networks. It is a data-driven model which reveals the hidden patterns or non-linear relationships in the data. A neural network model is composed of its basic units called neurons, which are then stacked to form layers. An ANN 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 backpropagation [68]. This process continues until the difference between the predic-

[2] The popular kernel functions used are $d^{th}$ degree polynomial, radial basis function (RBF), and hyperbolic tangent. A penalty coefficient C that controls the strength of the penalty term (loss function) 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, $\gamma$ for RBF) that also needs to be optimized [67].

tion and the actual target is within the tolerance limit. If the number of neurons/layers 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 trial and error, which results in a network that yields a perfect prediction with minimization of a loss function.

### INTERPRETABILITY OF ML MODELS

With the increase in complexity of ML models ranging from the simplest Lasso regression to the most complex ANN the "black box" character of the model increases, and the interpretation of the models becomes more challenging. For a "simple" model such as linear regression, the weights or coefficients of the independent variables give 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 practically uninterpretable [69]. To this purpose different interpretation techniques such as permutation importance [70] and Shap values [71–73] are now being developed and used.

In the pursuit for discovery of catalysts and optimization of performance, these machine learning methods are an alternative for traditional catalyst optimization as it is based on generalization and summarization. However, due to the complexity involved in the reactions like FTS, factors affecting the catalyst performance are diverse and, hence need to be further understood at a fundamental level. Unfortunately, the contemporary studies mainly focus at either a mechanistic approach that demands expert-level knowledge on the topic, or ML-based data-driven methods that are used as black box models. Most of these studies do not exploit the advances made in data interpretation and analysis techniques in ML (see section ) to systematically analyze the kinetic data. These techniques can also be used to understand the predictions from black box models at a detailed level, thus building confidence in the ML models.

1

## 1.8. SCOPE AND OUTLINE OF THE THESIS

In this thesis, it is investigated how data-driven computational methods could be useful for the re-investigation of many existing chemical processes, especially with the increasing demand towards sustainability. Owing to the role of sustainable production of plastics in circular economy, chemical recycling of plastics and consequently Fischer-Tropsch synthesis (FTS) play a key role in it. In this thesis, a systematic analysis of kinetic data obtained from mechanistic SEMK model for FTS is carried out using interpretation techniques developed for ML models. This thesis aims at automated identification of optimal catalyst properties for enhanced light olefin production, from SEMK kinetic simulations. The capability of different ML models to identify optimum operating conditions for enhanced light olefin production is also investigated. Each chapter in the thesis has been accepted for publication in peer- reviewed journal. In this introductory chapter, the general concepts of the employed methodology has been briefly discussed and in the coming chapters the detailed methodology/applications of this will be discussed.

In Chapter 2, the application of microkinetics for FTS catalyst design is illustrated. The impact of promoters in iron based catalyst on the process (at a single set of operating condition) performance is analyzed with the SEMK model. The model consist of kinetic and catalyst parameters, referred to as kinetic and catalyst descriptors respectively. Here, an extension of the capabilities of the SEMK modelling approach to investigate the influence of variation in catalyst properties i.e. catalyst descriptors, on the yield of desired component, light olefins ($C_2 - C_4 =$) is carried out. A virtual catalyst screening combined with 3D graphical visualization approach is used to determine the catalyst descriptors, i.e. atomic chemisorption enthalpies of hydrogen , carbon and oxygen. The impact of the catalyst descriptors on conversion and selectively is rationalized in terms of relative surface coverages of different species, leading to different dominant reaction pathways, and thus resulting in product yield variations. Using this approach, a "promising potential catalyst" with catalyst descriptors, QH, QC and QO resulting in enhanced light olefins

yield with lower methanation and long-chain hydrocarbon formation, is identified.

In Chapter 3, the exploration of the synergies between microkinetic and data-driven machine learning models to provide fundamental insights on catalyst development is carried out. In this work, a clustering approach (unsupervised machine learning) coupled with microkinetic simulations is employed to determine the catalyst properties for achieving the desired target output such as light olefins. The application of the proposed methodology is demonstrated for Fischer-Tropsch Synthesis (FTS) on an iron-based catalyst, with the yield of light olefins as the desired target product. The approach demonstrated, automates the process of finding desired catalyst parameters (descriptors) irrespective of their dimensions and thus constitute an integral part in catalyst design and, hence, in chemical recycling technologies.

Chapter 4 tries to unravel the black box nature of the machine learning model based on Artificial Neural Network in its prediction, simultaneously with the verification of absence of any bias as compared to the micro-kinetic generated data for a single dominant output (methane) Fischer Tropsch process. Interpretation techniques such as permutation importance, Shap values and partial dependence plots is used for a more systematic (model agnostic) analysis of these data. For the purpose of this case study, the dataset required for training the ANN model is synthetically generated using a Single-Event MicroKinetic (SEMK) model. The work demonstrates how an ANN model with a simpler network is able to adequately reproduce the SEMK results. The relative ranking of the process variables, as learnt by the ANN model, is then identified using the interpretation techniques, and consistency with the physico-chemical understanding from SEMK is verified.

Chapter 5 showcases, how well the non-linearity in the kinetic data can be captured by different ML models. Multi-output Fischer-Tropsch Synthesis data, generated (at different operating conditions) via mechanistic SEMK model 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 interpretation technique based on Shap values is then applied on the best performing ML model and consequently ranked the input variables according to their process impact.

Finally in Chapter 6, the general conclusions concerning the insights gained from Microkinetic and machine learning models are presented along with the future perspectives for modelling applications in chemical processes.

# REFERENCES

[1] *Circular economy action plan (accessed in May 2022),* https://ec.europa.eu/environment/strategy/circular-economy-action-plan_en.

[2] M. Geissdoerfer, M. P. Pieroni, D. C. Pigosso, and K. Soufani, *Circular business models: A review,* Journal of Cleaner Production **277**, 123741 (2020).

[3] *Production of basic chemicals from plastic waste for reuse in the chemical industry,* https://psycheplastics.eu/.

[4] *Annual production of plastics worldwide from 1950 to 2020,* https://www.statista.com/statistics/282732/global-production-of-plastics-since-1950/.

[5] *The New Plastics Economy,* https://www3.weforum.org/docs/WEF_The_New_Plastics_Economy.pdf.

[6] *Web of Science, Citation report,* https://www.webofscience.com/.

[7] M. B. A. Rayaan, *Recent advancements of thermochemical conversion of plastic waste to biofuel-a review,* Cleaner Engineering and Technology **2**, 100062 (2021).

[8] *Vortex reactor technology, UGent,* https://www.lct.ugent.be/assets/vortex.

[9] G. Lopez, M. Artetxe, M. Amutio, J. Alvarez, J. Bilbao, and M. Olazar, *Recent advances in the gasification of waste plastics. a critical overview,* Renewable and Sustainable Energy Reviews **82**, 576 (2018).

[10] K. Ragaert, L. Delva, and K. V. Geem, *Mechanical and chemical recycling of solid plastic waste,* Waste Management **69**, 24 (2017).

[11] K. Liu, C. Song, and V. Subramani, eds., *Hydrogen and Syngas Production and Purification Technologies* (John Wiley & Sons, Inc., 2009).

[12] W. Zhang, H. Liu, I. U. Hai, Y. Neubauer, P. Schröder, H. Oldenburg, A. Seilkopf, and A. Kölling, *Gas cleaning strategies for biomass gasification product gas,* International Journal of Low-Carbon Technologies **7**, 69 (2012).

[13] *Facilities, SusCat, Université Lille,* `https://www.suscat.org/`.

[14] M. E. Dry, *The fischer–tropsch process: 1950–2000,* Catalysis Today **71**, 227 (2002).

[15] A. Steynberg, R. Espinoza, B. Jager, and A. Vosloo, *High temperature fischer–tropsch synthesis in commercial practice,* Applied Catalysis A: General **186**, 41 (1999).

[16] S. Saeidi, M. K. Nikoo, A. Mirvakili, S. Bahrani, N. A. S. Amin, and M. R. Rahimpour, *Recent advances in reactors for low-temperature fischer-tropsch synthesis: process intensification perspective,* Reviews in Chemical Engineering **31** (2015), 10.1515/revce-2014-0042.

[17] F. Botes, L. Dancuart, H. Nel, A. Steynberg, A. Vogel, B. Breman, and J. F. Freide, *Middle distillate fuel production from synthesis gas via the fischer–tropsch process,* in *Advances in Clean Hydrocarbon Fuel Processing* (Elsevier, 2011) pp. 329–362.

[18] R. C. Brady and R. Pettit, *Reactions of diazomethane on transition-metal surfaces and their relationship to the mechanism of the fischer-tropsch reaction,* Journal of the American Chemical Society **102**, 6181 (1980).

[19] R. C. Brady and R. Pettit, *Mechanism of the fischer-tropsch reaction. the chain propagation step,* Journal of the American Chemical Society **103**, 1287 (1981).

[20] H. Pichler and H. Schulz, *Neuere erkenntnisse auf dem gebiet der synthese von kohlenwasserstoffen aus CO und h2,* Chemie Ingenieur Technik - CIT **42**, 1162 (1970).

[21] M. Röper, *Fischer-tropsch synthesis,* in *Catalysis in C1 Chemistry* (Springer Netherlands, 1983) pp. 41–88.

[22] J. S. Albuquerque, F. O. Costa, and B. V. S. Barbosa, *Fischer–tropsch synthesis: Analysis of products by anderson–schulz–flory distribution using promoted cobalt catalyst,* Catalysis Letters **149**, 831 (2019).

[23] Q. Zhang, K. Cheng, J. Kang, W. Deng, and Y. Wang, *Fischer-tropsch catalysts for the production of hydrocarbon fuels with high selectivity,* ChemSusChem **7**, 1251 (2013).

[24] A. N. Pour, S. M. K. Shahri, H. R. Bozorgzadeh, Y. Zamani, A. Tavasoli, and M. A. Marvast, *Effect of mg, la and ca promoters on the structure and catalytic behavior of iron-based catalysts in fischer–tropsch synthesis,* Applied Catalysis A: General **348**, 201 (2008).

[25] H. Wan, B. Wu, C. Zhang, H. Xiang, and Y. Li, *Promotional effects of cu and k on precipitated iron-based catalysts for fischer–tropsch synthesis,* Journal of Molecular Catalysis A: Chemical **283**, 33 (2008).

[26] K. Cheng, V. V. Ordomsky, B. Legras, M. Virginie, S. Paul, Y. Wang, and A. Y. Khodakov, *Sodium-promoted iron catalysts prepared on different supports for high temperature fischer–tropsch synthesis,* Applied Catalysis A: General **502**, 204 (2015).

[27] B. Gu, V. V. Ordomsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, and 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**, 153 (2018).

[28] D. J. Dooling and L. J. Broadbelt, *Microkinetic models and dynamic Monte Carlo simulations of nonuniform catalytic systems,* AIChE Journal **47**, 1193 (2001).

[29] A. G. Hansen, W. J. M. van Well, and P. Stoltze, *Microkinetic modeling as a tool in catalyst discovery,* Topics in Catalysis **45**, 219 (2007).

[30] A. H. Motagamwala, M. R. Ball, and J. A. Dumesic, *Microkinetic analysis and scaling relations for catalyst design,* Annual Review of Chemical and Biomolecular Engineering **9**, 413 (2018).

[31] D. J. Klinke II and L. J. Broadbelt, *Construction of a mechanistic model of Fischer–Tropsch synthesis on Ni(1 1 1) and Co(0 0 0 1) surfaces,* Chemical Engineering Science **54**, 3379 (1999).

[32] S. Vernuccio, E. E. Bickel, R. Gounder, and L. J. Broadbelt, *Microkinetic model of propylene oligomerization on Brønsted acidic zeolites at low conversion,* ACS Catalysis **9**, 8996 (2019).

[33] L. H. Oakley, F. Casadio, K. R. Shull, and L. J. Broadbelt, *Microkinetic modeling of the autoxidative curing of an alkyd and oil-based paint model system,* Applied Physics A **121**, 869 (2015).

[34] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts,* Industrial & Engineering Chemistry Research **47**, 5879 (2008).

[35] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Fischer-tropsch synthesis: Development of a microkinetic model for metal catalysis,* Oil & Gas Science and Technology - Revue de l'IFP **61**, 489 (2006).

[36] J. Thybaut and G. Marin, *Single-event MicroKinetics: Catalyst design for complex reaction networks,* Journal of Catalysis **308**, 352 (2013).

[37] A. V. Zeigarnik and E. Shustorovich, *The UBI-QEP method: Mechanistic and kinetic studies of heterogeneous catalytic reactions,* Russian Journal of Physical Chemistry B **1**, 330 (2007).

[38] E. M. Shustorovich and A. V. Zeigarnik, *The UBI-QEP method: Basic formalism and*

*applications to chemisorption phenomena on transition metal surfaces. chemisorption energetics,* Russian Journal of Physical Chemistry **80**, 4 (2006).

[39] G. Lozano-Blanco, K. Surla, J. Thybaut, and 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**, 423 (2010).

[40] J. A. Dumesic, D. F. Rudd, D. F. Rudd, L. M. Aparicio, J. E. Rekoske, and A. A. Trevino, *The microkinetics of heterogeneous catalysis* (Wiley-Vch, 1993).

[41] P. Linstorm, *Nist chemistry webbook, nist standard reference database number 69,* J. Phys. Chem. Ref. Data, Monograph **9**, 1 (1998).

[42] N. Cohen, *Revised Group Additivity Values for Enthalpies of Formation (at 298 K) of Carbon–Hydrogen and Carbon–Hydrogen–Oxygen Compounds,* Journal of Physical and Chemical Reference Data **25**, 1411 (1996).

[43] P. W. Atkins and J. De Paula, *Physical chemistry,* (1998).

[44] Y. Yan, T. Mattisson, P. Moldenhauer, E. J. Anthony, and P. T. Clough, *Applying machine learning algorithms in estimating the performance of heterogeneous, multicomponent materials as oxygen carriers for chemical-looping processes,* Chemical Engineering Journal **387**, 124072 (2020).

[45] C. I. Méndez and J. Ancheyta, *Modeling and control of a Fischer-Tropsch synthesis fixed-bed reactor with a novel mechanistic kinetic approach,* Chemical Engineering Journal **390**, 124489 (2020).

[46] P. J. Donaubauer, D. M. Melzer, K. Wanninger, G. Mestl, M. Sanchez-Sanchez, J. A. Lercher, and O. Hinrichsen, *Intrinsic kinetic model for oxidative dehydrogenation of ethane over MoVTeNb mixed metal oxides: A mechanistic approach,* Chemical Engineering Journal **383**, 123195 (2020).

[47] C. Casado, J. Moreno-SanSegundo, I. D. la Obra, B. E. García, J. A. S. Pérez, and J. Marugán, *Mechanistic modelling of wastewater disinfection by the photo-fenton process at circumneutral pH,* Chemical Engineering Journal **403**, 126335 (2021).

[48] M. R. Dobbelaere, P. P. Plehiers, R. V. de Vijver, C. V. Stevens, and K. M. V. Geem, *Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats,* Engineering **7**, 1201 (2021).

[49] G. Hinton and T. J. Sejnowski, *Unsupervised learning: foundations of neural computation* (MIT press, 1999).

[50] H.-H. Bock, *Clustering methods: A history of k-means algorithms,* in *Selected Contributions in Data Analysis and Classification* (Springer Berlin Heidelberg, 2007) pp. 161–172.

[51] C. A. M. López, S. Bhonsale, K. Peeters, and J. F. M. V. Impe, *Manifold learning and clustering for automated phase identification and alignment in data driven modeling of batch processes,* Frontiers in Chemical Engineering **2** (2020), 10.3389/fceng.2020.582126.

[52] M. Rovira, K. Engvall, and C. Duwig, *Identifying key features in reactive flows: A tutorial on combining dimensionality reduction, unsupervised clustering, and feature correlation,* Chemical Engineering Journal **438**, 135250 (2022).

[53] N. A. Abd Majid, B. R. Young, M. P. Taylor, and J. J. Chen, *K-means clustering pre-analysis for fault diagnosis in an aluminium smelting process,* in *2012 4th Conference on Data Mining and Optimization (DMO)* (IEEE, 2012) pp. 43–46.

[54] R. Palkovits and S. Palkovits, *Using artificial intelligence to forecast water oxidation catalysts,* ACS Catalysis **9**, 8383 (2019).

[55] X. Zhao, X. Li, H. Lu, H. Yue, C. Liu, S. Zhong, K. Ma, S. Tang, and 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,* Chemical Engineering Journal **422**, 130389 (2021).

[56] Q. Dai, L. Xie, Z. Guo, L. Ma, J. Yang, N. Ren, G. Tian, and P. Ning, *Research on the mechanism of synergistic-dehydration/detoxification for sludge under treatment with double-modified potato residue,* Chemical Engineering Journal **420**, 127699 (2021).

[57] S. Bansal, S. Roy, and F. Larachi, *Support vector regression models for trickle bed reactors,* Chemical Engineering Journal **207-208**, 822 (2012).

[58] A. Nandy, C. Duan, J. P. Janet, S. Gugler, and H. J. Kulik, *Strategies and software for machine learning accelerated discovery in transition metal chemistry,* Industrial & Engineering Chemistry Research **57**, 13973 (2018).

[59] D. S. Palmer, N. M. O'Boyle, R. C. Glen, and J. B. O. Mitchell, *Random forest models to predict aqueous solubility,* Journal of Chemical Information and Modeling **47**, 150 (2006).

[60] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, and J. B. O. Mitchell, *Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization,* Journal of Chemical Information and Modeling **46**, 2412 (2006).

[61] N. S. Kaveh, F. Mohammadi, and 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**, 161 (2009).

[62] F. Kartal and U. Özveren, *A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus,* Energy (2020), 10.1016/j.energy.2020.118457.

[63] 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, and T. Tani-

ike, *The rise of catalyst informatics: Towards catalyst genomics,* ChemCatChem **11**, 1146 (2019).

[64] R. Tibshirani, *Regression shrinkage and selection via the lasso,* Journal of the Royal Statistical Society: Series B (Methodological) **58**, 267 (1996).

[65] Y. Song, J. Liang, J. Lu, and X. Zhao, *An efficient instance selection algorithm for k nearest neighbor regression,* Neurocomputing **251**, 26 (2017).

[66] G. Smits and 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).

[67] B. Boehmke and B. M. Greenwell, *Hands-on machine learning with R* (CRC Press, 2019).

[68] W. S. Sarle, *Neural networks and statistical models,* (1994).

[69] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, *Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions,* IEEE Access **7**, 137184 (2019).

[70] L. Breiman, *Random forests,* Machine learning **45**, 5 (2001).

[71] L. S. Shapley, *Stochastic Games,* Proceedings of the National Academy of Sciences **39**, 1095 (1953).

[72] S. Lundberg and S.-I. Lee, *An unexpected unity among methods for interpreting model predictions,* arXiv , 1 (2016), arXiv:1611.07478 .

[73] S. M. Lundberg and S. I. Lee, *A unified approach to interpreting model predictions,* Advances in Neural Information Processing Systems **2017-December**, 4766 (2017), arXiv:1705.07874 .

# 2

# INFLUENCE OF CATALYST PROPERTIES ON LIGHT OLEFIN PRODUCTION

*The first goal of this thesis is to identify a potential catalyst with enhanced light olefin selectivity. To achieve this, we extend the capabilities of the SEMK modelling approach to investigate the influence of variation in catalyst properties i.e. catalyst descriptors, on the yield of desired component, light olefins ($C_2 - C_4 =$). We explore the catalyst descriptor space around three literature-reported Bi and Pb promoted and non-promoted Fe catalysts. These are used as a benchmark to explore the parametric catalyst descriptor space.*

*The three catalyst descriptors, i.e. atomic chemisorption enthalpies of hydrogen ($Q_H$),carbon ($Q_C$), and oxygen ($Q_O$) in the SEMK modelling approach have a combined effect on the conversion, whereas the selectivity to light olefins is found to be less sensitive to $Q_O$. These effects can be rationalized in terms of relative surface coverages of different species, leading to different dominant reaction pathways, and thus resulting in product yield variations. Using this approach, a "promising catalyst" with catalyst descriptors, $Q_H \approx 234$ kJ/mol, $Q_C \approx 622$ kJ/mol and $Q_O \approx 575$ kJ/mol resulting in 55% light olefins yield with lower methanation and long-chain hydrocarbon formation, is identified* [1] .

[1] This work is published in Chemical Engineering Journal with doi: https://doi.org/10.1016/j.cej.2021.129633

## 2.1. INTRODUCTION



Fischer-Tropsch synthesis (FTS) is one of the most promising approaches to obtain non-petroleum-based hydrocarbons. While, most of the current plants rely on coal [1] or natural gas [2, 3] to produce syngas, the feedstock for FTS, various chemical recycling techniques can be used as an alternative to produce syngas with non-conventional $H_2$/CO (e.g. from plastic waste via gasification) [4]. With the current urge for circular economy [5, 6] and, hence, the need to reduce plastic waste and associated pollution, save primary resources and preserve our natural ecosystem, the use of syngas with non-conventional $H_2$/CO has gained popularity. Technological developments with emphasis on better catalysts can help to increase the competitiveness of the FTS process [7] to produce value-added chemicals such as light olefins from plastic waste. The light olefins so produced could in turn be used for plastic production [8], thus leading to a true circular economy. Therefore, high-temperature FTS has been of interest in the last decades due to the opportunity to convert plastics and organic waste into value-added chemicals such as light olefins rather than other hydrocarbon products such as methane, paraffins, and other long-chain hydrocarbons [9, 10]. Studies also concentrate on the influence of different catalysts in enhancing the activity and selectivity. A comprehensive review of early catalyst development is reported by Vannice et al. [11] and Anderson et al. [12].

At present, the FTS process mostly uses iron-based (low- and high- temperature operation modes) and cobalt-based (low-temperature operation mode) catalysts. Both cobalt [13–15] and iron [16, 17] catalysts yield higher hydrocarbons. Higher activity, light olefin selectivity, and lower costs make iron-based catalysts a better candidate in FTS even though they also lead to proportionally more oxygenates. Due to their potential water gas shift (WGS) activity, iron-based catalysts find merit in processing syngas with varying H2/CO ratio, more particularly hydrogen deficient syngas [12, 18, 19].

The investigations on iron-based catalysts have focused mainly on operating conditions such as temperature, gas flow rate, pressure, gas type, as well as on the promoters.

The improvement in catalyst performance with addition of promoters in iron-based catalysts [16, 17] has attracted recent research interests. The study of the promoter effect is rather complex as it is done with different supports and at different reaction conditions. Recent studies investigate the effect of iron-promoted catalysts [9, 10] on performance. Gu et.al [9] report a higher light olefin selectivity, specifically on promotion by Bi and Pb on carbon nanotube support, over unpromoted Fe catalyst.

In order to provide fundamental insights on catalyst development, kinetic models can be recurred to account for properties in the modelling procedure [20]. Microkinetic models, which represent the occuring chemistry at the elementary step level [21–23] help to study the reaction kinetics of complex reactions such as Fischer Tropsch reaction [24], oligomerization [25], autoxidative curing [26], etc. To model reaction systems with a large number of elementary steps using a microkinetic approach, the single event concept can be employed. In the Single-Event MicroKinetic (SEMK) methodology the reactive moieties are considered to determine the reactivity of individual molecules. Accordingly, the elementary reactions are classified into reaction families to reduce the number of parameters [27]. The single-event kinetic coefficient is unique for a reaction family and the adsorption enthalpies of surface species are calculated via the UBI-QEP method [28, 29]. In the SEMK framework, the catalysts are differentiated in terms of model parameters referred to as catalyst descriptors. Single-Event MicroKinetic models have been developed for alkylation [30], hydrocracking [31], catalytic cracking [32] and reforming [33]. Specifically for FTS, studies using the SEMK modelling approach concentrate mostly on identifying the descriptors values corresponding to the catalysts used in a particular experimental investigation [34–39]. However less attention was dedicated to explain how a change in the catalyst properties and, hence, in the descriptor values, would influence the performances. Screening of the descriptor space could help us in identifying the surface reactions that mostly impact the selectivity towards the desired products.

The scope of the current chapter is, hence, to identify the catalyst descriptor values that enhance the yield of light olefins, within a broad, yet realistic descriptor space. For our study, the descriptor space is identified by taking into account the experimental performances of three iron-based catalysts. We carry out SEMK simulations using different combinations of descriptors, also denoted as virtual catalysts [40]. The SEMK simulations with virtual catalysts aim at assessing how the model explains the difference in the behaviour of catalysts in terms of the microkinetic phenomena and identifying a "promising catalyst" within the descriptor space.

## 2.2. SINGLE EVENT MICROKINETIC MODELLING APPROACH

The SEMK methodology as applied to FTS of iron based catalysts is briefly explained. The elementary steps and reaction families considered in the SEMK reaction network, are summarized in Table 5.1. The forward activation energies listed in Table 5.1 constitute the kinetic descriptors of the model, which are independent of the catalyst descriptors. The forward activation energies reported in the table, along with the standard surface reaction enthalpies, are used to calculate the reverse activation energies for the elementary steps by applying the principle of microscopic reversibility. The surface reaction enthalpies are calculated starting from those of the corresponding gas phase reactions and by accounting for the enthalpy change induced by chemisorption [37]. Chemisorption enthalpies of various surface species are calculated as a function the atomic chemisorption enthalpies using the UBI-QEP method [29, 41]. The single-event forward pre-exponential factors reported in Table 5.1 are determined via statistical thermodynamics calculations. The calculation involves gas phase entropies of the components and the single-event standard entropy change related to chemisorption [42]. The gas phase entropies are obtained either from databases [43] or using group additivity methods [44]. The calculation of the entropy change related to the chemisorption step is based on the loss of translational entropy of a gas phase component, which is inturn

calculated with the Sackur-Tetrode equation [45]. The reverse pre-exponential factors are calculated using the forward pre-exponential factors ($\tilde{A}^{for}$) and single event surface reaction entropies ($\Delta\tilde{S}_r^0$) as:



$$\tilde{A_j}^{rev} = \frac{\tilde{A_j}^{for}}{exp\left(\frac{\Delta\tilde{S}^0_{r,j}}{R}\right)} \tag{2.1}$$

The pre-exponential factors are used for determining the rate coefficients of the reaction families in Table 5.1. All the above calculations are explained in detail by Lozano-Blanco et al. [34, 36].

In the SEMK model, the considered catalyst and corresponding reaction mechanism are quantified in terms of both catalyst and kinetic [27] descriptors. Kinetic descriptors are model parameters that are specific to the reaction families considered and are independent of the catalyst properties. While catalyst descriptors are model parameters that specifically account for the impact of the catalyst on the kinetics. In the case of Fischer-Tropsch synthesis, atomic chemisorption enthalpies of carbon ($Q_C$ in carbide phase), hydrogen ($Q_H$ and $Q_Fe_3O_4-H$ corresponding to carbide and oxide phase respectively), and oxygen ($Q_O$ in carbide phase) are the catalyst descriptors [36] [2]. The atomic chemisorption enthalpies are reported in absolute value, with an increase corresponding to more pronounced exothermicity. The atomic chemisorption enthalpies of the iron catalyst discussed by Lozano-Blanco et al. [36] i.e. $Q_C$, $Q_H$, $Q_Fe_3O_4-H$ and $Q_O$ are 630.03 ± 2.47 kJ/mol, 252.40 ± 0.63 kJ/mol, 233.83 ± 1.28 kJ/mol and 579.19 ± 1.77 kJ/mol, respectively. In our discussion below, a "virtual catalyst" corresponds with a specific combinations of these catalyst descriptors, while the ones actually synthesised

[2] It is however to be noted that the number of catalyst descriptors potentially depends on the specificities of the reaction. For e.g. in case of $CO_2$ hydrogenation where we expect the reverse water gas shift reaction to be dominant and the oxide phase of iron catalyst will play a dominant role in this respect. This will be accounted for in the model through the parameter $Q_Fe_3O_4-H$. When $CO_2$ hydrogenation would be carried out using a cobalt catalyst, there will be minimal WGS reaction. The values of $Q_H$ are then expected to be higher than in case of iron to have good activity and those of the other catalyst descriptors to be low. Methane is expected to be the dominant product.

Table 2.1: Elementary reactions and reaction families in the reaction network [36], where $E_a^{for}$ represent the kinetic descriptors, M represents the metal surface, and $\tilde{A}^{for}$, represent the forward pre-exponential factors, and $\Delta\tilde{S}_r^0$ represent single-event surface reaction entropies, respectively. The surface reaction enthalpies is a function of catalyst descriptors. The calculations of enthalpies and entropies are reported by Lozano-Blanco et al. [34, 36].

| ***Reaction family/ elementary reaction*** | $E_a^{for}$ **(kJ/mol)** | $\tilde{A}^{for}$ ($s^{-1}$ **or** $(MPas)^{-1}$) | $\Delta\tilde{S}_r^0$ ($J(Kmol)^{-1}$) |
|---|---|---|---|
| 1. $H_2$+ 2M $\rightleftharpoons$ 2MH | 0 | $3.1 \times 10^8$ | −63.5 |
| 2. CO + 2M $\rightleftharpoons$ MMCO | 0 | $2.2 \times 10^7$ | −165.7 |
| 3. MMCO + 3M $\rightleftharpoons$ MMMC + MMO | 56.81 ± 0.53 | $1.3 \times 10^{13}$ | −17.34 |
| 4. MMMC + MH $\rightleftharpoons$ MMMCH + M | 77.66 ± 0.70 | $8.8 \times 10^{14}$ | +35.08 |
| 5. MMMCH + MH $\rightleftharpoons$ $MMCH_2$+ 2M | 11.94 ± 0.10 | $5.9 \times 10^{11}$ | −25.6 |
| 6. $MMCH_2$ + MH $\rightleftharpoons$ $MCH_3$ + 2M | 61.88 ± 0.50 | $2.2 \times 10^{11}$ | −33.5 |
| 7. $MC_nH_{2n+1}$ + $MMCH_2$ $\rightleftharpoons$ $MC_{n+1}H_{2n+3}$ + 2M | 44.79 ± 0.43 | $8.3 \times 10^9$ | −60.69 |
| 8. $MC_nH_{2n+1}$ + MH $\rightleftharpoons$ $C_nH_{2n+2}$ + 2M | 117.75 ± 0.67 | $3.3 \times 10^{10}$ (n=2)<br>$2.0 \times 10^{10}$ (n=3-10) | +64.3 |
| 9. $MC_nH_{2n+1}$ + M $\rightleftharpoons$ $MC_nH_{2n}$ + MH | 96.27 ± 0.50 | $1.0 \times 10^{10}$ | +12.32 |
| 10. $MC_nH_{2n}$ $\rightleftharpoons$ $C_nH_{2n}$ + M | 63(n=2)<br>61(n=3-10) | $1.3 \times 10^{13}$ | +118.6 |
| 11. O − CHO − M + M − OH +O $\rightleftharpoons$ O − COOH − M + O − H + M | 138.95 ± 1.15 | $1.7 \times 10^{14}$ | +53.2 |
| 12. MMO + MH $\rightleftharpoons$ MOH + 2M | 103.80 ± 0.96 | $1.3 \times 10^{12}$ | +33.92 |
| 13. MOH + MH $\rightleftharpoons$ H2O + 2M | 86.22 ± 0.62 | $2.4 \times 10^{11}$ | +73.7 |

and experimentally tested are referred to as real catalysts.



## 2.3. METHODOLOGY

The current study on FTS using SEMK simulations aims at identifying a "promising catalyst" (catalyst descriptors) that leads to enhanced yield of light olefins. The kinetic descriptor values listed in Table 5.1 based on previous work [36]. The catalyst descriptors are identified as discussed later in this section. The experimental results reported by Gu et al.[9] are used to determine the catalyst descriptor values corresponding to Bi and Pb promoted and non-promoted Fe catalysts. These are used as a benchmark to explore the parametric catalyst descriptor space. The comparison between the numerical simulation results and the experimental data serves to identify the relevant part of the space within which the real catalysts are situated. A "promising catalyst" that leads to enhanced yield of light olefins, is then identified within the catalyst descriptor space enclosing the three literature-reported iron-based catalysts. Thus our methodology is split into two subsections.

### 2.3.1. IDENTIFICATION OF REALISTIC DESCRIPTOR SPACE

To identify and analyse the catalyst descriptors space corresponding to Fe/CNT, FeBi/CNT, and FePb/CNT catalysts used by Gu et al.[9], we follow the steps discussed below:

- **Step 1: Generation of virtual catalyst library:**

  A library with a large number of virtual catalysts is generated by varying the catalyst descriptors (Fig.2.1). The investigated range of catalyst descriptors is chosen around the descriptor values for an Fe catalyst, as previously investigated [36]. To generate a diverse virtual catalyst library, the virtual space is sampled via an experimental design. In order to account for this deterministic system with various factors (descriptors), we make use of space-filling design [3] [40, 46]. The combina-

[3]The space-filling design was used to ensure that within the randomness that is applied, all relevant combina-

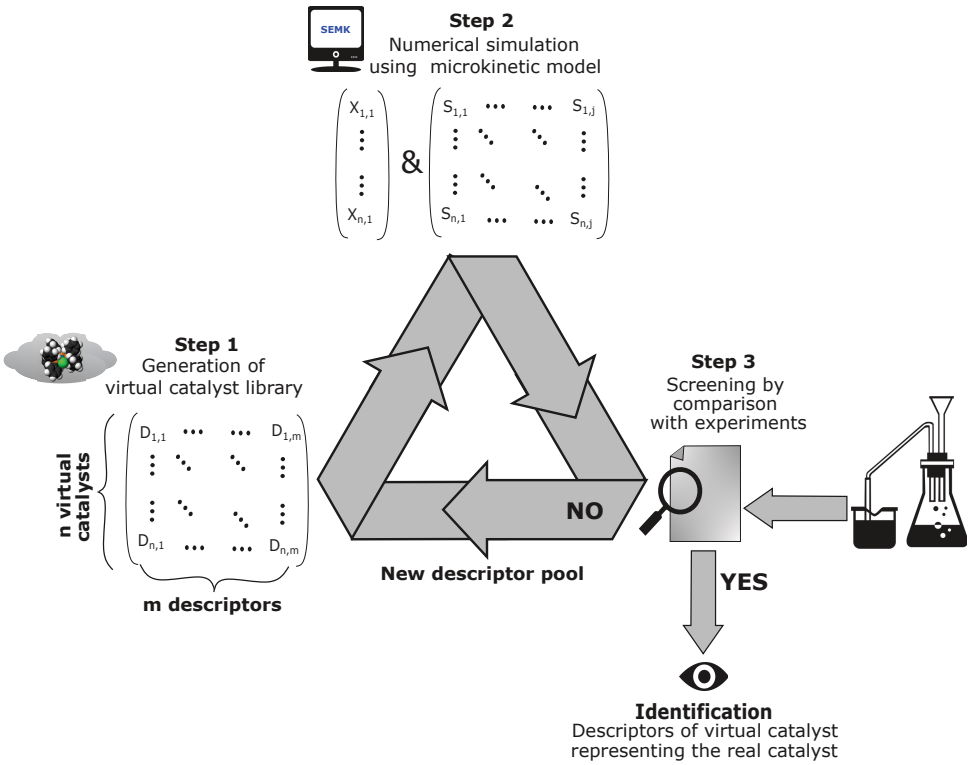



Figure 2.1: Identification of descriptors in SEMK simulations that explain the performance behaviour of real experimental catalysts. SEMK simulations are carried out at different combinations of descriptors and the resulting performance is compared with experimental results. The descriptor combinations explaining the performance obtained with experiments using different catalysts are identified.

tion of descriptors generated by this procedure forms a virtual catalyst library. The virtual catalysts are used to carry out numerical simulations.

- **Step 2: Numerical simulations with a virtual catalyst library:**

  In the second step of the proposed methodology numerical simulations are performed with FTS Single-Event MicroKinetic model incorporated in a plug flow reactor model for all virtual catalysts generated in **Step 1**. The set of ordinary differential equations (mass balances for the molecules) and nonlinear algebraic equations (pseudo-steady state approximation for the surface species) in the reactor

tions that are effectively tried out in the analysis. It guarantee an optimal coverage of the experimental space. These designs allow for a spread of points encompassing the entire design space leading to more accurate predictions at out-of-sample testing locations

model is solved with the DASPK [47] from the Netlib [48] software library as discussed by Lozano-Blanco et al. [36]. In order to ensure convergence, the numerical subroutine DNSQE within Netlib library is used to initialize the variables associated with the algebraic equations (solved using DASPK). Here, DASPK is used with variable-stepsize backward differentiation formulas applying a direct linear method.

The simulations are performed at the following operating conditions 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g, and $H_2$/CO molar inlet ratio of 1, as reported in [9].

- **Step 3: Comparison and screening to identify realistic catalyst descriptors:** From the simulation results, we find a pool of virtual catalysts with different combinations of catalyst descriptors resulting in comparable conversion ($\Delta < 5\%$) and light olefin selectivity ($\Delta < 10\%$) with respect to the experimentally observed ones. This reduces the number of virtual catalyst candidates by $\approx 90\%$. This is followed by a screening of the selectivities towards other components namely, $CO_2$, $CH_4$, and long-chain hydrocarbons (on a carbon-dioxide-free basis), to identify the range of realistic catalyst descriptors.This screening process in the order mentioned above leaves us with less than 5%, 2% and 0.5% virtual catalysts initially generated, respectively. We identify the range of catalyst descriptors that represent the above 0.5% of virtual catalysts. More virtual catalysts are generated in this confined range, and the **Step 1-Step 3** are carried out iteratively, until there is no difference between consecutive iterations. The catalyst descriptors which result in a match of conversion and selectivity between experiments and simulations are thus identified. The process discussed above is carried out to identify the catalyst descriptors of the three real catalysts, Fe/CNT, FePb/CNT, and FeBi/CNT.

### 2.3.2. IDENTIFICATION OF A "PROMISING CATALYST"



To identify a "promising catalyst" with maximum light olefins yield, the descriptor space around the catalyst descriptors of the three literature-reported real catalysts is analysed. The descriptor space is analysed using 2D parametric planes and 3D iso-surfaces of conversion, selectivity, and yield. We also compare reaction pathways as simulated for various virtual catalyst to understand how the corresponding descriptor values determine the selectivities/yields. Reaction pathway analysis (RPA) serves as an important tool to analyze the occurring phenomena, e.g., elementary steps, both in a qualitative and quantitative manner, and identify the prevailing reaction routes [39]. here, the reaction pathway analysis is performed by determining the affinities of elementary reactions.

For an elementary reaction, $aA + bB \rightleftharpoons cC + dD$, we define affinity of the elementary reaction as,

$$A_i = RT \times ln\left(K_{eq}\frac{A^a B^b}{C^c D^d}\right), \tag{2.2}$$

where, R, T and $K_{eq}$ are universal gas constant, temperature and thermodynamic equilibrium coefficient, respectively. If the affinity is close to zero, the elementary reaction occurs very fast and can be assumed to be in quasi-equilibrium. In the reaction pathway analysis carried out, an elementary reaction is considered to be in quasi-equilibrium when the affinity, $|A_i| < 1000$ Jmol$^{-1}$ [39]. Kinetically relevant reactions will have an affinity exceeding this value. The corresponding reaction rate will then allow determining what fraction of the reactant is consumed via this elementary reaction. In the current study, the reactions in quasi-equilibrium are represented with solid black arrows, while the thickness of the kinetically relevant steps (colored arrows) are scaled logarithmically. Thickness of the colored arrows are chosen such that, $Thickness,\ t = k \times |logx|^{-1}$, where, $x$= forward or backward reaction rate and $k$= multiplication factor. The value of $k$ is chosen to be 11.5 for discrimination of arrows. Each reaction family is given a separate

color.

From our analysis above, we then identify a "promising catalyst" with maximum light olefins yield, within the investigated descriptor space.



## 2.4. RESULTS AND DISCUSSION

### 2.4.1. IDENTIFICATION OF REALISTIC CATALYST DESCRIPTORS

To find the catalyst descriptor range corresponding to the catalysts Fe/CNT, FePb/CNT, and FeBi/CNT reported in [9], we compare the simulated performance of virtual catalysts with the experimental ones. In Fig.2.2(a) each blue dot corresponds to the conversion and light olefin selectivity obtained with a virtual catalyst. As seen from the scatter plot (Fig.2.2(a)), different virtual catalysts, result in a wide conversion and selectivity range. However, only a few virtual catalysts result in conversion and light olefins selectivity comparable to that with the real catalysts (experiments) namely, Fe/CNT (brown), FePb/CNT (green) and FeBi/CNT (red).

Table 2.2: Catalyst descriptors, i.e atomic chemisorption enthalpies for Fe, FeBi/CNT and FePb/CNT catalysts.

| **Atomic Chemisorption enthalpies** | **Fe/CNT (kJ/mol)** | **FeBi/CNT (kJ/mol)** | **FePb/CNT (kJ/mol)** |
|---|---|---|---|
| $Q_H$ ($Fe_xC-H$) | 249.5 | 247.7 | 248.4 |
| $Q_C$ ($Fe_xC-C$) | 644.1 | 632.1 | 641.5 |
| $Q_O$ ($Fe_xC-O$) | 601.0 | 589.1 | 577.1 |

As discussed in **Step 3** of our methodology, the catalyst descriptors which best reproduced the experimental performances (see Fig.2.2(b), Fig.2.2(c)-(e)) are identified for both non-promoted and promoted catalysts. The descriptor values for the three real catalysts are reported in Table 2.2. The atomic chemisorption enthalpy of oxygen (Table 2.2) is lower for the promoted Fe based catalysts compared to the non-promoted Fe catalyst. It leads to an enhanced CO dissociation and, hence, higher conversion, in line with the

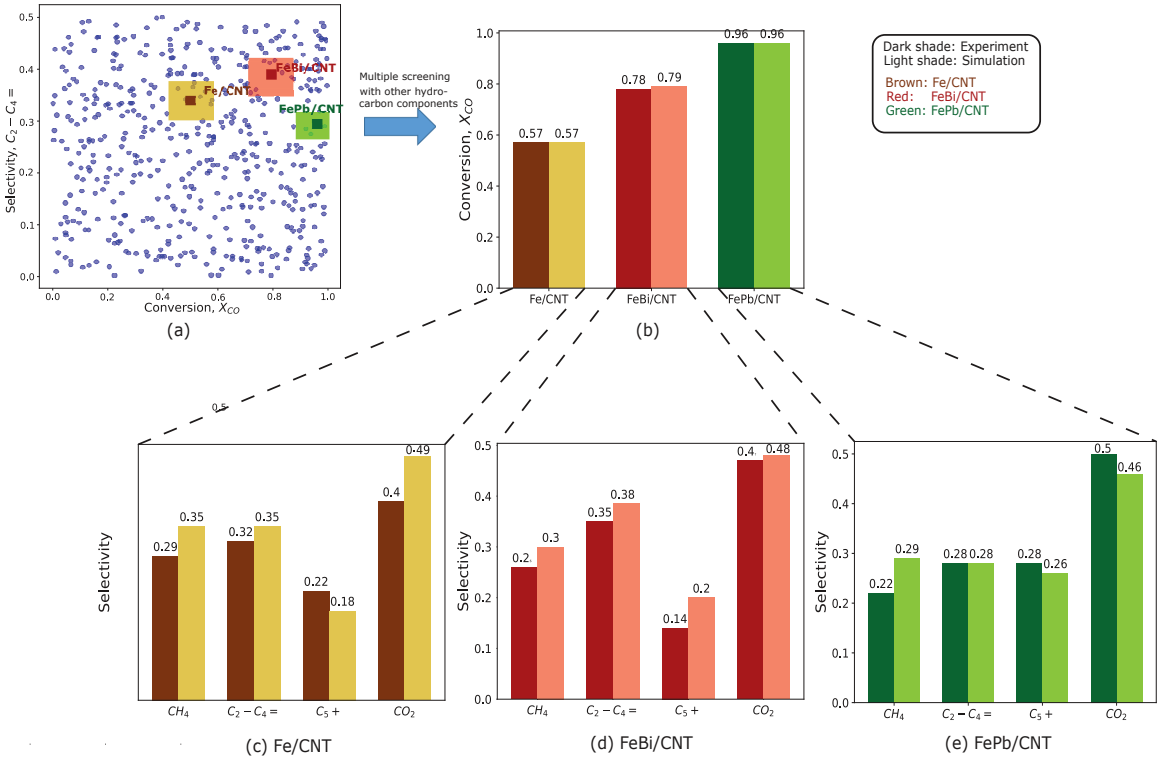



Figure 2.2: (a) CO Conversion vs light olefin selectivity ($C_2 - C_4 =$) obtained with different virtual catalysts (blue dots). The experimental results obtained with real catalysts Fe/CNT, FePb/CNT and FeBi/CNT are represented with brown, green and red square dots respectively. The light coloured boxes around real catalysts represent the virtual catalysts with comparable conversion and light olefin selectivity. The screening of the virtual catalysts within the light coloured boxes are carried out to obtain the virtual catalyst matching real catalyst. (b) Conversion obtained with experiments (real catalyst) are compared with simulated conversion obtained with the best matching virtual catalyst after screening. Selectivity of $CH_4$ (Methane), $C_2 - C_4 =$ (light olefin), $C_5+$ (long-chain hydrocarbons) and $CO_2$ (Carbon dioxide) with (c) Fe/CNT catalyst (conversion 57%), (d) FeBi/CNT catalyst(conversion 79%), (e) FePb/CNT catalyst (conversion 96%)obtained with experiments are matched with that of simulations using the best matching virtual catalyst. The experiments and simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2/CO$ molar inlet ratio of 1. The values in the figure are reported in a scale between 0 and 1.

experimental observations. This has been denoted in the literature as an increase in the scavenging of oxygen atoms [9]. It is ensured from Fig.2.2(c)-(e) that the simulated selectivity towards the other hydrocarbons and $CO_2$ (Fig.2.2(c)-(e)) using the catalyst descriptor values in Table 2.2 is also comparable with experimental results of Gu et.al. [9].

From Table 2.2 we could see that, in addition to the difference in atomic chemisorption enthalpy of oxygen$Q_O$, the non-promoted and two promoted catalysts also exhibit differences in atomic chemisorption enthalpies of hydrogen $Q_H$ and carbon $Q_C$. It was observed in some preliminary simulations that the atomic chemisorption enthalpy of hydrogen in oxide phase ($Fe_3O_4 - H = 220$ kJ/mol), affecting the WGS reaction did not lead to major differences in the targeting results. Hence, this value was fixed throughout the procedure. As expected, from the above discussion, we see that the catalyst performance is well captured by the catalyst descriptors without the adjustment of any of the kinetic descriptors discussed in Table 5.1. To assess the effect of the differences in catalyst properties on their performance, we thus examine the catalyst descriptor space of $Q_H$, $Q_C$, and $Q_O$.

### 2.4.2. ANALYSIS OF CATALYST PERFORMANCE

In this section we analyse the catalyst performance as a function of catalyst descriptors to understand their influence on conversion and light olefins selectivity.

#### CONVERSION

To analyse the catalyst descriptor effect on the CO conversion, we investigate $Q_H - Q_C$ descriptor planes at three different $Q_O$ (Fig.2.3). The planes are considered at atomic chemisorption enthalpies of oxygen, $Q_O = (601, 589.1, 577.1)$ kJ/mol, corresponding to the three real catalysts (Fe/CNT, FeBi/CNT, and FePb/CNT) reported in Table 2.2. The ranges of $Q_H$ and $Q_C$ are chosen such that the plane encloses the three real catalysts, Fe/CNT, FeBi/CNT, and FePb/CNT. From Fig.2.3 (a)-(c), we see a combined influence of the catalyst descriptors on the simulated conversions. In the descriptor planes (Fig.2.3

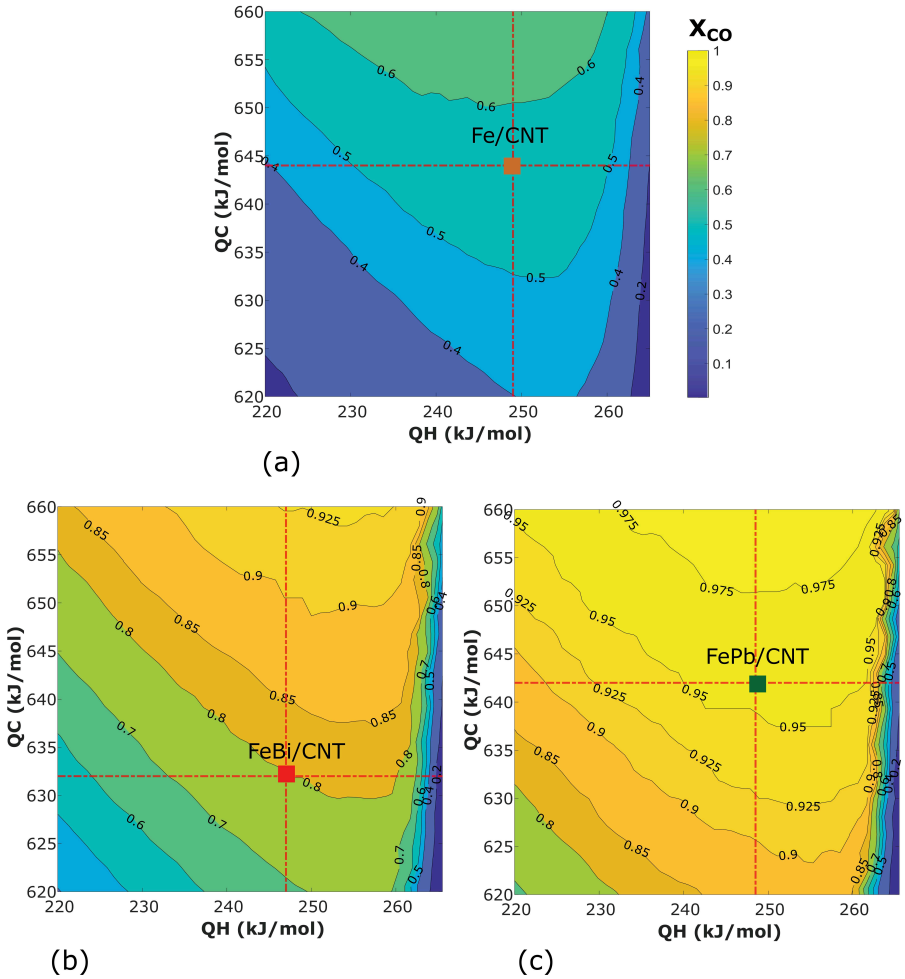



Figure 2.3: Conversion as a function of atomic chemisorption enthalpy of hydrogen ($Q_H$) and carbon ($Q_C$) for atomic chemisorption enthalpy of oxygen ($Q_O$) at 601 kJ/mol (a), 589.1 kJ/mol (b) and 577.1 kJ/mol (c). The intersection of dotted red lines corresponds to Fe/CNT (a), FeBi/CNT(b) FePB/CNT (c) respectively. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2$/CO molar inlet ratio of 1. The values in the figure are reported in a scale between 0 and 1.

(a)-(c)), the intersection of dotted lines (red) indicate the location of the real catalysts, namely Fe/CNT, FeBi/CNT, FePb/CNT. As we traverse along the horizontal dotted line with fixed atomic chemisorption enthalpies of carbon and oxygen, respectively we observe an increase in conversion followed by a decrease. As expected, this observation follows Sabatier's principle also referred to as a volcano-curve [39]. It is also seen from

the parametric plots of conversion (Fig.2.3) that the effect of the atomic chemisorption enthalpy of hydrogen ($Q_H$) is more pronounced (especially at $Q_H > 260$ kJ/mol) than that of the atomic chemisorption enthalpy of carbon ($Q_C$). A too high value of $Q_H$ results in higher hydrogen surface coverages, which causes the coverage of CO to decrease and, hence, results in a decrease in CO conversion, while a too low value of $Q_H$ leads to lower surface coverage of hydrogen thus adversely affecting the initiation and chain growth reactions discussed in Table 5.1. On the other hand, too high values of atomic chemisorption enthalpy of carbon ($Q_C > 660$ kJ/mol, not shown in Fig.2.3) lead to lower availability of surface hydrogen (due to higher CO coverage), whereas, too low values lead to lower availability of surface carbon. Both these situations result in a decrease in conversion. However, the conversion is less sensitive to a change in $Q_C$ as compared to $Q_H$. Along different $Q_H - Q_C$ planes (Fig.2.3 (a)-(c)) with the decrease in atomic chemisorption enthalpy of oxygen, we obtain a higher conversion for a fixed $Q_H$ and $Q_C$. Thus, a decrease in $Q_O$ with the addition of Bi and Pb as promoters, means that oxygen is more loosely bound to the surface. This reduces the concentration of metal sites being blocked by oxygenates [4], permitting adequate coverages of surface carbon/hydrogen, in turn increasing the conversion.

### SELECTIVITY

In Fig.2.4 we look at the light olefin selectivity along the $Q_H - Q_C$ planes, for different values of atomic chemisorption enthalpy of oxygen corresponding to non promoted and promoted Fe catalysts, as discussed earlier in Section 2.4.2. As in the case of conversion, catalyst descriptors have a combined role in the selectivity values (Fig.2.4), but the effect of $Q_O$ on the light olefin selectivity is limited, i.e. Fig.2.4(a)-(c) are quite comparable. It is observed that higher selectivity for light olefins is observed with $Q_H$ in the range of 230-245 kJ and $Q_C$ in range of 625-630 kJ. However, this not necessarily ensures a higher yield of light olefins ($C_2 - C_4 =$) as the CO conversion should also be considered when

[4] For $Q_H$=248.4 kJ/mol and $Q_C$=641.5 kJ/mol, with decrease in $Q_O$ from 601 kJ/mol to 577.1 kJ/mol the free metal sites increases by a factor of 2.6.

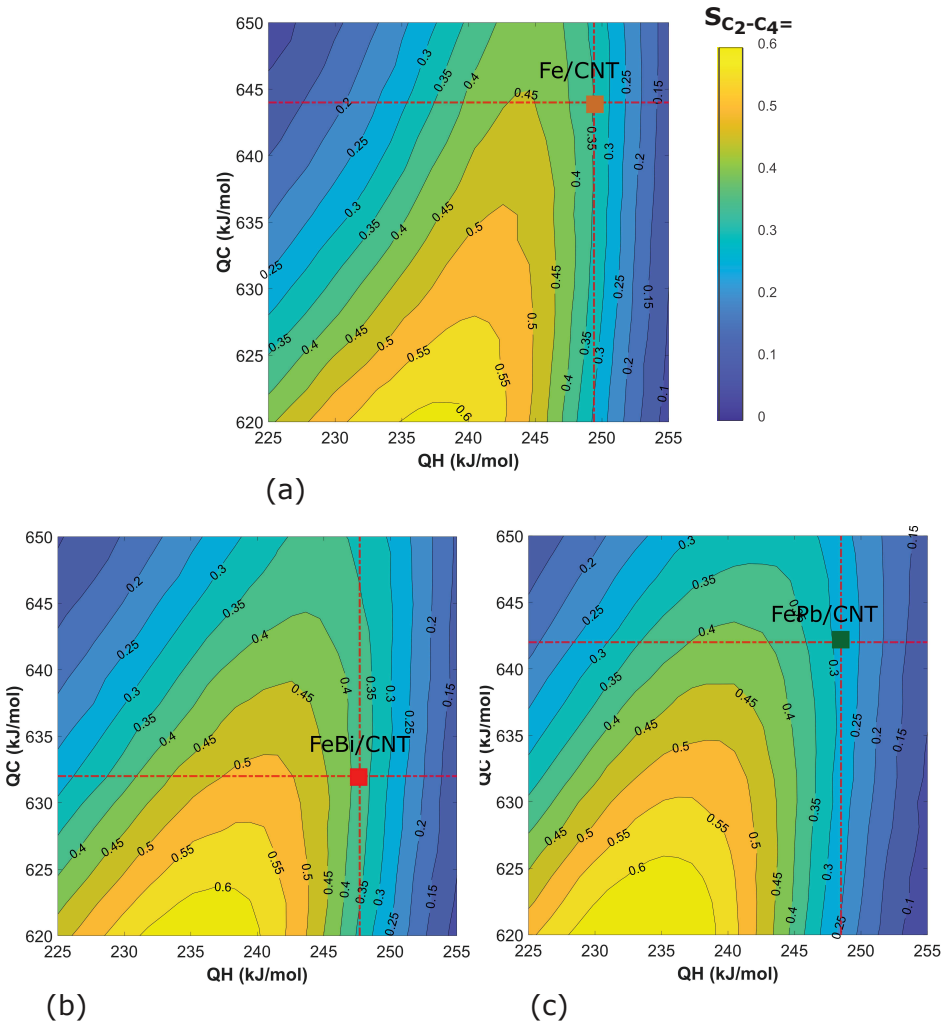



Figure 2.4: Selectivity of light olefins as a function of atomic chemisorption enthalpy of hydrogen ($Q_H$) and carbon ($Q_C$) for atomic chemisorption enthalpy of oxygen ($Q_O$) at 601 kJ/mol (a), 589.1 kJ/mol (b) and 577.1 kJ/mol (c). The intersection of dotted red lines corresponds to Fe/CNT (a), FeBi/CNT(b) and FePb/CNT (c) respectively. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2$/CO molar inlet ratio of 1. The values in the figure are reported in a scale between 0 and 1.

determining the latter. This trade-off between light olefin selectivity and CO conversion should also be taken into account when engineering promoted catalysts.

### 2.4.3. UNDERSTANDING THE DIFFERENCES IN PERFORMANCES USING SURFACE LEVEL PHENOMENA



In this section, we first compare the change in performance between selected virtual catalysts. It is then followed by discussions at an in-depth level, explaining the differences in their performance in terms of surface coverages and reaction pathways.

#### COMPARISON OF PERFORMANCE

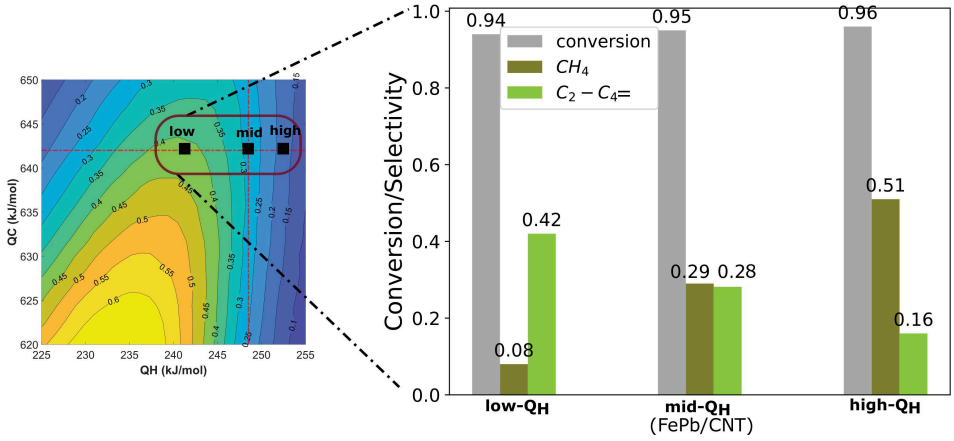



Figure 2.5: Conversion and selectivity as a function of atomic chemisorption enthalpy of hydrogen $Q_H$ for points **low-$Q_H$** ($Q_H$= 240 kJ/mol), **mid-$Q_H$** ($Q_H$= 248 kJ/mol) and **high-$Q_H$** ($Q_H$=252 kJ/mol) at a constant value of $Q_C$ = 642 kJ/mol and $Q_O$ = 577.1 kJ/mol. The contour plane to the left of the histogram corresponds to the $Q_H - Q_C$ plane of selectivity containing FePb/CNT catalyst. The FePb/CNT catalyst is represented by **mid-$Q_H$**. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2$/CO molar inlet ratio of 1. The values in the figure are reported in a scale between 0 and 1.

To better understand the impact of descriptors at iso-conversion, in Fig.2.5 we more deeply analyze the simulation results at three points in the $Q_H - Q_C$ descriptor plane containing the catalyst with higher conversion, i.e. FePb/CNT. The FePb/CNT catalyst is indicated by **mid-$Q_H$**. The points **low-$Q_H$**, **mid-$Q_H$** and **high-$Q_H$** are chosen such that the conversion obtained with the virtual catalysts represented by these three points is comparable (Fig.2.5). Thus the variation in the selectivity of these three points directly indicates the variation in light olefins yield because of the change in $Q_H$ at a constant value of $Q_C$ = 642 kJ/mol and $Q_O$ = 577.1 kJ/mol. The $Q_H$ values corresponding to points

**low-$Q_H$**, **mid-$Q_H$** and **high-$Q_H$** are 240 kJ/mol, 248 kJ/mol, and 252 kJ/mol, respectively. The selectivity of light olefins ($C_2 - C_4 =$) and the most common product of FTS reaction, methane ($CH_4$) is compared at these three points, to understand the competing effects among these two products. From the selectivity contour and histogram in Fig.2.5, for the three points considered, we observe that the light olefin selectivity is highest at **low-$Q_H$** (left) and lowest at **high-$Q_H$** (right) when compared to the selectivity with FePb/CNT catalyst (**mid-$Q_H$**). The simulations with these three set of catalyst descriptors in the $Q_H - Q_C$ plane containing FePb/CNT catalyst (Fig.2.5), shows that the lower light olefin selectivity at **high-$Q_H$** occurs at the expense of more pronounced methanation. To better understand the differences at these three points, we look at the surface coverage of different species.

### INFLUENCE OF CATALYSIS DESCRIPTORS ON SURFACE COVERAGE

Table 2.3: Relative surface coverage of species MH, $MMCH_2$ and $MCH_3$ for the three virtual points **low-$Q_H$**, **mid-$Q_H$** and **high-$Q_H$** calculated as reported in [39]. The points are located in the $Q_H - Q_C$ plane containing FePb/CNT catalyst. The atomic chemisorption enthalpies of carbon and oxygen are kept at a constant value of $Q_C$= 642 kJ/mol and $Q_O$= 577.1 kJ/mol. The points **low-$Q_H$** ($Q_H$= 240 kJ/mol), **mid-$Q_H$** ($Q_H$= 248 kJ/mol) and **high-$Q_H$** ($Q_H$= 252 kJ/mol) have different atomic chemisorption enthalpy of hydrogen. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2$/CO molar inlet ratio of 1.

| | **Relative surface coverage [-]** | | |
|---|---|---|---|
| **Surface species** | **low-$Q_H$** - | **mid-$Q_H$** - | **high-$Q_H$** - |
| MH | 0.18 | 0.54 | 1.0 |
| $MMCH_2$ | 1.0 | 0.40 | 0.21 |
| $MCH_3$ | 1.0 | 0.47 | 0.25 |

An analysis on the relation between surface coverages and catalyst performance at different operating conditions is discussed by Van Belleghem et al [39]. Here we carry out a similar analysis for different (virtual) catalysts at the same operating conditions. The difference in selectivity can be interpreted in terms of evolutions in the relative surface coverage of the dominant species (Table 2.3) among different surface species reported

in Table 5.1. The relative surface coverage is obtained by dividing the surface coverage of surface species i at a particular $Q_H$ by the maximum surface coverage of surface species i found among the points **low-$Q_H$**, **mid-$Q_H$** and **high-$Q_H$**. The maximum surface coverage of the surface species MH, $MMCH_2$ and $MCH_3$ are of order $10^{-2}$, $10^{-9}$ and $10^{-4}$ respectively. It can be observed that relative surface coverage of MH, i.e hydrogen on the metal surface follows the trend, **high-$Q_H$** > **mid-$Q_H$**> **low-$Q_H$**. Thus, the relative surface coverage of the surface species MH increases with an increase in $Q_H$. The higher MH relative surface coverage for **high-$Q_H$** results in increased hydrogenation for **high-$Q_H$** followed by **mid-$Q_H$** and **low-$Q_H$**. The surface coverages of $MMCH_2$ and $MCH_3$ show a reverse trend, i.e. **low-$Q_H$** > **mid-$Q_H$** > **high-$Q_H$**. Thus, the relative surface coverage of metal alkyls (see Table 5.1) decreases with an increase in $Q_H$. This trend can result in higher alkanes or alkenes production depending on MH surface coverage. From Table 2.4 we see that the relative selectivity of total alkenes to total alkanes is higher for **low-$Q_H$** followed by **mid-$Q_H$** and **high-$Q_H$**. Thus, alkene production is higher if MH surface coverage is lower and vice-versa. Higher surface coverage of MH leads to lower surface coverage of $MCH_2$ and lower availability of free sites for beta hydride elimination (Table 2.2, Eq.9) and thus results in increased hydrogenation [39]. It is also observed that the 1-alkane (methane) to 2-alkane (ethane) selectivity ratio follows a reverse trend, i.e. **high-$Q_H$** has higher methane to ethane production, followed by **mid-$Q_H$** and **low-$Q_H$**. This indicates that for a fixed value of $Q_c$, the tendency for increase in chain length decrease with increasing $Q_H$.

A similar analysis on performance at two different atomic chemisorption enthalpies of carbon $Q_C$ ($Q_C$=630 kJ/mol and 665 kJ/mol) at constant values of $Q_H$ (240 kJ/mol) and $Q_O$ (570 kJ/mol) is carried out. It is observed that the yield of light olefins increases with decreasing $Q_C$, whereas it follows a reverse trend for the yield of long-chain hydrocarbons. This is attributed to the increased relative surface coverage of $MMCH_2$ at higher values of $Q_C$. For the two $Q_C$ values investigated the relative surface coverage of $MMCH_2$

Table 2.4: Relative selectivity of alkanes to alkenes and 1-alkane (methane) to 2-alkane (ethane) at three virtual points **low-$Q_H$**, **mid-$Q_H$** and **high-$Q_H$**. The points are located in the $Q_H - Q_C$ plane containing FePb/CNT catalyst. The atomic chemisorption enthalpies of carbon and oxygen are kept at a constant value of $Q_C$= 642 kJ/mol and $Q_O$= 577.1 kJ/mol. The points **low-$Q_H$** ($Q_H$= 240 kJ/mol), **mid-$Q_H$** ($Q_H$= 248 kJ/mol) and **high-$Q_H$** ($Q_H$= 252 kJ/mol) have different atomic chemisorption enthalpy of hydrogen. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2$/CO molar inlet ratio of 1.

| | **Relative selectivity** [-] | | |
|---|---|---|---|
| | **low-$Q_H$** - | **mid-$Q_H$** - | **high-$Q_H$** - |
| $\frac{\sum \text{alkenes}}{\sum \text{alkanes}}$ | 4.0 | 0.88 | 0.31 |
| $\frac{1-\text{alkane}}{2-\text{alkane}}$ | 2.41 | 3.08 | 4.39 |

is 0.61 and 1 respectively where maximum surface coverage is of order of $10^{-9}$.

Thus, the yield of light olefins becomes lower due to increased methanation at high values of $Q_H$ and due to long chain hydrocarbon formation at high values of $Q_C$. At very low values of $Q_C$, unavailability of surface carbon also leads to lower conversion and thus lower light olefins yield.

### COMPARISON OF REACTION PATHWAYS

A reaction pathway analysis is carried out at virtual catalyst points **low-$Q_H$** and **high-$Q_H$**, to indicate the difference in trend in hydrocarbon selectivity as we traverse from lower to higher hydrogenative catalysts with respect to the FePb/CNT catalyst. The affinity calculations are used in Fig.2.6 to differentiate the elementary surface reactions which are at quasi-equilibrium (black arrows) and those that are kinetically relevant (colored arrows). The reactions :

- chemisorption of CO, $H_2$ and alkenes
- CO dissociation
- hydrogenation of surface oxygen and carbide

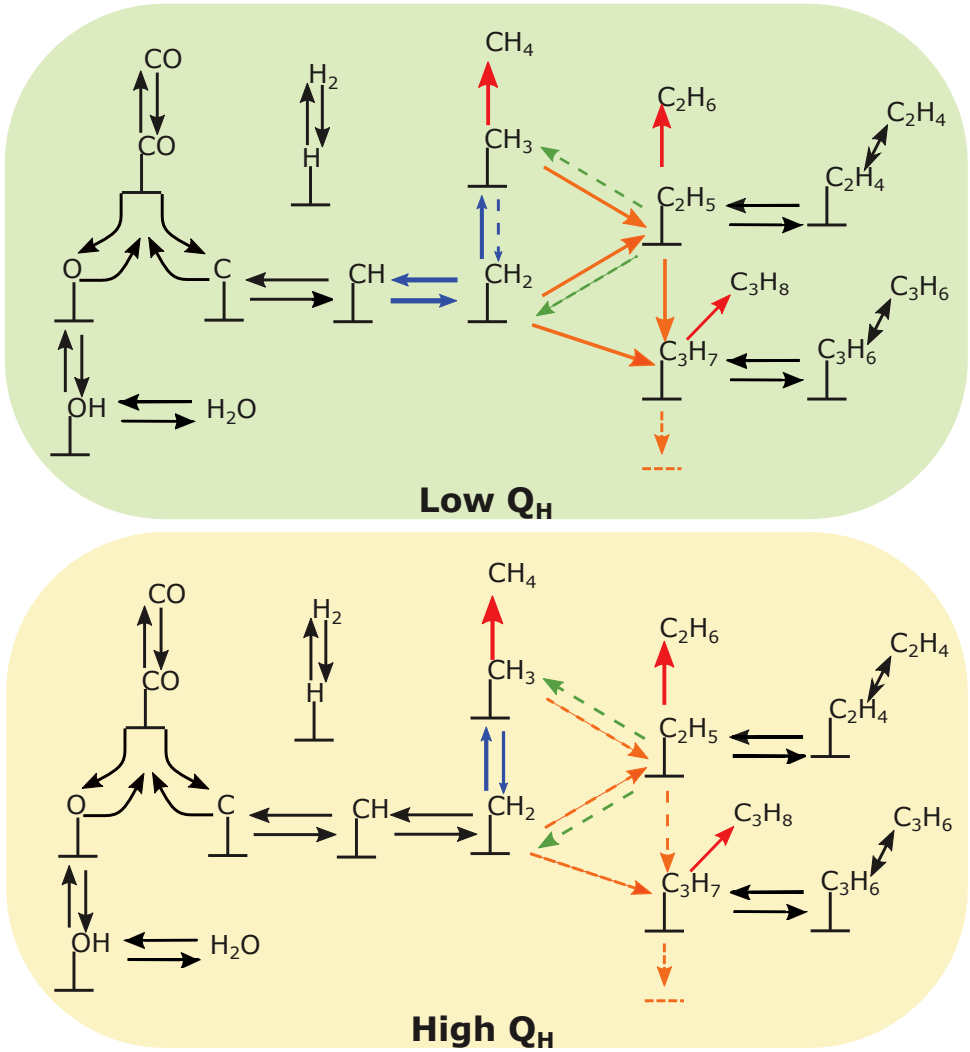



Figure 2.6: Reaction pathway analysis at **low-$Q_H$** ($Q_H$= 240 kJ/mol) and **high-$Q_H$** ($Q_H$=252 kJ/mol) is carried out at temperature of 623 K, pressure of 10 bar, GHSV of 3.4 $Lg^{-1}h^{-1}$,$W_{cat}$ of 0.2 g and $H_2$/CO molar inlet ratio of 1. The colored arrows with a thickness of less than 1 unit are represented with dotted arrows. The rate corresponding to oxidative addition of $MCH_3$ (in the case of **high-$Q_H$**), is used as the base (with lines thickness of 1 unit) for the scaling of colored arrows. The reaction between hydrogen attached to the metal surface and other surface species are not shown, for better visualization.

were found to be in quasi-equilibrium for both the virtual catalysts, **low-$Q_H$** and **high-$Q_H$**.

At **low-$Q_H$** (Fig.2.6), it could be noted that MCH hydrogenation is kinetically relevant

when compared at **high-$Q_H$**, where the reaction is in quasi-equilibrium. This leads to a reduced surface coverage of MH at **low-$Q_H$** (Table 2.3) and, hence, lower hydrogenation of metal alkyls. Since the $MMCH_2$ insertion reaction is more prominent at **low-$Q_H$** along with lower hydrogenation, an enhanced production of chain growth products, more particularly olefins, is observed. However, at **high-$Q_H$** (Fig.2.6), since the MCH hydrogenation is in quasi-equilibrium, the surface coverage of MH is higher, leading to an increased alkane production, particularly methane. Thus, it is explained how a change in catalyst descriptor values can influence the surface coverages (Table 2.3) of key intermediates, thus altering the prevailing reaction pathways (Fig.2.6) and ultimately impacting the product selectivities (Fig.2.5).

## 2.5. IDENTIFICATION OF OPTIMUM CATALYST

From our discussion above, we observe that the effects of each catalyst descriptor on conversion and selectivity are inter connected to each other. To identify the trends in conversion and selectivity due to the simultaneous variation of the three catalyst descriptors and, hence, to identify the best catalyst leading to maximum yield of light olefins, iso-performance surfaces are presented in Fig.2.7. Each sub-figure in Fig.2.7 contains two iso-surfaces, one corresponding to a higher value and another to a lower value of the analyzed performance indicator. The values of iso-surfaces are chosen such that a clear distinction in trend can be observed. The three catalysts, Fe/CNT, FeBi/CNT, and FePb/CNT are represented with black dots. Depending on the relative location of catalysts *w.r.t.* the iso-surface the black dots representing the catalysts, appear dark (not enclosed by the surface) or light (enclosed by the iso-surface).

As was already evident from Fig.2.4(a)-(c) and Fig.2.7(b), the light olefins selectivity weakly depends on $Q_O$ with other descriptors being kept constant. However, the conversion significantly increases with decreasing $Q_O$, for the range of catalyst descriptors investigated (see also Fig.2.3(a)-(c) and Fig.2.7(a)). This leads to an increased light olefins

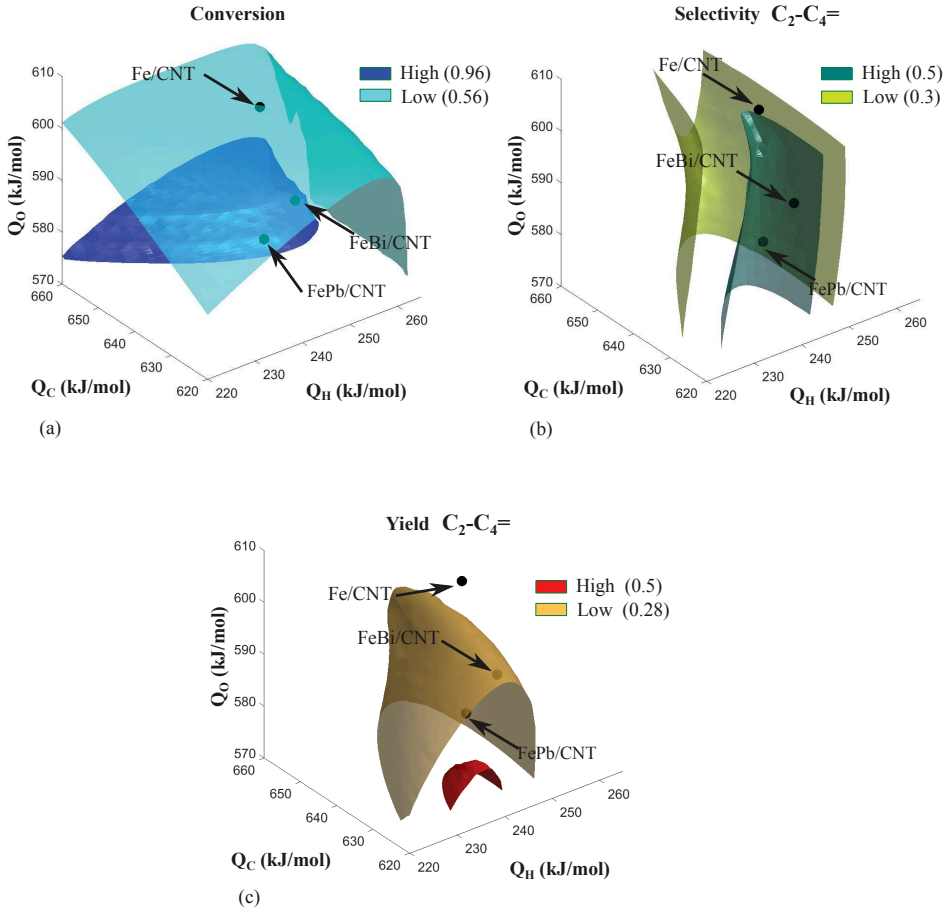



Figure 2.7: Iso-surfaces of conversion (a), selectivity- $C_2 - C_4 =$ (b) and yield- $C_2 - C_4 =$ (c) as a function of atomic chemisorption enthalpy of $Q_H$, $Q_C$ and $Q_O$. Each figure has 2 iso-surfaces corresponding to high and low values respectively. The catalysts Fe/CNT, FBi/CNT and FePb/CNT are represented by black dots. The simulations are reported at operating condition of 623 K, 10 bar, GHSV =3.4 $Lg^{-1}h^{-1}$, W=0.2 g and $H_2/CO$ molar inlet ratio of 1. The iso-surface values are reported in a scale between 0 and 1.

yield at lower $Q_O$. As discussed in Section 2.4.3, the lower light olefins yield at very low values of $Q_H$ and $Q_C$ is attributed to lower surface coverage of hydrogen and carbon, respectively. Increased production of methane at high values of $Q_H$, and that of long-chain hydrocarbons at high values of $Q_C$ also adversely affect the light olefins yield.

The combined effect of the descriptors mentioned above is evident from the iso-surface of higher yield (yield=0.5) being enclosed within the iso-surface of lower yield

(yield=0.28). We observe that the FePb/CNT and FeBi/CNT catalysts are close to the iso-surface corresponding to yield=0.28, while the non-promoted Fe catalyst is situated high above from this iso-surface. From Fig.2.7(c) we could infer that the yield of light olefins ($C_2 - C_4 =$) for the given operation conditions, may be increased to a value exceeding 0.5, by engineering the catalyst in such a way that $Q_H \approx 234$ kJ/mol, $Q_C \approx 622$ kJ/mol and $Q_O \approx 575$ kJ/mol. A catalyst with the above descriptors would result in an increased yield of light olefins close to 55% when compared to Fe/CNT (20%), FeBi/CNT (27%), and FePb/CNT (30%) catalysts.

## 2.6. CONCLUSIONS

The single-event methodology developed for Fe catalyzed Fischer Tropsch synthesis is used to differentiate the performance of different catalysts from a micro-kinetic perspective, using catalyst descriptors. The experimental results reported in literature are used to determine the catalyst descriptor values corresponding to Bi and Pb promoted and non-promoted Fe catalysts. These are used as a benchmark to explore the parametric catalyst descriptor space. The atomic chemisorption enthalpy of hydrogen, $Q_H$ effects the formation of methane vs longer hydrocarbons, and olefins vs paraffins. A very high value of $Q_H$ leads to higher hydrogenation and thus, higher methanation and lower olefin production. The atomic chemisorption enthalpy of carbon, $Q_C$ relates to carbon species on the surface and, hence, to chain growth by increased methylene insertion at higher values. Whereas, atomic chemisorption enthalpy of oxygen, $Q_O$ in the investigated range plays an important role in the availability of free metal sites. A lower $Q_O$, reduces the metal sites being blocked by oxygenates and thus permitting adequate coverage of metal- carbon and hydrogen, resulting in increased CO conversion. With the decrease in $Q_H$ (w.r.t. the real catalysts), the availability of free sites increases, that can provoke enhanced beta hydride elimination. This leads to an increased light olefins yield. At lower value of $Q_C$ (w.r.t. the real catalysts), the yield of long-chain hydrocarbons

is reduced. In the investigated range of catalyst descriptors, an enhanced light olefin production is observed at lower values of $Q_O$, $Q_H$ and $Q_C$ (w.r.t the real catalysts). For the given operating conditions, we identify that a "promising catalyst" with catalyst descriptors, $Q_H \approx 234$ kJ/mol, $Q_C \approx 622$ kJ/mol and $Q_O \approx 575$ kJ/mol which would result in an increased yield of light olefins (55%) compared to the catalysts Fe/CNT (20%), FePb/CNT (27%), and FeBi/CNT (30%).

The further identification of promoters that modify the catalyst descriptor values close to the values mentioned above can help us to attain enhanced yield of light olefins. The above study thus represents a first step to unravel the causes behind the relative performances of catalysts, and thus act as a guideline to engineer promoted catalysts with desired performance traits.

## REFERENCES

[1] Y. Cao, Z. Gao, J. Jin, H. Zhou, M. Cohron, H. Zhao, H. Liu, and W. Pan, *Synthesis Gas Production with an Adjustable H2/CO Ratio through the Coal Gasification Process: Effects of Coal Ranks And Methane Addition,* Energy & Fuels **22**, 1720 (2008).

[2] P. Jaramillo, W. M. Griffin, and H. S. Matthews, *Comparative Analysis of the Production Costs and Life-Cycle GHG Emissions of FT Liquid Fuels from Coal and Natural Gas,* Environmental Science & Technology **42**, 7559 (2008).

[3] I. Dimitriou, P. García-Gutiérrez, R. H. Elder, R. M. Cuéllar-Franca, A. Azapagic, and R. W. K. Allen, *Carbon dioxide utilisation for production of transport fuels: process and economic analysis,* Energy & Environmental Science **8**, 1775 (2015).

[4] M. E. Dry, *The Fischer–Tropsch process: 1950–2000,* Catalysis Today **71**, 227 (2002).

[5] PlasticsEurope, *Plastics – the Facts 2018, An analysis of European plastics production, demand and waste data,* (2018).

[6] E. M. Foundation, *New Plastics Economy, Catalysing Action,* (2017).

[7] R. Cuéllar-Franca, P. García-Gutiérrez, I. Dimitriou, R. H. Elder, R. W. Allen, and A. Azapagic, *Utilising carbon dioxide for transport fuels: The economic and environmental sustainability of different Fischer-Tropsch process designs,* Applied Energy **253**, 113560 (2019).

[8] N. Board, *Modern Technology of Plastic & Polymer Processing Industries* (Asia Pacific Business Press Inc., 2003).

[9] B. Gu, V. V. Ordomsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, and 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**, 153 (2018).

[10] A. J. Barrios, B. Gu, Y. Luo, D. V. Peron, P. A. Chernavskii, M. Virginie, R. Wojcieszak, J. W. Thybaut, V. V. Ordomsky, and 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**, 119028 (2020).

[11] M. A. Vannice, *Catalytic activation of carbon monoxide on metal surfaces,* in *Catalysis: Science and Technology* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1982) pp. 139–198.

[12] R. B. Anderson, *Forty Years With The Fischer-Tropsch Synthesis 1944-1984,* in *Catalysis on the Energy Scene* (Elsevier, 1984) pp. 457–461.

[13] Y. Qi, C. Aaserud, A. Holmen, J. Yang, and D. Chen, *Promotional effect of in situ generated hydroxyl on olefin selectivity of Co-catalyzed Fischer–Tropsch synthesis,* Physical Chemistry Chemical Physics **21**, 24441 (2019).

[14] Y. Qi, J. Yang, A. Holmen, and D. Chen, *Investigation of C1 + C1 Coupling Reactions in Cobalt-Catalyzed Fischer-Tropsch Synthesis by a Combined DFT and Kinetic Isotope Study,* Catalysts **9**, 551 (2019).

[15] J. Yang, C. L. Rodriguez, Y. Qi, H. Ma, A. Holmen, and D. Chen, *The effect of CO-feeding ethene on Fischer-Tropsch synthesis to olefins over Co-based catalysts,* Applied Catalysis A: General **598**, 117564 (2020).

[16] K. Cheng, V. V. Ordomsky, B. Legras, M. Virginie, S. Paul, Y. Wang, and A. Y. Khodakov, *Sodium-promoted iron catalysts prepared on different supports for high temperature Fischer–Tropsch synthesis,* Applied Catalysis A: General **502**, 204 (2015).

[17] J. Xie, J. Yang, A. I. Dugulan, A. Holmen, D. Chen, K. P. de Jong, and M. J. Louwerse, *Size and Promoter Effects in Supported Iron Fischer–Tropsch Catalysts: Insights from Experiment and Theory,* ACS Catalysis **6**, 3147 (2016).

[18] M. E. Dry, *High quality diesel via the Fischer-Tropsch process - a review,* Journal of Chemical Technology & Biotechnology **77**, 43 (2001).

[19] D. H. Chun, J. C. Park, S. Y. Hong, J. T. Lim, C. S. Kim, H.-T. Lee, J.-I. Yang, S. Hong, and H. Jung, *Highly selective iron-based Fischer–Tropsch catalysts activated by CO2-containing syngas,* Journal of Catalysis **317**, 135 (2014).

[20] I. A. W. Filot, R. A. van Santen, and E. J. M. Hensen, *The Optimally Performing Fischer-Tropsch Catalyst,* Angewandte Chemie International Edition **53**, 12746 (2014).

[21] D. J. Dooling and L. J. Broadbelt, *Microkinetic models and dynamic Monte Carlo simulations of nonuniform catalytic systems,* AIChE Journal **47**, 1193 (2001).

[22] A. G. Hansen, W. J. M. van Well, and P. Stoltze, *Microkinetic modeling as a tool in catalyst discovery,* Topics in Catalysis **45**, 219 (2007).

[23] A. H. Motagamwala, M. R. Ball, and J. A. Dumesic, *Microkinetic Analysis and Scaling Relations for Catalyst Design,* Annual Review of Chemical and Biomolecular Engineering **9**, 413 (2018).

[24] D. J. Klinke II and L. J. Broadbelt, *Construction of a mechanistic model of Fischer–Tropsch synthesis on Ni(1 1 1) and Co(0 0 0 1) surfaces,* Chemical Engineering Science **54**, 3379 (1999).

[25] S. Vernuccio, E. E. Bickel, R. Gounder, and L. J. Broadbelt, *Microkinetic model of propylene oligomerization on Brønsted acidic zeolites at low conversion,* ACS Catalysis **9**, 8996 (2019).

[26] L. H. Oakley, F. Casadio, K. R. Shull, and L. J. Broadbelt, *Microkinetic modeling of the autoxidative curing of an alkyd and oil-based paint model system,* Applied Physics A **121**, 869 (2015).

[27] J. Thybaut and G. Marin, *Single-Event MicroKinetics: Catalyst design for complex reaction networks,* Journal of Catalysis **308**, 352 (2013).

[28] E. Shustrovich, *The UBI-QEP method: A practical theoretical approach to understanding chemistry on transition metal surfaces,* Surface Science Reports **31**, 1 (1998).

[29] A. V. Zeigarnik and E. Shustorovich, *The UBI-QEP method: Mechanistic and kinetic studies of heterogeneous catalytic reactions,* Russian Journal of Physical Chemistry B **1**, 330 (2007).

[30] J. M. Martinis and G. F. Froment, *Alkylation on Solid Acids. Part 2. Single-Event Kinetic Modeling,* Industrial & Engineering Chemistry Research **45**, 954 (2006).

[31] J.-M. Schweitzer, P. Galtier, and D. Schweich, *A single events kinetic model for the hydrocracking of paraffins in a three-phase reactor,* Chemical Engineering Science **54**, 2441 (1999).

[32] N. Dewachtere, F. Santaella, and G. Froment, *Application of a single-event kinetic model in the simulation of an industrial riser reactor for the catalytic cracking of vacuum gas oil,* Chemical Engineering Science **54**, 3653 (1999).

[33] J. Verstraete, *Kinetic study of the catalytic reforming of naphtha on a Pt-Sn* /$Al_2O_3$ *catalyst*, Ph.D. thesis, Ghent University (1997).

[34] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Fischer-Tropsch Synthesis: Development of a Microkinetic Model for Metal Catalysis,* Oil & Gas Science and Technology - Revue de l'IFP **61**, 489 (2006).

[35] D. Sorescu, *First-principles calculations of the adsorption and hydrogenation reactions of CHx(x=0, 4)species on aFe(100) surface,* Physical Review B **73** (2006), 10.1103/physrevb.73.155420.

[36] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts,* Industrial & Engineering Chemistry Research **47**, 5879 (2008).

[37] G. Lozano-Blanco, K. Surla, J. Thybaut, and 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**, 423 (2010).

[38] M. Moqadam, M. Rahmani, Z. Karimi, and A. Naderifar, *A UBI-QEP Microkinetic Study for Fischer-Tropsch Synthesis on Iron Catalysts,* Procedia Engineering **42**, 34 (2012).

[39] J. V. Belleghem, C. Ledesma, J. Yang, K. Toch, D. Chen, J. W. Thybaut, and G. B. Marin, *A Single-Event MicroKinetic model for the cobalt catalyzed Fischer-Tropsch Synthesis,* Applied Catalysis A: General **524**, 149 (2016).

[40] L. Pirro, P. S. F. Mendes, S. Paret, B. D. Vandegehuchte, G. B. Marin, and J. W. Thybaut, *Descriptor–property relationships in heterogeneous catalysis: exploiting synergies between statistics and fundamental kinetic modelling,* Catalysis Science & Technology **9**, 3109 (2019).

[41] E. M. Shustorovich and A. V. Zeigarnik, *The UBI-QEP method: Basic formalism and applications to chemisorption phenomena on transition metal surfaces. chemisorption energetics,* Russian Journal of Physical Chemistry **80**, 4 (2006).

[42] J. A. Dumesic, D. F. Rudd, D. F. Rudd, L. M. Aparicio, J. E. Rekoske, and A. A. Trevino, *The microkinetics of heterogeneous catalysis* (Wiley-Vch, 1993).

[43] P. Linstorm, *Nist chemistry webbook, nist standard reference database number 69,* J. Phys. Chem. Ref. Data, Monograph **9**, 1 (1998).

[44] N. Cohen, *Revised Group Additivity Values for Enthalpies of Formation (at 298 K) of*

*Carbon–Hydrogen and Carbon–Hydrogen–Oxygen Compounds,* Journal of Physical and Chemical Reference Data **25**, 1411 (1996).

[45] P. W. Atkins and J. De Paula, *Physical chemistry,* (1998).

[46] SAS Institute Inc., Cary, NC, *JMP, Version 14,* (2018).

[47] U. M. Ascher and L. R. Petzold, *Computer methods for ordinary differential equations and differential-algebraic equations*, Vol. 61 (Siam, 1998).

[48] The Netlib, *http://netlib.org/,* .

# 3

# BRIDGING KINETIC MODELLING AND MACHINE LEARNING FOR CATALYST DESIGN

*Kinetic and data-driven machine learning models are useful tools to provide fundamental insights on catalyst development. However, there has been a limited interest to explore the synergies between both modelling approaches. In Chapter 2 we identified the catalyst properties of a potential catalyst with enhanced light olefin selectivity, using a visual identification approach. In this chapter, a clustering approach (unsupervised machine learning) coupled with microkinetic simulations is employed for automated identification of catalyst properties for achieving the desired target output, light olefins. This approach enables to overcome the limitation of existing visualization-based approaches (discussed in Chapter 2) limited to 3 dimensions, for identification of optimal catalysts. The approach demonstrated here automates the process of finding desired catalyst parameters (descriptors) irrespective of their dimensions and thus constitute an integral part in catalyst design and, hence, in chemical recycling technologies* [1].

[1] This work is published in Chemical Technology Symposium

## 3.1. INTRODUCTION

Plastics are used in our daily life and engineering applications such as packaging, electronic devices, engineering constructions, etc.[1]. More than 6% of the total oil produced globally is used for plastic production, a figure which is steadily growing [2]. Most of the resulting plastic waste at the "end-of-life" of the above products ends up as landfill. Various waste management techniques such as incineration or composting (bio-degradable plastics) are also utilized and further developed to reduce the plastic waste ending in our natural environment [3]. However, these waste management present certain limitations, e.g., due to the presence of contaminants or additives in the plastic waste leading to water contamination, slower degradation, etc. With the current urge for a circular economy [1] to reduce plastic waste and associated pollution, retain our resources and preserve our natural ecosystem; various alternatives such as mechanical recycling and chemical recycling are being adopted. Chemical recycling, utilizing various chemical processes such as pyrolysis and catalysis, has specific advantages over mechanical recycling owing to the preservation of material properties, lower material downgrading, etc. [4]. The use of a gasification process followed by a catalytic one involves the production of syngas which have different ratios of H2 and CO, which are then utilized for the production of liquid hydrocarbons using Fischer-Tropsch synthesis [5]. Technological developments with emphasis on better catalyst design to produce value-added chemicals such as light olefins via the FTS process [6] and, hence, to produce plastics will lead to a truly circular economy. To provide fundamental insights on catalyst development, kinetic models can be recurred to when properly accounting for the catalyst properties in the modelling procedure [7]. A detailed microkinetic analysis of the properties of iron-based catalysts to synthesize light olefins has previously been reported by Chakkingal et al [6]. Alternative techniques such as using machine learning, with a focus on the data, for catalyst design and predicting catalyst performances are also being investigated. Currently, there are a limited number of studies that combine the benefits of the mechanistic

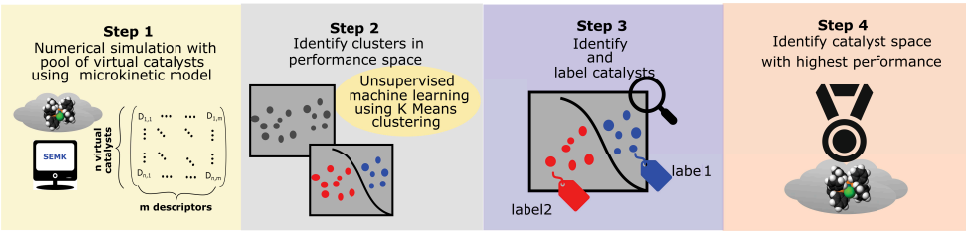



Figure 3.1: Steps involved in the identification of catalyst descriptor space with high performance using unsupervised machine learning approach, K- means clustering.



model along with the usage of machine learning techniques to address the catalyst design [8]. In our previous work, performance relations to obtain enhanced light olefin production via FTS were investigated using microkinetic simulations. The role played by catalyst descriptors, (i.e, atomic chemisorption enthalpy of hydrogen, carbon, and oxygen) were investigated. The identification of the best catalyst was aided with the help of three-dimensional graphical visualization of the catalyst descriptor space. However, this approach relying on data visualization becomes impractical when the number of catalyst descriptors exceeds 3. In the current work, we extend our previous studies on the Single event Microkinetic (SEMK) [6] framework by coupling it with a clustering-based unsupervised machine learning model to analyze catalyst performance. Here we propose a generalized clustering-based unsupervised machine learning approach suitable for any n-dimensional space, to identify catalyst descriptor space with higher performances.

## 3.2. METHODOLOGY

The current work on FTS coupling SEMK and unsupervised machine learning methodology aims at identifying the regime of catalyst descriptors (i.e a promising catalyst) that enhance light olefin production. The kinetic descriptor values used are based on the work by Lozano-Blanco et al. [9] for an iron-based catalyst. The methodology can be divided into 4 steps as shown in Fig. 3.1 :

- **Step 1: Generation of virtual catalyst library:**

A virtual library of catalysts is generated by varying the catalyst descriptors and fed to the SEMK model integrated into a PFR reactor model. The simulations are performed at the following operating conditions: 623 K, 10 bar, GHSV = 3.4 $Lg^{-1}h^{-1}$, W = 0.2 g, and H2/CO molar inlet ratio of 1, as reported in [10]. A detailed explanation of the simulations is reported by Chakkingal et. al [6].



- **Step 2: Identifying clusters in the performanace space:**

  The performance space of desired products, i.e., CO conversion and light olefin yield is divided into groups or clusters using a k-means clustering algorithm. K-means is an unsupervised machine learning technique designed to partition unlabelled data into distinct k groupings and, hence, identify hidden patterns in the data. Here each cluster is represented by its centroid which corresponds to the mean of the points assigned to the cluster. The clusters are designed so that total intra-cluster variation is minimized. After the random selection of k centre of clusters, an iteration step is followed to update the centroid until convergence is achieved. One of the prominant methods to determine the optimal k-value is the elbow method. In the elbow method plot variance within clusters is plotted as a function of k. The optimal number of clusters is indicated by the location of a bend (knee) in the plot.

- **Step 3: Labelling of virtual points to the corresponding cluster:**

  Once the virtual points/catalysts (represented by a combination of catalyst descriptors) in the performance space are grouped into clusters, the label or cluster identity is attached or tagged to the corresponding virtual point in the performance space. Thus, each virtual catalyst is identified in terms of its cluster identity.

- **Step 3: Identifying virtual catalyst space with highest performances:**

  After labelling virtual catalysts with their respective cluster identities, the cluster with the highest performance is selected. The range of catalyst descriptor values

for this cluster is then identified.

## 3.3. RESULTS AND DISCUSSION



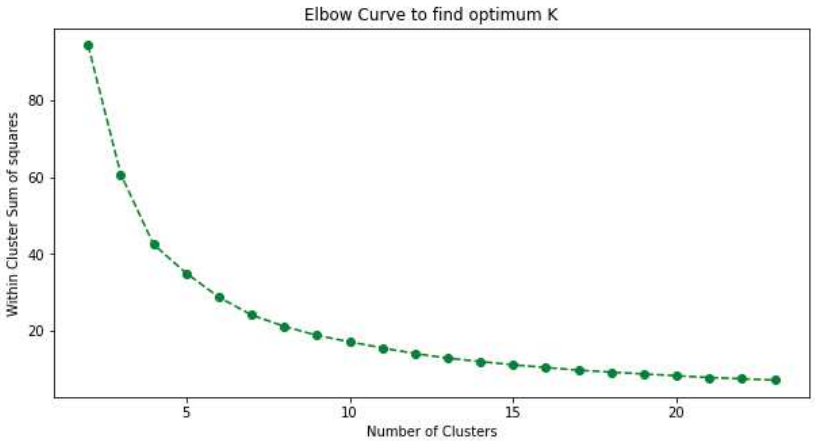



Figure 3.2: Elbow plot of within cluster sum of squares vs number of clusters in the conversion-yield performance plane, to identify the clusters required.

To find the optimal catalyst descriptor space for enhanced light olefin selectivity, the cluster groups of "Conversion-Yield of light olefins space" are identified. The elbow method is used to find the optimum number of clusters. Here, the within-cluster sum of squares, which is a measure of the variability of the observations within each cluster, is plotted as a function of the number of clusters. The minimum number of clusters is thus identified as 8 (see Fig.3.2), after which the variability decrease is minimal. With the obtained minimal number of clusters, the performance plane with light olefin yield on the y-axis and conversion on the x-axis is plotted in Fig.3.3. From the figure, it can be observed that cluster 1 has the highest light olefin yield with the highest conversion. The conversion and selectivity towards light olefins observed in this cluster are shown in Fig.3.4.

The upper and lower limits of each catalyst descriptor corresponding to the virtual catalysts in this cluster are compared with the descriptor space corresponding to the

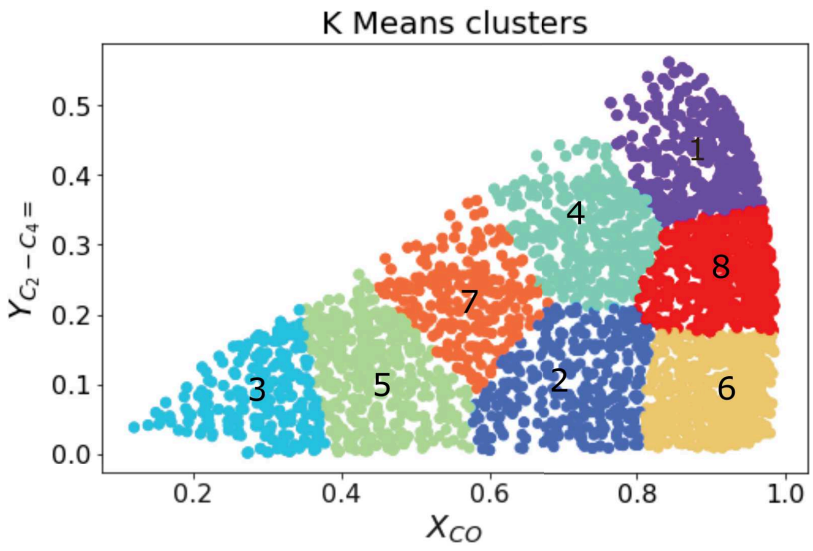



Figure 3.3: CO Conversion vs light olefin yield ( $Y_{C_2-C_4=}$ ) obtained with different virtual catalysts grouped into clusters (indicated by the colors) using K means clustering.

higher yield as reported in Chakkingal et al. (see Fig.3.5). The range of catalyst descriptor values within the cluster with the highest yield of light olefins is obtained to be $Q_H$ (225-246 kJ/mol), $Q_C$ (620-645kJ/mol) and $Q_O$ (575-590 kJ/mol). This is in line with the optimal catalyst descriptors reported in our previous work. A too high value of $Q_H$ results in higher hydrogen surface coverages, which causes the CO coverage to decrease and, hence, results in a decrease in CO conversion. while a too low value of $Q_H$ leads to a lower surface coverage of hydrogen thus adversely affecting the initiation and chain growth. On the other hand, too high values of atomic chemisorption enthalpy of carbon ($Q_C$ > 650 kJ/mol ) lead to lower availability of surface hydrogen (due to higher CO coverage), whereas, too low values lead to lower availability of surface carbon. The catalyst descriptor space values with the highest yield obtained with the automated unsupervised clustering approach are very close to that obtained using the visualization approach. Thus, compared to the current standard approaches, automated clustering with less manual intervention is a promising approach for identification of optimal catalyst

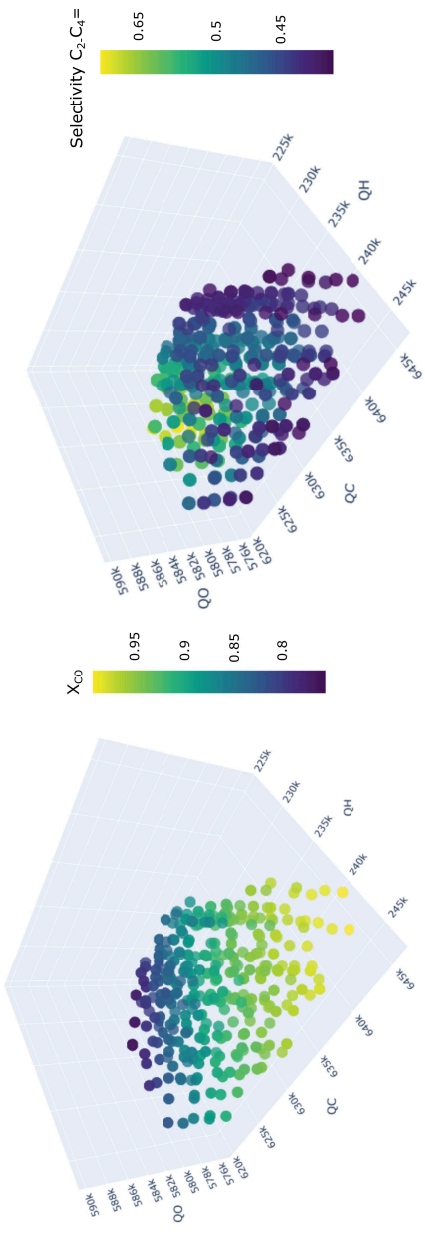



Figure 3.4: Conversion and selectivity as a function of catalyst descriptors obtained for the highest yield cluster.

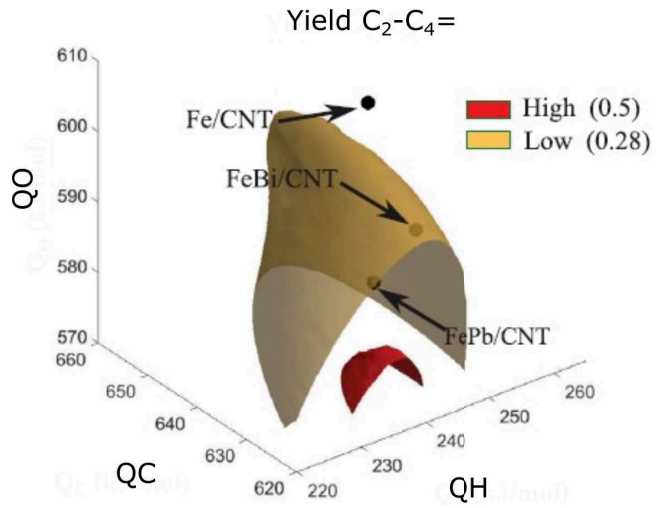



Figure 3.5: The iso-surfaces of light olefin yield obtained as a function of catalyst descriptors using visualisation approach with manual intervention, as reported by Chakkingal et al. [6].

space

## 3.4. CONCLUSIONS

A coupled "micro-kinetic-machine learning" approach for catalyst design and examining in detail the catalyst properties, reducing the human intervention, has been proposed. It has been used to assess the impact of catalyst properties (descriptors) on light olefin production via Fischer Tropsch synthesis, and thus to achieve optimal catalyst performance. This hybrid approach could help in understanding and realization of properties relevant to catalyst design and thus help in improving the chemical recycling technologies for plastic waste. The descriptor space generated by microkinetic data (virtual catalysts) is explored using a systematic unsupervised clustering (ML) and labelling approach. The performance space is grouped into clusters and the minimum number of clusters is identified to be 8. Each virtual catalyst (represented by the combination of catalyst descriptors) is then identified in terms of its cluster. The range of catalyst

descriptor values within the cluster with the highest yield of light olefins is obtained to be $Q_H$ (225-246 kJ/mol), $Q_C$ (620-645kJ/mol), and $Q_O$ (575-590 kJ/mol). This is in line with the optimal catalyst descriptors for enhanced light olefin production reported in our previous work [6]. The approach automates the process of finding the optimal catalyst descriptors irrespective of the dimension of the descriptor space and hence enabling to overcome the limitation of the manual visual inspection approach.

# REFERENCES

[1] PlasticsEurope, *Plastics – the Facts 2018, An analysis of European plastics production, demand and waste data,* (2018).

[2] R. C. Thompson, S. H. Swan, C. J. Moore, and F. S. vom Saal, *Our plastic age,* Philosophical Transactions of the Royal Society B: Biological Sciences **364**, 1973 (2009).

[3] K. S. Rebeiz and A. P. Craft, *Plastic waste management in construction: technological and institutional issues,* Resources Conservation and Recycling **15**, 245 (1995).

[4] J. Hopewell, R. Dvorak, and E. Kosior, *Plastics recycling: challenges and opportunities,* Philosophical Transactions of the Royal Society B: Biological Sciences **364**, 2115 (2009).

[5] Y. Cao, Z. Gao, J. Jin, H. Zhou, M. Cohron, H. Zhao, H. Liu, and W. Pan, *Synthesis Gas Production with an Adjustable H2/CO Ratio through the Coal Gasification Process: Effects of Coal Ranks And Methane Addition,* Energy & Fuels **22**, 1720 (2008).

[6] A. Chakkingal, L. Pirro, A. C. da Cruz, A. J. Barrios, M. Virginie, A. Y. Khodakov, and 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**, 129633 (2021).

[7] J. Thybaut and G. Marin, *Single-Event MicroKinetics: Catalyst design for complex reaction networks,* Journal of Catalysis **308**, 352 (2013).

[8] J. R. Kitchin, *Machine learning in catalysis,* Nature Catalysis **1**, 230 (2018).

[9] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts,* Industrial & Engineering Chemistry Research **47**, 5879 (2008).

[10] B. Gu, V. V. Ordomsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, and 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**, 153 (2018).

# 4

# MACHINE LEARNING BASED INTERPRETATION OF FTS MICROKINETIC DATA

*Machine-Learning (ML) methods, such as Artificial Neural Networks (ANN) bring the data-driven design of chemical reactions within reach. Simultaneously with the verification of the absence of any bias in the machine learning model as compared to the micro-kinetic data, interpretation techniques such as permutation importance, Shap values and partial dependence plots allow for a more systematic (model agnostic) analysis of these data. To achieve this aim, an investigation is carried out with FTS kinetic data generated using cobalt catalyst, with only single dominant output, methane yield. After gaining confidence on the investigated interpretation techniques, for an FTS reaction with single dominant output, a similar investigation on the potential iron based catalyst with enhanced light olefin selectivity is carried out in the next chapter.*



*For the purpose of the study in this chapter, the dataset required for training an ANN based ML model is synthetically generated using a Single-Event MicroKinetic (SEMK) model. With a number of 3 hidden layers with 20 nodes, the ANN model is able to adequately reproduce the SEMK results. The relative ranking of the process variables, as learnt by the ANN model, is identified using the interpretation techniques, the methane yield being most dependent on the temperature, followed by the space-time and syngas molar inlet ratio, in the investigated range of operating conditions. This is in line with the physico-chemical understanding from SEMK. A systematic approach for analysing micro-kinetic data, generally analysed on a case-specific basis, is thus developed by combining more widely used interpretation techniques in data science with the ANN*[1].

[1]This work is published in Reaction Chemistry & Engineering with doi: DOIhttps://doi.org/10.1039/D1RE00351H

## 4.1. INTRODUCTION

With increase in computational capacity and the ability to handle large volumes of data, high throughput experimental trials, etc., there has been an increased interest in applying machine learning models to chemical engineering problems. Machine Learning (ML) methods are rapidly gaining popularity for modelling complex nonlinear process phenomena in the field of chemical engineering [1–4]. ML is a sub-field of artificial intelligence, where information from the data is learnt using algorithms. Usage of different ML methods such as support vector machine (SVM), random forest and neural networks, etc. for classification and regression are being extensively investigated in different areas of chemical engineering such as electro-synthesis[5], biomass gasification [6], catalysis [7, 8], molecular drug discovery [9, 10], etc. The data required for developing ML based models are obtained from experiments [4, 11] or from synthetically generated using computational models [12]. These studies point at the increasing interest on ML based models in different sub-fields in chemical engineering.



Among different techniques in machine learning mentioned above, Artificial Neural Network (ANN) is one of the powerful predictive tools, which works on the principle based on the Universal Approximation Theorem [13]. ANN is used for both regression and classification. ANN relies on the collective working of the building units, i.e. the neurons [14]. The functioning of these neurons is inspired by that of biological ones. Here, relationships or patterns are established from the dataset between the input and output in the training stage, and the ANN model uses this information in the prediction stage. A neural network is considered a "black box" model, as it is difficult to interpret it in a fundamental manner when compared to models such as linear regression (Fig.4.1).

A machine learning algorithm such as ANN predicts the outcome based on the information learnt from the training set. The applicability and validity of such a model for a process are currently based on accuracy measurements such as the mean square error, mean absolute error, etc. However, relying on these metrics alone can make them biased

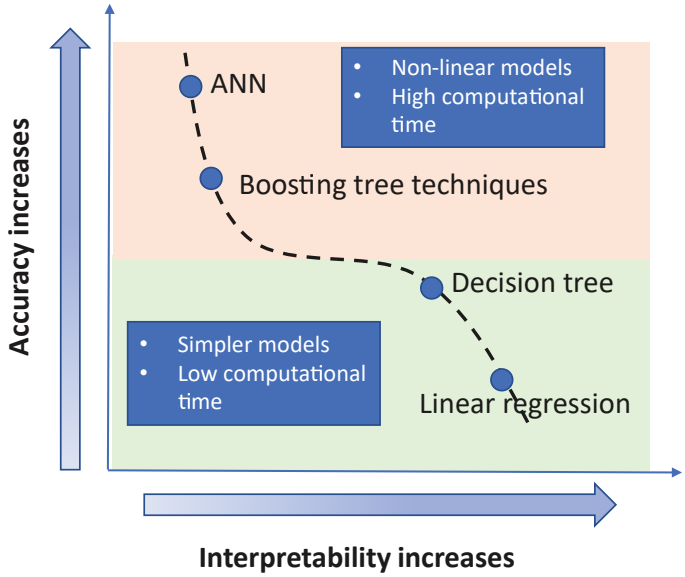



Figure 4.1: Graphical representation of variation in interpretability of a model with change in accuracy as reported in literature [15]. Interpretability of a model decreases with increase in accuracy and complexity of a model.

towards certain input features. Any bias in the dataset will be reflected in the model obtained from the algorithm. Thus, trusting the model also demands understanding why a certain decision is made by the model.

The identification of input features which exhibit the most pronounced contribution towards the target output prediction in the learning process of ANN is not straightforward. There is always a trade-off between prediction accuracy and interpretability of a model, see Fig.4.1. For a "simple" model such as linear regression, the weights or coefficients associated with the independent features provide a direct quantitative measure for their importance in the model. As the model complexity increases from linear/logistic regression to neural networks, the prediction accuracy increases, but the interpretability decreases [15]. To address this issue, with the recent advances in interpretation techniques [16], the interpretability of complex models such as ANN is being extensively investigated.

The most prominent interpretation techniques reported in literature are permuta-

tion importance [17], Shap values [18–20] and partial dependence plots [21]. The validity and explanation quality of these techniques depend on the situations in which they are used. Insights gained from different interpretation techniques allow revealing the relative impact of each input feature on the output. The combined effects and correlation of different input features may also be identified. The contribution of a particular feature in a multidimensional dataset can then be evaluated based on process expertise [22]. At present, these interpretation techniques, which are model agnostic in nature, are mostly used in medicine [23–26], finance [27, 28], etc. However, the application of these techniques in the field of chemical modelling is currently under-explored [22]. These techniques along with ML models can help in unravelling the hidden trends in kinetic data obtained from different chemical reactions, and for their systematic analysis.



Fischer-Tropsch Synthesis (FTS) is one such interesting chemical reaction where these techniques can be useful. FTS is widely investigated from synthesis gas that can be obtained from a wide variety of origins to synthesize hydrocarbons [29, 30]. The composition of non-petroleum-based hydrocarbons obtained via FTS depends upon a number of process features, such as the feed-stock (syngas ratios obtained after gasification) nature, the catalyst used, and the process operating conditions. The FTS reaction has been widely investigates experimentally [29, 30], via density functional theory [31], and by different kinetic models [32, 33]. A Single-Event MicroKinetic model is one such versatile, comprehensive kinetic model developed to deal with complex mixtures [34] in reactions such as hydrocracking [35], catalytic cracking [36], Fischer-Tropsch Synthesis [37, 38], etc. The analysis of the kinetic data obtained with such models is usually carried out on a case-specific basis, demanding expertise working with these models. In the recent decade, literature has also been reported on the use of ANN based models for modelling the FTS [39–41]. However, these studies are limited to the building of the neural network to predict the output components with limited focus on how each input feature plays a role in the prediction process.

In this chapter, a case study on interpretation of a "black-box" ANN regression model developed from the micro-kinetic data corresponding to Fischer-Tropsch Synthesis (FTS) is assessed with the help of different interpretation techniques mentioned above. This opens up the possibility of more systematic analysis and interpretation of kinetic data with the help of methods currently used widely in data science. Interpretability of ML models such as ANN can also build confidence in them to accurately predict results and draw chemical trends/insights. We could thus use them as an alternative to (micro)kinetic modeling and even to analyse the behavior of existing (micro)kinetic models using 'non-classical' contribution analysis techniques.



## 4.2. PROCEDURES

### 4.2.1. THEORETICAL BACKGROUND

An ANN is an efficient data-driven model which can learn the hidden patterns in a dataset, and transform input data into output [14].

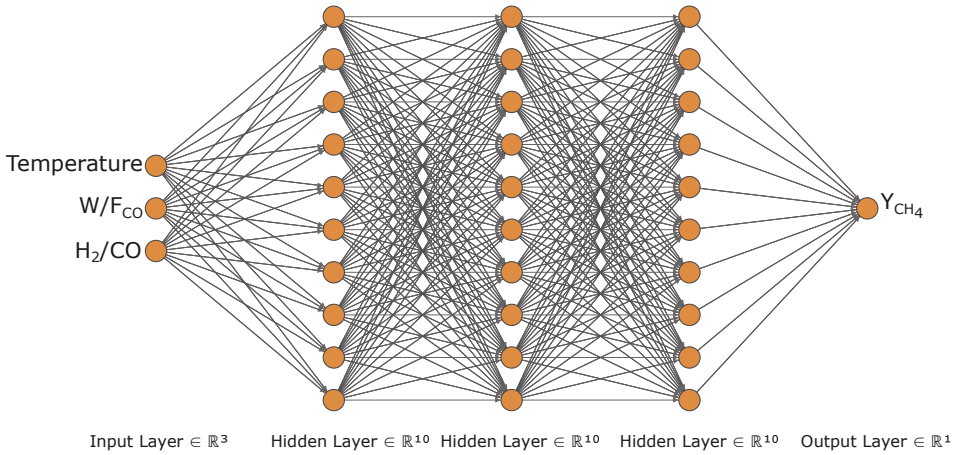



Figure 4.2: Schematic representation of an ANN consisting of 3 input features in the input layer, 3 hidden layers with 10 neurons in each layer and 1 output feature in the output layer. The input features in the representation are temperature, space-time ($W/F_{CO}$) and syngas molar inlet ratio ($H_2/CO$) and the output is methane yield ($Y_{CH_4}$).

As shown in Fig.4.2, an ANN is composed of components called neurons (colored

circles). A set of neurons is subsequently stacked to form layers, which are classified as:

- Input layer: contains the input features, i.e., for the FTS process: temperature, space-time and syngas molar inlet ratio.
- Hidden layer(s): The layer(s) of neurons between the input and output layers.
- Output layer: The layer of neurons that corresponds to the (predicted) output of the model, i.e. methane yield for the current FTS process.



The output is generated by assigning weights to the neurons, and applying activation functions to the input, output and hidden layers. The connections between the neurons have a weight that does a linear transformation on the input value, while the activation function does a non-linear transformation. Although there are different types of activation functions, the most conventional ones are the sigmoid and the Rectified Linear Unit (ReLU) for input (and output), and hidden layers, respectively [14]. The sigmoid activation function ensures that the network captures the non-linearity of the input-output relation, while the ReLU activation function in the hidden layers effectively avoids the vanishing gradient problem [42].

The output values at each iteration, also denoted as epoch in the field, are obtained after the input information is fed via a feed-forward propagation. A back-propagation algorithm is used to train the neural network by recalculating the revised weights based on the error obtained from the output value.

### 4.2.2. ARTIFICIAL NEURAL NETWORK CONSTRUCTION AND ANALYSIS

An experimentally validated, Single-Event MicroKinetic model developed for the cobalt catalyst based FTS process [37] with single dominant output, methane, is incorporated into a 1-D pseudo homogeneous plug flow reactor model, and is used to *in-silico* generate kinetic data, to develop an ANN model. Insights on the decision-making process of the ANN model are obtained with the help of different interpretation techniques such as

permutation importance, Shap value and partial dependence plots. The steps involved in the development of the ANN model, and the explanation (interpretation) of results are shown in Fig.4.3. The important steps involved in the process are:

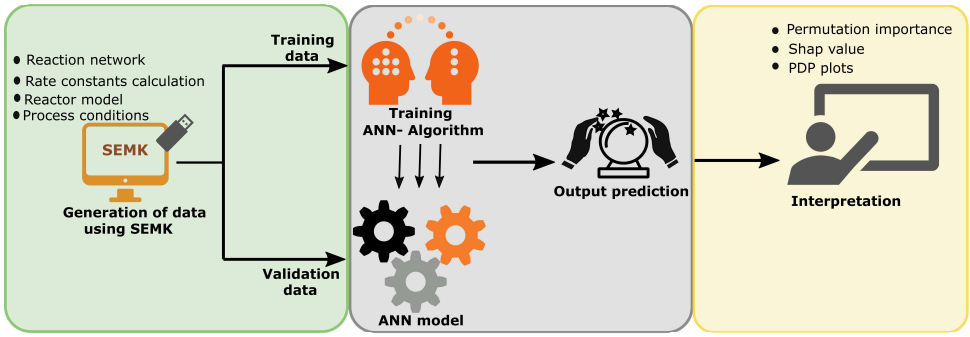



Figure 4.3: Schematic representation of the different steps involved in ANN model development and interpretation of the model. Data from SEMK simulations (green box) are used for training the ANN model (grey box) which is then analysed using different interpretation techniques (yellow box).

- **Step 1: Generation of datasets:**

  A dataset comprising 120 data points is generated at the following operating conditions: space-time 9 - 22 $(kg_{cat}s)mol_{CO}^{-1}$ , syngas molar inlet ratios 3 - 10 $mol.mol^{-1}$, temperature 483 - 503 K, and a total pressure of 1.85 bar, as reported by Van Belleghem et al.[37][2]. The catalyst and operating conditions in the cited work [37] are such that, a single dominant output, i.e. methane is produced. Detailed physico-chemical insights of this data, from a kinetic model's perspective is discussed in the cited work [37].

  The dataset is split into training and validation datasets with 75 and 45 data points, respectively. The input features used for training the network are space-time $(W/F_{CO})$, syngas molar inlet ratio $(H_2/CO)$ and temperature, with methane yield $(Y_{CH_4})$ as

[2]The experiments were reported at a single operating pressure condition of 1.85 bar which was fixed for the cited work.To ensure the repeatability of the experimental conditions reported, the effect of pressure is not accounted.

output. As the dataset is composed of input features with different units, a standardization is performed, where each feature is centered and scaled before training the ANN:

$$z = \frac{i - \mu}{\sigma} \tag{4.1}$$

where $i$ is the input feature, $\mu$ and $\sigma$ are the mean and standard deviation of the input feature, respectively, in the training dataset, and $z$ is the transformed input feature.



The validation dataset is then also transformed using $\mu$ and $\sigma$ obtained for the training dataset. Once the scaled datasets have been created, the ANN is trained in the next step.

- **Step 2: Training and prediction:**

  The scaled input features of the training dataset are fed into the neural network and the model is trained as follows:

  1. The weights associated with the neurons are initialized [43].
  2. Information is shared from one layer to the other to calculate a prediction $\hat{y}_i$ via a feed-forward propagation.
  3. After the feed-forward propagation step, a loss function is calculated which, in this study, equals the mean square error, MSE (see Eq.4.2) based on the methane yield.

$$MSE = \sum_{j=1}^{n} \frac{(\hat{y}_j - y_j)^2}{n} \tag{4.2}$$

  where $n$ represents the total number of observations, $\hat{y}_j$ is the ANN predicted output and $y_j$ is the observed output.

4. The loss function, MSE, is minimised via a back-propagation step. In this step the gradients of the loss function are calculated, and the error is used to update the new weights associated with the neurons.

5. The feed-forward propagation and the back-propagation steps are repeated iteratively (epochs) until the global minimum of the loss function is obtained. Here, an Adam optimizer [44] is used for reaching the global minimum.

6. The final weights associated with the neurons in the network are then used for making predictions using the ANN model.



Once the model is trained, the predictions are made on the so-called validation dataset.

- **Step 3: Interpretability of the learning process of ANN:**

  The interpretation of the performance of the ANN model developed for FTS is analysed using model agnostic interpretation techniques. The analysis is carried out by investigating the training set using different interpretation techniques such as permutation importance, Shap value and partial dependence plots (PD plots). With the help of permutation importance, the importance of each feature across the entire dataset is obtained. Next, with the help of Shap values the relevance of each feature in each set of operating conditions is obtained. The combined impact of different input features as interpreted by the developed ANN model is then discussed with the help of partial dependence plots. The steps involved in the calculation of each of these interpretation techniques is further explained in detail below.

PERMUTATION IMPORTANCE

The features which the ANN algorithm has assigned higher weighting or prominence to in its prediction are identified via determining the permutation importance [17, 45]. The importance of each input feature can be different for the yield of each output com-

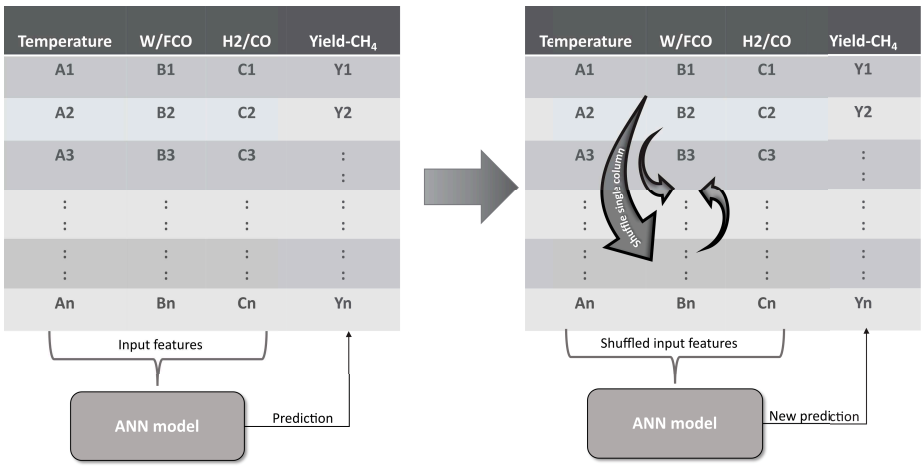



Figure 4.4: Graphical representation of the permutation importance principle. The model error is calculated using the ANN prediction without permutation (left) and ANN prediction with permutation (right) of input features such as $W/F_{CO}$.

ponent. Thus identifying the input feature importance can assist in achieving the targeted enhancement of desired output products. The permutation importance of the ANN model for FTS is calculated as follows:

1. A neural network model is made using the training dataset containing different input features, and the model error [3] for the training dataset is calculated.

2. To calculate the permutation importance of an input feature (for example space-time, $W/F_{CO}$), a new dataset is created by shuffling the rows of that feature in the training dataset (Fig.4.4).

3. A prediction for this new dataset is made using the model developed in step 1 above, and the model error is calculated.

4. The permutation importance of the feature is then calculated as the difference of model errors obtained in step 3 and step 1, above.

5. The above steps are repeated for the other input features (i.e for temperature and

[3] The model error is the difference between the actual output and the prediction.

syngas molar inlet ratio- $H_2/CO$) to obtain their permutation importance.

The feature resulting in the biggest difference in model error contributes the most to the model prediction, while the feature with smallest difference contributes the least.

SHAP VALUES

Unlike permutation importance, which represents the feature importance across the entire dataset, Shap values (Shapley values) are meant for a more local interpretation by pinpointing the contribution of each feature in each set of operating conditions. These values are used to explain the complex decision making of an ANN model with the help of simplified linear models. A complex ANN model that identifies non-linear patterns in the data is developed using the training dataset. Multiple linear, more easily interpretable models, also see Fig.5.3, which describe individual data points are then built to interpret the complex ANN model.

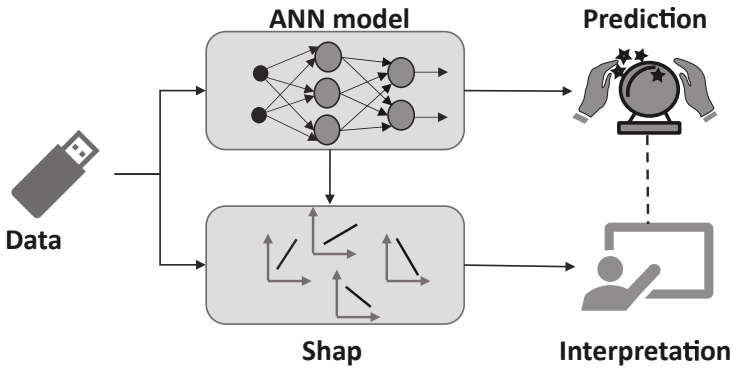



Figure 4.5: Graphical representation of how a Shap model assists in the interpretation of complex ANN models, with the help of linear models.

Instead of trying to explain the model in all its complexity, Shap values focus on how a complex model such as ANN behaves around a single data point. By considering the impact of features on individual data points, and then aggregating them, the interplay of combinations of features can be revealed. The Shap value gives the importance of a feature by comparing the model output obtained with and without that feature. The

Shap value for each feature is calculated as follows:

1. To calculate the Shap value of a feature $i$, create all possible subsets of features ($SS$), from features $F$. i.e. $SS \subset F$, the feature $i$ being excluded from $F$. After creating the subsets:

   (a) Train a model $M(SS_1 \cup i)$ including the feature $i$, and another model $M(SS_1)$ without it, where $SS_1$ is one of the subsets of input features.

   (b) Predict the output, $Y_{M(SS_1 \cup i)}$ with the model $M(SS_1 \cup i)$, using input features in $SS_1$ and the feature $i$.

   (c) Predict the output, $Y_{M(SS_1)}$ with the model $M(SS_1)$, using input features $SS_1$, and calculate the difference from the model prediction including the feature $i$ obtained in step 3.

2. Step 1 is repeated for all possible subsets of input features (without the feature $i$), i.e. $SS_1$, $SS_2$, $SS_3$, etc. as the effect of excluding the feature $i$ also depends on other input features.

3. Shap value (Shapley score) for feature $i$, $\phi_i$ is then calculated as:

$$\phi_i = \sum_{SS_i \subset F} \frac{|SS_n|!(|F_n| - |SS_n| - 1)!}{|F_n|!} \left[ Y_{M(SS_1 \cup i)} - Y_{M(SS_1)} \right] \tag{4.3}$$

   where, $SS_n$ represents the number of features in the subset, $F_n$ represents the total number of features.

4. Repeat the above steps for all other features to calculate their Shap values.

The above calculation can be carried out using the Shap library [46], which calculates Shap values significantly faster than calculating them via all possible combination of features.

PARTIAL DEPENDENCE PLOT

Partial dependence plots (PD plot) represent the marginal effect of (a combination of) selected input features on the output of a machine learning model, such as ANN. Visualizing the impact of higher-dimensional feature space on output prediction is difficult, when compared to 1D and 2D visualizations. This is overcome by visualizing the partial dependence of the output on selected small subsets of the input features. Though, a collection of these plots for different input features can be made, it does not provide a comprehensive depiction. However, PD plots are extremely useful for preliminary identification of trends, especially when the input features have lower order interactions, and when the features not included in the plot have less impact on output. The steps involved in the calculations for visualizing using PD plots are mathematically intensive and are discussed in detail in the cited work [47].



## 4.3. RESULTS AND DISCUSSION

### 4.3.1. NEURAL NETWORK IDENTIFICATION AND COMPARISON WITH SEMK

To identify the best performing network, multiple ANN configurations with a different number of neurons and hidden layers were trained using the dataset generated via SEMK simulations. As typically done, sigmoid activation functions are used in the input and output layer, whereas ReLU activation functions are used in the hidden layers, owing to their better performance compared to the other activation function combinations [14].

The number of hidden layers and neurons in each hidden layer are systematically varied to obtain the best performing model. This is assessed via the parity plot and $R^2$ value. In Fig.4.6(a), the variation in the $R^2$ value of the methane yield for the validation dataset is presented as a function of the number of hidden layers and the number of neurons within a hidden layer. A maximum $R^2$ value of 0.99 is obtained using a neural network composed of 3 hidden layers with 20 neurons in each layer. With an increase

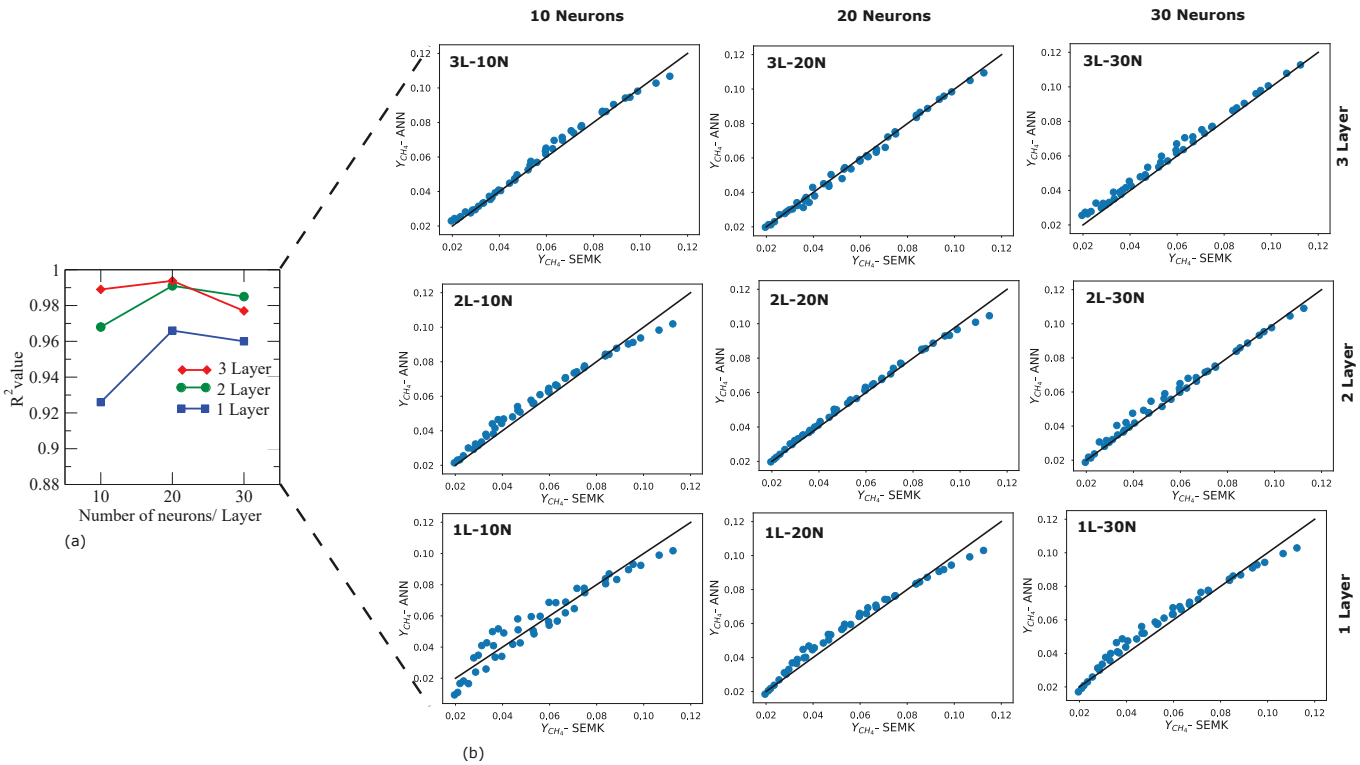



Figure 4.6: $R^2$ value (a) and parity diagrams (b) of methane yield obtained with SEMK simulations and ANN predictions, when using the validation dataset. Different ANN models with 1-3 hidden layers and 10-30 neurons in each layer are shown.

in the number of neurons, for a fixed number of hidden layers, the $R^2$ value initially increases and attains an optimum value, depending on the number of hidden layers. With a further increase in the number of neurons, the $R^2$ value as calculated against the validation dataset decreases, indicating an over-fitting.

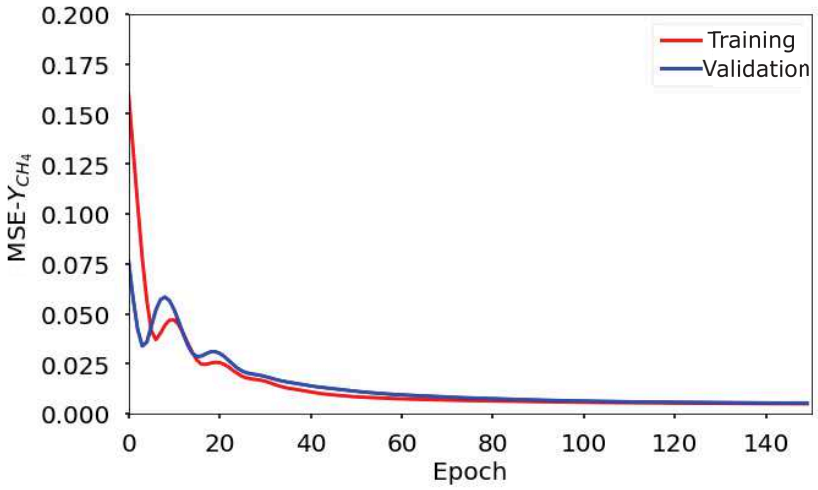



Figure 4.7: Convergence of the mean square error (MSE) of the methane yield, $Y_{CH_4}$ obtained with the ANN model consisting of 3 hidden layers with 20 neurons in each layer. The MSE for the ANN model converges after 80 epochs (iterations).

The neural network configuration that yields the highest $R^2$ value is chosen as the optimal one. Thus, all the analyses reported here are carried out using a neural network with 3 hidden layers containing each 20 neurons. This information is more explicit from the parity diagram, Fig.4.6(b). From Fig.4.7 it is observed that for the optimal neural network configuration, the mean square value of output yield (MSE), for both training and validation dataset converges to a stable value (indicating best learning) after 80 epochs.

To show the predictive capability of the ANN, the methane yield ($Y_{CH_4}$) obtained with ANN and SEMK simulations are compared in Fig.4.8, in which the methane yield is plotted as a function of space-time, at a syngas molar inlet ratio of 10 mol.mol$^{-1}$. It is observed that the methane yield increases with both space-time and temperature. As indicated by the slope of the lines (constant temperature) in Fig.4.8, the influence of the

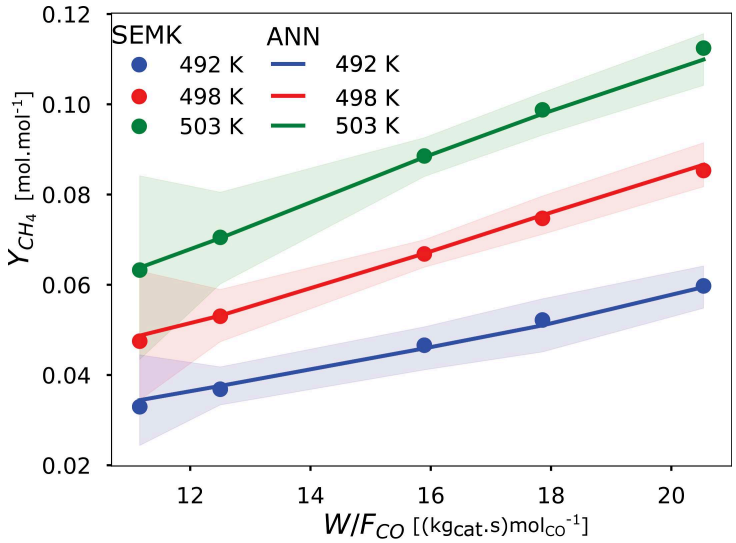



Figure 4.8: Comparison of the methane yield, $Y_{CH_4}$ obtained by SEMK simulations (·) and that with ANN model (–) as a function of space-time ($W/F_{CO}$) for varying temperatures at a syngas molar inlet ratio ($H_2/CO$) of 10 mol.mol$^{-1}$. The ANN model consists of 3 hidden layers with 20 neurons in each layer. The 95% confidence limit for the yield obtained with different intializations of the ANN model is represented by the shaded region around the mean ANN prediction (represented by solid line).

space-time on the methane yield increases with temperature. The methane yield obtained at the highest temperature and space-time for a syngas molar inlet ratio of 10 mol.mol$^{-1}$, is about the triple of that at the lowest temperature and space-time. As the results obtained from SEMK simulations and ANN predictions show a similar trend, it is concluded that the developed ANN model accurately predicts the output generated by the SEMK model in terms of the methane yield.

### 4.3.2. INTERPRETATION OF THE ANN MODEL

#### GLOBAL INTERPRETATION USING PERMUTATION IMPORTANCE

The feature resulting in the biggest difference in model error contributes overall the most to the model predictions, while the feature with smallest difference contributes the least. Within the investigated range of operating conditions, it can be observed from Fig.4.9 that the most prominent feature in the ANN model is the temperature. However, no

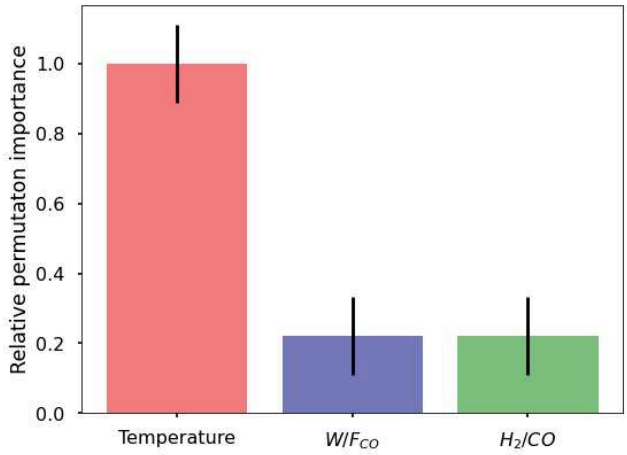



Figure 4.9: Relative importance of each input feature on a global level (i.e. averaged over the entire range of operating conditions in the training dataset). The relative importance is obtained by scaling the results with that of temperature. The relative importance is calculated using the Python package, Eli5 [48].

confirmation on the prominence of space-time over syngas molar inlet ratio or vice-versa, is obtained. The relative feature importance of temperature in the investigated training dataset is approximately 5 times that of both the space-time and the syngas molar inlet ratio. As discussed in Section 4.2.2 and 4.2.2, though the information on the global influence of a feature is obtained, the permutation importance analysis does not yield any information on the impact of the value of each feature with respect to the other features, for a specific set of inputs (i.e. locally). For example, no information can be extracted on whether the temperature has the same importance at a different space-time and a syngas molar inlet ratio.

#### LOCAL INTERPRETATION OF THE MODEL USING SHAP VALUE

The Shap values obtained for each input feature considered in the ANN model of the FTS process are shown in Fig.5.12. These values represent the local contribution of each feature in each set of operating conditions. The features are arranged in the order of their importance in the FTS process, for the range of operating conditions in the training dataset. The feature with the highest importance is the one with the widest range of Shap

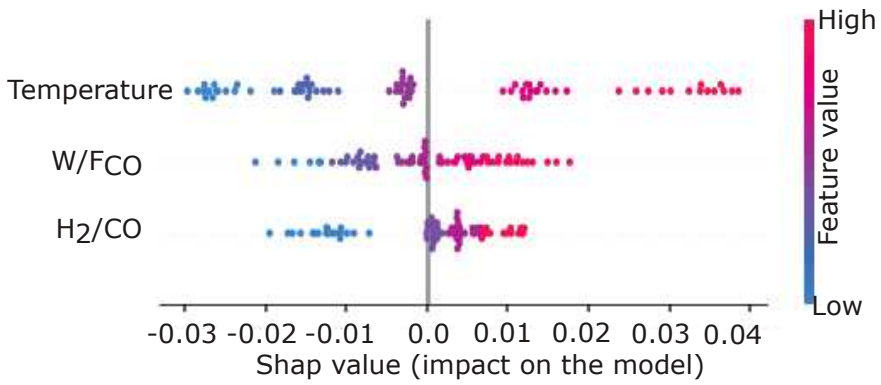



Figure 4.10: Local interpretation of importance of each input on the training dataset. The local feature importance is quantified in terms of contribution towards yield with respect to the average methane yield obtained from the entire training dataset is generated using the Python package, Shap [46].



values obtained for that feature. Also based on the Shap values the temperature is identified as the most influential feature, followed by space-time and syngas molar inlet ratio. The average yield in the training dataset, indicated by a Shap value of zero, serves as the base value for the analysis. The relative contribution is calculated with respect to this base value. Each dot (·) indicates the contribution of that feature towards the methane yield. The coloring used for the dots are indicative for the value of the respective feature, with blue representing lower values, while red represents higher values. From Fig.5.12, it is observed that the impact of each feature is monotonous in nature and, thus, an increase in each feature-value leads to an increase in yield. As the temperature, space-time or syngas molar inlet ratio increase from a low value (blue dots) to a high value (red dots), the contribution of that input feature to the methane yield which is initially low with respect to the base value (represented by a negative Shap value) increases. Thus, an increase in the input feature results in a positive contribution (represented by a positive Shap value) to the methane yield. The dispersion of the data points with comparable feature values (indicated by the spread of same coloured dots), also indicates a strong combined influence of features on the methane yield. These are inline with the results obtained with experimentally validated SEMK simulations [37].

INTERPRETATION OF THE MODEL USING PARTIAL DEPENDENCE PLOT



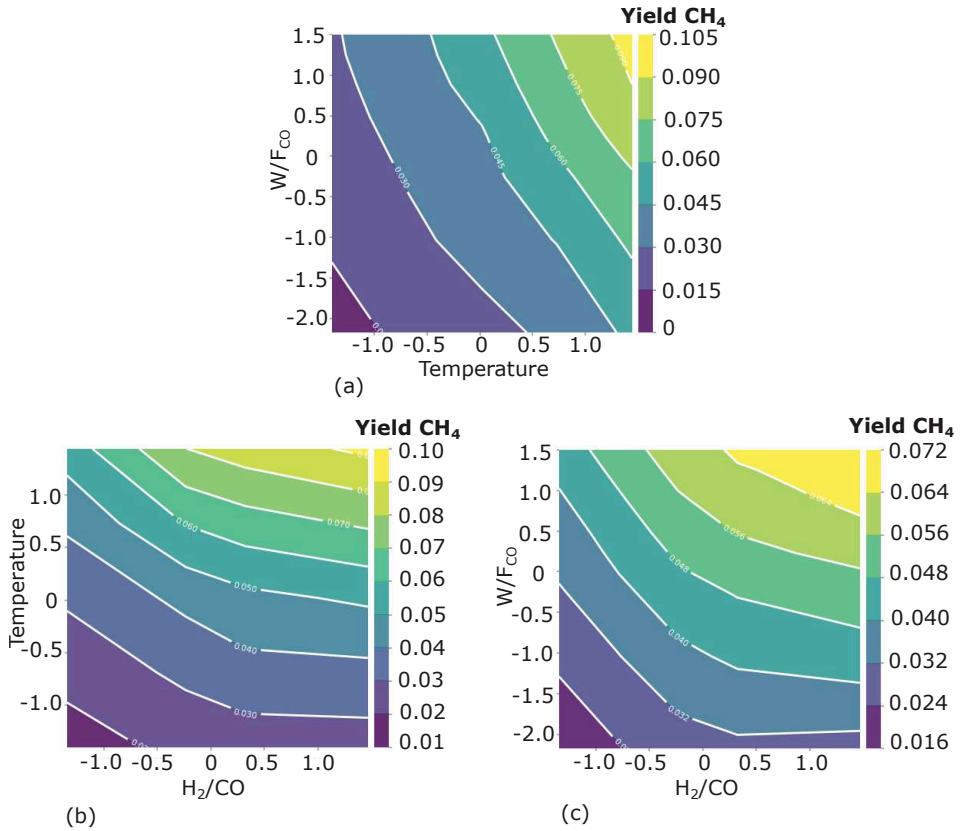



Figure 4.11: Marginal effect of 2 input features simultaneously considered, towards the methane yield. The effect of the 3$^{rd}$ feature is averaged out in each subplot, i.e. the results are plane-averaged along the 3$^{rd}$ feature. Effect of space-time ($W/F_{CO}$) and temperature (a), temperature and syngas molar inlet ratio ($H_2/CO$) (b), and space-time and syngas molar inlet ratio (c), in the range of investigated operating conditions. The input features are standardized for better visualization using mean and standard deviation of each input feature in the training dataset: temperature (492.8 ± 7.07 K), $W/F_{CO}$ (17 ± 3.5 ($kg_{cat}s$)$mol_{CO}^{-1}$), $H_2/CO$ (7.4 ± 1.8 $mol.mol^{-1}$). The PD plots are visualized using the Python package, pdpbox [49].

After the preliminary identification of the most important features, their combined impact is analysed with the help of partial dependence (PD) plots, thus targeting the most relevant features. From the analysis on permutation importance and Shap values, it is evident that the importance of the features in the ANN model prediction (methane yield) follows the order: temperature, space-time and syngas molar inlet ratio.

The combined impact of two input features on the methane yield is shown in Fig.4.11.

The effect of the 3rd feature is averaged out (by plane averaging the results along the 3rd feature). The combined influence of space-time and temperature is considered in Fig.4.11(a). The methane yield increases with both an increase of temperature and space-time. The increase in yield with temperature is more prominent at a higher space-time. The maximum methane yield averaged along the syngas molar inlet ratio (0.11 mol.mol$^{-1}$) is obtained at the highest temperature and the highest space-time. Though the yield increases with space-time at a fixed temperature, the change in yield is less prominent when compared to the change in methane yield with the change in temperature. These observations are in line with the SEMK simulation results reported in Fig.4.8. The combined influence of syngas molar inlet ratio and temperature in Fig.4.11(b) also indicates the monotonous increase in methane yield with temperature. It is observed that at a lower temperature, the yield is almost unaffected by the syngas molar inlet ratio while at a higher temperature, the effect of the syngas molar inlet ratio becomes pronounced. From Fig.4.11(c), it is observed that at lower space-time the yield is almost unaffected by the syngas molar inlet ratio. However, with an increase in space-time, the dependence of the methane yield on syngas molar inlet ratio increases. From the PD plot analysis it is, however, observed that the maximum yield observed in Fig.4.11(a), (b) and (c) varies, depending on the feature importance of the features, whose effect is averaged out in each plot. Although the influence of each feature on the methane yield is determined using the PD plots, it however remains important to check the impact of the averaged feature, to confirm the results. This observation is consistent with the nature of these plots, as discussed in Section 4.2.2.



## 4.4. CONCLUSIONS AND PERSPECTIVES

A machine learning model based on ANN for cobalt catalysed Fischer-Tropsch Synthesis with single dominant output, methane, is developed using a synthetic dataset generated via a Single-Event MicroKinetic (SEMK) model. The optimal ANN model for the FTS

process in the investigated range of operating conditions consists of 3 hidden layers with 20 neurons in each layer. This optimal network has a $R^2$ value of 0.99. After confirming that the methane yield obtained with the ANN model represents the one obtained with SEMK simulations, a systematic analysis of the kinetic data is carried out using different interpretation techniques.

For the Fischer-Tropsch process, in the range of investigated operating conditions, analysis of the ANN model using interpretation techniques shows that the prominent features follow the order: temperature > space-time > syngas molar inlet ratio. The global importance of temperature is roughly 5 times that of space-time and syngas molar inlet ratio. The investigation of the local contribution of each feature (Shap value) shows a monotonous increase in methane yield with increasing feature values. The coupled impact of input features on the methane yield is observed in the partial dependence plots, with the maximum yield (averaged along syngas molar inlet ratio) of 0.11 $mol.mol^{-1}$ obtained at the highest temperature and space-time. It is confirmed that a analysis of kinetic data can be carried out with the help of an interpretable ANN model. A deeper understanding of the FTS reaction mechanism, with the help of these techniques can be achieved by a multi-stage ANN, with process variables as the initial input to predict intermediate outputs such as surface coverages. These coverages can then be fed as an input into next stage of ANN to predict the performances.

The current study thus shows that more widely applied techniques in data science can now be applied for systematic analysis and interpretation of kinetic data. Similar analysis using experimental data can also help experimenters in their preliminary analysis, to detect hidden trends in the data, and thus to identify importance features. Extensive studies using the different techniques used in this study, for different chemical processes, will also help to identify the most important features. The understanding gained on the decision making by "black-box" ML models such as ANN, also enhances the confidence in using these techniques for FTS reaction with multiple outputs.

# REFERENCES

[1] A. Thakkar, S. Johansson, K. Jorner, D. Buttar, J.-L. Reymond, and O. Engkvist, *Artificial intelligence and automation in computer aided synthesis planning,* Reaction Chemistry & Engineering **6**, 27 (2021).

[2] N. S. Eyke, W. H. Green, and K. F. Jensen, *Iterative experimental design based on active machine learning reduces the experimental burden associated with reaction screening,* Reaction Chemistry & Engineering **5**, 1963 (2020).

[3] Y. Yan, T. Mattisson, P. Moldenhauer, E. J. Anthony, and P. T. Clough, *Applying machine learning algorithms in estimating the performance of heterogeneous, multicomponent materials as oxygen carriers for chemical-looping processes,* Chemical Engineering Journal **387**, 124072 (2020).

[4] S. Mittal, S. Pathak, H. Dhawan, and 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**, 127385 (2021).

[5] N. S. Kaveh, F. Mohammadi, and 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**, 161 (2009).

[6] F. Kartal and U. Özveren, *A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus,* Energy (2020), 10.1016/j.energy.2020.118457.

[7] J. Fujima, Y. Tanaka, I. Miyazato, L. Takahashi, and K. Takahashi, *Catalyst acquisition by data science (CADS): a web-based catalyst informatics platform for discovering catalysts,* Reaction Chemistry & Engineering **5**, 903 (2020).

[8] C. A. Vandervelden, S. A. Khan, S. L. Scott, and B. Peters, *Site-averaged kinetics*

*for catalysts on amorphous supports: an importance learning algorithm,* Reaction Chemistry & Engineering **5**, 77 (2020).

[9] D. S. Palmer, N. M. O'Boyle, R. C. Glen, and J. B. O. Mitchell, *Random forest models to predict aqueous solubility,* Journal of Chemical Information and Modeling **47**, 150 (2006).

[10] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, *Generating focused molecule libraries for drug discovery with recurrent neural networks,* ACS Central Science **4**, 120 (2017).

[11] P. P. Plehiers, S. H. Symoens, I. Amghizar, G. B. Marin, C. V. Stevens, and K. M. V. Geem, *Artificial intelligence in steam cracking modeling: A deep learning algorithm for detailed effluent prediction,* Engineering **5**, 1027 (2019).

[12] J. Athavale, Y. Joshi, and M. Yoda, *Artificial Neural Network Based Prediction of Temperature and Flow Profile in Data Centers,* Proceedings of the 17th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, ITherm 2018 , 871 (2018).

[13] B. C. Csáji *et al.*, *Approximation with artificial neural networks,* Faculty of Sciences, Etvs Lornd University, Hungary **24**, 7 (2001).

[14] D. Graupe, *Principles of Artificial Neural Networks,* 3rd ed., Advanced Series in Circuits and Systems, Vol. 7 (WORLD SCIENTIFIC, 2013).

[15] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, *Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions,* IEEE Access **7**, 137184 (2019).

[16] C. Molnar, *Interpretable machine learning : a guide for making Black Box Models interpretable* (Lulu, Morisville, North Carolina, 2019).

[17] L. Breiman, *Random forests,* Machine learning **45**, 5 (2001).

[18] L. S. Shapley, *Stochastic Games,* Proceedings of the National Academy of Sciences **39**, 1095 (1953).

[19] S. Lundberg and S.-I. Lee, *An unexpected unity among methods for interpreting model predictions,* arXiv , 1 (2016), arXiv:1611.07478 .

[20] S. M. Lundberg and S. I. Lee, *A unified approach to interpreting model predictions,* Advances in Neural Information Processing Systems **2017-December**, 4766 (2017), arXiv:1705.07874 .

[21] Q. Zhao and T. Hastie, *Causal interpretations of black-box models,* Journal of Business & Economic Statistics **39**, 272 (2019).

[22] S. Zhong, K. Zhang, D. Wang, and H. Zhang, *Shedding light on a Black Box machine learning models for predicting the reactivity of HO[rad] radicals toward organic compounds,* Chemical Engineering Journal **405**, 126627 (2021).

[23] Z. Zhang, M. W. Beck, D. A. Winkler, B. Huang, W. Sibanda, and H. Goyal, *Opening the black box of neural networks: methods for interpreting neural network models in clinical applications,* Annals of Translational Medicine **6**, 216 (2018).

[24] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K. W. Low, S. F. Newman, J. Kim, and S. I. Lee, *Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,* Nature Biomedical Engineering **2**, 749 (2018).

[25] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, *From local explanations to global understanding with explainable AI for trees,* Nature Machine Intelligence **2**, 56 (2020), arXiv:1905.04610 .

[26] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, *What clinicians want: Contextualizing explainable machine learning for clinical end use,* arXiv , 1 (2019), arXiv:1905.05134 .

[27] D. Brigo, X. Huang, A. Pallavicini, and H. S. d. O. Borde, *Interpretability in deep learning for finance: a case study for the Heston model,* arXiv , 1 (2021), arXiv:2104.09476 .



[28] K. Futagami, Y. Fukazawa, N. Kapoor, and T. Kito, *Pairwise acquisition prediction with SHAP value interpretation,* Journal of Finance and Data Science **7**, 22 (2021).

[29] A. Y. Khodakov, W. Chu, and P. Fongarland, *Advances in the development of novel cobalt fischer-tropsch catalysts for synthesis of long-chain hydrocarbons and clean fuels,* Chemical Reviews **107**, 1692 (2007).

[30] A. J. Barrios, B. Gu, Y. Luo, D. V. Peron, P. A. Chernavskii, M. Virginie, R. Wojcieszak, J. W. Thybaut, V. V. Ordomsky, and 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**, 119028 (2020).

[31] J. Cheng, P. Hu, P. Ellis, S. French, G. Kelly, and C. M. Lok, *Some Understanding of Fischer–Tropsch Synthesis from Density Functional Theory Calculations,* Topics in Catalysis **53**, 326 (2010).

[32] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts,* Industrial & Engineering Chemistry Research **47**, 5879 (2008).

[33] C. G. Visconti, E. Tronconi, L. Lietti, R. Zennaro, and P. Forzatti, *Development of a complete kinetic model for the fischer–tropsch synthesis over co/al2o3 catalysts,* Chemical Engineering Science **62**, 5338 (2007).

[34] J. Thybaut and G. Marin, *Single-event MicroKinetics: Catalyst design for complex reaction networks,* Journal of Catalysis **308**, 352 (2013).

[35] G. G. Martens, J. W. Thybaut, and G. B. Marin, *Single-event rate parameters for the hydrocracking of cycloalkanes on pt/US-y zeolites,* Industrial & Engineering Chemistry Research **40**, 1832 (2001).

[36] W. Feng, E. Vynckier, and G. F. Froment, *Single event kinetics of catalytic cracking,* Industrial & Engineering Chemistry Research **32**, 2997 (1993).



[37] J. V. Belleghem, C. Ledesma, J. Yang, K. Toch, D. Chen, J. W. Thybaut, and G. B. Marin, *A Single-Event MicroKinetic model for the cobalt catalyzed Fischer-Tropsch Synthesis,* Applied Catalysis A: General **524**, 149 (2016).

[38] A. Chakkingal, L. Pirro, A. C. da Cruz, A. J. Barrios, M. Virginie, A. Y. Khodakov, and 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**, 129633 (2021).

[39] F. A. N. Fernandes, *Optimization of fischer-tropsch synthesis using neural networks,* Chemical Engineering & Technology **29**, 449 (2006).

[40] H. Adib, R. Haghbakhsh, M. Saidi, M. A. Takassi, F. Sharifi, M. Koolivand, M. R. Rahimpour, and S. Keshtkari, *Modeling and optimization of fischer–tropsch synthesis in the presence of co (III)/al2o3 catalyst using artificial neural networks and genetic algorithm,* Journal of Natural Gas Science and Engineering **10**, 14 (2013).

[41] F. A. N. Fernandes, F. E. Linhares-Junior, and S. J. M. Cartaxo, *Prediction of fischer–tropsch synthesis kinetic parameters using neural networks,* Chemical Product and Process Modeling **9**, 97 (2014).

[42] X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks,* in *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (JMLR Workshop and Conference Proceedings, 2011) pp. 315–323.

[43] J. Guo, *AI Notes: Initializing neural networks,* .

[44] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization,* (2017), arXiv:1412.6980 [cs.LG] .

[45] A. Fisher, C. Rudin, and F. Dominici, *All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously,* arXiv (2018), arXiv:1801.01489 .

[46] *Welcome to the shap documentationÂ¶,* .

[47] J. H. Friedman, *Greedy function approximation: A gradient boosting machine.* The Annals of Statistics **29**, 106834 (2001).

[48] *Overview â€" eli5 0.11.0 documentation,* .

[49] *Pdpbox - latest documentation,* .

# 5

# MULTI-OUTPUT MACHINE LEARNING MODELS FOR FTS KINETIC DATA EVALUATION

*Predicting the impact of input process variables on chemical processes is key to assess their performance of the latter. Models explaining this impact through a mechanistic approach are rarely readily available, complex in nature and/or require long development time. 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 Chapter 4 a discussion on machine learning based technique to interpret kinetic data of an FTS reaction with single dominant output was carried out. With the confidence gained on different interpretation techniques used for analysing the kinetic data in Chapter 4, an extended investigation on the optimal light olefin selective catalyst identified in Chapter 2 is carried out using different machine learning models. In this chapter, multi-output Fischer-Tropsch Synthesis data generated via mechanistic SEMK model 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. Temperature and pressure are identified as the dominant input variables. Among the considered ML models, ANN emerged as the superior performing one with respect to 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 (0.22x) > space-time (0.1x) > syngas ratio (0.03x). Temperature (1x) and pressure (0.26x) remain the dominant input variables for light olefin selectivity, but that of space-time (0.03x) and syngas ratio (0.03x) becomes comparable. This study provides a reference method for the identification of suitable ML models for multi-output prediction in chemical processes* [1].

5

[1] This work is is published in Chemical Engineering Journal with doi: DOIhttps://doi.org/10.1016/j.cej.2022.137186

## 5.1. INTRODUCTION

With the increased volume of data obtained via automation and high-throughput experimentation, data-driven models are becoming increasingly popular [1, 2]. As a result, machine learning (ML) methods are extensively investigated to assess chemical engineering problems [3–5] where the impact of input variables is highly non-linear in nature [6]. Depending on the process data complexity, the specific type of machine learning model suited to model the phenomena of interest varies.

At present, from the modeling perspectives, chemical engineering research challenges are mostly addressed via mechanistic models [7–9] which rely on physico-chemically meaningful relations in the data. However, not all features (mechanistic aspects) contribute equally 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 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. E.g. microkinetic modelling of Fischer- Tropsch Synthesis, accounting for a maximum hydrocarbon chain length of 8 carbon atoms, there are 131 number of metal-alkyl species, 36 alkanes, and 94 alkenes. All these are generated via 922 elementary steps. If the investigation is carried out by setting a higher value for the maximum hydrocarbon chain length, the number of species also increases, correspondingly [10]. Alternative ML models, when properly trained, allow for a faster and more efficient calculation of the (complex) reaction performance. This presents an opportunity for the development of ML models which are easier [11] to develop as compared to a detailed mechanistic model.



Machine learning models can be classified as :

- **unsupervised learning models**: Patterns are inferred from unlabeled input data, i.e. without output variables. Unsupervised ML models identify structure and pat-

terns from the input data by themselves, e.g. for clustering [12] or dimensionality reduction [13].

- **supervised learning models**: Models are trained using labeled data. Supervised ML models determine a mapping function between the input variables and the output variables, e.g., classification [14] or regression [15].

In chemical engineering, supervised learning methods (mainly regression-based methods) are widely used, for applications such as absorption [16], sludge treatment [17], reactor modelling [18], etc. ML regression methods reported in literature include the use of Lasso regression [19], decision tree [20], K-Nearest Neighbor (KNN) regression [21], Support Vector Regression (SVR) [22], Artificial Neural Network (ANN) [23] etc. The wide application of these techniques in catalysis has recently been reported by Takahashi et al. [24]. As the model complexity increases from Lasso to ANN, the interpretability of the model becomes less straightforward [25]. To mitigate this, aiming at a systematic analysis and interpretation of kinetic data, a model agnostic interpretation technique denoted as Shap has been developed [26]. This helps in building confidence in the model's ability to accurately draw chemical insights.



Data-driven modeling in catalysis is becoming an alternative modeling approach [24, 27] 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 [28]. Within the plastic chemical recycling processes, the Fischer-Tropsch Synthesis (FTS) may play an important role. In FTS [29], 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 [30]. 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 [31]. Several

experimental [32–34] and mechanistic/kinetic studies [8, 35] that aim at identifying the properties which steer these selectivities in FTS have been reported. Among the mechanistic models, kinetic models based on Single Event MicroKinetics (SEMK) have been widely applied in the field of oligomerization [36], cracking [37], alkylation [38] etc. and have emerged as a versatile tool to model the FTS [39, 40]. A Single Event MicroKinetic evaluation of the key catalyst properties for enhanced light olefin synthesis via FTS was reported in our previous work [41].

Studies using machine learning models such as neural network (NN) [42, 43] and Support Vector Regression (SVR) [44] to predict a single response FTS process have been reported in literature, as an alternative to detailed kinetic models. 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. A detailed review on multi-output machine learning is reported by Xu *et. al* [45]. It discusses challenges to multi-output learning by taking inspiration from big data.



Compared to single response scenarios, only a limited number publications on ML models for multi-response scenarios, such as conversion and selectivities in the FTS process, have been reported [46, 47]. 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 study, machine-learning algorithms such as Lasso algorithms [48], K-Nearest Neighbor (KNN) [49], Support Vector Regression (SVR) [50], and Artificial Neural Network (ANN) [51, 52] 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 [41]. 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 selec-

tivity 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 [53, 54].

## 5.2. THEORY

### 5.2.1. SINGLE EVENT MICROKINETIC MODEL

To quantify the contributions of competing reaction pathways in a complex mixture process, the Single Event MicroKinetic methodology serves as a versatile fundamental modelling tool to unravel the complex kinetics of a process [55]. In the Single-Event MicroKinetic (SEMK) methodology the reactive moieties are considered to determine the reactivity of individual molecules. Accordingly, the elementary reactions are classified into reaction families to reduce the number of parameters [55]. The model-parameters of a SEMK model are classified as kinetic and catalyst descriptors. The kinetic descriptors relate to the kinetic parameters [56] of the chemical reactions such as activation energies and rate coefficients, and are catalyst invariant. To quantify the effect of catalyst properties on the chemical reaction, catalyst descriptors such as atomic chemisorption enthalpies are utilized [39]. The parameters governing the microkinetic model are generally obtained from multi-response model regression to experimental data. In particular for FTS, the Single Event MicroKinetic methodology has been reported by Lozano-Blanco et al.[10, 39, 56]. The elementary steps and reaction families considered in the SEMK reaction network for FTS and the associated kinetic parameters, are summarized in Table 5.1.

In the case of FTS, atomic chemisorption enthalpies of carbon ($Q_C$ in carbide phase), hydrogen ($Q_H$ and $Fe_3O_4-H$ corresponding to carbide and oxide phase respectively), and oxygen ($Q_O$ in carbide phase) are the catalyst descriptors [56]. The atomic chemisorp-

Table 5.1: Elementary reactions and reaction families in the reaction network [56], where $E_a^{for}$ represents the kinetic descriptors and M represents the metal surface.

| ***Reaction family/elementary reaction*** | $E_a^{for}$ **(kJ/mol)** |
|---|---|
| 1. $H_2$+ 2M $\rightleftharpoons$ 2MH | 0 |
| 2. CO + 2M $\rightleftharpoons$ MMCO | 0 |
| 3. MMCO + 3M $\rightleftharpoons$ MMMC + MMO | 56.81 ± 0.53 |
| 4. MMMC + MH $\rightleftharpoons$ MMMCH + M | 77.66 ± 0.70 |
| 5. MMMCH + MH $\rightleftharpoons$ $MMCH_2$+ 2M | 11.94 ± 0.10 |
| 6. $MMCH_2$ + MH $\rightleftharpoons$ $MCH_3$ + 2M | 61.88 ± 0.50 |
| 7. $MC_nH_{2n+1}$ + $MMCH_2$ $\rightleftharpoons$ $MC_{n+1}H_{2n+3}$ + 2M | 44.79 ± 0.43 |
| 8. $MC_nH_{2n+1}$ + MH $\rightleftharpoons$ $C_nH_{2n+2}$ + 2M | 117.75 ± 0.67 |
| 9. $MC_nH_{2n+1}$ + M $\rightleftharpoons$ $MC_nH_{2n}$ + MH | 96.27 ± 0.50 |
| 10. $MC_nH_{2n}$ $\rightleftharpoons$ $C_nH_{2n}$ + M | 62.09 (n=2), 59.08 (n=3-10) |
| 11. O − CHO − M + M − OH +O $\rightleftharpoons$ O − COOH − M + O − H + M | 138.95 ± 1.15 |
| 12. MMO + MH $\rightleftharpoons$ MOH + 2M | 103.80 ± 0.96 |
| 13. MOH + MH $\rightleftharpoons$ H2O + 2M | 86.22 ± 0.62 |



tion enthalpies of the virtual catalyst with enhanced light olefin selectivity i.e. $Q_C$, $Q_H$, $Fe_3O_4 - H$ and $Q_O$ are 622 kJ/mol, 234 kJ/mol, 220 kJ/mol and 575 kJ/mol, respectively [41]. The SEMK simulations are carried out after setting these parameters. For carrying out the SEMK simulations the input process variables are: temperature, total pressure, weight of catalyst, flow-rate of carbon monoxide and flow-rate of hydrogen. The output components include methane, paraffins (C2-C5), olefins (C2-C5) and longer chain hydrocarbons.

### 5.2.2. MACHINE LEARNING MODELS

The increase in the available volume of data and the urge for less complex models has also led to an increased interest in machine learning based models. Most of the ML studies in modelling a chemical reaction such as FTS, with single and multiple outputs are based on neural network models [46, 47, 57]. No extensive studies on the suitability of different machine learning models for multi-output FTS reaction are reported in the lit-

erature. However, different machine learning models could be utilized for multi-output predictions in general. The most popular ML methods for regression are Lasso regression, decision trees, random forests, artificial neural networks (ANN), support vector regression (SVR), and k-nearest neighbor regression(KNN) [58].

In the current work, a selection of the above mentioned machine learning models, see (Fig.5.1) are compared to determine the requirements of a model to be suitable for predicting the impact of operating conditions in FTS for multiple outputs. Decision trees are highly sensitive to dataset being used to train and can easily become unstable with small changes in the dataset [58] and have, therefore, not be considered further. Although, random forests are considered good for regression, they are highly conservative with unknown data [58]. Hence, for the above reasons, the ML methods based on tree algorithms i.e., decision trees and random forest, are not considered in this thesis. It is also to be noted that the different machine learning models considered in this thesis are of varying complexity to interpret [25]. The simplest ML regression technique for a non-classification problem is linear regression, owing to its easy interpretability. One of the most complex models to interpret is ANN. The information exchange between different layers of neurons makes it a "black box" [26]. In the current work, we consider these two as the extreme models and also investigate the capabilities of two other ML-based regression models, namely KNN and SVR. The complexity of the machine learning models considered in this thesis, as reported in the literature, follows the order ANN > SVR > KNN > Lasso. ML methods such as Lasso regression, KNN regression, and ANN support multi-output regression directly and the implementation is similar to that with single output [58]. However, SVR does not support multiple output regression directly and needs a workaround. The multi-output problem is split into multiple single output sub-problems, and then the regression is carried out using SVR [58]. The details of these ML regression models: Lasso, KNN, SVR, and ANN are introduced in more detail in this section. The predictions made by each of these models are then compared with

the microkinetic model predictions and physico-chemical understanding in Section 5.4.

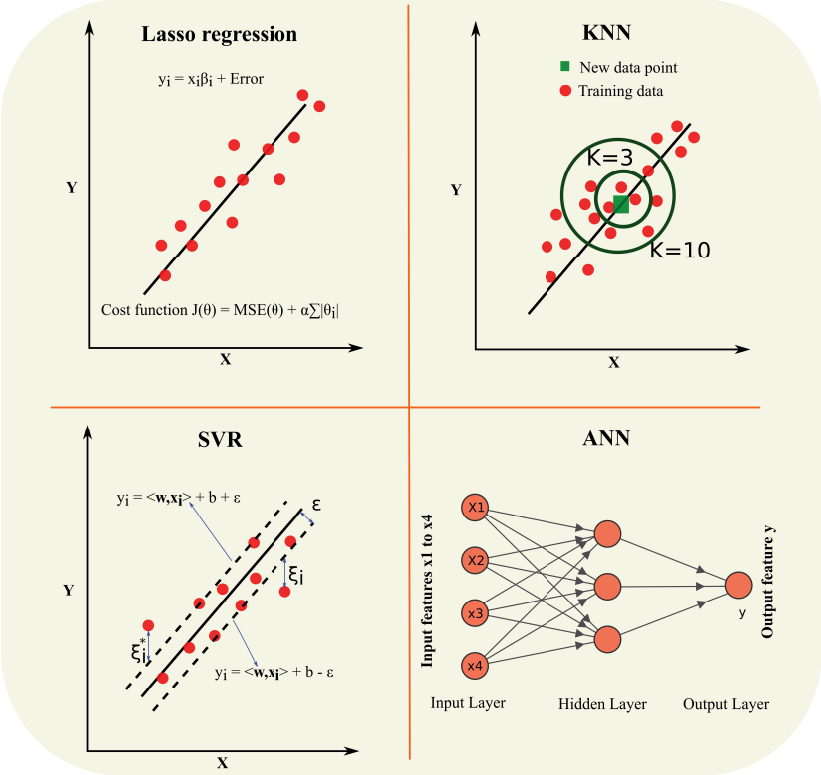



Figure 5.1: Different machine learning models investigated to analyse multi-output FTS reaction.

5

### LASSO REGRESSION

The most general supervised ML model 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 $\beta_i$. 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 with independent variables of lesser impact. This issue is addressed by Lasso regression [48], where regularization or shrinkage of the parameters is done to reduce over-fitting. It penalizes the

less important inputs in the dataset and thus creates a simple model by reducing their respective weights. Here the L1 norm [2] is used for regularization. The cost function containing the sum of mean square error (MSE) of the actual and predicted output, and the penalty as shown in Eq.5.1, is minimized to find the parameters in the Lasso regression.

$$J = MSE + \alpha \sum_{i=1}^{n} |\beta_i| \tag{5.1}$$

where $\alpha$ is the hyperparameter that controls the intensity of regularization.

#### K NEAREST NEIGHBORS (KNN) REGRESSION



KNN is a memory-based model [49] that makes predictions for new observations based on its "similarity" to the data used for training. Though KNN is mostly used for classification [59], it is also suited for regression [21]. It identifies k observations that are nearest or similar to a newly considered observation, 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 potentially leading to over-fitting, while using large values of k considers more neighbors may result in not capturing sufficient variation. 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:

$$d = \sqrt{\sum_{i=1}^{n} (x_a - x_b)^2} \tag{5.2}$$

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

[2] L1 norm is defined as the sum of the absolute values of the weights, $\beta_i$

### SUPPORT VECTOR REGRESSION

Support Vector Machine (SVM) [50] 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. SVR uses e-insensitive loss as the loss metric. The e-insensitive loss is given as:

$$Le = max(0, |r(x,y)| - \epsilon) \tag{5.3}$$

where $r(x,y)$ is the residual, i.e. the difference between the actual and the predicted output. A margin of width, $\epsilon$, as set by the modeller, 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. Thus, in essence, the aim is to contain maximum data points within the margin, with a minimal number of violations. 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. [3]

For multi-response regression, the SVR model requires a wrapper function (Fig.5.2) 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 model.

### ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is a machine learning algorithm [51] inspired by biological neural networks. It is a data-driven model which reveals the hidden patterns or non-linear relationships in the data. A neural network model is composed of its ba-

[3] The popular kernel functions used are $d^{th}$ degree polynomial, radial basis function (RBF), and hyperbolic tangent. A penalty coefficient C that controls the strength of the penalty term (loss function) 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, $\gamma$ for RBF) that also needs to be optimized [60].

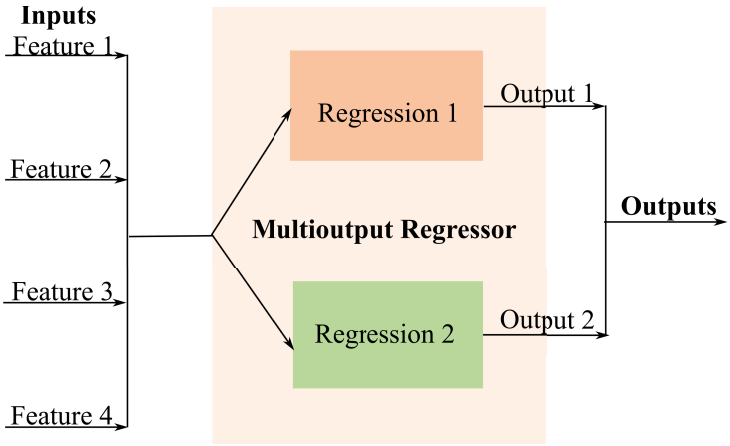



Figure 5.2: Wrapper function in multi-response Support Vector Regression. Separate regression results for each output are wrapped together by this model.



sic units called neurons, which are then stacked to form layers. An ANN 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 backpropagation [51]. This process continues until the difference between the prediction and the actual target is within the tolerance limit. If the number of neurons/layers 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 trial and error, which results in a network that yields a perfect prediction with minimization of a loss function such as RMSE.

INTERPRETABILITY OF ML MODELS

With the increase in complexity of ML models ranging from the simplest Lasso regression to the most complex ANN the "black box" character of the model increases, and the interpretation of the models becomes more challenging. For a "simple" model such as linear regression, the weights or coefficients of the independent variables give 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 practically uninterpretable [25]. To this purpose different interpretation techniques such as permutation importance [61] and Shap values [53, 54, 62] are now being developed and used. In the current study, 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.5.3, which describe the individual data points and thus helps to interpret a complex ML model.



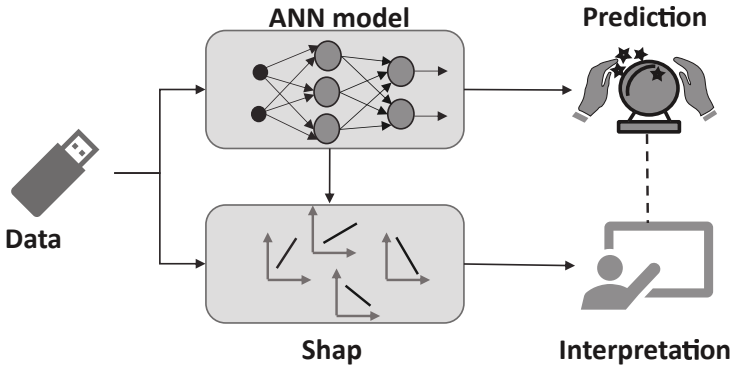



Figure 5.3: Graphical representation of how a Shap model assists in the interpretation of a complex model with the help of linear models. An ANN model is chosen as an example to represent a complex model.

The Shap value gives the importance of a feature by comparing the model output obtained with and without that feature. These calculations are carried out using the Shap library [63], which calculates Shap values significantly faster than calculating them via all possible combinations of features. A simplified interpretation of the algorithm is

reported in our previous work [43].

## 5.3. METHODS AND DATA

In the current study, a dataset is generated using a Single Event MicroKinetic model, validated for Fischer-Tropsch Synthesis experiments [56]. In our previous work [41], the catalyst descriptors were obtained for an optimally performing virtual catalyst for light olefin production. The focus in that work was on catalyst screening at a single set of operating conditions. Here, we extend our analysis to obtain a multi-response output within a wide range of operating conditions [4].



- 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

The data is uniformly distributed within this range i.e., there is no concentration of data at a particular condition, which could lead to a biased prediction. A total of 2450 synthetic experiment responses were generated and a train to test split ratio of 80:20 (non-stratified) is used in the analysis. Apart from the 4 input variables (temperature, pressure, space-time, syngas ratio), 4 output variables (conversion, selectivity of methane, paraffins, and light olefins) are considered in the dataset. The generated dataset is scaled and centered using standardization [60] to facilitate the learning by the machine learning models. In the training phase, actual model parameters are learned

[4] The kinetic and catalyst descriptors of the catalyst are fixed in the current study. The data thus generated does not include any uncertainty on kinetics or characteristic model parameters or catalyst descriptors. The investigation of the capabilities of the investigated ML models taking into account the uncertainty in data, due to sensitivity of output towards model parameters and operating conditions, is out of scope of the current study. The ML models are developed using the Python package sklearn [64]. The codes are provided as supplementary material in the corresponding published manuscript.

from the data. The hyperparameters control the learning process and critically influences the model parameters the algorithm ends up learning [5], is used as the re-sampling procedure to evaluate ML models on the data sample and thus prevent over-fitting [58]. 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: $\epsilon$=0.01, C = 10, $\gamma$ 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.



For the ANN, the number of nodes and layers for the ANN is selected by looking at the corresponding performance metric (R-square and RMSE values) for different combinations of neurons/ layers. Different combinations of layers (1 and 2 layers) and neurons (5, 10, 20, 30) in each layer were considered. The number of neurons and layers were increased in order to find the best values of performance metric. The R-square values were comparable for different combinations of neurons and layers, hence the RMSE values were compared, for the simplest scenario, 1 layer with different neurons. An RMSE threshold of $10^{-2}$ was chosen to determine the number of neurons. With 1 layer and 20 neurons in the hidden layer, the RMSE of the test dataset was less than $10^{-2}$ and is thus chosen as the optimal number of neurons. Higher values of neurons and layers which gave the same performance were disregarded to avoid over-fitting of the model. It is also to be noted that the computational time (which can become important when the volume of data is very high as in high-throughput experiments) decreases with decrease in

[5] Hyperparameters 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, where K = 5.

number of hidden layers, as the computations are more parallelised with less number of layers.

## 5.4. RESULTS AND DISCUSSION

### 5.4.1. 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.5.4 is shown in Fig.5.4.



$$r_p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5.4}$$

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.5.4) 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 within the investigated range of operating conditions. 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 the light olefin selectivity increases upon a decrease in pressure (blue) [6].

The above relation is also visible from the joint plot (Fig.5.5) drawn for the most sensitive input variable, i.e. temperature. Each joint plot in Fig.5.5 consists of 3 separate plots, 1 relationship plot, and 2 marginal distribution plots. For e.g. Fig.5.5(a), the grey-

[6] A similar Pearson correlation matrix with yield of the different hydrocarbons is available in Appendix A.

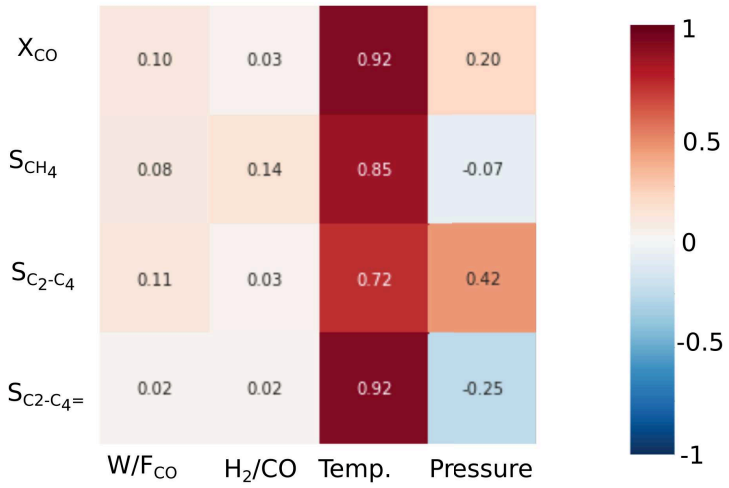



Figure 5.4: Heatmap of Pearson correlation between input variables: temperature, pressure, space-time and syngas ratio with output variables: conversion and selectivity of methane, paraffins and light olefins.



colored contours show the relationship between 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.5.5(a) give us the marginal distribution of temperature and conversion. The marginal distribution of temperature is obtained by projecting the relationship plot to a new horizontal axis. Similarly, the marginal distribution of conversion is obtained by projecting the data to a new vertical axis.

In Fig.5.5(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.5.5(b)). Higher conversion and light olefin selectivity are obtained in the temperature range 600-650 K. It is also observed that the conversion and light olefin selectivity do not further increase in this temperature range. This behavior is in line with that observed by Garona et.al [47].

From Fig.5.5 it is observed that the general trend of change in conversion and light

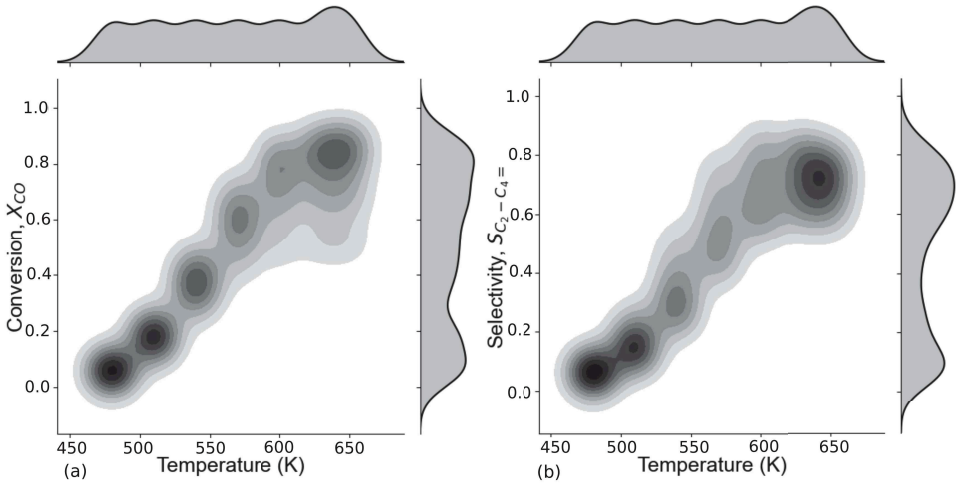



Figure 5.5: Conversion (a) and selectivity (b) behaviour as a function of temperature for the entire dataset. It is to be noted that the highest conversion and light olefin selectivity are not obtained at the same set of operating condition.



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.5.6). The variation of light olefin yield with the temperature at different syngas ratios ($H_2$/CO- 2 and 5) is shown in Fig.5.6(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 at different space-time ($W/F_{CO}$ - 24 and 64) is shown in Fig.5.6(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.5.6(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.

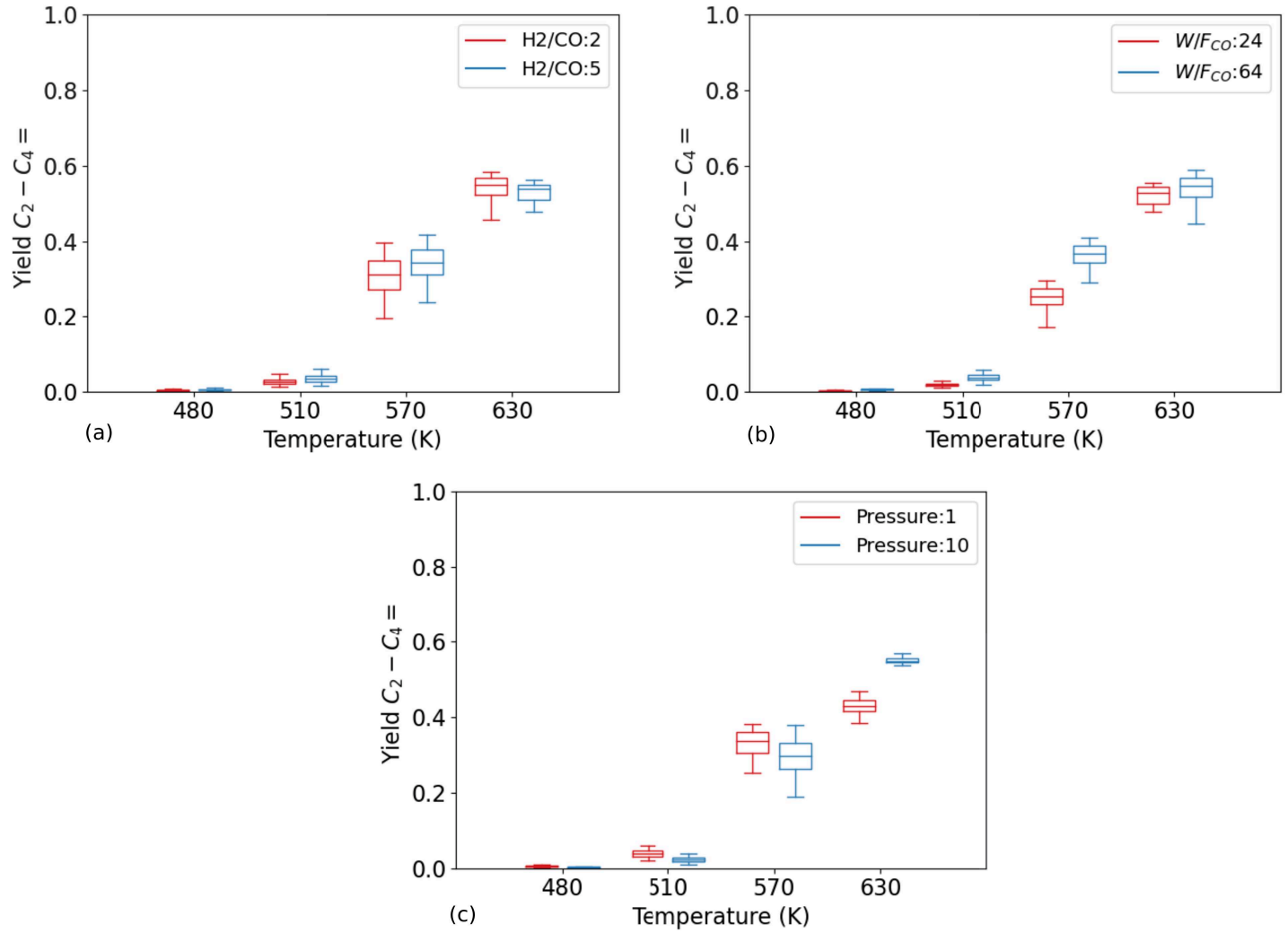



Figure 5.6: Yield as a function of temperature and syngas ratio (a),temperature and space time (b), temperature and pressure (c) for the entire range of operating conditions in the dataset.

### 5.4.2. PARITY PLOT COMPARISON OF THE ML MODELS

From the analysis above on the SEMK benchmark dataset, the most important input variables and their impact on conversion and selectivity have been determined. 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.5.7-5.8. The training dataset is indicated by the green colored dots in the parity diagrams whereas the blue dots correspond to the testing or prediction dataset. The parity diagram with reference to the training dataset shows how adequate the model is fitted to the data. The parity diagram for the testing dataset shows how well the model can predict a new dataset.



It is observed from the conversion behavior in Fig.5.7(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 5.2). 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.5.7(b-d), it is observed that the KNN, SVR, and ANN better fit the data than the Lasso regression. In addition to the R-square plots, standardised residuals are also compared (see Appendix A). It is observed that though the R-square values obtained with different ML models are comparable, the residual at low conversion is higher with Lasso regression and KNN regression.

It is observed from Fig.5.8(a-d), that the light olefin selectivity prediction better matches with the parity lines for KNN, SVR, and ANN than for the Lasso regression model. The R-square values for the selectivity of light olefins (ranging between 0.89 and 0.99) are comparable for all the ML models. The test root mean squared (RMSE) value of light

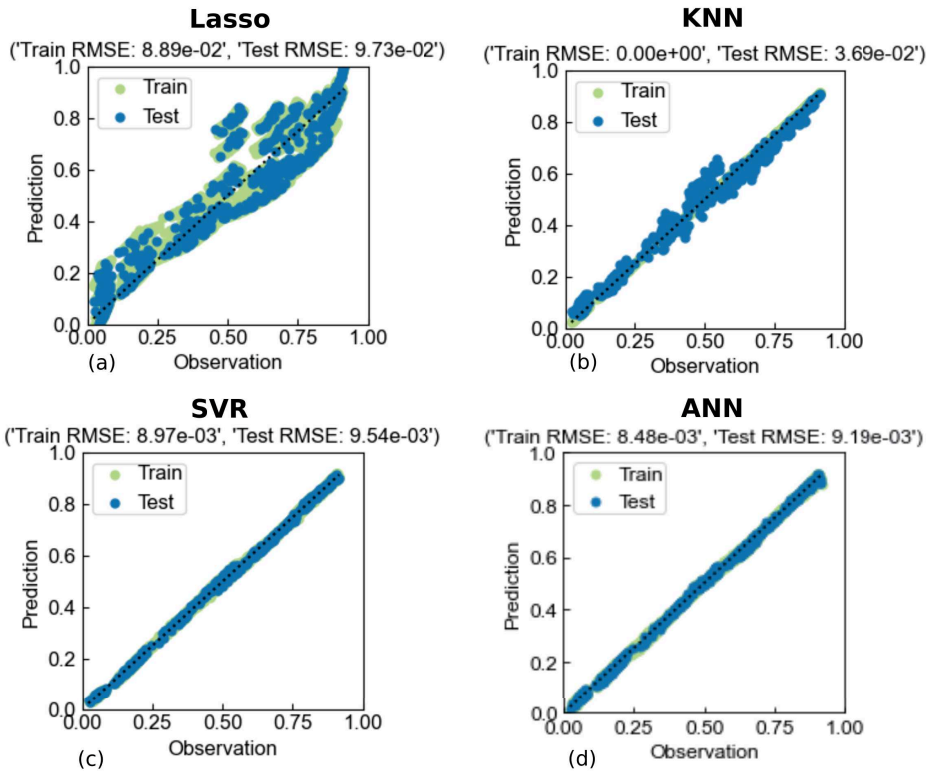



Figure 5.7: Parity plot of conversion for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression (d).

Table 5.2: 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** | $8.9 \times 10^{-2}$ | $9.7 \times 10^{-2}$ | $8.0 \times 10^{-2}$ | $8.6 \times 10^{-2}$ |
| **KNN** | 0 | $3.7 \times 10^{-2}$ | 0 | $2.7 \times 10^{-2}$ |
| **SVR** | $8.9 \times 10^{-3}$ | $9.5 \times 10^{-3}$ | $8.7 \times 10^{-3}$ | $9.0 \times 10^{-3}$ |
| **ANN** | $8.5 \times 10^{-3}$ | $9.2 \times 10^{-3}$ | $4.9 \times 10^{-3}$ | $5.0 \times 10^{-3}$ |

olefin selectivity of SVR and ANN are slightly lower than that of KNN[7] (Table 5.2). Further comparison of parity plots of methane and paraffin selectivity, obtained with these

[7] KNN 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.

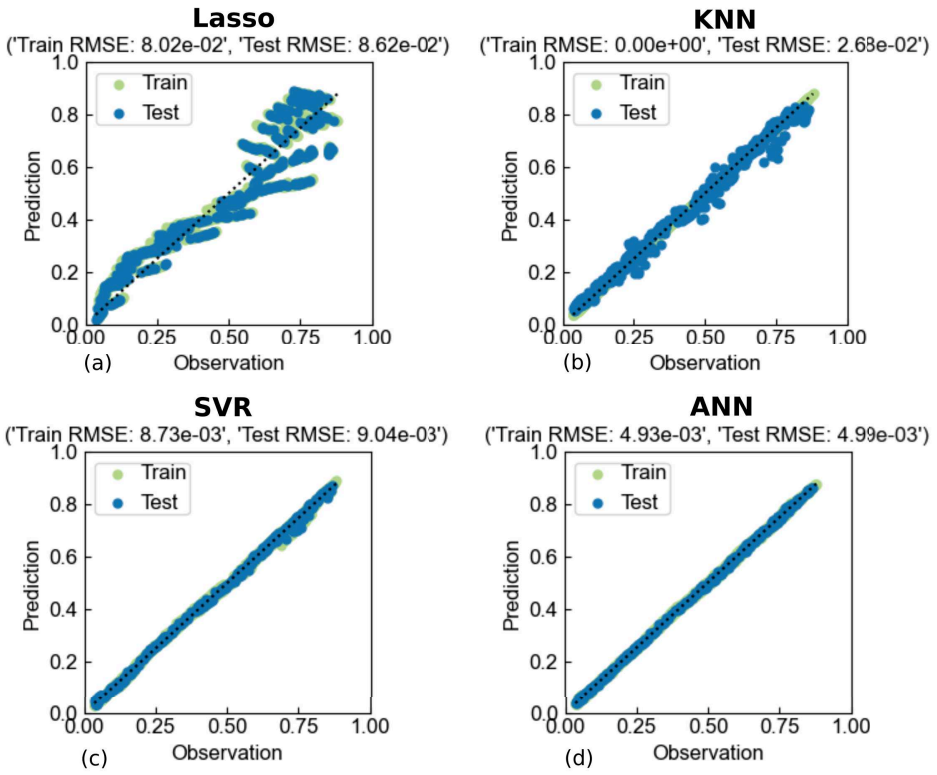



Figure 5.8: Parity plot of light olefin selectivity obtained for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression (d).

ML models are carried out to obtain the best candidate among the four models (see Appendix A). A comparison of residuals for the selectivity of methane, olefins, and paraffins is also reported in Appendix A, which shows that though the RMSE values are lower with all ML models, significant differences are observed while predicting lower values of output. The performance of these models is further assessed with the help of contour plot analysis discussed in the section below.

### 5.4.3. 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

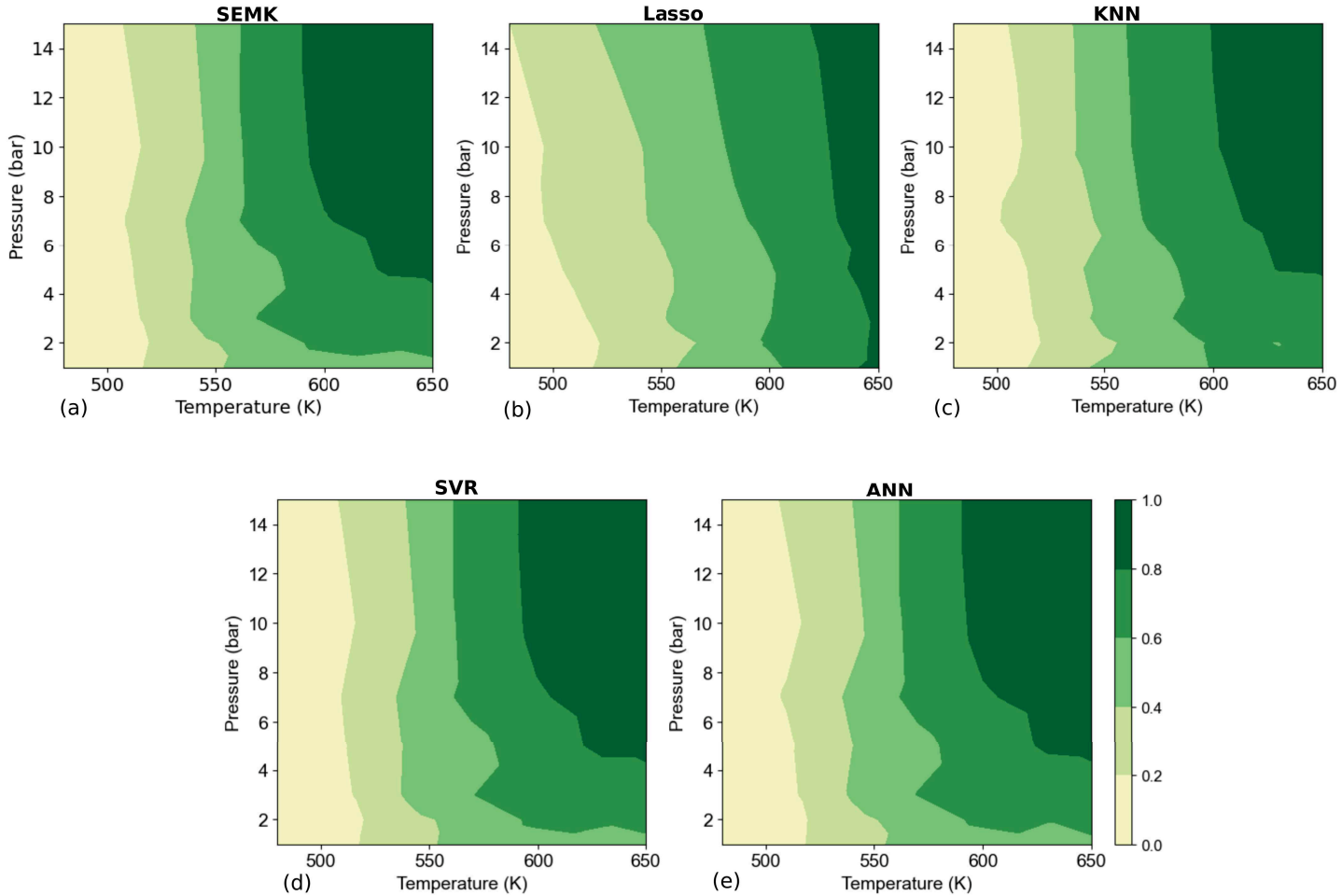



Figure 5.9: Contours of conversion obtained with SEMK (a) and with the machine learning models: Lasso regression (b), KNN regression (c), Support Vector Regression (d), and ANN regression (e) for the entire range of operating conditions.

detailed analysis of variation in conversion with the input variables, temperature and pressure is carried out on the testing dataset with the ML models. The CO conversion obtained at different temperatures and pressure using the microkinetic model (Fig.5.9(a)) is compared with the predictions using different machine learning models Fig.5.9(b-e) at the same operating conditions.

From Fig.5.9(a), it can 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 the plot with increasing temperature and with increasing 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.5.9(b), it can be observed that the prediction contours obtained with Lasso regression are linear in nature and the non-linear behavior of the conversion (as evident from the evolution of curvature of contours) as observed with the microkinetic model is not captured. The KNN model (Fig.5.9(c)) performs slightly better than Lasso regression, but the evolution of curvature of contours is still less pronounced than in the contours obtained by SEMK simulation. Conversely, the non-linear trend is captured well by SVR (Fig.5.9(d)), and ANN (Fig.5.9(e)), with ANN matching the SEMK results very closely.



### 5.4.4. ANALYSIS OF LIGHT OLEFIN SELECTIVITY USING MACHINE LEARNING MODELS

Similar to Fig.5.9 for the conversion, the light olefin selectivity at different temperatures and pressures obtained using the microkinetic model (Fig.5.10(a)) is compared to the predictions using different ML models Fig.5.10(b-e). Similar to the conversion plots discussed in the previous section, Fig.5.10(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 the plot with increasing temperature and with de-

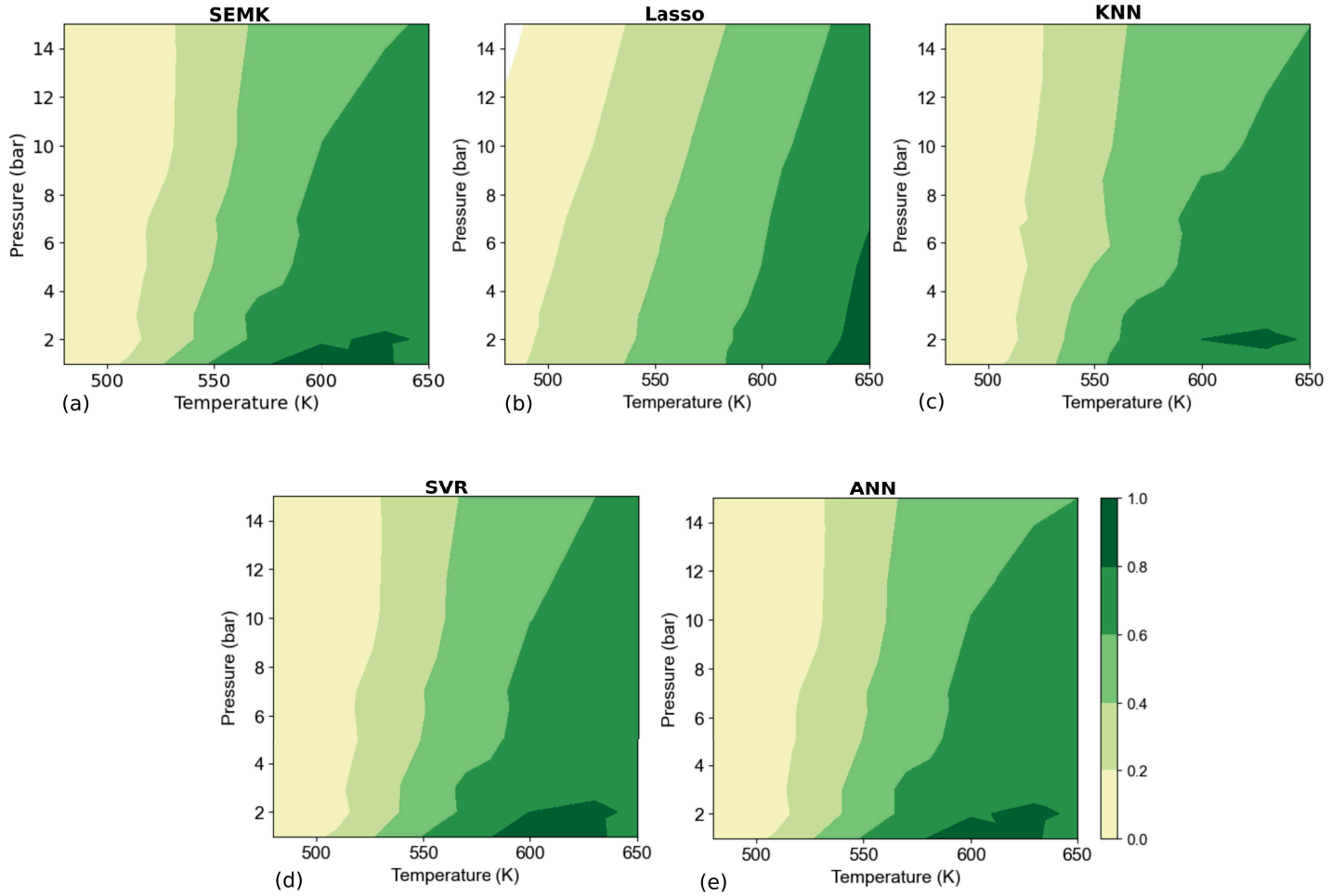



Figure 5.10: Contours of light olefin selectivity obtained with SEMK (a) and with the machine learning models: Lasso regression (b), KNN regression (c), Support Vector Regression (d), and ANN regression (e) for the entire range of operating conditions.

creasing pressure. Unlike in Fig.5.9, where conversion increases with pressure, here in Fig.5.10, 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.5.10(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.5.10(c)), SVR (Fig.5.10(d)), and ANN (Fig.5.10(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 as adequately by KNN, as compared to SVR and ANN, which hence out-perform the Lasso and KNN models. Thus, SVR and ANN models seem promising to make multi-output prediction, and match closely with SEMK results. ANN and SVR perform superior to the other methods owing to their inherent capability to handle non-linearity. In ANN, neurons have a weight that does a linear transformation on the input value, while the activation function does a non-linear transformation of the data. Similar to the ANN models, In SVR similarity functions (kernels) are used to transform input vectors into higher dimensional spaces. This enables solving a non-separable problem by converting it into a separable mathematical problem [65]. However it is to be noted that the SVR model requires a work-around using a wrapper function, as discussed in Section 5.2.2, to make multi-output predictions. It is also observed from Table 5.2 that the ANN model slightly outperforms the SVR model.



### 5.4.5. INTERPRETABILITY OF THE ANN MODEL

As discussed above, the model complexity increases from Lasso regression to ANN, meaning that the accuracy of prediction increases at the expense of interpretability. As a result, potentially a trade-off is to be made between the accuracy and interpretability of machine learning models. To mitigate this issue, model agnostic interpretation techniques

have been developed which explain the prediction results of complex models, and allow systemic analysis and ranking of the important contributing input variables [54]. 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 interpretability or importance of input variables on the FTS process aimed at selective light olefin production are discussed for the best performing ML model, i.e ANN by using Shap values.

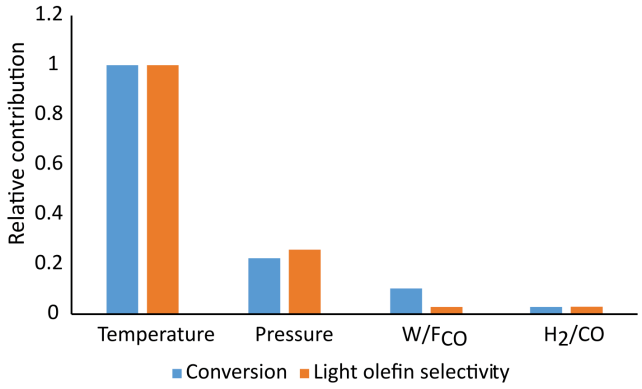
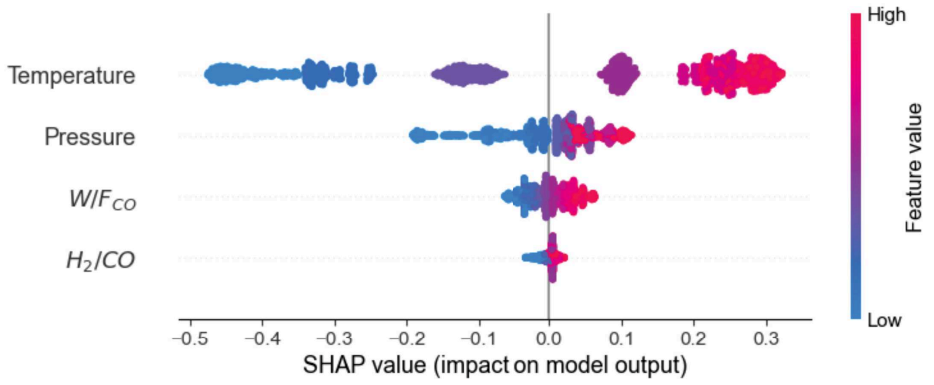


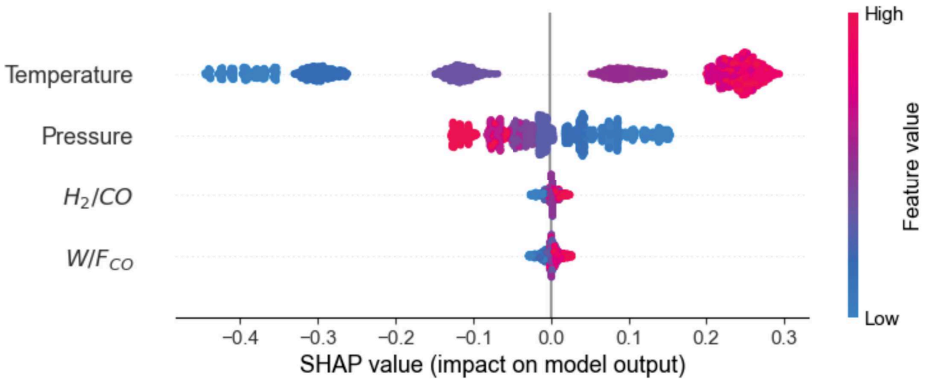

Figure 5.11: Shap bar plot indicating the global feature importance of the input variables on ANN prediction. The global feature importance is quantified in terms of contribution towards conversion and light olefin selectivity, from the entire training dataset, using the Python package, Shap [63].

From the Shap bar plot (Fig.5.11), the importance of each input variable on a global level could be observed. The input variables in Fig.5.11 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 (0.22x) > space-time (0.1x) > syngas ratio (0.03x). For the selectivity towards light olefins, the temperature remains the most contributing input variable followed by pressure. However, with respect to light olefin selectivity, the order of importance of space-time and syngas ratio becomes comparable. In the case of light olefin selectivity it follows the order: temperature (1x) > pressure (0.26x) > syngas ratio (0.03x) ≈ space time (0.03x). The order of importance obtained with Shap bar plot

is in line with the Pearson correlation matrix, Fig.5.4.




(a) Conversion




(b) Selectivity of light olefins

Figure 5.12: 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 [63].

In addition to the information on the global relative importance of each input variable towards conversion/selectivity from Fig.5.11, the relative importance at the local level is shown in Fig.5.12. The order of input variables in Fig.5.12, 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.5.12(a)) and light olefin selectivity (Fig.5.12(b)). The average conversion/light olefin selectivity in the training dataset, indicated by a Shap value of zero (in Fig.5.12(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.5.12 it can 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 can 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 opposite effect on light olefin selectivity. The dispersion in y axis in Fig.5.12(a)-(b), corresponding to a particular input feature is related to the less impact of other input features. For e.g. at higher temperature the influence of the input features like pressure, space time and syngas ratio is minimal. So, different combinations of these input features along with a fixed value of temperature leads to the same Shap value, thus leading to multiple dots of same colour (indicating same temperature) arranged vertically. These observations are in line with the results reported in Section 5.4.1, 5.4.3 and 5.4.4.



## 5.5. CONCLUSIONS

Multi-response machine learning models based on Lasso, K-Nearest Neighbors, Support Vector, and Artificial Neural Network regression were developed for Fischer-Tropsch

Synthesis using a synthetic dataset generated with a Single Event MicroKinetic model within a wide range of operating conditions. The non-linearity in the kinetic data could not be properly 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. Though the SVR and ANN models are promising, from the RMSE, ANN model slightly outperforms SVR, making it the preferred model for the current investigation as it matches closely with the results from the SEMK model.

The widely used interpretation technique based on Shap values on the (best preferred 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 follows the order of temperature (1x) > pressure (0.22x) > space-time (0.1x) > syngas ratio (0.03x). On the other hand, the influence of temperature (1x) and pressure (0.26x) remain the same for light olefin selectivity, but that of space-time (0.03x) and syngas ratio (0.03x) becomes comparable. The obtained results reproduce the insights obtained from the physico-chemical point of view with the SEMK methodology.



The present study opens up perspectives for kinetics data analysis with ML models. Indeed, ML models allow reproducing the insights gained based on fundamental models, such as the Single-Event MicroKinetic one. Hence, provided that the ML model can be trained against a sufficiently extended data set[8], it will allow acquiring essential insights directly on these data, i.e., without the need for fundamental modeling. In this respect, the definition of "sufficiently extended" will need to be evaluated in detail.

[8]The data set employed in chapter 5 (multi-response) is more than 10 times more extended compared toas that of the number of data points discussed in chapter 4 (single response). This has guaranteed to achieve a good accuracy of the models developed as part of chapter 5. The impact of the number of data points to achieve a reasonably accurate result (especially for multi-response catalysis problems) needs to be further evaluated in detail.

# REFERENCES

[1] J. H. Lee, J. Shin, and M. J. Realff, *Machine learning: Overview of the recent progresses and implications for the process systems engineering field,* Computers & Chemical Engineering **114**, 111 (2018).

[2] Y. Yan, T. Borhani, and P. Clough, *Machine learning applications in chemical engineering,* Mach Learn Chem **17**, 340 (2020).

[3] N. A. Jose, M. Kovalev, E. Bradford, A. M. Schweidtmann, H. C. Zeng, and 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**, 131345 (2021).

[4] R. Upadhya, S. Kosuri, M. Tamasi, T. A. Meyer, S. Atta, M. A. Webb, and A. J. Gormley, *Automation and data-driven design of polymer therapeutics,* Advanced Drug Delivery Reviews **171**, 1 (2021).

[5] L. C. Xue, D. Dobbs, A. M. Bonvin, and V. Honavar, *Computational prediction of protein interfaces: A review of data driven methods,* FEBS Letters **589**, 3516 (2015).

[6] A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne, and A. A. Lapkin, *Machine learning meets continuous flow chemistry: Automated optimization towards the pareto front of multiple objectives,* Chemical Engineering Journal **352**, 277 (2018).

[7] C. Casado, J. Moreno-SanSegundo, I. D. la Obra, B. E. García, J. A. S. Pérez, and J. Marugán, *Mechanistic modelling of wastewater disinfection by the photo-fenton process at circumneutral pH,* Chemical Engineering Journal **403**, 126335 (2021).

[8] C. I. Méndez and J. Ancheyta, *Modeling and control of a Fischer-Tropsch synthesis fixed-bed reactor with a novel mechanistic kinetic approach,* Chemical Engineering Journal **390**, 124489 (2020).

[9] P. J. Donaubauer, D. M. Melzer, K. Wanninger, G. Mestl, M. Sanchez-Sanchez, J. A. Lercher, and O. Hinrichsen, *Intrinsic kinetic model for oxidative dehydrogenation of ethane over MoVTeNb mixed metal oxides: A mechanistic approach,* Chemical Engineering Journal **383**, 123195 (2020).

[10] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Fischer-tropsch synthesis: Development of a microkinetic model for metal catalysis,* Oil & Gas Science and Technology - Revue de l'IFP **61**, 489 (2006).

[11] M. R. Dobbelaere, P. P. Plehiers, R. V. de Vijver, C. V. Stevens, and K. M. V. Geem, *Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats,* Engineering **7**, 1201 (2021).



[12] D. Greene, P. Cunningham, and R. Mayer, *Unsupervised learning and clustering,* in *Machine Learning Techniques for Multimedia* (Springer Berlin Heidelberg) pp. 51–90.

[13] M. Dash, H. Liu, and J. Yao, *Dimensionality reduction of unsupervised data,* in *Proceedings ninth ieee international conference on tools with artificial intelligence* (IEEE, 1997) pp. 532–539.

[14] M. Castelli, L. Vanneschi, and Á. R. Largo, *Supervised learning: Classification,* in *Encyclopedia of Bioinformatics and Computational Biology* (Elsevier, 2019) pp. 342–349.

[15] X. Zheng, Z. Jiang, Z. Ying, J. Song, W. Chen, and 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**, 117609 (2020).

[16] X. Zhao, X. Li, H. Lu, H. Yue, C. Liu, S. Zhong, K. Ma, S. Tang, and 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,* Chemical Engineering Journal **422**, 130389 (2021).

[17] Q. Dai, L. Xie, Z. Guo, L. Ma, J. Yang, N. Ren, G. Tian, and P. Ning, *Research on the mechanism of synergistic-dehydration/detoxification for sludge under treatment with double-modified potato residue,* Chemical Engineering Journal **420**, 127699 (2021).

[18] S. Bansal, S. Roy, and F. Larachi, *Support vector regression models for trickle bed reactors,* Chemical Engineering Journal **207-208**, 822 (2012).

[19] A. Nandy, C. Duan, J. P. Janet, S. Gugler, and H. J. Kulik, *Strategies and software for machine learning accelerated discovery in transition metal chemistry,* Industrial & Engineering Chemistry Research **57**, 13973 (2018).

[20] D. S. Palmer, N. M. O'Boyle, R. C. Glen, and J. B. O. Mitchell, *Random forest models to predict aqueous solubility,* Journal of Chemical Information and Modeling **47**, 150 (2006).

[21] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, and J. B. O. Mitchell, *Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization,* Journal of Chemical Information and Modeling **46**, 2412 (2006).

[22] N. S. Kaveh, F. Mohammadi, and 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**, 161 (2009).

[23] F. Kartal and U. Özveren, *A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus,* Energy (2020), 10.1016/j.energy.2020.118457.

[24] 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, and T. Tani-

ike, *The rise of catalyst informatics: Towards catalyst genomics,* ChemCatChem **11**, 1146 (2019).

[25] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, *Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions,* IEEE Access **7**, 137184 (2019).

[26] C. Molnar, *Interpretable machine learning : a guide for making Black Box Models interpretable* (Lulu, Morisville, North Carolina, 2019).

[27] L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike, and K. Takahashi, *Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts,* Chemical Science **12**, 12546 (2021).

[28] J. M. Garcia and M. L. Robertson, *The future of plastics recycling,* Science **358**, 870 (2017).

[29] M. E. Dry, *The fischer–tropsch process: 1950–2000,* Catalysis Today **71**, 227 (2002).

[30] M. Dry, *Catalytic aspects of industrial fischer-tropsch synthesis,* Journal of Molecular Catalysis **17**, 133 (1982).

[31] Q. Zhang, J. Kang, and Y. Wang, *Development of novel catalysts for fischer-tropsch synthesis: Tuning the product selectivity,* ChemCatChem **2**, 1030 (2010).

[32] E. Iglesia, S. C. Reyes, R. J. Madon, and 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.

[33] B. Gu, V. V. Ordomsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, and 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**, 153 (2018).

[34] A. J. Barrios, B. Gu, Y. Luo, D. V. Peron, P. A. Chernavskii, M. Virginie, R. Wojcieszak, J. W. Thybaut, V. V. Ordomsky, and 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**, 119028 (2020).

[35] A. Aguirre, E. Scholman, J. van der Shaaf, and 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**, 128139 (2021).

[36] K. Toch, J. Thybaut, and G. Marin, *Ethene oligomerization on ni-SiO2-al2o3: Experimental investigation and single-event MicroKinetic modeling,* Applied Catalysis A: General **489**, 292 (2015).

[37] T. von Aretin and O. Hinrichsen, *Single-event kinetic model for cracking and isomerization of 1-hexene on ZSM-5,* Industrial & Engineering Chemistry Research **53**, 19460 (2014).

[38] J. M. Martinis and G. F. Froment, *Alkylation on solid acids. part 2. single-event kinetic modeling,* Industrial & Engineering Chemistry Research **45**, 954 (2006).

[39] G. Lozano-Blanco, K. Surla, J. Thybaut, and 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**, 423 (2010).

[40] J. V. Belleghem, C. Ledesma, J. Yang, K. Toch, D. Chen, J. W. Thybaut, and G. B. Marin, *A Single-Event MicroKinetic model for the cobalt catalyzed Fischer-Tropsch Synthesis,* Applied Catalysis A: General **524**, 149 (2016).

[41] A. Chakkingal, L. Pirro, A. C. da Cruz, A. J. Barrios, M. Virginie, A. Y. Khodakov, and 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**, 129633 (2021).

[42] M. Esfandyari, M. Amiri, and M. K. Salooki, *Neural network prediction of the fischer-tropsch synthesis of natural gas with co (III)/al2o3 catalyst,* Chemical Engineering Research Bulletin **17** (2015), 10.3329/cerb.v17i1.22915.

[43] A. Chakkingal, P. Janssens, J. Poissonnier, A. J. Barrios, M. Virginie, A. Y. Khodakov, and J. W. Thybaut, *Machine learning based interpretation of microkinetic data: a fischer–tropsch synthesis case study,* React. Chem. Eng. **7**, 101 (2022).

[44] A. Mirzaei, S. Golestan, and S.-M. Barakati, *Prediction of fe-co-mn/mgo catalytic activity in fischer-tropsch synthesis using nu-support vector regression,* Physical Chemistry Research **4** (2016), 10.22036/pcr.2016.14776.



[45] D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong, and X. Shen, *Survey on multi-output learning,* IEEE Transactions on Neural Networks and Learning Systems , 1 (2019).

[46] M.-C. Kang, D.-Y. Yoo, and R. Gupta, *Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete,* Construction and Building Materials **266**, 121117 (2021).

[47] H. A. Garona, F. M. Cavalcanti, T. F. de Abreu, M. Schmal, and 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**, 129003 (2021).

[48] R. Tibshirani, *Regression shrinkage and selection via the lasso,* Journal of the Royal Statistical Society: Series B (Methodological) **58**, 267 (1996).

[49] Y. Song, J. Liang, J. Lu, and X. Zhao, *An efficient instance selection algorithm for k nearest neighbor regression,* Neurocomputing **251**, 26 (2017).

[50] G. Smits and 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).

[51] W. S. Sarle, *Neural networks and statistical models,* (1994).

[52] M. Mishra and M. Srivastava, *A view of artificial neural network,* in *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)* (IEEE, 2014).

[53] L. S. Shapley, *Stochastic Games,* Proceedings of the National Academy of Sciences **39**, 1095 (1953).

[54] S. M. Lundberg and S. I. Lee, *A unified approach to interpreting model predictions,* Advances in Neural Information Processing Systems **2017-December**, 4766 (2017), arXiv:1705.07874 .

[55] J. Thybaut and G. Marin, *Single-event MicroKinetics: Catalyst design for complex reaction networks,* Journal of Catalysis **308**, 352 (2013).

[56] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, *Single-Event Microkinetic Model for Fischer-Tropsch Synthesis on Iron-Based Catalysts,* Industrial & Engineering Chemistry Research **47**, 5879 (2008).

[57] H. Adib, R. Haghbakhsh, M. Saidi, M. A. Takassi, F. Sharifi, M. Koolivand, M. R. Rahimpour, and S. Keshtkari, *Modeling and optimization of fischer–tropsch synthesis in the presence of co (III)/al2o3 catalyst using artificial neural networks and genetic algorithm,* Journal of Natural Gas Science and Engineering **10**, 14 (2013).

[58] J. Brownlee, *Machine learning mastery with python,* Machine Learning Mastery Pty Ltd **527**, 100 (2016).

[59] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, *Knn model-based approach in classification,* in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (Springer, 2003) pp. 986–996.

[60] B. Boehmke and B. M. Greenwell, *Hands-on machine learning with R* (CRC Press, 2019).

[61] L. Breiman, *Random forests,* Machine learning **45**, 5 (2001).

[62] S. Lundberg and S.-I. Lee, *An unexpected unity among methods for interpreting model predictions,* arXiv , 1 (2016), arXiv:1611.07478 .

[63] *Welcome to the shap documentation¶*, .

[64] *scikit-learn machine learning in python,* .

[65] B. Yildiz, J. Bilbao, and A. Sproul, *A review and analysis of regression and machine learning models on commercial building electricity load forecasting,* Renewable and Sustainable Energy Reviews **73**, 1104 (2017).

# 6

# CONCLUSION AND OUTLOOK

Fischer Tropsch synthesis (FTS) represents a compelling reaction for syngas valorization generated from plastic waste gasification. Despite having been investigated for decades, several challenges remain of which the steering the catalyst's selectivity are of primary interest. The present work aims at model-based catalyst design and process optimization, specifically for the synthesis of light olefins via the FTS reaction. These components are of interest for the production of plastics and would, hence, contribute to the implementation of a circular economy via chemical recycling. However, the current mechanistic studies on catalytic processes marginally focus on the catalyst properties affecting their performance.

In Chapter 2, with screening and identification of relevant catalyst descriptors (i.e. atomic chemisorption enthalpies), a guideline for light olefin synthesis is proposed. The descriptor space for the analysis was identified using the experimental results, reported by Gu et al.[1], for Bi and Pb promoted and non-promoted Fe catalysts. These are used as a benchmark to explore the parametric catalyst descriptor space. To better understand

the impact of the catalyst descriptors on the catalyst performance, relative surface coverages were computed and a reaction pathway analysis was performed and the catalyst properties that are responsible for the it's selectivity could be identified using graphical visualizations (limited to 3-D domain). This enabled the identification of the preferential direction in the descriptor space to enhance light olefin production on a fundamental (microkinetic) level, and thus setting the path towards more promising FT catalysts. In a more general perspective, Chapter 2 showcases the possibilities and the value arising from the application of microkinetics to heterogeneous catalysis, by providing guidance for improved catalyst design based on the insights obtained through the modelling efforts.

In Chapter 3, the parameter-space limitations of visualization-based methods (i.e., 3D) are dealt with by applying a clustering approach (unsupervised machine learning) to the microkinetic simulations for the identification of optimal catalyst properties in an N-dimensional parameter space. The presented methodology effectively automates the aforementioned 3D process and thereby offers valuable insights in the, potentially correlated, non-linear behavior of the catalyst descriptors.



Due to the ever-increasing computational capacity and, in particular, the amount of available data, has popularized the use of Machine Learning (ML) techniques (such as Artificial Neural Network (ANN)) for modelling complex non-linear process phenomena in the field of chemical engineering, as well as to explore kinetic data. The contemporary used methods for kinetic experimental data are based on detailed (micro-)kinetic models which demands expert-level knowledge and a vast amount of manual effort. Also, these kinetic models are reaction-specific and the interpretation of their outcome has to be done in a case-specific manner. The studies using ML models have their own challenges. For instance, some of these ML models are hard to interpret and, hence, are denoted as "black-box" models. To fully relying on such models requires an understanding of why and how decisions are made using a non-classical contribution analysis. The

analysis of kinetic data based on ML models and interpretation techniques provide a more systematic approach to obtain more insight into the kinetic process. Chapter 4 aims at understanding the decision-making process of such a "black-box" model based on ANN, using model-agnostic interpretation techniques that are widely used in data science. A case study is put forward involving kinetic data of the FTS. The methane yield was targeted in a single dominant output scenario and was the therefore the focus when evaluating the interpretability of the ANN model. The kinetic data was generated using experimentally validated microkinetic simulations by varying the operating conditions: space-time, temperature and molar inlet ratio of CO and hydrogen. These inputs were then used to explore the possibility of systematically interpreting the kinetic data, using the interpretation techniques such as permutation importance, Shap-values and partial dependence plots. The input features could then be ranked according to their impact on the output (i.e., methane yield) and as such shed light on the complex decision-making process made by the multi-layer ANN model. The presented ML-based analysis can therefore act as an intermediate step to reduce the barrier to perform a full-scale and systematic data analysis.



The study was extended to multi-output response prediction in Chapter 5 with primary focus on light olefin production at different operating conditions in the FTS. The non-linear behavior of conversion and selectivity profiles as a function of the process variables as generated by an SEMK model is compared to the output of four machine learning models: Lasso regression, K-nearest neighbor regression, Support vector regression and ANN regression. It was found that the ANN prediction was closely matched the SEMK benchmark. The ANN-based model was then further analyzed using the Shap interpretation technique to investigate the feature importance on the multi-response outcome. This dual SEMK and ML approach yielded key insights which can guide catalyst design strategies as well as elucidate optimal FTS operating conditions aimed at light olefin production. It was also shown that (high-throughput) experimental data could

be analyzed with machine learning models and this significantly reduces the computational and manual effort due to their simplicity as compared to a detailed mechanistic model. In conclusion, for large volumes of data, a preliminary analysis could be carried out using ML followed by an in-depth analysis using a mechanistic model.

It is clear that the full potential of the presented modelling methodology has not yet been explored. In the search for optimal catalyst and process performance, several opportunities for optimizing the prediction process remain. For instance:

- The comparative study between various ML models was carried out within the framework of a synthetic (*in-silico*) SEMK-generated dataset. However, real experimentation often results in a higher variability within the obtained data. An extensive investigation of the applicability of ML models could be carried out on such data to evaluate the impact of such variability in the ultimate performance of these ML models.



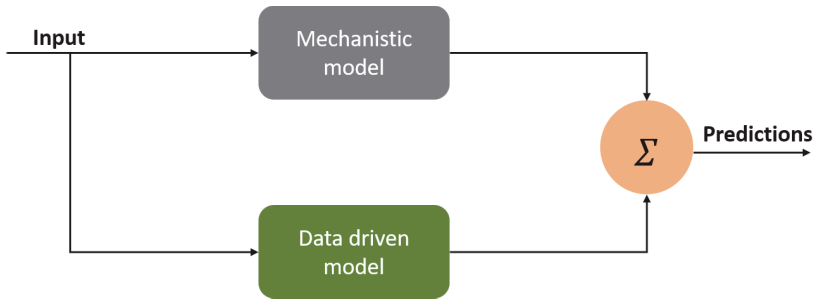



Figure 6.1: Schematics of a hybrid model.

- A hybrid approach (Fig.6.1) comprising a mechanistic and a ML model could bring together the best of both worlds. Thus, we can realize a trade off between the benefit of prior knowledge (mechanistic aspects) when training data is scarce and with the flexibility of ML model approaches when training data is abundant. This would benefit result in achieving improved prediction-extrapolation capabilities, in particular for scenarios where the reaction mechanism is not yet fully understood, e.g. in catalyst deactivation, the loss in number of active sites could be modelled using

data driven model discovery (Python Pysindy package) method and consequently coupled with a mechanistic model.

## REFERENCES

[1] B. Gu, V. V. Ordomsky, M. Bahri, O. Ersen, P. A. Chernavskii, D. Filimonov, and 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**, 153 (2018).

# A

# SUPPLEMENTARY MATERIAL, ML MODELS WITH MULTIPLE OUTPUT

## A.1. PEARSON CORRELATION MATRIX

The figure and discussion below relate to the Pearson correlation matrix with conversion and yield of the different hydrocarbons.

The correlation plot (Fig.A.1) 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 yields is limited within the investigated range of operating conditions. The impact of total pressure on the FTS reaction varies according to the targeted output variable.

A

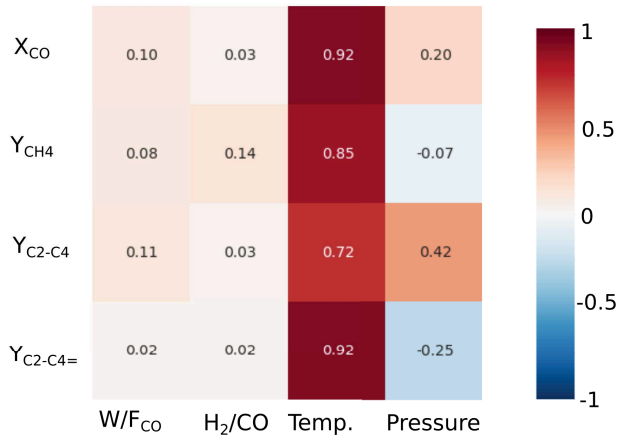



Figure A.1: Heatmap of Pearson correlation between input variables: temperature, pressure, space-time and syngas ratio with output variables: conversion and yield of methane, paraffins and light olefins.

## A.2. MULTI-OUTPUT PREDICTION USING DIFFERENT ML MODELS: METHANE SELECTIVITY

The figure and discussion below relate to the parity plot of methane selectivity for train and test data. The training dataset is indicated by the green colored dots in the parity diagrams whereas the blue dots correspond to the testing or prediction dataset. It is observed from Fig.A.2(a) , that lasso model deviates from the parity line for both the training dataset and the testing dataset. This indicates, simpler ml model like lasso is not sufficient for the analysis of multi response data. It is observed from Fig.A.2(b-d), as the model complexity increase, the model is able to fit the data.

## A.3. MULTI-OUTPUT PREDICTION USING DIFFERENT ML MODELS : PARAFFIN SELECTIVITY

The training dataset is indicated by the green colored dots in the parity diagrams whereas the blue dots correspond to the testing or prediction dataset. It is observed from Fig.A.3(a), that lasso model deviates from the parity line for both the training dataset and the test-

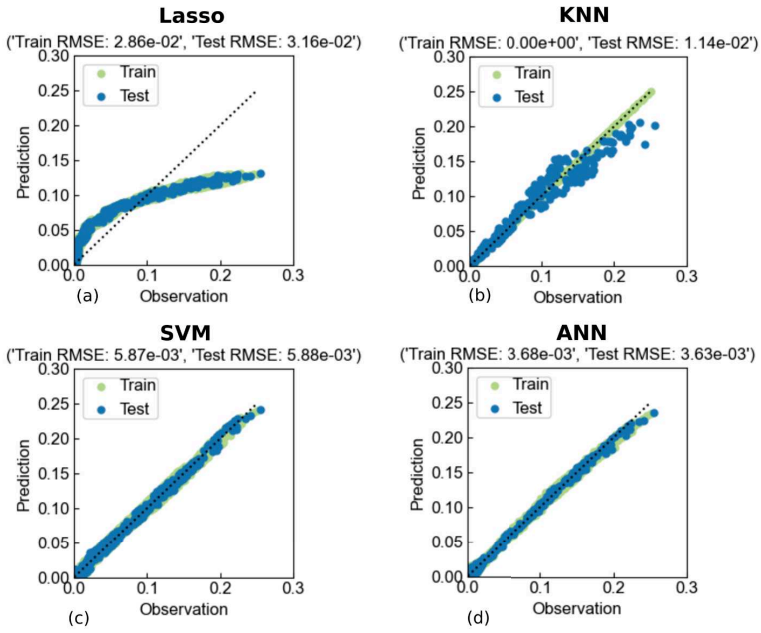



Figure A.2: Parity plot of methane selectivity for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression (d).

ing dataset. This indicates, simpler ml model like lasso is not sufficient for the analysis of multi response data. It is observed from Fig.A.3(b-d), as the model complexity increase, the model is able to fit the data.

## A.4. RESIDUAL PLOTS

The residual plots for conversion and selectivity of methane, paraffins and light olefins obtained with different ML models are plotted. The residual along y axis is standardized. The standardized residuals are calculated as:

$$Residual = \frac{y_{actual} - y_{predicted}}{y_actual}$$

A

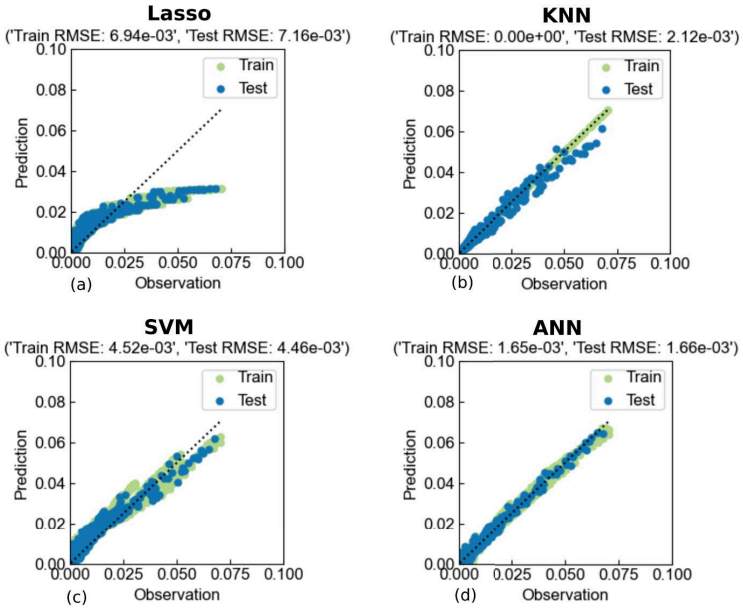



Figure A.3: Parity plot of paraffin selectivity for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression (d).

## A.5. R SQUARE VALUES

The R-squared values for conversion and selectivities obtained via different ML models are listed above. Since the values are high for obtaining the best model, the residual plots and contour plots are referred to. The nonlinearity in output is best captured by the contour plots.

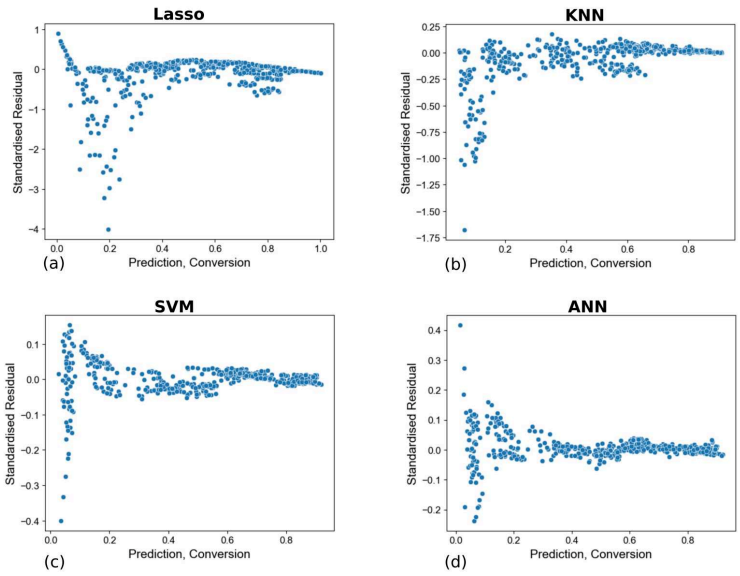



Figure A.4: Standardized residual plot of Conversion for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression(d).

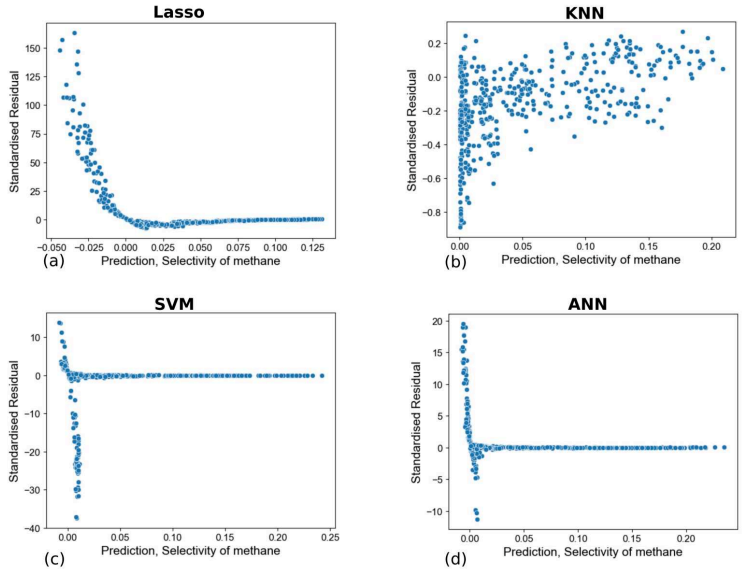



Figure A.5: Standardized residual plot of methane selectivity for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression(d).

A

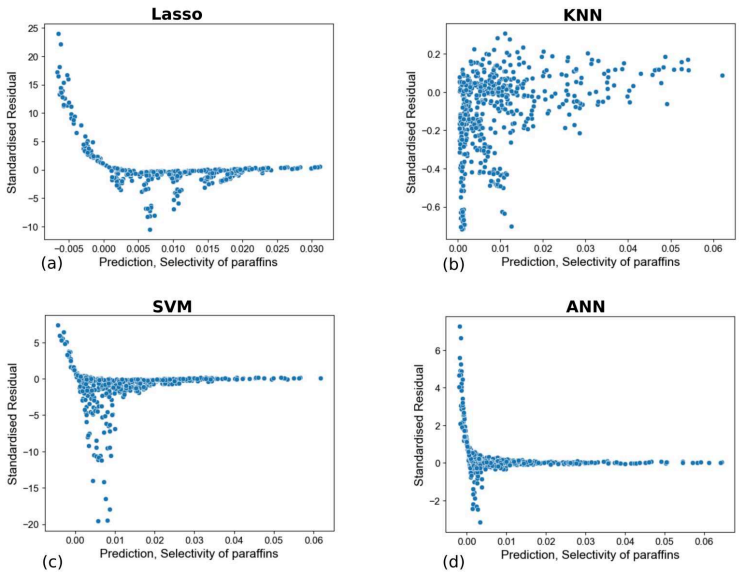



Figure A.6: Standardized residual plot of paraffin selectivity for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression(d).

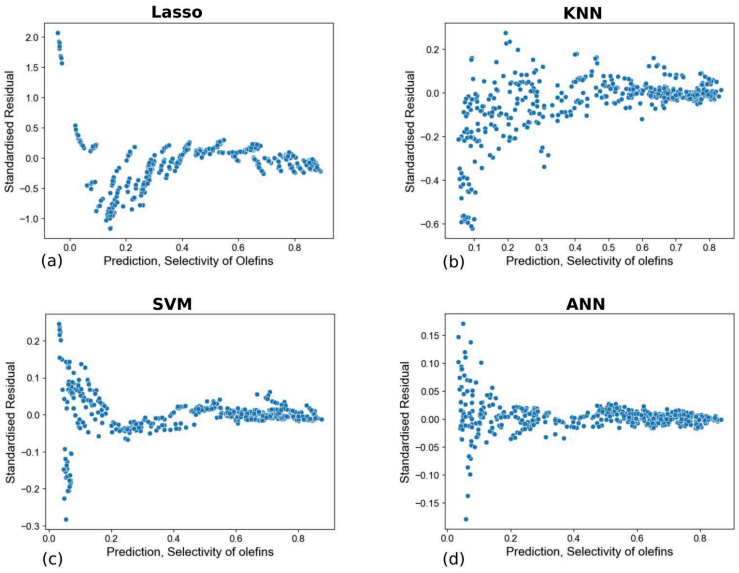



Figure A.7: Standardized residual plot of light olefin selectivity for the machine learning models: Lasso regression (a), KNN regression (b), Support Vector Regression (c), and ANN regression(d).

Table A.1: R square values of conversion and selectivity obtained using different ML models for train and test data.

| R square values | Lasso | KNN | SVR | ANN |
|---|---|---|---|---|
| **Conversion** | | | | |
| Train | 0.90 | 1 | 0.99 | 0.99 |
| Test | 0.87 | 0.98 | 0.99 | 0.99 |
| **Selectivity methane** | | | | |
| Train | 0.75 | 1 | 0.99 | 0.99 |
| Test | 0.73 | 0.96 | 0.99 | 0.99 |
| **Selectivity paraffins** | | | | |
| Train | 0.70 | 1 | 0.87 | 0.99 |
| Test | 0.67 | 0.97 | 0.87 | 0.99 |
| **Selectivity light olefins** | | | | |
| Train | 0.91 | 1 | 0.99 | 0.99 |
| Test | 0.89 | 0.99 | 0.99 | 0.99 |

# B

# DERIVATION OF UBI/QEP CHEMISORPTION ENTHALPIES

The detailed derivation of the UBI/QEP chemisorption enthalpy for a weakly chemisorbed AB molecule is given here as example. Assuming that the molecule is bonded to a metal via the A atom, the $M-B$ interactions can be neglected. Hence, the UBI/QEP potential energy expression is

$$E_{pot}(n) = Q_{0A} \sum_{i=1}^{n} \left( x_{A,i}^2 - 2x_{A,i} \right) + D_{AB} \left( x_{AB}^2 - 2x_{AB} \right) \tag{B.1}$$

Where n is the coordination site, $Q_{0A}$ is the chemisorption enthalpy of an atom A on an on-top site (n=1), x is the bond order and $D_{AB}$ is the required enthalpy to break the AB bond. Assuming that the contact atom A is situated in the center of the binding site, i.e.,$x_{A,1} = x_{A,2} = x_{A,3} = ...$ etc, the previous expression remains as:

$$E_{pot}(n) = nQ_{0A}\left(x_{A,1}^2 - 2x_{A,1}\right) + D_{AB}\left(x_{AB}^2 - 2x_{AB}\right) \tag{B.2}$$

The optimization is carried out accounting for the UBI/QEP constraint, i.e.:

$$nx_{A,1} + x_{AB} = 1 \tag{B.3}$$

Substituting now $x_{AB}$ from equation B.3 into equation B.2 and setting the derivative of the potential energy with respect to the bond order $x_{A,1}$ equal to zero, the result is the bond order $x_{A,1}$ at the potential minimum:

$$x_{A,1} = \frac{Q_{0A}}{(nD_{AB} + Q_{0A})} \tag{B.4}$$

Substituting this value into the potential energy expression, equation B.2 and subtracting the enthalpy of formation of AB in the gas phase $D_{AB}$, results in the chemisorption enthalpy of AB on a coordination site n:

$$Q_{AB,n} = \frac{Q_{0A}^2}{(\frac{Q_{0A}}{n} + D_{AB})} \tag{B.5}$$

# LIST OF PUBLICATIONS

1. **A. Chakkingal**, L. Pirro, A.R.Costa 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 (2021) .

2. **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*, Reaction Chemistry & Engineering (2021).

3. A. J. Barrios, D. V. Peron, **A. Chakkingal**, A. Dugulan, K. Nakouri, J. Thuriot-Roukos, R. Wojcieszak,J. W. Thybaut, M. Virginie, A. Y. Khodakov. *Efficient Promoters and Reaction Paths in the CO2 Hydrogenation to Light Olefins over Zirconia-Supported Iron Catalysts*, ACS Catalysis (2022) .

4. **A. Chakkingal**, P. Janssens, J. Poissonnier, M. Virginie, A. Y. Khodakov, J. W. Thybaut *Multi-output machine learning models for kinetic data evaluation : A Fischer-Tropsch Synthesis Case Study*, Chemical Engineering Journal (2022).

# ABSTRACT


## TAILORING FISCHER TROPSCH SYNTHESIS PRODUCT SELECTIVITIES: INSIGHTS FROM MICROKINETIC MODELLING AND MACHINE LEARNING

Kinetic and data-driven machine learning models are useful tools to provide fundamen- tal insights on catalyst and process development. Fischer Tropsch synthesis (FTS) pro- cess constitute an important strategic process in chemical recycling of plastics. There has been attempts to model via kinetic and data driven approaches. In this work, the synergy between detailed Single event Microkinetic model (to guide catalyst design) and machine learning models are used to explore the FTS process, and achieve targeted lightolefin synthesis.



Keywords: *Fischer-Tropsch, Process, Alkenes Synthesis (Chemistry), Waste plastics, machine learning, Artificial intelligence*



## RÉGLAGE DE LA SÉLECTIVITÉ DE LA SYNTHÈSE FISCHER-TROPSCH : APERÇU DE LA MODÉLISATION MICROCINÉTIQUE ET DE L'APPRENTISSAGE AUTOMATIQUE

Les modèles d'apprentissage automatique cinétiques et basés sur les données sont des outils utiles pour fournir des informations fondamentales sur le développement de ca- talyseurs et de processus. Le procédé de synthèse Fischer Tropsch (FTS) constitue un procédé stratégique important dans le recyclage chimique des plastiques. Il y a eu des tentatives de modélisation via des approches cinétiques et axées sur les données. Dans ce travail, la synergie entre le modèle microcinétique détaillé à événement unique (pour guider la conception du catalyseur) et les modèles d'apprentissage automatique est uti- lisée pour explorer le processus FTS et réaliser une synthèse ciblée d'oléfines légères.



mots clés: *Fischer-Tropsch, Procédé, Alcènes Synthèse (Chimie), Matières plastiques Déchets, Apprentissage automatique, Intelligence artificielle*