

Review

Hydrocracking: A Perspective towards Digitalization

Esin Iplik *, Ioanna Aslanidou and Konstantinos Kyprianidis

School of Business, Society and Engineering, Mälardalen University, 721 23 Västerås, Sweden; ioanna.aslanidou@mdh.se (I.A.); konstantinos.kyprianidis@mdh.se (K.K.)

* Correspondence: esin.iplik@mdh.se

Received: 31 July 2020; Accepted: 27 August 2020; Published: 29 August 2020



Abstract: In a world of fast technological advancements, it is increasingly important to see how hydrocracking applications can benefit from and adapt to digitalization. A review of hydrocracking processes from the perspective of modeling and characterization methods is presented next to an investigation on digitalization trends. Both physics-based and data-based models are discussed according to their scope of use, needs, and capabilities based on open literature. Discrete and continuous lumping, structure-oriented lumping, and single event micro-kinetic models are reported as well as artificial neural networks, convolutional neural networks, and surrogate models. Infrared, near-infrared, ultra-violet and Raman spectroscopic methods are given with their examples for the characterization of feed or product streams of hydrocracking processes regarding boiling point curve, API, SARA, sulfur, nitrogen and metal content. The critical points to consider while modeling the system and the soft sensor are reported as well as the problems to be addressed. Optimization, control, and diagnostics applications are presented together with suggested future directions of interdisciplinary studies. The links required between the models, soft sensors, optimization, control, and diagnostics are suggested to achieve the automation goals and, therefore, a sustainable operation.

Keywords: hydrocracking; modeling; soft sensor; optimization; control; digitalization

1. Introduction

Petroleum processing is one of the oldest industries which can benefit technological improvements in automation and data analytics. With the increasing capability of these tools, thanks to the advances in computational power, problems of refineries can be addressed. One of the problems is the nature of the crude oil, a mixture of many compounds, the ratio of which depends on the source location. The processing of this mixture results in a diverse range of products. These products differ in structure and density.

In 2009, the European Council has added a low carbon fuel standard into its quality of petrol and diesel fuels directive to decrease the greenhouse gas emissions [1]. With the stricter environmental laws and legislations, market demand shifts towards lighter hydrocarbons. Therefore, heavier products have a lower market value and limited use. Production of lighter hydrocarbons is not a new interest in this industry; the coal liquefaction process dates back to the beginning of 20th century [2]. However, the increasing severity of the cracking motivates further research. From process development to retrofitting, from optimization to decision support systems, refinery research progresses in multiple domains for energy and resource efficiency.

Long hydrocarbons can be cracked into shorter chains with thermal (e.g., visbreaking, coking) or catalytic methods (e.g., fluid catalytic cracking, hydrocracking). Visbreaking is one of the oldest processes of thermal cracking. Oil is heated up to 500 °C, and after a short residence time, cracking reactions are interrupted before completion with quenching; therefore, this process has high yields of heavy distillates [3]. In the coking process, thermal cracking reactions are completed, but the products

need to be further processed for sulfur removal [4]. These drawbacks of the thermal cracking lead the industry towards catalytic cracking.

Although fluid catalytic cracking (FCC) units have a lower middle distillate yield compared to hydrocrackers (HC) [5], in European and U.S.A. refineries, FCC units have a higher total capacity than HCs. Therefore, the investment shows an increasing trend for HC. Table 1 shows the total capacity of FCC and HC units in EU countries and the U.S.A. for 2008 and 2015.

Table 1. Total capacity of FCC and HC in 2008 and 2015 [6,7].

	FCC Capacity (Mt/Year)		HC Capacity (Mt/Year)	
	2008	2015	2008	2015
EU	132	124	75	97
U.S.A.	368	353	103	133

The direction of the research supports the industrial trend. A search in Web of Science was performed for keywords “hydrocracking” and “fluid catalytic cracking”. The number of publications for each year is shown in Figure 1.

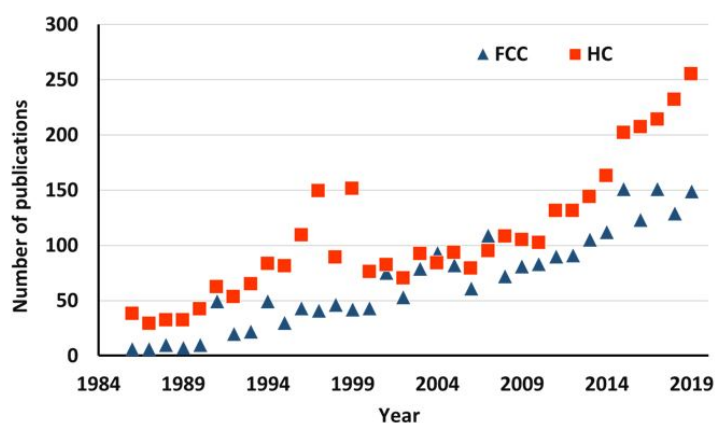


Figure 1. Number of publications in Web of Science.

A more in-depth analysis of the “hydrocrack*” search results in Web of Science (<http://webofknowledge.com>) shows that 30% of the research is about physical or applied chemistry, most of which are catalyst studies. Industrial applications of hydrocracking correspond to 58% of the published articles, and it is classified as chemical engineering research. Automation, control systems, and the interdisciplinary work of chemical engineering and computer engineering are represented by 1.2% of the publications. The rest of the research consists of environmental studies, biotechnology, nanoscience, and some other topics. Web of Science covers the publications from 1986 up to now, but there are highly cited articles from the 60s and 70s which reveal the reaction mechanisms of hydrocracking and provide the fundamentals for process development and modeling efforts [8–11]. After the catalyst research dominated the 80s, the 90s was the decade with high impact modeling works and process development [12–15]. In the last two decades, bio-oil hydrocracking gained importance [16–18]. Furthermore, the vast majority of the automation, optimization, and control research dates back to this period, which shows the focus shifting from process development to process improvement. Most of the publications originate from China, the U.S.A., and Europe. In Europe, France and Spain have the highest number of published articles. Next to them, Japan and Canada contributed to HC research, especially in the 90s.

According to the literature, the most recent review paper that covered the catalyst and process development was published in 2007 [19], where Rana et al. focused on the industrial applications

of hydrocracking and discussed each process route. Later, Casteneda et al. discussed the combined process schemes to improve the yield of heavy oil cracking [20]. Furthermore, a dedicated book chapter on hydrocracking catalysts was published in 2019 [21]. In the current study, process development and catalyst research is not covered, and the reader is referred to the referenced studies.

Considering the recent advances of HC, the authors have decided to focus on physics-based models, data-based models, soft sensors, and their integrated use for optimization, control, diagnostics, and decision support systems. The last review article on discrete and continuous lumping models was published in 2005 [22]. This paper gathered the most accepted physics-based modeling methods together and scrutinized their needs and applicability. However, since then, different models have been introduced to explain the particular behavior of newly developed hydrocracking processes. Aside from the first principle modeling, data-based models have been developed and utilized in industrial applications for real-time optimization and control due to their speed and simplicity.

Regardless of the type, development and utilization of models need quality data. Sensors are often used in hydrocracking operations for various measurements to acquire this data. Soft sensors are those, which do not measure the property directly but use a model to convert what is measured to what is desired. The review articles for soft sensors are analyzer specific. Some of the most important ones are near-infrared (NIR) spectroscopy [23] and nuclear magnetic resonance (NMR) spectroscopy [24].

Soft sensors are important to supply the required data to the models, and the model outputs are used in the optimization and advanced control systems. A review of refinery control systems by Ogden-Swift was published in 1996 [25], where soft sensors were suggested for the missing measurements because, as stated in the article, if a variable cannot be measured, it cannot be controlled. It was also claimed that the big oil companies, which invested in automation, advanced control, and real-time optimization, had significant amounts of savings thanks to these systems.

The structure of this paper is as follows. Brief information on different routes of industrial processes is first given to enable the reader to understand the necessities of the operation. Physics-based and data-based models are then discussed in detail, focusing on the need and expectation of the models versus available data, modeling, and computational effort. In the next section, the use of soft sensors to predict important characteristics of HC feed and products are shown. Reference methods which are relevant to similar feedstock are introduced to be able to specify the improvement capability of the soft sensor type. Finally, optimization, control, and diagnostics works are given, and possible future directions utilizing all the covered elements together for process improvement are discussed.

2. Models and Processes

Hydrocarbons are found in mixtures in nature, and crude oils consist of different amounts of heavy and light compounds. As previously mentioned, the heavy compounds are less desired; therefore, the need to crack these into lighter ones arises. Hydrocracking can be basically described as a chain of cracking reactions occurring with the help of a catalyst in an excess oxygen environment. In this process, C-C bond breaking reactions occur next to saturation reactions; and so, both the long chains and naphthenes (heavy cyclic compounds) are cracked into lighter molecules. Temperatures of 300–450 °C and pressures of 150–200 bar are preferred for these reactions; while the lower values within these ranges are sufficient for light feedstock, only the plants operating on the upper limits can process heavy mixtures. The fundamentals of hydrocracking, including pressure and temperature requirements for different feed properties and the important reaction mechanisms, can be found in [26]. The conditions which are required to favor hydrocracking reactions can be realized in different configurations. Figure 2 shows the different reactor types.

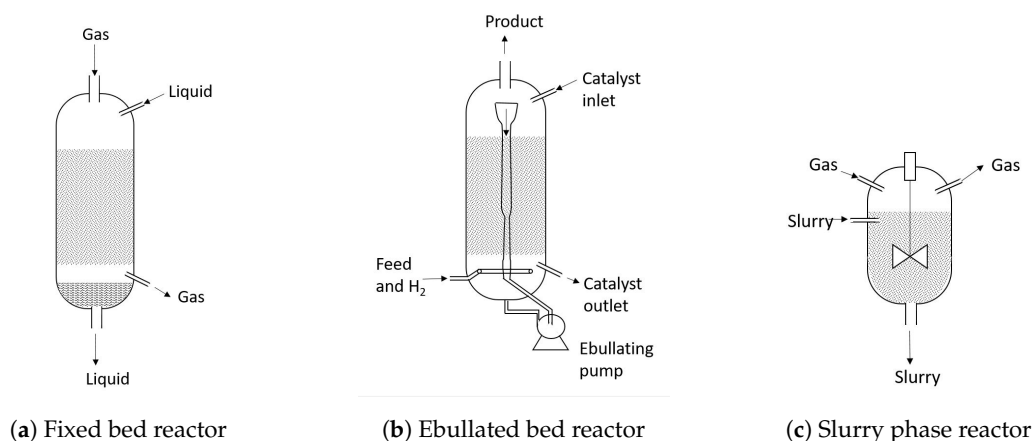


Figure 2. Three reactor designs used in hydrocracking.

A fixed bed reactor (FBR) is filled with the catalyst, which is to be used long term. In the years of operation, catalyst activity decreases significantly over time. As in its name, catalyst pellets are fixed in the tank such that a change of the deactivated catalyst requires a shutdown of the operation. Although FBRs were devised to process gas oil, currently, they are used to crack heavier feedstock due to the need. Another important reason for their use is the simplicity of both the design and the operation. Up to 50% feed conversion to middle distillates can be achieved depending on the feedstock, catalyst, and operation conditions. Mass transfer is the limiting factor of the conversion due to the size of catalyst particles, which are in a range of 1.2–3.0 mm [21].

As shown in Figure 2b, an ebullated bed reactor (EBR) has a moving catalyst bed which allows the addition of fresh catalyst while withdrawing the deactivated particles. EBRs can handle heavier feedstock than FBRs. Due to lower concern of pressure drop, the catalyst particles of this process are typically smaller, around 0.8 mm [21], which results in a higher conversion of up to 90%. The most significant disadvantage of EBRs in HC is the higher amount of catalyst need [27]. Al-Dalama and Stanislaus studied the deactivation pattern of FBR and EBR catalysts and gave an insight into excess catalyst use of EBRs [28]. A considerable dimension reduction was observed in the catalyst samples taken from such reactors. Moreover, it was claimed that the particles, which are not fully deactivated, were also withdrawn, which increases the need further. EBRs are more complicated to design and scale-up due to their hydrodynamics [29].

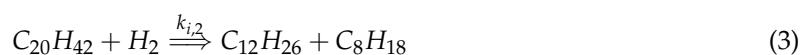
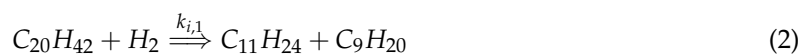
A slurry phase reactor (SPR) is an agitated tank reactor, which is designed to use very fine catalyst particles (0.002 mm) that are dispersed in the feed and form a slurry together [21]. Decreasing the particle size decreases the mass transfer limitation while increasing the conversion up to 90%; however, the recovery of such small particles is not preferred. Therefore, the material must be cheap and disposable. Only two refineries are operating a slurry phase hydrocracking on a commercial scale, one is in Venezuela using a local ore with high iron content [30], and the other one is in Italy using molybdenum disulfide as a catalyst [31]. Deeper information about slurry phase hydrocracking can be found in literature [32,33].

2.1. Physics-Based Models

Modeling the hydrocarbon mixture has been ongoing research in the long history of crude oil refining. These models are necessary to predict, simulate, and control the chemical reaction kinetics of a complex mixture. With the technological progress and the increasing need for cracking, more sophisticated processes gain importance. Therefore the need for detail increases. Most of the models use an entity called “lump” to describe multiple components at a time because hydrocarbon mixtures include hundreds of compounds that make it hard to have a reaction kinetic equation (Equation (1)), a mass balance and an energy balance for each and every one of them. Every compound can be cracked into smaller chains with different reactions (Equations (2) and (3)), so there are more reactions than

compounds in the mixture. The rate of reaction, r_i , is dependent on the concentration of the compound i , c_i , and the reaction constant, k_i . It is hard to measure or calculate each compound's concentration and the reaction constant for every possible reaction. Lumps have the same rate equation, the concentration is the total concentration of the compounds included in the lump, and the reactions represent the conversion of one lump into another.

$$r = k_{i,n}c_i \quad (1)$$



While the earlier models had a focus on the feed conversion (global lumping or single lumping) [9], later research shifted towards lumping the mixture according to their similarities (discrete lumping). Increasing the number of lumps, increased the amount of information that can be gathered at the expense of higher computational effort. One of the earliest lumped models was suggested by Mosby et al. with seven lumps [34]. Their suggestion of lumps, which served as a base network for other research, is given in Figure 3. With its three feed lumps (feed gas oil, soft and hard residue) and their conversion to the four product lumps (gas oil, middle distillate, naphtha, and gases), this model aimed to represent the inhomogeneous nature of the mixture using different chemical reaction kinetic coefficients. Every directed arc of the network shows a conversion of one lump into the other; therefore, the reaction network given in Figure 3 considers 14 reactions between its lumps with different kinetic coefficients.

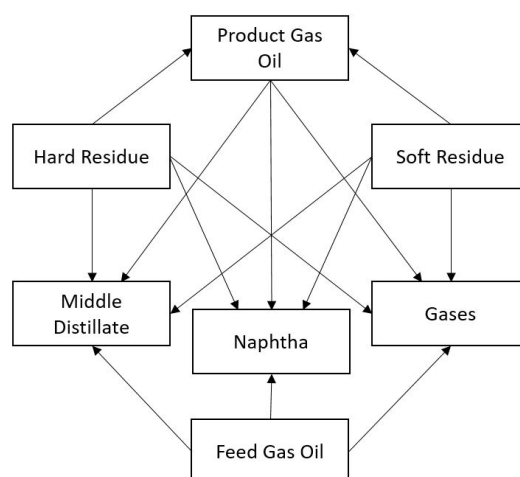


Figure 3. Seven-lump model by Mosby et al. [34].

Later, three variations with the reduced number of feed lumps were developed from the aforementioned seven-lump model [14]. Comparison of the original model and the reduced models proved better data fitting capabilities of the models with a higher number of lumps. Next to that, it was strongly emphasized that the kinetic coefficients of the lumping models are dependent on the feed and cannot be used for different feeds.

A work by Laxminarasimhan et al. suggested another method called continuous lumping, which was an approach taking the mixture as a continuum [12]. An integral term was used to represent the compounds according to their reactivity and enabled the model to carry the information of the entire concentration distribution. A similar approach was used to develop pseudo-component models [35]. The discretization of the integral term with respect to the reactivity leads to lumping, and these so-called pseudo-components comprised the real components in their neighborhood.

A detailed analysis of global lumping, discrete lumping, and continuous lumping models was done by Ancheyta et al. [22]. The models which were included in their work were Qader and Hill's

global lumping model [9], a three-lump model from Callejas and Martinez [36], a pseudo-component model from Stangeland [35] and the continuous lumping model by Laxminarasimhan et al. [12]. All four models were based on the mass and energy balances of the hydrocarbons. Discretely lumped models were systems of ordinary differential equations. The continuous approach led to a set of integro-differential equations. This work showed the capability of different kinds of models based on the same data such that it allowed a comparison of the trade-off between prediction accuracy and computational effort. The global lump model gives only a feed conversion prediction, and the three lump model gives a prediction for gases and light oils as its two product lumps. On the other hand, pseudo-component and continuous lumping models can give a detailed product distribution on the boiling point curve. Ancheyta et al. stated that the parameter estimation procedure was more complicated for the latter two models, requiring more data on the feed mixture characteristics. The detailed information of the product distribution might not always be necessary, and the time needed to obtain this information makes detailed models infeasible to be used for real-time approaches.

It should be noted that the data taken from El-Kady et al. [37] was used in the comparison study [22]. It belongs to a bench-scale FBR system that suffers from catalyst degradation. However, calculated parameters for the reaction rates do not take the degradation into account. According to Martinez et al. HC catalyst degradation is linear after the fast initial phase, and it should be modeled to have a better prediction accuracy [38]. In this study, they introduced an effectiveness factor into a discrete lumping kinetic equation to show the effects of catalyst deactivation. Another work showed the way to consider the deactivation of the catalyst for a continuous lumping model [39]. The model they used had both the fast initial activity loss and the linear degradation pattern following it.

A devoted work for the solution of continuous lumping models was published in 2011 [40]. Suggested methodology for the system of integro-differential equations with a probabilistic distribution (for cracked molecules) was discretization by using a linear approximation in short intervals and then using Lobatto quadrature and Runge-Kutta methods. Their concern about the linear approach was the number of intervals and its optimization. The use of a higher-order polynomial was also examined, and the possibility of oscillation was anticipated. To reach the conservation of mass, 150 intervals were employed. However, it should not be forgotten that this number can change according to the range of components in the real feed. An iterative procedure was suggested to find the optimal parameters. They pointed out the need for future research to distinguish the effects of thermal and catalytic reactions on the model parameters.

Parameter estimation studies are often formulated as optimization problems. When the system of equations is complex and highly parameterized, the solution needs a methodology. While the aforementioned work of Elizalde et al. [40] focused only on continuous lumping models, Alcazar and Ancheyta published their method suggestion also considering the other types of nonlinear models [41]. They pointed out the importance of a well-posed optimization problem and proposed a strategy that was different from a straight-forward parameter estimation. What made it distinct was the integration of sensitivity analysis into the optimization problem. They suggested an iterative use of perturbations on parameters, unlike the conventional way of utilizing it at the end for testing purposes. Another important suggestion was using the Monte Carlo method for the parameter initialization, which ensured a global parameter set.

Increasing computational power and the development of continuous models have not eliminated the use of discrete lumped models. They are faster than continuous models and flexible with varying numbers of lumps adjusted according to the need of the application. More recent works can be found in the literature with different numbers of lumps. In one study, a four-lump model was developed for a hydrocracker with a vacuum gas oil as its feed stream and middle distillates, naphtha, and gas as the product streams [42]. All six reactions of these four lumps were considered to represent the secondary cracking mechanisms (cracking of the product lumps with the increasing residence time), unlike previously mentioned seven-lump model considering the secondary cracking of only

the product gas oil stream. This network, given in Figure 4, was found to have a good data fitting capability, and later, it was used by other researchers.

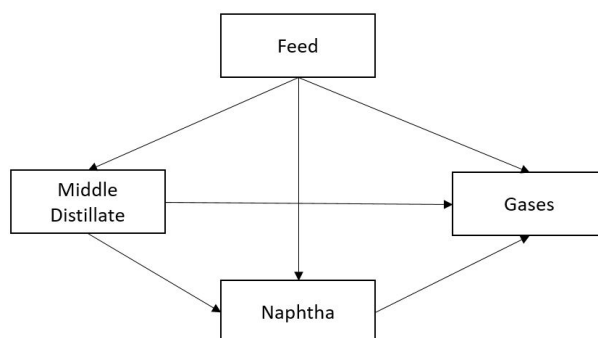


Figure 4. Four lump model by Valavarasu et al. [42] including secondary cracking mechanisms.

Most of these models use a boiling point curve approach, meaning that the lumps are based on the true boiling point (TBP) curve of the hydrocarbons. This curve shows the percentage of evaporated volume at certain temperatures and indirectly represents the distribution of compounds. An example is given in Figure 5. A less common criterion used for lumping is the carbon number. The reason of its rare usage might be the problems in measurements to be used in model validation. Balasubramanian and Pushpavanam showed its applicability for a continuous lumping system [43]. While the TBP based model had 40 kinetic parameters, the carbon number based model only had 12, and the latter model showed a higher error than the former.

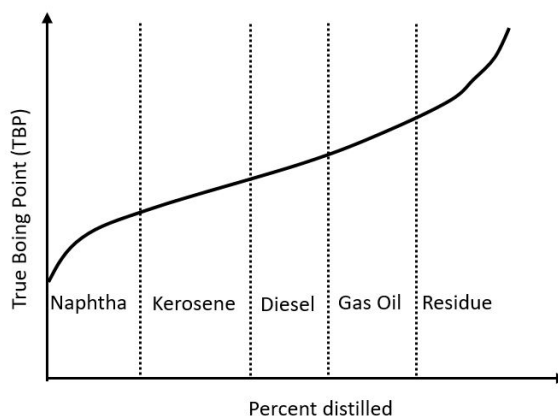


Figure 5. A representative true boiling point curve of crude oil.

A unique approach to hydrocracking is the single event microkinetic modeling (SEMM). Unlike any of the lumping methods, SEMM takes every compound into account separately and considers every possible reaction. Becker et al. compared SEMM against the continuous lumping concept [44]. They discussed the apparent disadvantage of any lumping model, feed dependency, and how this is not a problem for a single events model. However, SEMM has disadvantages, such as high computational effort and arduous solution procedures for single event coefficients' calculation and relumping. Although it was a comparison study, they stated that these models are not comparable; instead, they are complementary. The same group later published their work on a single event model for hydrocracking of a network of paraffins, naphthenes, and aromatics with fundamental observations on the reaction mechanisms of these compound classes [45]. It was shown that for conversion over 50%, a vast amount of the normal paraffins of longer than 15 carbon chains goes under isomerization

reactions to form iso-paraffins. Next to that, the lateral chain reactions were proven to be the most important mechanism of cracking. Even though SEMM is computationally heavy and requires detailed data, these findings cannot be observed using any other type of model.

Obtaining fundamental chemistry knowledge is not the only use of SEMMs. Tyhbaut and Marin discussed the utilization of these models for catalyst design [46]. They showed a selectivity and product yield adjustment by fine-tuning the catalyst according to the need by which the catalysts could be designed for specific feeds, products, or refineries.

Building a SEMM is not a trivial task. Next to the computational effort, it is highly demanding in terms of molecular analysis. In Figure 6, some C₈ compounds are given. With the increasing number of carbons, the number of compounds increases together with the analysis intricacy. Furthermore, these compounds differ in reactivity, which affects the rate coefficients of their reactions. Svoboda et al. built a network of more than 400 possible reactions of 154 compounds to describe octane hydrocracking [47]. A full-scale SEMM includes thousands of compounds and their reactions.

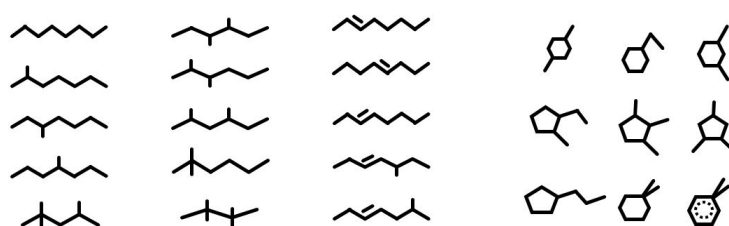







Figure 6. Some C₈ hydrocarbons.

Structure oriented lumping (SOL) offers a solution compromising from the detail for feasibility. Quann and Jaffe discussed the impracticality of SEMMs by giving the number of C₄₃ isomers [48]. A trillion and these are only the paraffins. Chromatographic methods are not capable of distinguishing each and every one of them in a mixture. SOL models map and partially lump the compounds by using a classification matrix according to their contents. A simple example is given in Table 2. As shown, isomers are lumped, but these models offer greater detail than traditional lumping methods. Quann and Jaffe built an SOL model for gas oil and lighter fractions [49], and later Jaffe et al. extended this model to include vacuum residua molecules with metal content [50].

Table 2. Structure oriented lumping example for some C₈ compounds.

Compound	−CH ₃	−CH ₂	Branch	Double Bond	Triple Bond	C-Hexane	Benzene
	0	8	0	0	0	0	0
	0	8	2	0	0	0	0
	0	8	0	1	0	0	0
	2	0	0	0	0	1	0
	2	0	0	0	0	0	1

As it was mentioned before, high conversion hydrocracking solutions have the disadvantage of their complex hydrodynamics for scale-up. To understand the effect of hydrodynamics on chemical reactions, Sadighi et al. developed a one-dimensional computational fluid dynamics (CFD) model for a four-lump discrete system [51]. Later, Sadighi developed a two dimensional model with the same approach [52]. Using the same lumps as [42], a reduced network excluding two secondary cracking reactions was also tested in this study. This network is given in Figure 7. It was shown that the reduced network could reach almost the same accuracy as the complete network, and the 2D results were proven to improve the prediction accuracy considerably. While the 1D model was enough to simulate the pressure drop, the 2D model showed that the radial velocity in this reactor is low.

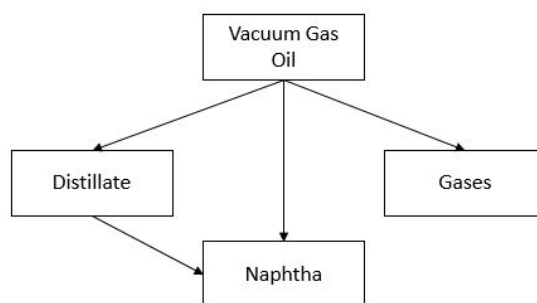


Figure 7. Reduced four-lump model by Sadighi [52].

CFD models are not the only works, which model the slurry phase reactor hydrodynamics. Calderon and Ancheyta used another approach, a dynamic axial dispersion model, at first 1D [53], then 2D [32]. In the same publication, they also analyzed the sensitivity of the model with respect to the axial dispersion coefficient, convective heat transfer coefficient, and the heat of reaction. The model showed a robust prediction trend with gradual changes in each parameter in the range of $\pm 50\%$. Next to parametric sensitivity, the effect of operating conditions was investigated. The process was found more sensitive to hydrogen to oil ratio than to space velocity. However, feed temperature, bed temperatures, and pressure had a more significant effect on conversion.

Physics-based models have no alternative when the physical meanings of the variables are crucial, such as in process design. A wide range of information can be obtained by using different types of models. The most suitable model can be picked only by analyzing the needed data next to the data in hand because the detailed models demand more data at the model building phase. Furthermore, it should be noted that these models need a high effort to build due to the underlying process understanding.

2.2. Data-Driven Models

In addition to the physics-based approach, HC is also modeled often with different data-driven techniques. This type of solution aims to decrease complexity and provide fast results, which can be used in real-time. One of the earliest examples was an artificial neural network (ANN) model by Elkamel et al. suggesting two different models according to the objective of the production [54]. The jet fuel and middle distillate models had a higher error in predicting C_4 and light naphtha yields. This model was also used to predict some critical product specifications like API gravity (relative density with respect to water), octane number, and smoke point next to product yields.

Bahmani et al. reported an ANN model predicting the product yields of an industrial hydrocracker [55]. In this work, they trained three different ANN models, compared their error and run time, and showed the network using more input variables performed better. Next to that, they compared the error values also with a physics-based four-lump model. According to the authors, the error was lower with the four-lump model, but the ANN model using various measurements of the unit was more versatile.

Sadighi and Zahedi did a similar comparison work with a six-lump model and an ANN model [56]. In this comparison, the ANN model performed better in product yield prediction, especially for diesel, but the residue prediction was found insufficient for both models. An axial-dispersion phenomenon that was not considered in the models was given as a possible explanation for the low accuracy of diesel yield prediction.

Later, another ANN model was proposed with better yield prediction capability [57]. By using the prediction of the yield of seven product streams and their prices, this work also demonstrated a careful

analysis based on the expected profit increase in real-time optimization (RTO). At the beginning of the catalyst life cycle, real control inputs did not show a different trend than the optimized values, which did not suggest an additional profit. However, with the aging catalyst, RTO suggestions deviated from the actual operating points. These points were shown to bring a considerable profit for this operation.

In another study, a deep neural network model for an industrial hydrocracker was investigated. A convolutional neural network (CNN) was compared with a feed-forward neural network [58]. The CNN model showed better prediction accuracy. This model was trained to predict the exit temperatures of each catalyst bed next to the product yields. More importantly, an updating strategy was proposed for the CNN model to be able to explain the catalyst degradation. For the models which do not take catalyst aging into account, this is an issue to be addressed to conclude the research gains.

An important use of models is steady-state detection. It is necessary to know where the stationary point of the operation is according to changing variables, disturbances, and even the noise. A data-driven steady-state detection model was proposed by Wang et al. based on a receding horizon principal component analysis (PCA) [59]. They reasoned their choice for PCA as the elimination of correlations between considered variables. As they stated, the importance of these variables might not be easy to calculate and lead to wrong identification. PCA minimizes this possibility. The demonstration of this model on a hydrocracker used 21 process variables and a five minute sampling time and gave better results than an R statistical steady-state detection model.

Data-driven models are very often used in industrial applications for their speed but rarely published, and most of the related literature focuses on the prediction of product characteristics instead of entire system representation. These models are given in the next section, characterization methods.

Surrogate modeling is another alternative to model complex systems in a more straight-forward way. However, the discontinuity in the operation space brings some difficulty and a compromise between conservative modeling or including the infeasible space in the model. Xue et al. suggested a model using continuous subspaces based on catalyst life [60]. This multi-block Kriging based surrogate was compared with a single Kriging model for process optimization. The results showed that the single block model optimization might lead to early catalyst degradation due to ignored catalyst life in its design space.

Table 3 summarizes the advantages and disadvantages of the different modeling methods used for hydrocracking. As mentioned earlier, different lumping methods are preferred due to the number of compounds involved in this process, and with the increasing number of lumps, more information can be gained from the models; however, this also increases the computational effort. SEMM is not a lumping model, and it is the most detailed model which can help contribute to fundamental understandings of the process. SOL uses a low level of lumping standing between traditional lumping methods and SEMM. Data-driven models are much faster than the physics-based models, and they are suitable for iterative calculations like optimization and control. There are different kinds of data-driven approaches, and it is hard to classify their advantages and disadvantages without the knowledge of the specific setting, but they have one thing in common, their dependency on quality data.

Table 3. Advantages and disadvantages of different modelling approaches.

Model	Advantage	Disadvantage
Global lumping	Fast	No product distribution
Discrete lumping	Flexible number of lumps	Limited product distribution
Pseudo-components	Full product distribution	Data need and arduous solution
Continuous lumping	Full product distribution	Data need and arduous solution
SEMM	Fundamental observations	Slow
SOL	Fundamental observations	Slow
Data-driven	Fast	No physical interpretation

3. Characterization Methods

The importance of fast and accurate detection of feedstock and product quality indicators increases with stricter environmental regulations. While the low-quality product needs further processing, thus consuming more energy, producing at a higher quality level than market specifications also means additional energy and resource consumption for this process due to the heat and hydrogen demand of the reactions. The range of deviation from optimality can be minimized by utilizing the sensors in a better way.

As mentioned earlier, the TBP curve (Figure 5) is often used to describe a hydrocarbon mixture. Reference methods of TBP curve characterization are ASTM D2892 [61] for atmospheric distillation and ASTM D5236 [62] for vacuum distillation. Atmospheric distillation has a temperature limit of 673 K because unwanted thermal cracking reactions start over this temperature limit. For the heavy residues, ASTM D5236 is used up to 838 K. Both of these methods are time and labor consuming. The most commonly used setting is the 15-theoretical-plate column, ASTM D2892, which might take up to 48 h for the characterization of a single sample. Another commonly used reference method for characterization is ASTM D86, which is a batch distillation taking roughly 30 min depending on the sample [63]. These reference methods are impossible to use for real-time applications, but they are used to calibrate soft sensor models.

Other important quality indicators are sulfur, nitrogen, and metal content. The reference methods used for the detection of sulfur and nitrogen compounds are ASTM D2622 [64] and ASTM D4629 [65] respectively. While the sulfur measurement is relatively fast with X-ray fluorescence, nitrogen content determination needs a longer sample preparation and analysis time. The most abundant metals in hydrocarbon mixtures are vanadium, nickel, and iron. One very commonly used reference method for these substances is ASTM D5863 [66]. This method has a long sample preparation procedure with dry ashing.

The reference methods are used internationally by the certified laboratories as proof of the quality of the feedstock or the product; however, they are mostly time consuming, which makes them impossible to use in real-time applications. Spectroscopic methods are fast, therefore, possible to use in real-time when fast action is important for a smooth operation. There are different types of spectroscopic methods that can be beneficial to ensure stable product quality and energy efficiency. In the literature research, nuclear magnetic resonance (NMR), infrared (IR), near-infrared (NIR), ultra-violet (UV), visible (Vis), and Raman spectroscopic methods are found for petroleum refinery applications. Fundamentals of these methods can be found, NMR [24], IR [67], NIR [68], UV-Vis [69], Raman [70].

The selection of the suitable spectroscopic method is not the only concern for soft sensor applications. Data treatment and multivariate statistical analysis have important roles in the prediction accuracy of the soft sensors. An extensive review of relevant spectroscopy techniques was published in three parts, focusing on different aspects [71–73]. In [71], the focus was on different spectroscopic methods and relevant chemometric applications. UV-Vis, IR, and NIR spectroscopy was given with regard to utilization of partial least squares (PLS) regression, principal component regression (PCR), multiple linear regression (MLR), and ANN with exemplar works. Data pretreatment was the main topic of [72]. In this part, multiplicative scatter correction (MSC), normalization, derivatives, and PCA were given next to other techniques. In the last part, [73], multi-way data analysis, and its use for several spectroscopic methods were discussed.

Targeting the petroleum applications, Silva et al. reviewed different spectroscopic methods with specific details of NMR spectroscopy for the determination of heavy oil characteristics [74]. They compared NMR against IR spectroscopy, Raman spectroscopy, and mass spectrometry in terms of sulfur, nitrogen, aromatic hydrocarbons, paraffins content, and API prediction. Within this discussion, IR was found better in qualitative analysis than quantitative, Raman was identified as unsuitable for crude oil mixtures, and mass spectrometry was found inadequate for certain contaminants' measurements.

NMR spectroscopy was tested for TBP curve characterization by multiple researchers. Filgueiras et al. used NMR to predict T10, T50, and T90 values using the support vector machine (SVM) and PLS regression models and obtained a better data fit by using SVM [75]. For the heavy feedstock, root mean square error of prediction (RMSEP) values were found significantly high for the TBP curve prediction. However, this does not confound the useability of NMR for refining applications. The chemical shift locations of ^1H NMR spectrum is given in Figure 8. Different H atoms are located separately on the spectrum due to their surrounding atoms, and thanks to that feature, different characteristics of the mixture can be calculated.

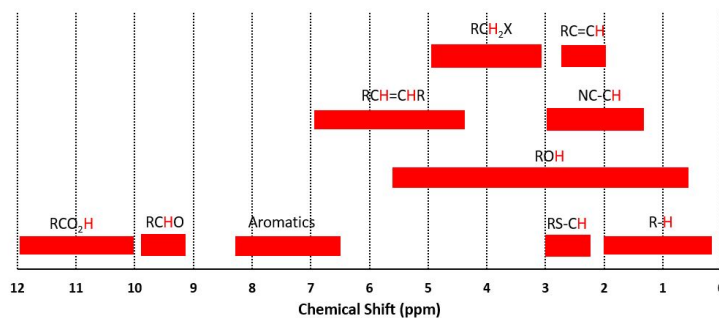


Figure 8. ^1H NMR chemical shift locations.

Another study showed a better data fit by using NMR with PLS regression models, built to predict product yields, sulfur, nitrogen, and metal content [76]. The high accuracy was obtained thanks to the selection of crude oil from a specific region. The lower accuracy was confirmed by other researchers for more generalized models [77]. A hydrocarbon mixture ^1H NMR spectra is given in Figure 9 and by using Figure 8, alkane, alkene, and aromatic peaks can easily be observed.

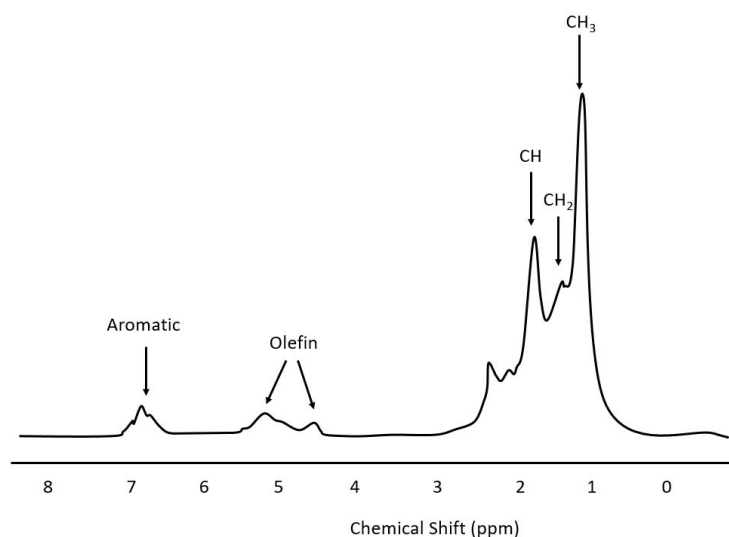


Figure 9. ^1H NMR spectrum of a hydrocarbon mixture.

NIR spectroscopy and its use in refinery applications are investigated deeply. Thanks to the new methods in statistical analysis, the prediction accuracy is improved. Pasquini and Bueno compared PLS regression and ANN models, which were built to predict the TBP curve and API gravity of hydrocarbon mixtures [78]. A PLS regression model showed better performance. In this work, it was also demonstrated that wavelength selection should be done carefully as it might bring a considerable improvement or a loss. They selected three wavelength regions to build their models on (2000–2500 nm (I), 1650–2500 nm (II), 1100–2500 nm (III)), and found that the model using region (I)

performed the best and (II) the worst. Although (II) has all the information included in (I), it performed worse due to the additional region without relevant information and hence more noise. A hydrocarbon mixture NIR spectrum is given in Figure 10.

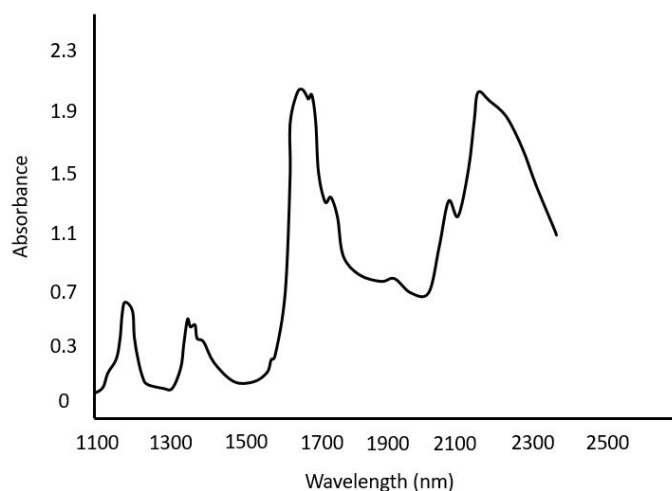


Figure 10. NIR spectrum of a hydrocarbon mixture.

Unlike NMR giving information of an individual atom in different molecules, the NIR spectrum carries chemical bond information of the analyzed substance. Different bond locations of the NIR region are given in Figure 11, which also shows the importance of wavelength selection. Some bond information can be found in multiple locations. For example, the C-H bond is observed in combination and overtone regions, unlike S-H, which is only observed in the first overtone region.

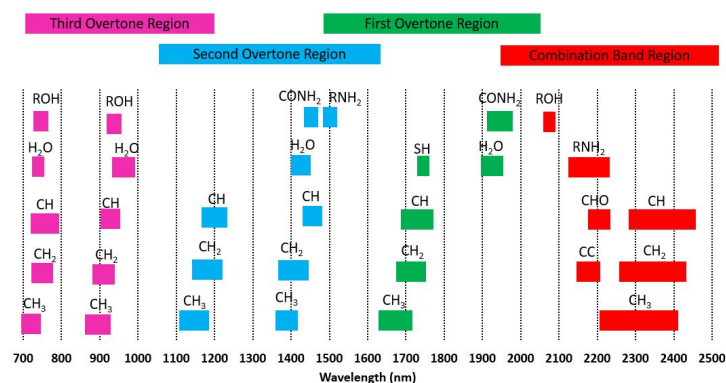


Figure 11. Peak locations of NIR spectrum.

Hidajat and Chong built PLS regression models for the TBP curve and density by using NIR spectra [79]. Instead of using similar crudes that come from the same region, they used 30 different samples from 12 different locations. This kind of work is vital to build a library and test the generalization capability and flexibility of NIR based models. These models proved the usefulness of NIR for refineries operating with different crudes or a mixed feedstock. However, all types of oils should be included while building the model. Losing accuracy is a possible outcome of a wide range of properties. Updating the model according to the oil or mixtures characteristics can be a solution to build fewer models and to keep accuracy acceptable.

Laxalde et al. studied the optimization of preprocessing methods and variable selection for heavy mid-product characterization by using NIR [80]. They used a genetic algorithm (GA) to perform this task. Atmospheric residue and vacuum gas oil samples were analyzed for their saturate, aromatic, resin, asphaltene (SARA) content. Smoothing, detrending, multiplicative scatter correction, normalization,

standard normal variate, Savitzky-Golay derivative, and centering (mean and median) were some of the tested data pretreatment methods. Using a GA for preprocessing method selection improved the accuracy of the saturates model. Optimization of both preprocessing method and the wavelength selection resulted in better models for aromatics and resins models.

In another study, NIR was compared with IR and Raman spectroscopy for predicting T50, T85, and sulfur content [81]. This study showed that Raman spectroscopy is a good alternative for diesel characterization, which is in contrast to the results claimed by [74]. There can be two reasons behind the different opinions; elimination of the heaviest compounds within the process from crude oil to diesel and different statistical analysis and modeling approaches. The best Raman spectroscopy models were obtained by using ANNs. Figure 12 shows a typical Raman spectrum for a hydrocarbon mixture, which has sharper peaks than the NIR spectrum of a similar sample, thus more specific on molecular structure information.

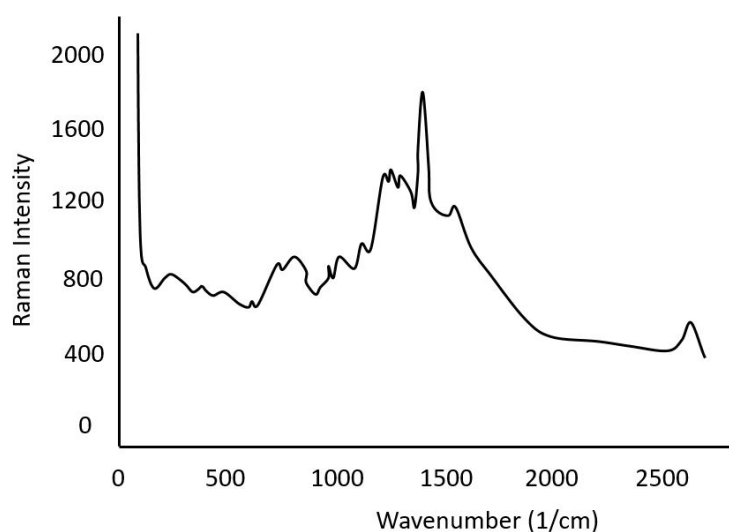


Figure 12. Raman spectrum of a hydrocarbon mixture.

Laxalde et al. showed the combined utilization of near and mid-infrared spectroscopy for heavy oil characterization and compared the capability with the individual use of each method [82]. To be able to use the different spectral data acquired from both methods together, two different PLS models are tested, multi-block PLS and serial PLS. Multi-block PLS requires creating a single model, whereas serial PLS uses two different models in series. This allows the use of different numbers of latent variables for each model. NIR results showed better prediction accuracy compared to IR, but the improvement of the RMSEP values (except for saturates model) with the combined spectral information proved that the IR spectra carry relevant information for the unsaturated compounds. C=C stretching vibrations were observed only in the IR, which makes it an important contributor to the asphaltene models.

Chung's review of NIR use in refineries provided a comprehensive summary in consideration of specific products [23]. To this point of view, accuracy was the biggest challenge to be addressed. However, it was also mentioned that other aspects that needed to be considered for the heavy oils, i.e., sampling system. For the online applications, this must be taken into account to be able to obtain the spectral data in good quality.

Buchanan reviewed the use of NIR in the petrochemical industry in the handbook of near-infrared analysis [83]. This chapter focused on oxygenates, octane number, composition, bitumen, and environmental analysis. Presenting the findings of composition analysis by Swarin and Drumm [84], Buchanan drew the attention to sulfur concentration prediction. The low r^2 value was explained with sulfur being a non-active species for NIR. Buchanan concluded the review with a remark, that even though the petrochemical industry is mature, further process enhancements are possible thanks to the technological advances in sensor development.

Baldrich worked on using UV spectra for physicochemical analysis of oil [85]. UV analyzers are cheaper than NIR and NMR devices; however, their use in the petrochemical industry is not wide. Classification studies can be found in the literature for crude oil and light gas oils [86]. Baldrich's work showed the usefulness of UV spectra to correlate sulfur, nitrogen, nickel, vanadium, saturates, resins, and aromatic content with PLS regression models. Aromatic content models showed better accuracy than saturates and resins models. Particular attention should be given to the sulfur prediction model. Considering Buchanan's aforementioned discussion on NIR use for sulfur prediction and relatively higher error of such models, the industry might tend towards the use of cheaper UV option.

Baldrich later published a comparison study of UV-Vis and NIR spectra together with Novoa and Bueno [87]. Aligned with the UV-only analysis, the UV-Vis model showed superior performance compared to the NIR model for sulfur prediction. Next to that, for microcarbon residue, nickel, and vanadium content UV-Vis was found to perform better. On the other hand, nitrogen content and density were predicted more accurately by using NIR. This study employed PLS regression for all its models.

Pasadakis et al. tested mid-IR use for diesel characterization [88]. Developed ANN models were used to detect the distillation curve next to the pour point and cloud point. They used a visual wavelength selection method and employed certain ranges in their model (1474–1464, 1314–1306, 810–800, 740–720 cm^{-1}). If these ranges are investigated on a typical mid-IR spectrum of diesel oil, the peaks clearly indicate the importance as in Figure 13. The accuracy of the models was found to be comparable to the repeatability of the reference methods.

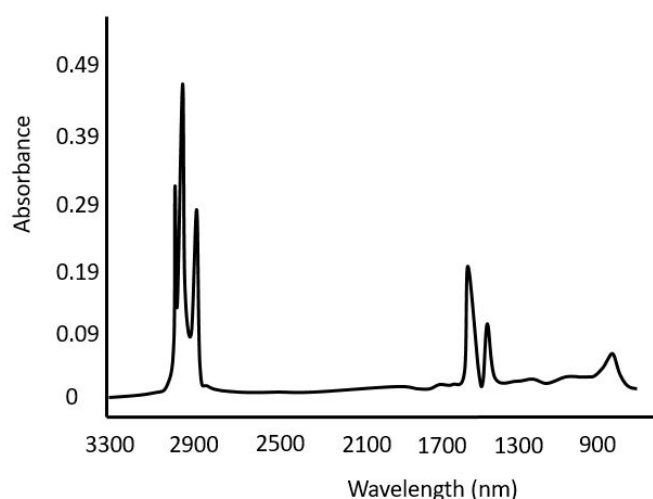


Figure 13. Mid-IR spectrum of a hydrocarbon mixture.

Table 4 shows the different analyzers and their advantages next to the selected work using them for characterization.

Table 4. Advantages of different analyzers and selected works for determination of hydrocarbon characteristics.

Analyzer	Advantage	TBP	SARA	N	S	Metal	API
NIR	No sampling requirements	[78]	[82]		[89]		[78]
NMR	No optic interference	[75] ⁱ	[90]	[90]	[90]	[90]	[91]
Raman	Not sensitive to temperature changes	[81] ⁱⁱ	[92]		[81]		
Mid-IR	Precise molecular data	[88]	[82]		[81]		
UV	Cheap equipment		[85]	[85]	[85]	[85]	

ⁱ Only T10, T50 and T90. ⁱⁱ Only T50 and T85.

Some researchers developed data-driven models to predict the TBP assay or only the specific points on it. These models can be considered similar to soft sensors, although there is no specific measurement taken for them; instead, they use the readily available process measurements. Lababidi et al. proposed using a simple model to predict the entire TBP curve by using the API measurements based on a gamma distribution [93]. API is relatively easier and faster to measure than the TBP curve. Hence, this type of prediction model can be useful for the operation. Moreover, the simplicity of the model makes it easy to calibrate and run fast. However, it should not be forgotten that it is simplicity, which makes it impossible to extrapolate. As suggested by the authors as well, the model has to be recalibrated according to feedstock, catalyst, and operating point change.

Yuan et al. developed adaptive soft sensors to overcome the recalibration need [94]. Their study demonstrated four different methods to modify the PLS regression models with an adaptive approach to be used to calculate T10. The recursive PLS regression and the moving window recursive PLS regression models performed equally well and better than their alternatives by using 43 process measurements. The common feature of all four models was the forgetting factor included in the algorithm. Higher importance was given to the newer data points.

A recent work by Wang et al. showed a detailed analysis of different data-driven modeling methods to predict T10, T50, and T90 [95]. They tested PLS regression, SVM, ANN, and ensemble learning models. Their results showed that the more sophisticated ensemble learning approach gives better accuracy, whereas the SVM model performed worse than the rest of the models in this work. ANN and PLS had comparable performances according to their parameters, such as the number of layers and number of latent variables. This result shows the importance of model selection with an understanding of the real system and the improvement margin of each model by tuning the parameters.

TBP point prediction is not the only interest of the data-driven soft sensor development. Lukec et al. tested an ANN approach for sulfur content prediction [96]. As mentioned earlier, catalyst deactivation is an important drawback of continuous operation that causes higher sulfur content product if the operation temperature is kept constant. In the suggested ANN models, they introduced a catalyst deactivation factor that was calculated by using the total flow rate and the inlet temperature. By using six inputs from the plant data (inlet temperature, four flow rates, density, and pressure) and the deactivation factor, Lukec et al. calculated the sulfur content for two different feeds (light gas oil and vacuum gas oil).

Soft sensors and data-driven prediction models overcome the problem of slow measurements with the reference methods. Different analyzers, data pretreatment, and modeling methods were tested by the researchers to predict the quality indicators. Real-time optimization or optimal control systems need an accurate prediction of the quality indicators to be able to react fast to process changes; therefore, it is necessary to employ fast characterization techniques next to the reference methods.

4. Optimization, Control, Diagnostics and Future Directions

As discussed previously, the interdisciplinary research of computer engineering and chemical engineering for hydrocracking is limited. Existing work has a focus on model predictive control (MPC) and RTO. Model predictive control is an advanced control method which uses an optimization algorithm on a model to calculate control inputs. Principles of MPCs can be found in literature [97]. It is commonly used in industrial applications and so in refineries. One reason for its extensive use over traditional proportional—integral—derivative (PID) control structures is the PID controller's error correction approach on a defined setpoint by using its three components comprising the name. The error correction works well for single input—single output systems, whereas for multiple input—multiple output systems with interference, it is often insufficient.

4.1. Optimization and Control

Lababidi et al. compared constrained MPC and PID control structures for a pilot hydrotreatment plant [98]. They defined four manipulated variables (three reactor skin temperatures and system

pressure) and three controlled variables (two reactor bed temperatures and pressure). Instead of a first principle model, a system identification model was used to optimize the sulfur concentration. They defined high sulfur content as 'off-spec' and low sulfur content as 'giveaway'. In this work, the constrained MPC showed better performance over the PID structure in the set point tracking, therefore in eliminating giveaway or off-spec production modes. MPC is not a novel technology and is already applied widely in industrial operations. However, as the authors stated, this kind of study is essential not only to demonstrate the capabilities of advanced control systems but also to have better control over the pilot plants and have more realistic experimental results.

Aydin et al. demonstrated a two-layer MPC model on a hydrotreatment plant [99]. In the upper level, an economic MPC was used with long sampling time and in the lower level two decoupled MPCs, one for feed blending and another for temperature control, with much shorter sampling time were employed. Both of the lower level MPCs got their setpoints from the regulatory economic MPC. This system was tested for set point tracking and disturbance rejection. They demonstrated a change in the feed price and the price of the products, which caused a setpoint change in the operating conditions.

Alhajree et al. modeled an industrial scale hydrocracker with an ANN and optimized it for different products [100]. The fractionation columns were included in the model next to the reactors, and this approach brought a broader control over the system. Although the cracking reactions occur in the reactors and therefore depend on the temperature and pressure of them, column temperature and pressure values change the yield and characteristics of the product streams. Having a higher temperature in a fractionation column increases the amount of the top product and pushes its distillation curve to higher temperatures. The sensitivity analysis showed the dependency of the yields on the reactor temperature. The heavy naphtha showed a more significant response to the temperature change, hence a higher sensitivity than the other product streams. Different objectives were tested for maximizing different products. As a result, gas oil maximization was found to need 16% less hydrogen due to lower cracking requirements. To maximize the kerosene production, reactor temperatures needed to be increased 8% next to a feed flow rate increase of 10%. These results show RTO's importance, considering that the maximization of a particular product needs a change in more than one parameter.

Zhou et al. developed a discrete lumped model and used it to optimize an industrial plant [101]. They compared sequential quadratic programming (SQP) and GA for two different objectives, diesel and kerosene maximization. They stated that the computational time required for SQP was 5 to 10 times lower than GA. One unexpected result was the temperature values of the optimized diesel maximization scenario, for which both algorithms calculated lower bed exit temperatures than the real operation. Lower temperatures should favor for less cracking and lower yields. However, this was explained by lower byproduct yields and eliminating the deep cracking in the optimized operation.

A Kriging based surrogate model was proposed by Zhong et al. [102]. This model was built to optimize the operating conditions to increase the profit with the changing feed and defined product values respecting the market specifications. Their demonstration of four different feed types showed a drastic change in optimal operating conditions. The objective of this optimization was economical within process-specific constraints such that it showed the necessity of operational adjustments in real-time according to feed characteristics to increase the profit.

Sildir et al. tested a plant-wide approach for optimization on an industrial hydrocracker plant [103]. They pointed out the varying economic and market conditions and how RTO can be useful for such a dynamic system. Just as Alhajree et al. [100], they also considered the fractionation in their system. While the reactor model was obtained with a step test applied to a continuous lumping model, the distillation column models were developed by using real data. On the integrated system, they applied MPCs in layers. In the management layer, feed, hydrogen, and product prices were calculated considering the market demand. A steady-state optimization layer took these values to calculate the set points of optimal product distribution. A supervisory MPC structure was used to optimize the integrated system according to the calculated product distribution. The final

layer consisted of two decentralized MPCs, which were used separately for the reactors and the fractionation columns.

Another study by Altissimi et al. showed the reason behind the need for simplification of the models for optimization problems [104]. Their work showed an ANN model for the hydrocracker separation unit against a first principle model used for profit maximization. An SQP algorithm was used for the optimization, and with the ANN model, it took one minute, much faster than the physics-based model's 70 min calculation time. The use of data-driven models is inevitable for RTO applications.

RTO and MPC applications gained importance for HC due to the increasing cost of energy, as this is an energy-intensive process carried out in high temperature and pressure environment. The operation cost can be significantly decreased by optimization. Furthermore, the feedstock cost is an important contributor to the total production cost that can be minimized with resource efficiency. Although MPC is not a novel technology, its applications show novelty with the newly developed methods and carry importance to demonstrate the capability of such systems contributing to the interdisciplinary research of chemical and computer engineering.

4.2. Diagnostics

Data-based diagnostics is another critical domain of interdisciplinary research. It gained importance for process industries with the growing data, and HC is a good candidate for applications because it has a high number of sensors installed for safety, and they can supply the required data for the diagnostics. A fault diagnosis system was tested for a hydrocracker by Zhao et al. [105]. They used a neural network structure for the detection of six identified faults related to feed flow rate (high, low), feed temperature (high, low), and characteristics of the feed (heavy, light). This model used the outlet temperature and the hydrogen flow rate as the input parameters. However, it only worked as long as the system ran on a steady-state due to the position of the selected inputs. Before the effects of the faults were observable on the outlet temperature, it was not possible to detect them.

A work by Stockmann et al. showed a source identification for plant-wide faults of an industrial hydrocracker [106]. A new method was developed based on the k-nearest neighbor algorithm to be able to deal with the nonlinearity. Two different sources for the faults were defined as oscillating disturbances and stiction. The HC unit was described as a system of multiple input single output subsystems. The obstruction to the real-time application was the long calculation time, as pointed out clearly in the study.

Zheng and Zhao developed an unsupervised data mining method for fault diagnosis [107]. Like most papers in this area, they demonstrated the capabilities of their method on the benchmark Tennessee-Eastman process, but they also included an industrial hydrocracker study. The fault detected in the hydrocracker was a change in the product. The model detected the change after one hour. This shows how useful such models can be because, as mentioned earlier, the traditional characterization methods take such a long time. Instead of an online analyzer, this type of model can be utilized to react faster to a troublesome change in the product.

This area of research needs further attention, especially to demonstrate the usefulness of developed methods on real systems. The limited number of studies show that the diagnostics systems need further improvement to be used effectively and help take precautions before the result of the fault is observed on the product quality.

4.3. Future Directions

There are some problems which are not addressed yet. A change in the feed characteristics, which often happens for the most refinery operations, causes non-optimal operation until the laboratory analysis is finalized. Slow and labor consuming analysis methods hinder the possible use of feed-forward control strategies. This was previously pointed out in literature [108]. Consolidation of the research directions can eradicate this problem. Using the soft sensors to supply the necessary

real-time data for the control strategies can lead to a more efficient process. However, a careful analysis is necessary in terms of the level of automation, the accuracy of the sensor model, and the optimality of the operation. Figure 14 summarizes the relationship between these major points. An optimal operation can be achieved with accurate real-time solutions. The accuracy is dependent on the specificity of the model, the generalization of the models causes a loss of accuracy but gain in automation as the obligation of recalibration or remodeling declines.

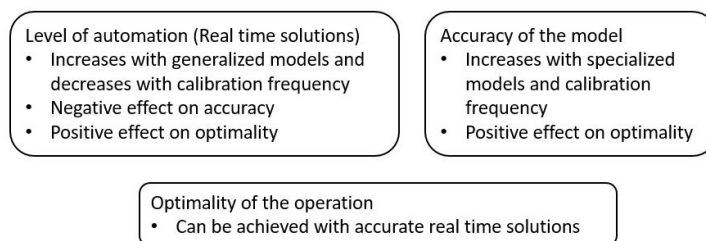


Figure 14. Essential points to consider for soft sensor applications.

As mentioned earlier, there are a limited number of studies that discuss automation and digitalization of HC. Although they are not high in number, they are notable examples demonstrating the capability of interdisciplinary work for energy and resource efficiency. Models and soft sensors are crucial elements of optimization and control studies as well as decision support systems. A flowchart showing possible information flow between these elements is given in Figure 15. Process measurements and soft sensor predictions are logged in a database that supplies data to the models. A data-based model generates process predictions for an optimization algorithm to find the best control actions, which is an iterative process, and a decision support system gets the necessary data from different models that help the operators in their decisions.

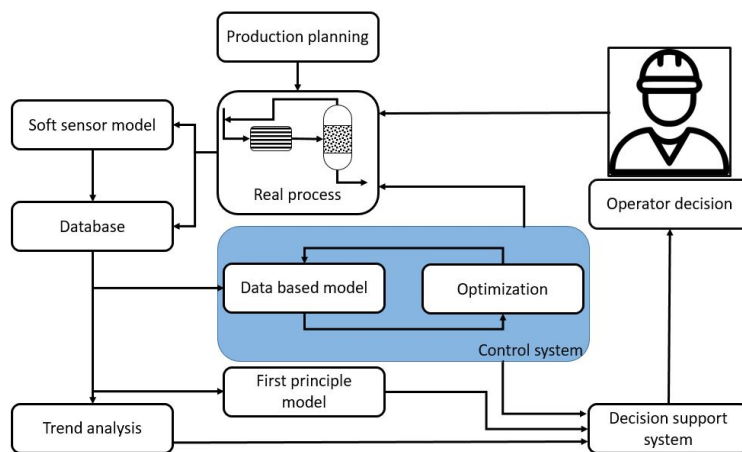


Figure 15. System architecture schematic showing the links between different elements covered in this review.

The identified critical aspects of each element should be taken into account to escalate the automation in a reliable manner. Until recently, each part needed its own improvement to reach a degree of maturity to be able to supply the demands of the connected units. The realization of this type of architecture is becoming feasible, considering the level attained at every point. The future studies that can integrate these elements together will reveal the real improvement capacity, and in view of covered literature, HC shows a promising competence. Furthermore, fundamental research should be supplemented with applied scenarios because only with such an approach can the integration of the developed schemes be achieved. Changes are seen as risky by the industry. However, the increasing number of scientific demonstrations will increase the credibility of new methods. HC is

a favorable process to test the capabilities of new data analysis techniques, optimization algorithms, and control structures.

5. Conclusions

The recent advances in hydrocracking are reviewed from a digitalization perspective. To keep pace with the changing needs towards a more sustainable process, mathematical models, soft sensors, control optimization, and diagnostics are selected as the main topics. Concluding remarks for this work are:

- Physics-based models still have high importance for design and analysis. The increasing complexity of the process increases the need for detail, which can only be supplied by first-principle models. The physical meaning of the variables brings a better understanding of the actual system by providing links between the changes and their potential causes. Next to that, these models have a better extrapolation capability hence a higher fitness for analysis. However, these models are time consuming both in the development phase and in the computation phase.
- The need should be analyzed carefully before deciding the type of model to avoid excessive computational effort or lack of data. It is not only the decision of a physics-based or a data-based model since both of these model classes also offer various options with different assets. A deliberate evaluation leads to a suitable model selection and maximizes the benefit from it.
- Data-driven models need more attention, as they are crucial for real-time optimization and control. These models are fast and robust when designed with system knowledge to evaluate the required inputs to predict the desired outputs. They are suitable for complex processes and employed in industrial applications but are rarely published. Their addition to literature may contribute to a better judgment of data analysis, model training, tuning, and testing for the refining industry.
- The models that do not consider catalyst aging have a need for regular recalibration. As most of the industrial scale hydrocracking units employ fixed bed reactors, catalyst degradation affects the model parameters. With the decreasing activity, bed temperatures increase, and models lose the prediction accuracy. To avoid ill-advised decisions based on outdated models, the affected model parameters should be assessed and updated systematically.
- The accuracy of the soft sensors is still a concern that needs to be addressed. Together with the data-based models, soft sensors can solve real-time monitoring problems of refineries. However, before depending on them for process decisions, they need to be tested with varying feedstock and proved to be reliable.
- Wavelength selection has to be done thoroughly. Using selected wavelengths instead of the entire spectra might help avoid over-fitting; however, it might lead to a loss of necessary data. A well-distributed training sample is necessary to observe the data carrying peaks and to avoid the elimination of them, hence a sacrifice of fidelity.

The integration of research directions is necessary to reach the potential of this process in terms of efficiency. The improvement of the models or the characterization methods naturally helps the progress, but their combined use should also be studied for further enhancements. Hydrocracking offers a spacious research ground not only for chemical engineering but also for computer engineering and data science. To prove the reliability of the new methods to the industry, applied interdisciplinary work should be supported.

Funding: This research was a part of FUDIPO project and funded by European Union's Horizon 2020 Research and Innovation Program under grant number 723523.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

ANN	Artificial Neural Network
SARA	Saturate, Aromatic, Resin, Asphaltene
API	American Petroleum Institute
FCC	Fluid Catalytic Cracking
HC	Hydrocracking
FBR	Fixed Bed Reactor
EBR	Ebullated Bed Reactor
SPR	Slurry Phase Reactor
TBP	True Boiling Point
SEMM	Single Event Microkinetic Modeling
CFD	Computational Fluid Dynamics
RTO	Real-Time Optimization
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
ASTM	American Society for Testing and Materials
NIR	Near Infrared
IR	Infrared
NMR	Nuclear Magnetic Resonance
UV	Ultra-Violet
PLS	Partial Least Squares
PCR	Principal Component Regression
MLR	Multiple Linear Regression
SVM	Support Vector Machine
RMSEP	Root Mean Square Error of Prediction
GA	Genetic Algorithm
MPC	Model Predictive Control
PID	Proportional - Integral - Derivative
SQP	Sequential Quadratic Programming

References

1. Directive 2009/30/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 98/70/EC as regards the specification of petrol, diesel and gas-oil and introducing a mechanism to monitor and reduce greenhouse gas emissions and amending Council Directive 1999/32/EC as regards the specification of fuel used by inland waterway vessels and repealing Directive 93/12/EEC. *Off. J. Eur. Union*. **2008**, *140*, 88–114.
2. Höök, M.; Aleklett, K. A review on coal-to-liquid fuels and its coal consumption. *Int. J. Energy Res.* **2010**, *34*, 848–864. [[CrossRef](#)]
3. Speight, J. Visbreaking: A technology of the past and the future. *Sci. Iran*. **2012**, *19*, 569–573. [[CrossRef](#)]
4. Talebian-Kiakalaieh, A.; Amin, N.A.S.; Mazaheri, H. A review on novel processes of biodiesel production from waste cooking oil. *Appl. Energy* **2013**, *104*, 683–710. [[CrossRef](#)]
5. Scherzer, J.; Gruia, A.J. *Hydrocracking Science and Technology*; CRC Press: Boca Raton, FL, USA, 1996.
6. Calzado Catala, F.; Flores de la Fuente, R.; Gardzinski, W.; Kawula, J.; Hille, A.; Iglesias Lopez, A.; Lambert, G.; Leveque, F.; Lyde, C.; Mackenzie, A.; et al. *Oil Refining in the EU in 2020, with Perspectives to 2030*; Technical Report; CONCAWE: Brussels, Belgium, 2013.
7. Petroleum and Other Liquids. Available online: <https://www.eia.gov/petroleum/data.php/summary> (accessed on 13 April 2020).
8. Coonradt, H.L.; Garwood, W.E. Mechanism of hydrocracking. Reactions of Paraffins and Olefins. *Ind. Eng. Chem. Process. Des. Dev.* **1964**, *3*, 38–45. [[CrossRef](#)]
9. Qader, S.; Hill, G. Hydrocracking of gas oil. *Ind. Eng. Chem. Process. Des. Dev.* **1969**, *8*, 98–105. [[CrossRef](#)]
10. Flinn, R.; Larson, O.; Beuther, H. The mechanism of catalytic hydrocracking. *Ind. Eng. Chem.* **1960**, *52*, 153–156. [[CrossRef](#)]

11. Sullivan, R.; Egan, C.J.; Langlois, G.; Sieg, R.P. A new reaction that occurs in the hydrocracking of certain aromatic hydrocarbons. *J. Am. Chem. Soc.* **1961**, *83*, 1156–1160. [[CrossRef](#)]
12. Laxminarasimhan, C.; Verma, R.; Ramachandran, P. Continuous lumping model for simulation of hydrocracking. *AIChE J.* **1996**, *42*, 2645–2653. [[CrossRef](#)]
13. Mohanty, S.; Saraf, D.; Kunzru, D. Modeling of a hydrocracking reactor. *Fuel Process. Technol.* **1991**, *29*, 1–17. [[CrossRef](#)]
14. Ayasse, A.R.; Nagaishi, H.; Chan, E.W.; Gray, M.R. Lumped kinetics of hydrocracking of bitumen. *Fuel* **1997**, *76*, 1025–1033. [[CrossRef](#)]
15. Schweitzer, J.M.; Galtier, P.; Schweich, D. A single events kinetic model for the hydrocracking of paraffins in a three-phase reactor. *Chem. Eng. Sci.* **1999**, *54*, 2441–2452. [[CrossRef](#)]
16. Elliott, D.C.; Hart, T.R.; Neuenschwander, G.G.; Rotness, L.J.; Zacher, A.H. Catalytic hydroprocessing of biomass fast pyrolysis bio-oil to produce hydrocarbon products. *Environ. Prog. Sustain. Energy* **2009**, *28*, 441–449. [[CrossRef](#)]
17. Tanneru, S.K.; Steele, P.H. Direct hydrocracking of oxidized bio-oil to hydrocarbons. *Fuel* **2015**, *154*, 268–274. [[CrossRef](#)]
18. Yang, C.; Jia, L.; Chen, C.; Liu, G.; Fang, W. Bio-oil from hydro-liquefaction of *Dunaliella salina* over Ni/REHY catalyst. *Bioresour. Technol.* **2011**, *102*, 4580–4584. [[CrossRef](#)]
19. Rana, M.S.; Sámano, V.; Ancheyta, J.; Diaz, J. A review of recent advances on process technologies for upgrading of heavy oils and residua. *Fuel* **2007**, *86*, 1216–1231. [[CrossRef](#)]
20. Castañeda, L.; Muñoz, J.; Ancheyta, J. Combined process schemes for upgrading of heavy petroleum. *Fuel* **2012**, *100*, 110–127. [[CrossRef](#)]
21. Tye, C.T., Catalysts for Hydroprocessing of Heavy Oils and Petroleum Residues. In *Processing of Heavy Crude Oils—Challenges and Opportunities*; Gounder, R.M., Ed.; IntechOpen: London, UK, 2019.
22. Ancheyta, J.; Sánchez, S.; Rodríguez, M.A. Kinetic modeling of hydrocracking of heavy oil fractions: A review. *Catal. Today* **2005**, *109*, 76–92. [[CrossRef](#)]
23. Chung, H. Applications of near-infrared spectroscopy in refineries and important issues to address. *Appl. Spectrosc. Rev.* **2007**, *42*, 251–285. [[CrossRef](#)]
24. Edwards, J.C., A Review of Applications of NMR Spectroscopy in the Petroleum Industry. In *Spectroscopic Analysis of Petroleum Products and Lubricants*; Nadkarni, K.R., Ed.; ASTM International: West Conshohocken, PA, USA, 2011; pp. 423–473.
25. Ogden-Swift, A. An industrial review of advanced process control of continuous processes on refineries and petrochemical plants. *Trans. Inst. Meas. Control* **1996**, *18*, 9–14. [[CrossRef](#)]
26. Robinson, P.R.; Dolbear, G.E., Hydrotreating and hydrocracking: Fundamentals. In *Practical Advances in Petroleum Processing*; Hsu, C.S., Robinson, P.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 177–218.
27. Ancheyta, J.; Alvarez-Majmutov, A.; Leyva, C. Hydrotreating of oil fractions. In *Multiphase Catalytic Reactors: Theory, Design, Manufacturing, and Applications*; Ilsen Onsan, Z., Avci, A.K., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2016.
28. Al-Dalama, K.; Stanislaus, A. Comparison between deactivation pattern of catalysts in fixed-bed and ebullating-bed residue hydroprocessing units. *Chem. Eng. J.* **2006**, *120*, 33–42. [[CrossRef](#)]
29. Martinez, J.; Sanchez, J.L.; Ancheyta, J.; Ruiz, R.S. A review of process aspects and modeling of ebullated bed reactors for hydrocracking of heavy oils. *Catal. Rev.* **2010**, *52*, 60–105. [[CrossRef](#)]
30. Zhang, S.; Liu, D.; Deng, W.; Que, G. A review of slurry-phase hydrocracking heavy oil technology. *Energy Fuels* **2007**, *21*, 3057–3062. [[CrossRef](#)]
31. Bellussi, G.; Rispoli, G.; Landoni, A.; Millini, R.; Molinari, D.; Montanari, E.; Moscotti, D.; Pollesel, P. Hydroconversion of heavy residues in slurry reactors: Developments and perspectives. *J. Catal.* **2013**, *308*, 189–200. [[CrossRef](#)]
32. Calderón, C.J.; Ancheyta, J. Modeling, simulation, and parametric sensitivity analysis of a commercial slurry-phase reactor for heavy oil hydrocracking. *Fuel* **2019**, *244*, 258–268. [[CrossRef](#)]
33. Sahu, R.; Song, B.J.; Im, J.S.; Jeon, Y.P.; Lee, C.W. A review of recent advances in catalytic hydrocracking of heavy residues. *J. Ind. Eng. Chem.* **2015**, *27*, 12–24. [[CrossRef](#)]
34. Mosby, J.; Buttke, R.; Cox, J.; Nikolaidis, C. Process characterization of expanded-bed reactors in series. *Chem. Eng. Sci.* **1986**, *41*, 989–995. [[CrossRef](#)]

35. Stangeland, B.E. A kinetic model for the prediction of hydrocracker yields. *Ind. Eng. Chem. Process. Des. Dev.* **1974**, *13*, 71–76. [[CrossRef](#)]
36. Callejas, M.A.; Martínez, M.T. Hydrocracking of a Maya residue. Kinetics and product yield distributions. *Ind. Eng. Chem. Res.* **1999**, *38*, 3285–3289. [[CrossRef](#)]
37. Elkady, F. Hydrocracking of vacuum distillate fraction over bifunctional molybdenum-nickel-silica-alumina catalyst. *Indian J. Technol.* **1979**, *17*, 176–183.
38. Martínez, J.; Ancheyta, J. Kinetic model for hydrocracking of heavy oil in a CSTR involving short term catalyst deactivation. *Fuel* **2012**, *100*, 193–199. [[CrossRef](#)]
39. Elizalde, I.; Ancheyta, J. Modeling catalyst deactivation during hydrocracking of atmospheric residue by using the continuous kinetic lumping model. *Fuel Process. Technol.* **2014**, *123*, 114–121. [[CrossRef](#)]
40. Elizalde, I.; Ancheyta, J. On the detailed solution and application of the continuous kinetic lumping modeling to hydrocracking of heavy oils. *Fuel* **2011**, *90*, 3542–3550. [[CrossRef](#)]
41. Alcázar, L.A.; Ancheyta, J. Sensitivity analysis based methodology to estimate the best set of parameters for heterogeneous kinetic models. *Chem. Eng. J.* **2007**, *128*, 85–93. [[CrossRef](#)]
42. Valavarasu, G.; Bhaskar, M.; Sairam, B.; Balaraman, K.; Balu, K. A four lump kinetic model for the simulation of the hydrocracking process. *Pet. Sci. Technol.* **2005**, *23*, 1323–1332. [[CrossRef](#)]
43. Balasubramanian, P.; Pushpavanam, S. Model discrimination in hydrocracking of vacuum gas oil using discrete lumped kinetics. *Fuel* **2008**, *87*, 1660–1672. [[CrossRef](#)]
44. Becker, P.; Serrand, N.; Celse, B.; Guillaume, D.; Dulot, H. Comparing hydrocracking models: Continuous lumping vs. single events. *Fuel* **2016**, *165*, 306–315. [[CrossRef](#)]
45. Becker, P.J.; Serrand, N.; Celse, B.; Guillaume, D.; Dulot, H. A single events microkinetic model for hydrocracking of vacuum gas oil. *Comput. Chem. Eng.* **2017**, *98*, 70–79. [[CrossRef](#)]
46. Thybaut, J.; Marin, G. Single-Event MicroKinetics: Catalyst design for complex reaction networks. *J. Catal.* **2013**, *308*, 352–362. [[CrossRef](#)]
47. Svoboda, G.D.; Vynckier, E.; Debrabandere, B.; Froment, G.F. Single-event rate parameters for paraffin hydrocracking on a Pt/US-Y zeolite. *Ind. Eng. Chem. Res.* **1995**, *34*, 3793–3800. [[CrossRef](#)]
48. Quann, R.; Jaffe, S. Building useful models of complex reaction systems in petroleum refining. *Chem. Eng. Sci.* **1996**, *51*, 1615–1635. [[CrossRef](#)]
49. Quann, R.J.; Jaffe, S.B. Structure-oriented lumping: Describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483–2497. [[CrossRef](#)]
50. Jaffe, S.B.; Freund, H.; Olmstead, W.N. Extension of structure-oriented lumping to vacuum residua. *Ind. Eng. Chem. Res.* **2005**, *44*, 9840–9852. [[CrossRef](#)]
51. Sadighi, S.; Ahmad, A.; Shirvani, M. Comparison of lumping approaches to predict the product yield in a dual bed VGO hydrocracker. *Int. J. Chem. React. Eng.* **2011**, *9*, 1–25. [[CrossRef](#)]
52. Sadighi, S. A two-dimensional discrete lumped model for a trickle-bed vacuum gas oil hydrocracking reactor. *Korean J. Chem. Eng.* **2016**, *33*, 2538–2546. [[CrossRef](#)]
53. Calderon, C.J.; Ancheyta, J. Dynamic modeling and simulation of a slurry-phase reactor for hydrotreating of oil fractions. *Energy Fuels* **2017**, *31*, 5691–5700. [[CrossRef](#)]
54. Elkamel, A.; Al-Ajmi, A.; Fahim, M. Modeling the hydrocracking process using artificial neural networks. *Pet. Sci. Technol.* **1999**, *17*, 931–954. [[CrossRef](#)]
55. Bahmani, M.; Sharifi, K.; Shirvani, M.; Street, M.; Street, N. Product yields prediction of tehran refinery hydrocracking unit using artificial neural networks. *Iran. J. Chem. Eng.* **2010**, *7*, 51.
56. Sadighi, S.; Zahedi, G.R. Comparison of kinetic-based and artificial neural network modeling methods for a pilot scale vacuum gas oil hydrocracking reactor. *Bull. Chem. React. Eng. Catal.* **2013**, *8*, 125. [[CrossRef](#)]
57. Sadighi, S.; Mohaddecy, S.R.S.; Abbasi, A. Modeling and Optimizing a Vacuum Gas Oil Hydrocracking Plant using an Artificial Neural Network. *Int. J. Technol.* **2018**, *9*, 99–109. [[CrossRef](#)]
58. Song, W.; Mahalec, V.; Long, J.; Yang, M.L.; Qian, F. Modeling the hydrocracking process with deep neural networks. *Ind. Eng. Chem. Res.* **2020**, *59*, 3077–3090. [[CrossRef](#)]
59. Wang, Y.; Sun, K.; Yuan, X.; Cao, Y.; Li, L.; Koivo, H.N. A novel sliding window PCA-IPF based steady-state detection framework and its industrial application. *IEEE Access* **2018**, *6*, 20995–21004. [[CrossRef](#)]
60. Xue, Y.; Wang, Y.; Shang, D. Parameter Optimization of Hydrocracker using Multi-block Kriging Metamodeling within Discontinuous Operating Space. In Proceedings of the 2019 12th Asian Control Conference (ASCC), Kitakyushu-shi, Japan, 9–12 June 2019; pp. 254–259.

61. ASTM-D2892-19. *Standard Test Method for Distillation of Crude Petroleum (15-Theoretical Plate Column)*; Standard, ASTM International: West Conshohocken, PA, USA, 2019.
62. ASTM-D5236-18a. *Standard Test Method for Distillation of Heavy Hydrocarbon Mixtures (Vacuum Potstill Method)*; Standard, ASTM International: West Conshohocken, PA, USA, 2018.
63. ASTM-D86-19. *Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure*; Standard, ASTM International: West Conshohocken, PA, USA, 2019.
64. ASTM-D2622-16. *Standard Test Method for Sulfur in Petroleum Products by Wavelength Dispersive X-ray Fluorescence Spectrometry*; Standard, ASTM International: West Conshohocken, PA, USA, 2016.
65. ASTM-D4629-17. *Standard Test Method for Trace Nitrogen in Liquid Hydrocarbons by Syringe/Inlet Oxidative Combustion and Chemiluminescence Detection*; Standard, ASTM International: West Conshohocken, PA, USA, 2017.
66. ASTM-D5863-00a. *Standard Test Methods for Determination of Nickel, Vanadium, Iron, and Sodium in Crude Oils and Residual Fuels by Flame Atomic Absorption Spectrometry*; Standard, ASTM International: West Conshohocken, PA, USA, 2016.
67. Ismail, A.A.; van de Voort, F.R.; Sedman, J. Fourier transform infrared spectroscopy: Principles and applications. In *Techniques and Instrumentation in Analytical Chemistry*; Elsevier: Amsterdam, The Netherlands, 1997; Volume 18, pp. 93–139.
68. Pasquini, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **2003**, *14*, 198–219. [[CrossRef](#)]
69. Förster, H. UV/vis spectroscopy. In *Characterization I*; Karge, H.G., Weitkamp, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 337–426.
70. Petry, R.; Schmitt, M.; Popp, J. Raman spectroscopy—a prospective tool in the life sciences. *ChemPhysChem* **2003**, *4*, 14–30. [[CrossRef](#)]
71. Workman JR, J.J.; Mobley, P.R.; Kowalski, B.R.; Bro, R. Review of chemometrics applied to spectroscopy: 1985–1995, Part I. *Appl. Spectrosc. Rev.* **1996**, *31*, 73–124. [[CrossRef](#)]
72. Mobley, P.R.; Kowalski, B.R.; Workman JR, J.J.; Bro, R. Review of chemometrics applied to spectroscopy: 1985–1995, part 2. *Appl. Spectrosc. Rev.* **1996**, *31*, 347–368. [[CrossRef](#)]
73. Bro, R.; Workman JR, J.J.; Mobley, P.R.; Kowalski, B.R. Review of chemometrics applied to spectroscopy: 1985–1995, Part 3—Multi-way analysis. *Appl. Spectrosc. Rev.* **1997**, *32*, 237–261. [[CrossRef](#)]
74. Silva, S.L.; Silva, A.M.; Ribeiro, J.C.; Martins, F.G.; Da Silva, F.A.; Silva, C.M. Chromatographic and spectroscopic analysis of heavy crude oil mixtures with emphasis in nuclear magnetic resonance spectroscopy: A review. *Anal. Chim. Acta* **2011**, *707*, 18–37. [[CrossRef](#)]
75. Filgueiras, P.R.; Terra, L.A.; Castro, E.V.; Oliveira, L.M.; Dias, J.C.; Poppi, R.J. Prediction of the distillation temperatures of crude oils using ¹H NMR and support vector regression with estimated confidence intervals. *Talanta* **2015**, *142*, 197–205. [[CrossRef](#)]
76. Molina, D.; Uribe, U.N.; Murgich, J. Partial least-squares (PLS) correlation between refined product yields and physicochemical properties with the ¹H nuclear magnetic resonance (NMR) spectra of Colombian crude oils. *Energy Fuels* **2007**, *21*, 1674–1680. [[CrossRef](#)]
77. Gilbert, W.R.; de Lima, S.G.; Bueno, A.F. Comparison of NIR and NMR Spectra Chemometrics for. In *Fluid Catalytic Cracking VI: Preparation and Characterization of Catalysts*; Elsevier: Amsterdam, The Netherlands, 2004; p. 203.
78. Pasquini, C.; Bueno, A.F. Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. *Fuel* **2007**, *86*, 1927–1934. [[CrossRef](#)]
79. Hidajat, K.; Chong, S. Quality characterisation of crude oils by partial least square calibration of NIR spectral profiles. *J. Near Infrared Spectrosc.* **2000**, *8*, 53–59. [[CrossRef](#)]
80. Laxalde, J.; Ruckebusch, C.; Devos, O.; Caillol, N.; Wahl, F.; Duponchel, L. Characterisation of heavy oils using near-infrared spectroscopy: Optimisation of pre-processing methods and variable selection. *Anal. Chim. Acta* **2011**, *705*, 227–234. [[CrossRef](#)]
81. Santos, V.O., Jr; Oliveira, F.C.; Lima, D.G.; Petry, A.C.; Garcia, E.; Suarez, P.A.; Rubim, J.C. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. *Anal. Chim. Acta* **2005**, *547*, 188–196. [[CrossRef](#)]
82. Laxalde, J.; Caillol, N.; Wahl, F.; Ruckebusch, C.; Duponchel, L. Combining near and mid infrared spectroscopy for heavy oil characterisation. *Fuel* **2014**, *133*, 310–316. [[CrossRef](#)]

83. Buchanan, B., Recent Advances in the Use of Near-IR Spectroscopy in the Petrochemical Industry. In *Handbook of Near-Infrared Analysis*, 3rd ed.; Burns, D.A., Ciurczak, E.W., Eds.; CRC Press: Boca Raton, FL, USA, 2007; pp. 521–529.
84. Swarin, S.J.; Drumm, C.A. Prediction of gasoline properties with near-infrared spectroscopy and chemometrics. *SAE Trans.* **1991**, *100*, 1110–1118.
85. Baldrich Ferrer, C.A. Prediction of physicochemical properties of FCC feedstock by chemometric analysis of their ultraviolet spectrum. *Ct&F-Cienc. Tecnol. Futuro* **2008**, *3*, 143–156.
86. Evdokimov, I.N.; Losev, A.P. Potential of UV-visible absorption spectroscopy for characterizing crude petroleum oils. *Oil Gas Bus.* **2007**, *1*, 1–21.
87. Baldrich, C.A.; Novoa, L.A.; Bueno, A. Comparison between NIR and UVVIS spectra chemometrics for predicting FCC feedstocks properties. *Ct&F-Cienc. Tecnol. Futuro* **2010**, *4*, 113–125.
88. Pasadakis, N.; Sourligas, S.; Foteinopoulos, C. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. *Fuel* **2006**, *85*, 1131–1137. [[CrossRef](#)]
89. Breikreitz, M.C.; Raimundo, I.M., Jr.; Rohwedder, J.J.; Pasquini, C.; Dantas Filho, H.A.; José, G.E.; Araújo, M.C. Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *Analyst* **2003**, *128*, 1204–1207. [[CrossRef](#)]
90. Molina, D.; Uribe, U.N.; Murgich, J. Correlations between SARA fractions and physicochemical properties with ¹H NMR spectra of vacuum residues from Colombian crude oils. *Fuel* **2010**, *89*, 185–192. [[CrossRef](#)]
91. Muhammad, A.; de Vasconcellos Azeredo, R.B. ¹H NMR spectroscopy and low-field relaxometry for predicting viscosity and API gravity of Brazilian crude oils—A comparative study. *Fuel* **2014**, *130*, 126–134. [[CrossRef](#)]
92. Flórez, M.; Guerrero, J.; Cabanzo, R.; Mejía-Ospino, E. SARA analysis and Conradson carbon residue prediction of Colombian crude oils using PLSR and Raman spectroscopy. *J. Pet. Sci. Eng.* **2017**, *156*, 966–970. [[CrossRef](#)]
93. Lababidi, H.M.; Chedadeh, D.; Riazi, M.; Al-Qattan, A.; Al-Adwani, H.A. Prediction of product quality for catalytic hydrocracking of vacuum gas oil. *Fuel* **2011**, *90*, 719–727. [[CrossRef](#)]
94. Yuan, X.; Zhou, J.; Wang, Y. A Comparative Study of Adaptive Soft Sensors for Quality Prediction in an Industrial Refining Hydrocracking Process. In Proceedings of the 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), Enshi, China, 25–27 May 2018; pp. 1064–1068.
95. Wang, Y.; Wu, D.; Yuan, X. A two-layer ensemble learning framework for data-driven soft sensor of the diesel attributes in an industrial hydrocracking process. *J. Chemom.* **2019**, *33*, e3185. [[CrossRef](#)]
96. Lukec, I.; Sertić-Bionda, K.; Lukec, D. Prediction of sulphur content in the industrial hydrotreatment process. *Fuel Process. Technol.* **2008**, *89*, 292–300. [[CrossRef](#)]
97. Camacho, E.F.; Alba, C.B. *Model Predictive Control*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
98. Lababidi, H.; Alatiqi, I.; Ali, Y. Constrained model predictive control for a pilot hydrotreating plant. *Chem. Eng. Res. Des.* **2004**, *82*, 1293–1304. [[CrossRef](#)]
99. Aydın, E.; Arkun, Y.; Is, G.; Mutlu, M.; Dikbas, M. Plant-wide optimization and control of an industrial diesel hydro-processing plant. *Comput. Chem. Eng.* **2016**, *87*, 234–245. [[CrossRef](#)]
100. Alhajree, I.; Zahedi, G.; Manan, Z.; Zadeh, S.M. Modeling and optimization of an industrial hydrocracker plant. *J. Pet. Sci. Eng.* **2011**, *78*, 627–636. [[CrossRef](#)]
101. Zhou, H.; Lu, J.; Cao, Z.; Shi, J.; Pan, M.; Li, W.; Jiang, Q. Modeling and optimization of an industrial hydrocracking unit to improve the yield of diesel or kerosene. *Fuel* **2011**, *90*, 3521–3530. [[CrossRef](#)]
102. Zhong, W.; Qiao, C.; Peng, X.; Li, Z.; Fan, C.; Qian, F. Operation optimization of hydrocracking process based on Kriging surrogate model. *Control Eng. Pract.* **2019**, *85*, 34–40. [[CrossRef](#)]
103. Sildir, H.; Arkun, Y.; Cakal, B.; Gokce, D.; Kuzu, E. Plant-wide hierarchical optimization and control of an industrial hydrocracking process. *J. Process Control* **2013**, *23*, 1229–1240. [[CrossRef](#)]
104. Altissimi, R.; Brambilla, A.; Deidda, A.; Semino, D. Optimal operation of a separation plant using artificial neural networks. *Comput. Chem. Eng.* **1998**, *22*, S939–S942. [[CrossRef](#)]
105. Zhao, J.; Chen, B.; Shen, J. Multidimensional non-orthogonal wavelet-sigmoid basis function neural network for dynamic process fault diagnosis. *Comput. Chem. Eng.* **1998**, *23*, 83–92. [[CrossRef](#)]
106. Stockmann, M.; Haber, R.; Schmitz, U. Source identification of plant-wide faults based on k nearest neighbor time delay estimation. *J. Process Control* **2012**, *22*, 583–598. [[CrossRef](#)]

107. Zheng, S.; Zhao, J. A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis. *Comput. Chem. Eng.* **2020**, *135*, 106755. [[CrossRef](#)]
108. AlGhazzawi, A.; Lennox, B. Monitoring a complex refining process using multivariate statistics. *Control Eng. Pract.* **2008**, *16*, 294–307. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).