

Article

A Physics-Informed Combinatorial Digital Twin for Value-Optimized Production of Petroleum Coke

Vladimir V. Bukhtoyarov ¹, Alexey A. Gorodov ¹, Natalia A. Shepeta ¹, Ivan S. Nekrasov ^{1,*}, Oleg A. Kolenchukov ¹, Svetlana S. Kositsyna ² and Artem Y. Mikhaylov ¹

¹ Department of Technological Machines and Equipment of Oil and Gas Complex, School of Petroleum and Natural Gas Engineering, Siberian Federal University, 660041 Krasnoyarsk, Russia; vladber@list.ru (V.V.B.); glexx84@mail.ru (A.A.G.); nshepeta@sfu-kras.ru (N.A.S.); okolenchukov@sfu-kras.ru (O.A.K.); amikhaylov@sfu-kras.ru (A.Y.M.)

² Department of Chemistry and Technology of Natural Energy Carriers and Carbon Materials, School of Petroleum and Natural Gas Engineering, Siberian Federal University, 660041 Krasnoyarsk, Russia

* Correspondence: nekrasov-is@ya.ru

Abstract

Petroleum coke quality strongly influences refinery economics and downstream energy use, yet real-time control is constrained by slow quality assays and a 24–48 h lag in laboratory results. This study introduces a physics-informed combinatorial digital twin for value-optimized coking, aimed at improving energy efficiency and environmental performance through adaptive quality forecasting. The approach builds a modular library of 32 candidate equations grouped into eight quality parameters and links them via cross-parameter dependencies. A two-level optimization scheme is applied: a genetic algorithm selects the best model combination, while a secondary loop tunes parameters under a multi-objective fitness function balancing accuracy, interpretability, and computational cost. Validation on five clustered operating regimes (industrial patterns augmented with noise-perturbed synthetic data) shows that optimal model ensembles outperform single best models, achieving typical cluster errors of ~7–13% NMAE. The developed digital twin framework enables accurate prediction of coke quality parameters that are critical for its energy applications, such as volatile matter and sulfur content, which serve as direct proxies for estimating the net calorific value and environmental footprint of coke as a fuel.

Keywords: physics-informed digital twin; delayed coking; petroleum coke; combinatorial modeling; adaptive model constructor; genetic algorithm; multi-objective optimization; volatile matter; sulfur content; cross-parameter dependence

Academic Editors: Shengyu Tao and Lianlei Lin

Received: 6 December 2025

Revised: 13 January 2026

Accepted: 14 January 2026

Published: 16 January 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

For modern facilities in the fuel and energy complex, such as oil refineries, diversification of product and resource bases has become one of the key directions for improving the efficiency of operational activities. One of the development paths is deepening of oil refining through the commissioning of new coking units or expanding the operational capabilities of existing ones.

The effectiveness of such measures can be substantially enhanced through the use of modern approaches, such as analytical support of technological processes using digital twins (DTs). Their application is possible both during operation and during the design

and development phases of industrial facilities, serving as tools for increasing energy efficiency and environmental performance in the fuel and energy sector.

Despite the ongoing energy transition, oil and gas value chains remain critical for global energy supply and industrial feedstocks, motivating continued research in flow assurance and risk prediction (e.g., hydrate formation in deep-water wellbore/subsea systems [1]), stimulation efficiency (e.g., fracture propagation in coal reservoirs [2]), and sub-surface re-use options such as CO₂ sequestration and hydrogen storage in depleted gas reservoirs [3].

In recent years, the share of light crude oils on the market has been decreasing, and they are increasingly being replaced with heavier fractions [4,5], which has led to the growing popularity of processes aimed at upgrading heavy fractions into more valuable light products [6,7].

As a result of these changes, the volume of extremely heavy residues has been increasing, with the majority (approximately 63 wt%) of petroleum residues processed through thermal conversion processes [6], such as coking—producing petroleum coke (petcoke), including through delayed coking units [8] and fluid coking units [9]. Coking is a thermal cracking process in which the feed decomposes into lower-boiling products, resulting in the formation of gas (≈13 wt%), naphtha (≈11 wt%), middle distillate (≈45 wt%), and petroleum coke (≈31 wt%).

Petroleum coke is a key product of deep refining heavy petroleum residues and plays a central role in the technological chain of modern refineries. Its quality characteristics directly determine both the economic efficiency of the coking process and the market value of the final product [10,11], with a sufficiently high yield of vacuum residue relative to the total volume of processed crude oil [12].

The industrial coking process can be implemented using three main types of units:

- Batch (in coke tubes);
- Semi-continuous (in unheated coke drums);
- Continuous (in fluidized beds of coke–heat carrier).

Delayed (semi-continuous) coking is the most widely used technology for petroleum coke production in refineries, due to its relatively low investment cost, flexibility in processing various residual feeds, and the ability to produce high-quality products [13].

Delayed coking is a thermal cracking process of heavy petroleum residues carried out under controlled residence time and temperature conditions. The physicochemical essence of the process involves stepwise transformation of high-molecular-weight components through stages of dealkylation, condensation, and aromatization with the formation of a solid carbonaceous residue. Critical stages include initiation of radical reactions at 400–450 °C, mesophase formation via orientation of polycyclic aromatic structures, and subsequent polycondensation leading to the formation of a coke matrix.

Modern trends in processing heavy and extra-heavy crude oils—characterized by elevated sulfur, metals, and asphaltene content—require removing larger amounts of residues, heavier oils, and sulfur [14].

The global petroleum coke market is showing stable growth, with production volumes projected to increase through 2037, which intensifies requirements for precise control of product quality and creates additional challenges for traditional monitoring methods.

The primary feedstock types used for coke production include high-molecular-weight petroleum (crude, heavy oils) and petroleum residues: vacuum residues, light and heavy gas oils (vacuum and secondary processing), pyrolysis tars, thermal cracking tar, deasphalting bitumens, and others. Heavy residues are mixtures of high-molecular-weight hydrocarbons and compounds containing not only carbon and hydrogen but also heteroatoms—sulfur, oxygen, nitrogen—and, to a lesser extent, metals such as vanadium,

nickel, cobalt, iron, molybdenum, titanium, etc. Due to the diversity of isomeric forms and combinations of aliphatic, hydroaromatic, aromatic hydrocarbons and their derivatives, a wide spectrum of high-molecular-weight components is formed. In industrial practice, residual products are commonly classified into structural groups (oils, resins, asphaltenes) [15].

Depending on feedstock properties and operating conditions of delayed coking, various types of coke can be obtained and classified using different criteria.

According to physical structure, petroleum coke may be categorized as sponge coke and shot coke [16–18].

From a morphological perspective:

- Spherical coke (isotropic, amorphous, nearly non-porous);
- Sponge coke (semi-isotropic);
- Needle coke (anisotropic, with regular crystalline structure, containing numerous micropores, 4–7 nm crystallite sizes).

Depending on application, coke may be categorized as fuel-grade (for cement and power industries), anode-grade (for aluminum production), and electrode-grade (for steelmaking) [19,20].

Petroleum coke can be of two types: green petroleum coke and calcined petroleum coke. Petroleum coke obtained without calcination is called green coke. Calcination is carried out in a rotary or shaft kiln to remove the remaining hydrocarbons by heating the green coke to about 1300–1500 °C. During calcination, coke undergoes further decomposition, and the carbon-to-hydrogen ratio increases from about 20 in green coke to about 1000 in calcined coke [7].

From an operational standpoint, petroleum coke value is determined not only by yield but also by a set of quality indicators that are expensive and slow to measure online.

Petroleum coke is a solid, porous, black–gray petroleum-derived material obtained through delayed coking of petroleum residues. The elemental composition of raw (uncalcined) petroleum coke is approximately (wt%): 91–99.5% carbon, with the remainder consisting of impurities including hydrogen (0.035–4%), sulfur (0.5–8%), nitrogen and oxygen (1.3–3.8%), as well as metals.

Key quality indicators are maximum sulfur, ash and moisture contents (<3 wt%), volatile matter yield (on heating), particle size distribution (screening), and mechanical strength (under load). Petroleum cokes are classified by the following:

- Sulfur content: low sulfur (up to 1%), medium sulfur (up to 2%), high sulfur (>2%);
- Ash content: low ash (up to 0.5%), medium ash (0.5–0.8%), high ash (>0.8%);
- Granulometric composition: lump coke (particle size > 25 mm), nut coke (6–25 mm), fines (<6 mm).

Additional coke characteristics influencing its application include porosity 16–56%; density at 20 °C—true density 2.04–2.13 g/cm³, apparent density 0.8–1.4 g/cm³; bulk density 400–500 kg/m³; specific electrical resistivity (800–1000) μΩ·m [21].

Coke obtained after coking is a chemically stable and inert material. To impart electrode properties, coke is calcined at 1200–1300 °C in calcination furnaces. The structure of calcined petroleum coke consists of crystallites of various sizes and orientations.

The main consumers of petroleum coke (provided its quality meets requirements) are the non-ferrous and ferrous metallurgy sectors, where coke acts as a base for conductive products—anodes and anode mass for aluminum smelting, and electrodes to produce electric steel, ferroalloys, carbides, chlorine, alkalis, abrasive materials, etc. For these purposes, low sulfur calcined cokes with sulfur content up to 1.5% are required. Petroleum coke with medium to high sulfur content (6–8 wt%) is also employed as an effective reducing and sulfidizing agent in the shaft smelting of oxidized nickel and copper ores.

Petroleum coke is used in the production of carbides (CaC_2 , SiC , TiC , B_4C , etc.), with particular attention to calcium and silicon carbides—important components for acetylene and abrasive production (manufactured from petroleum coke with particle size less than 8 mm), and in ferroalloys—as alloying additives that improve steel quality. One of the primary requirements for carbonaceous reductants is high porosity, which ensures good gas permeability and uniform descent of the burden while preserving sorption and filtration properties.

Special purpose cokes (for structural graphite products, lining of nuclear reactors) must have an isotropic (point, spherulitic) structure.

Low-grade petroleum cokes can be used as fuel. The heat of combustion of carbonaceous materials primarily depends on the C/H ratio, and therefore on the volatile matter content. As the volatile matter content increases from 3.2 to 7.0% (i.e., as the hydrogen fraction in the coke increases), the specific heat of combustion increases from 7100 to 8500 kcal/kg (29.7 to 35.6 MJ/kg). Pulverized petroleum coke is most widely used as fuel in the cement industry [22,23].

In the metallurgical industry, coke is the base component for anode mass and electrode production. Here, the requirements for sulfur content, volatile matter, and mechanical strength are extremely stringent—even minor deviations from specifications can cause multimillion losses. In the energy sector, coke is used to produce fuel cells, where parameters such as porosity and reactivity are critical for process stability [24]. Hybrid fuels based on petroleum coke and natural gas are also gaining popularity, improving combustion and reducing CO_2 emissions [25].

In anode production for the aluminum industry, sponge coke with high volatile content and large specific surface area is used. Requirements include sulfur content up to 2%, true density 2.02–2.12 g/cm³ and ash content not higher than 0.6%.

For the production of electrodes for steelmaking and synthetic graphite, only needle coke can be used—the most valuable product of the delayed coking process, with tightened quality requirements. Important characteristics of needle (anisotropic) coke are coefficient of linear thermal expansion (CTE) up to $4 \times 10^{-7} (\text{°C})^{-1}$, density greater than 2.12 g/cm³, microstructure rated 5–6 points, content of micro impurities (vanadium and nickel) not more than 10 ppm, sulfur not more than 0.8%, ash less than 0.2% [21].

Elevated sulfur content in needle coke used for electrode manufacture can lead to cracking of electrodes under high temperature operation due to rupture of C–S bonds. The presence of vanadium and nickel compounds deteriorates the thermophysical properties of electrodes. Furthermore, vanadium and nickel (especially in the presence of sodium) catalyze oxidation of carbon by oxygen, thus increasing anode consumption during electrolysis. In high-quality needle coke, sulfur content must not exceed about 0.6%, vanadium and nickel contents must be low (typically not more than ~10 ppm each).

High ash content in coke limits its application in the electrode industry and worsens its electrical performance (conductivity, specific electrical resistance) [26–28].

As a feedstock for anodes and electrodes, usually the coarse fraction (25–200 mm) of pitch and petroleum cokes is used. On the other hand, fine-grained coke has higher specific resistivity and higher mechanical strength. Isotropic coke has higher mechanical strength than needle coke [29].

The properties of produced petroleum coke depend on the type and origin of the feed, as well as on process parameters.

The most important variables are the hydrocarbon components present in the feed, which range from highly resinous and asphaltenic to aromatic compounds [30,31].

Coke from oils with a predominance of paraffinic structures has a needle like structure, with wide slit-shaped pores and predominantly elongated particles, reflecting growth of carbonaceous structures in one direction; asphaltenic coke has a continuous

isotropic structure, resulting from uniform growth of dense colloidal structures in all directions; tar-derived coke occupies an intermediate position between oil and asphaltene cokes, has a sponge structure with low anisotropy and inclusions of fibrous domains with high thermal expansion coefficient.

A high concentration of asphaltenes and resins leads to formation of coke with rounded pores; conversely, a low content of these components yields highly porous coke with elongated large pores [32].

Sulfur content in coke mainly depends on the composition and origin of the feed. Depending on the nature of the crude, coke contains 20–35% of the sulfur present in the original residue. As the content of paraffinic naphthenic hydrocarbons in the feed increases, sulfur content in the coke tends to rise. In some cases, the sulfur content in coke can be significantly higher than in the feed, and in other cases, lower. During coking of straight run residues, the sulfur content in the coke relative to that in the feed usually increases; during coking of cracking residues, it decreases. Therefore, quantitative sulfur contents in feed and coke can only be determined experimentally (in the laboratory).

Ash content of coke and metal content directly depend on the presence of mechanical impurities, salts, and organometallic compounds in the feed, which almost completely accumulate in the solid coking residue [23]. Fine-grained coke (particles < 8 mm) is characterized by higher ash and sulfur contents compared with lump coke [29].

To predict coke yield, the most common indicator used is the feed's coking tendency determined by the Conradson carbon residue method (CCR). However, if only CCR is used to predict coke yield in delayed coking units, the error can reach $\pm 25\%$. In this case, the following pattern is observed: as CCR increases, the yields of coke, gasoline, and gas in delayed coking increase, while the yield of gas oil fractions decreases [29].

As density of residual products increases, their coking tendency also increases, due to the higher content of resins and asphaltenes. Coke yield is predicted more accurately when both CCR and feed density are taken into account. However, these dependencies are nonlinear. When coking highly resinous residues, the coke yield exceeds the coking tendency value by about 1.5 times, whereas when coking residues with lower coking tendency, the coke yield exceeds it by 1.8–2 times [33].

The chemical and physical properties taken into account when choosing suitable feed for producing needle coke can be described as follows: the feed should be highly aromatic, containing 60–85% aromatic hydrocarbons; the initial boiling point should be above 250 °C, with no more than 25–30% of the material boiling below 360 °C; the feed should have a low API gravity (i.e., high density); sulfur content should be low, preferably less than 1 wt% [4].

The main process parameters influencing coke yield and quality are coking temperature, drum pressure, coking time, and recycling ratio [34].

Coking temperature and coking time determine the volatile matter content in green coke. The higher these parameters, the lower the volatile matter content. A normal volatile matter content in coke is 8–12%. Elevated volatile content indicates incomplete coking and negatively affects coke quality. Such coke is known as “under baked” coke. It is characterized by lower density, reduced mechanical strength, and increased reactivity.

High coking temperature leads to reduced volatile matter and increased strength, but also to greater fissuring. Higher temperature also increases the risk of coke formation in furnace tubes and shortens their lifetime. An optimal coking temperature is around 490 °C. For example, when vacuum residue is used as feed and coked at 490–495 °C, volatile matter in coke is 13%. When temperature is lowered to 440–460 °C, the volatile matter yield increases by more than 30%, to about 19%. As temperature rises above 520 °C, volatile matter decreases to about 6%.

For coke obtained from lighter feed, achieving an acceptable volatile matter level (around 10%) requires raising the coking temperature by 5–10 °C compared to heavy feed. At low coking temperature, volatile matter content and grindability of the coke increase. Calcination of such coke produces a material with high porosity and low bulk density. Reduced hardness leads to smaller mean particle size and greater content of fines (<25 mm) [16,35].

With increasing pressure, yields of coke and gas increase, but the total yield of liquid products decreases.

As delayed coking drum pressure increases from 0.10 to 2.50 MPa, coke yield grows from 9.1 to 44.8%, distillate yield decreases from 84.5 to 4.3%, and gas yield rises from 6.4 to 50.9%. The increase in coke yield and decrease in distillate yield with pressure are due to the involvement in coking of additional fractions that would otherwise evaporate at lower pressure and not participate in coke formation [36].

At the same time, coke quality characteristics do not depend linearly on pressure. At 0.8 MPa, coke with the best microstructure (5.2 points) and reduced ash content is obtained [37].

When vacuum residue is used as feed and pressure is increased from 0.1 to 2.5 MPa, coke yield rises significantly from 14.0 to 46.2%, while distillate yield decreases from 73.7 to 14.3%, with a higher fraction of light components in the distillate. The best microstructure of coke from vacuum residue (4.8 points) is obtained at a pressure of 0.4 MPa [19,38].

The recycle ratio is defined as the ratio of the total feed to the coking unit (fresh feed plus recycle) to the amount of fresh feed supplied from the upstream refinery units. An increase in the recycle ratio reduces the vanadium and nickel content in the coke, increases its anisotropy, and decreases its mechanical strength. In delayed coking units, the recycle ratio typically ranges from 1.03 to 1.30 [7,29]. The highest values are used in industrial units producing needle coke, while the lowest are used in units where coke yield should be minimized.

From a data perspective, delayed coking is characterized by sparse laboratory targets, measurement noise, and substantial delays between process changes and quality assays, which motivates robust predictive soft-sensing.

The coking process is a complex, multi-parameter system with pronounced nonlinear dependencies, where final product properties are formed under the influence of hundreds of interrelated factors, many of which are mutually dependent.

In practice, monitoring the coking process is often difficult and faces significant methodological limitations [29]. Prompt determination of parameters such as volatile matter is complicated by the duration of standard tests, which makes it impossible to use results for real-time control. Sulfur analysis by X-ray fluorescence offers acceptable accuracy ($\pm 0.1\%$) but requires calibration for each specific feed type. Measurement of porosity provides detailed information about pore size distribution but is expensive and cannot be implemented online. Similar difficulties arise with other product quality parameters when they must be determined in a timely manner for operating control and decision-making for process correction.

Another serious issue in modeling the coking process is incompleteness and noise in industrial data. This remains a critical challenge [39], as operating measurements often contain significant errors. Moreover, key parameters such as particle size distribution or content of specific metals may not be available in real-time [40]. A major limitation is also the time lag between changes in process parameters and receipt of laboratory coke quality data, which can reach 24–48 h, ruling out real-time control [41,42].

Additional complexity arises from variability in feed streams: modern refineries process blends of different residues [43,44], whose composition can change several times a

day. This imposes requirements on models to be adaptive, able to account for dynamic changes in input without loss of predictive accuracy.

Particularly challenging is assigning coke to its appropriate quality class. This typically requires expert assessment performed on the entire product obtained in a single coke drum. Therefore, when building predictive estimates of petroleum coke quality, statistical methods are usually employed.

These operational and data constraints motivate an adaptive, interpretable modeling framework that can switch between mechanism-consistent equation families depending on regime and data availability.

The objective of this work is to develop a combinatorial approach enabling intelligent control of processes for processing petroleum waste at the stages of design and re-engineering. The approach combines the accuracy of modern mathematical methods with the practicality and interpretability required for industrial implementation, to improve energy efficiency and environmental performance of plants that use coke as a raw material base.

In addition to product quality, the decarbonization agenda makes greenhouse-gas (GHG) performance a key operational KPI for delayed coking units. Delayed coking is associated with high energy demand and, consequently, significant carbon emissions, which motivates optimization frameworks that explicitly include Global Warming Potential (GWP) and energy-intensity indicators alongside economic objectives.

Existing delayed-coking prediction models can be grouped into the following:

- Empirical correlations;
- Mechanistic/dynamic models;
- Purely data-driven machine-learning (ML) models;
- Hybrid approaches.

Empirical correlations are typically interpretable but degrade under feedstock shifts and outside the calibration domain. Mechanistic/dynamic models can be physically rigorous, yet their parameterization is labor-intensive, and they may be computationally costly for routine plant deployment. ML models often provide strong accuracy on large datasets but remain difficult to validate and explain for process control (“black-box” barrier) and may require more labeled cycles than are available at most refineries. Hybrid models mitigate some weaknesses but usually couple one mechanism model with one data-driven estimator and rarely formalize cross-parameter dependence among multiple quality indicators within a unified, selectable model library. These limitations motivate a framework that is adaptive across regimes, interpretable by design, and computationally feasible for industrial operation.

In contrast to the traditional methods discussed above, this study proposes a fundamentally new concept for constructing a digital twin, based on the following key innovations:

- Combinatorial library of physics-informed models.

Instead of searching for a single universal model, we propose an adaptive selection system operating on a library of 32 specialized parametric models designed to predict eight key petroleum coke quality indicators. Each model in the library possesses a clear physicochemical interpretation, thereby overcoming the “black-box” limitation inherent to neural-network-based approaches.

- Two-level heuristic optimization architecture.

To efficiently explore a search space comprising millions of possible model combinations, a two-level optimization scheme is developed. The upper level employs a genetic algorithm to identify the optimal structure of the model ensemble, while the lower level

performs fine tuning of the parameters of the selected models. This architecture ensures an effective balance between global exploration and local accuracy.

- Explicit mechanism for cross-parameter dependency handling.

The system does not treat predictions of individual quality indicators as independent. A formal mechanism is introduced to account for mutual interactions between models through shared intermediate variables (e.g., the influence of predicted porosity on the calculation of mechanical strength). This approach enables representation of the coupled physical nature of the process and improves the internal consistency of predictions.

- Multi-objective fitness function balancing accuracy, complexity, and interpretability.

The optimality criterion of the combinatorial model constructor extends beyond pure error minimization. It explicitly incorporates penalties for computational complexity and includes an expert-based interpretability score, which is critically important for practical industrial deployment and for supporting technology-driven decision-making in refinery operations.

2. Modern Approaches to Modeling Coking Processes

The main challenges of contemporary modeling include the multifactorial nature of the process, where temperature regimes, pressure, heating rate, and feed properties create complex cross effects.

Mathematical descriptions of these parameters and models are often expressed as nonlinear systems of differential equations, including not only ordinary but also parabolic and hyperbolic partial differential equations. To make qualitative or numerical analysis feasible, such models must first be simplified [45].

Experimental data themselves show that a change in maximum coking temperature by 10 °C can result in up to a 15% variation in volatile matter content [29]; moreover, this dependence is strongly nonlinear and depends on feed composition [46,47].

In addition, a number of works model not so much the coking process itself as the subsequent processing of coke and its conversion into a marketable product [40], often using rather complex mathematical descriptions [48].

Continuously increasing requirements for product quality and the focus on resource saving and greater energy efficiency of processes determine the need to equip production with advanced prediction, control, and management tools. At the same time, it is not always possible to obtain data on the characteristics and parameters of dynamically changing processes in units in real time or with sufficient detail, which reduces the timeliness and accuracy of management decisions. This limits the system's ability to rapidly adapt to process disturbances, which is critical for maintaining optimal product quality and yield.

Predictive modeling can compensate for these deficiencies by enabling selection of necessary parameters based on mathematical models and providing process data without physical measurements. Virtual models can effectively predict changes in any refining process and in the quality of the resulting products [49].

Modeling of chemical technological processes makes it possible to minimize intervention in operating units, since modeling involves creating a virtual model, and changes to its operation pose no danger to personnel or the environment. This is one of the advantages of modeling chemical and technological processes.

Using virtual model descriptions offers significant advantages: based on data on feed variability and process behavior, it is possible to predict product yields and characteristics, and to develop optimization proposals without the cost of additional measuring equipment and laboratory tests.

Traditional approaches to modeling the coking process can be conventionally divided into several categories: empirical correlations [50–52]; physicochemical models [46,53]; models using fuzzy logic [54]; hybrid models [55]; and machine learning (ML) models [56]. Each of these approaches exhibits systemic limitations in industrial applications.

Empirical models based on statistical processing of historical data offer good interpretability but often prove inadequate outside the calibration range or when process conditions deviate from the original range. Analysis of the literature shows that standard correlations for predicting sulfur and metal contents in coke can introduce additional error when feed composition varies, due to catalytic effects of metals that are not accounted for in simplified models. Furthermore, some of these correlations are difficult to distinguish from the next categories of models.

Physicochemical models describing mechanisms of pyrolysis and condensation are theoretically capable of high accuracy, but their practical use is limited by extremely complex parameterization and high computational costs. Attempts to build deterministic models that take into account all significant factors lead to systems of hundreds of differential equations, whose solution demands specialized software and significant computation time [45].

Models based on fuzzy logic have a number of drawbacks when applied to modeling the coking process. These limitations are related both to the method itself and to the complexity of the process and data specifics, and they require high qualifications from decision makers. One of the ways to overcome these issues is hybridization of the modeling process, including the use of ML models.

Machine learning models, despite impressive results in academic studies, face implementation problems under industrial conditions. The “black box” nature of neural networks becomes a barrier to adoption for process control, where understanding cause and effect relationships is critical. In addition, these architectures often require large amounts of data that are not available at most industrial units.

2.1. *Physicomathematical and Kinetic Models*

Modeling of chemical transformations involves developing mathematical models that describe mechanisms, kinetics, and dynamics of chemical reactions without treating the detailed design and operating modes of equipment. Mathematical description of reaction rates is based on the reaction mechanism (sequence of elementary steps), kinetic parameters (rate constants, reaction orders, activation energies), and dependencies of rate on reactant concentrations, temperature, and other conditions.

The key stages of modeling are as follows:

- Determining the reaction mechanism (sequence of elementary steps and intermediates);
- Writing kinetic equations (for each step, rate equations are formulated using the law of mass action);
- Parameter estimation (experimental or calculated determination of rate constants k , often described by the Arrhenius equation, activation energy, and other characteristics);
- Solving the system of equations (numerical integration of differential equations that describe concentration changes with time);
- Verification of model adequacy (comparison of calculated data with experimental results).

Kinetic modeling enables the following:

- Prediction of reaction progress (calculation of product yields, side reactions, and concentration time profiles);

- Optimization of process conditions (determining optimal temperature, pressure, and concentrations to maximize target product yield);
- Analysis of reaction mechanisms (identification of sequence of steps and rate limiting steps);
- Design of industrial units (modeling spatial distributions of parameters such as temperature and concentrations, e.g., in reactors);
- Prediction of safety issues (risk assessment, e.g., thermal runaway).

The laws of chemical kinetics have long been applied to modeling refining processes, as reflected in a large body of work [57,58], including for delayed coking [59]. It has been established that delayed coking proceeds via parallel reactions of feed conversion into distillate and an intermediate mesophase, which then interacts and forms coke. Asphaltenes in the feed are the main coke-forming fraction [60,61].

In most refining processes (catalytic reforming, hydrotreating, hydroprocessing, catalytic cracking, thermal coking, etc.), the feed has a complex composition whose molecular characterization is practically impossible. In such cases, lumping methods are used, where feed constituents are grouped into pseudo components based on similar properties. The number of kinetic reactions (and therefore equations) is determined by the number of lumped components, which are considered as homogeneous ensembles [62].

In [63], a 15-component kinetic model of catalytic cracking is considered, which accounts for both quantitative (including coke yield) and qualitative characteristics of obtained products (light gas oil and catalytic cracking gasoline).

For delayed coking, kinetic modeling with division of feed and products into several lumped components has also been used. In [64], an 11-component model is proposed: 6 feed components (including asphaltenes, heavy and soft resins) and 5 product components (gas, naphtha, two mid cut fractions, and coke) to track yield curves in a delayed coking unit. Satisfactory agreement was obtained between the model and laboratory measurements for a wide range of feed compositions and coking temperatures.

In [65], results are presented for modeling the rate of coke formation and accumulation inside furnace tubes of a crude distillation unit. The model is based on heat transfer laws through the coke layer and tube wall (conductivity, convection from fired gas, radiation) and on non-elementary kinetic reaction laws. Comparing measured data from various units with model predicted coke deposits inside the furnace tubes gave deviations from 3 to 15%. The modeling results were used to optimize furnace maintenance (decoking) schedules, reducing operating costs by 20%.

2.2. Machine Learning Models

Machine learning (ML) and artificial intelligence (AI) models use ML/AI algorithms (neural networks, decision trees, clustering algorithms, support vector machines, etc.) to model complex relationships between process parameters and product properties. Such models can adapt to changes in refining processes and can improve their accuracy over time.

The development of software for various refining and petrochemical units based on neural-network modeling makes it possible to describe the behavior of nonlinear dynamic systems by extracting knowledge from all available process measurements [66,67].

For example, ref. [68] presents a nonlinear model based on multilayer neural networks to estimate diesel fuel quality produced in an atmospheric crude distillation unit. Such models can be used as an alternative to laboratory testing for quality assessment of refinery products.

In [69], delayed coking is modeled using artificial neural networks: a multilayer perceptron (MLP) and a radial basis function network (RBF). Input variables for the networks were feed API gravity and CCR (wt%). Network outputs (predictions) were coke yield

(wt%) and percentage yields of light gases, gasoline, gas oil and C₅+ fraction. Among MLP architectures, the network with 31 hidden neurons gave the best results. The RBF network with 20 centers (radii) was the best for delayed coking unit evaluation. Prediction error did not exceed 1.1%.

In [70], delayed coking is modeled for process prediction and optimization using two approaches: a generalized regression neural network and a double-strand DNA-based genetic algorithm for smoothing parameter optimization.

2.3. Hybrid Models

Hybrid models combine elements of physics, mathematical, and statistical models with ML/AI methods to improve prediction accuracy and robustness. Such models compensate limitations of one approach with the strengths of another, e.g., combining the accuracy of physical models with the adaptability of machine learning.

This type of model has found applications in related areas of refining, in particular in diesel fuel production. In [55], a hybrid model is used to predict diesel fuel quality in a hydrodesulfurization (HDS) process, integrating vector quantization (VQ) with a support vector regression (SVR) model. Model parameters were determined by hybrid optimization based on a genetic algorithm (GA) and sequential quadratic programming (SQP). A wide set of experimental HDS data was used for training and testing the SVR model. The resulting model predicted sulfur content in diesel with high accuracy (absolute relative error 0.0745, $R^2 = 0.997$) and minimal computation time ($T = 56$ s).

For continuous assessment of diesel fuel characteristics, a hybrid model based on linear regression and artificial neural networks was developed using real process temperature measurements [71].

The modeling process generally includes several stages. First, data collection is required: process parameters (e.g., temperature, pressure, feed rate, etc.) and laboratory analysis of feed and products. Then, preliminary data processing, integration from different sources, and data quality analysis are performed using statistical methods and/or ML/AI. Next, a model is selected (from simple regression to more complex neural network, ML, decision tree or deep learning models) based on requirements for accuracy, computation speed, process complexity, etc. At this stage, missing/erroneous data must be handled using specialized algorithms [72], such as the expectation maximization algorithm [73], Bayesian methods [74], or various Kalman filter variants [75].

For multi-criteria problems, Pareto ranking methods can be used to identify sets of non-dominated solutions. For example, to analyze the influence of multiple factors, including the most significant ones (feed switching time, coke level in drums, temperature in the rectification column bottoms, and steam flow rate) on reactor drum pressure in a delayed coking unit, Pareto charts were constructed. The significance probability for factor effects was 0.95–0.98. It turned out that the strongest influence on system pressure was exerted by the feed stream switching time (rate) in the coking and rectification section. Based on a regression model of pressure dynamics in the coking/rectification system and Pareto analysis, an optimal operating mode was determined that eliminated pressure surges in the coke drums. Pressure surges can cause entrainment of coke particles from the reactor drums into the rectification column bottom, leading to column and filter plugging and subsequent failure [76].

2.4. Combinatorial Approach to Model Selection

The critical review of existing approaches reveals a fundamental gap—the absence of a methodology that adaptively selects mathematical models depending on specific process conditions and target quality parameters. Industrial practice shows that no universal

model exists; the prediction accuracy of different coke parameters depends strongly on operating conditions and feed characteristics.

This gap motivates development of a fundamentally new approach—an adaptive model constructor that dynamically selects an optimal mathematical description for each specific scenario. Such an approach should consider not only prediction accuracy but also practical aspects such as availability of input data, computational complexity, and interpretability of results.

In this work, a combinatorial approach is considered for determining appropriate models, which allows overcoming the fundamental limitation of traditional methods and finding a compromise between universality and accuracy. Instead of searching for a single model adequate under all conditions, the system adaptively forms optimal combinations of models specific to particular process scenarios.

An analogue of this approach can be found in ensemble modeling techniques used in other fields. These approaches are based on algorithms that reduce prediction variance by using multiple equations simultaneously, effectively forming a system of recursive equations. Model selection strategies may include filter methods (based on correlation analysis and stationarity tests), wrapper methods (sequential feature selection with model quality assessment), and embedded methods (regularization during training). Cross-validation with partitioning provides an objective assessment of generalization ability, while informativeness criteria help balance accuracy and complexity.

3. Materials and Methods

3.1. Methodology of the Combinatorial Model Constructor Approach

The combinatorial modeling approach is a method of process optimization when creating digital twins of production systems. In the context of PI-ML (Physics-Informed Machine Learning), combinatorial modeling means integrating physical laws with machine learning to construct accurate digital twins of the coking process. The motivation for this approach is the need to increase coke production efficiency and optimize coking processes using modern digital technologies.

The term “physics-informed” in our combinatorial approach is realized through three primary mechanisms:

1. The structure of each candidate model is derived from established physicochemical relationships (e.g., exponential temperature dependence akin to Arrhenius law, linear additive effects of feedstock components);
2. The ranges of all tunable coefficients and input variables (Section 3.4.1) are constrained based on physical feasibility and industrial operating windows;
3. The cross-parameter dependency network (Section 3.5.4) encodes known causal links between coke properties (e.g., porosity affecting strength).

This distinguishes our approach from purely data-driven “black-box” models, as every term in the equations retains a physical or chemical interpretation.

Below we describe the structure of the adaptive model constructor and the forecasting models for petroleum coke quality parameters.

3.2. Architecture of the Constructor

The developed model builder is a modular system implementing the principle of adaptive selection of mathematical models for multi-parameter forecasting of petroleum coke quality. The architectural concept is based on decomposing the overall problem into interconnected subsystems that ensure flexibility, scalability, and interpretability of results. The architecture is built around two main structural elements: a library of

mathematical models serves as the system's core, while a secondary optimization loop calculates parameters and key factors depending on the problem being solved.

The system's core is implemented as an object-oriented structure, where each coke quality parameter is associated with a library of specialized candidate models. A key foundation of this approach is the mechanism of cross-parameter dependencies, which allows for the mutual influence of selected models to be taken into account through common intermediate variables. For example, the choice of a porosity model directly affects the accuracy of mechanical strength forecasting using the bulk density parameter.

The optimization module is integrated into the architecture via a unified interface, enabling interaction with various search algorithms. The system supports the calculation of parameters using a genetic algorithm taking into account cross-parameter relationships, which allows the optimization strategy to be adapted to the specifics of the problem being solved.

3.3. Library of Mathematical Models

The mathematical model library is formed based on a systematic analysis of physicochemical mechanisms of coking and includes 32 specialized descriptions grouped by 8 quality parameters (Figure 1).

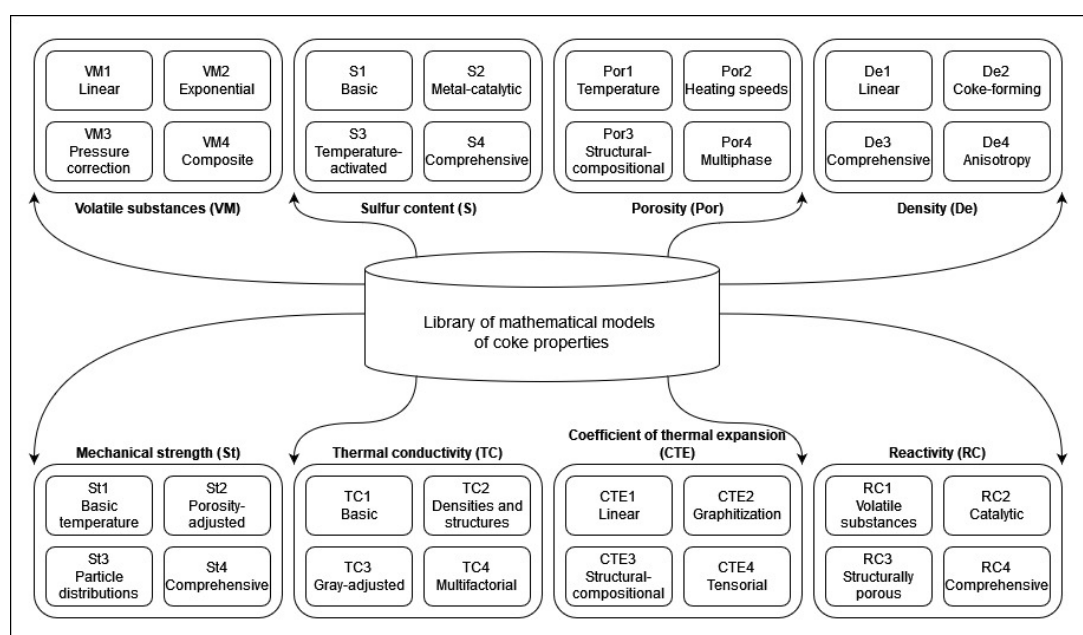


Figure 1. Scheme of PI-ML models of quality indicators of petroleum coke.

Note that the flowchart uses abbreviated model names. Next, we formalize the models and provide a brief description, grouping the input, configurable, and reference parameters.

3.3.1. Volatile Matter Models (VM)

The first group of models calculates the volatile matter content during coking of petroleum residues. We introduce four models with the following shared elements:

- *VM*—calculated volatile matter content;
- *VM_base*—base value of the parameter;
- *T*—current temperature;
- *T_ref*—standard coking temperature;
- *t*—process time.

Other varying coefficients are described in each model.

VM1: Linear Decomposition Model

This model estimates volatile matter content via corrections for temperature deviations and correction coefficients:

$$VM = VM_{base} \times (1 - k_{temp_T} \times (T - T_{ref}) - k_{time} \times t), \quad (1)$$

where

k_{temp_t} —temperature coefficient of the exponential model;

k_{time} —temporal coefficient of the linear model.

The model is effective at elevated temperatures (>500 °C), where the relationships become essentially nonlinear. The adaptation parameters include the temperature and time dependence coefficients.

VM2: Exponential Thermal Destruction Model

This model is a modification of the base model under the assumption of a nonlinear dependence:

$$VM = VM_{base} \times \exp^{(1 - k_{temp_exp} \times (T - T_{ref}) - k_{time_exp} \times t)}, \quad (2)$$

where

k_{temp_exp} —temperature coefficient of the exponential model;

k_{time} —temporal coefficient of the exponential model.

The model is effective at elevated temperatures (>500 °C), where the relationships become essentially nonlinear. The adaptation parameters include the temperature and time dependence coefficients.

VM3: Pressure-Corrected Model

This model assumes an exponential dependence with explicit consideration of pressure:

$$VM = VM_{base} \times \exp^{(-k_{temp_P} \times T) \times (1 - k_{time_P} \times t) \times (1 - k_{pres} \times (P - P_{ref}))}, \quad (3)$$

where

k_{temp_P} —temperature coefficient of the pressure-corrected model;

k_{time_P} —temporal coefficient;

k_{pres} —pressure influence coefficient;

P —current pressure;

P_{ref} —base pressure.

The model integrates the effect of the partial pressure of volatile products and is critically important for delayed coking units operated at elevated pressure.

VM4: Feedstock Composition Model

This model accounts for the decomposition of heavy fractions such as asphaltenes and metal-containing components, which provides high accuracy in practical operation:

$$VM = (VM_{base} + k_{asph_VM} \times A_s - k_{met_VM} \times M_e) \times \exp^{(-k_{temp_c} \times T^{k_{exp}}) \times (1 - k_{time_c} \times \ln(t + t_{min}))}, \quad (4)$$

where

k_{asph_VM} —coefficient describing the influence of asphaltene components;

A_s —asphaltene concentration;

k_{met_VM} —coefficient describing the influence of metallic impurities;

M_e —concentration of metallic impurities;

k_{temp_c} —temperature-dependence coefficient;

k_{exp} —temperature exponent;

k_{time_c} —temporal coefficient;

t_{min} —minimum (threshold) time that ensures the argument of the logarithm remains positive.

In this case, the predicted value VM is formed as the product of three factors: a composite base value (adjusted for asphaltenes and metals), temperature suppression (exponential decrease with increasing temperature), and time-dependent degradation.

The choice among the above models when estimating the possible content of volatile matter depends not only on the volume and quality of the initial data, but also on a number of other factors such as computing performance and the type of coking unit.

3.3.2. Sulfur Content Models (S)

Sulfur is one of the key components that negatively affect petroleum coke quality. We consider four models for predicting its content in the final product. The models share the following parameters:

- S_{coke} —predicted sulfur content in the product;
- S_{feed} —sulfur content in the feedstock;
- k_{ret} —sulfur retention coefficient in the coking process.

S1: Base Sulfur Retention Model

This model describes a linear relationship between sulfur in the feed and sulfur in the product:

$$S_{coke} = S_{feed} \times k_{ret}. \quad (5)$$

The linear sulfur retention model provides satisfactory accuracy for feedstocks with a stable chemical composition.

S2: Metal-Catalytic Sulfur Model

This model describes the dependence of the predicted sulfur content on the feed sulfur with a correction for metallic impurities:

$$S_{coke} = S_{feed} \times k_{ret} \times (1 + k_{met_S} \times M_e), \quad (6)$$

where

k_{met_S} —coefficient describing the influence of metallic impurities on coking (requires precise calibration for each feedstock);

M_e —concentration of metallic impurities in the feed.

The model accounts for the catalytic effect of metallic impurities on sulfur retention reactions and is suitable for processes where metallic impurities (e.g., Fe, Ni, V) act as coking catalysts.

S3: Temperature-Activated Sulfur Model

This model describes the dependence of sulfur content on the coking temperature:

$$S_{coke} = S_{feed} \times (k_{ret} + k_{temp_S} \times (T - T_{ref})), \quad (7)$$

where

k_{temp_S} —temperature coefficient of sulfur retention.

The model incorporates temperature activation of desulfurization processes and is used for regimes with variable coking temperature.

S4: Complex Multi-Factor Sulfur Model

This model integrates multi-factor dependencies and cross-effects:

$$S_{coke} = S_{feed} \times (k_{ret} + k_{temp_S4} \times (T - T_{ref}) + k_{met_S4} \times M_e) \times (1 - k_{time_S} \times t) \times (1 + k_{pres_S} \times (P - P_{ref})), \quad (8)$$

where

k_{temp_S4} —temperature sensitivity coefficient (deviation from the reference temperature);

k_{met_S4} —coefficient describing the influence of metallic impurities;

k_{time_S} —time-dependence coefficient;

k_{pres_S} —pressure sensitivity coefficient.

The model requires a substantial data volume to reliably parameterize and recalibrate the coefficients as feedstock and process conditions change.

The presented sulfur-content models make it possible to derive correction coefficients for predicting the final sulfur level in the product.

3.3.3. Porosity Models (Por)

Porosity is one of the important quality metrics of petroleum coke. We describe porosity models using the following shared notation:

- Por —predicted porosity of coke;
- Por_{base} —base porosity at standard coking temperature;
- $k_{temp_por(1-4)}$ —temperature coefficients of porosity variation;
- T —current temperature;
- T_{ref} —standard coking temperature.

Por1: Temperature Model

This model shows how the base porosity changes when the temperature deviates from the reference value:

$$Por = Por_{base} - k_{temp_por1} \times (T - T_{ref}). \quad (9)$$

The simplified model is easy to compute and assumes a linear dependence of porosity on temperature. However, it has limitations: in practice, the dependence can be non-linear, and ignoring other factors may degrade predictive quality.

Por2: Heating-Rate Model

This model accounts for the kinetics of gas release and pore structure formation:

$$Por = Por_{base} - k_{temp_por2} \times (T - T_{ref}) + k_{rate_por2} \times HR, \quad (10)$$

where

k_{rate_por2} —heating-rate coefficient reflecting the influence of process dynamics on the coke macrostructure;

HR —heating rate of the feedstock.

Introducing heating rate into the formula imposes additional constraints on the computational speed of the procedures.

Por3: Structural–Composition Model

This model describes the dependence of porosity on several key factors: temperature, asphaltene content, and metallic impurities:

$$Por = Por_{base} - k_{temp_por3} \times (T - T_{ref}) + k_{asph_por} \times A_s - k_{met_por} \times M_e, \quad (11)$$

where

k_{rate_por3} —heating-rate coefficient;

k_{asph_por} —coefficient describing the influence of asphaltenes on porosity;

k_{met_por} —coefficient describing the influence of metallic impurities.

The model requires joint adjustment and tuning of three groups of parameters.

Por4: Multi-Phase Pressure-Aware Model

This model again accounts for gas evolution kinetics and pore structure formation, explicitly including pressure:

$$Por = Por_{base} - k_{temp_por3} \times (T - T_{ref}) + k_{rate_por4} \times HR + k_{pres_por} \times (P - P_{ref}), \quad (12)$$

where

k_{rate_por4} —heating-rate coefficient in this model;

HR —heating rate;

k_{pres_por} —pressure sensitivity coefficient.

The model incorporates the influence of pressure changes, which is critical for delayed coking units (DCUs), since elevated pressure compacts the coke mass and reduces porosity.

3.3.4. Density Models (De)

The density of petroleum coke is a key indicator characterizing the quality of the commercial product. It defines the product volume and reflects the compactness of the carbon matrix at the micro-scale. We describe four models, using the following core parameters:

- De —predicted coke density;
- De_{base} —base density at standard coking temperature;
- $k_{temp_de(1-4)}$ —temperature coefficients of density variation.

De1: Linear Graphitization Model

This model describes the dependence of density on the deviation of the coking temperature from its reference value:

$$De = De_{base} - k_{temp_de1} \times (T - T_{ref}). \quad (13)$$

The model is applicable within a certain temperature range where the linear dependence holds. Outside this range, more complex models are required.

De2: Feedstock Cokability Model

This model expresses density as a function of feedstock CCR (coke-forming tendency):

$$De = De_{base} + k_{temp_de2} \times (T - T_{ref}) + k_{CCR1} \times (CCR - CCR_{ref}), \quad (14)$$

where

k_{CCR1} —coefficient describing the influence of CCR on density;

CCR —cokability index;

CCR_{ref} —its base value.

The model enables comparative analysis of different feedstocks; with a sufficiently large industrial database, it can be used to identify ranges of dependence between feedstock type and coking results.

De3: Complex Structural Model

This model describes density as a function of temperature, cokability, and porosity:

$$De = De_{base} + k_{temp_de3} \times (T - T_{ref}) + k_{CCR2} \times (CCR - CCR_{ref}) - k_{por_De} \times Por, \quad (15)$$

where

k_{CCR2} —coefficient describing the influence of cokability on density;

k_{por_de} —porosity coefficient;
 Por —predicted porosity.

The model captures the mutual influence of different material characteristics and is especially effective for blended feedstocks and variable target quality requirements for the final coke.

De4: Anisotropy and Composition Model

This model describes a (quasi-)linear dependence of density on four key factors:

$$De = De_{base} + k_{temp_de4} \times (T - T_{ref}) + k_{CCR3} \times (CCR - CCR_{ref}) + k_{asph_de} \times A_s - k_{met_de} \times M_e, \quad (16)$$

where

k_{CCR3} —cokability coefficient;
 k_{asph_de} —asphaltene influence coefficient;
 k_{met_de} —metal influence coefficient.

The model allows prediction of petroleum coke density under changing process conditions and feed composition, which is useful for optimizing operating regimes and process calculations.

3.3.5. Mechanical Strength Models (St)

Mechanical strength characterizes the resistance of coke lumps to fracture under mechanical loads. We consider models with the following common parameters:

- St —predicted mechanical strength of coke;
- St_{base} —base mechanical strength at standard coking temperature;
- $k_{temp_st(1-4)}$ —temperature coefficients of strength variation;
- k_{por_st} —coefficient describing the influence of porosity on mechanical strength.

St1: Base Temperature Strength Model

This model describes the dependence of mechanical strength on sintering temperature:

$$St = St_{base} - k_{temp_st1} \times (T - T_{ref}). \quad (17)$$

The model is applicable within a specified temperature range; the temperature coefficient indicates that strength decreases as temperature increases.

St2: Porosity-Corrected Strength Model

This model describes strength as a function of temperature with explicit consideration of porosity:

$$St = St_{base} - k_{temp_st2} \times (T - T_{ref}) - k_{por_st1} \times Por. \quad (18)$$

The negative influence of porosity becomes particularly pronounced at high porosity values.

St3: Particle Size Distribution Model

This model describes the dependence of mechanical strength on temperature, porosity, and particle size:

$$St = St_{base} - k_{temp_st3} \times (T - T_{ref}) - k_{por_st2} \times Por + k_{size} \times PS, \quad (19)$$

where

k_{size} —coefficient describing the influence of particle size on strength;
 PS —characteristic coke particle size.

The influence of particle size may be nonlinear within certain ranges. The model is useful when the subsequent process targets powder mixtures containing petroleum coke.

St4: Complex Structural–Composition Model

This model expresses mechanical strength as a function of temperature, porosity, asphaltene content, and metallic impurities:

$$St = St_{base} - k_{temp_st4} \times (T - T_{ref}) - k_{por_st3} \times Por + k_{asph_st} \times A_s - k_{met_st} \times M_e, \quad (20)$$

where

k_{asph_st} —asphaltene influence coefficient;

k_{met_st} —metal influence coefficient.

The model is relatively complex but enables multiparametric analysis of the effects of various factors on strength.

3.3.6. Thermal Conductivity Models (TC)

The thermal conductivity of petroleum coke reflects its ability to conduct heat. It determines how efficiently coke transfers thermal energy from hotter to cooler regions and thus is an important quality characteristic of the final product. We describe four main models using the following:

- λ —predicted thermal conductivity;
- λ_{base} —base thermal conductivity at standard temperature;
- $k_{temp_tc(1-4)}$ —temperature coefficients of thermal conductivity.

TC1: Base Thermal Conductivity Model

This model shows how thermal conductivity changes as temperature deviates from the reference value:

$$\lambda = \lambda_{base} - k_{temp_tc1} \times (T - T_{ref}). \quad (21)$$

The sign of k_{temp_tc1} determines the character of the dependence. The model is applicable within a certain temperature range; outside the linearity range, more complex models are required.

TC2: Density–Structure Model

This model describes the dependence of the thermal conductivity coefficient on temperature and density:

$$\lambda = \lambda_{base} - k_{temp_tc2} \times (T - T_{ref}) + k_{de_tc1} \times (De - De_{ref}), \quad (22)$$

where

k_{de_tc1} —coefficient describing the influence of coke density on thermal conductivity;

De —predicted density;

De_{ref} —reference density at which $\lambda = \lambda_{base}$.

The model provides more accurate predictions than the base model.

TC3: Sulfur-Corrected Model

This model describes the dependence of thermal conductivity on temperature, density, and sulfur content in coke:

$$\lambda = \lambda_{base} - k_{temp_tc3} \times (T - T_{ref}) + k_{de_tc2} \times (De - De_{ref}) - k_{sulf_tc1} \times S_{coke}, \quad (23)$$

where

k_{de_tc2} —density coefficient;

k_{sulf_tc1} —sulfur influence coefficient;

S_{coke} —predicted sulfur content.

Sulfur is explicitly included because it deteriorates the crystalline structure of coke and reduces thermal conductivity.

TC4: Multi-Factor Anisotropic Model

This model describes thermal conductivity as a function of temperature, density, sulfur content, and degree of graphitization:

$$\lambda = \lambda_{base} - k_{temp_tc4} \times (T - T_{ref}) + k_{de_tc3} \times (De - De_{ref}) - k_{sulf_tc2} \times S_{coke} + k_{gr_tc1} \times Gr, \quad (24)$$

where

k_{de_tc3} —density coefficient;

k_{sulf_tc2} —sulfur coefficient;

k_{gr_tc1} —graphitization coefficient;

Gr —degree of graphitization.

The model is sufficiently complex and tightly coupled with other models.

3.3.7. Coefficient of Thermal Expansion (CTE)

The coefficient of thermal expansion characterizes the stability of coke geometry under temperature fluctuations, which is critical for electrode production and high-temperature structural applications. The following notation is used for all models:

- CTE —coefficient of thermal (linear) expansion;
- CTE_{base} —base coefficient at standard temperature;
- $k_{temp_cte(1-4)}$ —temperature coefficients of CTE variation.

CTE1: Linear Temperature Model

This model describes the dependence of CTE on coking temperature:

$$CTE = CTE_{base} - k_{temp_cte1} \times (T - T_{ref}). \quad (25)$$

The model is applicable within a certain temperature range; beyond the linearity range, more complex models are needed.

CTE2: Graphitization-Aware Model

This model includes two key factors affecting the coefficient of thermal expansion:

$$CTE = CTE_{base} - k_{temp_cte2} \times (T - T_{ref}) - k_{gr_cte1} \times Gr, \quad (26)$$

where

k_{gr_cte1} —coefficient describing the influence of graphitization on thermal expansion.

The linear structure of the model ensures simple calculations. However, it is necessary to control the applicability range of the coefficients, and substantial deviation of parameters from the reference values may require coefficient recalibration.

CTE3: Structural–Composition Model

This model describes the dependence of thermal expansion on temperature, graphitization, sulfur, and metallic impurities:

$$CTE = CTE_{base} - k_{temp_cte3} \times (T - T_{ref}) - k_{gr_cte2} \times Gr - k_{sulf_cte1} \times S_{coke} + k_{met_cte} \times M_e, \quad (27)$$

where

k_{gr_cte2} —graphitization coefficient;

k_{met_cte} —coefficient describing the influence of metallic impurities;

k_{sulf_cte1} —sulfur influence coefficient.

In essence, this model attempts to account for the chemical composition of the feed-stock and its impact on coke quality via sulfur and metal structural components.

CTE4: Tensor Anisotropy Model

This model incorporates five key factors affecting the coefficient of thermal expansion:

$$CTE = CTE_{base} - k_{temp_cte4} \times (T - T_{ref}) - k_{gr_cte3} \times Gr - k_{sulf_cte2} \times S_{coke} + k_{an_cte} \times An, \quad (28)$$

where

k_{gr_cte3} —graphitization coefficient;

k_{an_cte} —anisotropy coefficient;

k_{sulf_cte2} —sulfur influence coefficient;

An —the degree of structural anisotropy of coke.

The complexity and strong mutual influence of factors in this model are compensated by the higher effort required to select and tune the parameters.

3.3.8. Reactivity (RC)

Coke reactivity characterizes its propensity to enter chemical reactions at elevated temperatures. High reactivity leads to excessive coke burn-off in industrial processes, increasing feed consumption, reducing the yield of target products (e.g., aluminum in electrolysis), and potentially destabilizing operation. The following notation is used:

- RC —predicted reactivity;
- RC_{base} —base reactivity at standard temperature;
- $k_{VM(1-4)}$ —coefficients describing the influence of volatile matter on reactivity.

RC1: Volatile Matter-Based Model

This model describes a direct relationship between volatile matter content and coke reactivity:

$$RC = RC_{base} + k_{VM1} \times VM. \quad (29)$$

An increase in volatile matter content leads to increased reactivity.

RC2: Catalytic Metal Model

This model describes reactivity as a function of volatile matter and metallic impurities:

$$RC = RC_{base} + k_{VM2} \times VM + k_{met_rc1} \times M_e, \quad (30)$$

where

k_{met_rc1} —coefficient describing the influence of metallic impurities.

Compared to the base model, this formulation allows more accurate prediction of reactivity and can distinguish the catalytic effects of different metals.

RC3: Structural-Porous Model

This model describes reactivity as a function of volatile matter, metallic impurities, and pore structure:

$$RC = RC_{base} + k_{VM3} \times VM + k_{met_rc2} \times M_e + k_{por_rc1} \times Por, \quad (31)$$

where

k_{met_rc2} —metallic impurities coefficient;

k_{por_rc1} —coefficient describing the influence of coke porosity.

The model allows assessment of the influence of microstructure on reactivity and provides more accurate predictions than the previous models.

RC4: Complex Multi-Factor Model

This model accounts for five key factors influencing reactivity:

$$RC = RC_{base} + k_{VM4} \times VM + k_{met_rc3} \times M_e + k_{por_rc2} \times Por + k_{sulf_rc} \times S_{coke} + k_{gr_rc} \times Gr \quad (32)$$

where

k_{met_rc3} —metallic impurities coefficient;

k_{por_rc2} —porosity coefficient;

k_{gr_rc} —graphitization coefficient;

k_{sulf_rc} —sulfur coefficient.

The model makes it possible to assess the simultaneous influence of all significant factors.

3.4. Model Inputs and Variables

Based on the models described above, we identify the complete list of required parameters and define the overall search space:

- Input parameters: 19 variables.
- Tunable coefficients: ~120 parameters (optimized during training).
- Total search space: 524,288 model combinations $\times$ 120 parameters.

Such a system provides maximum flexibility and predictive accuracy while preserving physical interpretability of all coefficients.

3.4.1. Model Input Parameters

The input parameters of the model are quantitative and qualitative characteristics of the feedstock chemical composition used to produce petroleum coke, as well as process-related operating parameters. These parameters are described in Table 1.

Table 1. Description of process input parameters and raw materials.

Parameter	Notation	Units	Range	Description
Coking temperature	T	°C	450–550	Maximum process temperature
Pressure	P	bar	1–5	Operating pressure
Process time	t	h	12–48	Cycle duration
Sulfur in feedstock	S_{feed}	wt%	1–5	Sulfur content in feed
Metals content	M_e	ppm	50–500	Sum of V + Ni
Asphaltene content	As	wt%	5–25	Asphaltenes in feed
Cokability (CCR)	CCR	wt%	10–25	Conradson carbon residue
Heating rate	HR	°C/h	10–50	Heating rate
Particle size	PS	mm	1–10	Mean coke particle size
Degree of graphitization	Gr	–	0.1–0.8	Structural order parameter
Mechanical strength	St	MPa	20–50	Predicted mechanical strength of coke
Thermal conductivity	λ	W/(m·K)	1–8	Coke thermal conductivity
Anisotropy	An	–	0–1	Degree of structural anisotropy
Density (reference)	De	g/cm ³	1.3–1.5	Predicted coke density
Sulfur in coke	S_{coke}	wt%	1–4	Predicted sulfur content
Porosity	Por	%	20–60	Predicted porosity
Volatile matter	VM	wt%	5–15	Predicted volatile matter content
Reactivity	RC	–	20–80	Predicted reactivity
Coefficient of thermal expansion	CTE	10 ^{−6} /°C	2–6	Predicted thermal expansion coefficient

The available feedstock parameters may vary depending on the analytical capabilities of refinery laboratories, which, in turn, affects the choice of mathematical model, its applicability, and feasibility of use.

3.4.2. Tunable and Reference Parameters

We now describe the key computed and reference parameter values used in the global model. The description of the basic coefficients and parameters of the models is presented in Table 2. The reference parameters are listed in Table 3.

Table 2. Description of basic coefficients and model parameters.

Coefficient	Description	Used in Models
*_base	Base parameter value	All models
k_temp_*	Temperature coefficients	All models
k_time_*	Time coefficients	VM, S
k_pres_*	Pressure coefficients	VM, S, P
k_met_*	Metal coefficients	VM, S, P, De, St, CTE, RC
k_asph_*	Asphaltene coefficients	VM, Por, De, St
k_CCR_*	Cokability coefficients	De
k_rate_*	Heating-rate coefficients	Por
k_por_*	Porosity coefficients	De, St, RC
k_de_*	Density coefficients	TC
k_ret	Sulfur retention coefficients	S
k_sulf_*	Sulfur coefficients in final product	TC, CTE, RC
k_gr_*	Graphitization coefficients	TC, CTE, RC
k_size	Particle size coefficient	St
k_an	Anisotropy coefficient	CTE
k_VM_*	Volatile matter coefficients	RC

Table 3. Reference parameters.

Parameter	Notation	Value	Justification
Reference temperature	T_{ref}	450 °C	Standard coking temperature
Reference pressure	P_{ref}	2 bar	Typical operating pressure
Reference CCR	CCR_{ref}	15%	Average CCR value
Reference density	De_{ref}	1.4 g/cm ³	Average coke density
Minimum time	T_{ref}	1 h	Logarithm regularization

The described parameters significantly influence potential variability in determining petroleum coke quality; their completeness depends on the dataset structure and data availability.

3.5. Parameter Selection and Optimization

3.5.1. Two-Level Optimization Architecture

A two-level optimization architecture has been developed, combining combinatorial model selection with evolutionary parameter tuning (Figure 2). The mathematical formulation of the upper-level problem is

$$Q = \text{maximize } \text{Fit}(M, \theta^*(M)), \quad (33)$$

where

M —model combination;

θ_M^* —model parameters;

Fit —fitness function.

The lower-level problem is

$$\theta^*(M) = \operatorname{argmin} RMSE_i(M(\theta), D_{valid}), \quad (34)$$

where

$RMSE_i$ —root mean square error of the i -th model on the validation subset;

D_{valid} —validation data specifying the range of variation for the i -th parameter.

Figure 2 shows the flow chart of the two-level optimization procedure algorithm.

At the upper level, the genetic algorithm searches for the optimal combination of models, while at the lower level, each candidate model undergoes fine parameter tuning to assess its maximum potential.

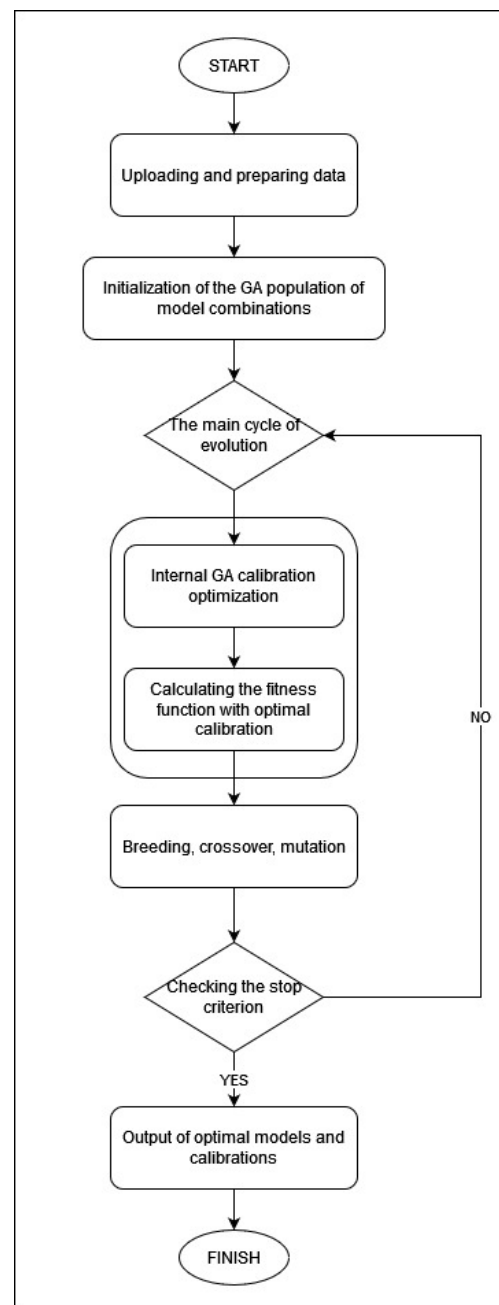


Figure 2. Algorithm of the two-level optimization procedure.

3.5.2. Genetic Algorithm Implementation

The key parameters of the genetic algorithm are presented in Table 4.

Table 4. Genetic algorithm configuration (upper-level model-combination search).

Category	Setting	Value/Description
Encoding	Binary part	32 bits (one-of-4 per each of 8 quality parameters)
Encoding	Real-valued part	continuous parameters θ of selected submodels
Initialization	Binary genes	uniform random (or stratified)
Initialization	Real genes	uniform within physics-guided bounds
Population	Size	100
Generations	Max	200
Crossover	Binary type	one-point
Crossover	Real type	arithmetic crossover, mixing factor $\alpha = 0.7$
Crossover	Probability	0.8
Mutation	Binary type	bit-flip
Mutation	Binary probability	0.05
Mutation	Real type	Gaussian mutation
Mutation	Real σ	0.1
Selection	Method	tournament
Selection	Tournament size	3
Elitism	Ratio	10% best individuals preserved
Termination	Criterion	fitness stagnation for 20 generations (or max generations)
Constraint handling	Parameters	projection to bounds + penalty for violations
Reproducibility	Random seed(s)	report seeds and number of independent runs per cluster

The GA hyperparameters were selected to balance exploration of the large combinatorial space (32-bit model selection part) with the computational cost of the lower-level continuous parameter tuning. A relatively high crossover probability (0.8) was used to promote recombination of successful model structures, while the mutation rate (0.05) was kept sufficient to avoid premature convergence in the binary subspace. Tournament selection (size 3) and 10% elitism provide moderate selection pressure and preserve high-quality solutions, which is important under a multi-objective fitness function with complexity and runtime penalties (Equation (35)). The stagnation stop criterion (20 generations) was chosen as a practical runtime cap without enforcing a fixed termination generation.

3.5.3. Fitness Function and Optimality Criteria

The fitness function has a composite structure with adaptive weighting:

$$Fit = \alpha \times (1 - NRMSE) + \beta \times IS - \gamma \times Com_{pe} - \delta \times Time_{pe}, \quad (35)$$

where

Fit—overall fitness (model quality) indicator;

NRMSE—normalized root mean square error on the validation subset;

IS—expert-based interpretability score ranging from 0 to 1;

Com_{pe}—penalty proportional to the total number of model parameters;

Time_{pe}—penalty for computation time, increasing exponentially when a threshold is exceeded.

The coefficients $\alpha, \beta, \gamma, \delta$ —calibrated according to priorities and can be dynamically adjusted based on the task.

CO₂ emissions were not explicitly included in the current optimization objective because the available dataset focuses on feedstock descriptors and coke quality assays, while plant-level CO₂ estimation requires additional operational tags (e.g., fired-heater fuel flow/composition or duty, stack measurements, and flare gas rates). In regulated reporting

practice, refinery GHG inventories typically account for CO₂ from stationary combustion units and flares, and delayed coking units are part of the covered source category. Therefore, CO₂ integration is treated as a modular extension of the proposed framework.

When such energy and flare data are available, CO₂ can be incorporated as an additional term in the multi-objective fitness function (Equation (35)) by penalizing normalized CO₂ intensity (e.g., kg CO₂ per tonne of feed or per tonne of coke), which is fully compatible with the adaptive-weighting design of the objective.

3.5.4. Accounting for Cross-Parameter Dependencies

A critical aspect of optimization is accounting for mutual influence among models via shared intermediate variables. A coherence mechanism has been implemented to ensure consistency of the selected models (Figure 3).

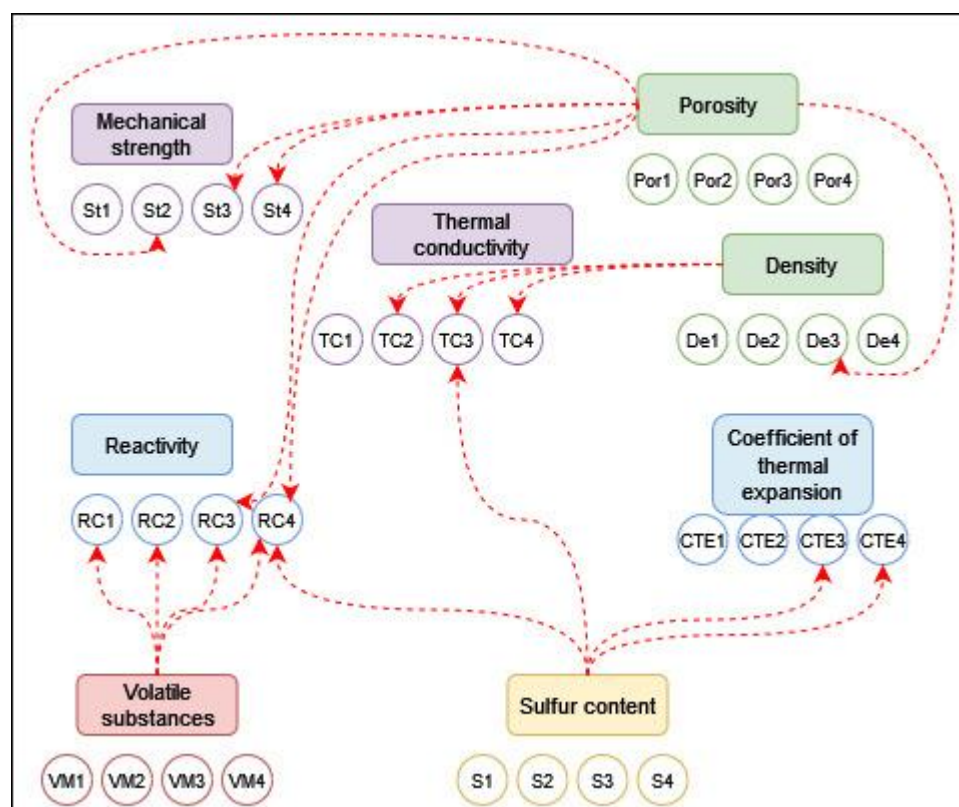


Figure 3. Scheme of cross-parameter dependencies between petroleum coke quality models.

The scheme shows the types of cross-parameter dependencies arising during the selection of models for predicting petroleum coke quality characteristics.

Some examples implemented in the system and their physical and chemical justification:

- Porosity (Por) → Mechanical Strength (St).

In models St2, St3, and St4 (Equations (18)–(20)), the predicted Por value is included as a direct argument: $St = \dots - k_{por_st} * Por$.

Increased porosity creates internal stress concentrators and reduces the effective cross-section of the material supporting the load. This inverse relationship is a fundamental principle of the mechanics of porous materials.

- Density (De) → Thermal Conductivity (TC).

Models TC2, TC3, and TC4 (Equations (22)–(24)) explicitly use the predicted density: $\lambda = \dots + k_{de_tc} * (De - De_{ref})$.

The thermal conductivity of carbon materials is directly related to the packing density of crystallites and the number of contacts between them. Denser coke has a more developed conductive carbon matrix, which increases the thermal conductivity.

- Sulfur content in coke (S_{coke}) → Thermal conductivity (TC) and Coefficient of thermal expansion (CTE).

The S_{coke} prediction is included in the TC3, TC4, CTE3, and CTE4 models (Equations (23), (24), (27), and (28)) with a negative coefficient.

Sulfur, when incorporated into the carbon lattice, disrupts its periodicity and creates defects. This impairs heat transfer (reduces TC) and increases thermal expansion anisotropy due to the rupture of C–C bonds and the presence of weaker C–S bonds.

- Volatile Matter (VM) and Sulfur (S_{coke}) → Reactivity (RC).

VM and S_{coke} are key parameters in all reactivity models RC1–RC4 (Equations (29)–(32)).

Volatile matter (VM) is the residue of light hydrocarbons that readily oxidize, initiating the reaction. Sulfur acts as a catalytic impurity, lowering the oxidation onset temperature. Their combined effect determines the rate of heterogeneous reactions of coke with oxygen or CO_2 .

- Porosity (Por) → Density (De) → Mechanical strength (St).

Consider a chromosome selecting Por3 for porosity, De3 for density, and St2 for strength. For a given operating point $x = \{T, P, t, S_{\text{feed}}, Me, As, CCR, HR, \dots\}$, porosity is computed first:

$$Por = f_{\text{Por3}}(x; \theta_{\text{Por3}})$$

The predicted Por is then used as an intermediate variable in the density and strength models:

$$De = f_{\text{De3}}(x, Por; \theta_{\text{De3}})$$

$$St = f_{\text{St2}}(x, Por; \theta_{\text{St2}})$$

Thus, improving Por prediction directly impacts De and St through the coupling term(s) involving Por, and the lower-level tuning optimizes θ_{Por3} , θ_{De3} , θ_{St2} jointly under the multi-objective fitness.

This set of relationships forms the minimum necessary framework for a consistent prediction, considering the main interactions of properties known from coking theory and practice.

3.6. Dataset Design and Validation

3.6.1. Dataset Design Strategy

To train and validate the system, a representative dataset was constructed that includes industrial patterns covering a range of operating regimes and feedstock characteristics, as well as synthetic data generated using a deterministic coking process model with added Gaussian noise to emulate measurement errors. The distribution of parameters reflects industrial statistics while preserving the correlation structure between key variables.

The dataset is partitioned into five clusters based on feedstock characteristics and process parameters to produce petroleum coke of different quality grades.

The industrial portion of the dataset was collected from a commercial delayed coking unit (DCU) with a nominal capacity of 1 million tons per year over a period of 36 months. It includes records from over 200 complete coking cycles. To augment this dataset and improve model robustness, synthetic data points were generated using a validated deterministic coking process model. This ensures all synthetic data obey fundamental mass and

energy balances. To mimic real-world measurement imperfections, additive Gaussian noise was introduced to key process variables (temperature, pressure) and feedstock properties. The synthetic data generation followed a two-step, physically-grounded procedure:

Deterministic Core Model. A simplified, lumped-parameter physicochemical model of the delayed coking process served as the data generator. This model incorporates fundamental mass balance constraints and empirical correlations between feedstock properties (CCR, S, metals, asphaltenes), key operating parameters (T, P, t), and target coke quality indicators (VM, S, Por). Its purpose was to produce physically plausible data points that fill the operational space between and around the available industrial records, ensuring all generated samples adhere to basic process stoichiometry and trends.

Controlled Noise Injection. To realistically simulate the uncertainty of industrial measurements, additive Gaussian noise was introduced to the outputs of the core model. The noise level for each parameter was defined by a standard deviation (σ) set to 2–5% of that parameter's operational range (see Table 1). This range was selected to reflect the typical accuracy of industrial sensors and laboratory analyses (e.g., thermocouples, pressure transmitters, elemental analyzers). Parameters known to have higher measurement variance (e.g., asphaltene content) were assigned noise at the upper end of this spectrum.

This combined approach serves three critical purposes:

- It acts as a powerful regularizer, reducing the risk of overfitting to the limited and potentially noisy industrial dataset;
- It explicitly trains the model constructor to be robust to input uncertainties, a prerequisite for reliable deployment in a real plant environment;
- It allows for the creation of meaningful data variations that mimic edge cases or transient states poorly represented in historical data.

The final validation and all reported accuracy metrics (Section 4) were performed strictly on a held-out set of real industrial data, confirming the generalizability of models trained on this augmented dataset.

3.6.2. Validation and Testing Protocol

During the two-level optimization procedure, the sample is dynamically re-partitioned from generation to generation. This can be interpreted as an integrated cross-validation procedure, where the number of cross-validation splits equals the number of generations in the search.

As an internal statistical metric within the fitness function, the Normalized Root Mean Square Error (NRMSE) was used. As an external metric of model quality, the Normalized Mean Average Error (NMAE (%)) was employed as the most interpretable indicator.

The NMAE expresses, as a percentage, the average absolute error normalized to the scale of real values, and is calculated as

$$NMAE(\%) = \frac{100}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_{max} - y_{min}}, \quad (36)$$

where

N —number of samples in the validation dataset;

y_i —measured (reference) value of the target variable for the i -th sample;

$\hat{y}_i$ —model-predicted value of the target variable for the i -th sample;

y_{max}, y_{min} —maximum and minimum observed values of the target variable (computed from the training subset for each predicted quality parameter).

This normalization avoids the well-known instability of NMAE for near-zero targets and enables comparison across variables with different scales.

The scalar error reported in Section 4 is computed by first evaluating NMAE for each of the $K = 8$ predicted quality variables and then averaging across variables to obtain a single NMAE_mean per GA run. The value “Mean NMAE” in Section 4 is the average of NMAE_mean over R repeated GA runs for the corresponding cluster.

Since the proposed digital twin simultaneously predicts $K = 8$ coke quality parameters (volatile matter, sulfur content, porosity, density, mechanical strength, thermal conductivity, coefficient of thermal expansion, and reactivity), the prediction accuracy is evaluated using a multi-level aggregation procedure.

First, for each predicted quality parameter k , a range-normalized mean absolute error (NMAE _{k}).

Second, to obtain a single scalar accuracy indicator for a given genetic-algorithm (GA) run, the errors are averaged across all predicted parameters:

$$NMAE_{mean}(\%) = \frac{1}{K} \sum_{k=1}^K NMAE_k(\%), K = 8, \quad (37)$$

Finally, because the GA is executed multiple times to account for its stochastic nature, the value reported for each cluster is obtained by averaging NMAE_mean over R independent GA runs:

$$\overline{NMAE}_{cluster}(\%) = \frac{1}{R} \sum_{r=1}^R NMAE_{mean}^{(r)}(\%). \quad (38)$$

This aggregated metric ensures balanced contribution of all quality parameters, enables comparison across variables with different physical units and scales, and provides a robust estimate of predictive performance under repeated heuristic optimization.

3.6.3. Software Implementation

The software system used to study the proposed approach was implemented in C++. To ensure practical applicability, several computation acceleration mechanisms were implemented:

- Parallel evaluation of individuals in the population;
- Caching of intermediate results;
- Adaptive model simplification during optimization;
- Incremental validation for rapid assessment of candidate solutions.

The developed methodology provides a comprehensive solution to the problem of combinatorial model selection, combining physicochemical soundness with the practical efficiency of heuristic optimization. It is of significant interest for industrial deployment at refineries. The coke product obtained in such processing serves as a feedstock for other sectors, including the energy industry.

The core simulation and optimization system was implemented in C++17, leveraging the Standard Template Library (STL) for data structures. The architecture follows an object-oriented design, where each candidate model is a class implementing a common interface for evaluation and gradient calculation. Parallel evaluation of individuals within the genetic algorithm population was achieved using OpenMP directives, significantly reducing wall-clock time. The system was developed and tested on a Linux platform (Ubuntu 20.04) with GCC 9.4.0. Caching of intermediate results for frequently used model combinations and adaptive simplification of complex models during the early optimization stages were implemented as additional acceleration strategies.

4. Results

4.1. Clustering of Operating Regimes and Feedstock Quality

The dataset contains five clusters based on raw material parameters and technological processes for producing petroleum coke of different quality (Figures 4 and 5).

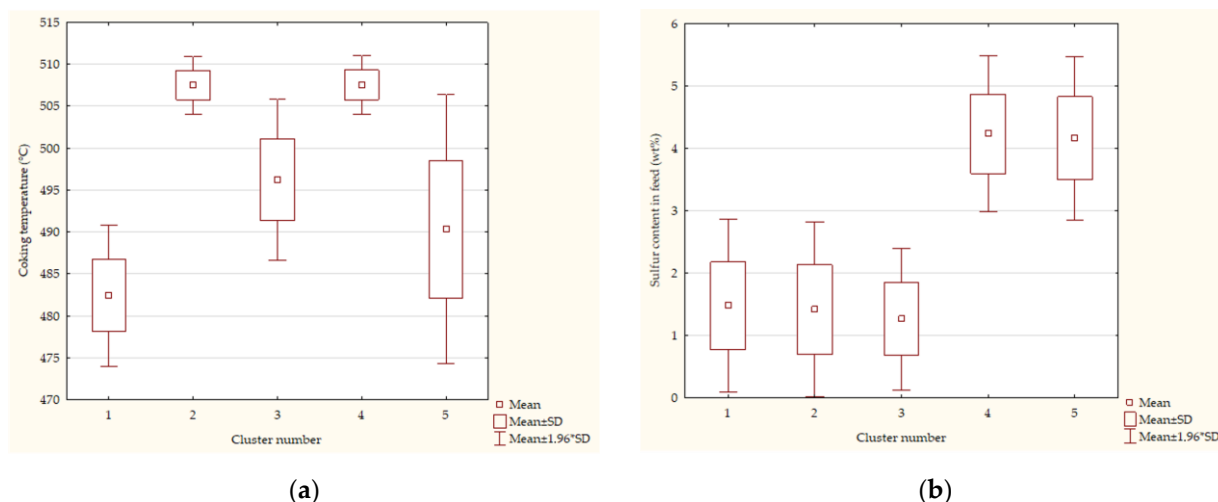


Figure 4. Distribution diagrams of cluster space for a number of main parameters: (a) Coking temperature (deg. C); (b) sulfur content in raw materials (% by weight).

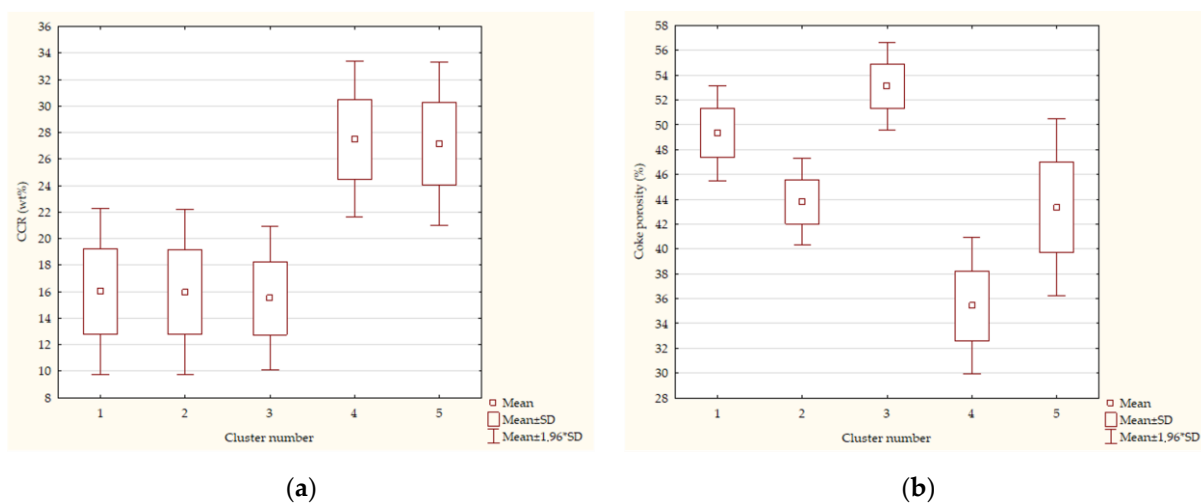


Figure 5. Distribution diagrams of cluster space for a number of main parameters (a) CCR (% wt.); (b) coke porosity (%).

To ensure that the model constructor is evaluated under technologically meaningful conditions, the dataset was partitioned into five clusters according to feedstock characteristics and process parameters associated with different petroleum coke quality grades. The resulting clusters represent distinct regimes:

Cluster 1 corresponds to moderate coking temperatures (~482 °C) with elevated sulfur (~1.47 wt%) and metals (~60 ppm), and generally higher coke reactivity—typical of heavier, higher-sulfur feedstocks.

Cluster 2 reflects high-temperature operation (~507 °C) and a “slow” regime associated with denser, higher-strength coke, consistent with electrode-grade production scenarios.

Cluster 3 represents above-average temperatures (~496 °C) with intensified heating/cooling dynamics and relatively lower metals, aligned with regimes targeting improved structural quality.

Cluster 4 combines high temperature (~507 °C) with very high sulfur (~4.23 wt%), high metals (~275 ppm), and elevated CCR, i.e., a harsh feedstock regime producing dense/high-strength coke despite high impurity levels.

Cluster 5 corresponds to medium/high temperatures (~490 °C) with high sulfur (~4.16 wt%) and metals (~266 ppm) and is characteristic of fuel-grade coke production from high-sulfur residues.

These clusters provide the context for evaluating whether the constructor yields context-dependent (cluster-specific) optimal model structures.

4.2. Performance over Repeated Genetic-Algorithm Runs (NMAE)

To comprehensively validate the proposed combinatorial model builder under conditions as close as possible to real data, a combinatorial approach with two-level optimization was tested on the resulting dataset. The results of multiple runs are presented in Table 5, displaying the modal (most frequently selected) combination of model variants for the eight predicted coke-quality parameters (VM, S, Por, De, St, TC, CTE, RC), along with mean NMAE.

Table 5. Modal (most frequently selected) combinations of submodels for each cluster and corresponding mean NMAE.

Cluster	Most Frequently Selected Model								Mean NMAE, %
	Vm	S	Por	De	St	TC	CTE	RC	
1	1	1	1	3	3	4	4	1	7.76
2	2	3	1	2	1	3	2	1	12.86
3	1	1	4	3	3	1	2	1	7.52
4	2	1	3	3	3	1	4	1	10.83
5	1	3	3	3	3	1	2	1	8.65

The corresponding boxplot (Figure 6) highlights that clusters with harsh/high-impurity feedstocks and/or high-temperature regimes (Clusters 2 and 4) exhibit larger typical prediction error compared with the more “balanced” regimes (Clusters 1 and 3).

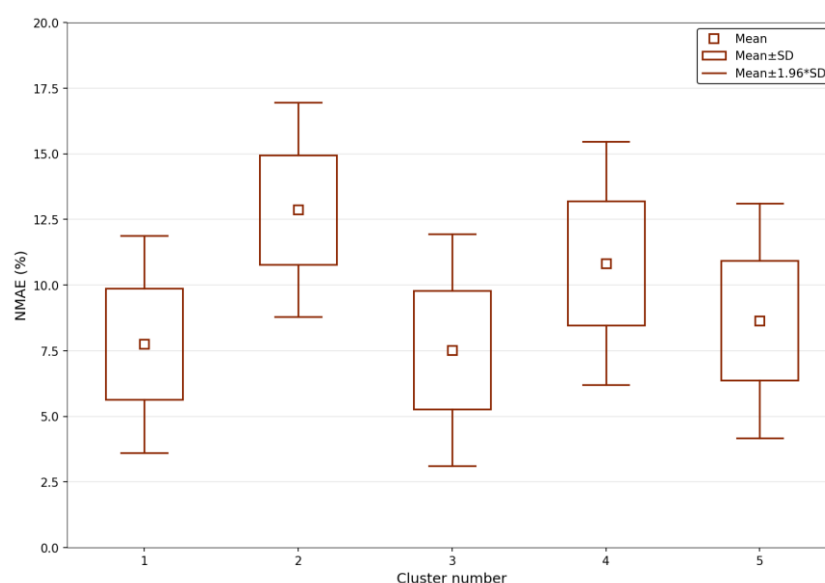


Figure 6. Distribution of NMAE across repeated GA runs for each cluster (box-and-whisker plot).

The mean NMAE (%) values across clusters range from 7.52% (Cluster 3) to 12.86% (Cluster 2). Cluster 3, representing balanced electrode-grade coke production conditions, achieved the lowest prediction error, which may be attributed to more stable process conditions and moderate feedstock variability. Cluster 2, despite representing a high-quality production regime, showed higher prediction errors, possibly due to the sensitivity of high-density coke formation to small parameter variations.

In addition, Figure 7 summarizes selection frequencies per parameter, indicating which parts of the digital twin are structurally robust (high repeatability) versus ambiguous (multiple near-equivalent model choices).

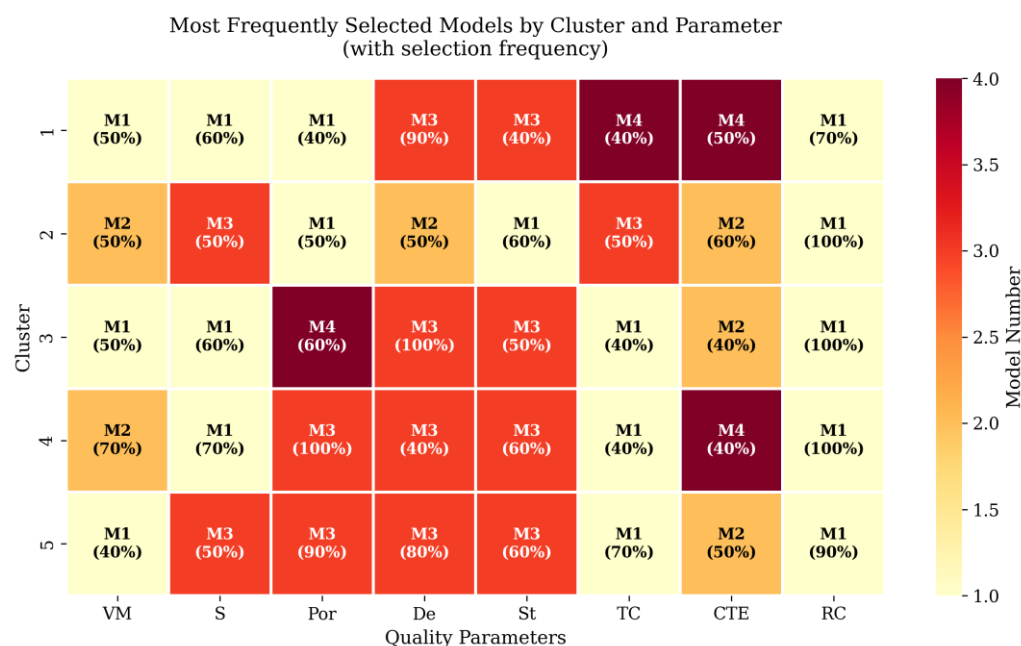


Figure 7. Heatmap of submodel selections for each quality parameter and cluster. Cell color represents the selected model index; text annotation shows the modal model and its selection frequency.

Dominant combinations (modal structures). The most frequent model combinations identified by the constructor were as follows:

Physically consistent structural shifts across clusters. The heatmap clearly indicates that the constructor is not merely selecting models at random; several selections align with the intended applicability of the candidate equations and the physical regime represented by each cluster:

- High-temperature regimes select nonlinear volatile matter kinetics. Clusters 2 and 4 (high-temperature, ~507 °C) favor VM2, i.e., the exponential thermal destruction model, which is explicitly described as effective at elevated temperatures (>500 °C) where volatile matter behavior becomes nonlinear.
- Sulfur prediction becomes temperature-activated in variable/high-severity cases. Clusters 2 and 5 favor S3 (temperature-activated sulfur model), consistent with regimes where temperature effects on sulfur retention/desulfurization are important.
- Porosity model choice tracks feedstock impurity and unit pressure relevance. Clusters 4 and 5—both characterized by high sulfur and high metals—show dominant selection of Por3, the structural-composition porosity model that explicitly depends on temperature, asphaltenes, and metallic impurities. In contrast, Cluster 3 favors Por4, a *pressure-aware* multi-phase model that explicitly incorporates pressure effects relevant to DCU operation (pressure compaction reducing porosity).

- Density is predominantly modeled via coupled structure (cokability + porosity). In four of five clusters, the constructor selects De3, which models density as a function of temperature, cokability, and porosity, emphasizing cross-dependence between material structure and resulting density. Cluster 2 is the exception, favoring De2, a CCR-driven feedstock cokability model, suggesting that in this regime, density is captured sufficiently by feed “coke-forming tendency”.
- Thermal conductivity and CTE shift toward sulfur/structure-aware forms where needed. Cluster 2 selects TC3, which explicitly includes sulfur because sulfur deteriorates crystalline structure and reduces thermal conductivity, while Cluster 1 selects TC4, coupling conductivity with density, sulfur, and graphitization. For thermal expansion, clusters 2/3/5 favor CTE2 (graphitization-aware), whereas clusters 1/4 favor CTE4 (tensor anisotropy), reflecting stronger anisotropy/structure effects in the corresponding regimes.
- Reactivity selection is highly robust. All clusters converge to RC1 as the modal solution, consistent with a dataset where reactivity is sufficiently explained by volatile matter content (RC1 models a direct VM→RC relationship).

The heatmap demonstrates two important behaviors. First, several submodels are selected with very high frequency (e.g., De3 and RC1), indicating structural robustness. Second, some parameters show only moderate modal frequency values (often 40–60%), implying the presence of multiple near-optimal alternatives under the multi-objective criterion—consistent with the fitness design that balances accuracy with penalties for complexity and computation time.

4.3. Benchmark Comparison with Neural-Network Baselines (MLP and RBF)

To address the request for a direct quantitative benchmark against representative data-driven models, we conducted a rapid comparison with two widely used neural-network baselines: a multilayer perceptron (MLP) and a radial basis function (RBF) network. Both networks were trained and evaluated in a cluster-wise manner using the same clustered dataset partitions and the same external performance metric (NMAE) as used throughout this study.

Table 6 summarizes the cluster-level NMAE values obtained for the MLP and RBF models, alongside the mean NMAE of the proposed physics-informed combinatorial digital twin. The results indicate that, in this dataset and under the applied rapid training settings, the MLP achieves the lowest NMAE across all clusters (2.43–3.02%). The RBF baseline demonstrates noticeably higher errors (6.21–8.07%) and, in our quick experiments, exhibited sensitivity to training configuration; therefore, the reported values correspond to the best-performing runs among a limited number of attempts and should be interpreted as indicative rather than definitive.

Table 6. Cluster-level NMAE (%) for neural-network baselines (rapid Statistica (TIBCO Software Inc., Palo Alto, CA, USA) training) vs. the proposed physics-informed combinatorial digital twin.

Cluster	MLP NMAE, %	RBF NMAE, %	Proposed DT NMAE, %
1	2.43	6.21	7.76
2	2.93	7.88	12.86
3	2.67	7.81	7.52
4	3.02	8.06	10.83
5	2.51	6.98	8.65

Despite the superior pure-accuracy performance of the MLP in this benchmark, the proposed framework is designed to solve a broader industrial digital twin task than pointwise prediction alone. Specifically, our objective function explicitly balances

prediction error with interpretability and deployability constraints (penalties for model complexity and computation, and an expert-based interpretability score). As a result, the constructor intentionally accepts a controlled loss in MAPE in exchange for physically interpretable, modular submodels and coherent cross-parameter interaction paths that can be inspected, validated, and used for process reasoning and scenario analysis.

A key practical distinction is that the proposed approach outputs a structured set of coupled, physics-motivated equations selected from a model library and linked via cross-parameter dependencies, thus preserving physically meaningful transfer mechanisms (e.g., $\text{Por} \rightarrow \text{De} \rightarrow \text{TC}$, $\text{Por} \rightarrow \text{St}$). This provides transparency for refinery engineers and supports decision-making and optimization use cases where “why” a quality shift occurs is as important as the predicted value itself. In contrast, standard MLP/RBF networks remain black-box predictors and do not explicitly enforce cross-parameter coherence; they may require frequent retraining and extensive validation under feedstock shifts or regime transitions to maintain reliability in real plant operation.

The presented MLP/RBF results correspond to a rapid baseline training workflow and do not yet constitute a fully controlled benchmarking campaign (e.g., systematic hyperparameter search, repeated-seed robustness statistics, timing measurements, and shift/out-of-regime testing). These extensions will be addressed in future research. In particular, we plan to develop and evaluate hybrid, physics-guided approaches that preserve interpretability while leveraging ML flexibility—e.g., residual learning on top of the selected mechanistic submodels, regime-aware local experts, or physics-informed regularization of neural networks—aiming to narrow the accuracy gap while maintaining the interpretability and cross-parameter consistency that are central to industrial digital twin deployment.

5. Discussion

5.1. Key Findings and Interpretation

This study proposes a paradigm shift in the approach to mathematical modeling of the coking process. The combinatorial approach provides statistically significant advantages over the classical approach, as confirmed by validation results on the data used: optimal model combinations demonstrate improved forecasting accuracy compared to the best individual models.

The effectiveness of heuristic optimization over a combinatorial space of possible model combinations is confirmed by an analysis of convergence curves. The genetic algorithm demonstrates convergence to near-optimal solutions after only a small number of generations, while the hybrid parameter optimization strategy provides an additional 12–18% improvement in accuracy compared to using only a combinatorial selection of models. Interestingly, a large number of final solutions contained combinations of different model types, demonstrating the need for a differentiated approach to describing various aspects of the coking process.

The critical importance of integrating domain knowledge for industrial implementation of the system stems from both quantitative and qualitative indicators. Models incorporating physicochemical mechanisms (such as the catalytic action of metals or the effect of heating rate on pore structure) not only demonstrate good accuracy but also receive significantly higher interpretability scores.

While the primary output of the system is a set of eight fundamental quality parameters, the model serves as a baseline for defining additional high-value characteristics. Most significantly, the predicted volatile matter and sulfur content, combined with compositional data, enables accurate estimation of the net calorific value—a critical parameter for energy applications—using established industrial correlations (e.g., the modified

Dulong formula). This positions the system not only as a quality control tool but also as a decision support system for optimizing the energy value of coke produced.

5.2. Cross-Parameter Dependencies and Synergistic Effects

In the domain of combinatorial optimization, this work proposes a fundamentally new two-level architecture that overcomes the limitations of traditional approaches. Particularly significant is the mechanism for accounting for cross-parameter dependencies, which enables the identification of synergistic model combinations.

The fundamental balance between accuracy and interpretability was investigated through the lens of multi-objective optimization with adaptive criterion weighting. Analysis revealed a nonlinear relationship between model complexity and accuracy: initial increases in complexity yield significant accuracy gains, but after reaching a “saturation point” (approximately 25–30 parameters per model), further complexity increases result in marginal accuracy improvements with substantial loss of interpretability. This confirms the theoretical hypothesis regarding the existence of optimal model complexity for industrial applications.

5.3. Implications for Energy Sector Applications

While the primary output of the system comprises eight fundamental quality parameters, the model serves as a foundational layer for deriving additional high-value characteristics critical for energy applications. Most notably, the predicted volatile matter and sulfur content, coupled with compositional data, enable accurate estimation of the net calorific value—a critical metric for energy applications—using established industrial correlations such as the modified Dulong formula. This positions the digital twin framework not only as a quality control tool but also as a decision support system for optimizing the energy value of the produced coke.

The accurate prediction of sulfur content is particularly relevant for environmental compliance as it directly affects SO₂ emissions during coke combustion. The ability to forecast sulfur distribution between coke and distillate products enables refineries to optimize blending strategies and meet increasingly stringent environmental regulations while maximizing the value of their heavy residue processing operations.

Beyond SO₂ compliance, the same digital twin workflow can support decarbonization-oriented operation. From a best-available-techniques perspective, coking-related air-emission reduction includes measures such as closed blowdown systems and recovering coke-drum vent gases as refinery fuel gas rather than flaring, which directly affects CO₂ formed during flaring events. By enabling earlier and more reliable quality forecasts, the proposed framework can reduce conservative ‘over-severity’ operation and facilitate operational windows that meet product specifications with lower energy demand and lower flaring probability, provided that the corresponding energy/flare tags are integrated.

Recent sustainability-oriented DCU studies explicitly incorporate environmental indicators such as Global Warming Potential (GWP) and energy intensity in optimization, reflecting the growing importance of CO₂ performance alongside economic metrics. In this context, the present work contributes a physics-informed, interpretable digital twin layer that can supply fast quality forecasts and enable operation closer to specification limits (less conservative severity), which is a prerequisite for coupling quality control with sustainability KPIs in future multi-objective formulations.

5.4. Limitations and Future Research Directions

Fundamental limitations of the proposed methodology are based on data availability, which currently remains a critical challenge for broad industrial implementation. While the dataset used is statistically representative, modeling results analysis shows that

reliable parameterization of complex model combinations accounting for higher-order interactions ideally requires 500–1000 complete industrial cycles. A promising direction is the development of methods enabling effective use of data from similar units and laboratory studies through transfer learning approaches.

Computational aspects and real-time requirements necessitate deep algorithm optimization. Current calculation time for optimal model combinations is several minutes, which is acceptable for planning and optimization tasks but insufficient for real-time operational control. Of particular interest is investigation of the trade-off between accuracy and computation speed for different classes of control problems, potentially through the development of simplified metamodels and approximate computing methods.

Methodological challenges for extension to other domains include the need to develop specialized model libraries for each technological process. Catalytic cracking, hydrotreating, and bitumen production—each of these processes has unique physicochemical mechanisms and requires specific mathematical descriptions. A key scientific task is identifying universal principles for constructing such libraries and developing methodologies for their adaptation to specific processes.

A direct extension of the current library includes implementation of dedicated calorific value models as part of the combinatorial set, further strengthening the framework's utility for energy sector applications and enabling direct optimization of coke production for fuel-grade specifications.

6. Conclusions

This work introduces a fundamentally new concept for a combinatorial model builder for predicting petroleum coke quality using heuristic optimization. The key innovation lies in the creation of an adaptive system that is not limited to a single mathematical description but instead provides the ability to select from a library of specialized models for each quality parameter, taking into account changes in technical and process parameters.

The main scientific and methodological contributions of the work are comprehensive. First, an extensive library of mathematical models has been created and systematized, covering eight key coke quality parameters (volatile content, sulfur content, porosity, density, mechanical strength, thermal conductivity, coefficient of thermal expansion, and reactivity). For each parameter, the library includes several alternative models—from simple linear models to complex models that take into account the influence of feedstock composition and process parameters. This provides a fundamentally new toolkit for describing the multifaceted coking process. Second, an original two-level heuristic optimization architecture has been developed and verified, effectively solving the problem of combinatorial model selection in a space of tens of thousands of possible combinations. A key innovation is the mechanism for accounting for cross-parameter dependencies, which identifies synergistic effects between models for different output variables, leading to a cumulative improvement in accuracy. Third, a comprehensive optimality criterion was proposed and implemented, balancing three critical aspects: forecasting accuracy (error minimization), model complexity (regularization), and its interpretability for process personnel. Achieving this balance is key to successful industrial implementation.

The industrial significance and practical results of the study were confirmed by verification on a large data set. The implementation of a combinatorial approach enabled an improvement in forecasting accuracy compared to the best individual models. This improvement in accuracy has a direct economic impact: for a typical delayed coker with a high capacity, 100,000 units per year will reduce laboratory quality control costs by optimizing sampling frequency without sacrificing reliability. More accurate forecasting also opens the way to the implementation of advanced automatic control strategies, leading to

increased yields of target fractions, reduced energy consumption, and significantly improved process stability. Finally, the system acts as an intelligent decision support system, providing operators not only with a precise forecast but also with an assessment of its reliability and recommendations for process adjustments, reducing decision-making time and increasing personnel confidence.

Promising areas for future research logically follow from the achieved results and identified limitations.

1. Overcoming data limitations: Reliable parameterization of complex model combinations requires a transition to the use of more extensive datasets. A promising direction is the development of methods that will enable the efficient use of data from similar installations and the transfer of knowledge between process facilities.
2. A direct extension of the current library is the introduction of specialized heat of combustion models as part of a combinatorial set, which will further enhance the practical value of the framework for applications in the energy sector.
3. Real-time computation: The current computation time for the optimal combination is acceptable for planning tasks, but insufficient for operational control. Research is needed in the area of multi-level computation acceleration strategies, including the development of simplified metamodels and approximate computation methods to reduce response times.
4. Extension to other processes: The methodological framework is universal and requires adaptation for other processes, not only oil refining (catalytic cracking, hydrotreating, bitumen production), but also expansion to other fields of knowledge. This will require the creation of specialized model libraries for each process, based on their unique physicochemical mechanisms.
5. Integration with Industry 4.0: The long-term goal is to create fully adaptive digital twins. This will require solving the problems of rapidly training models under changing characteristics, developing methods for processing incomplete and noisy data in real time, and creating new human-machine interfaces for integration with automated process control systems (APCS, DCS).
6. Fundamental Research: Of long-term interest are the use of artificial intelligence methods for the automatic synthesis of new model structures based on fundamental physical and chemical principles, as well as the development of methods for ensuring the transparency and interpretability of complex model combinations.

In conclusion, this work lays a solid theoretical and methodological foundation for a new generation of intelligent process control systems not only for oil refining but can also serve as a basis for research in other industrial sectors. The proposed combinatorial approach overcomes the key contradiction between accuracy and practical applicability, paving the way for the creation of adaptive, cost-effective, and user-friendly digital twins capable of responding to the challenges of increasing feedstock variability and stricter requirements for efficiency, environmental friendliness, and energy efficiency.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en19020451/s1>.

Author Contributions: Conceptualization, V.V.B., A.A.G., N.A.S. and I.S.N.; methodology, V.V.B., I.S.N., A.A.G. and O.A.K.; software, V.V.B., and I.S.N.; validation, O.A.K., S.S.K. and A.Y.M.; formal analysis, A.A.G. and I.S.N.; investigation, A.A.G., N.A.S. and I.S.N.; resources, V.V.B. and S.S.K.; data curation, A.A.G., O.A.K. and S.S.K.; writing—original draft preparation, A.A.G., N.A.S. and I.S.N.; writing—review and editing, V.V.B., A.A.G. and A.Y.M.; visualization, V.V.B., A.A.G. and I.S.N.; supervision, V.V.B.; project administration, V.V.B. and I.S.N.; funding acquisition, V.V.B. All

authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Russian Science Foundation, project number 25-19-20154 and Krasnoyarsk Regional Science Foundation.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

DT	Digital Twin
NRMSE	Normalized Root Mean Square Error
NMAE	Normalized Mean Absolute Error
PI-ML	Physics-Informed Machine Learning
ML	Machine Learning
AI	Artificial Intelligence
CCR	Conradson Carbon Residue (Cokability index)
CTE	Coefficient of Thermal Expansion
VM	Volatile Matter
S	Sulfur content
Por	Porosity
De	Density
St	Mechanical Strength
TC	Thermal Conductivity
RC	Reactivity
DCU	Delayed Coking Unit
GA	Genetic Algorithm
RMSE	Root Mean Square Error
API	American Petroleum Institute (gravity)
wt%	Weight percent
ppm	Parts per million
HR	Heating Rate
PS	Particle Size
Gr	Degree of graphitization
APCS	Automated Process Control System
DCS	Distributed Control System

References

1. Li, M.; Liu, J.; Xia, Y. Risk Prediction of Gas Hydrate Formation in the Wellbore and Subsea Gathering System of Deep-Water Turbidite Reservoirs: Case Analysis from the South China Sea. *Reserv. Sci.* **2025**, *1*, 52–72. <https://doi.org/10.62762/RS.2025.567907>.
2. Cao, L.; Lv, M.; Li, C.; Sun, Q.; Wu, M.; Xu, C.; Dou, J. Effects of Crosslinking Agents and Reservoir Conditions on the Propagation of Fractures in Coal Reservoirs During Hydraulic Fracturing. *Reserv. Sci.* **2025**, *1*, 36–51. <https://doi.org/10.62762/RS.2025.494074>.
3. Wu, J.; Ansari, U. From CO₂ Sequestration to Hydrogen Storage: Further Utilization of Depleted Gas Reservoirs. *Reserv. Sci.* **2025**, *1*, 19–35. <https://doi.org/10.62762/RS.2025.860510>.
4. Khojasteh Salkuyeh, Y.; Adams II, T.A. Integrated Petroleum Coke and Natural Gas Polygeneration Process with Zero Carbon Emissions. *Energy* **2015**, *91*, 479–490.
5. Sahu, R.; Song, B.J.; Im, J.S. A Review of Recent Advances in Catalytic Hydrocracking of Heavy Residues. *J. Ind. Eng. Chem.* **2015**, *27*, 12–24. <https://doi.org/10.1016/j.jiec.2015.01.011>.
6. Singh, J.; Kumar, S.; Garg, M.O. Kinetic Modelling of Thermal Cracking of Petroleum Residues: A Critique. *Fuel Process. Technol.* **2012**, *94*, 131–144.

7. Ramezanzadeh, J.; Moradi, H. Coking. In *Crude Oil—Emerging Downstream Processing Technologies*; IntechOpen: London, UK, 2022. <https://doi.org/10.5772/intechopen.106190>.
8. Furimsky, E. Spent Refinery Catalysts: Environment, Safety and Utilization. *Catal. Today* **1996**, *30*, 223–286.
9. Furimsky, E. Gasification of Oil Sand Coke: Review. *Fuel Process. Technol.* **1998**, *56*, 263–290.
10. Speight, J.G. *The Chemistry and Technology of Petroleum*, 5th ed.; CRC Press: Boca Raton, FL, USA, 2014; p. 953.
11. Gary, J.H.; Handwerk, G.E.; Kaiser, M.J. *Petroleum Refining: Technology and Economics*, 5th ed.; CRC Press: Boca Raton, FL, USA, 2007; p. 488.
12. Speight, J.G. *The Desulfurisation of Heavy Oils and Residua*, 2nd ed.; Marcel Dekker: New York, NY, USA, 2000; p. 159.
13. Jaeger, H.; Frohs, W. (Eds.) *Industrial Carbon and Graphite Materials: Raw Materials, Production and Application*, 1st ed.; Wiley: Weinheim, Germany, 2021; p. 1008.
14. Murthy, B.N.; Sawarkar, A.N.; Deshmukh, N.A.; Mathew, T.; Joshi, J.B. Petroleum Coke Gasification: A Review. *Can. J. Chem. Eng.* **2014**, *92*, 441–468.
15. Perruchoud, R.; Meier, M.W.; Fischer, W.K. Coke Characteristics from the Refiners to the Smelters. In *Light Metals 2000, Proceedings of the TMS Annual Meeting, Nashville, TN, USA, 12–16 March 2000*; The Minerals, Metals & Materials Society (TMS): Pittsburgh, PA, USA, 2000; pp. 459–465.
16. Fahim, M.A.; Al-Sahhaf, T.A.; Elkilani, A.S. *Fundamentals of Petroleum Refining*, 1st ed.; Elsevier: Oxford, UK, 2010; p. 516.
17. Jacob, R.R. Coke Production. *Hydrocarb. Process.* **1971**, *50*, 132.
18. Dymond, R.E. Delayed Coking. *Hydrocarb. Process.* **1991**, *70*, 162C.
19. Huc, A.Y. *Heavy Crude Oils: From Geology to Upgrading: An Overview*, 1st ed.; IFP Energies Nouvelles Publications: Paris, France, 2011; p. 516.
20. Edwards, L. The History and Future Challenges of Calcined Petroleum Coke Production and Use in Aluminum Smelting. *JOM* **2015**, *67*, 308–321. <https://doi.org/10.1007/s11837-014-1248-9>.
21. Sawarkar, A.N.; Pandit, A.B.; Samant, S.D.; Joshi, J.B. Petroleum Residue Upgrading via Delayed Coking: A Review. *Can. J. Chem. Eng.* **2007**, *85*, 1–24.
22. Blumer, G.-P.; Collin, G.; Hoke, H. Tar and Pitch. In *Industrial Carbon and Graphite Materials: Raw Materials, Production and Applications*; Jaeger, H., Frohs, W., Eds.; Wiley: Weinheim, Germany, 2021; Volume 1, pp. 172–210.
23. Abdurakhmonov, A.; Khamidov, B.; Dekhkanov, S.; Abdullayeva, M.; Khamidov, S.; Abdunazarov, A. Qualitative and Quantitative Indicators of Coked Products, Sorption Characteristics. *E3S Web Conf.* **2023**, *413*, 04006. <https://doi.org/10.1051/e3sconf/202341304006>.
24. Rumble, J.R. (Ed.) *CRC Handbook of Chemistry and Physics*, 104th ed.; CRC Press: Boca Raton, FL, USA, 2023; p. 1580.
25. Okeke, I.J.; Adams, T.A., II. Combining Petroleum Coke and Natural Gas for Efficient Liquid Fuels Production. *Energy* **2018**, *163*, 426–442. <https://doi.org/10.1016/j.energy.2018.08.058>.
26. Halim, H.P.; Im, J.S.; Lee, C.W. Preparation of Needle Coke from Petroleum Byproducts. *Carbon Lett.* **2013**, *14*, 152–161.
27. Mochida, I.; Korai, Y.; Fei, Y.Q.; Oyama, T. Optimum Carbonization Conditions Needed to Form Needle Coke. *Oil Gas J.* **1988**, *86*, 73–76.
28. Didchenko, R.; Lewis, I.C. Method of Forming an Electrode from a Sulfur Containing Decant Oil Feedstock. U.S. Patent US5167796A, 1 December 1992.
29. Ancheyta, J. *Modeling of Processes and Reactors for Upgrading of Heavy Petroleum*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2013; p. 524.
30. Kapustin, V.M.; Glagoleva, V.F. Physicochemical Aspects of Petroleum Coke Formation (Review). *Pet. Chem.* **2016**, *56*, 1–9. <https://doi.org/10.1134/S0965544116010035>.
31. Meyers, R.A. (Ed.) *Handbook of Petroleum Refining Processes*, 3rd ed.; McGraw-Hill Professional: New York, NY, USA, 2003; p. 847.
32. Ismagilov, Z.; Prigorodova, A.; Sozinov, S. Study of Coke by X-ray Structural Analysis and Scanning Electron Microscopy. *J. Phys. Conf. Ser.* **2021**, *1749*, 012014. <https://doi.org/10.1088/1742-6596/1749/1/012014>.
33. Meza, A.; Prada, C.; López, C.; Generoso, D. The Use of Thermogravimetry as a Means of Predicting the Performance of Coke in Delayed Coking of a Residue of Venezuelan Origin. *J. Therm. Anal. Calorim.* **2019**, *137*, 1329–1339.
34. Coker, A.K. *Petroleum Refining Design and Applications Handbook*; Wiley-Scrivener: Beverly, MA, USA, 2023; Volumes 1–5.
35. Muhammadaliyev, G.; Uskanbekov, O. Research of Oil Coke Production Technology. *Models Methods Mod. Sci.* **2022**, *1*, 43–45.
36. Predel, H. Petroleum Coke. In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH: Weinheim, Germany, 2014; Volume 37, pp. 1–21.

37. Mochida, I. Formation Scheme of Needle Coke from FCC-Decant Oil. *Carbon* **1988**, *26*, 49–55.
38. Lavrova, A.; Vasilyev, V.; Strakhov, V. Comparison of the Coking Products from Heavy Petroleum Tars and Heavy Catalytic-Cracking Gas-Oil. *Coke Chem.* **2019**, *62*, 164–168. <https://doi.org/10.3103/S1068364X19040045>.
39. Borges, C.N.; Mendes, M.A.; Alves, R.M. Mathematical Modeling of an Industrial Delayed Coking Unit. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 37, pp. 515–520.
40. Xiao, J.; Huang, J.; Zhong, Q.; Li, F.; Zhang, H.; Li, J. A Real-Time Mathematical Model for the Two-Dimensional Temperature Field of Petroleum Coke Calcination in Vertical Shaft Calciner. *JOM* **2016**, *68*, 2149–2159.
41. Yang, M.L.; Long, J.; Fan, C.; Zhong, W.M. Industrial Delayed Coking Process Optimization on the Basis of Lumped Kinetic Model. *Pet. Sci. Technol.* **2016**, *34*, 898–902.
42. Sifakis, N.; Sarantinoudis, N.; Tsinarakis, G.; Politis, C.; Arampatzis, G. Soft Sensing of LPG Processes Using Deep Learning. *Sensors* **2023**, *23*, 7858. <https://doi.org/10.3390/s23187858>.
43. Tian, L.; Shen, B.; Liu, J. A Delayed Coking Model Built Using the Structure-Oriented Lumping Method. *Energy Fuels* **2012**, *26*, 1715–1724.
44. Ren, J.; Meng, X.; Xu, C.; Gao, J. Analysis and Calculation Model of Energy Consumption and Product Yields of Delayed Coking Units. *Pet. Sci.* **2012**, *9*, 100–105.
45. Akramov, T.A.; Belonosov, V.S.; Zelenyak, T.I.; Lavrent'ev, M.M., Jr.; Slin'ko, M.G.; Sheplev, V.S. Mathematical Foundations of Modeling of Catalytic Processes: A Review. *Theor. Found. Chem. Eng.* **2000**, *34*, 263–273.
46. Voloshin, N.D. Petroleum Coke Dewatering. *Khim. Tekhnol. Topl. Masel* **1981**, *3*, 17.
47. Ivanchina, E.D.; Ivashkina, E.N.; Dolganova, I.O.; Belinskaya, N.S. Mathematical Modeling of Multicomponent Catalytic Processes of Petroleum Refining and Petrochemistry. *Rev. Chem. Eng.* **2021**, *37*, 163–191.
48. Sharikov, Y.V.; Sharikov, F.Y.; Krylov, K.A. Mathematical Model of Optimum Control for Petroleum Coke Production in a Rotary Tube Kiln. *Theor. Found. Chem. Eng.* **2021**, *55*, 711–719.
49. Iplik, E.; Aslanidou, I.; Kyprianidis, K. Hydrocracking: A Perspective towards Digitalization. *Sustainability* **2020**, *12*, 7058. <https://doi.org/10.3390/su12177058>.
50. Chistyakova, T.B.; Lavrova, A.S.; Novozhilova, I.V.; Dronov, S.V. Predicting Coke Characteristics from the Properties of the Raw Materials and the Coking Conditions. *Coke Chem.* **2024**, *67*, 325–330.
51. Muñoz, J.A.D.; Aguilar-López, R.; Castañeda, L.C.; Ancheyta, J. Comparison of Correlations for Estimating Product Yields from Delayed Coking. *Energy Fuels* **2013**, *27*, 7179–7190.
52. Ghashghae, M. Thorough Assessment of Delayed Coking Correlations against Literature Data: Development of Improved Alternative Models. *React. Kinet. Mech. Catal.* **2019**, *126*, 83–102.
53. Xiao, J.; Huang, J.; Zhong, Q.; Zhang, H.; Li, J. Modeling and Simulation of Petroleum Coke Calcination in Pot Calciner Using Two-Fluid Model. *JOM* **2016**, *68*, 643–655.
54. Orazbayev, B.; Orazbayeva, K.; Uskenbayeva, G.; Dyusseimbina, E.; Shukirova, A.; Rzayeva, L.; Tuleuova, R. System of Models for Simulation and Optimization of Operating Modes of a Delayed Coking Unit in a Fuzzy Environment. *Sci. Rep.* **2023**, *13*, 14317.
55. Shokri, S.; Sadeghi, M.; Ahmadi Marvast, M.; Narasimhan, S. Improvement of the Prediction Performance of a Soft Sensor Model Based on Support Vector Regression for Production of Ultra-Low Sulfur Diesel. *Pet. Sci.* **2015**, *12*, 177–182. <https://doi.org/10.1007/s12182-014-0010-9>.
56. Chen, J.; Gui, W.; Dai, J.; Jiang, Z.; Chen, N.; Li, X. A Hybrid Model Combining Mechanism with Semi-Supervised Learning and Its Application for Temperature Prediction in Roller Hearth Kiln. *J. Process Control* **2021**, *98*, 18–29.
57. de Oliveira, L.P.; Hudebine, D.; Guillaume, D.; Verstraete, J.J. A Review of Kinetic Modeling Methodologies for Complex Processes. *Oil Gas Sci. Technol.—Rev. IFP Energ. Nouv.* **2016**, *71*, 45. <https://doi.org/10.2516/ogst/2016011>.
58. Wang, J.; Anthony, E.J. A Study of Thermal-Cracking Behavior of Asphaltenes. *Chem. Eng. Sci.* **2003**, *58*, 157–162. [https://doi.org/10.1016/S0009-2509\(02\)00430-X](https://doi.org/10.1016/S0009-2509(02)00430-X).
59. Quintana-Solórzano, R.; Thybaut, J.W.; Marin, G.B.; Lødeng, R.; Holmen, A. Single-Event MicroKinetics for Coke Formation in Catalytic Cracking. *Catal. Today* **2005**, *107–108*, 619–629. <https://doi.org/10.1016/j.cattod.2005.07.036>.
60. Quliyeva, A.N.; Allahverdiyeva, G.Z. Modeling of the Coking Process. *Theor. Appl. Sci.* **2021**, *95*, 36–41. <https://doi.org/10.15863/TAS.2021.03.95.4>.
61. Speybroeck, V.; Hemelsoet, K.; Minner, B.; Marin, G.B.; Waroquier, M. Modeling Elementary Reactions in Coke Formation from First Principles. *Mol. Simul.* **2007**, *33*, 879–887. <https://doi.org/10.1080/08927020701308315>.

62. He, Z.; Xu, H.; Zhou, C.; Xin, Z.; Liu, J.; Shen, B. A Kinetic Model for in situ Coking Denitrification of Heavy Oil with High Nitrogen Content Based on Starch Using a Structure-Oriented Lumping Method. *RSC Adv.* **2018**, *8*, 32707–32718. <https://doi.org/10.1039/C8RA05561K>.
63. Mannanova, G.I.; Gubaydullin, I.M. Development of 15-Lump Kinetic Model of the Catalytic Cracking Process. *Comput. Math. Inf. Technol.* **2019**, *3*, 104–117. <https://doi.org/10.23947/2587-8999-2019-2-2-104-117>.
64. Zhou, X.; Chen, S.; Li, C.-L. A Predictive Kinetic Model for Delayed Coking. *Pet. Sci. Technol.* **2007**, *25*, 1539–1548. <https://doi.org/10.1080/10916460500529001>.
65. Anantpinijwatna, A. Coke Formation Model in Crude Oil Furnace for Maintenance Scheduling. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 43, pp. 79–84. <https://doi.org/10.1016/B978-0-444-64235-6.50015-2>.
66. Bakhtadze, N.N.; Lototsky, V.A. Knowledge-Based Models of Nonlinear Systems Based on Inductive Learning. In *New Frontiers in Information and Production Systems Modelling and Analysis*; Rózewski, P., Novikov, D., Bakhtadze, N., Zaikin, O., Eds.; Springer: Cham, Switzerland, 2016; Volume 98, pp. 83–106. https://doi.org/10.1007/978-3-319-23338-3_4.
67. Tun, M.; Emoto, G. Data Selection and Regression Method and Its Application to Softsensing Using Multirate Industrial Data. *J. Chem. Eng. Jpn.* **2008**, *41*, 374–383. <https://doi.org/10.1252/jcej.06WE047>.
68. Bolf, N.; Galinec, G.; Baksa, T. Development of Soft Sensor for Diesel Fuel Quality Estimation. *Chem. Eng. Technol.* **2010**, *33*, 405–413. <https://doi.org/10.1002/ceat.200900435>.
69. Zahedi, G.; Lohi, A.; Karami, Z. A Neural Network Approach for Identification and Modeling of Delayed Coking Plant. *Int. J. Chem. React. Eng.* **2009**, *7*, A68. <https://doi.org/10.2202/1542-6580.1832>.
70. Chen, X.; Wang, N. Modeling a Delayed Coking Process with GRNN and Double-Chain Based DNA Genetic Algorithm. *Int. J. Chem. React. Eng.* **2010**, *8*, A66. <https://doi.org/10.2202/1542-6580.2192>.
71. Mohler, I.; Galinec, G.; Hölbling, N.; Bolf, N.; Andrijić, Ž. Soft Sensors for Diesel Fuel Property Estimation. *Chem. Eng. Trans.* **2010**, *21*, 1423–1428. <https://doi.org/10.3303/CET1021238>.
72. Khan, S.I.; Hoque, A.S.M.L. SICE: An Improved Missing Data Imputation Technique. *J. Big Data* **2020**, *7*, 37. <https://doi.org/10.1186/s40537-020-00313-w>.
73. Jin, X.; Wang, S.; Huang, B.; Forbes, F. Multiple Model Based LPV Soft Sensor Development with Irregular/Missing Process Output Measurement. *Control Eng. Pract.* **2012**, *20*, 367–381. <https://doi.org/10.1016/j.conengprac.2011.10.007>.
74. Khatibisepehr, S.; Huang, B. Dealing with Irregular Data in Soft Sensors: Bayesian Method and Comparative Study. *Ind. Eng. Chem. Res.* **2008**, *47*, 8713–8723. <https://doi.org/10.1021/ie800386v>.
75. Kadlec, P.; Gabrys, B.; Strandt, S. Data-Driven Soft Sensors in the Process Industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814. <https://doi.org/10.1016/j.compchemeng.2008.12.012>.
76. Yusubov, F.V.; Aghamaliyev, Z.Z.; Namazova, S.R.; Aydinsoy, E.A. Mathematical Modelling and Optimization of Delayed Coking Unit Operation. *Proc. Petrochem. Oil Refin.* **2025**, *26*, 160–177. <https://doi.org/10.62972/1726-4685.2025.1.160>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.