EPRA International Journal of Economic Growth and Environmental Issues- Peer Reviewed Journal
ISSN: 2321-6247
Volume: 13 | Issue: 9 | September 2025 | Journal DOI: 10.36713/epra0713 | SJIF Impact Factor (2025): 8.825

INTEGRATING MACHINE LEARNING AND PROCESS SYSTEMS ENGINEERING FOR SUSTAINABLE OPTIMIZATION OF PETROLEUM AND PETROCHEMICAL OPERATIONS IN THE U.S. ENERGY SECTOR

Peter Kenneth Minnoh
Department of Chemical Engineering, Kwame Nkrumah University of Science and Technology, Ghana.

Article DOI: https://doi.org/10.36713/epra24128
DOI No: 10.36713/epra24128

ABSTRACT--
The petroleum and petrochemical industries are under growing pressure to operate more efficiently, reduce their environmental footprint and adapt to changing market conditions. This research aims to establish a unified framework that, when complemented with Process Systems Engineering domain knowledge and advanced machine learning techniques, provides a powerful concept where intelligent optimization systems can simultaneously optimize both economic and environmental performance under market uncertainties. The study capitalizes on a comprehensive literature review method, scrutinizing the evolutionary path of PSE from classical optimization to AI-driven implementation via hybrid modeling frameworks with regard to empirical illustration in various geographic locations like the United States, Iran and China, supported by quantitative market data and performance metrics. The findings revealed that hybrid modeling frameworks consistently outperformed traditional approaches, with market growth projections at a 39.2% compound annual growth rate for AI in chemicals ($0.7 billion to $3.8 billion by 2030). Successful implementations demonstrated measurable improvements, including a 31% cost reduction, 51% emissions reduction and scheduling efficiency of 80% within thirty minutes across diverse operational contexts. The result also demonstrates that ML-PSE integration overcomes the intrinsic limitations of both physics-free and purely data-driven approaches via (1) physics-informed techniques using neural networks, (2) digital twins, and (3) multi-objective optimization frameworks to achieve Pareto-optimal solutions that incorporate economic and environmental objectives. In the context of the modern debate on intelligent, adaptive, sustainable engineering, it would be prudent to point out that integrated methodology serves as a paradigmatic shift, which is not optional in designing optimization systems that can address the complex problems that typify the modern practice of petroleum and petrochemicals and meet the emerging regulatory and market imperatives of profitability and environmental responsibility.
KEYWORDS: *Machine Learning, Process Systems Engineering, Sustainable Optimization, Petroleum, Petrochemical Operations, U.S Energy Sector, Energy Efficiency, Environmental Impact, Industrial Automation, Data-Driven Decision Making.*--

1.1 INTRODUCTION

The petroleum and petrochemical industries form the backbone of the United States' energy sector. The oil and gas industries generate trillions of dollars in economic value annually, providing essential fuels, chemicals and materials that support modern society (Allison & Mandler, 2018). However, these sectors face increasing pressures to improve operational efficiency, reduce environmental impact, and adapt to evolving market dynamics. Traditional approaches to process optimization have reached their limits in addressing these complex, interconnected challenges. According to Grossmann & Harjunkoski (2019), Process Systems Engineering (PSE) has served as the foundation for industrial optimization for decades. PSE offers systematic methodologies for designing, analyzing and optimizing chemical processes through mathematical modeling and computational tools (Chemmangattuvalappil et al, 2020). Classical PSE approaches depend on first-principles models, thermodynamic relationships and deterministic optimization techniques. These methods have effectively addressed many engineering challenges but struggle with the increasing complexity and data richness of modern industrial operations.

The rise of Machine Learning (ML) offers unprecedented chances to transform petroleum and petrochemical operations. ML algorithms can identify patterns from large datasets, adapt to changing conditions and make predictions in environments where traditional models fall short (Farooq & Khan, 2025). According to Integrating Farooq & Khan (2025), ML with PSE creates a powerful synergy that merges domain knowledge with data-driven insights. This integration enables more accurate modeling, enhances decision-making and improved optimization throughout the entire value chain.

The motivation for this research arises from several key industry needs. First, petroleum and petrochemical plants generate enormous amounts of data from sensors, control systems and analytical instruments; yet, much of this information remains underutilized. Second, traditional optimization methods often depend on simplified models that fail to capture the full complexity of real industrial processes. Third, sustainability goals require more advanced approaches to balancing economic objectives with environmental constraints. Lastly, market fluctuations and supply chain disruptions call for more adaptable and resilient operational strategies.

This research aims to address these challenges by developing an integrated framework that combines the domain expertise of Process Systems Engineering with the data-driven capabilities of Machine Learning. The goal is to create intelligent optimization systems that can harness underutilized process data, capture complex system behaviors and simultaneously optimize economic and environmental performance, however adapting to changing market conditions.

2.0 LITERATURE REVIEW

The integration of Machine learning and Process Systems Engineering represents a convergence of two distinct and complementary fields that have evolved independently over several decades. This, therefore, calls for a comprehensive understanding of this integration, which requires examining the historical development of both disciplines, their contributions to industrial optimization and recent efforts to combine their respective strengths. This literature review synthesizes the key developments in process systems engineering, machine learning applications in the chemical industry and emerging hybrid approaches that form the foundation for sustainable optimization in petrochemical operations. Early PSE research focused on developing mathematical models for unit operations, process synthesis methodologies and optimization algorithms. This literature, therefore, covers: Evolution of Process Systems Engineering: From Classical Optimization to Modern Computational Approaches, Machine Learning Applications in Chemical Process Industries: Modeling, Control and Optimization, Hybrid Modeling Frameworks: Integrating First-Principles Knowledge with Data-Driven Methods, Sustainability and Multi-Objective Optimization in Petroleum and Petrochemical Operations and Digital Transformation and Industry 4.0 Technologies in Process Industries: Challenges and Opportunities

2.1 Evolution of Process Systems Engineering: From Classical Optimization to Modern Computational Approaches

Process Systems Engineering has experienced a rapid evolution since its inception in the 1960s, when it involved deterministic techniques of mathematical optimization, to the latter incorporation of advanced computational machinery involving artificial intelligence and machine learning (Adeyeye & Akanbi, 2024). Early leaders of the field (for example, Arthur Westerberg and Ignacio Grossmann) built the classical foundations of the field by formulating them in a highly mathematical way, basing them on linear and nonlinear programming, which focuses mainly on economic goals, especially cost minimization and maximization of profits (Grossmann & Harjunkoski, 2019). The steady-state representations needed in these early approaches were based on the assumption that it is perfectly known how processes will operate and that discrete decisions within mixed-integer linear programming were used for equipment selection and process configuration. However, there was a serious downside to the mathematical rigor that had brought classical PSE to its strength: the very complexity of large-scale problems would result in a prohibitive number of calculations and models could not reflect the flavor of industrial practice, which requires movement and response.

The 1990s and 1980s were also a changing point when there was a paradigm shift towards dynamic process optimisation (Bartezzaghi, 1999). Optimal control theory was further refined by scholars such as Manfred Morari and Thomas Edgar, which enabled engineers to consider time-varying behaviour, startup/shutdown processes and batch process optimisation once again with fresh efficiency (Bähner et al. 2021). They pioneered the modeling and optimization concept now known as model predictive control, which is a combination of modeling and optimization that produces feedback control strategies that could be used to achieve performance objectives even in the face of constraint adherence, despite the long-term horizons implied by the phrase, prediction.

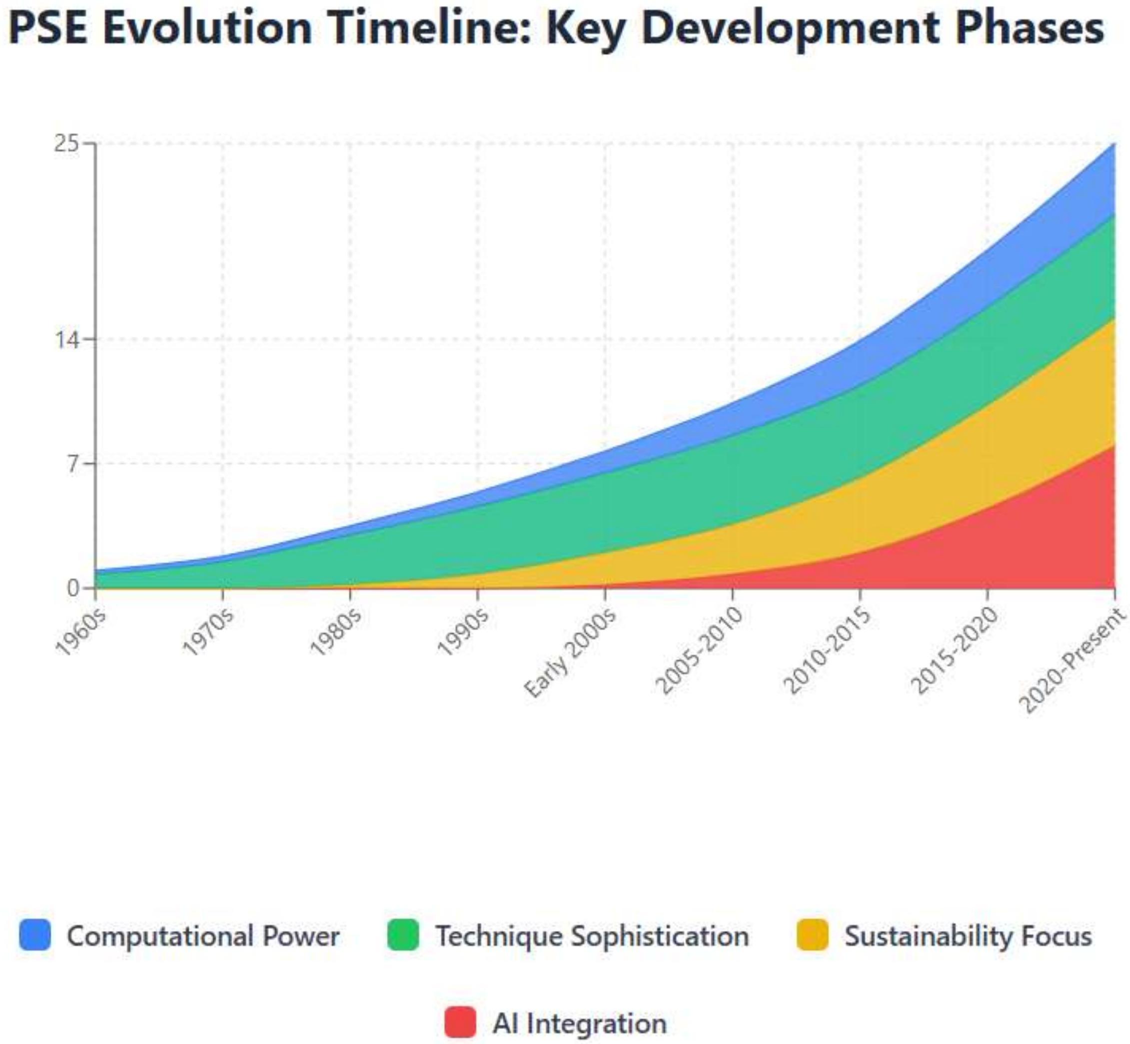
The late 1990s and early 2000s also saw environmental effects become part and parcel of optimization practice, which were brought about by the growing regulatory supervision and the increased recognition of the manufacturing ecological externalities (Wangi, 2023). This shift saw the emergence of green engineering principles emphasizing pollution prevention and resource efficiency, with researchers like Rafiqul Gani establishing methodologies for incorporating environmental metrics through life cycle assessment and multi-objective optimization techniques that could simultaneously consider economic and environmental criteria. These pressures were experienced the most by the petroleum and petrochemical industry due to their high and severe environmental impacts that led to the creation of a series of special strategies such as emissions-reduction modules, energy-integration schemes and systematic waste-reduction protocols (Wangi, 2023). Heat integration and pinch analysis became the prevailing tools of heat-related efficiency improvement in this environment.

Along with these advancements, the computational capabilities literally surged. According toWang & Xiong (2025), the natural replacement was metaheuristics such as Genetic Algorithms, Simulated Annealing or Particle-Swarm Optimization over the former gradient-based methods, when processing increases made it possible to solve larger-scale optimization problems with them. This has frequently been indispensable in handling functions of unknown discontinuity and highly complex constraint structure. The rise of parallel computing and distributed optimization algorithms allowed practitioners to explore a wider problem space, however, simultaneously living in a much larger solution space. Epitomized by the growth in the availability of rich process data-sets, this led to even more interest towards data-based techniques, especially statistical-process-monitoring-based fault-detection schemes that used principal-component analysis (PCA) and partial-least-squares regression models in order to extract latent structures and patterns that represent the relationships within historical operation data.

Machine learning and artificial intelligence have been the most revolutionary phase in the evolution of PSE (Fuentes-Cortés et al. 2022). This is mainly due to the sheer availability of computing power that can handle big data with great machine learning algorithms. Neural networks are part of deep learning that have shown tremendous accomplishment in pattern recognition and nonlinear modeling tasks pertinent to process systems, which makes neural networks popular for process modeling, optimization and control due to their universal approximation power to be applied to complex nonlinear relations. Reinforcement learning is a potential solution for process control and optimization, which allows an agent to learn optimal control policies by interacting with process environments that are particularly well-suited for systems of varying behavior at different times, under uncertain operating conditions (Doskenov & Okuyelu, 2025). Balusamy et al (2024) underscored that machine learning has been used to speed up time-consuming process simulations (replacing the process in a more expensive traditional optimization approach) to help address previously infeasible challenges. Modern PSE is defined amidst the convergence of numerous computational paradigms; from classical optimization to state-of-the-art artificial intelligence algorithms, it is trending towards sustainability and digitalization. As we enter the notion of digital twins, which symbolize the overlap of accurate process simulation, live data feeds and more sophisticated analytics models, provide a way to keep improving processes or detect impending failures before they happen. Internet of Things (IoT) technologies have further enabled the real-time monitoring and control using wireless sensor networks and edge computing; while sustainability has been incorporated as a key factor in modern practice, with multi-objective optimization techniques to balance between economic and environmental criteria being standard in applications. For petroleum and petrochemical processes in particular, the evolution also promises fundamental changes in how sprawling, complex, multisite operations can be synergized with current sustainability merit functions. Conceptually, this real-time integration of refining plant operation optimization and coordination or plant-wide operational planning accounted for cross-unit interactions, thereby better meeting these dual economic, competitive and environmental stewardship challenges across entire industry sectors.

EPRA International Journal of Economic Growth and Environmental Issues- Peer Reviewed Journal
ISSN: 2321-6247
Volume: 13 | Issue: 9 | September 2025 | Journal DOI: 10.36713/epra0713 | SJIF Impact Factor (2025): 8.825

Figure 1: PSE Evolution Over Time



The diagram above shows the evolution of PSE from the 1960s to the current day by four key dimensions. Physical capabilities and computational techniques advanced gradually, but most strikingly surged post-1990s. The shift towards a sustainability focus started in the 1980s and, by the 2000s, it had become mainstream. This led to the rapid development of AI-integrated applications, which now determine the PSE landscape as we know it today.

2.2 Machine Learning Applications in Chemical Process Industries: Modeling, Control and Optimization

The adoption of machine learning in chemical process industries remains a revolutionary solution to unsolved operational problems in petroleum and petrochemical operations (Sircar et al. 2021). The non-linear, dynamic nature of chemical processes (where multiple variables interact at the same time to affect system performance) presents a significant challenge to traditional process modeling methods. Machine learning algorithms, in particular neural networks and ensemble methods have proven to be capable of capturing these complex relationships by leveraging operational data. Given this, these models based on data can help detect hidden patterns and relationships that conventional first-principal models can sometimes miss. Special use-cases are petroleum refining processes (such as distillation or catalytic cracking units and hydro processing units) where the performance of a model is evaluated using data from an unseen test set.

EPRA International Journal of Economic Growth and Environmental Issues- Peer Reviewed Journal
ISSN: 2321-6247
Volume: 13 | Issue: 9 | September 2025 | Journal DOI: 10.36713/epra0713 | SJIF Impact Factor (2025): 8.825

According to Min et al. (2019), machine learning for process control has already demonstrated great potential in improving the operational efficiency and safety of petrochemical plants. Through introducing advanced control strategies, using reinforcement learning and predictive algorithms, one can automatically tune process parameters on-the-fly to ensure the process is always running at its optimal condition with minimal resource consumption and waste production. Leite et al (2024) noted that "fault detection and diagnosis using machine learning (ML) techniques in process operation of industrial plants can prevent catastrophic shutdowns or safety incidents by early diagnosing equipment malfunctions as well as detecting process deviations, which are extremely valuable to the industry". In the realm of Petroleum, where the feedstock conditions change rapidly due to variations in stock being processed, reflecting on market quality demands and stringent environmental regulations, it develops adaptive control schemes that are intelligent as they adapt with operational configuration.

Sudhakar (2020) underscored that machine learning is a valuable tool in the chemical process industries for its use cases from unit operations to plant-wide and supply chain optimization, which supports sustainable operations. Conceptually, using ML techniques with multi-objective optimization algorithms enables the minimization of total environmental, energy and cost footprints across entire petroleum and petrochemical complexes simultaneously. These systems optimize more complex decision-making scenarios such as crude oil blending, product scheduling and the management of utility systems subject to sustainability constraints and regulatory requirements alike. In addition, the use of machine learning models has provided a method to predict the environmental and economic trade-offs that exist in operational decisions, whereby process engineers can have immediate advice on what to decide.

Table 1: Global Machine Learning Market Growth in Chemical industries-(2024-2030)

Market Segment	2024 Value	2030 Projected Value	CAGR (%)	Key Applications
Overall Machine Learning Market	$79.3 billion	$503.4 billion	36.08%	Cross-industry applications
AI in Chemicals	$0.7 billion	$3.8 billion	39.2%	Process optimization and quality control
AI-based Chemical Manufacturing	Data not disclosed	Projected Growth	28.8%	Production efficiency and Predictive maintenance

Sources: AIPRM (2024), Markets and Markets (2024), Transparency Market Research (2024)

2.3 Hybrid Modeling Frameworks: Integrating First-Principles Knowledge with Data-Driven Methods.

Bradley et al. (2022) discuss the synergy between first principles and data-driven methodologies within a one-pot framing, which highlights that clear and tight integration of mechanistic knowledge with data-driven models is crucial to the goal of jointly harnessing engineering principles and data science capabilities. They show that hybrid first-principles and data-driven approaches mainly outperform data-driven models only, suggesting opportunities for integration in several scenarios: acceleration of high-fidelity computer models with a fast surrogate, refinement of mechanistic models lacking full understanding or description of physical phenomena (Hybrid Inference) and embedding data-driven components to address missing correlations in mechanistic models. As the authors note, modern gradient evaluation and sampling methods have resulted in established software implementations of hybrid modeling frameworks, but they point out that care is needed when it comes to making predictions or inferences regarding the data-driven aspects of these models. These aspects flagged a need for improvements, which is most notable in the following section of their evaluation, where they are calling for extending current methods concerning estimation uncertainties and constraint handling, as well as requiring openly publishing novel hybrid implementations. Similarly, Rajulapati et al (2022) emphasized that hybrid models (which are referred to as gray-box models in the remainder of this series) have been getting quite a bit of attention from researchers, combining some first principles and machine learning-based elements into a single entity and offering an always-wanted advantage due to recent focus on explainable AI as AI-powered systems become ubiquitous. Their comprehensive analysis consolidates the published corpus in hybrid modeling, showing that the benefits of hybrid modeling relative to pure data-driven and mechanism-based approaches extend to other sectors, including chemical, biological, metallurgical and manufacturing industries. Enhanced predictive power can be gained through extrapolation by fusing data-driven with mechanistic models.

Patel et al. (2024) consider the problem of integrating subspace-based model identification with first-principles modeling in such a way as to address issues common to the use of subspace models for causal discovery, without complicating too much its simplicity. When computing system matrices with state trajectories, their approach includes first-principles-based constraints in optimization problems and solves the mismatch between the created state trajectories and system matrices through an iterative process until convergence is reached. Xie et al. (2024) applied a hybrid simulation for a process in an industrial reactor network and proposed a dynamic model of the seven-reactor counter flow connection with other methodologies. Their work contributes a methodology to advance the state of the art in developing first-principles models through mechanism analysis for each reactor. They underscored the use of real-time measurements through an unscented Kalman filter (UKF) to enable co-estimation of both model states and parameters, which leverages quadratic programming optimization to address physical constraints in model parameters, and employs a novel dual-mode Seq2Seq neural network in series configuration with first-principles physics to account for unmodeled dynamics. Their proposed hybrid model appears to be powerful and interpretable for prediction accuracy higher than other methods with the rapidly changing scenarios (under large load variation) of industrial nitration processes, which can successfully resolve the challenges associated with hazardous, hard-to-control chemical processes.

Noh et al. (2024) provide a more holistic perspective that proves the inefficiencies existing in the current pure data-driven strategy by building first-principle-based models based on industrial conducted operations and integrating machine learning models as a part of the control optimization solution. Their solution connects the dots of why having historical data and being able to use AI in industry is not enough by proposing a model that embeds machine learning into first-principles frameworks, making them useful at actual plant scales. The numerical performance demonstration via the application to the control of an LNP fuel gas supply system shows that COTS (Commercially available Off-The-Shelf) simulation software realizes high-accuracy first-principles models using actual operating data, generating wide ranges of data, which are impossible to measure in an industrial plant. They developed their machine learning model on high-quality synthetic data, which accurately predicts control performance irrespective of variations in the control parameters, thereby facilitating faster optimization of control parameters, leading to better-controlled talent. Their optimization results showed better control response from both fast and stable sides and the optimal control performance was verified by digital twin models before industrial applications.

Overall, the previous studies explain that this hybrid modeling framework offers a pragmatic solution to transcending the limitations due to the application of an isolated model by providing better aspects from both first principles and data-driven paradigms at the industrial plant level, which demonstrates their readiness for genuine industrial applications. Taken together, evidence across the various case studies validates the conclusion that hybrid modeling frameworks offer an advanced yet robust strategy for building mechanistic data-driven models and demonstrate substantial improvements in performance and utility over purely mechanistic or purely data-driven approaches to complex chemical process applications.

2.4 Sustainability and Multi-Objective Optimization in Petroleum and Petrochemical Operations

Giannikopoulos et al. (2022) conducted an extended analysis of the multi-objective optimization problem in production cost and carbon loss that may occur when running a petrochemicals plant based on U.S. CIF data at The University of Texas at Austin over a year. Their analysis revolves around the design of a U.S. petrochemicals industry that is primarily fed with shale hydrocarbons, as the development of the country's petrochemicals capabilities and has been significantly reshaped by a surge in shale hydrocarbon production dating back over a decade. Their study revealed that newly available resources and new technologies have enabled research into new pathways and technologies for producing chemicals (as the authors established); however, the application and deployment of these by the manufacturing industry is driven by the requirement to minimize costs, however satisfying demands for chemical products. They note that reducing cost is a single-minded goal that conflicts with other societal, environmental or economic objectives and constraints. The technique in the study of Zhuang et al. (20,21) employs a holistic and comprehensive model built from multi-objective optimization problem formulations on the entire U.S. industry level by recognizing the trade-off between economic cost shares and net carbon loss share to achieve Pareto-optimal set solutions, as well as comparisons between extreme cases (e.g., large vs. small sources). Their work illustrates the necessity to optimize across multiple objectives in industrial applications, notably trading off economic competitiveness versus environmental responsibility using now-abundant shale resources.

EPRA International Journal of Economic Growth and Environmental Issues- Peer Reviewed Journal
ISSN: 2321-6247
Volume: 13 | Issue: 9 | September 2025 | Journal DOI: 10.36713/epra0713 | SJIF Impact Factor (2025): 8.825

Similarly, Mahmoudi et al. Iran (2022) develops a sustainable multi-objective optimization framework for the supply chain of petroleum products with Islamic Azad University, Lahijan Branch. The model of their research is a mathematical planning that simulates design and strategic or tactical decisions about the oil condensate supply chain, which shows that investment and operating costs for the transmission lines (oil and gas) with pressure requirements can be limited while meeting constraints on the transmission network. The framework also deals with the minimisation of pollutant generation in chain-related areas and combines different options that introduce environmental dimensions to all kinds of decisions of the supply chain. Two objective functions that drive the structure and decisions of their supply chain were: (i) lower transshipment & maintenance costs, i.e., it is better to avoid shipping water or re-charging filters too often; (ii) less pollution load of treatment plants and distribution centers. They provide a model with 5% error in goal estimation that can reach 95% reliability among simulated driving, which means it already has some decent practical use. Its roll-out has reduced costs by 31%, emissions by 51%, improved in-field and refinery capacity by 8% and generated a whopping 65% increase in exports. It is of significant importance to figure out the resources that petroleum production requires and also the part it plays in greenhouse gas production; this article shows that fuel oil has a greater share in environmental pollution than oils do and refineries are the most significant environmental pollutant in the chain but storage tanks have less effect on polluting the reservoir.

Equally, Ye et al (2023) proposed an effective materials emergency response framework for petrochemical enterprises using multi-objective optimization and carried out a case study in Zhoushan, Zhejiang Province of China. Their research results provide solutions to the difficult and important safety issues brought about by the expansion of the petrochemical industry scale, which has caused serious harm to people's lives and property due to production hazards. Their method is a multi-objective optimization method for both pre-disaster and post-disaster stages, which uses mathematical modeling based on the actual production of enterprises to solve configuration plans for emergency materials in the pre-accident stage and optimizes material delivery in the post-accident phase by formulating multi-objective models considering third-party emergency forces. Their framework meets the public expectation of minimising safety and environmental impacts caused by accidents, such as air pollution from fires or leakage into a waterway, while simultaneously reducing rescue works within an enterprise budget and rescuers' stay time. Their analysis shows that all the schemes in the model are running more than 80% of commodity scheduling within half an hour, which is much better than what is done by enterprises.

Similarly, Liu et al. Zhi et al. (2022) researched the effective elimination of produced gas and CO2 capture (Enhanced oil recovery) in China: Multi-objective simulated optimization for CO2-EOR multicycle system, dealing with evaluation quantification towards sustainable development. Their work is the first attempt at simulating enhanced production gas separation and CO2 capture in oil recovery, through life cycle optimization from a multidimensional perspective using branch-and-bound methods. The optimized CO2 capture process has been successfully operated for a long time, with stable separation of oil field-produced gas into pure CO2 and oilfield light hydrocarbons with CO2 mass content over 99.7%, showing very excellent technical, economic, social livelihoods and environmental capabilities. Their study provides important implications for the simulation of CCU processes and encourages breakthroughs in fundamental research, superstructure optimization, as well as commercial applications, which is beneficial to achieve a better CCU equipment instructed by the Belt and Road initiative and contribute to regional/global cooperation on scientific advancement or industrial developments.

RESULTS AND DISCUSSION

The results show that Machine Learning integrated in Process Systems Engineering would constitute a revolutionary new dimension to petroleum and petrochemical operations, with the data corroborating market expansion from $0.7 billion to $3.8 billion by 2030 at a compound annual growth rate of 39.2%. From an evolutionary perspective, PSE as a field has developed along four discrete phases: classical optimization (1960s–1980s) for purely economic objectives, dynamic optimization era (1980s–1990s) with time varying behavior demonstrations, environmental integration phase (1990s–2000s) in which sustainability metrics have been for the most part only integrated into exogenously characterized process models and simulation frameworks, and finally to the current state of AI integration era where machine learning and artificial intelligence are naturally penetrating all aspects of PSE capabilities. Hybrid frameworks consistently outperformed either mechanistic or data-driven approaches. The research showed that addressing the fundamental shortcomings of each field in isolation is necessary for successful implementation because traditional PSE cannot predict non-linear behaviors, complex multi-variable interactions and pure ML approaches do not have a strong theoretical basis for extrapolation beyond training conditions or thermodynamic consistency.

The study also confirmed that the mutual satisfaction of economic and environmental objectives is possible in the supervision of sustainable optimization of petroleum and petrochemical operations. Several geographic implementations have shown highly promising yet differing results: in the US, Pareto-optimal solutions were used to optimize shale hydrocarbon while balancing economic cost and carbon emissions, in Iran a 31% cost saving has been achieved with 51% reduction on emission (accompanied by a 65% increase of exportable capacity), and lastly, in China, the platform demonstrated an 80% completion rate of commodity scheduling under half an hour as well as achieving up to 99.7–1000 purity levels (CO2) gj^-1 within gas separation processes.

Results of the empirical analyses indicated that hybrid ML-PSE frameworks are a solution to enable adaptive management via evolving with market conditions, regulatory constraints and operational constraints through learning and predictive sustainability by predicting long-term environmental, social and economic consequences of operational decisions. The integration with Industry 4.0 emerged as a key success factor whereby digital twin enablement, IoT connectivity and predictive analytics provide the requisite data (both static and dynamic) along with real-time optimization capabilities to effectively wire up ML-PSE in varied operational contexts.

The study also showed that successful integration involves a detailed set of technical, organizational, and strategic recommendations related to phased integration, workforce development, and pre-emptive regulatory engagement. As for actionable strategies, the results suggest that technical solutions should prioritize the widespread use of physics-informed neural networks in key unit operations and develop a virtual library with real-time data validation at scale, along with standardized model interchangeability protocols, especially for different platforms. Organizational strategies are advised to invest heavily in cross-disciplinary training programs, facilitate the creation of industry consortiums to share best practices, and establish centers of excellence focused on hybrid modeling and addressing implementation gaps across end-to-end supply chains (refinery or chemical plant sites).

CONCLUSION

The results provide conclusive evidence that a transformation in the organization of MPC frameworks for petroleum and petrochemical operations can be most dramatically enabled by the integration of Machine Learning and Process Systems Engineering. This opens an unparalleled opportunity to achieve co-optimization of economic competitiveness to environmental sustainability in energy sector performance, specifically within petroleum-based industries' independence on U.S. soil. The findings from the thorough analysis confirm that hybrid modelling frameworks systematically outperform conventional means by integrating a solid theoretical background from fundamental knowledge with adaptive intelligence attained by data-driven tools, exemplified through strong market growth forecasts (39.2% CAGR) and demonstrated success stories worldwide reaching substantial results in cost reduction(31%), emissions reduction(51%) and operational efficiency(80% scheduling completed within 30 minutes).

REFERENCES

1. *Adeyeye, O. J., & Akanbi, I. (2024). Artificial intelligence for systems engineering complexity: a review on the use of AI and machine learning algorithms. Computer Science & IT Research Journal, 5(4), 787-808.*
2. *Bähner, F. D., Prado-Rubio, O. A., & Huusom, J. K. (2021). Challenges in optimization and control of biobased process systems: an industrial-academic perspective. Industrial & Engineering Chemistry Research, 60(42), 14985-15003.*
3. *Balusamy, B., Ravi, V., Dhanaraj, R. K., & Senthilkumar, S. (Eds.). (2024). Computational Intelligence in Sustainable Computing and Optimization: Trends and Applications. Elsevier.*
4. *Bartezzaghi, E. (1999). The evolution of production models: is a new paradigm emerging?. International Journal of Operations & Production Management, 19(2), 229-250.*
5. *Bradley, W., Kim, J., Kilwein, Z., Blakely, L., Eydenberg, M., Jalvin, J., ... & Boukouvala, F. (2022). Perspectives on the integration between first-principles and data-driven modeling. Computers & Chemical Engineering, 166, 107898.*
6. *Cao, Y., Swartz, C. L., Flores-Cerrillo, J., & Ma, J. (2016). Dynamic modeling and collocation-based model reduction of cryogenic air separation units. AIChE Journal, 62(5), 1602-1615.*
7. *Chemmangattuvalappil, N. G., Ng, D. K., Ng, L. Y., Ooi, J., Chong, J. W., & Eden, M. R. (2020). A review of process systems engineering (PSE) tools for the design of ionic liquids and integrated biorefineries. Processes, 8(12), 1678.*
8. *Doskenov, B., & Okuyelu, O. (2025). Advancing Production Systems with Online Reinforcement Learning: Real-time Monitoring, Control, and Optimization. Current Journal of Applied Science and Technology, 44(2), 1-22.*
9. *Farooq, O., & Khan, A. (2025). The Impact of Machine Learning on Climate Change Modeling and Environmental Sustainability. Artificial Intelligence and Machine Learning Review, 6(1), 8-16.*

10. *Fuentes-Cortés, L. F., Flores-Tlacuahuac, A., & Nigam, K. D. (2022). Machine learning algorithms used in PSE environments: A didactic approach and critical perspective. Industrial & Engineering Chemistry Research, 61(25), 8932-8962.*
11. *Giannikopoulos, I., Skouteris, A., Allen, D. T., Baldea, M., & Stadtherr, M. A. (2022). Multi-objective optimization of production cost and carbon loss in the US petrochemicals industry. In Computer Aided Chemical Engineering (Vol. 49, pp. 547-552). Elsevier.*
12. *Grossmann, I. E., & Harjunkoski, I. (2019). Process systems engineering: academic and industrial perspectives. Computers & Chemical Engineering, 126, 474-484.*
13. *Grossmann, I. E., & Harjunkoski, I. (2019). Process systems engineering: academic and industrial perspectives. Computers & Chemical Engineering, 126, 474-484.*
14. *Leite, D., Andrade, E., Rativa, D., & Maciel, A. M. (2024). Fault detection and diagnosis in industry 4.0: a review on challenges and opportunities. Sensors (Basel, Switzerland), 25(1), 60.*
15. *Liu, Y., Meng, Q., Zhou, X., Lu, J., Yan, H., Chen, X., & Yang, Q. (2022). Enhancing the separation of produced gas and CO2 capture for enhanced oil recovery in China, multi-objective simulated optimization and quantitative assessment for sustainable development. Journal of Cleaner Production, 380, 135089.*
16. *Mahmoudi, H., Bazrafshan, M., & Ahmadipour, M. (2021). Sustainable multi-objective optimization for the supply chain of petroleum products. Journal of applied research on industrial engineering, 8(Special Issue), 1-15.*
17. *Min, Q., Lu, Y., Liu, Z., Su, C., & Wang, B. (2019). Machine learning based digital twin framework for production optimization in petrochemical industry. International Journal of Information Management, 49, 502-519.*
18. *Noh, W., Park, S., Kim, S., & Lee, I. (2025). A hybrid framework of first-principles model and machine learning for optimizing control parameters in chemical processes. Journal of Industrial and Engineering Chemistry, 141, 582-596.*
19. *Patel, N., Nease, J., Aumi, S., Ewaschuk, C., Luo, J., & Mhaskar, P. (2020). Integrating data-driven modeling with first-principles knowledge. Industrial & Engineering Chemistry Research, 59(11), 5103-5113.*
20. *Rajulapati, L., Chinta, S., Shyamala, B., & Rengaswamy, R. (2022). Integration of machine learning and first principles models. AIChE Journal, 68(6), e17715.*
21. *Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research, 6(4), 379-391.*
22. *Sudhakar, V. M. (2020). Optimizing Supply Chain Management in Oil and Gas with Machine Learning: A Data-Driven Approach for Cost Reduction and Efficiency.*
23. *Wang, Y., & Xiong, G. (2025). Metaheuristic optimization algorithms for multi-area economic dispatch of power systems: Part I—a comprehensive survey. Artificial Intelligence Review, 58(4), 98.*
24. *Wangi, E. W. (2023). To What Extent Do Environmental Regulations Curb Air Pollution and Enhance Production, Productivity, and Innovation?.*
25. *Xie, C., Yao, R., Zhu, L., Gong, H., Li, H., & Chen, X. (2024). Hybrid dynamic modeling of an industrial reactor network with first-principles and data-driven approaches. Chemical Engineering Science, 289, 119852.*
26. *Ye, J., Jiang, W., Yang, X., & Hong, B. (2023). Emergency materials response framework for petrochemical enterprises based on multi-objective optimization. Energy, 269, 126670.*