Surrogate Based Mixed Integer Linear Programming Model for Decarbonization of an Integrated Gas-Oil Separation Network

Abdullah Bahamdana, Nilay Shaha, Antonio del Rio-Chanonaa\*

aSargent Centre for Process Systems Engineering, South Kensington.

London, SW7 2AZ, United Kingdom

\*Corresponding author: a.del-rio-chanona@imperial.ac.uk

Abstract

Energy companies seek to decarbonize their business operations and reduce emissions. However, achieving this goal while maintaining a solid financial performance is challenging. This research focuses on the upstream sector of the oil and gas industry, where a cluster of producing wells supply a network of interconnected gas-oil separation (GOSP) facilities. Existing methods in literature exhibit several key limitations, such as lack of sustainability objectives, utilization of simplified approaches, or adoption of computationally expensive models to optimize network operations. To overcome such challenges, a novel data-driven approach is proposed. It employs artificial neural networks within a multi-objective mixed integer linear programming (MILP) framework to optimize feed allocation and equipment utilization while reducing emissions across the network. Moreover, the surrogate models are trained using a high-fidelity model to capture possible feed uncertainties. The proposed framework is tested using real-world industrial production scenarios. The solution demonstrated acceptable capabilities in reducing greenhouse gas emissions while maximizing profitability.

**Keywords**: upstream processing, machine learning, multi-objective optimization, enterprise-wide optimization, supply chain.

* 1. Introduction

Since the Paris Agreement came into effect, the pressure on the energy industry to achieve sustainability objectives has escalated. This is mainly attributed to the conflict between minimizing emissions and maximizing the profitability required to finance capital-intensive energy transition projects. Accordingly, this work is proposed to support practitioners in optimizing the trade-off between sustainability and profitability metrics.

This project focuses on the upstream and midstream sectors of the oil and gas industry. In particular, the interface between the production wells and gas-oil separation facilities (GOSP). In a conventional oil and gas reservoir that is spread across a vast geographical space, production wells are clustered into groups, while each group of wells supplies wet crude oil to a network of GOSPs. The GOSPs are interconnected via swing pipelines, typically used during maintenance shutdowns and operational disturbances to divert feed flow from one facility to another.

Figure 1 illustrates a block flow diagram of a typical GOSP configuration. The feed, which consists of crude oil, oily water, and natural gas, is fed into the high-pressure production trap (HPPT) and low-pressure production trap (LPPT) in series to separate the three phases. The gas streams are mixed, compressed in the high and low-pressure compressors, and transferred to downstream facilities for further processing. Meanwhile, the crude oil stream is dehydrated, desalted, and transferred to either refining for further processing or terminals for export. The oily water stream is collected throughout the system and processed in the water-oil separation section (WOSEP) to remove hydrocarbon traces before being injected into the reservoir to maintain its pressure profile.



Figure 1: Block flow diagram of a typical GOSP

The literature covers numerous contributions from academics and practitioners in applying legacy chemical engineering tools to improve the efficiency and performance of processing industries. This includes optimizing unit process parameters, retrofitting equipment, and upgrading technologies. In recent years, enterprise-wide optimization (EWO) has emerged as an exciting research field within process systems engineering (PSE). It seeks to achieve higher combined rewards by expanding the envelope and solving complex optimization problems considering several supply chain functions (Grossmann, 2012). Consequently, researchers have explored improving the overall performance of a network of GOSPs by utilizing swing pipelines during normal operations to optimally allocate loads between the facilities. The first effort was instigated by developing a mixed-integer linear programming (MILP) model to minimize the network's overall operating expenditure (OPEX) (Liu et al., 2016). Each facility was modeled using a state-task network (STN) approach, whereas simplified linear equations were used to model the nonlinear equipment behaviour. However, this work only considers a profit-driven objective and the simplified linear equations cannot fully capture process nonlinearity. Leveraged by this work, researchers explored the possibility of utilizing a mixed-integer nonlinear programming (MINLP) model to minimize OPEX (Al-Ghazal et al., 2020). In contrast, physics-based simulations were used to model each facility and achieve higher savings. However, this approach presents drawbacks, primarily due to utilizing a profit-driven objective and hindering scalability due to expensive computational costs.

This work aspires to overcome the drawbacks by proposing a generalizable and robust framework that couples machine learning algorithms in a multi-objective MILP model to maximize profitability while minimizing emissions. The paper is structured as follows: in section 2, the proposed methodology is thoroughly explained; in section 3, the case study is introduced; in section 4, the results are discussed; in section 5, the work is concluded with a research outlook.

* 1. Methodology

This work aims to maximize profitability and minimize emissions by solving a multi-objective optimization problem. According to domain knowledge expertise, the profitability objective is achieved by minimizing OPEX, resulting mainly from power consumption. Meanwhile, the emissions objective is realized by minimizing emissions from energy consumption tasks and minimizing/eliminating process venting and flaring events. The latter occurs when the gas feed flow rate exceeds the facility's maximum capacity.

Given the goal of this work, the proposed methodology consists of four main phases: the development of high-fidelity models, the training and selection of surrogate models, the formulation of the optimization problem, and solving and validating the results. Figure 2 demonstrates and summarizes the four phases in a visual format.

The first phase of the methodology concerns the development of high-fidelity models to capture the complexity of an operating facility. Equipment items highly correlated with the objective function are rigorously modeled, while others are simplified for computational ease. Process simulation software is then used to generate an appropriately large dataset for a variety of operating scenarios to ensure information-rich data. In the second phase, the dataset is used to train and test several machine learning algorithms. A suitable surrogate model is selected following several iterations of training/testing and hyperparameters tuning according to predefined performance metrics (van de Berg et al., 2022). In the third phase, the case study is formulated as a multi-objective optimization problem consisting of an objective function, constraints, and binary and continuous variables. In the fourth phase, the problem is solved using a MILP approach, which requires linearization techniques for problem reformulation (Grossmann, 2021). The results are then generated and validated.



Figure 2: Flowchart of the proposed methodology

* 1. Case Study

An abstract network consisting of three interconnected facilities (A, B, C) is employed to demonstrate the performance of the proposed methodology. The network specifications are based on a real-world example presented in (Liu et al., 2016). The facilities are assumed to have identical layout configurations and component capacities according to the following: crude oil is 330 “kbdoe”, water is 160 “kbdoe”, gas is 30 “kbdoe”. Moreover, they are assumed to operate at a steady state, with optimal operating conditions.

Starting with the high-fidelity model, Aspen HYSYS was used to simulate the flowsheet. According to domain knowledge expertise, approximately 80 % of the facility’s OPEX is attributed to the power consumed by rotating equipment. This includes the five major types: high and low-pressure compressors, dry crude shipper pumps, and sour water injection pumps. Major equipment has been rigorously modeled using historical data and power consumption curves from a real-world facility to capture the nonlinearity. The case study feature in Aspen HYSYS was used to generate a rich dataset of 18,479 data points. The dataset covers 25 feed quality scenarios by varying the gas-to-oil ratio and water cut percentages. This is essential to ensure the robustness of the solution, as data-driven models, or surrogate models in this case, perform optimally when interpolating. To overcome the imbalanced dataset resulting from the gas flow rate values being magnitudes higher than the liquid component flow rates, the dataset values were rescaled to meet the characteristics of the standard normal distribution. This will significantly enhance the efficiency and accuracy of the machine learning-based surrogate models. This standardization, given in Eq. (1), uses the z-score normalization method and was applied to the dataset:

|  |  |
| --- | --- |
|  | (1) |

Where *x’* is the normalized value, is the mean, and is the standard deviation. In selecting the surrogate model type, three machine-learning algorithms have been investigated. The first is the Artificial Neural Network (ANN) with a rectified linear unit (ReLU) activation function to ease computational cost-effectively and support complex pattern recognition. The second algorithm is RandomForest. Which is based on the bagging method, a statistical technique used in machine learning to reduce variance and improve prediction accuracy by combining multiple models trained on different subsets of the available data. The third algorithm is based on a boosting method called XGBoost. It uses a numerical technique to minimize the model's loss function by adding weak learners using gradient descent. Each of the three potential surrogate models was trained and tested according to a five-fold cross-validation technique. In this work, TensorFlow/Keras was used to develop the surrogate models for software compatibility purposes. After that, the three models were compared and evaluated according to the following four metrics: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R2). Given the results, the artificial neural network was selected as the surrogate model to represent GOSPs in the optimization problem.

The optimization problem is formulated using an open-source Python-based package, Pyomo (Bynum et al., 2021). Pyomo aligns with the notation used in mathematical optimization and adopts an object-oriented approach for modeling elements such as parameters, decision variables, objectives, and constraints. To overcome the complexity of translating the trained machine learning-based surrogate models into Pyomo’s environment, we used a Python-based package called OMLT: Optimization & Machine Learning Toolkit (Ceccon et al., 2022). It provides the flexibility of formulating the activation functions differently based on the application needed.

The optimization problem features two sets of binary decision variables: *Xsd* denotes the selection of swing pipeline originating from source *s* and terminating in destination *d*; the second binary variable is *Xf*, which denotes the selection of facility *f* in the network. Additionally, the optimization problem includes a set of continuous decision variables *Vc,sd* representing the flow rate of component *c* in swing pipeline *sd*. The multi-objective function of the optimization problem is formulated as given below in Eq. (2):

|  |  |
| --- | --- |
|  | (2) |

Where *OPEX* is the total operating expenditure cost of the entire network and calculated by adding the power consumption value of each equipment *ei* in facility *f*. *Emm* is the total emissions generated from the whole network and calculated by adding the emissions produced by power generation activities from equipment *ei* in facility *f* with the surplus gas feed flow rate in facility *f*. The optimization problem is subject to several constraints; however, this paper will highlight the major ones. Eq. (3) demonstrates the first set of constraints related to the design capacities of the facilities and swing pipelines:

|  |  |
| --- | --- |
| m | (3) |

Where is the design capacity of facility *f*, and is the design capacity of swing pipeline *sd*. Eq. (4) defines a constraint that ensures the selection of a single flow direction in bi-directional swing pipelines. Furthermore, Eq. (5) describes the mass balance constraint, which calculates the material balances across each facility in the network.

|  |  |
| --- | --- |
|   | (4) |

|  |  |
| --- | --- |
|  | (5) |

Meanwhile, the flow composition ratios of each component in the swing pipelines were reformulated from a division of continuous variables to a linear expression as given in Eqs. (6):

|  |  |
| --- | --- |
|   | (6) |

Similarly, the product of binary variable with continuous variable in the mass balance constraint results in a nonlinear term that was linearized using the big-M method, as per the following:

|  |  |
| --- | --- |
|  | (7) |

* 1. Results

Following the description of section 3, a large optimization problem is formulated. The problem features 727 constraints, 126 binary variables, and 333 continuous variables. In Pyomo, the IBM CPLEX optimizer was used to solve the model. Accordingly, the proposed methodology has contributed to a total reduction of 6.3 % in OPEX and 10.7 % in total emissions by shifting loading between facilities using swing pipelines. The magnitude of the results depends on the initial planning feed flow rates. However, the solution confirms the results are directionally correct. Figure 3 summarizes the difference between the current and proposed operating scenarios once swing pipelines are utilized.



Figure 3: Difference of objective function values between current and proposed methods

* 1. Conclusion and Outlook

In this work, we proposed a multi-objective MILP optimization framework to maximize profitability and minimize emissions of an integrated network of GOSPs in the oil and gas industry. Embedding machine learning-based surrogates in a multi-objective optimization framework alleviates the drawbacks of alternative methods in the literature. The presented case study demonstrates the framework's capability to yield robust results for the desired objectives. Future work will pursue expanding the formulation into a complete real-world application while exploring alternative multi-objective methods.

* 1. References
1. Al-Ghazal, A. H., He, Y., Al-Hurai, M. A., & White, R. J. (2020). A Rigorous Mixed Integer Nonlinear Programming Model to Optimize the Operations of an Integrated Gas-Oil Separation Network. The Aramco Journal of Technology.
2. Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B., Siirola, J.D., Watson, J.-P., Woodruff, D.L. (2021). Pyomo – Optimization Modeling in Python. Third Edition. Springer.
3. Ceccon, F., Jalving, J., Haddad, J., Thebelt, A., Tsay, C., Laird, C. D., & Misener, R. (2022). OMLT: Optimization & Machine Learning Toolkit. In Journal of Machine Learning Research (Vol. 23).
4. Grossmann, I. E. (2012). Advances in mathematical programming models for enterprise-wide optimization. Computers and Chemical Engineering, 47, 2–18.
5. Grossmann, I. E. (2021). Advanced Optimization for Process Systems Engineering. Cambridge University Press.
6. Liu, S., Alhasan, I., & Papageorgiou, L. G. (2016). A mixed integer linear programming model for the optimal operation of a network of gas oil separation plants. Chemical Engineering Research and Design, 111, 147–160.
7. van de Berg, D., Savage, T., Petsagkourakis, P., Zhang, D., Shah, N., & del Rio-Chanona, E. A. (2022). Data-driven optimization for process systems engineering applications. Chemical Engineering Science, 248.