



OPEN ACCESS

EDITED BY

José Ezequiel Santibañez-Aguilar,
Monterrey Institute of Technology and Higher
Education, Mexico

REVIEWED BY

Ao Yang,
Chongqing University of Science and
Technology, China
Yu Zhuang,
Dalian University of Technology, China

*CORRESPONDENCE

Daniela Galatro,
✉ daniela.galatro@utoronto.ca

RECEIVED 04 September 2024

ACCEPTED 02 January 2025

PUBLISHED 20 January 2025

CITATION

Gupta R, Navas G and Galatro D (2025) Data
analysis-based framework for the design and
assessment of chemical process plants: a case
study in amine gas-treating systems.
Front. Chem. Eng. 7:1490825.
doi: 10.3389/fceng.2025.1490825

COPYRIGHT

© 2025 Gupta, Navas and Galatro. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Data analysis-based framework for the design and assessment of chemical process plants: a case study in amine gas-treating systems

Rahul Gupta¹, Gladys Navas² and Daniela Galatro^{1*}

¹Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada, ²Material Science, IUTFRP/UNETRANS, Caracas, Venezuela

This work presents a process-integrity assessment framework to chemical process design that combines first principles, heuristics, vendor specifications, standards/codes, data analysis, and machine learning modelling, hypothesized as an efficient route for optimal process design. Our case study, a gas treating unit, illustrates its implementation compared with traditional process guidelines. Surrogate models are fitted with hybrid data from process simulation and plant values, supporting the integration between process and integrity values, as well as equipment sizing and cost estimation. Considerable errors are obtained when estimating design duty (1.4%–8.7%) and power requirements (11.1%–33.5%) of the main equipment. Potential sources of these deviations might be attributable to the inherent simplification of process guidelines and intrinsic noise of the plant data used for fitting surrogate models. The process design is then assessed by evaluating process variables and corrosion rate within an operational envelope, showing the synergy and integration of these variables. The benefits and challenges of this approach are drawn while future work in engineering education is presented for its future implementation and effectiveness assessment in enhancing the process design workflow.

KEYWORDS

data analysis, chemical process design, surrogate models, plant integrity, amine gas treating

1 Introduction

Chemical process design aims to develop manufacturing processes that can produce chemicals cost-effectively and safely (Mody and Strong, 2011). At any engineering stage, chemical process design involves creating deliverables such as design basis, process flow diagram, heat and material balances, piping and instrumentation diagrams, equipment sizing, and safety and economic analyses (Sinnott and Towler, 2020; Fengqi, 2024). Process engineering deliverables can vary depending on the project scope, risk, driving factors, and client requirements and are created by integrating first principles, heuristics, and vendor specifications (Radcliffe, 2014; Erwin, 2002). In preliminary design calculations, equipment sizing can heavily rely on heuristics, leading to considerable errors. As heat and material balances are further developed mostly using process simulation tools, the uncertainty around flow rate, temperatures, pressures, and heat duties decreases, and shortcut and robust sizing methods are then employed to enhance equipment sizing. The latest design

stages also include checking and evaluating vendor specifications to “fit” and/or adjusting our equipment sizing against constructible items.

Process simulation is perhaps the main tool used for engineers in chemical process design, as the model-based representation of processes and unit operations provide insights on the heat and material balance, therefore providing information regarding thermophysical properties, properties of the unit operations, kinetics, and environmental and safety-related data (Rhodes, 1996; Gani and Pistikopoulos, 2002). While traditional modelling is mostly based on first principles and some heuristics, the integration of first principles models and machine learning has proven to be an efficient way of representing complex processes and increasing its prediction accuracy, including improvements when predicting the dynamics of the system (Nazemzadeh et al., 2021; Sun et al., 2020; Daoutidis et al., 2024). Thus, machine learning can find potential applications in surrogate modelling for simulation and optimization, process monitoring, and fault operation (Daoutidis et al., 2024), in different areas such as kinetics and material selection, and ultimately, process design (Daoutidis et al., 2024). Industrial cases studies mostly include plant monitoring, integrated decision making, and multiscale modelling (Daoutidis et al., 2024). Chemical engineering, and therefore, process design, is driven by reducing costs and increasing profit, providing unit operations or blocks that can be easily used to design a plant (Sinnott and Towler, 2020), and is motivated by the principles of circular design for circular economy.

In a traditional process design workflow, a design base scenario is simulated at design and/or typical process conditions, and the equipment sizing proceeds considering the heat and mass balance generated from the process simulation while using sizing procedures integrating first principles and heuristics, as provided in textbooks (Sinnott and Towler, 2020; Smith, 2016). In the industry, this is completed using codes, standards, and recommended practices, which, in many cases, are implicitly included in undergraduate textbooks of process design. The design base scenario is then evaluated based on capital and operating costs, safety, flexibility, reliability, maintainability, controllability, and sustainability (Smith, 2023). Hence, plant integrity is crucial when designing chemical process plants to ensure their safety and reliability. Some specific design aspects related to plant integrity include principles to minimize hazards, proper material selection (corrosion-resistant), redundancy and reliability, effective control systems, regular inspections and preventive maintenance actions, and compliance with industry standards and regulations. Therefore, assessing chemical process designs requires evaluating process and integrity variables by considering energy consumption, yield, and efficiency, conducting hazard and risk assessments, assessing reliability and risk of equipment, and assessing environmental impact (Tugnoli et al., 2012). For this assessment, process engineers typically employ process simulation tools. With the advancement of artificial intelligence and, in particular, machine learning can enhance the process-integrity design assessment by looking at optimizing process conditions and energy consumption, predicting equipment failures, and predicting potential safety hazards.

The state-of-the-art chemical process design includes the application of machine learning and artificial intelligence (AI), used for surrogate modeling in simulation and optimization, as

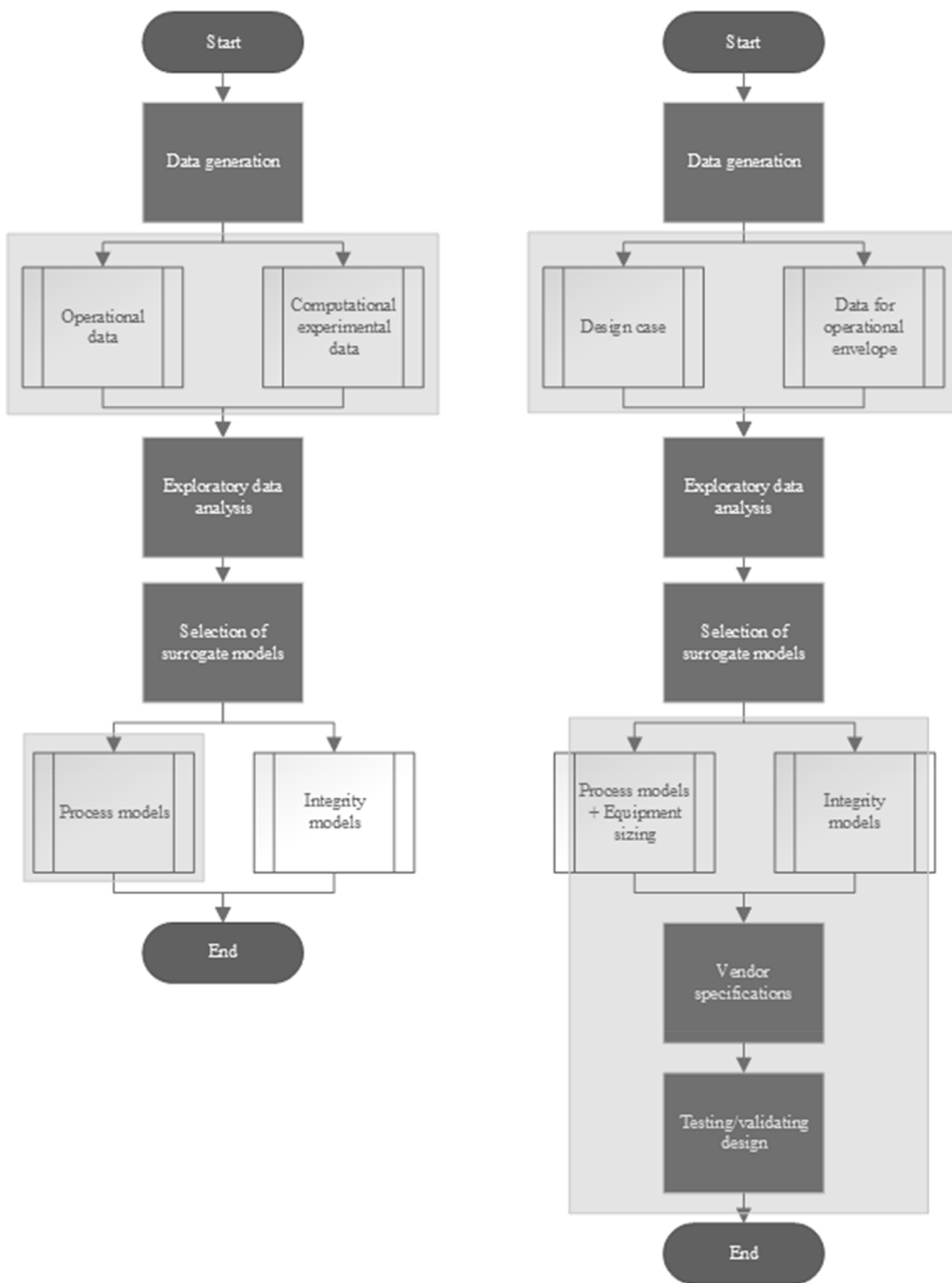
well as process monitoring and fault detection, in addition to contributing to areas such as kinetics and material selection (Mowbray et al., 2022). Moreover, recent versions of process simulation software incorporate AI algorithms for more accurate predictions and optimization of complex processes (Gao et al., 2022). Finally, modern chemical process design focuses on sustainability and circular economy, for which design strategies prioritize reducing energy consumption, minimizing waste and carbon footprint, and using renewable resources, in alignment with global efforts to develop sustainable industries.

In this work, we present a data analysis-based framework that integrates first principles, heuristics, and machine learning when simulating and sizing equipment, aiming at optimizing design drivers and principles in chemical process design. As a proof-of-concept and to validate our approach, our case study (a gas treating unit) includes a process simulation, which was used to create the design base scenario and generate data to build surrogate process and integrity models that convey optimal equipment sizing within a feasible operational envelope. In real-world process design, complex processes and dynamics might not be fully reproduced in process simulation software, from which data-based and intrinsically machine-learning approaches are required for modelling purposes instead. Our framework is also a paradigm shift in the context of engineering education and field process engineering, integrating features that enhance and validate the chemical process design, for which we discuss the benefits and challenges of applying it in the Capstone Project of the undergraduate program of Chemical Engineering.

2 Methodology

2.1 Overview of our data analysis-based framework

Our data analysis-based framework enhances the original work published by (Galatro and Navas, 2022), in which surrogate models for the same case study combining operational and computational experimental data were obtained, complying with the exploratory data analysis and model selection based on statistical metrics. Figure 1 illustrates the workflow of our framework, where differences between the original (Galatro and Navas, 2022) and modified works are indicated in shaded rectangles. This framework was used by (Galatro and Navas, 2022) to predict and assess corrosion damages in existing amine systems. In this work, a process simulation is performed to represent the design base scenario, and “clean” (without noise) data is generated via sensitivity analysis within a feasible operational envelope. The completed data pool is then generated by adding plant data to the clean data. This data is then used to build surrogate models to predict process and integrity-related variables, such as duty, power, and equipment surfaces/areas, fed as input to estimate capital expenditures, equipment dimensions, and corrosion rate in critical plant locations. Equipment dimensions are preliminarily contrasted against vendor specifications. The preliminary process design is then assessed using our surrogate models within a feasible operational envelope to provide insight into the plant operation, as well as predicted plant integrity issues, such as corrosion rates in



(a)

(b)

FIGURE 1 Overview of our data analysis-based framework: **(A)** original framework from reference (Galatro and Navas, 2022) and **(B)** adapted framework for design and assessment of chemical process plants.

critical locations. Surrogate models can also be used as standalone models for monitoring purposes and can be updated with plant data, once in operation.

2.2 Case study

Our framework was implemented in a case study including an amine gas-treating unit. Amine gas-treating involves scrubbing the gas with various alkyl amines, which removes acidic gases from the acid natural gas, thereby generating “sweet natural gas.” The bond between the impurity, such as H₂S, and the amine is very weak. Thus, we can take advantage of this weakness by regenerating the used amine solution in the stripping section and mixing the regenerated amine with the fresh amine to recycle it back to the absorption column. Methyl diethanolamine (MDEA) is an amine used to selectively remove H₂S to pipeline concentrations (<4 ppm in sweet gas). Critical process parameters in MDEA units include pressure to the separator, an absorber, and a regenerator, temperatures in the lean/rich heat exchanger, and regenerator reboiler duty. The Gas Processors Suppliers Association (GPSA) provides process guidelines on how to operate these plants (Galatro and Navas, 2022; Gas Engineering Data Book, 2016) as well as typical equipment sizing data for these units.

Our design base scenario considers:

- Flow rate of acid gas to be processed: 0.28 MSm³/d.
- Amine type and concentration in liquid solution: 45% weight MDEA.
- Acid gas concentration in gas: 7.5% mole (H₂S: 2.5% and CO₂: 5%).
- Inlet pressure: 5,585 kPag.
- Sweet gas specification: 4 ppm H₂S.

2.3 Data generation and surrogate models

A set of surrogate models is created to predict process-corrosion variables and equipment sizing models. Capital expenditure (CAPEX) models for the equipment are taken from the literature (Cooper et al., 2005).

The selection of the surrogate models is made by comparing the prediction performance of different models, for instance, linear/polynomial regression, response surface analysis (RSA) and multivariate adaptive regression spline (MARS). Statistical metrics include individual and overall p-values, F-statistics, relative standard error (RSE), and R².

2.3.1 Process surrogate models

Galatro and Navas (2022) performed a sensitivity analysis using a commercial process simulation software to obtain “clean” process data by changing the values of MDEA concentration, MDEA flow rate, and reboiler duty while recording the effect on the H₂S and CO₂ composition in the sweet gas, acid gas loading (regenerator bottom), temperature, and fluid velocity at the regenerator feed and regenerator bottom streams.

Consequently (Galatro and Navas, 2022), obtained surrogate models (MARS) from the combined simulated/existing plant data to

predict the acid gas loading and fluid temperature. These variables are used as input for the surrogate model used to estimate the corrosion rates in critical streams. In this work, we generated the process data and fitted to surrogate models, as prescribed by (Galatro and Navas, 2022). To reduce the computational time required during the sensitivity analysis in the process simulation software, we have completed clean data using Random Forest, an unsupervised machine learning algorithm useful for regression/prediction purposes (Galatro and Dawe, 2024). This task led to generating updated versions of the existing process surrogate models.

2.3.2 Corrosion surrogate model

Process integrity (PI) in manufacturing includes (i) hazard identification, (ii) material selection, (iii) designing, operating, and maintaining pressure relief and blowdown systems, (iv) plant operation guidelines, (v) establishing safe operating limits, (vi) defining plant layout and area classification, among other topics. In our case study, integrity was assessed by obtaining the corrosion rate in critical streams (feed to regenerator and bottom from regenerator), as suggested by (Galatro and Navas, 2022). Corrosion is an integrity threat that costs \$1.4 billion in the oil and gas production industry (Bhawna and Rakesh, 2023).

Amine corrosion is a form of localized corrosion that occurs mostly on carbon steel and is a function of several factors, including the concentration of the amine solution, the acid gas loading, the heat-stable amine salts, temperature, and velocity (Galatro and Navas, 2022). High lean loadings with a considerable amount of regenerated amine lead to corrosion; In H₂S/CO₂ MDEA systems, acid gas loading is set to a maximum of 0.55 mol/mol (API RP 581, 2016). Higher temperatures, in combination with high rich loading, drastically promote corrosion (API RP 581, 2016). Heat-stable amine salts (HSAS), comprised of amine degradation products, are corrosive; furthermore, they cause corrosion by reducing the amount of active amine available for absorption, resulting in higher acid gas loadings (API RP 581, 2016). Higher fluid velocities contribute to localized corrosion in places such as valves and elbows; for this reason, velocity is limited to 1.5 m/s for rich amine and 6 m/s for lean amine in carbon steel. Galatro and Navas (2022) obtained a neural network-based surrogate model for the corrosion rate in MDEA systems, and it is used in this work. The validity range of their model is for temperatures between 80°C–140°C, a maximum HSAS content of 4.5 weight%, maximum acid gas loading of 0.7 mol/mol, and fluid velocities between 1.0–6.0 m/s, with a maximum error estimated of 4.2% (Gas Engineering Data Book, 2016).

2.3.3 Equipment sizing surrogate models

A sensitivity analysis was performed using process simulation software, varying the input data (as per the design case) to record their effect on equipment duty, surface area, and power, and vessel dimensions within the following ranges:

- Flow rate of acid gas to be processed: 0.15–0.30 MSm³/d.
- Amine flow rate: adjusted to obtain 4 ppm H₂S in the sweet gas.
- Amine concentration in liquid solution: same as design case.

- H₂S/CO₂ composition in the feed gas: 1 to 2.5/1%–10% mole.
- Inlet pressure: 4,500–6,000 kPag.

The decision-making process on which independent variables are required to size each equipment can be based on first principles, heuristics, and/or data. In our case study, the reference for comparison purposes is the GPSA handbook, which presents heuristics-based correlations. Alternatively, in the case of unavailable heuristics or lack of first principles governing equations for a system and/or process, we suggest data-driven approaches involving typical data analytics stages, such as exploratory data analysis, feature-importance, and preliminary interaction between variables, and fitting the data to regression models (using machine learning algorithms).

Several process simulation software solutions are already equipped with routines allowing for equipment sizing; nevertheless, we emphasize the importance of obtaining surrogate models since (i) they are standalone solutions that can serve modularization purposes, (ii) there are equipment which are not embedded in the software database from which dimensions are typically obtained from vendors and/or licensors, and (iii) they can be integrated into optimization algorithms.

2.3.4 Capital expenditure models

In this work, we employed a set of correlations to estimate equipment cost as per (Couper et al., 2005). The equipment CAPEX estimated with these correlations has been adjusted to the current cost using the corresponding Chemical Engineering Plant Cost Index (CEPCI) (Mignard, 2014).

2.4 Process-integrity assessment

A correlogram was used to visually assess the correlation strength and interaction between the variables affecting the corrosion rate in MDEA systems, assuming that the relationship between variables is linear. The “level of importance” of these correlations and interactions allowed for selecting the critical process streams where the corrosion rate shall be estimated. Thus, the process-integrity assessment was performed on the streams “Regen Feed” (feed stream to the regenerator) and “Regen bottom” (bottom stream from the regenerator) by estimating the corrosion rate as a function of the temperature, HSAS content, acid gas loading, and fluid velocity, as suggested by (Galatro and Navas, 2022). The acid gas loading and temperature are estimated using the corresponding process surrogate models. Either when designing or operating the plant, the corrosion rate allows monitoring, for instance, (i) regeneration efficiency, which is affected by the acid gas loading, (ii) amine degradation, which is impacted by the HSAS content, and (iii) localized corrosion, which is caused by high fluid velocities (>6 m/s).

2.5 Calculation tools

Aspen Hysys[®] was used as process simulation software. All calculations in this work were performed in R, a robust free software typically used in data analytics.

3 Results

3.1 Process and corrosion surrogate models

Figure 2A shows a simplified correlogram for some of the process design-related variables. Correlograms are used for a preliminary assessment of the correlation strength and interaction between process design variables to facilitate the dimensionality reduction for the equipment-sizing surrogate models.

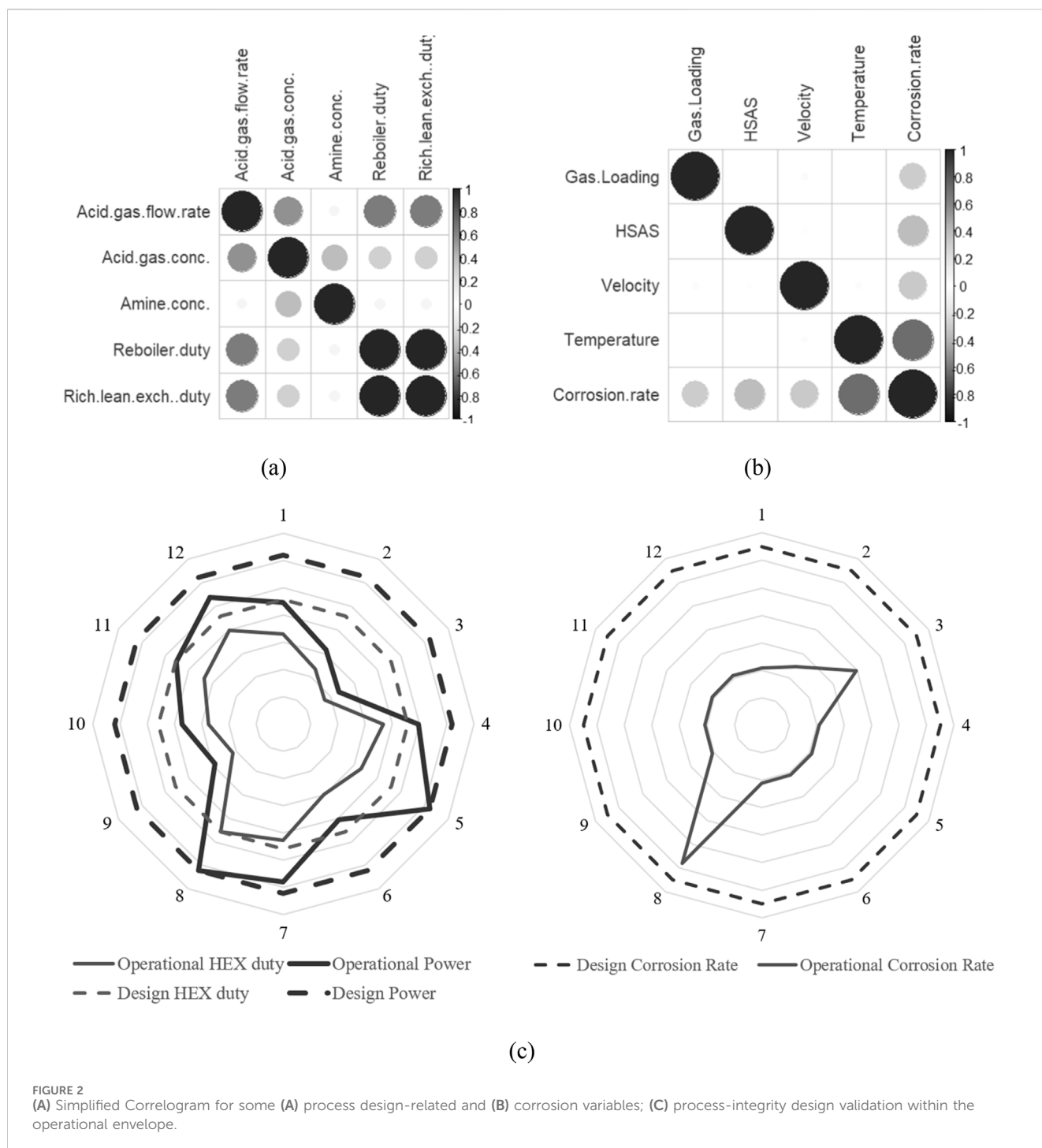
Design-related variables, such as the reboiler duty, strongly depend on the acid gas flow rate, followed by the acid gas concentration, and weakly on the amine concentration. In our case study, we can confirm that our data-based models considered the required independent variables, matching most of the strength/weak correlations shown in several correlograms. Similarly to the GPSA, most of the equipment sizing models are linear and dependent on one variable, except for the pumps power and the amine flow rate, which depends on the acid gas concentration, acid gas flow rate, and amine concentration in liquid solution. These models exhibit R² close to 1, with RSEs less than 0.001. The corrosion rate model was taken from (Galatro and Navas, 2022) and is a neural network-based model with R² = 0.991 and RSE = 0.02. The updated acid gas loading and temperature models, which fitted the hybrid data to MARS and RSA models, respectively, report R² close to 1 and RSEs of 0.0001 and 0.026, respectively.

Figure 2B illustrates a correlogram relating the variables affecting the corrosion rate. The circles are proportional to the correlation coefficients, showing positive correlations for all variables; this proportionality allows us to rank the importance of the variables affecting the corrosion rate, starting with the temperature, followed by the HSAS content, acid gas loading, and fluid velocity. Assuming an HSAS content constant in all process streams, looking at their remaining variables in the heat and material balance generated by the process simulation, and estimating the corresponding acid gas loading as per the fitted surrogate model, it is possible to select the highest expected corrosion rates of the plant in the streams “Feed Regen” and “Bottom regen,” as describe by (Galatro and Navas, 2022).

3.2 Design base scenario

The main results regarding the design base scenario are shown in Table 1.

The total duty comprising the reboiler, rich-lean amine heat exchanger, amine cooler, and reflux condenser is 4,581 kW, totaling 586.2 K\$. Regarding power requirements comprising main amine, amine booster, and reflux pumps, the total is 62.1 kW, costing 119.1 K\$. Finally, the contactor, surge tank, reflux accumulator, and regenerator dimensions were estimated to cost 155.7 K\$. A comparison between our surrogate models and the GPSA correlations reveals errors of 3.0% (duty), –29.2% (power), and 4.2% (total CAPEX, which included a design factor of +30%) in average. Duty errors lie between 1.4% and 8.7%, while power requirements were estimated between +11.1 and



+33.5%. Potential sources of deviation for these requirements might be mostly due to the nature of the GPSA estimations, which are considered simplified for conventional monoethanolamine (MEA) and diethanolamine (DEA) amine gas units (Gas Engineering Data Book, 2016). Moreover, intrinsic noise in the plant data might contribute to these deviations. The obtained equipment sizing surrogate models were linear and/or polynomial, like the models reported in GPSA (Gas Engineering Data Book, 2016), with R^2 of

0.86–0.95; dimensions for vessels obtained with our surrogate models (contactor, surge tank, reflux accumulator, and regenerator) are the same as per the GPSA guidelines, as standard equipment sizing is expected. The amine flow rate via surrogate model obtained from simulated data was estimated to be 92% of the calculated using the GPSA guidelines, which might be attributed to intrinsic noise in the plant data and/or intrinsic differences derived from the thermodynamic method used for the process simulation software.

TABLE 1 Equipment sizing for the design base scenario.

Heat exchangers						
	GPSA Duty, kW	Duty, kW	% Deviation	Area, m ²	Duty, Btu/h	Cost, K\$
Reboiler (direct fired)	2098.5	1969.3	6.2	96.1	6047452.8	131.9
Rich-lean amine HEX	1,308.8	1,271.1	2.9	95.5	3903495.0	48.9
Amine cooler (air cooled)	435.5	473.3	8.7	86.8	2422551.2	379.4
Reflux condenser	880.0	867.3	1.4	44.2	2959479.2	26.0
		4,581.0				586.2
Power requirements						
	GPSA Power, kW	Power, kW	% Deviation	Area, m ²	Duty, Btu/h	Cost, K\$
Main amine pumps, kW	39.1	52.2	33.5			1.9 (2 pumps)
Amine booster pumps, kW	4.5	5.0	11.1			0.2 (2 pumps)
Reflux pumps, kW	4.5	5.0	11.1			0.2 (2 pumps)
		62.1				119.1
Contactor						
Still diameter, mm		750		636.0		
Height, mm		2,250				
Weight, lb		186.2				31.2
Surge Tank						
Diameter, mm		1,524				
Length, mm		4,877				
Weight, lb		794.2				50.7
Reflux Accumulator						
Diameter, mm		1,067				
Length, mm		2,438				
Weight, lb		303.3				35.6
Regenerator						
Still diameter, mm		1,067				
Height, mm		3,201				
Weight, lb		372.7				38.2
						155.7
						1,119.3

3.3 Design validation within the operational envelope

Table 2 presents a series of scenarios defined for design validation and visualization purposes. Acid gas flow rate to be processed, amine concentration in liquid solution and H₂S/CO₂ composition are varied to evaluate the impact on the total duty and power of the plant. These scenarios are contrasted against their corresponding design values (4,581 kW and 62.1 k, respectively) to

validate that expected values within the operational envelope can be “handled” by the design.

Table 2 supports evaluating how the design from Table 1 holds up under different process conditions; multiple scenarios provide insights into the flexibility and robustness of the chemical process design. Moreover, Table 2 supports identifying how changes in process parameters, such as H₂S concentration or reboiler temperature, affect performance indicators, such as equipment duty, size, or power requirements, allowing for assessing whether

TABLE 2 Process scenarios for design validation.

Scenario	Acid gas to be processed, MSm ³ /d	Amine concentration in liquid solution, mass%	H ₂ S, % mol (inlet)	CO ₂ , % mol (inlet)	Reboiler temperature (bottom), °C
1	0.283	50	1.0	5.0	119.4
2	0.200	50	1.0	5.0	113.5
3	0.150	50	1.0	5.0	110.0
4	0.283	45	1.0	5.0	121.5
5	0.283	50	1.0	5.0	119.3
6	0.283	55	1.0	5.0	117.5
7	0.283	45	2.0	5.0	125.1
8	0.283	45	2.5	5.0	127.0
9	0.283	45	2.5	1.0	112.3
10	0.283	45	2.5	2.0	116.0
11	0.283	45	2.5	3.0	119.6
12	0.283	45	2.5	4.0	123.3

equipment are adequately sized to handle different operational/process conditions. Finally, Table 2 might support fine-tuning decisions leading to process optimization when comparing expected results with actual process behaviour.

Figure 2C shows the process design validation within the operational envelope for twelve random scenarios (this is only for visualization purposes, as our approach allows for several iterations, hence, theoretically an infinite number of operational scenarios).

The dashed contours depict the design duty and power, while the solid contours represent the maximum values of duty and power for each scenario, which never cross the dashed boundaries. Figure 2B shows the integrity design validation within the operational envelope, with operational corrosion rates never crossing the boundaries of the maximum corrosion rate of 1.3 mm/y, set at a maximum gas recommended acid loading in rich amine (mol/mol) of 0.55 mol/mol (as per GPSA), average recommended reboiler temperature of 116°C (as per GPSA), and average HSAS of 2%.

3.4 Importance of our framework in chemical process design and engineering education

Our process-integrity design assessment approach looks at optimizing process conditions and energy consumption while potentially predicting safety hazards and equipment failures. The synergy between first principles (via process simulation), heuristics, data analysis, and complementary modelling/simulation using machine learning tools allows for assessing several operating scenarios. In our case study, for instance, we quantified considerable differences when estimating duty and power requirements compared to rough estimations, while we also assessed the integrity of our design in key sections of our plant using a machine learning model integrated into our process

models, when estimating corrosion rate at several operating conditions.

Therefore, we foresee the use of our framework as an event-based approach where process/plant design is seen dynamically. Field engineers and chemical engineering students could test the design for different operational scenarios (including off-specs/failure-inducing ones) and study, through data-based causation, the interaction between variables when analyzing these scenarios (refer to the correlogram in a previous section of this paper).

Some of the challenges that we identified for future implementation of our framework in field process engineering include (i) defining/scoping feasible scenarios might require a solid understanding of the process, and potentially be resource-consuming to execute; and (ii) requiring multidisciplinary expertise as some integrity-reliability aspects might not be easily identifiable by process engineers but experts in areas such as corrosion, failure analysis, etc. In our classrooms, we identified challenges such as (i) defining fair rubrics to reflect the students' understanding of the process design, and (ii) assessing the effectiveness of implementing this framework, as instructors might require surveying its impact on the workplace. Future work will examine implementing this framework for the capstone project in our undergraduate chemical engineering curriculum and longitudinally assessing its effectiveness in enhancing the design workflow process of students.

4 Conclusion

In this work, we presented a process-integrity assessment framework for chemical process design that combines first principles, heuristics, vendor specifications, standards/codes, data analysis, and machine learning (ML) modelling. A case study on a gas treating unit was used to implement and assess our framework. A set of surrogate models was fitted from process simulation data and plant data for process, integrity, equipment sizing, and equipment cost variables that allow for designing the plant, as

well as assessing the design based on several conditions within an operational envelope. Equipment power requirements showed considerable differences when compared to values obtained from traditional process guidelines, with errors up to 33.5%. While we perceive notable integrating and synergy benefits of applying hybrid frameworks, including ML modelling, we foresee implementation challenges associated with scoping scenarios, gathering data, and combined process/integrity expertise. Finally, we also present insights into future work for its implementation and effectiveness assessment in engineering education, looking at enhancing the design workflow process of students.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RG: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing—original draft. GN: Conceptualization, Formal Analysis, Investigation, Methodology, Writing—review and editing. DG:

Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Supervision, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- API RP 581 (2016). Risk-based inspection methodology.
- Bhawna, B., and Rakesh, K. (2023). An overview of corrosion in the oil and gas industry: challenges and solutions. *i-manager's J. Material Sci.* 10 (4), 56. doi:10.26634/jms.10.4.19455
- Cooper, J., Penney, W. R., Fair, J. R., and Walas, S. M. (2005). *Chemical process equipment*. Second. Elsevier.
- Couper, J. R., Penney, W. R., Fair, J. R., and Walas, S. M. (2005). *Chemical process equipment*. Second Edition. United States: Gulf Professional Publishing.
- Daoutidis, P., Lee, J. H., Rangarajan, S., Chiang, L., Gopaluni, B., Schweidtmann, A. M., et al. (2024). Machine learning in process systems engineering: challenges and opportunities. *Comput. Chem. Eng.* 181 (Feb), 108523. doi:10.1016/j.compchemeng.2023.108523
- Erwin, D. (2002). *Industrial chemical process design*. McGraw-Hill.
- Fengqi, Y. (2024). "Northwestern university chemical process design open textbook."
- Galatro, D., and Dawe, S. (2024). *Data analytics for process engineers: prediction, control and optimization*. Springer Nature.
- Galatro, D., and Navas, G. (2022). "Framework for the prediction and assessment of corrosion damages in amine systems using plant data, process simulation and data analysis," in *SPE western regional meeting*. doi:10.2118/209330-MS
- Gani, R., and Pistikopoulos, E. N. (2002). Property modelling and simulation for product and process design. *Fluid Phase Equilib.* 194-197, 43-59. doi:10.1016/s0378-3812(01)00680-x
- Gao, H., Zhu, L. T., Luo, Z. H., Fraga, M. A., and Hsing, I. M. (2022). Machine learning and data science in chemical engineering. *Am. Chem. Soc.* 61, 8357-8358. doi:10.1021/acs.iecr.2c01788
- Gas engineering data Book. *Gas Processors Suppliers association*, (2016).
- Mignard, D. (2014). Correlating the chemical engineering plant cost index with macro-economic indicators. *Chem. Eng. Res. Des.* 92 (2), 285-294. doi:10.1016/j.cherd.2013.07.022
- Mody, D., and Strong, D. S. (2011). "An overview of chemical process design engineering," in *Proceedings of the Canadian engineering education association (CEEA)*. doi:10.24908/pceea.v0i0.3824
- Mowbray, M., Vallerio, M., Perez-Galvan, C., Zhang, D., Del Rio Chanona, A., and Navarro-Brull, F. J. (2022). *Industrial data science - a review of machine learning applications for chemical and process industries*. United States: Royal Society of Chemistry. doi:10.1039/d1re00541c
- Nazemzadeh, N., Malanca, A. A., Nielsen, R. F., Gernaey, K. V., Andersson, M. P., and Mansouri, S. S. (2021). Integration of first-principle models and machine learning in a modeling framework: an application to flocculation. *Chem. Eng. Sci.* 245 (Dec), 116864. doi:10.1016/j.ces.2021.116864
- Radcliffe, D. (2014). *Multiple perspectives on engineering design*. United States: Purdue University Press.
- Rhodes, C. L. (1996). The process simulation revolution: thermophysical property needs and concerns. *J. Chem. Eng. Data* 41, 947-950. doi:10.1021/jc960029b
- Available at: <https://pubs.acs.org/sharingguidelines>.
- Sinnott, R., and Towler, G. (2020). United States: Chemical Engineering Design.
- Smith, R. (2016). *Chemical process design and integration*. Second. John Wiley and Sons.
- Smith, R. (2023). *Process plant design*. Wiley.
- Sun, B., Yang, C., Wang, Y., Gui, W., Craig, I., and Olivier, L. (2020). A comprehensive hybrid first principles/machine learning modeling framework for complex industrial processes. *J. Process Control* 86, 30-43. doi:10.1016/j.procont.2019.11.012
- Tugnoli, A., Landucci, G., Salzano, E., and Cozzani, V. (2012). Supporting the selection of process and plant design options by Inherent Safety KPIs. *J. Loss Prev. Process Ind.* 25 (5), 830-842. doi:10.1016/j.jlp.2012.03.008