CHEMICAL ENGINEERING TRANSACTIONS
VOL. 118, 2025

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Laura Piazza, Francesco Donsì, Giorgia Spigno, Paolo Ciambelli, Luca di Palma

ISBN 979-12-81206-19-9; **ISSN** 2283-9216

DOI: 10.3303/CET25118066

Novel Gas-Oil Separation Plant Design in Mediterranean and Synthetic Data Generation for Machine learning Model

Ahmed A. Elsheemy[a,*], Mamdouh A. Gadalla[b,c], Sahar M. El-Marsafy[a], Fatma H. Ashour[a]

[a]Chemical engineering department, Faculty of Engineering, Cairo University, Egypt
[b]Chemical engineering department, Faculty of Engineering,Portsaid University, Egypt
[c]Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa
ahmedashraf_1989@hotmail.com

This study presents the development of a Gas Oil Separation Plant (GOSP) simulation using ASPEN HYSYS and the Peng Robinson equation of state to model the thermodynamic behavior of hydrocarbon mixtures. The design was tested under a range of feed conditions, including varying gas and condensate flow rates, to improve separation and processing efficiency. Field data from the Mediterranean concession, including reservoir properties, well characteristics, and fluid compositions, formed the basis of the model. Key parameters such as temperature, pressure, gas to oil ratio, and the mole fractions of methane and ethane were considered. Fluid characterization was carried out using PVT modeling software by Schlumberger. A Python script was developed to automate HYSYS simulations via the "win32com" interface, generating a synthetic dataset of 1,700 operational scenarios. Three experimental cases explored the effects of separator pressures, compressor limitations, and Reid Vapor Pressure on system performance. The resulting synthetic data achieved over 91 percent agreement with simulation outputs using linear regression proxy model, while terminal gas heating value reached 9691 kcal per cubic meter and condensate Reid Vapor Pressure remained within the 82.7 kPa limit. These outcomes confirm the model's accuracy and its potential for supporting design optimization and operational decision-making.

1. Introduction

Natural gas consists primarily of methane (CH_4), along with ethane, propane, butane, and heavier hydrocarbons (C_{5+}), collectively known as natural gas liquids (NGLs, C_{2+}). Impurities such as CO_2, N_2, H_2S, and water vapor must be removed to meet commercial specifications. Given their high economic value, NGLs are typically recovered through structured gas processing systems. These include slug catchers, high-pressure separators, and downstream treatment units: Gas Sweetening (GSU), Sulfur Recovery (SRU), Gas Dehydration (GDU), Mercaptans Removal (MRU), and mercury removal stages, ensuring safety in cryogenic operations (Mokhatab et al., 2006; Elsheemy et al., 2018; Abdel-Aal et al., 2003; Kidnay et al., 2006; GPSA, 2004; Lynch et al., 2007; Sinnott, 1999; Ghorbani et al., 2016; Al-Dogail et al., 2023). Dew Point Control Units (DPCUs) recover NGLs, while final compression prepares the gas for transport. Condensates and water are stabilized in the Condensate Stabilization Unit (CSU), ensuring product readiness (Mokhatab et al., 2006; Abdel-Aal et al., 2003; Kidnay et al., 2006; GPSA, 2004; Lynch et al., 2007; Sinnott, 1999; Ghorbani et al., 2016; Al-Dogail et al., 2023).
Mediterranean concession in the Mediterranean poses specific design and operational challenges. Traditional experimental methods for GOSP development are often costly and time-intensive. To address this, the present study introduces an advanced methodology that integrates ASPEN HYSYS process simulation with synthetic data generation for enhanced Gas-Oil Separation Plant (GOSP) design and optimization (Bartolome and Van Gerven, 2022; Rovira Herrero, 2022; Noh et al., 2025). The approach combines thermodynamic modeling with automated data synthesis to support machine learning applications, enabling more adaptive and efficient plant performance tuning.

Paper Received: 13 May 2025; Revised: 14 July 2025; Accepted: 09 September 2025
Please cite this article as: Elsheemy A.A., Gadalla M.A., El-Marsafy S.M., Ashour F.H., 2025, Novel Gas-oil Separation Plant Design in Mediterranean and Synthetic Data Generation for Machine Learning Model, Chemical Engineering Transactions, 118, 391-396 DOI:10.3303/CET25118066

2. Design of the GOSP Using ASPEN HYSYS

GOSP simulation was conducted in ASPEN HYSYS using the Peng-Robinson equation of state to represent the thermodynamic behavior of hydrocarbon mixtures. Design scenarios incorporated varying gas and condensate flow rates to achieve optimal separation across operating conditions. Model input was based on data from the Mediterranean concession, including reservoir characteristics (Table 1), well profiles (Table 2), fluid compositions (Tables 3 and 4), and feed conditions (Table 5). Key parameters such as temperature, pressure, gas-oil ratio, and molar fractions of methane, ethane, and heavier hydrocarbons were essential to reflect the field's unique feed properties

2.1 Reservoir Properties and Well Information

Table 1: Reservoir Properties

Property	WEB	Papyrus
Reservoir Fluid Type	Gas	Gas
Reservoir Temperature (°C)	73 @ 2087m TVDss	66 @ 1760m TVDss
Initial Reservoir Pressure (kPa)	33900	28200
CGR ($m^3/MMsm^3$)	0.128	0.07
Wellhead Shut-in Pressure (kPa)	28400	23900
CO_2 (mol%)	0.108 (dry)	0.099 (dry)
H_2S (ppm)	0	0

Reservoir properties for the WEB and Papyrus prospects define the fluid behavior and operational conditions influencing GOSP design. Differences in temperature, pressure, and gas-to-liquid ratios between the sites necessitate distinct processing strategies. The WEB platform includes more wells, reflecting higher feed volumes and capacity requirements. Compositional analyses by Schlumberger (2009, 2011) and further characterization using PVT modeling (Tables 3 and 4) provided critical input. Accurate fluid compositions and pseudo-component data were essential for reliable GOSP simulation and separation optimization.

Table 3: Fluid Composition of WEB and Papyrus Fluids

Component	Mole% (WEB Fluid)	Mole% (Papyrus Fluid)
N_2	0.276	0.040
CO_2	0.108	0.099
C_1 (Methane)	90.134	90.002
C_2 (Ethane)	4.693	5.049
C_3 (Propane)	1.648	1.916
iC_4	0.464	0.589
nC_4	0.387	0.508
iC_5	0.213	0.303
nC_5	0.143	0.201
nC_6	0.276	0.314
nC_7	0.355	0.279
nC_8	0.356	0.241
C_{9+} (Total)	1.247	0.959

Table 4: Pseudo-Component Properties

Fluid	Component	Normal Boiling Point (°C)	Molecular Weight (g/mol)	Liquid Density (kg/m³)
WEB	C_{16}-C_{17}	281.0	230.4	846.7
WEB	C_{18}-C_{19}	305.9	254.3	850.8
WEB	C_{20}-C_{51}	368.0	316.5	878.8
Papyrus	C_{30+}	508.0	484.0	925.0

2.2 Development of the GOSP Simulation Model

The ASPEN HYSYS simulation incorporated all core GOSP process units to ensure efficient separation of gas, oil, and water from mixed feeds. Design benchmarks: gas flow, condensate, and water production are detailed

in Table 5 and define operational capacity. The facility transitions from an early-life mode with higher pressures to a late-life mode reflecting reservoir depletion, accommodating changes across the plant's operational lifecycle.
As shown in Figure 1, The work procedure of gas processing initiates the separation of the feed fluid within the slug catcher, where the major split occurs between the bulk gas and condensate. Separated gas exiting the slug catcher passes through a high-pressure scrubber to remove entrained liquids, ensuring a clean gas stream. It then enters the tri-ethylene glycol (TEG) dehydration unit to remove water completely.

Table 5. Design Criteria

Design Criteria	Value
Gas (kgmol/hr)	6226
Condensate (m^3/hr)	16.56
Water (m^3/hr)	0.24
Onshore Arrival Pressure (Early Life) (kPa)	6000
Onshore Arrival Pressure (Late Life) (kPa)	2200

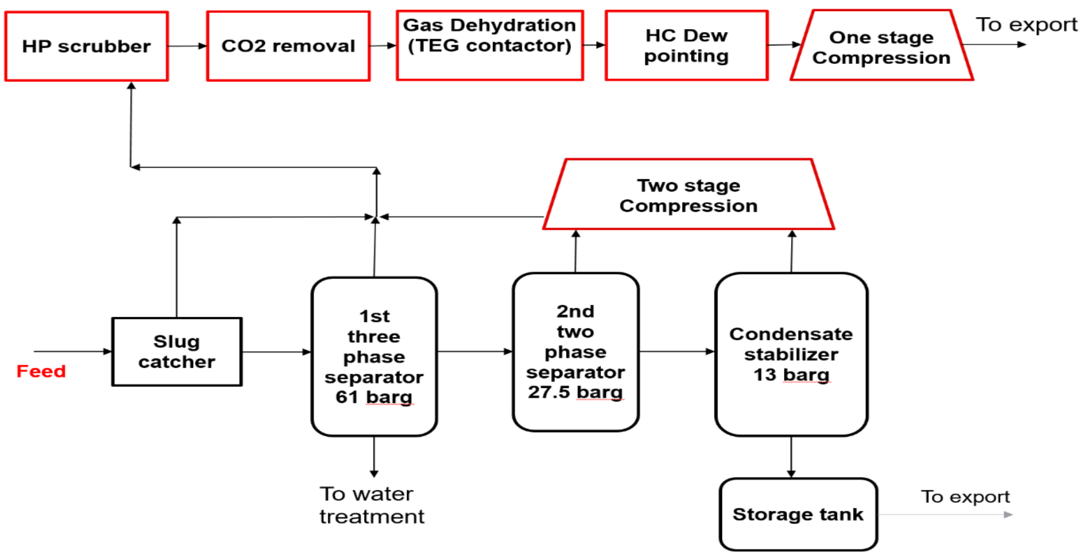

Figure 1: GOSP Process Flow Diagram

Following dehydration, the gas undergoes dew-point control through a self-refrigeration process using a Joule-Thomson (J-T) valve. This cools the gas and condenses C_{3+} components, which are separated in a low-temperature separator. The resulting dry gas, free of heavier hydrocarbons, is compressed via a gas compression unit for export. As the CO_2 concentration is below the export limit, a CO_2 removal unit is not required, simplifying the process and reducing costs (Mokhatab et al., 2006; Elsheemy et al., 2018; Abdel-Aal et al., 2003). While, the liquid collected in the slug catcher undergoes two stages of separation to remove lighter fractions. The remaining condensate is sent to the Condensate Stabilizer Unit, where it is treated to meet export specifications, ensuring the Reid Vapor Pressure (RVP) remains below 82.7 kPa (Kidnay et al., 2006). The stabilized condensate is then stored at standard conditions in dedicated storage tanks, ready for export or further use. Off-gas from the stabilizer is compressed in a two-stage compressor and reintroduced into the main gas stream. This integrated approach ensures efficient and safe operation of the GOSP, maximizing hydrocarbon recovery and meeting product quality standards (Mokhatab et al., 2006; Elsheemy et al., 2018; Abdel-Aal et al., 2003; Kidnay et al., 2006).

3. Synthetic Data Generation for Machine Learning Development

Synthetic data was generated to optimize plant performance using machine learning models. A Python script, integrated with ASPEN HYSYS via the "win32com" library, enabled direct manipulation and extraction of simulation data for multiple operational scenarios (Bartolome and Van Gerven, 2022; Rovira Herrero, 2022; Noh et al., 2025). Through the COM interface, the script accessed HYSYS case files to retrieve key input variables such as feed flow rate, pressure, and temperature, and the output variable, gas production (Goh et al., 2024; Ahmad, 2024; Zhu et al., 2025). Extracted data, including molar flow rates, pressures, and temperatures, were organized into a structured format such as a Pandas DataFrame to support systematic simulation analysis and parameter adjustments (Galeazzi et al., 2023).

3.1 Constraint establishment

Synthetic data generation in the HYSYS simulations followed strict operational constraints to reflect practical field conditions. Separator pressures were set between 2500 and 5000 kPa, with a minimum stabilizer pressure of 100 kPa to ensure effective feed separation. Compressor K-100 and K-102 operated under a 4:1 discharge-to-inlet pressure ratio to maintain system performance. Reid Vapor Pressure (RVP) for condensate was confined to 62–82.7 kPa, aligning with safety and export compliance standards (Bartolome and Van Gerven, 2022; Rovira Herrero, 2022; Noh et al., 2025).

3.2 Generation of synthetic data set

A synthetic dataset was developed through three experiments designed to explore GOSP operational parameters. Experiment One used random variation in feed pressure, temperature, flow rate, and separator pressures across 650 runs. Experiment Two systematically varied feed flow while randomly adjusting other inputs, generating 1,050 runs. Experiment Three combined both datasets into 1,700 runs, enhancing statistical robustness. After removing outliers and noise, key variables were selected and feature engineering was applied to improve the dataset's predictive quality.

3.3 Feature engineering for enhanced predictive modeling

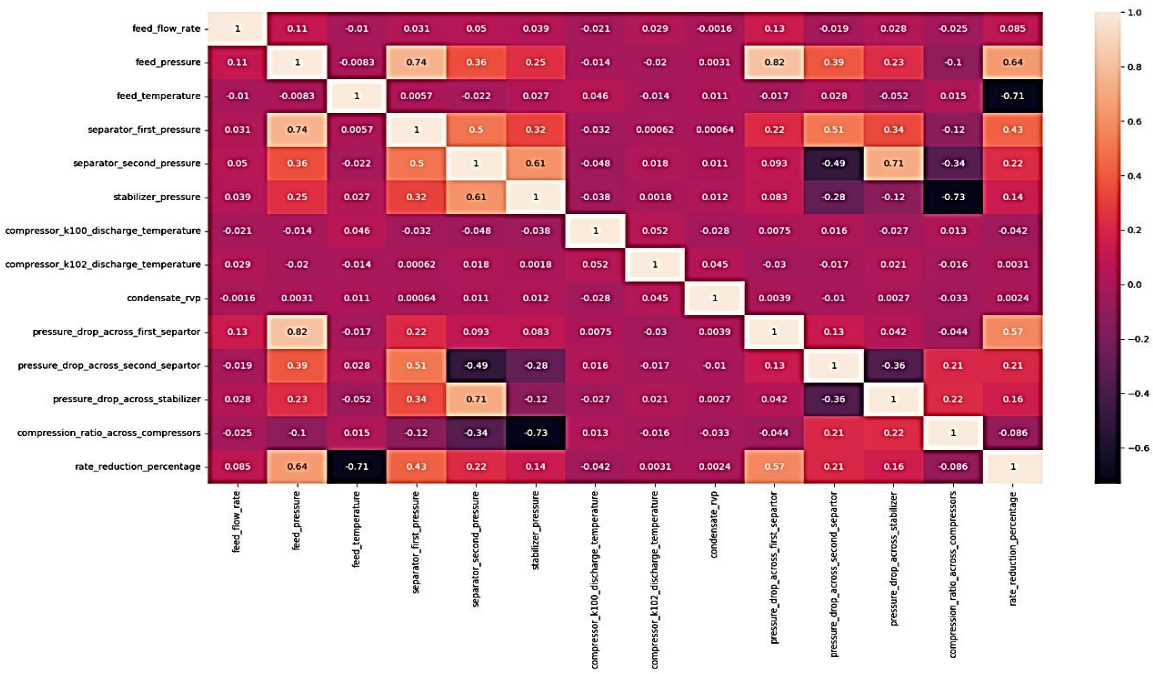

Figure 2: Experiment Three Heat map after cleaning data

This study enhances the reliability and accuracy of predictive models for Gas Oil Separation Plant (GOSP) performance by engineering key operational features. These include compression ratios across compressors K-100 and K-102, pressure differentials across sequential separators and the stabilizer, and the percentage flow rate loss from feed to gas production. These features capture critical system dynamics affecting separation efficiency, compression performance, and operational stability. Their integration enables machine learning models to uncover nonlinear relationships between input conditions and output flow behavior, improving model accuracy.

3.4 Experimental analysis and data insights

Three experiments evaluated GOSP performance under varied operating conditions. Experiment One used and rate reduction) but high variability. Experiment Two applied structured variations in feed flow (1,050 runs), yielding stronger relationships (e.g., 0.82 between feed pressure and separator pressure drop, -0.71 with temperature), improving insight clarity. Experiment Three, combining both datasets (1,700 runs), enhanced model accuracy and robustness, confirming key trends and supporting future proxy model development.

4. Results and Discussion

4.1 HYSYS simulation

The HYSYS simulation confirmed that the terminal gas and condensate met or exceeded industry specifications. As shown in Table 6, the gas stream processed at 91°C and 5400 kPa had a gross heating value of 9691 kcal/m³, surpassing the typical minimum of 9255 kcal/m³. It also exhibited low water content (0.025 kg/m³) and a hydrocarbon dew point of -22.7°C, ensuring dehydration efficiency and thermal stability for pipeline transport. According to Table 6, the condensate, stabilized at 44°C and 12 kPa, displayed high quality with a molecular weight of 119, API gravity of 67, and a Reid Vapor Pressure of 68.9 kPa within the 82.7 kPa specification. BS&W and salt content were 0.15 wt% and 70 kg/km³, respectively, confirming product readiness for export. These outcomes validate the effectiveness of the GOSP's separation and stabilization design.

Table 6: HYSYS Terminal Gas Vs National Grid Specs and HYSYS Terminal Condensate Properties

HYSYS Terminal Gas Properties	Value	National Grid Specs	HYSYS Terminal Condensate Properties	Value
Temperature, °C	91	No spec defined	Temperature, °C	44
Pressure, kPa	5400	No spec defined	Pressure, kPa	12
Gross Heating Value, kcal/m³	9691	≥ 9255 kcal/m³	Molecular Weight	119
Molecular Weight	17.72	No spec defined	API	67
Water Content (Gas), kg/m³	0.0252	≤ 0.06 – 0.11 kg/m³	Reid VP at 37.8 C (kPa)	68.9
HC Dew Point (Gas), °C	-22.7	≤ -10 to 4 °C		

4.2 Suitability for Proxy Models

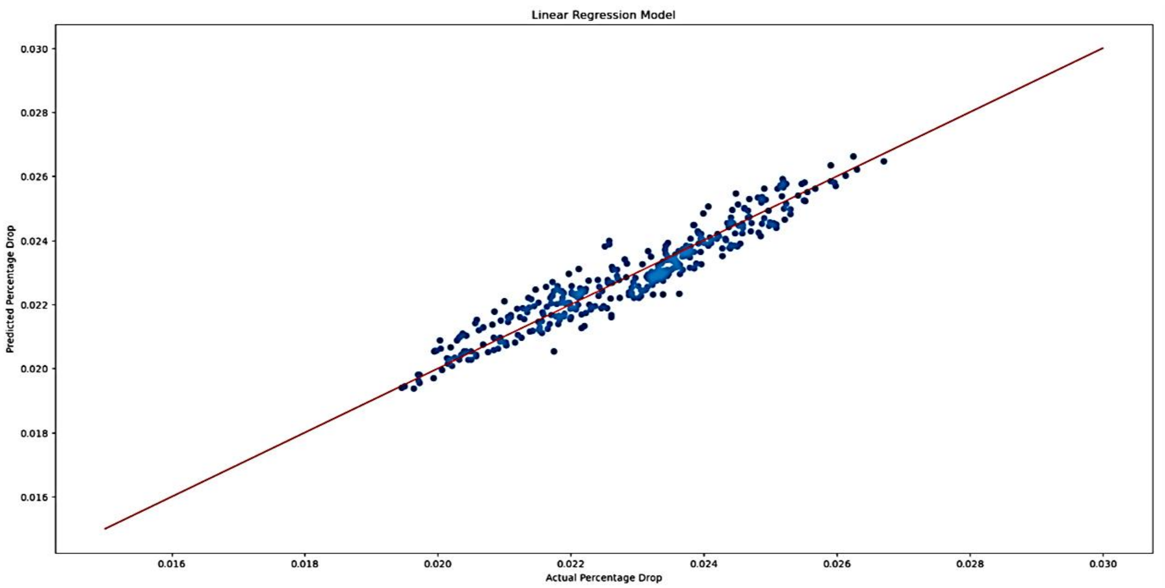

Figure 3: A scatter plot to visualize the performance of Linear Regression Model

Experiment Three provided the most representative dataset for proxy model development, covering diverse operating scenarios with minimal randomness. Preprocessing involved separating features from the target (rate reduction percentage), followed by stratified splitting into training (70%) and test/validation (30%) subsets and standardizing input variables (mean = 0, SD = 1) to enhance model performance. Linear regression achieved R^2 values between 0.913 and 0.926, capturing over 91% of output variance. Figure 3 presents a scatter plot of predicted vs. actual values from the linear regression model, key predictors included feed temperature

(0.001152), pressure drop across the first separator (0.000486), and feed pressure (0.000431). The resulting model is:

$$Rate\ reduction\ percentage\ across\ GOSP\ \approx\ 0.022732 + (0.001152 \times Feed\ Temperature)\ + (0.000486 \times Pressure\ Drop\ Across\ First\ Separator)\ + (0.000431 \times\ Feed\ Pressure) \quad (1)$$

This equation identifies the primary operational factors driving performance variability, offering a foundation for process optimization.

5. Conclusion

This study introduces a practical approach for designing a Gas Oil Separation Plant (GOSP) using ASPEN HYSYS and the Peng Robinson equation of state to accurately simulate the behavior of hydrocarbon mixtures under various feed conditions. The model, developed using data from the Mediterranean concession (including reservoir characteristics, well details, and fluid composition) confirmed that the final gas and condensate properties meet industry standards. Critical design variables included temperature, pressure, and the gas to oil ratio. A key advancement was the creation of synthetic datasets through a Python-based interface using the "win32com" library, allowing automated data extraction from HYSYS simulations. Three sets of simulations tested a range of operating limits for separator pressures, compressor conditions, and Reid Vapor Pressure. Validation using linear regression showed a strong correlation (R^2 above 0.91), confirming the reliability of the generated data for simulating plant performance and supporting future use in data-driven optimization models. Future work may explore nonlinear models or real-time optimization based on field deployment.

References

Abdel-Aal H.K., Aggour M.A., Fahim M.A., 2003, Petroleum and Gas Field Processing, Marcel Dekker, New York, USA.

Ahmad S., 2024, Sustainable Hydrogen Production from Sour Gas: Integrating Machine Learning for Process Optimization and Prediction, PhD diss., Khalifa University of Science.

AL-Dogail A., Gajbhiye R., Al-Shammari H., Alnaser M., Kamerkar T., 2023, Maximization of Gas-Oil Separation Plant Oil Recovery by Operation Parameter Optimization, SPE Production & Operations, 38(04), 666–677.

Bartolome P.S., Van Gerven T., 2022, A comparative study on Aspen Hysys interconnection methodologies, Computers & Chemical Engineering, 162, 107785.

Elsheemy A.A., Ashour F.H., Gadalla M.A., 2018, Maximization of Condensate Production by Revamping of Gas-Oil Separation Plant in Gulf of Suez, Chemical Engineering Transactions, 70, 343–348.

Galeazzi A., Prifti K., Cortellini C., Di Pretoro A., Gallo F., Manenti F., 2023, Development of a surrogate model of an amine scrubbing digital twin using machine learning methods, Computers & Chemical Engineering, 174.

Ghorbani B., Hamedi M.H., Amidpour M., 2016, Development and optimization of an integrated process configuration for natural gas liquefaction (LNG) and natural gas liquids (NGL) recovery with a nitrogen rejection unit (NRU), Journal of Natural Gas Science and Engineering, 31, 590–603.

Goh B.Z.H., Elyas R., Hamid A., Foo D.C.Y., 2024, Artificial neural network modelling of natural gas dehydration process, Discover Chemical Engineering, 4(1).

GPSA, 2004, Engineering Data Book, Gas Processors Supply Association, 12th ed., Sec. 14, Tulsa.

Kidnay A.J., Parrish W.R., McCartney D.G., 2006, Fundamentals of Natural Gas Processing, Vol. 218, CRC Press, Boca Raton, New York, USA.

Lynch J.T., Lousberg N.B., Pierce C.M., 2007, How to compare cryogenic process design alternatives for a new project, Presented at the 86th Annual Convention of the Gas Processors Association, San Antonio, Texas, USA.

Mokhatab S., Poe W.A., Speight J.G., 2006, Handbook of Natural Gas Transmission and Processing, Gulf Professional Publishing, UK.

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.

Rovira Herrero G., 2022, Systematic methodology to develop surrogate models for chemical processes using artificial neural networks, Bachelor’s thesis, Universitat Politècnica de Catalunya.

Sinnott R.K., 1999, Chemical Engineering Design, Coulson & Richardson’s Chemical Engineering Series, Vol. 6, Third edition, Elsevier, UK.

Zhu J., Fan C., Yang M., Qian F., Mahalec V., 2025, A semi-supervised learning algorithm for high and low-frequency variable imbalances in industrial data, Computers & Chemical Engineering, 193.