A comprehensive gray-box framework for high-fidelity process simulation calibration

Jaime David Ponce-Rochaa\* , David Camilo Corralesb

aDirección de Ingeniería y Construcción de Plantas, CIATEQ A.C. Centro de Tecnología Avanzada, Av. del Retablo 150, Constituyentes Fovissste, Queretaro, Querétaro, 76150, México.

bToulouse White Biotechnology (TWB), INRAE, UMS (1337), 135 Avenue de Rangueil, Toulouse, 31077, France.

[jaime.ponce@ciateq.mx](mailto:jaime.ponce@ciateq.mx)

Abstract

The hybrid models (gray-box models) architecture allows for obtaining a simple and robust tool for chemical process description, where a process simulator represents the white-box model, and an artificial neural network (ANN) involves the black-box model. This study incorporates a case study of calibrating a simulation model based on conventional distillation using a hybrid model that associates operational data from a chemical plant with the results of a process simulator, and six operative scenarios were evaluated for the framework validation. The arrangement of the gray-box model involves a white-box with a black-box model parallel sequence, and later, the results validation was performed in the white-box model. The results showed that a hybrid model allowed obtaining a simulation with a better approximation to the behaviours of the system compared to conventional thermodynamic models, as well as identifying the optimal range of feed flow (425,994.98-475,938.10 kg/h) and the approximated optimal flow design (469,738.80 kg/h).

**Keywords**: Hybrid models, data-driven models, distillation column, artificial neural networks.

* 1. Introduction

A chemical process description can be achieved through three ways: a) First-Principia models (White-box models), b) Data-driven models (Black-box models), and c) Hybrid approach model (Gray-box models), which combine the advantages of the a) and b) models into a joint architecture, obtaining a simple and robust model capable of harnessing the advantages of both models (Kurz et al., 2022).

Process simulators as computer-aided tools allow the conceptualization, evaluation, and optimization of any chemical process through a priori knowledge (First-Principia models: Material/Energy balances, thermodynamics, transport laws, kinetic laws, etcetera). Nevertheless, the integration of operational data with a process simulator represents added value for any company because it enables continuous improvement of the simulation model, process understanding, identification of relevant factors, relationships between operational variables, and ensuring an economically competitive operation (Asprion et al., 2019; Foo & Elyas, 2023).

On the other hand, the use of machine learning has become a powerful tool with applications in industrial scenarios, where the data-driven models based on technologies such as artificial neural networks (ANN) have been consolidated as appropriate tools, thus allowing for carry out the calibration and optimization of process systems mentioned above, handling the nonlinearity of process with multivariable inputs and outputs, and often having better results than more sophisticated statistical techniques (Su et al., 1992).

A grey-box model can be represented by different configurations between the black-box and white-box models (serial or parallel arrangements). However, when input and output operational data and the simulation model are available, an integrated model based on a semi-serial arrangement (illustrated in section 3) represents a suitable integrated model, where a first preprocessing of the data inputs is done in a white-box and black box parallel sequence, and later the outputs are taken like a new fed in the white box model for the final validation (Zapf & Wallek, 2021). These model arrangements offer advantages associated with lower computational expenses because they can substitute unknown components of a pure white-box model like finding non-available parameters, fitting numerical deviations of models, or finding a fast model for process optimization. Thus, this work aims to develop a comprehensive framework for gray-box models with high-fidelity results, using computer-aided tools like a process simulator, and data-driven models, as well as to identify the best operational point based on the relevant variables of the process.

* 1. Case study

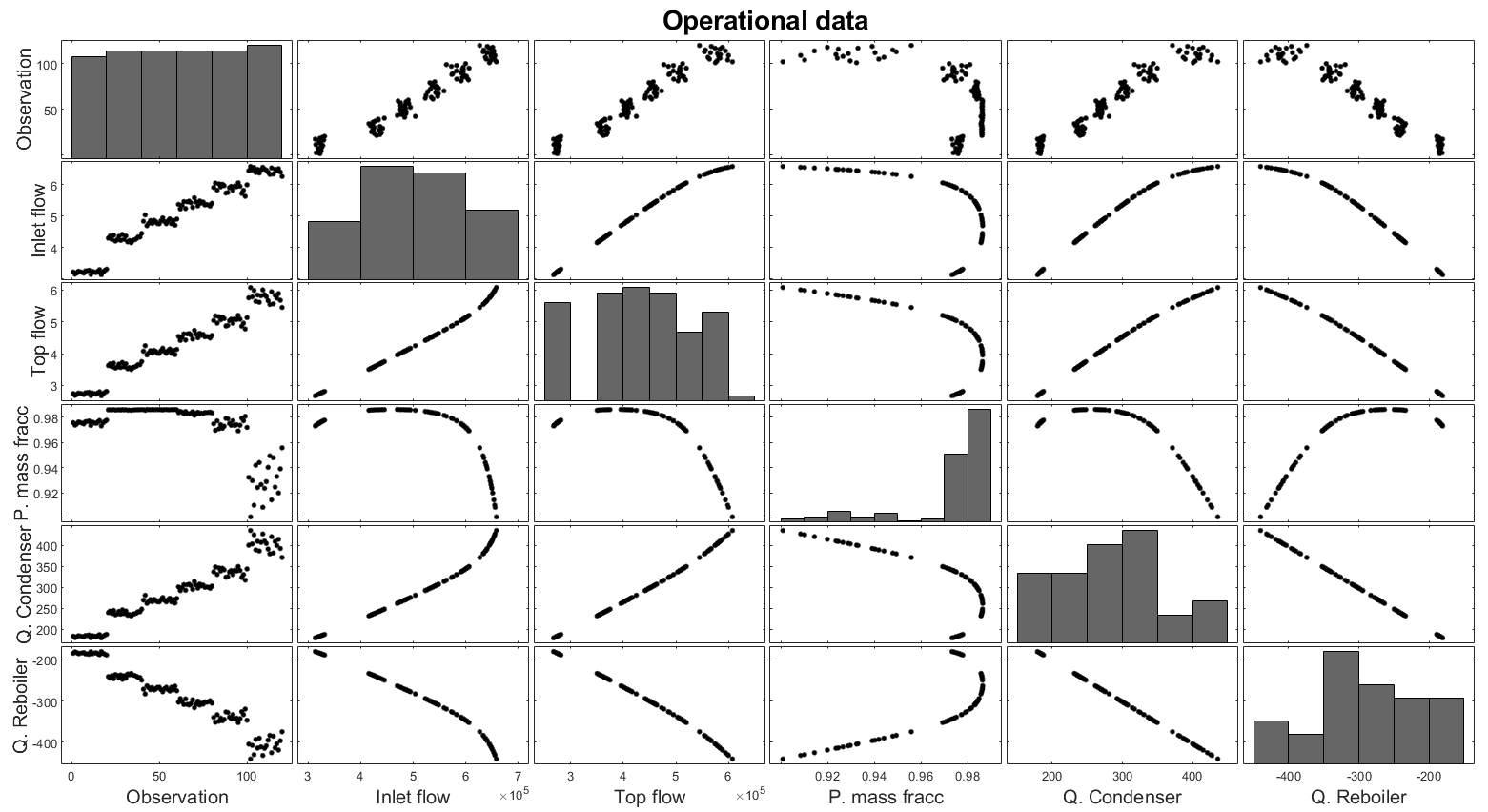
This work involves a case study of calibrating a simulation model based on conventional distillation using a grey-box model that incorporates operational data from a chemical plant and the results of a process simulator. The study evaluates six different operating conditions for framework validation. The calibration framework focuses on improving the prediction of liquid-vapor equilibrium using the Soave Redlich Kwong (SRK) equation of state. Figure 1 shows the distillation column diagram, where the separation and purification of methanol in a quaternary mixture is carried out (methanol-MeOH, water-H2O, dimethyl ether-DME, and ethanol-EtOH) (Adams II et al., 2018).

Diagram

Description automatically generated

Figure 1. Case study: distillation column with fifty number of stages.

A set of one hundred twenty-two operational data was obtained, where the main variables with upper and lower bounds were: inlet mass flow (313,850.00, 658,095.00 kg/h), top mass flow (267,526.65, 607,342.00 kg/h), top mass composition (0.90, 0.99), condenser duty (-439.96, -180.54 MW), and reboiler duty (178.97, 435.87 MW). Figure 2 shows the operational behavior and correlation of all variables mentioned above. The reflux ratio (RR=1.00), top pressure (1.36 bar), top temperature (117.00 °C), and input mass fractions (MeOH-0.8312, H2O-0.1661, DME-0.0019, EtOH-0.0007) were held constant without variations.

Figure 2. Operational data: observation, inlet mass flow (kg/h), top mass flow (kg/h), product mass fraction, condenser duty (MW), reboiler duty (MW).

* 1. Methodology
     1. A comprehensive gray-box framework for high-fidelity process simulation calibration

The framework involves the integration of four systematic steps, where a process simulation and operational data are available: 1) Selection of scenarios, 2) Calibration and validation model in black-box model, 3) Validation of results, and 4) Identification of optimal operative zone. The white-box model was represented through a model simulation developed in Aspen Plus, while the black-box model and the framework were implemented in Python.

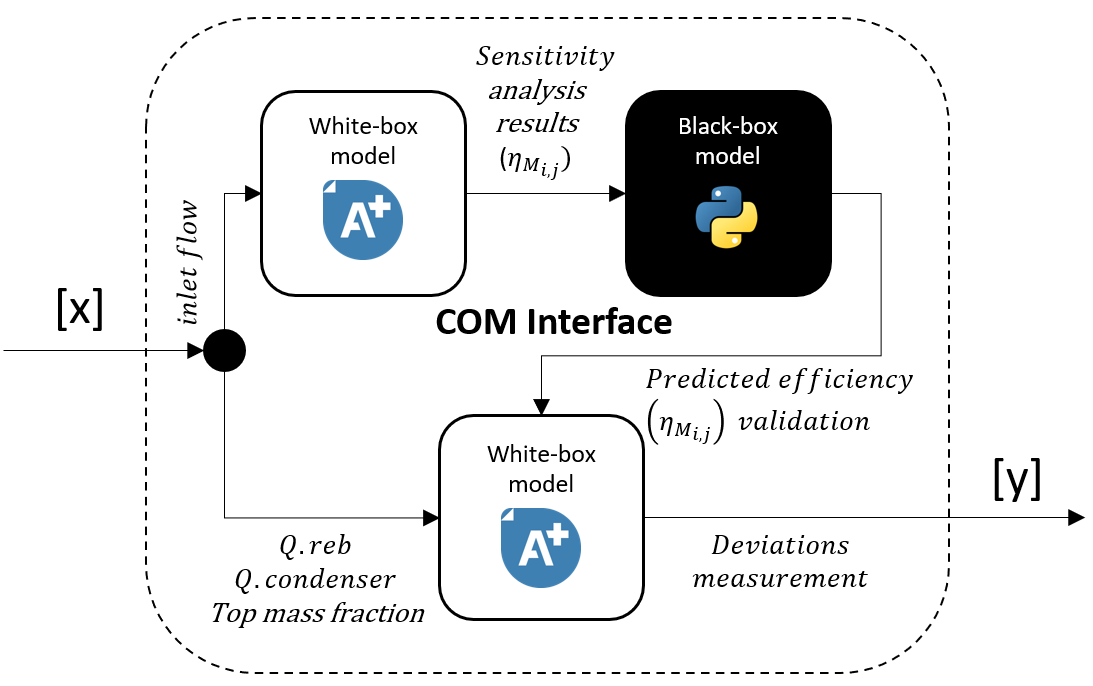


Figure 3. Gray-box model arrangement.

* + 1. Selection of scenarios

Based on the graphical results shown in Figure 2, it is possible to observe six representative operating zones considering the inlet mass flows. Therefore, Table 1 shows the selected operational points used as input data for the gray-box model.

Table 1. Operational scenarios.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario | Inlet flow  (kg/h) | Top flow (kg/h) | Product  mass frac | Condenser duty (MW) | Reboiler duty (MW) |
| I | 328,292.00 | 278,757.00 | 0.9771 | 185.80 | -187.40 |
| II | 425,995.00 | 358,720.00 | 0.9859 | 237.20 | -238.90 |
| III | 475,938.00 | 400,769.00 | 0.9862 | 265.00 | -266.90 |
| IV | 558,616.00 | 472,852.00 | 0.9816 | 314.20 | -316.40 |
| V | 590,502.00 | 503,250.00 | 0.9751 | 336.60 | -339.00 |
| VI | 649,336.00 | 584,499.00 | 0.9236 | 410.80 | -414.40 |

* + 1. Calibration and validation model

Although rigorous models based on non-equilibrium (mass transfer) for process simulation may be available or developed, these represent a significant computational and mathematical expense mainly, when these are components of a plant simulation or when an optimization task is required. An alternative to detailed mass transfer models involves the use of thermodynamic models based on phase equilibrium calculations, as well as the implementation of correction factors based on theoretical fundaments. For distillation columns, the Murphree efficiencies (Equation 1) is the most widely used equation:

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

Where is the Murphree vapor efficiency for component (*i*) on stage (*j*), and can be seen as a simplified mass transfer parameter that describes, how much the vapor composition changes from the inlet composition ( to the outlet composition () while approaching the equilibrium composition () for each column stage (Brunazzi et al., 2018).

For calibration and validation tasks, a COM interface between Python and Aspen Plus was developed to automate the framework steps. Initially, a sensitivity analysis was performed on the process simulator, generating a dataset with different values of the Murphree efficiency (0.10-0.99). This dataset was used in order to train the Artificial Neural Network (ANN) (test size: 30 %) and the obtention of the estimated efficiency value. Subsequently, this estimated value underwent validation in the process simulation, and the Root Mean Square Error (RMSE) was used as a performance metric to measure the differences between predicted and real values.

The neural networks employed were multi-layer perceptron (3) using a rectified linear unit activation function (ReLU) with a sequential model and 750 epochs for each evaluated scenario.

* + 1. Identification of optimal operative zones

For the identification of optimal operative zones, the reboiler duty, condenser duty, and product mass fraction were considered relevant variables. Therefore, the use of the utopia-tracking approach was implemented to identify the condition with minimum reboiler and condenser duty and a maximum purity of the product.

* 1. Results

Table 2 shows a comparative summary between the operative data, results of the hybrid model (calibrated), and the First-Principia model (process simulator results), the six scenarios report the value of the Murphree efficiencies, reboiler duty, condenser duty, top mass fraction, and the RMSE.

Table 2. Summary of the main results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Scenario | Murphree efficiency (ηM) (%) | Reboiler  duty  (MW) | Condenser duty  (MW) | Top mass fraction | RMSE |  |
| I | - | 185.80 | -187.40 | 0.9771 | - | OD |
| 0.67 | 184.58 | -186.03 | 0.9702 | 1.06 | C |
| 1.00 | 181.79 | -182.97 | 0.9841 | 3.45 | NC |
| II | - | 237.20 | -238.90 | 0.9859 | - | OD |
| 0.79 | 237.58 | -239.24 | 0.9778 | 0.29 | C |
| 1.00 | 235.69 | -237.31 | 0.9846 | 1.27 | NC |
| III | - | 265.00 | -266.90 | 0.9862 | - | OD |
| 0.79 | 265.44 | -267.22 | 0.9780 | 0.31 | C |
| 1.00 | 263.35 | -265.08 | 0.9847 | 1.42 | NC |
| IV | - | 314.20 | -316.40 | 0.9816 | - | OD |
| 0.71 | 313.11 | -315.15 | 0.9740 | 0.63 | C |
| 1.00 | 309.13 | -311.07 | 0.9845 | 4.25 | NC |
| V | - | 336.60 | -339.00 | 0.9751 | - | OD |
| 0.66 | 332.31 | -334.46 | 0.9706 | 3.61 | C |
| 1.00 | 326.78 | -328.80 | 0.9850 | 8.17 | NC |
| VI | - | 410.80 | -414.40 | 0.9236 | - | OD |
| 0.60 | 367.85 | -370.20 | 0.9651 | 35.58 | C |
| 1.00 | 359.36 | -361.53 | 0.9851 | 42.59 | NC |
| OD-Operative data, C-Calibrated simulation, NC-Non-calibrated simulation | | | | | | |

The RMSE implementation allowed quantifying the deviation between operational data and the hybrid model or process simulation results. The predictions for scenarios I to IV exhibit the lowest deviation values, accompanied by the highest product mass fraction. For inlet mass flows greater than 590,502.00 kg/h, the RMSE values for the gray-box model and process simulation show a significant increase. These increases are associated with an overflow in the distillation column; consequently, the prediction of the liquid-vapor equilibrium begins to be affected even with the gray-box model implementation.

The implementation of the utopia-tracking approach enables the identification of the best operative point, characterized by a product mass purity of 0.9863, reboiler duty of 261.46 MW, condenser duty of -263.35 MW, and an inlet mass flow of 469,739.0 kg/h. These conditions correspond to an optimal operative zone between scenario II and III values.

* 1. Conclusions

The results showed that a hybrid model allowed obtaining a simulation with a better approximation to the system behavior concerning conventional thermodynamic models (SRK), as well as estimating Murphree efficiencies values quickly, accurately, and continuously for the model, reducing the future instrumentation investment and experimental determination of unknown parameters.

On the other hand, it is possible to infer that the increase in the feed flow within the column presents a direct relationship with higher operating costs; however, there is a point at which the purity and, therefore, the efficiency of the stages reaches a maximum, which is associated with an operation close to the column design flow (approximately 469,739.00 kg/h). Finally, the proposed tools and methodology allow obtaining an area of opportunity for the generation of more complex analyses extrapolated towards the dynamic operation of the system and integration of more variables to generate greater robustness in the predictions of the hybrid model.

Acknowledgments

Jaime David Ponce-Rocha acknowledges the financial support from CIATEQ-Advanced Technology Center for the development of this project.

References

T. A. Adams II, T. Thatho, M. C. Le Feuvre, & C. L. E. Swartz, 2018, The Optimal Design of a Distillation System for the Flexible Polygeneration of Dimethyl Ether and Methanol Under Uncertainty, Frontiers in Energy Research, 6(Jun), 41.

N. Asprion, R. Böttcher, R. Pack, M. E. Stavrou, J. Höller, J. Schwientek, & M. Bortz, 2019, Gray-Box Modeling for the Optimization of Chemical Processes, Chemie Ingenieur Technik, 91(3), 305–313.

E. Brunazzi, E. Sorensen, J. Bausa, & J. Steimel, 2018, Extending Murphree Tray Efficiency from Mass to Heat Transfer in Distillation, Chemical engineering transactions, 69.

D. C. Y. Foo, & R. Elyas, 2023, Introduction to process simulation, Chemical Engineering Process Simulation, 3–28.

S. Kurz, H. De Gersem, A. Galetzka, A. Klaedtke, M. Liebsch, D. Loukrezis, S. Russenschuck, & M. Schmidt, 2022, Hybrid modeling: towards the next level of scientific computing in engineering, Journal of Mathematics in Industry, 12(1), 1–12.

H.-T. Su, N. Bhat, P.A. Minderman & T. J. McAvoy, 1992, Integrating Neural Networks with First Principles Models for Dynamic Modeling, IFAC Proceedings Volumes, 25(5), 327–332.

F. Zapf & T. Wallek, 2021, Gray-box surrogate models for flash, distillation and compression units of chemical processes, Computers & Chemical Engineering, 155, 107510.