

ARTICLE

Exploring Alternatives to Create Digital Twins from and for Process Simulation

Jaime Barbero-Sánchez^{1*} , Alicia Megía-Ortega¹, Víctor R. Ferro² , Jose-Luis Valverde¹ 

¹ Department of Chemical Engineering, University of Castilla La Mancha, Avda. Camilo José Cela 12, Ciudad Real, 13071, Spain

² Department of Chemical Engineering, Autonomous University of Madrid, Cantoblanco, Madrid, 28049, Spain

ABSTRACT

In this work, Digital Twins based on Neural Networks for the steady state production of styrene were generated. Thus, both the Aspen Technology AI Model Builder (alternative 1) and a homemade MS Excel VBA code connected to Aspen HYSYS and Aspen Plus (alternative 2) were used with this same aim. The raw data used for generating the Digital Twins were obtained from process simulations using Aspen HYSYS and/or Aspen Plus, which were connected through a recycle-like stream via automation for solving the entire simulation flowsheet. Aspen HYSYS was used for solving the pre-heating, reaction, and stabilization sections of the process whereas Aspen Plus ensured the computing of the separation and purification columns. Both alternatives led to an excellent prediction showing the capability of creating Digital Twins from and for process simulation.

Keywords: Digital Twin; Aspen Hybrid Model Builder; Aspen HYSYS; Aspen Plus; Automation; MS Excel-VBA

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) technologies, together with vast amounts of data obtained from modern digital technologies,

have emerged as a cornerstone of the Industry 4.0 revolution. Currently, almost every industry sector, including chemicals, is pursuing AI-enabled processes^[1-3]. Early examples of Industry Digitalization in the chemical industry are predictive maintenance

*CORRESPONDING AUTHOR:

Jaime Barbero-Sánchez, Department of Chemical Engineering, University of Castilla La Mancha, Avda. Camilo José Cela 12, Ciudad Real, 13071, Spain; Email: jaime.barbero@uclm.es

ARTICLE INFO

Received: 19 December 2023 | Revised: 10 January 2024 | Accepted: 12 January 2024 | Published Online: 19 January 2024

DOI: <https://doi.org/10.30564/jcsr.v6i1.6168>

CITATION

Barbero-Sánchez, J., Megía-Ortega, A., Ferro, V.R., et al., 2024. Exploring Alternatives to Create Digital Twins from and for Process Simulation. *Journal of Computer Science Research*. 6(1): 16–30. DOI: <https://doi.org/10.30564/jcsr.v6i1.6168>

COPYRIGHT

Copyright © 2024 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

and process automation. Supervised learning, unsupervised learning, and reinforcement learning are paradigms related to ML^[4]. Specifically, supervised Learning (SL) is a type of ML in which a model is constructed based on data that include the inputs (independent variables) and known outputs (dependent variables) for a potentially large number of examples^[1].

Digital Twins can provide digital copies of the production lines while offering real-time process monitoring, control, and optimization^[5,6]. They establish feedback between real physical systems and the digital model^[7] and allow to simulation of the physical systems more realistically^[8]. To perform an accurate representation of a real system using a Digital Twin, as much information as possible is required^[9]. However, acquiring data is a challenging task^[8]: a multitude of sensor technologies based on the Internet of Things (IoT) exist, and which one to rely on is a common source of concern; data acquired from sensors is not always reliable and is prone to noise and randomness; acquiring and validating data in a timely manner is a difficult task, etc. Consequently, some AI-based solutions must be addressed to avoid these issues. Among them, one could consider the following ones: data validation by utilizing ML classifiers to organize sensors as faulty; generation of synthetic data to simulate the behavior of physical systems which can later be utilized by ML models for testing purposes, etc.

In other words, AI has opened a new dimension to model and simulate chemical processes^[3]. It has been widely used as a tool for predictive analysis and has been successfully used to model processes including crystallization^[10-13], adsorption, distillation^[13], gasification^[14], dry reforming^[15] and filtration^[16]. In chemical engineering industries and many academic laboratories, AI is already in use, mostly to monitor, predict and control the outcomes of unit operations^[17]. To date, AI is most widely used in engineering to find the relationship between a dependent variable and several independent variables through regression algorithms^[17]. Alternatively, AI can be built using classification or clustering algorithms depending on the inputs available about the

problem that allow the prediction of the desired output.

Recently, Aspen Technology Inc. has incorporated into its products the AI Model Builder, which can generate Hybrid Models of processes and operations^[18-22]. The models obtained this way can be deployed to the flowsheets at the process simulators Aspen Plus and Aspen HYSYS, allowing the performance of mass and energy balances, sensitivity analyses, process optimizations and other calculations of relevant significance in design and analysis at the process engineering step of any project. Aspen Hybrid Models conveniently combine first principles and machine learning algorithms resulting in three types of them: those driven by AI or by pure first principles and, finally, the reduced order ones.

The pure first-principles models use theoretical equations and a limited set of empirical data to obtain their adjustable parameters. They generally extrapolate well over a wide range of conditions. However, they may not be available for all phenomena or operations of interest in the chemical industry and can be time-consuming to create and run.

AI Hybrid Models have pure machine learning character. The data used for training them can proceed from plants or experiments. They need lots of “good” data for training as already mentioned. If this condition is ensured, the model is accurate over the range of operating conditions used for its training. From this, it is evident that this kind of model may not extrapolate well or may violate physical constraints which in several cases are difficult to interpret. Positively it can stand out on these models with two noticeable features: i) they may simulate operations for which first-principles formulations are not available or result very difficult to describe in this way, and ii) they run very fast ensuring good convergencies within the overall process flowsheet.

Reduced order workflow is used to create an empirical model based on data from numerous simulation runs, where certain constraints are satisfied (e.g., mass balance). The creation of reduced order models follows two steps (**Figure 1**). Firstly, a case study on the process or operation is run, and second-

ly machine learning procedures are used to generate a Neural Network (NN) by using the data obtained in the former step. Finally, the model is deployed in the simulation flowsheet as any other process or logical operation. To fulfill this operational process, Aspen Technology has introduced two computational tools: i) the Aspen Multi-Case which takes advantage of the multicore architecture of modern CPUs to run a lot of cases (in the order of thousands) in much less time than the original Case Study tool implemented in Aspen HYSYS or the Sensitivity analysis available in Aspen Plus, ii) the Aspen AI Model Builder, a web-based service which allows the user creating NNs based on the data supply from plant operation, experiments or multi-case runs of rigorous process simulations.

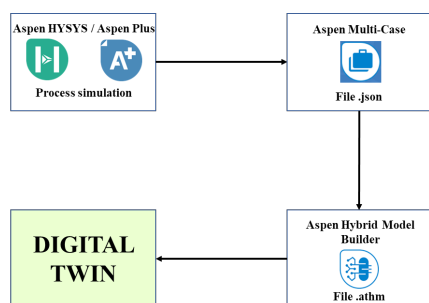


Figure 1. Reduced order workflow used to create Hybrid Models according to the new AI resources implemented in the Aspen Technology Process Engineering suite of programs.

Aspen Technology's products offer another interesting and confident way to create processes' Digital Twins. This latter one is supported by the automation strategies, where the process simulators (Aspen Plus and/or Aspen HYSYS) can be linked to third-party codes built in MS Excel-VBA, Python, etc. In this alternative, the NN is created and managed by third-party software, obtaining the process information from the simulators via automation. An advantage of this strategy is that the NN can be run by the third-party software selected without requiring the use of the process simulator. On the other hand, as the NN is created by a user code, its configuration, structure, and other features can be freely manipulated.

Artificial Neural Networks (Artificial NN) are commonly used in Supervised Learning. An artificial

neuron, also called perceptron, is the single processing unit that composes the Artificial NN and acts as a transfer function of inputs to generate outputs. A feedforward network might contain at least one layer. After the network is built, it must be trained. The training process simply refers to the process of repeatedly feeding the inputs and outputs, followed by adjusting the weights and biases using a suitable algorithm. The larger the data used for training, the greater the accuracy of the network. Once the network is trained, it should be tested to predict the outputs for new inputs.

As above mentioned, to demonstrate the capability of Digital Twins as a surrogate of the physical system, real and reliable data should be acquired, which is not an easy task. Alternatively, the generation of valuable data that simulates the behavior of that system can be used for configuring the Digital Twin of the real process. However, the large number of scenarios to be generated requires manually modifying the same number of times a specific simulation for accounting for results, using some tools like Aspen Multi-Case as commented later to automate the simulation process with third-party software.

In this work, the two alternatives mentioned will be applied to create Digital Twins of a well-known chemical process: the production of styrene. To work with a well-known process allowed the authors to focus on the development of both alternatives rather than on the process itself. Thus, both the Aspen Technology AI Model Builder and an MS Excel VBA code will be used with this aim. The conceptual and basic developments of the process reported by Luyben^[23] and Haydary^[24] were taken as references for the current process formulation. The Digital Twins here were developed to reflect the stationary operation of the styrene plant. Correspondingly, the input and output selected variables are mainly related to the operability and the controllability of the industrial process. The raw data used for generating the digital twins were obtained from process simulations using Aspen HYSYS and/or Aspen Plus. In the second alternative explored in this work and because of the specificities of the process simulators employed,

the interconnection of both, Aspen HYSYS and Aspen Plus, was required for solving the entire simulation flowsheet i.e., Aspen HYSYS was used for solving the pre-heating, reaction, and stabilization sections of the process meanwhile whereas Aspen Plus ensured the computing of the separation and purification columns. The coupling of both simulators was solved by the automation methodology through an MS Excel-VBA code. This way, a recycle-like stream emerged when connecting both programs which demanded a code for performing a direct iteration algorithm that fixed the internal recycle stream of the process.

2. Process description

Styrene is an aromatic hydrocarbon derived from benzene which is obtained in the oil refining process [25]. There are different ways of producing styrene, the most widely used being the catalytic dehydrogenation of ethylbenzene, which often takes place in the gas phase and requires an excess of water vapor apart from the presence of solid catalysts, generally based on Fe_2O_3 . This main reaction is accompanied by several secondary reactions such as the pyrolysis of ethylbenzene to benzene and ethylene, and the dealkylation of ethylbenzene to toluene and methane. Two main technologies are commercialized to produce styrene: that developed by the German engineering BASF where the reactor operates near isothermal conditions and the one proposed by the American Dow Chemical in which the reactor operates under a *quasi*-adiabatic regime [24]. In the current

work, the adiabatic operation considered two reactor beds including an inter-bed heating (Figure 2).

Ethylbenzene conversions typically range from 65% to 75%, with selectivity towards styrene being between 93% and 97%. However, styrene and ethylbenzene have very similar boiling points, which will require the use of distillation columns with many stages operating at high reflux ratios. This fact determines that the separation and purification section of the process is strongly energy-demanding being unavoidable in its exhaustive description, for which the knowledge of the pressure profiles along the column is critical. The reactor outlet mixture (Stream S8) is cooled before further separation and purification. The heat released was used to generate high- and low-pressure steam. The most volatile components are then separated into two flash drums. The liquid streams coming from these drums are conducted to a decanter, where two streams: aqueous and organic ones, are obtained. The latter one (Stream FT100) was conducted to the separation and purification column train which is composed of two distillation columns (T-100 and T-101). In the first column (T-100), pure styrene was obtained at the bottom. This first column operated under vacuum conditions to avoid undesired reactions such as styrene polymerization. The top stream from this column is constituted by a mixture of ethylbenzene and byproducts such as toluene and benzene. This mixture was conducted to a second distillation column where the unconverted ethylbenzene was extracted from the bottoms with high purity and then recirculated to the conditioning unit.

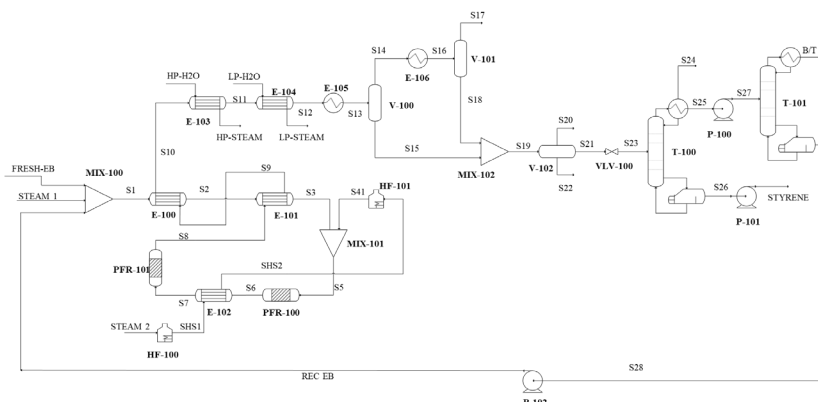


Figure 2. Flowsheet diagram of the styrene production process.

The two strategies considered in this work to create a Digital Twin of the styrene process are markedly different from the point of view of the computational details. However, the specifications used to define the process model are the same. All the calculations were performed with the Aspen Plus and/or Aspen HYSYS v12.1 and the corresponding version of the Aspen Multi-Case.

2.1 Process model specifications

The feed to the process (**Table 1**) consisted of pure ethylbenzene (Stream FRESH EB), steam (Stream STEAM1) and the recycled ethylbenzene stream coming (REC EB) from the bottom of the second distillation column (T-101). These three streams are mixed resulting in an effluent (S1) that was thermally conditioned in the heat exchange train (exchangers E-100 and E-101) which uses either the process fluid (S8 and S9). The conditioned mixture (S5) is fed to the reactions section which, as previously mentioned, is composed of two tubular reactors operating adiabatically with inter-bed heating (E-102). Both reactors are similar in dimensions and operating conditions.

Table 1 shows the input stream data to the reaction unit whereas **Table 2** shows the reaction unit decanter output current data, with stream S23 being the input to the separation unit (**Figure 3**).

The conditions and composition of the reactor outlet mixture are given in **Table 2**.

Columns T-100 and T-101 were specified by the purity of styrene and ethylbenzene, respectively, in the bottom streams (**Table 3**). The degree of freedom in both columns was completed with the reflux ratio

and the power supplied to the boiler.

The initial values for the simulation convergence were taken from Haydary^[24].

2.2 Input scenarios and response variables

In this study, 6 input variables and 4 levels of variation for each one were considered resulting in 4096 scenarios. The selected input variables were the mass flow rates of fresh steam entering the reaction unit, the top pressures of columns T-100 and T-101, and the pressures of the cooling water streams (HP-H₂O and LP-H₂O) yielding steam after passing through the corresponding heat exchangers. A variation of the input variables of 2.5% above and below the values reported in the reference simulations (**Table S1**)^[24] was assumed. The narrowest of the variation intervals for the input variables is consistent with the fact that the Digital Twin, as mentioned in the Introduction, is created to describe the stationary operation of the process.

The variables to be obtained by the Hybrid Model (output variables) were: styrene mass flow rate (STYRENE), T-100 column reflux ratio, ethylbenzene recirculation flow rate (REC EB), total ethylbenzene conversion (%) at the outlet of the two reactors, total boiler power (HFE-100 and HFE-101), T-100 column reboiler power, mass flow rate of the low pressure cooling water (LP-H₂O), mass flow rate of stream S17, H₂ mole fraction in stream S17, mass flow rate of the S24 stream (top stream in column T-100), mass flow rate of stream B/T (distillate from the T-101 column, containing benzene and toluene), and mass flow rate of stream S22 (aqueous phase at the outlet of the decanter).

Table 1. Input streams data to the conditioning and reaction units.

Streams	FRESH EB	STEAM1	STEAM2	LP-H ₂ O	HP-H ₂ O
Vapor fraction	0	1	1	1	1
Temperature [°C]	20.0	120.0	133.5	20.0	20.0
Pressure [kPa]	180	180	300	180	300
Mass flow [kg/h]	5308	1856	13721	704	4143
Mole Fractions					
E-Benzene	1	0	0	0	0
Water	0	1	1	1	1

Table 2. Data of the effluent streams from the reaction unit.

Streams	S17	S23	S22
Vapor fraction	1	0	0
Temperature [°C]	5.0	25.0	25.0
Pressure [kPa]	100	60	100
Mass flow [kg/h]	212	8018	15573
Mole Fraction			
E-Benzene	0.00	0.38	0.00
Water	0.01	0.00	1.00
Benzene	0.00	0.02	0.00
Ethylene	0.04	0.00	0.00
Hydrogen	0.89	0.00	0.00
Methane	0.05	0.00	0.00
Styrene	0.00	0.56	0.00
Toluene	0.00	0.03	0.00

Table 3. Operating variables for the distillation columns.

Operating Variable	T-100	T-101
Stages	80	36
Condenser	Partial	Total
Reboiler duty [MW]	3.252	0.376
Reflux ratio	6.92	9.63
Pressure on stage 1 [kPa]	10	120
Bottom pressure [kPa]	40	130
Feed stage	36	16
Column packing	Mellapak Plus	Mellapak Plus

2.3 Alternative 1. Using the Aspen Technology AI Model Builder

In this alternative, the Aspen Multi-Case launched from an Aspen Plus model of the process was used to generate massive results in the process (Paragraph 2.2). The property model Peng Robinson was used to estimate the fluid properties. Further, the Aspen AI Model Builder was used to generate the NN. Finally, the Hybrid Model was deployed to the simulation flowsheet as a sensor (**Figure 1**). In the current version of the Aspen AI Model Builder, the Lasso regression algorithm^[26] was used to create the NNs. The process was repeated several times trying to improve the quality of the NN generated. As part of this improvement, non-linear terms were included for certain variables when no good regressions were obtained with the linear ones. As recommended, 80% of the original data was used to train the neural

network whereas the remainder ones were used for its validation. Other percentages were tested aiming to improve the network capability without observing significant improvements of the results.

The model created was deployed to the process simulator as “sensor” according to the nomenclature adopted in this application. Correspondingly, only the sensor is necessary to calculate the response variables. No connections between process units or sections are permitted for the sensors.

2.4 Alternative 2. Using the MS Excel-VBA automation of Aspen HYSYS and Aspen Plus to generate the neural network

Aspen HYSYS was chosen for calculating the feed conditioning and reaction sections as well as the cooling of the gases overcoming the reactor and their separation. Meanwhile, the columns were com-

puted in Aspen Plus. This decision is determined by the fact that updating the information related to the column pressure profile in automated calculations, like sensitivity analyses, is a key factor in the current work. It significantly impacts the correct prediction of both the temperature at the bottom and the reboiler duties. In Aspen Plus this functionality is fully integrated into the internal analysis calculations. However, in Aspen HYSYS is possible to export the pressure profile in an individual calculation, but not automatically when case studies are performed. The Peng Robinson equation from the Aspen Properties was selected as the property model in both programs for predicting the fluid properties.

Integrating Aspen HYSYS and Aspen Plus in a single simulation demands continuous and bilateral information transference between both programs. This issue was solved here by connecting both programs through a homemade MS Excel-VBA application specially programmed for this purpose. It also solved the algebraic loop related to the ethylbenzene recycle (**Figure 3**). Thus, the subroutine used to converge the corresponding tear stream was included in the MS Excel-VBA code. In practice, the connection between simulators was made through the stream REB in Aspen Plus and the stream REC EB in Aspen HYSYS, and streams S23 in both simulators. When a new scenario must be computed, the MS Excel-VBA application sends an initial estimation of stream REC EB to Aspen HYSYS. After reaching the convergence, the information of stream S23 in Aspen HYSYS is transferred to stream S23 in Aspen

Plus. Once the simulation is converged, the corresponding results obtained for stream REB in Aspen Plus were compared to the input values of stream REC EB in Aspen HYSYS through a VBA code. If relative errors were higher than 10^{-4} (tolerance), the values of stream REB computed with Aspen Plus were transferred to REC EB in Aspen HYSYS. The process was repeated until the relative errors were lower than the tolerance.

Once the case study calculations were complete, the output variables produced by the coupled simulations were captured through another VBA subroutine and transferred to a third one which was responsible for creating the Neural Network.

2.5 Neural network training

Levenberg-Marquardt algorithm [27] coupled with an ANOVA analysis was used for training the network [28]. This algorithm has been also proposed for predictive inferential control of distillation. It was built in an MS Excel-VBA application which was sensitive to both the number of layers, the number of neurons in each hidden layer and the propagation or active function used in each layer. Pearson's ratio coefficient, r , defined similarly to the correlation index used in linear regressions, was also computed.

To generate a consistent neural network, the data generated by the process simulations were randomly divided into two groups: 80% of the data obtained were used for training the neural network, and 20% were used for its validation. This percentage is the same as that used by the Aspen AI Model Builder.

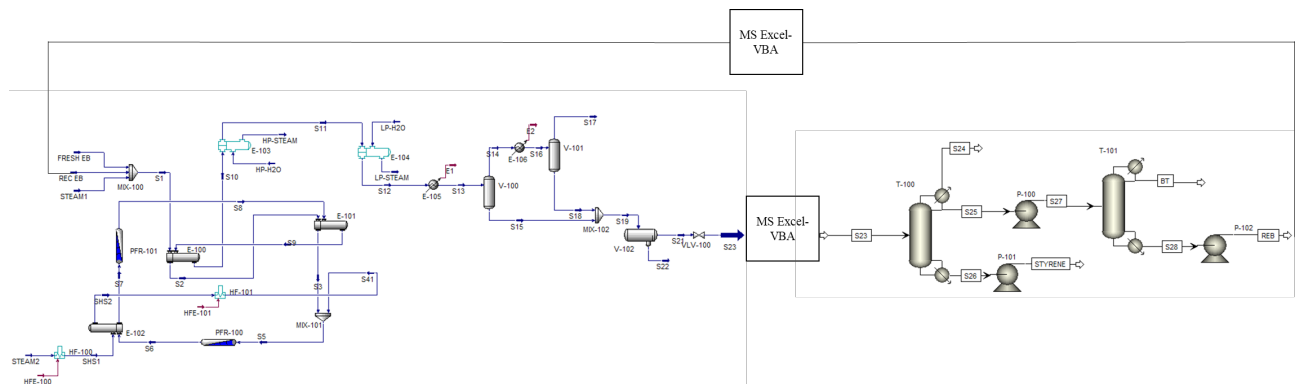


Figure 3. Combined Aspen HYSYS and Aspen Plus simulation flowsheet used in the second alternative explored in the current work to generate the data needed for training the Neural Network.

In this work, a multilayer perceptron was considered. It was composed of an input layer, an output layer and one or more hidden layers. The connections between neurons go from the neurons of one layer to the neurons of the next, with no possibility of feedback (non-recurrent Neural Network). Accordingly, information will only be transferred from the input layer to the output layer.

The processing unit is characterized by a weighted sum of inputs (p_i), an output (a_j) and weights (w_{ij}). The scalar input of a specific single neuron (p_i) in a layer of neuron is multiplied by the scalar weight (w_{ij}) to form $w_{ij}p_i$, one of the terms that is sent to the summer [29]. The other input (I) is multiplied by a bias (b_i) and then added to the total sum. The weight associated with the connection from unit i to unit j and the bias are continuously modified during the learning process. The output (b_i), often referred to as the net input, goes into a transfer function (f) which produces the scalar neuron output (a_i). This way:

$$n_j = \sum_{i=1}^N w_{ij}p_i + b_j \tag{1}$$

Three transfer functions were considered:

$$\text{Linear: } a_j = n_j \tag{2}$$

$$\text{Log-Sigmoid: } a_j = \frac{1}{1+e^{-n_j}} \tag{3}$$

$$\text{Polynomial: } a_i^{layer} = \left(n_i^{layer}\right)^i \tag{4}$$

where i ranges from 1 to the total number of neurons in layer-1.

For the training procedure, the number of layers and the number of neurons in each of the layers must be established. The number of neurons in the input layer is fixed by the number of input variables that constitute the independent variables. The number of neurons in the output layer is similarly fixed by the number of output variables that constitute the dependent variables. Finally, the number of hidden layers and the neurons defined in each layer will determine the learning capacity of the neural network. **Table 4** shows the different neural networks that were evaluated, as well as the number of layers and the number of neurons belonging to each layer. **Table 5** also displays the nomenclature used to identify the NNs. Thus, 7•P-3•P-2•L-14 means a neural network with three layers. The first received seven inputs (independent variables) connected with three neurons (polynomial function transfer) yielding three outputs. The second one took the three outputs of the first layer which are conducted to the other two neurons (polynomial function transfer) yielding two outputs. Finally, the third layer took these two outputs which were conducted to fourteen linear neurons yielding the same number of outputs (dependent variables).

Table 4. Neural networks tested. In all the cases the same 7 input and 14 output variables were considered.

Neural Network Name	Layers	Hidden Neural Networks	Type of Neural Network	Type of Output Neural Network
7•S-1•L-14	3	1	Sigmoidal	Lineal
7•S-2•L-14	3	2	Sigmoidal	Lineal
7•S-3•L-14	3	3	Sigmoidal	Lineal
7•P-1•L-14	3	1	Polynomial	Lineal
7•P-2•L-14	3	2	Polynomial	Lineal
7•P-3•L-14	3	3	Polynomial	Lineal
7•P-3•P-1•L-14	4	4	Polynomial	Lineal
7•P-3•P-2•L-14	4	5	Polynomial	Lineal

The training process consists of minimizing the following function, χ^2 , by nonlinear regression through the Levenberg-Marquardt algorithm:

$$\chi^2 = \sum_{i=1}^m \left[\sum_{k=1}^N (y_{ik} - a_{ij}^s)^2 \right] + \left[\sum_{k=1}^E F_k - \sum_{k=1}^S F_k \right]^2 \quad (5)$$

where y_{ik} and a_{ij}^s is the output variable k for the scenario i obtained by the automation process and the prediction made by the neural network, respectively. The second term considers the sum of the mass flowrates (F_k) entering the process (E) and those existing from it (S). This term forces that the mass balance was satisfied. After completing the nonlinear regression process, the optimal values of the neural network parameters were obtained.

Figure 4 shows a schematic representation of the whole computational procedure employed in the second alternative explored in the current work, from the generation of data through the automation process to the neural network validation.

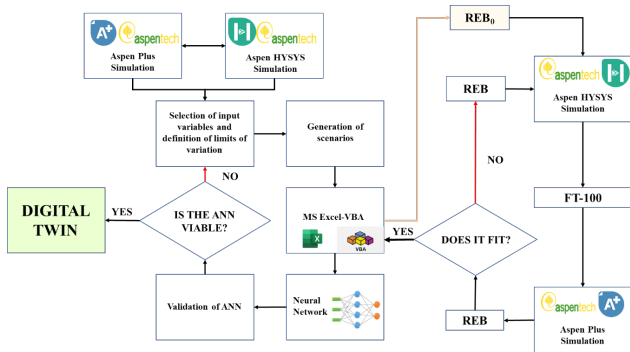


Figure 4. Block diagram of the process of generation of the Digital Twin for alternative 2: automation strategy.

3. Results and discussion

3.1 Aspen Hybrid Model

The Reduced Order Hybrid Model obtained by the procedure described previously (Figure 1) was deployed in a new Aspen HYSYS case (Figure 5).

Table 5 shows the statistics of the best model attained for the 12 output variables selected to describe the process performance. Both the reproducibility (R^2) and the predictability (Q^2) are higher than 0.990 for all the variables. As an example, Figure 6

displays the correlation between the reflux ratio calculated by the rigorous process model computed by Aspen Plus and those predicted by the Hybrid Model HM-100 for column T-100.

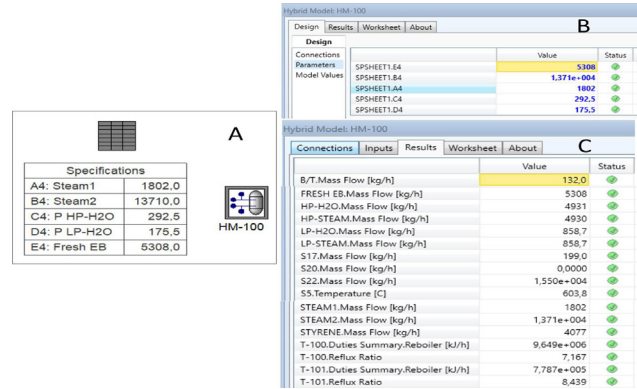


Figure 5 (A) Aspen HYSYS simulation flowsheet which incorporates the Hybrid Model (HM-100) created for reproducing the styrene process shown in Figure 2. Spreadsheet “Specifications” was used to generate the input values of the independent variables. They were exported as specifications to the Hybrid Model. (B) Specification sheet for the Hybrid Model. (C) Results page of the Hybrid Model.

Table 5. Reproducibility and predictability coefficients of the regressions supported by the Hybrid Model HM-100 obtained by the Aspen AI Model Builder.

Process Variable	R ²	Q ²
B/T.mass flow	0.993	0.993
S28.mass flow	0.993	0.993
Fresh EB.mass flow	1	1
HP-H ₂ O.mass flow	1	1
HP-Steam.mass flow	1	1
LP-H ₂ O.mass flow	0.996	0.996
LP-Steam.mass flow	0.996	0.996
S17.mass flow	0.993	0.993
S20.mass flow	1	1
S22.mass flow	1	1
S5. temperature	0.995	0.994
LP-H ₂ O.pressure	0.998	0.997
Total conversion EB	0.993	0.993
STEAM1.mass flow	1	1
STEAM2.mass flow	1	1
Styrene.mass flow	0.993	0.992

Table 5 and Figure 6 show that the Hybrid Model accurately reproduce the results obtained by the rigorous Aspen Plus model of the process. In fact,

for the conjunct of all the output variables selected, the mean relative error was lower than 1%.

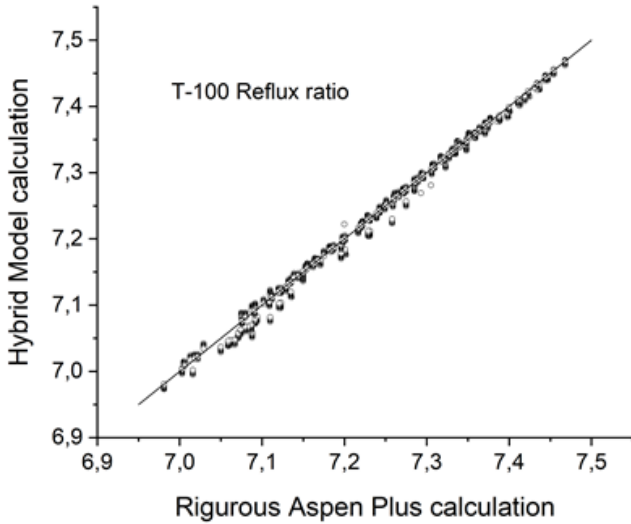


Figure 6. T-100 reflux ratio predicted by the Hybrid Model vs. those rigorously computed by the full process model in Aspen Plus. All the 4096 cases calculated are included in the graphic.

Relevant results are the differences in time for executing the 4096 cases using the rigorous Aspen Plus model and the Hybrid Model. Using the same machine, for the first one 516 minutes were required, whereas the second one demanded only 3 minutes. The Aspen Multi-case required 42 minutes employing 4 parallel runs and 16 CPUs of an I5 computer processing unit.

3.2 MS Excel-VBA supported Hybrid Model

Table 6 shows the parameters calculated for each of the Neural Networks tested, the χ^2 value, the mean error (%) and Pearson’s r correlation index.

The total number of parameters optimized depends on the number of neurons considered. The Neural Network that best fit the data generated using the automation methodology (lower χ^2 value and r value closer to 1) was the one that considered 3 polynomial neurons in layer 1, 2 polynomial neurons in layer 2 and 14 linear neurons in the output layer (7•6-3•6-2•7-14). **Figure 7** shows the architecture of the selected Neural Network, the numerical value of all the adjusted parameters and the transfer function considered for each neuron.

Once the most appropriate neural network was selected, it was validated with 20% of the total scenarios generated by automation. **Table 7** shows the statistics of the best model attained for the output variables selected. Both the reproducibility (R^2) and the predictability (Q^2) are higher than 0.990 for all the variables. In **Figure 8**, the output variables estimated by the chosen Neural Network are plotted against the same variables obtained by the rigorous calculations with Aspen HYSYS and Aspen Plus linked by the automation methodology shown in **Figure 3**.

Table 6. Results of the nonlinear regression process for the different NN tested in this work. Alternative 2: automation strategy.

Neural Network Name	Parameters	χ^2	Average Error (%)	r
7•0-1•7-14	36	1.46E+08	3.6	0
7•0-2•7-14	58	1.46E+08	3.6	0
7•0-3•7-14	80	1.46E+08	3.6	0
7•6-1•7-14	36	8.43E+07	1.5	0.651
7•6-2•7-14	58	5.11E+07	0.2	0.807
7•6-3•7-14	80	5.15E+07	9.3	0.805
7•6-3•6-1•7-14	56	7.58E+06	0.3	0.7298
7•6-3•6-2•7-14	74	6.89E+06	0	0.987

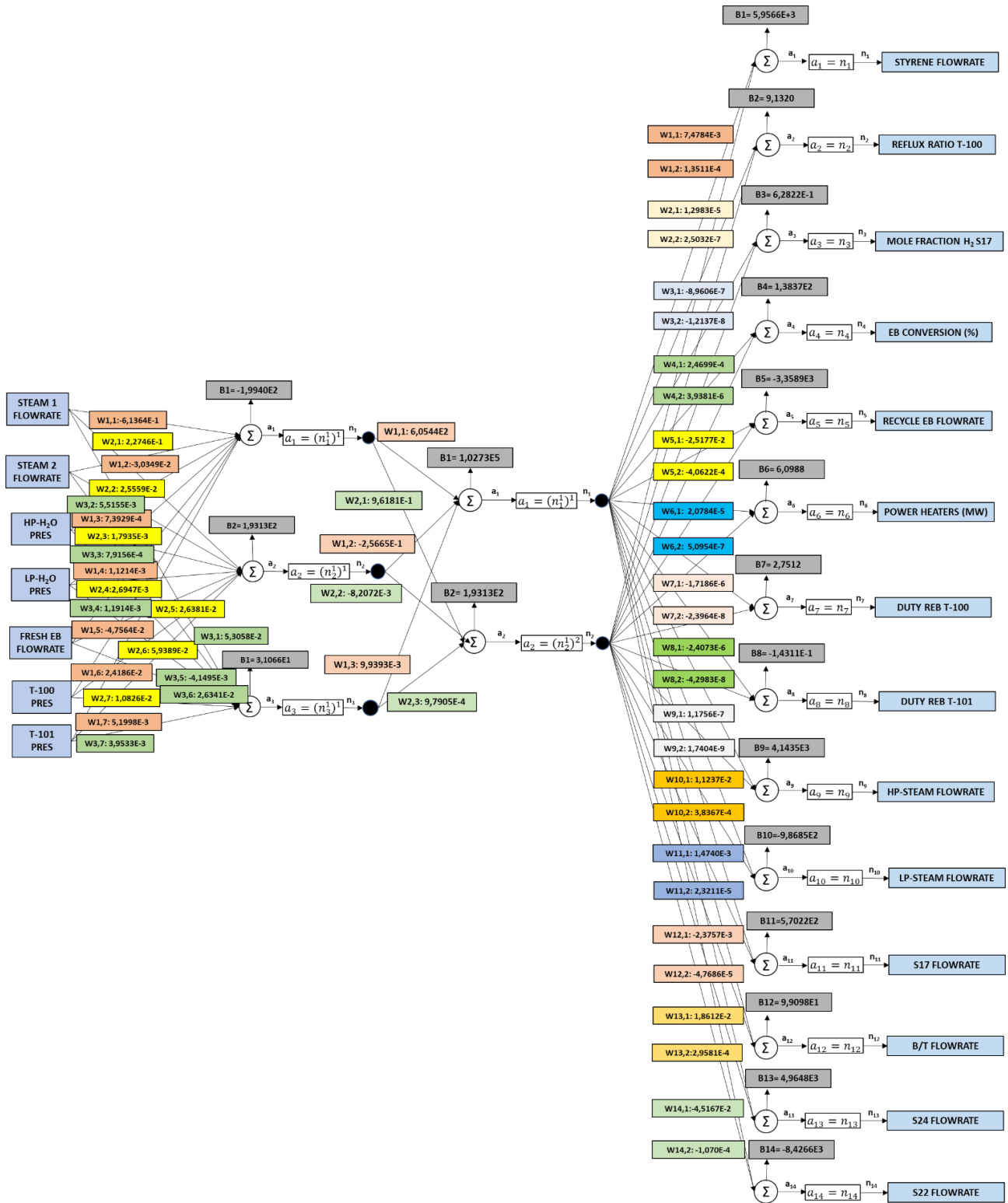


Figure 7. Neuronal Network selected (7•6•3•6•2•7•14). This NN consisted of three layers. The first received seven inputs connected with three neurons yielding three outputs. The second one took the three outputs of the first layer yielding two outputs. Finally, the third layer took these outputs yielding twelve outputs.

Table 7. Reproducibility and predictability coefficients of the regressions supported by the Hybrid Model HM-100 obtained by the automation strategy.

Process Variable	R ²	Q ²
Styrene flowrate	1.000	1.000
Reflux ratio	1.000	1.000
Total EB conversion	0.999	0.999
H ₂ mole fraction on S17	0.995	0.995
Recycle EB flowrate	0.999	0.999
Power heaters	1.000	1.000
T-100 reboiler duty	0.997	0.997
LP-steam flowrate	1.000	1.000
S17 flowrate	0.999	0.999
B/T flowrate	0.990	0.990
S24 flowrate	0.999	0.999
S22 flowrate	1.000	1.000

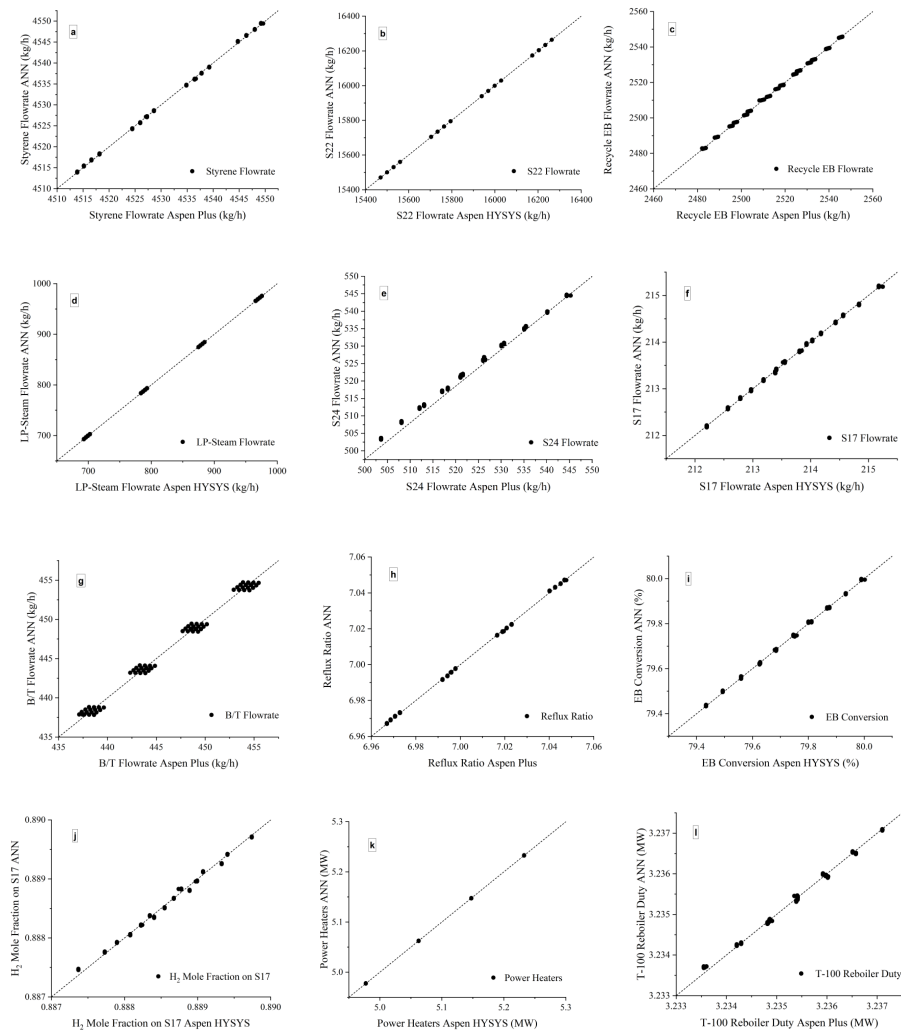


Figure 8. Validation of the Neural Network. a) styrene stream flow rate; b) S22 stream flow rate, c) recirculation stream flow rate (bottom flow rate of T-101); d) low pressure steam stream flow rate; e) S24 stream flow rate (gas flow rate of T-100); f) S17 stream flow rate; g) B/T stream flow rate (distillate flow rate of T-101); h) reflux ratio of T-100; i) total EB conversion; j) hydrogen purity of S17 stream; k) furnace power; l) T-100 boiler consumption.

Similarly, to that observed with the Aspen Hybrid Model, an excellent prediction was achieved with the Neural Network selected. This NN can be considered representative as a Digital Twin of the whole process of styrene production.

4. Conclusions

Both the Aspen Technology AI Model Builder and a MS Excel VBA code were successfully used for generating a Digital Twins based on a Neural Network to describe the steady state production of styrene. The raw data used for generating the Digital Twins were obtained from process simulations using Aspen HYSYS and/or Aspen Plus. The interconnection of both, Aspen HYSYS and Aspen Plus, via Automation was required for solving the entire simulation flow-sheet when the second alternative was considered. In this case, Aspen HYSYS was used for solving the pre-heating, reaction, and stabilization sections of the process meanwhile whereas Aspen Plus ensured the computing of the separation and purification columns. Both procedures led to excellent predictions.

Authors Contributions

This work was done jointly by all authors. Alicia Megía and Jaime Barbero prepared the simulations. Jaime Babero wrote the manuscript. Víctor Ferro developed the Aspen Hybrid Model. Jose Luis Valverde developed the Neural Network on Microsoft Office Excel. Víctor Ferro and Jose Luis Valverde coordinated the whole study, the manuscript review and revision. Jaime Barbero and Alicia Megía coordinated the data interpretation and result discussion.

Conflict of Interests

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.

Data Availability Statement

The authors declared that all data on the styrene

process are available in the bibliography. However, the results obtained are in an Excel spreadsheet that can be requested from the authors.

Acknowledgments

V.R.F. thanks to the Aspen Technology Inc. the possibility to participate in the training course “EHM 101: Introduction to Aspen Hybrid Models for Engineering”, where, during the trial time available for AIMB he carried out the case presented in the current paper.

References

- [1] Lee, J., Davari, H., Singh, J., et al., 2018. Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*. 18, 20–23.
DOI: <https://doi.org/10.1016/j.mfglet.2018.09.002>
- [2] Jan, Z., Verma, B., 2020. Multicluster class-balanced ensemble. *IEEE Transactions on Neural Networks and Learning Systems*. 32(3), 1014–1025.
DOI: <https://doi.org/10.1109/TNNLS.2020.2979839>
- [3] Feise, H.J., Schaer, E., 2021. Mastering digitized chemical engineering. *Education for Chemical Engineers*. 34, 78–86.
DOI: <https://doi.org/10.1016/j.ece.2020.11.011>
- [4] Murphy, K.P., 2012. *Machine learning: A probabilistic perspective*. MIT Press: Cambridge.
- [5] Mashaly, M., 2021. Connecting the twins: A review on digital twin technology & its networking requirements. *Procedia Computer Science*. 184, 299–305.
DOI: <https://doi.org/10.1016/j.procs.2021.03.039>
- [6] Rojek, I., Macko, M., Mikołajewski, D., et al., 2021. Modern methods in the field of machine modelling and simulation as a research and practical issue related to Industry 4.0. *Bulletin of the Polish Academy of Sciences. Technical Sciences*. 69(2), e136717.
DOI: <https://doi.org/10.24425/bpasts.2021.136717>
- [7] Hehenberger, P., Bradley, D., 2016. *Mechatronic futures: Challenges and solutions for mecha-*

- tronic systems and their designers. Springer: Berlin.
- [8] Cheng, J., Zhang, H., Tao, F., et al., 2020. DT-II: Digital twin enhanced Industrial Internet reference framework towards smart manufacturing. *Robotics and Computer-Integrated Manufacturing*. 62, 101881. DOI: <https://doi.org/10.1016/j.rcim.2019.101881>
- [9] Madni, A.M., Madni, C.C., Lucero, S.D., 2019. Leveraging digital twin technology in model-based systems engineering. *Systems*. 7(1), 7. DOI: <https://doi.org/10.3390/systems7010007>
- [10] Damour, C., Benne, M., Grondin-Perez, B., et al., 2010. Nonlinear predictive control based on artificial neural network model for industrial crystallization. *Journal of Food Engineering*. 99(2), 225–231. DOI: <https://doi.org/10.1016/j.jfoodeng.2010.02.027>
- [11] Velásco-Mejía, A., Vallejo-Becerra, V., Chávez-Ramírez, A.U., et al., 2016. Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. *Powder Technology*. 292, 122–128. DOI: <https://doi.org/10.1016/j.powtec.2016.01.028>
- [12] Yang, M., Wei, H., 2006. Application of a neural network for the prediction of crystallization kinetics. *Industrial & Engineering Chemistry Research*. 45(1), 70–75. DOI: <https://doi.org/10.1021/ie0487944>
- [13] Singh, V., Gupta, I., Gupta, H.O., 2007. ANN-based estimator for distillation using Levenberg-Marquardt approach. *Engineering Applications of Artificial Intelligence*. 20(2), 249–259. DOI: <https://doi.org/10.1016/j.engappai.2006.06.017>
- [14] Pandey, D.S., Das, S., Pan, I., et al., 2016. Artificial neural network based modelling approach for municipal solid waste gasification in a fluidized bed reactor. *Waste Management*. 58, 202–213. DOI: <https://doi.org/10.1016/j.wasman.2016.08.023>
- [15] Azzam, M., Aramouni, N.A.K., Ahmad, M.N., et al., 2018. Dynamic optimization of dry reformer under catalyst sintering using neural networks. *Energy Conversion and Management*. 157, 146–156. DOI: <https://doi.org/10.1016/j.enconman.2017.11.089>
- [16] Bagheri, M., Akbari, A., Mirbagheri, S.A., 2019. Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Safety and Environmental Protection*. 123, 229–252. DOI: <https://doi.org/10.1016/j.psep.2019.01.013>
- [17] Kakkar, S., Kwapinski, W., Howard, C.A., et al., 2021. Deep neural networks in chemical engineering classrooms to accurately model adsorption equilibrium data. *Education for Chemical Engineers*. 36, 115–127. DOI: <https://doi.org/10.1016/j.ece.2021.04.003>
- [18] Digital Twins: Essential to Driving Sustainable Operations for Chemical Producers [Internet]. Aspen Technology Inc.; 2021. Available from: https://www.aspentech.com/-/media/aspentech/home/resources/white-papers/pdfs/fy22/q2/at-07361-wp-digitaltwins_final.pdf
- [19] Course EHM101: Introduction to Aspen Hybrid Models for Engineering [Internet]. Aspen Technology Inc.; 2020. Available from: <https://esupport.aspentech.com/University-Course?Id=a3p4P000000Y4WVQA0>
- [20] Hybrid Modeling: AI and Domain Expertise Combine to Optimize Assets [Internet]. Aspen Technology Inc.; 2020. Available from: http://www.bluesp.co.za/wp-content/uploads/2021/02/AT-06425_POV_Hybrid_Modeling.pdf
- [21] Aspen Hybrid Models V12.1 [Internet]. Aspen Technology Inc.; 2020. Available from: https://www.aspentech.com/-/media/aspentech/home/resources/faq-documents/pdfs/fy22/q2/at-06496-hybrid-models-customer-faq.pdf?sc_lang=en
- [22] Aspen Hybrid Models: The Next Generation of Process Simulation Technology [Internet].

- Aspen Technology Inc.; 2020. Available from: <https://www.aspentech.com/en/resources/blog/aspen-hybrid-models-the-next-generation-of-process-simulation-technology>
- [23] Luyben, W.L., 2011. Design and control of the styrene process. *Industrial & Engineering Chemistry Research*. 50(3), 1231–1246. DOI: <https://doi.org/10.1021/ie100023s>
- [24] Haydary, J., 2019. Chemical process design and simulation: Aspen Plus and Aspen HYSYS applications. Wiley-AIChE: Hoboken, New Jersey.
- [25] Cui, C., Li, X., Guo, D., et al., 2017. Towards energy efficient styrene distillation scheme: From grassroots design to retrofit. *Energy*. 134, 193–205. DOI: <https://doi.org/10.1016/j.energy.2017.06.031>
- [26] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 58(1), 267–288. DOI: <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [27] Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*. 5(6), 989–993. DOI: <https://doi.org/10.1109/72.329697>
- [28] Valverde, J.L., Ferro, V.R., Giroir-Fendler, A., 2022. Estimation of e-NRTL binary interaction parameters and its impact on the prediction of thermodynamic properties of multicomponent electrolyte systems. *Fluid Phase Equilibria*. 551, 113264. DOI: <https://doi.org/10.1016/j.fluid.2021.113264>
- [29] Hagan, M.T., Demuth, H.B., Beale, M., 1997. Neural network design. PWS Publishing Co.: Worcester.

Supplementary information

Table S1. Limiting values of the input variables considered for creating the Digital Twin of the styrene process.

Operating Variable	Lowest Value	Initial Value	Highest Value
Fresh EB flow [kg/h]	5170	5308	5446
Steam 1 flow [kg/h]	1755	1802	1849
Steam 2 flow [kg/h]	13710	14070	14440
Pressure HP-H ₂ O [kPa]	292.2	300	307.8
Pressure LP-H ₂ O [kPa]	175.3	180	184.7

Source: [24].