Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Intelligent Soft Sensor Design for Composition Prediction in a Debutanizer Column Using ANN

Bassam Alhamad[1], Raed AlJowder[2], Wadeeah Almahaari[3], Hawraa Alattar[4]
[1]*University of Bahrain*
[2]*University of Bahrain*
[3]*University of Bahrain*
[4]*University of Bahrain*

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Chemical industries, such as natural gas and refinery industrial processes, face delays in measuring the composition of butane, as normal measurements depend on lab sampling. This delay affects the operation of the debutanizer to maintain the product quality. The single dynamic neural network model is used to overcome the lab-measurement delay through the development of a data-driven soft sensor, which enables the monitoring of the product quality at the top and the bottom of the column. The process of the debutanizer is simulated using Aspen-HYSYS in both steady-state mode and dynamic mode. The model is validated with life data from a natural gas plant, which is then used to develop the neural network model. The principal component analysis is performed using RStudio as a data reduction tool to optimize the number of variables used to create the soft-sensor model. The MATLAB neural network time series toolbox is used to build the neural network model and train it to give the future measurement of isobutane and normal butane. The function of the network is deployed using HYSYS-MATLAB interface code. According to the results from the test data, the built soft sensor showed accurate measurements of the product quality. The developed ANN-based soft sensor demonstrated exceptional accuracy, achieving a regression coefficient of 0.9999 for all data and 0.9987 for validation data, with a mean squared error of 1.07 and an RMSD of 1.88. The prediction accuracy exceeded 99%, and the minimal autocorrelation of error within the 95% confidence limit confirmed the robustness and reliability of the proposed model.

Keywords: *Soft sensor; Artificial neural network (ANN); Debutanizer column; Butane composition (iso-butane, n-butane); Aspen HYSYS; Principal Component Analysis (PCA); NARX; Online quality monitoring*.

INTRODUCTION

Natural gas processing companies regulate liquified petroleum gas (LPG) plant facilities to recover propane, butane, and naphtha from associated gas. Primarily, the associated gas with low pressure is collected and separated in gas-oil separators at a controlled pressure. Then, the gas is distributed to compressor stations to raise the pressure, along with air fan cooling. The compressed gas contains water vapor and acidic gases like carbon dioxide and hydrogen sulfide. So, when it condenses in the pipeline due to cooling, it causes acid corrosion. Thus, the compressed gas is dried in a GDU unit (Glycol Dehydration Unit), which is then sent to the gas processing units that consist of several operations within different unit operations: condensation, absorption, fractionation, treatment, and intermediate storage (**Figure 1**).

One of the fractionation units is the debutanizer column, which is used to separate the hydrocarbons based on the boiling point differences. It consists of several parts: a reboiler, a condenser, and the internals, which are a series of stacked trays/plates to enrich the separation. The operation of the distillation column begins when the feed (mixture of hydrocarbons) enters the column through the feed tray (mostly this tray is at the middle of the column). The

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

mixture of hydrocarbons introduced to the column is separated based on several factors related to the feed itself, the temperature of the reboiler and the condenser, and some other factors related to the boiling point of butane and other hydrocarbons. The boiling points at atmospheric pressure (14.7 psia/760 mmHg) are as follows: Butane (-0.5°C), Naphtha (100-160°C). The product quality of a debutanizer column is the focus of the operation, and controlling the product composition (product's purity) is crucial, especially for economy. Regularly the analyzers are used to measure the products' composition through laboratory tests, which are costly and time-consuming and may alter the product composition due to delays in the operation.

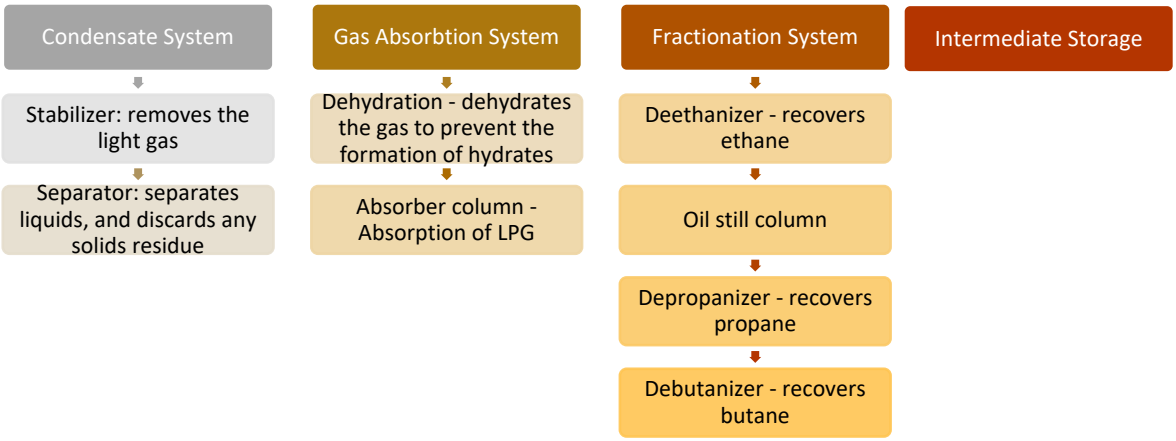

Figure 1. Natural Gas Plant Constituent Systems

The project's objective is to:

- Develop a soft sensor for a debutanizer column using an ANN that estimates top-product isobutane and n-butane compositions in (near) real time, reducing reliance on delayed lab analyses.
- Quantify predictive performance (R, MSE/RMSD, error autocorrelation) under unseen operating scenarios and demonstrate ≥99% accuracy with uncorrelated residuals.
- Identify and retain the minimum informative inputs via PCA, showing that a reduced variable set (e.g., reflux/boilup temperatures, top/bottom temperatures, key flow rates) preserves prediction fidelity.
- Validate the end-to-end workflow by simulating steady-state and dynamic cases in Aspen HYSYS, cross-checking against plant data, and training/testing the ANN in MATLAB.
- Demonstrate deployability through a MATLAB–HYSYS interface for online inference, highlighting latency reduction and suitability for integration with control/monitoring systems.
- Benchmark against conventional approaches (e.g., regression/PLS from prior studies) to evidence the advantages of nonlinear, dynamic ANN structures (e.g., NARX) for debutanizer quality prediction.

Operators of LPG/refining units often wait hours for lab results, forcing conservative operation, energy overuse, and occasional off-spec production. While traditional inferential models and linear methods struggle with strong nonlinearity and process dynamics, modern neural architectures can learn these relationships directly from data. Building on evidence that ANNs outperform linear soft sensors for butane composition—and noting the practical need for fast, robust, low-maintenance quality estimation—this work aims to deliver a validated, lightweight, and deployable ANN-based soft sensor. The goal is to close the information gap between process changes and composition awareness, enabling tighter quality control, reduced sampling costs, improved energy efficiency, and a practical pathway from simulation and plant data to an online, real-time estimator suitable for industrial debutanizer operations.

LITERATURE REVIEW

In many of the processes related to chemical industries, online processing had always been a challenge. The delay in the process causes poor control of the process variables. This may lower the product quality as control actions are delayed due to sampling lab techniques. Soft sensors had been adopted in several industries (Sun et al., 2014; Yan et

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

al., 2012; Yu et al., 2020; Yuan, Qi, Shardt, et al., 2020; Yuan, Qi, Wang, et al., 2020; Zeng & Ge, 2021). Soft sensing in industrial process and specifically for the debutanizer column, had been widely expedited (Barbosa, 2014; Kottavalasa & Snidaro, 2025; Peterson et al., 2025; Shang-Guan et al., 2025; Zhang et al., 2026; Zheng et al., 2025). (Sun et al., 2014) developed a new variable selection method to develop the soft sensor, which depends on a well-trained ANN, and then used the non-negative garrote (NNG) to shrink the number of input weights. (Yu et al., 2020) captures the causal factors that have the highest impacts by using the methodology of orthogonality and de-noising in feature analysis, where better performance could be achieved. (Yuan, Qi, Shardt, et al., 2020) expedites the importance of robust soft sensors, and that is to be achieved through the extraction of the dynamics and the non-linear nature of the system. (Yuan, Qi, Shardt, et al., 2020; Yuan, Qi, Wang, et al., 2020) adopted the multichannel convolutional neural network that grasps the correlations of different variable combinations and applied it on the industrial debutanizer column and hydrocracking process. (Zeng & Ge, 2021) obtained more information about the relation between the variables at different times by using the dynamic Bayesian network (DBN). (Zeng & Ge, 2021) used the regression analysis (vector regression) to increase the accuracy of the soft sensor. The methodology using the sequence of data at different times can be employed through the adoption of the ANN.

There had been various applications for developing a soft sensor for the debutanizer units in oil and gas industries. The recent efforts showed several in developing soft sensors to monitor the quality of the product (Mohamed Ramli et al., 2014; Ramli et al., 2016) showed that soft sensors of n-butane composition based on ANN achieved higher accuracy than partial least square (PLS) and regression analysis (RA). (Aimin et al., 2015; Kadlec et al., 2011; Liu et al., 2018; Saptoro, 2014; Shao et al., 2013; Shao & Tian, 2017; Singh et al., 2019). Soft-sensing techniques vary, using different methodologies to adapt to the non-linearity nature of the problem. However, there are efforts that show the accurate predictions of the composition of n-butane using neighborhood preserving regression embedding (NRPE) to capture the non-linearity in the process. This can be captured with the new models at ANN, which include nonlinear autoregressive networks with exogenous inputs (NARX) (Behnasr & Jazayeri-Rad, 2015; Hadian et al., 2021; Kadlec et al., 2011; Mohler et al., 2011; Ramli et al., 2016).

More advanced techniques included adaptive soft sensing methods, which consider the dynamic nature of the problem, where additional learning adapts the neural network model (Kadlec et al., 2011; Liu et al., 2018; Saptoro, 2014; Shao et al., 2013; Shao & Tian, 2017; Shi & Xiong, 2020; Sun et al., 2014; Urhan & Alakent, 2020; Xiong et al., 2017). (Kadlec et al., 2011) discussed the different methodologies used in the adaptive data-driven soft sensing techniques. Areas that need adaptive data-driven soft sensing techniques include moving from offline to online processes, adapting to the non-linearities, and coping with future and unpredictable changes within the process. Active and passive learning, as well as using both industrial and synthesis data, increase the accuracy of soft-sensing (Urhan & Alakent, 2020). ANN provides additional features on regression analysis-based models in terms of reliability, adaptation, and robustness. ANN with equation-based methods had been used as good estimators for the composition of butane in debutanizer columns (Mohamed Ramli et al., 2014; Ramli et al., 2016).

This paper presents a comprehensive study on developing a data-driven soft sensor for a debutanizer column using artificial neural networks (ANN). The process was simulated in Aspen HYSYS under both steady-state and dynamic modes and validated against real plant data to ensure model reliability. Principal Component Analysis (PCA) was employed in RStudio to identify the most influential variables and reduce data dimensionality, while the MATLAB Neural Network Time Series Toolbox was used to construct and train the ANN model. The integration between MATLAB and HYSYS enabled real-time prediction of isobutane and normal butane compositions. The paper details the simulation methodology, data preprocessing, network training, validation procedures, and performance evaluation metrics, highlighting the model's effectiveness in minimizing prediction delays and enhancing product quality monitoring in refinery operations.

DESIGN AND EXECUTION

The problem formulation in developing the soft sensor consists of several stages. First, a simulator is built in the Aspen Hysys platform, which includes the details of the process, including the mass and energy balance equations with all related physical and chemical properties of all components. The built simulator is validated with real-life data in both steady-state and dynamic modes to ensure the accuracy of the generated data. The simulator is then used to develop sufficient data for the ANN training and validation process. The ANN model is built with the details of the

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

weights and activation functions. To reduce the complexity of the system while preserving the required predicted accuracy, PCA was used to grasp most of the information, hence finding the causal factors. An interface code was written to update the weights in the ANN model. The details of the problem formulation are explained below.

The ANN is a machine learning method to comprehend the desired output based on the given input. It uses an optimization technique to attempt matching the desired targets, which are part of the network outputs. For every iteration, the network is trained to update the weights value to develop the model of the neural network. The data used for training the network must be analyzed before proceeding to using the principal component analysis (PCA). PCA is used to reduce the dimensionality of a large set of inputs by transforming it to a smaller set that explains the majority of variance in the data, without losing the accuracy. As the PCA is sensitive to variance, and to prevent the dominance of large ranges of variables, the standardization should be the first step to carry out so all variables are converted to the same scale. The next step is to find the correlation between the variables. To find the correlation matrix points with the independence relationship between multi-variables, the variance is calculated as follows:

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \quad (1)$$

The variance varies between -1 and 1, where 1 means strong positive correlation and -1 means strong inverse correlation; the closer it is to 0, the more uncorrelated those variables are.

The next step is to compute the eigenvalues and eigenvectors to identify the principal components. The PC is a linear combination of the variables. The eigenvectors are ranked from high to low, and the less significant ones are discarded to form the feature vector. The final data is the multiplication of the transposed feature vector with the initial data transposed (Jaadi, 2021)

To proceed with this analysis, the information is extracted, and the results are visualized based on coordinates for the variables, correlations between variables and principal components (dimensions), squared cosine, and contributions of the variables.

The squared cosine for variables implies the quality of each variable related to variances. It is calculated using the variables' coordinates. The contribution of a variable (var) to a given principal component (in percentage) is calculated as follows:

$$Contribution\ (\%) = \frac{Variable\ squared\ cosine \times 100}{Total\ squared\ csine\ of\ the\ component} \quad (2)$$

The simulator is built in the steady-state mode on the Aspen HYSYS platform (see **Figure 2**). The steady state allows for multiple plant scenario analyses. The process model is created to simulate realistic debutanizer column, which consists of 42 trays bottom to top with the required specifications (the column top/bottom temperatures and pressures along with feed composition, temperature, pressure, and flow.) Verification of the model include the implementation of the model against real data. The validation is simply done by comparing the inputs and outputs. The percentage error equation is calculated as follows:

$$E = \left|\frac{x_H - x_B}{x_B}\right| \times 100 \quad (3)$$

where X_H represents the top product purity (light components, like iso-butane/n-butane), X_B represents the bottom product purity (heavier components, like naphtha).

They are usually the controlled variables in process control and soft-sensing models, as maintaining X_H and X_B within target limits ensures the correct product specifications. After validating the data of simulator 1, another simulator is developed using modified HYSIM inside-out to add the details of the pumps and valves. (**Figure 3**). The steady-state simulator starts solving by verifying the degree of freedom (DOF) and solves based on the assigned specifications. The column values are calculated by the system on convergence. The simulator is then developed in the dynamic mode. The simulation of the plant provides detailed understanding, as well as testing and modification of the simulated process, which confirms the performance of the plant.

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

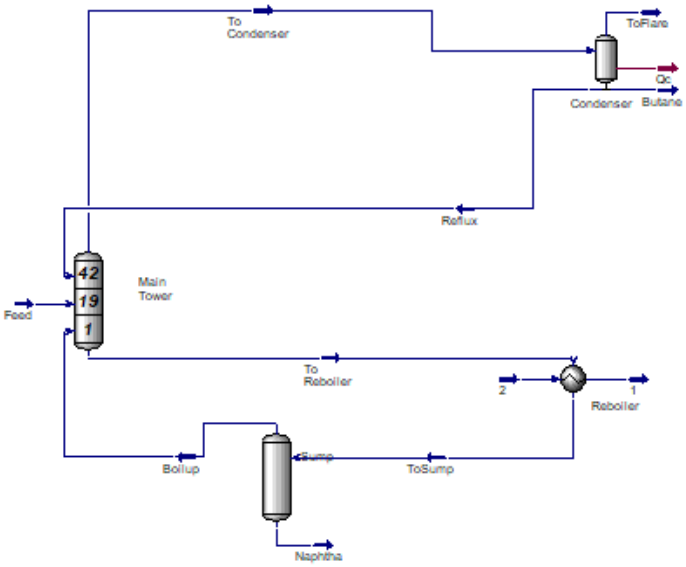

Figure 2. Debutanizer Column Steady-State Diagram (simulator 1)

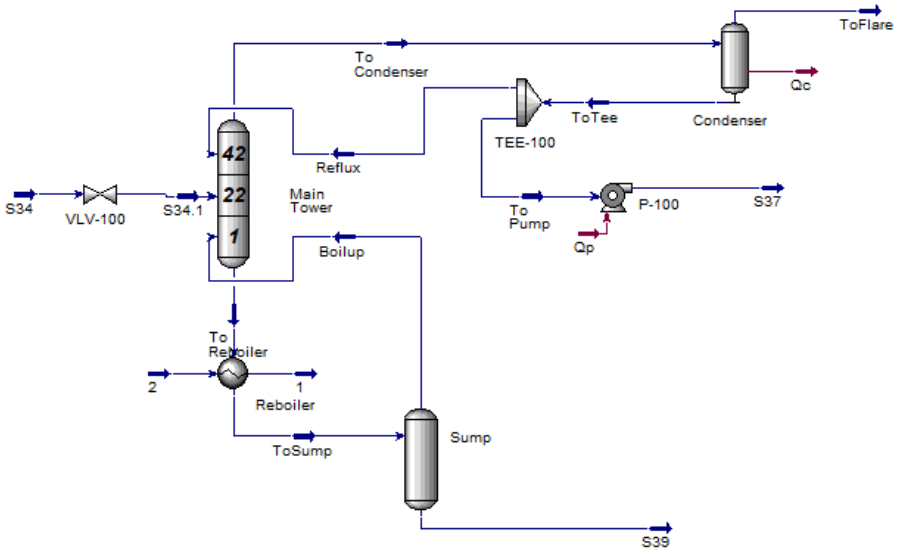

Figure 3. Debutanizer Column Steady-State Diagram (simulator 2)

To develop the ANN model, small increment changes in the inputs are applied to capture all the detailed changes in the outputs. This will also allow us to obtain the detailed dynamic change for every parameter. The 'case studies' tool in Aspen-HYSYS is used to add the independent variables and their boundaries, step size, or number of steps, as well as to add the dependent variables to simulate every case in the list and calculate the variables. Not all case study types are important, so the sensitivity and nested type are identified. The sensitivity will take every variable change alone and calculate accordingly. The nested type examines every possible combination of variable changes. The boundaries for each variable in the case studies are varied within ±10% of its set value, as most operational limits are within this range.

Before supporting NN with the data generated from the obtained case studies, the data is analyzed. Exploring the data will be performed using the RStudio program. This will aid in specifying the appropriate inputs for NN training. This will also provide an overview to examine the unchanged variables, which need to be excluded due to their insignificant correlation to the outputs. Meanwhile, the correlation plot shows the directly and inversely correlated variables between each other.

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To capture the non-linearity and the multi-dimensional nature within the system, the NARX with feedback connections and feedforward networks. As suggested by the equation above, the output signal y(t) is retreated (regressed) relying on the previous values of the output signal and previous values of an independent (exogenous) input signal. The diagram below represents the layers along with the weights while calculating the R (regression) values to obtain the best weights. As stated earlier and supported through literature, the PCA is used to reduce the dimensionality in the system, that is, to capture most of the information through identifying the causal factors. The NARX also achieved the denoised signal of the output. The output is fed back to the input of the feedforward neural network as part of the standard NARX architecture. Because the true output is available during the training of the network, a series-parallel architecture can be created. The first is that the input to the feedforward network is more accurate. The second is that the resulting network has a purely feedforward architecture and static backpropagation.

The data is loaded to the MATLAB workspace, and a matrix is created for the inputs (feed mass rate, feed i-butane, feed n-butane, top temperature, bottom temperature, reflux temperature, and boilup temperature) and another matrix for the outputs (top product i-butane composition and n-butane composition). These matrices of data are randomly divided, where 70% of the training data is used to train the network and update the weights, 15% of the training data is used for testing, and the remaining 15% is used for validation. It is to be noted that the simulator developed in Aspen Hysys is already validated with real-life data. The validation data is used to check the network generalization; however, overfitting is avoided, because over-training the data may ruin the network optimization.

This architecture consists of two layers, the hidden layer and the output layer nodes. The number of inputs and outputs to the network are 8 and 2, respectively. The Levenberg-Marquardt training algorithm is recommended, as it usually solves nonlinear problems, where the optimization is found through finding the minimum of a function. The hidden nodes are selected through the trial-and-error method. Several plots are generated to indicate the excellence of the network for best prediction outputs by minimizing the mean squared error.

In the first step, 70% of the data were randomly taken as a training set, while the remainder were used as a testing set. This was done by calculating the sample size, then using the sample function in RStudio to assign a random sample. Also, the 'set.seed' command was run to ensure that whenever we execute the sample function, a fixed random sample is generated to maintain consistency. The second step was scaling the data since the scaling is crucial to improve the NN learning process performance and stability by removing the scaling impact from all data variables. There are various techniques of data scaling, but during this step, the normalization scheme is used to rescale the dataset values from their original range and make it within the range of 0 and 1. In RStudio, scale function is employed to obtain the normalized data.

After the preprocessing steps, the neuralnet package was loaded to train and test the NN model. Where i-Butane and n-Butane compositions were set to be the dependent variables, and the feed temperature, i-Butane feed composition, n-Butane feed composition, and feed mass flow were situated to be the independent variables. Then, the neuralnet command was summoned to train the NN by providing it with dependent and independent variables, specifying the threshold to be 0.01, the backpropagation algorithm, and the activation function of sigmoid, the error function of "sse", the number of hidden layers with their neurons, which are all factors that have been manipulated multiple times to obtain the lowest possible error, ending up with an error of 1.88.

Thereafter, the constructed NN model was verified by executing the compute function to predict i-butane and n-butane compositions for the test data set. Additionally, the root mean square deviation (RMSD) was calculated to get the error between the actual and predicted data. Additionally, the root mean square deviation (RMSD) was calculated to get the error between the actual and predicted data.

PCA analysis is used to find the causal factors that contribute to the main changes between the input and output variables. This reduces the complexity that is inputted to the network. This analysis is utilized for nested data and sensitive data. There are 17 variables that could affect the output variables. The PCA returned 17 principal components (PC), where the first two PCs contained the most information of data. PC1 and PC2 contribute to 80.4% of the variation for nested data and 82% of the variation for the sensitivity. (See **Figure 4** and **Figure 5**). The PCA can be visualized using the factor map as shown in **Figure 6** and **Figure 7** as the quality of the variable is distended

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

from the origin. The positively correlated variables are close to each other, and the inversely correlated variables are located oppositely.

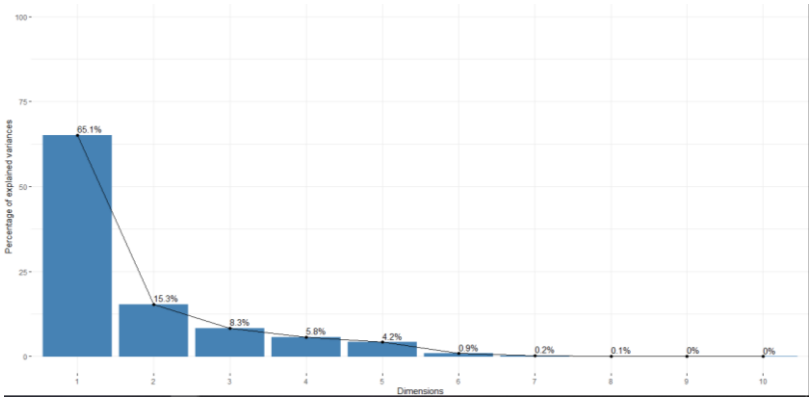

Figure 4. Percentage of Explained Variances Vs. Principal Component for Nested Data

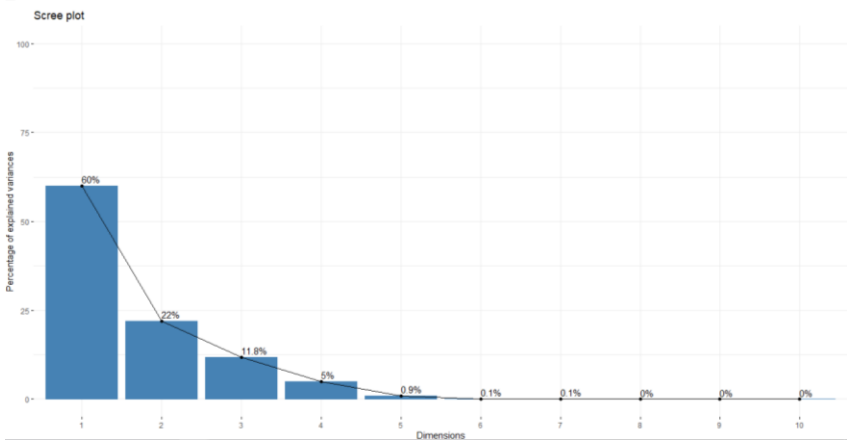

Figure 5. Percentage of Explained Variances vs. Principal Component for sensitivity Data

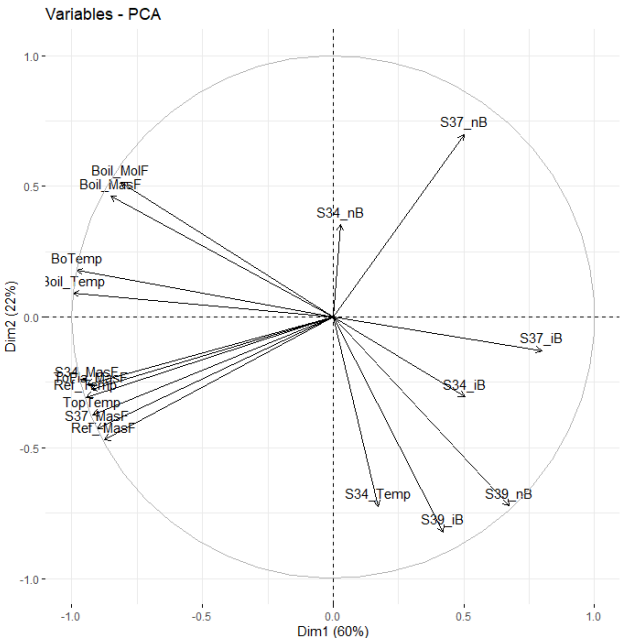

Figure 6. Factor Map (PCA) – Sensitivity Data

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

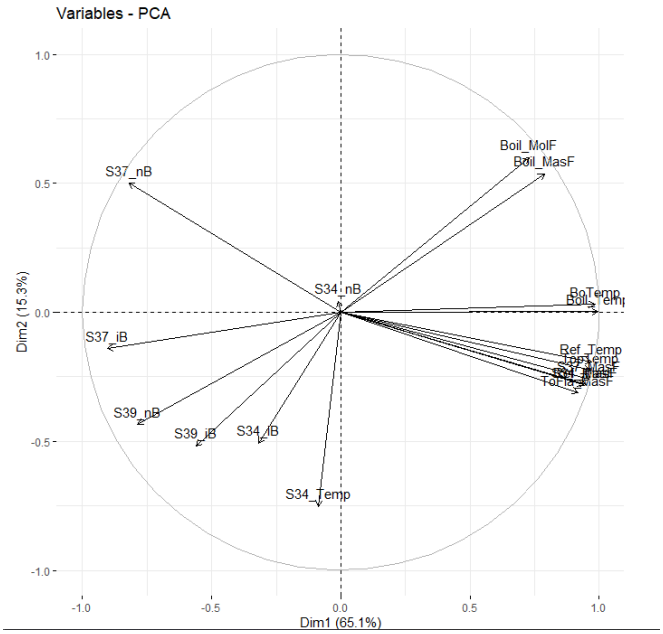

Figure 7. Factor Map (PCA) - Nested Data

To comprehend the PCA, other plots are performed and examined to identify the importance of data variables. The squared cosine plot explains the variable's quality, and the contribution plot explains the reliability and variability (See **Figure 8**, **Figure 9**, **Figure 10** and **Figure 11**). The plots show high quality for most of the variables, which increases the confidence of high correlation to the nonlinearity nature of the process. From PCA, the variables that are valuable to generate the neural network are the feed mass flow rate, butane mass flow rate, reflux mass flow rate, top temperature, bottom temperature, reflux temperature, and boilup temperature.

Figure 12 and **Figure 13** show the variable importance plot (VIP) that supports the PCA and indicates which variables are important regarding top product butane composition. Aspen-HYSYS are linked through an interface code to connect the simulated validated plant (HYSYS) to the artificial neural network-based soft sensor and to send data back and forth for better prediction.

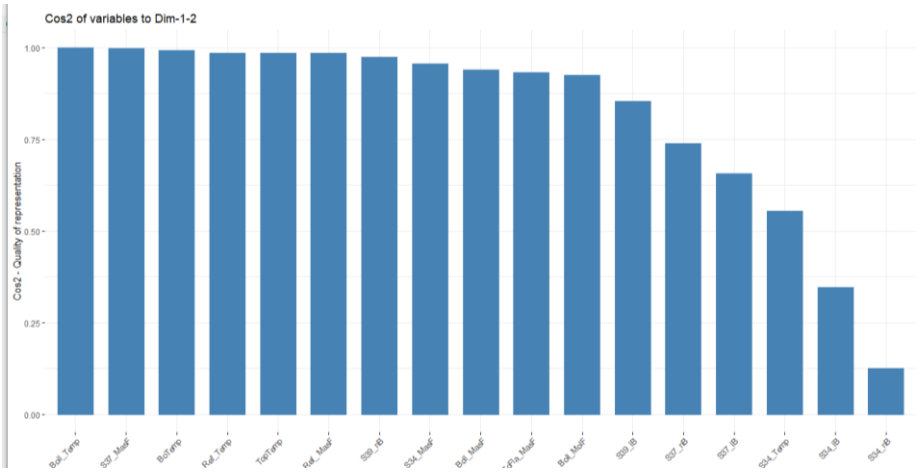

Figure 8. Quality of Representation vs. Variables (Nested Data)

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

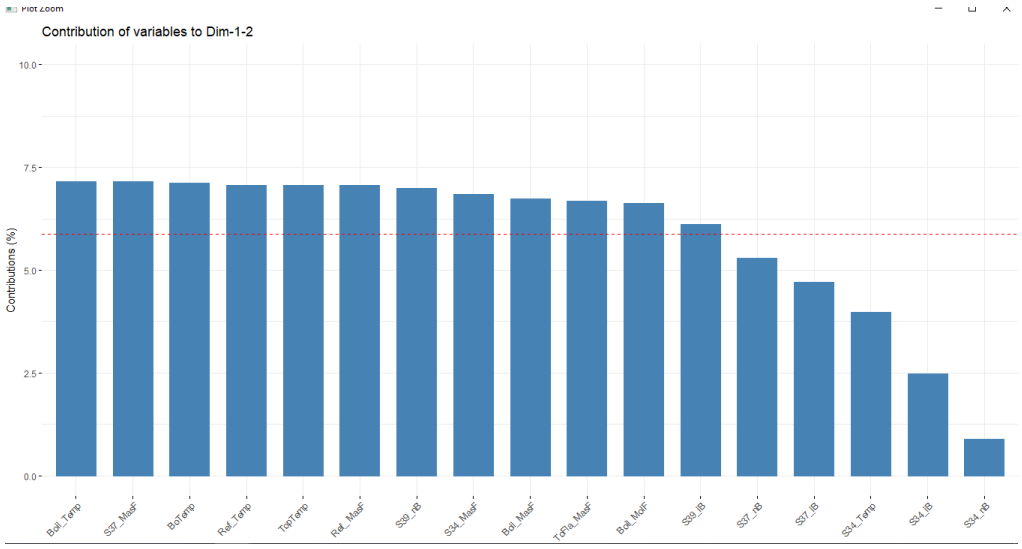

Figure 9. Contributions of Representation vs. Variables (Nested Data)

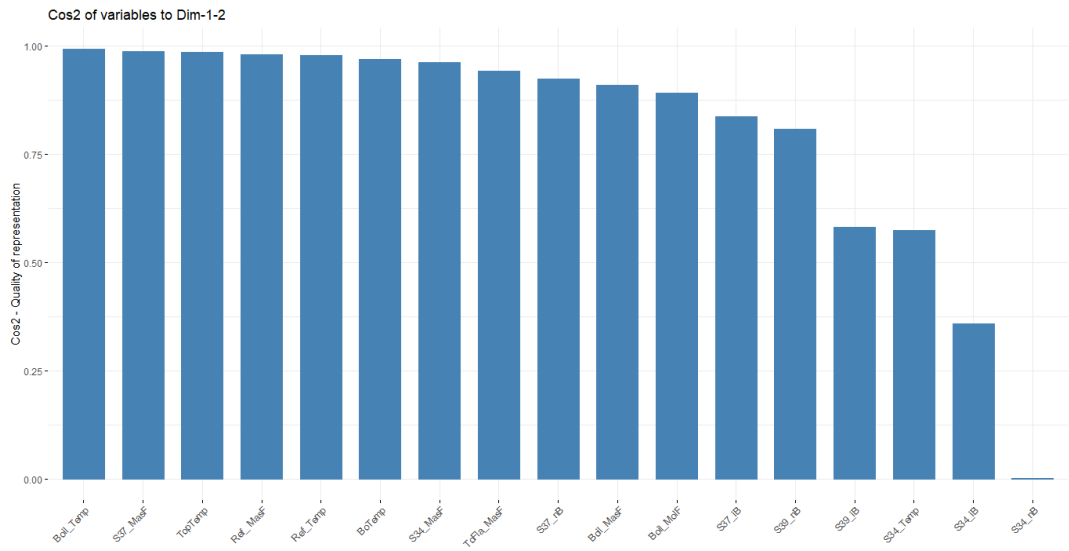

Figure 10. Quality of representation vs. Variables (Sensitivity Data)

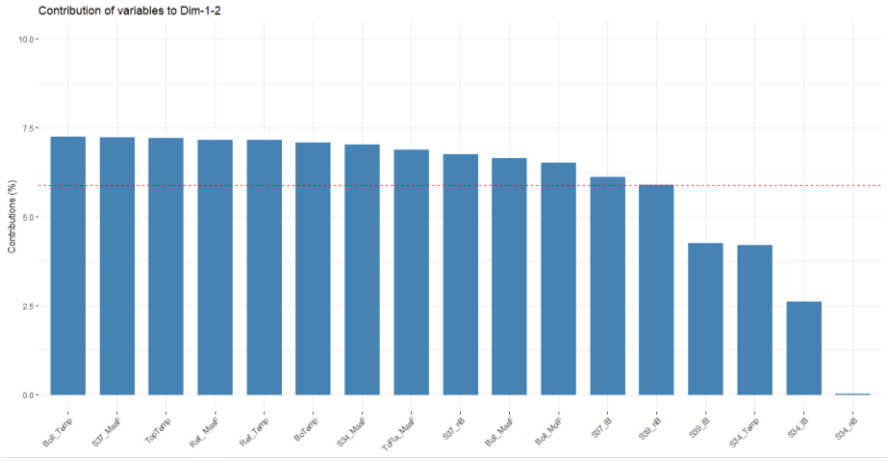

Figure 11. Contributions of representation vs. Variables (Sensitivity Data)

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

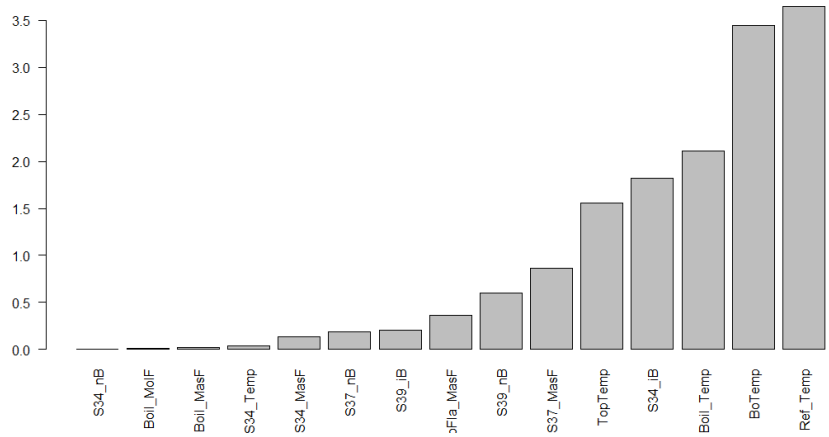

Figure 12. Variable Importance Plot for Top Product – i-Butane Composition

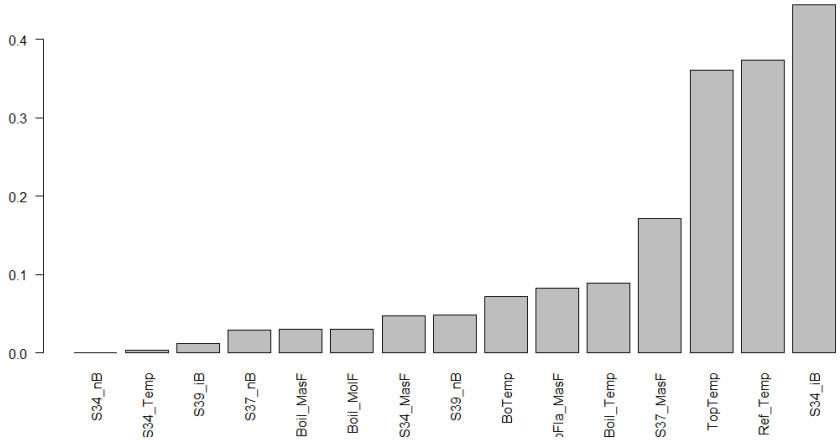

Figure 13. Variable Importance Plot for Top Product – n-Butane Composition

RESULTS AND DISCUSSION

Initially, the model in HYSYS is developed, simulated, and tested for validation. The process data imitates the debutanizer and shows a minimal percentage error. The data is used to build the soft sensor and predict i-butane and n-butane composition in the top product. The progress to find a better network has been implemented through analyzing the nonlinearity and the high correlation, as shown in **Figure 14**. There is a high correlation between the mentioned variables (boil-up temperature, reflux temperature, bottom temperature, top temperature, and feed mass flow rate) and the butane composition.

A network for inputs (feed mass flow rate, butane mass flow rate, reflux mass flow rate, top temperature, bottom temperature, reflux temperature, and boilup temperature) has been trained, tested, and judged with additional tests in MATLAB and shows the following results (see **Figure 15)**.

A better performance is retained with the feedforward network, which consisted of the inputs (feed mass flow rate, feed temperature, feed iBu, feed nBu, top temperature, bottom temperature, reflux temperature, and boilup temperature) and 2 layers with 8 hidden nodes and 2 delays (see **Figure 16)**.

The training stops at 230 epochs with 1.0711×10^{-5} of a mean square error (MSE) for validation, where the validation checks are 6. The regression for validation data is 0.99873, and the regression for all data is 0.99989. The regression indicates how much the output is equal to the target, where all the data should fall in a line of 450 (**Figure 17**).

The following graph (**Figure 18**) shows the time series response. All output and target data are plotted with respect to time. As the data are divided randomly, this plot demonstrates the validation, tests, and training points across the data, as well as the determination of the error between the target and the output. The error in the histogram chart

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(**Figure 19**) determines the instances of sample data points that fall in a certain range of error. The following histogram includes 20 bins, where each bin has a range of error equal to 0.0079. The autocorrelation of error is minimal (15×10^{-6}) as non-zero occurs at zero lag. The plot indicates that prediction error is not highly correlated, as it is within the confidence limit (95%), as shown in **Figure 20**.

Artificial neural network-based soft sensors should be tested to ensure they predict the composition faultlessly, to be able to increase monitoring quality and enhance effectiveness of control in the debutanizer column. The network is primarily tested in the MATLAB neural network time series toolbox, as the nested data is used to form the network. The sensitivity data is used as an additional test for the network. The data is carried out for a range of independent variables at each time and calculated respectively. 200 timesteps of data are implemented in the toolbox additional test data pane. The test proved that the network is accurate, as it showed a mean squared error of 3.1×10^{-6} and a regression of 0.9987. **Figure 21** shows the success of the network response.

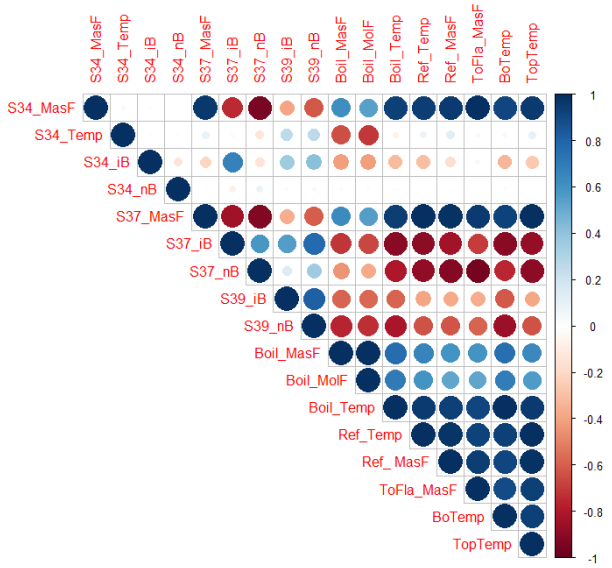

Figure 14. Nested Data Correlation Plot (Provide Legends)

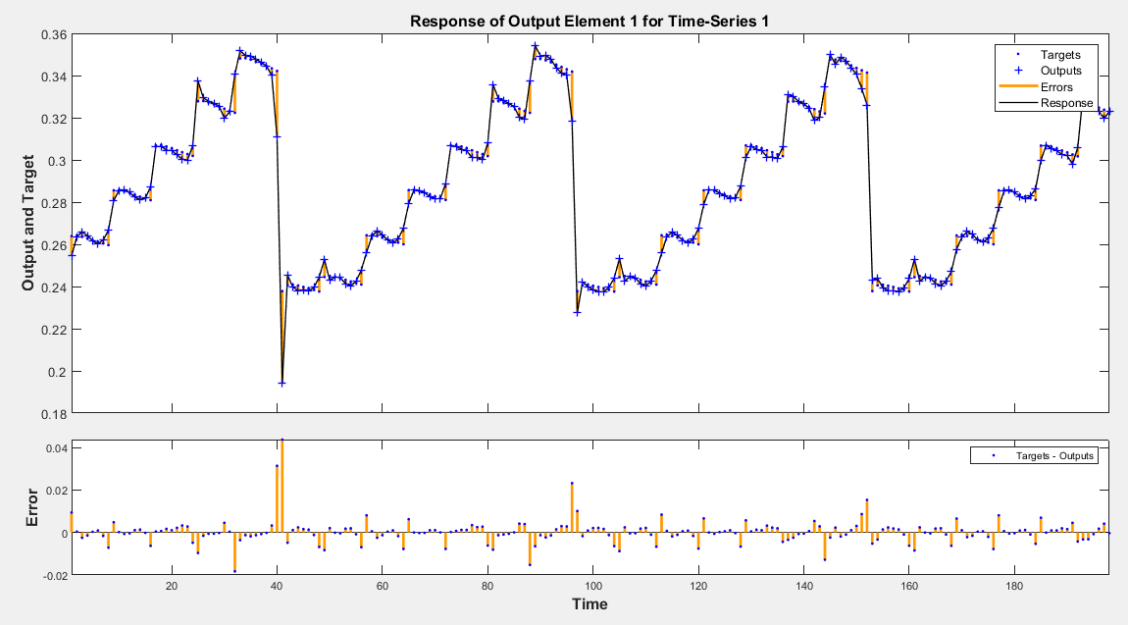

Figure 15. Neural Network Response Test Data

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

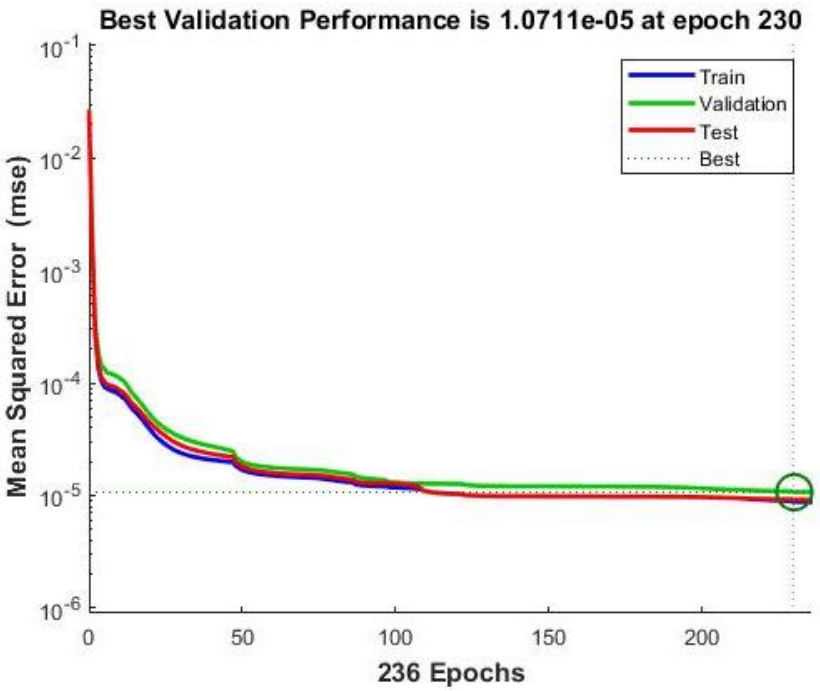

Figure 16. Neural Network Performance

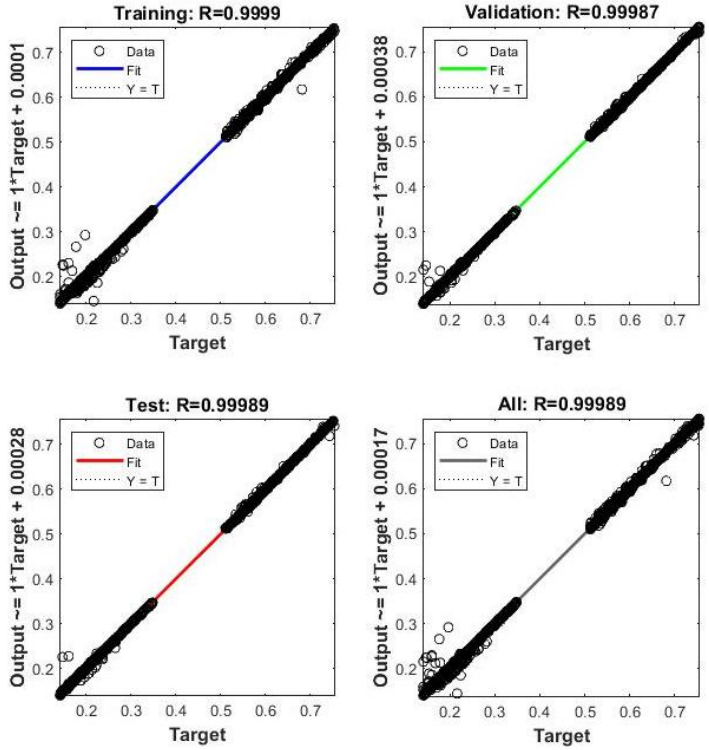

Figure 17. Neural Network Data Regression

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

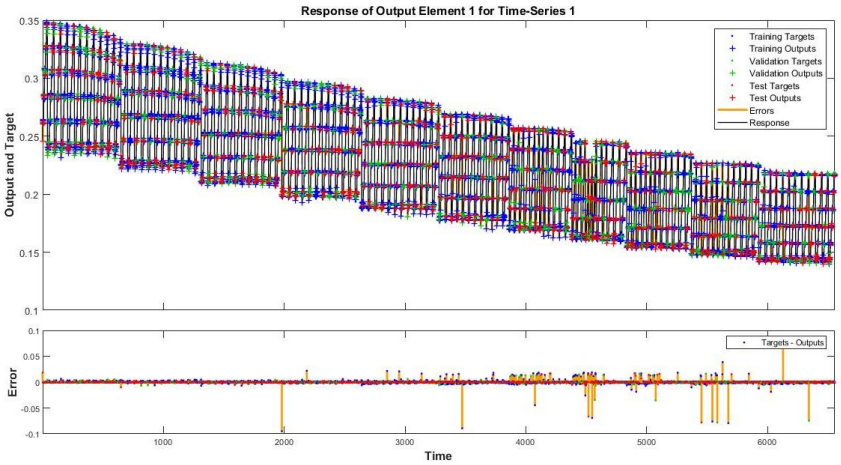

Figure 18. Neural Network Time Series Response

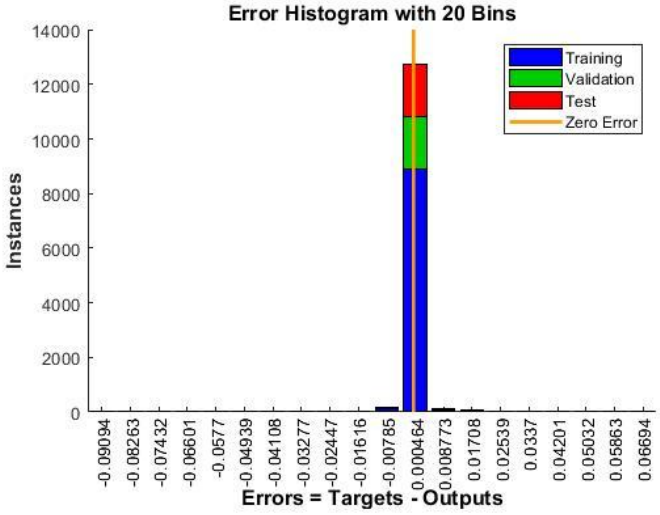

Figure 19. Error Histogram with 20 Bins

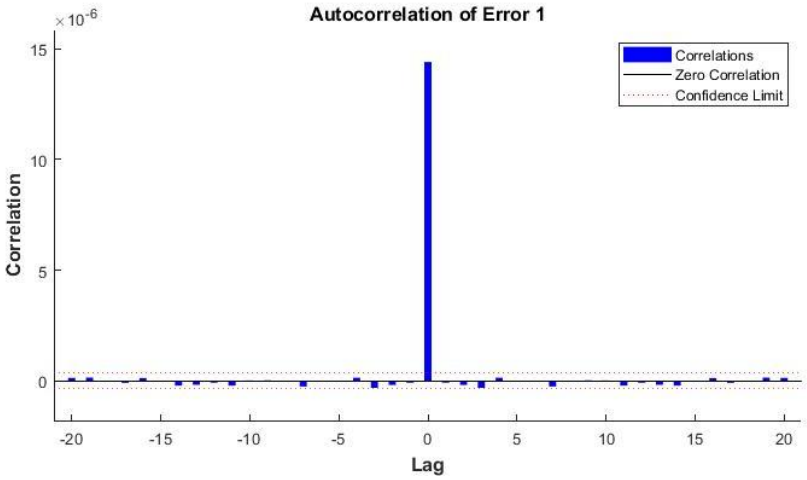

Figure 20. Autocorrelation Error

Journal of Information Systems Engineering and Management
2025, 10(61s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

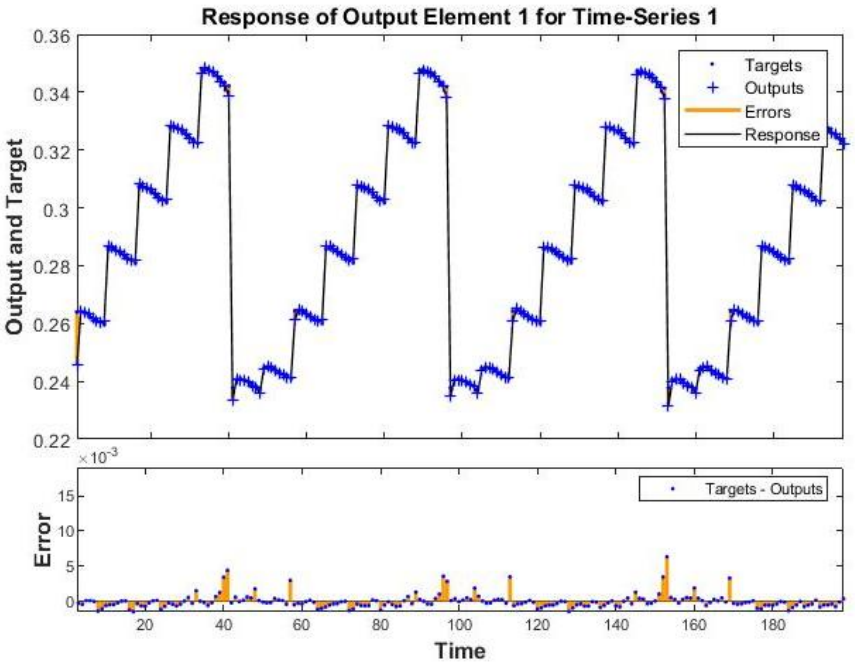

Figure 21. Neural Network Time Series Response (Additional Data Test)

CONFLICT OF INTEREST

There are no conflicts of interest.

CONCLUSIONS

With respect to the debutanizer process's control challenges due to delays in identifying the butane composition, a soft sensor based on an artificial neural network model was built to monitor the quality of the product. The debutanizer column was simulated by Aspen-HYSYS, and it was validated with real plant data. The simulated validated plant study was used to generate a wide data range to build the neural network model. The generated data were analyzed and amended by normalizing them and by applying the PCA analysis using RStudio to assist in selecting the input variables that were fed to train the NN in the MATLAB platform to read the required compositions.

The soft sensor was designed by the MATLAB time series neural network toolbox. High performance of composition prediction was achieved, with a minimal mean square error to the epoch number, as it shows how the data lines change as the epoch number increases in comparison with the best line of performance. As well, the neural network time series response plot, the error histogram, and the autocorrelation error plot have been considered to improve the soft-sensor model. Besides, the neural network data regression plot implies how the output data is equal to the target. The time series neural network is promising to be deployed and used to monitor the butane composition.

REFERENCES

[1] Aimin, M., Peng, L., & Lingjian, Y. (2015). Neighborhood preserving regression embedding based data regression and its applications on soft sensor modeling. Chemometrics and Intelligent Laboratory Systems, 147, 86–94. https://doi.org/10.1016/j.chemolab.2015.07.012

[2] Behnasr, M., & Jazayeri-Rad, H. (2015). Robust data-driven soft sensor based on iteratively weighted least squares support vector regression optimized by the cuckoo optimization algorithm. Journal of Natural Gas Science and Engineering, 22, 35–41. https://doi.org/10.1016/j.jngse.2014.11.017

[3] Hadian, M., Saryazdi, S. M. E., Mohammadzadeh, A., & Babaei, M. (2021). Application of artificial intelligence in modeling, control, and fault diagnosis. In J. Ren, W. Shen, Y. Man, & L. Dong (Eds.), Applications of Artificial Intelligence in Process Systems Engineering (pp. 255–323). Elsevier. https://doi.org/10.1016/B978-0-12-821092-5.00006-1

[4] Jaadi, Z. (2021). A step-by-step explanation of principal component analysis (PCA).
[5] Kadlec, P., Grbić, R., & Gabrys, B. (2011). Review of adaptation mechanisms for data-driven soft sensors. Computers & Chemical Engineering, 35(1), 1–24. https://doi.org/10.1016/j.compchemeng.2010.07.034
[6] Liu, Z., Ge, Z., Chen, G., & Song, Z. (2018). Adaptive soft sensors for quality prediction under the framework of Bayesian network. Control Engineering Practice, 72, 19–28. https://doi.org/10.1016/j.conengprac.2017.10.018
[7] Mohamed Ramli, N., Hussain, M. A., Mohamed Jan, B., & Abdullah, B. (2014). Composition prediction of a debutanizer column using equation based artificial neural network model. Neurocomputing, 131, 59–76. https://doi.org/10.1016/j.neucom.2013.10.039
[8] Mohler, I., Andrijić, Ž. U., & Bolf, N. (2011). Development of soft sensors for crude distillation unit control. IFAC Proceedings Volumes, 44(1), 5407–5412. https://doi.org/10.3182/20110828-6-IT-1002.01882
[9] Ramli, N. M., Hussain, M. A., & Jan, B. M. (2016). Multivariable control of a debutanizer column using equation based artificial neural network model inverse control strategies. Neurocomputing, 194, 135–150. https://doi.org/10.1016/j.neucom.2016.02.026
[10] Saptoro, A. (2014). State of the art in the development of adaptive soft sensors based on just-in-time models. Procedia Chemistry, 9, 226–234. https://doi.org/10.1016/j.proche.2014.05.027
[11] Shao, W., & Tian, X. (2017). Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development. Neurocomputing, 222, 91–104. https://doi.org/10.1016/j.neucom.2016.10.005
[12] Shao, W., Tian, X., & Chen, H. (2013). Adaptive anti-over-fitting soft sensing method based on local learning. IFAC Proceedings Volumes, 46(32), 415–420. https://doi.org/10.3182/20131218-3-IN-2045.00021
[13] Shi, X., & Xiong, W. (2020). Adaptive ensemble learning strategy for semi-supervised soft sensing. Journal of the Franklin Institute, 357(6), 3753–3770. https://doi.org/10.1016/j.jfranklin.2019.07.008
[14] Singh, H., Pani, A. K., & Mohanta, H. K. (2019). Quality monitoring in petroleum refinery with regression neural network: Improving prediction accuracy with appropriate design of training set. Measurement, 134, 698–709. https://doi.org/10.1016/j.measurement.2018.11.005
[15] Sun, K., Liu, J., Kang, J.-L., Jang, S.-S., Wong, D. S.-H., & Chen, D.-S. (2014). Soft sensor development with nonlinear variable selection using nonnegative garrote and artificial neural network. In J. J. Klemeš, P. S. Varbanov, & P. Y. Liew (Eds.), Computer Aided Chemical Engineering (Vol. 33, pp. 883–888). Elsevier. https://doi.org/10.1016/B978-0-444-63456-6.50148-4
[16] Urhan, A., & Alakent, B. (2020). Integrating adaptive moving window and just-in-time learning paradigms for soft-sensor design. Neurocomputing, 392, 23–37. https://doi.org/10.1016/j.neucom.2020.01.083
[17] Xiong, W., Li, Y., Zhao, Y., & Huang, B. (2017). Adaptive soft sensor based on time difference Gaussian process regression with local time-delay reconstruction. Chemical Engineering Research and Design, 117, 670–680. https://doi.org/10.1016/j.cherd.2016.11.020
[18] Yan, X., Yang, W., Ma, H., & Shi, H. (2012). Soft sensor for ammonia concentration at the ammonia converter outlet based on an improved group search optimization and BP neural network. Chinese Journal of Chemical Engineering, 20(6), 1184–1190. https://doi.org/10.1016/S1004-9541(12)60606-5
[19] Yu, F., Cao, L., Li, W., Yang, F., & Shang, C. (2020). Feature based causality analysis and its applications in soft sensor modeling. IFAC-PapersOnLine, 53(2), 138–143. https://doi.org/10.1016/j.ifacol.2020.12.111
[20] Yuan, X., Qi, S., Shardt, Y. A. W., Wang, Y., Yang, C., & Gui, W. (2020). Soft sensor model for dynamic processes based on multichannel convolutional neural network. Chemometrics and Intelligent Laboratory Systems, 203, 104050. https://doi.org/10.1016/j.chemolab.2020.104050
[21] Yuan, X., Qi, S., Wang, Y., & Xia, H. (2020). A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. Control Engineering Practice, 104, 104614. https://doi.org/10.1016/j.conengprac.2020.104614
[22] Zeng, L., & Ge, Z. (2021). Bayesian network for dynamic variable structure learning and transfer modeling of probabilistic soft sensor. Journal of Process Control, 100, 20–29. https://doi.org/10.1016/j.jprocont.2021.02.004