

Research

Artificial neural network modelling of natural gas dehydration process

Brandon Z. H. Goh¹ · Rafil Elyas² · Anuar Hamid² · Dominic C. Y. Foo¹

Received: 20 February 2024 / Accepted: 13 May 2024

Published online: 28 May 2024

© The Author(s) 2024 [OPEN](#)

Abstract

A multi-input–multi-output artificial neuron network (MIMO-ANN) model has been developed for process monitoring and improvement on a natural gas glycol dehydration process. The MIMO-ANN model was based on a steady-state process simulation model constructed in commercial software Aspen HYSYS. A set of training data was generated with the converged simulation model for the training of the MIMO-ANN model in Python. The process input of the model includes lean glycol recirculation rate and purity along with wet gas inlet pressure. On the other hand, the process output considered includes dehydrated gas water and aromatics content, dehydrated gas hydrate formation temperature and water dew point, stripping gas flow rate as well as reboiler duty. The overall mean squared error (MSE) of the MIMO-ANN model was calculated as 1.79. The best-fit line that highly overlaps with a 45° diagonal line is constructed with a correlation coefficient (R^2) score of more than 0.999 for all studied process output testing datasets. The mean absolute percentage error (MAPE) of the predicted outputs is generally less than 1%, except for dehydrated gas water dew point (16.63%) and stripping gas flow rate (11.55%), due to error predictions attempted on pseudo-zero values. This successfully displays an exceptional predictive performance of the MIMO-ANN model developed in this work which can be further deployed as an online dashboard for real-time monitoring tool.

1 Introduction

Natural gas is one of the well-established primary energy sources acquired from natural underground reservoirs. In comparison to other energy sources such as gasoline and diesel oils, natural gas possesses several superior properties such as cost-effectiveness, vast abundance, and lower emission rates that enable it to be more attractive when it comes to the application of energy generation [1]. Furthermore, the combustion of natural gas emits lesser greenhouse gas emissions together with minimal suspended solids, thereby inherently reducing air pollution as one of the global environmental concerns nowadays upon being utilised as a fuel source [2]. The global utilisation of natural gas demonstrated a compound annual growth rate of 2.87% from the year 2009 to 2019, in which the global natural gas consumption surged from 57 to 76 EJ after a decade [3]. This prominently displays a higher demand for natural gas as one of the heavily relied energy sources on Earth forecasted in the future. Therefore, extensive research has been carried out on natural gas processing due to its high potential in the application of energy generation as a relatively

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43938-024-00047-8>.

✉ Dominic C. Y. Foo, dominic.foo@nottingham.edu.my; Brandon Z. H. Goh, brandn1128@gmail.com; Rafil Elyas, rafil@east101.com; Anuar Hamid, anuar@east101.com | ¹Department of Chemical and Environmental Engineering, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia. ²East One-Zero-One Sdn Bhd, MR1-0205, Sri Acappella Commercial Annexe, 1, Jalan Lompat Tinggi 13/33, Tadisma Business Park, 40100 Shah Alam, Selangor, Malaysia.



clean energy source compared to other primary energy sources such as oil and coal that effectively exacerbate environmental pollution upon combustion to generate energy. Furthermore, natural gas is also one of the non-renewable energies that human heavily utilised in transitioning to net-zero carbon footprint target at the year 2050 [4].

Raw natural gas can be sourced from three types of underground reservoirs, namely oil wells, gas wells and condensate wells. Oil well produces natural gas known as associated gas that can either naturally exist as oil-free (free gas) or dissolved in crude oil (dissolved gas) [5]. On the other hand, gas wells produce solely raw natural gas and condensate wells yield natural gas containing semi-liquid hydrocarbon condensate along with a minute to an absent amount of crude oil which is known as non-associated gas [6].

Prior to natural gas extraction and processing, mixtures of light (gas) and heavy (oil) hydrocarbons are first extracted from a reservoir well followed by a slug catcher that distributes the outlet phases into gas along with light and heavy liquid phases under equilibrium conditions [7]. Once the gas exits the slug catcher, further downstream processes are mandatory to pre-treat the raw natural gas before sending it to export/reinjection/fuel gas. Natural gases are water-saturated as water is always present in the slug catcher during the separation process [8].

Natural gases that contain significant amounts of water content may cause severe problems such as pipeline corrosion and plugging [9]. The abundant presence of water in the natural gas may combine with carbon dioxide and hydrogen sulphide in the natural gas, leading to accelerated corrosion of the piping, lower flow efficiency due to slugging flow caused by liquid water condensation, and accumulation of liquid carryover may increase piping flow and operating pressure that results in a higher risk of equipment damage [10]. The presence of water vapour will prominently hamper the heating value of the natural gas as well [11].

Another critical issue regarding the presence of water in natural gas is the formation of hydrate solids that may cause pipeline plugging due to solid deposition in the interior of the pipelines [8]. Natural gas hydrates are a polyhedral solid structure of water (host) molecules caging gas (guest) molecules such as carbon dioxide, nitrogen, hydrogen sulphide and hydrocarbons such as methane, propane, and butane via Van der Waals forces. The hydrate structure consists of a highly stable cyclic cluster of water molecules held by hydrogen bonds, with a vacant cavity for guest molecules to be engaged [12]. Under high-pressure and low-temperature conditions, the host and guest molecules are forced to be in contact and engaged, forming solid hydrates with densities higher than fluid hydrocarbons thus depositing the pipelines and causing blockages [13].

Natural gas has to be dehydrated to overcome the above-mentioned issues. A typical acceptable water content is 4–7 lb/MMscf (million standard cubic feet of natural gas) for dehydrated natural gas after the absorption process [14]. Several techniques that have been established for natural gas dehydration purposes are absorption, adsorption, membrane separation and cooling [15]. Herein, the absorption technique is focused on and discussed in detail.

In typical natural gas dehydration via absorption processes, glycols such as diethylene glycol and tri-ethylene glycol (TEG) are usually employed as a water desiccant to dehydrate natural gas via a counter-current configuration in an absorber known as glycol contactor owing to their outstanding chemical affinity of water [14]. The absorption process is carried out physically through a strong hydrogen bonding between the glycol and water molecules that contain oxygen and hydrogen atoms [16]. TEG is commonly used industrially due to its low operating cost, superior water dew point depression, minimal recirculation losses, high hygroscopicity and chemical stability [6].

As the measurement of moisture properties of the dehydrated natural gas such as water content, water dew point, and hydrate formation temperature is often circuitous, tedious and time-consuming, the modelling of the dehydration process is often used to simplify and accelerate the effort in conducting process monitoring and improvement. Therefore, first principle models are often used to conduct process simulations in commercial software (e.g. Aspen HYSYS, Aspen Plus, UniSim Design). However, these models are highly dependent on computational calculations based on physical laws such as mass/energy balance, and equilibrium/kinetic-based models. This arises a bottleneck whereby a large amount of computational effort and time is required to perform process output calculations upon dynamic fluctuations land on process inputs. Thus, a need in further simplification of the process modelling is required.

In this regard, Artificial Neural Networks (ANN) emerge as one of the potential machine learning algorithms to serve as a reduced order model in providing real-time monitoring of the natural gas dehydration process. Complex calculations with various formulas based on physical laws are all simplified in terms of plain numbers by forcefully relating the process inputs and outputs into a simplified expression. The aftereffect, of course, sacrifices a minute degree of accuracy compared to first principle models but, drastically shortens the computational time of the model in predicting process outputs upon dynamic fluctuations on process inputs which is one of the utmost priorities in process monitoring and improvement.

In real-time process monitoring, the direction of the control actions taken (i.e., increase or decrease manipulated variables) is often highly prioritised compared to the degree of control accuracy at the given timestamp as process noises and fluctuations take place all the time. This results in an acceptable operating range at a given setpoint and therefore the priority of model accuracy level is much less prominent compared to model prediction time. However, the compromise of accuracy level should not be too large which may result in unreliable model results and therefore model validation and evaluation must be carried out to corroborate its validity.

Numerous modelling studies by ANN have been carried out on TEG natural gas dehydration such as dehydrated natural gas equilibrium water dew point prediction via particle swarm optimisation-artificial neural network [11], feedforward artificial neural network model with Bayesian training [17] and multi-layer perceptron artificial neural network model [18]. Another work has also reported a classification approach artificial neural network on predicting the output class of natural gas water content in which the model classifies the outcomes of the water content into two separate classes, namely Class 0 and Class 1. Class 0 indicates water content < 7 lb/MMscf that satisfies the water content specification while Class 1 implies water content > 7 lb/MMscf that does not meet the specification [19]. The above work successfully predicted the desired outcomes with outstanding correlation scores, however, suffers from one limitation which is not designated for multi-output predictions on other dehydrated natural gas properties. Important dehydrated natural gas properties such as water content, aromatics content, hydrate formation temperature and water dew point are crucial to be determined in order to meet the sales gas specifications and to avoid pipeline plugging issues.

The above studies have successfully demonstrated the prediction of single-process output prediction when it comes to dehydrated gas moisture properties. However, this greatly limits the application of these models in the potential of being employed as a real-time process monitoring and improvement tool. A robust process monitoring and improvement tool should be able to consider broader ranges of process inputs and outputs simultaneously as one process output can be influenced by several process inputs, and the competency in predicting multiple outputs endows the prediction model to be deployed as an online dashboard tool to perform real-time monitoring and improvement on the natural gas dehydration process. This facilitates real-time control of the process and enhances the efficiency and productivity of the plant.

Therefore, this work provides a novel insight into the development of a multi-input multi-output ANN (MIMO-ANN) model by resembling multiple single input–single output ANN models into one to overcome the limitations of having single output prediction from the above work. Additionally, this work may also unravel the potential of the MIMO-ANN model to be applied as an inferential sensor which creates a digital twin of the TEG natural gas dehydration process to perform real-time predictions on measurable and unmeasurable process output predictions.

Herein, multiple outputs of the dehydrated natural gas properties namely (1) dehydrated gas water content, (2) dehydrated gas aromatics content, (3) dehydrated gas hydrate formation temperature, (4) dehydrated gas water dew point, along with other interesting parameters such as (5) stripping gas flow rate and (6) reboiler duty.

To the best of the authors' knowledge, the absence of ANN multiple output predictions on the TEG natural gas dehydration process remains to date. Thus, this work can provide novelty that unravels potential insights on multiple output prediction ANN models on other industrial processes by venturing into the development of a MIMO-ANN model structure in predicting the outputs of interest in a TEG natural gas dehydration process.

The structure of this paper is organised whereby a problem statement is first provided in Sect. 2. The introduction of the TEG natural gas dehydration process and ANN theory along with its process simulation, ANN development and ANN model evaluation methodologies are elucidated in Sect. 3. The results of the case study and the MIMOANN model development are discussed in Sect. 4. The MIMO-ANN model evaluation, prediction performance and limitations are discussed in Sect. 5 respectively. A conclusion in Sect. 6 epitomises the overall work in this paper.

2 Problem statement

The problem statement of this work is elucidated below:

1. Process input of the TEG natural gas dehydration process fluctuates which induces uncertainty to the process outputs such as dehydrated natural gas properties.
2. Prediction of the dehydrated natural gas properties (water content, aromatics content, hydrate formation temperature and water dew point) along with other parameters such as stripping gas flow rate and reboiler duty is required.

Herein, the work aims to predict multiple outputs of the TEG natural gas dehydration process by developing a MIMO-ANN model which can be potentially installed as an inferential sensor on the process line.

3 Methodology

3.1 Process description of TEG dehydration process

Industrial TEG natural gas dehydration processes normally consist of two parts, which are TEG dehydration and TEG regeneration, as shown in Fig. 1. In the TEG dehydration section, the lean glycol and wet natural gas are fed counter-currently into a glycol contactor where water absorption is carried out. The rich glycol that flows out from the bottom of the glycol contactor then is subjected to a series of separation units (glycol still unit and flash drum) for the regeneration of lean glycol, before being recycled to the dehydration unit. The installation of filtration units such as sock filters and activated carbon filters after the glycol flash drum is implemented to further eliminate the accumulation of solid particles that may clog up pipelines of the downstream processes.

The process is then simulated in Aspen HYSYS and cross-validated with the given actual plant data to ensure that the actual dehydration process in the plant can be confidently resembled in the process simulation. Subsequently, a nested case study is used to generate large datasets by considering all possible combinations of process input conditions that yield various process outputs used for ANN training.

3.2 Artificial Neural Network (ANN) model development and evaluation

A multi-input multi-output artificial neural network (MIMO-ANN) model is created in Keras Python and trained by the dataset generated by Aspen HYSYS. A loss profile that consists of training and validation losses of the MIMO-ANN model is plotted and evaluated throughout the training process to ensure that the model is well-trained and no overfitting is observed (i.e., the loss profile eventually reaches a plateau with increasing training epochs).

The trained MIMO-ANN model is then used to predict unseen data inputs that are held back from the training process and linear regression is performed between the predicted and actual process outputs on both training and testing datasets. The correlation coefficient (R^2) score is calculated based on the regressed linear fit model to evaluate the prediction

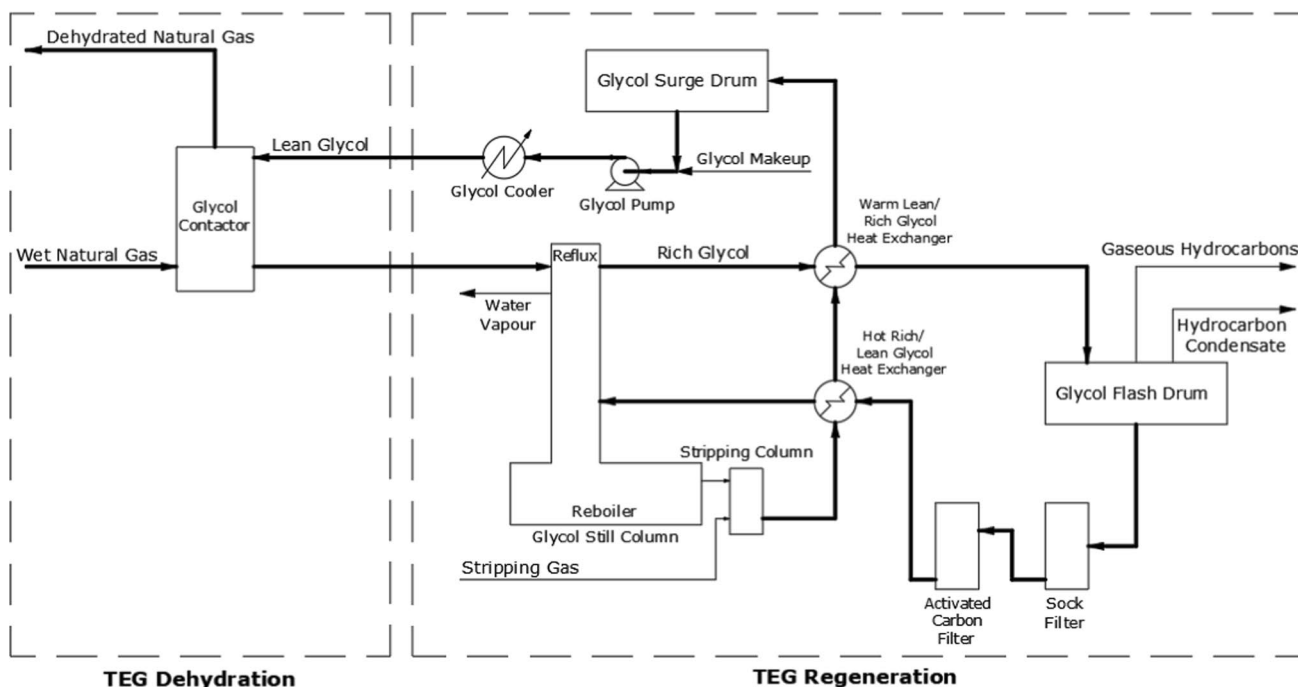


Fig. 1 Process flow diagram of a simplified TEG natural gas dehydration process

performance of the MIMO-ANN model. Percentage error and mean absolute percentage error (MAPE) are then used to quantify the error between the predicted and the actual data outputs.

Detailed information on R^2 score, percentage error and MAPE calculation can be found in the Supplementary file.

The overall methodology of the report is depicted in Fig. 2.

4 Case study

A case study of a natural gas dehydration plant located in Malaysia is simulated using Aspen HYSYS V11 [20], with the simulation flowsheet shown in Fig. 3. Peng-Robinson Equation of State is applied in this case study. The wet natural gas and recirculating lean glycol composition are elucidated in Table 1.

The wet natural gas (WG1) is fed into the bottom of the glycol contactor (C-100) at 39.85 °C and 119.61 bar(a). A water saturator unit (SAT-100) is implemented to ensure the natural gas is saturated with water at the given composition and thermodynamic conditions to resemble scenarios whereby the natural gas is saturated with water after passing through an upstream production separator that separates the full well-stream fluid into gas, oil and aqueous phases [8]. Lean glycol (LG-1) enters the top of C-100 to dehydrate the wet natural gas. The dehydrated natural gas (DG1) then exits from the top of the contactor to be delivered for further downstream processing. The entire TEG dehydration part is carried out under high-pressure and low-temperature conditions to favour the water absorption in C-100. It is important to note that the water absorption of the wet natural gas is carried out under hydrocarbon supercritical conditions in this case study.

The rich glycol (RG1) exiting C-100 is then let down and preheated by a glycol reflux condenser (E-C101) and heat exchangers (E-100 and E-101). The E-C101 created in the simulation is used to artificially resemble the integration of condenser duty from glycol still column (C-101). The preheated rich glycol then enters a glycol flash drum (V-100) whereby gaseous hydrocarbons, rich glycol, and hydrocarbon condensates are separated into three individual phases. It is noted that for the compositions in this study, there is no flow of hydrocarbon condensate. The rich glycol is then further let down and preheated to enter C-101 to boil off the remaining water contained in the rich glycol. The reboiler temperature is controlled within the decomposition temperature limit of TEG (206.67 °C) [14]. The boil-off water vapour (OG1) is then cooled before being sent to the flare system. The concentrated glycol (LG7) exits from the bottom C-101 at 98.28 wt% and is further purified in a stripping column (C-102) to achieve a target purity of 98.56 wt%. The stripping gas utilised is a minute fraction of the dehydrated natural gas (0.00153%) that is recycled from a dry gas - stripping gas splitter (DGSG-SPLIT) unit. The entire TEG regeneration process is carried out under low-pressure and high-temperature conditions to facilitate water removal in the glycol still column. A makeup glycol stream (MG) consists of pure TEG under the conditions of 30 °C and 1.38 bar(a) is used to replenish any loss of TEG in the dehydration system.

The hot lean glycol (LG6) is then cooled through a series of heat exchangers (E-100 and E-101) and a glycol cooler (E-102). A glycol circulation pump (P-001) is used to pressurise the lean glycol (LG2) back to complete the recirculation loop for the dehydration process.

The detailed column settings for the glycol contactor, glycol still column and stripping column are given in Table 2. Table 3 elucidates the assumed pressure drop and temperature change for the heat exchange units.

The dehydrated natural gas stream (DG1) obtained from the simulation is compared with the original plant data as shown in Table 4 to ensure the simulation results successfully resemble the data of the natural gas dehydration plant. The resulting simulated stream is similar (within 0.03% error) to that of the case study plant data in terms of thermodynamic properties, molecular weight, composition and flow rate. Thus, it can be concluded that the actual plant dehydration process can be resembled in the process simulation conducted in Aspen HYSYS. The process simulation results for the rest of the process streams are appended in Table S1 in the supplementary file.

Herein, the process input studies are (1) lean glycol recirculation rate, (2) lean glycol purity (TEG mass fraction) and (3) wet gas inlet pressure. On the other hand, the process output studies are (1) dehydrated gas water content, (2) dehydrated gas aromatics content, (3) dehydrated gas hydrate formation temperature, (4) dehydrated gas water dew point, (5) stripping gas flow rate and (6) reboiler duty. A base case is constructed based on given field data, and the values of process inputs and outputs of interest are summarised in Table 5.

4.1 Analysis of process variables

A nested case study is set up to study the effect of process inputs on the process outputs. The variables are monitored to ensure the results of the process simulation are logical and do not deviate from conventional engineering knowledge.

Fig. 2 Overall methodology for process simulation and MIMO-ANN model training of the TEG natural gas dehydration process

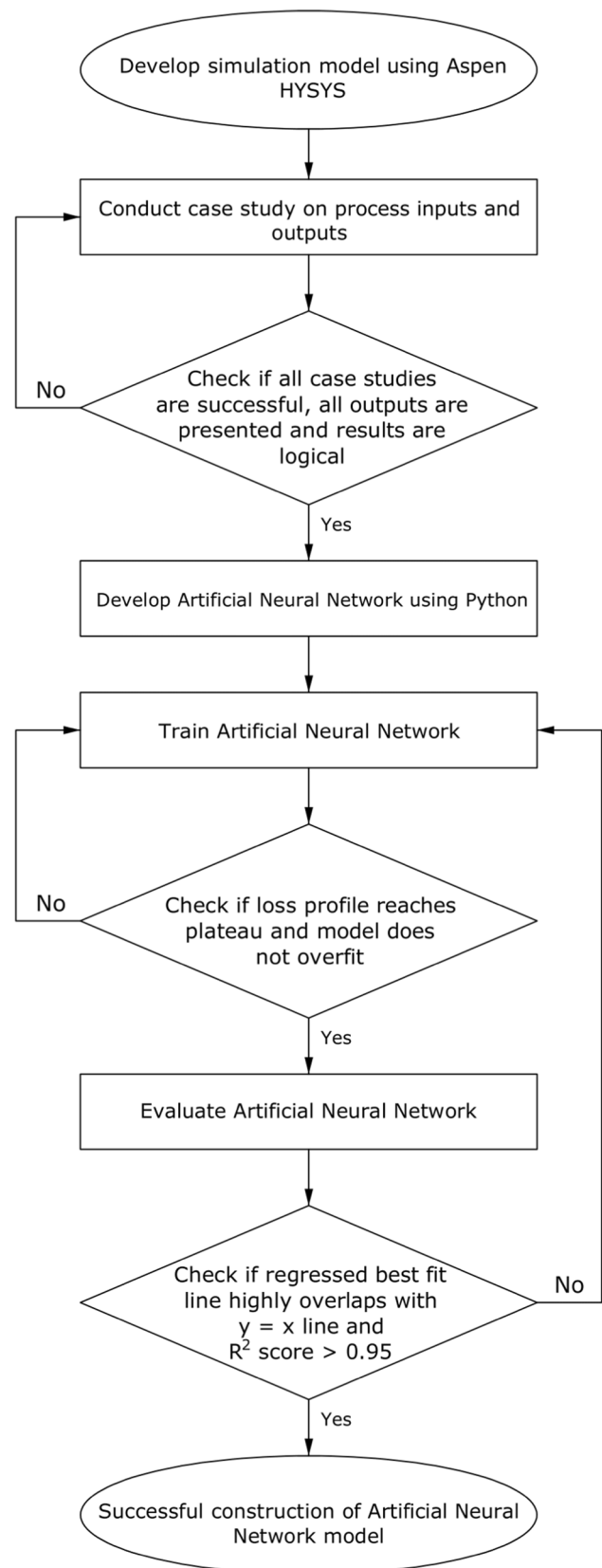


Table 1 Wet natural gas and recirculating lean glycol composition

| Component | Composition (%-wt) |
|---------------------------|--------------------|
| Wet natural gas | |
| Methane | 70.40 |
| CO ₂ | 9.00 |
| Ethane | 6.40 |
| Propane | 6.26 |
| <i>n</i> -Butane | 2.13 |
| <i>i</i> -Butane | 1.83 |
| <i>i</i> -Pentane | 0.99 |
| Nitrogen | 0.78 |
| <i>n</i> -Pentane | 0.66 |
| <i>n</i> -Hexane | 0.57 |
| Benzene | 0.32 |
| Mycyclopentane | 0.17 |
| Mycyclohexane | 0.15 |
| <i>n</i> -Heptane | 0.10 |
| H ₂ O | 0.09 |
| Cyclohexane | 0.09 |
| Toluene | 0.05 |
| Total | 100.00 |
| Recirculating lean glycol | |
| TEGlycol | 98.56 |
| H ₂ O | 1.09 |
| Mycyclohexane | 0.12 |
| <i>n</i> -Hexane | 0.07 |
| <i>n</i> -Heptane | 0.06 |
| Mycyclopentane | 0.03 |
| Cyclohexane | 0.03 |
| <i>n</i> -Pentane | 0.01 |
| <i>i</i> -Pentane | 0.01 |
| <i>n</i> -Butane | 4.10E-03 |
| Benzene | 3.61E-03 |
| Methane | 3.31E-03 |
| <i>i</i> -Butane | 1.96E-03 |
| Toluene | 1.72E-03 |
| Ethane | 1.37E-03 |
| CO ₂ | 1.10E-03 |
| Propane | 7.22E-04 |
| Nitrogen | 6.21E-05 |
| Total | 100.00 |

Table 2 Simulation setting for column units

| Column units | Number of stages | Settings |
|-----------------------------|------------------|--|
| Glycol contactor (C-100) | 3 | Top pressure: 119.11 bar(a) Bottom pressure: 119.61 bar(a) |
| Glycol still column (C-101) | 2 | Top pressure: 1.42 bar(a) Bottom pressure: 1.44 bar(a) Top temperature: 107.80 °C Bottom temperature: 204.40 °C |
| Stripping column (C-102) | 1 | Top pressure: 1.44 bar(a) Bottom pressure: 1.44 bar(a) |

Table 3 Simulation setting for heat exchange units

| Heat exchange units | Settings |
|--|--|
| Hot rich/lean glycol heat exchanger (E-100) | Shell inlet/outlet: RG5/RG6 Tube inlet/outlet: LG6/LG5 Shell side temperature: 79.99 to 165.00 °C Tube side temperature: 202.70 to 108.40 °C Shell side pressure drop: 0.70 bar(a) Tube side pressure drop: 0.03 bar(a) |
| Warm lean/rich glycol heat exchanger (E-101) | Shell inlet/outlet: RG3/RG4 Tube inlet/outlet: LG5/LG4 Shell side temperature: 53.73 to 80.00 °C Tube side temperature: 108.40 to 78.40 °C Shell side pressure drop: 0.70 bar(a) Tube side pressure drop: 0.03 bar(a) |
| Glycol reflux condenser (E-C101) | Inlet/Outlet: RG2/RG3 Temperature rise: 49.08 to 53.73 °C Pressure drop: 0.60 bar(a) |
| Glycol cooler (E-102) | Inlet/Outlet: LG2/LG1 Temperature drop: 74.49 to 45.00 °C Pressure drop: 0.60 bar(a) |

The effect of increasing lean glycol purity is similar to that of increasing lean glycol recirculation rate. However, the reboiler duty decreases at elevated lean glycol purity which is attributed to the enhanced column feed temperature (or feed enthalpy) after isenthalpic valve throttling. This results in lesser reboiler duty required to maintain the bottom temperature at the required set point. The increase in feed enthalpy can be explained by the lesser vapour fraction of the column feed after valve throttling at higher lean glycol purity (note that glycol is heavier than water). Therefore, more heat is trapped in the liquid fraction of the feed mixture due to its higher heat capacity thus resulting in a lower temperature drop (higher temperature) in the feed mixture after the valve throttling process. It is noted that the minute increase in lean glycol purity which increases the glycol mass flow at the bottom of the column does not affect the reboiler duty demand (as discussed earlier) which is due to the enhanced feed enthalpy having a more pronounced effect in decreasing the heating requirement of the reboiler.

Similarly, as more glycol is being separated at the bottom column and has a higher lean glycol purity requirement, more stripping gas is required to strip off excess water to attain the target lean glycol purity.

The effect of lean glycol recirculation rate and purity on dehydrated gas water and aromatics content, hydrate formation temperature, water dew point, stripping gas flow rate and reboiler duty are shown in Figs. 4 and 5.

4.1.2 Effect of wet gas pressure

Increasing wet natural gas pressure favours the water absorption process, resulting in decreasing water content of the dehydrated natural gas. However, the aromatics absorption is unfavoured under high pressures (> 63.00 bar(a)), which may be due to the decreased solubility of aromatics in lean glycol upon elevating pressure under the hydrocarbon supercritical state of the wet natural gas inlet as shown in Fig. 6. Similarly, the decreased water content in the dehydrated natural gas leads to depressed gas water dew point and lower hydrate formation temperature.

As lesser aromatics content is absorbed under elevated wet natural gas pressure, the purity of the rich glycol increases after exits the bottom of the glycol contactor. This arises the same consequence as increasing lean glycol purity in which the vapour fraction of the column feed decreases during the valve throttling process and thus has an elevated feed enthalpy which requires lesser reboiler duty to maintain the bottom temperature at the desired set point.

Additionally, lesser stripping gas is also required to strip off water from the lean glycol to achieve target purity due to higher initial lean glycol purity before entering the stripping column. The effects of lean glycol recirculation rate and purity on dehydrated gas water and aromatics content, hydrate formation temperature, water dew point, stripping gas flow rate and reboiler duty are shown in Fig. 7.

Table 4 Comparison between the simulation results and the case study plant data for dehydrated natural gas stream DG1

| Name | Simulated DG1 | Plant data DG1 |
|--|---------------|----------------|
| Vapour fraction | 1.00 | 1.00 |
| Temperature (C) | 40.45 | 40.42 |
| Pressure (bar) | 119.11 | 119.11 |
| Molar flow (kgmole/h) | 24,520.98 | 24,537.04 |
| Mass flow (kg/h) | 481,329.33 | 482,087.19 |
| Molecular weight | 19.63 | 19.65 |
| Heat flow (kW) | -623,858.53 | -624,000.00 |
| Mass density (kg/m ³) | 114.85 | 115.05 |
| Mass heat capacity (kJ/kg-C) | 2.98 | 2.98 |
| Thermal conductivity (W/m-K) | 0.05 | 0.05 |
| Std gas flow (STD_m ³ /h) | 579,784.74 | 580,164.62 |
| Actual gas flow (ACT_m ³ /h) | 4190.98 | 4190.26 |
| Viscosity (cP) | 0.02 | 0.02 |
| Mass lower heating value (kJ/kg) | 44,401.17 | 44,398.98 |
| Z Factor | 0.78 | 0.78 |
| Master comp mass frac (TEGlycol) | 0.0000 | 0.0000 |
| Master comp mass frac (H ₂ O) | 0.0001 | 0.0001 |
| Master comp mole frac (TEGlycol) | 0.0000 | 0.0000 |
| Master comp mole frac (H ₂ O) | 0.0001 | 0.0001 |
| Master comp mole frac (Nitrogen) | 0.0055 | 0.0055 |
| Master comp mole frac (CO ₂) | 0.0402 | 0.0402 |
| Master comp mole frac (H ₂ S) | 0.0000 | 0.0000 |
| Master comp mole frac (Methane) | 0.8637 | 0.8633 |
| Master comp mole frac (Ethane) | 0.0417 | 0.0418 |
| Master comp mole frac (Propane) | 0.0279 | 0.0280 |
| Master comp mole frac (<i>i</i> -Butane) | 0.0061 | 0.0062 |
| Master comp mole frac (<i>n</i> -Butane) | 0.0071 | 0.0072 |
| Master comp mole frac (<i>i</i> -Pentane) | 0.0027 | 0.0027 |
| Master comp mole frac (<i>n</i> -Pentane) | 0.0018 | 0.0018 |
| Master comp mole frac (<i>n</i> -Hexane) | 0.0013 | 0.0013 |
| Master comp mole frac (<i>n</i> -Heptane) | 0.0002 | 0.0002 |
| Master comp mole frac (Benzene) | 0.0008 | 0.0008 |
| Master comp mole frac (Toluene) | 0.0001 | 0.0000 |
| Master comp mole frac (<i>m</i> -Xylene) | 0.0000 | 0.0001 |
| Master comp mole frac (<i>o</i> -Xylene) | 0.0000 | 0.0000 |
| Master comp mole frac (Cyclopentane) | 0.0000 | 0.0000 |
| Master comp mole frac (Mycyclopentane) | 0.0004 | 0.0004 |
| Master comp mole frac (Cyclohexane) | 0.0002 | 0.0002 |
| Master comp mole frac (Mycyclohexane) | 0.0003 | 0.0000 |
| Total | 1.0000 | 0.9999 |

4.2 MIMO-ANN model development

A detailed explanation of the ANN model concept can be found in section S1.1 and illustrations are shown in Figure S1 in supplementary file.

An ANN model with a multi-input–multi-output (MIMO) structure is constructed via Python Keras Functional API library package. The MIMO-ANN structure along with the input and output layer terms are defined in Fig. 8 and Table 7 respectively.

Table 5 Process simulation inputs and outputs obtained under a base case

| | Values | Units |
|--|----------|-------------------|
| Process inputs | | |
| Lean glycol recirculation rate | 9.782 | m ³ /h |
| Lean glycol purity | 0.9856 | – |
| Wet gas pressure | 119.61 | bar(a) |
| Process outputs | | |
| Dehydrated gas water content | 5.017 | lb/MMSCF |
| Dehydrated gas aromatics content | 869.7 | ppm |
| Dehydrated gas hydrate formation temperature | 1.02 | °C |
| Dehydrated gas water dew point | 8.69 | °C |
| Stripping gas flow | 7.43E–03 | MMSCFD |
| Reboiler duty | 589.1 | kW |

LG Lean glycol, WG Wet gas, DG Dehydrated gas, SG Stripping gas

Table 6 Case study setup with lower bound, upper bound and step size of process inputs

| Process inputs | Lower bound | Upper bound | Step size | Steps |
|-----------------------|-------------|-------------|-----------|-------|
| LG recirculation rate | 6 | 12 | 0.4286 | 15 |
| LG purity | 0.9850 | 0.9900 | 3.57E–04 | 15 |
| WG pressure | 80 | 120 | 2.86 | 15 |

4.2.1 Hidden layer, nodes and activation functions

The number of hidden layers, nodes and type of activation functions in each hidden layer is determined via a trial-and-error method to obtain satisfactory results.

Three hidden layers are used for all the process outputs with a descending node number of 16, 8, and 4 for better results. Rectifying Linear Unit (ReLU) is used as the activation function in all the hidden layers whereas the first, second, and third hidden layer activation functions for Dep_v2 are ReLU, hyperbolic tangent (tanh) and Leaky ReLU respectively. For all the output layers, the activation functions implemented are Leaky ReLU. The total number of trainable parameters for the MIMO ANN model is 1446. Detailed illustrations of other activation functions can be found in Figure S2 in the supplementary file.

4.2.2 Loss function

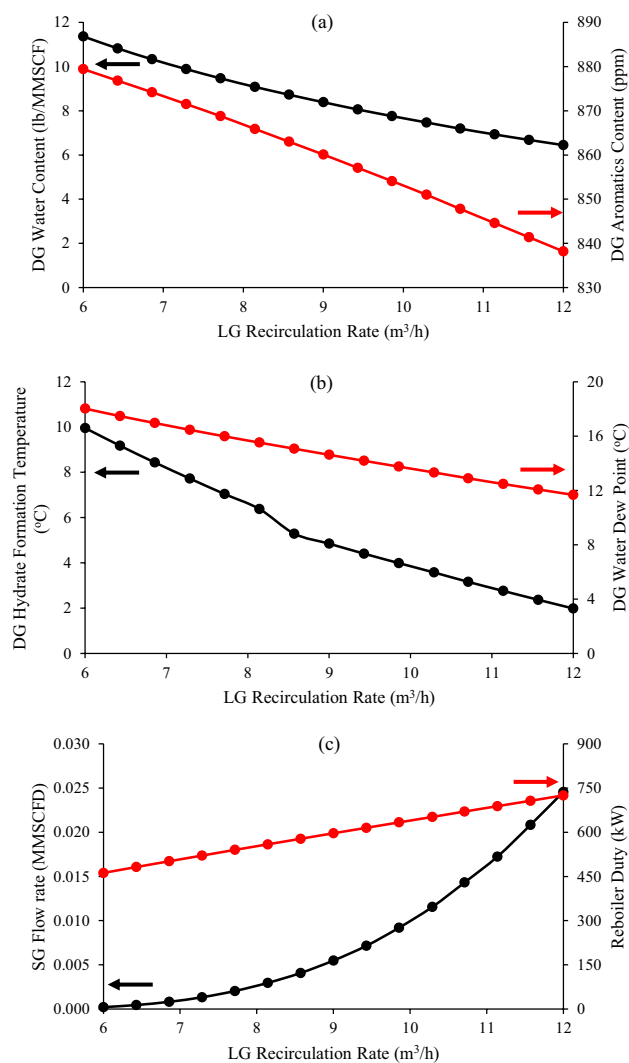
Mean squared error (MSE) is chosen as the loss function in the ANN model training. MSE measures the average squared difference between the estimated value and the actual value, which is useful in detecting large errors. Large errors are heavily penalised due to the quadratic exponential in the MSE function, therefore is beneficial when it comes to detecting outliers and minimising large errors compared to other loss functions [21].

The formula for calculating MSE is shown in Eq. 1.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

where n , Y_i , \hat{Y}_i indicates the number of data points, observed values and predicted values respectively.

Fig. 4 Effect of lean glycol recirculation rate on **a** dehydrated gas water content and aromatics content, **b** dehydrated gas hydrate formation temperature water dew point, **c** stripping gas flow rate, and reboiler duty



4.2.3 Data splitting

The 3375 sets of training data from Aspen HYSYS are divided into 2025 for training, 675 for validation and 675 for testing (unseen) datasets with a splitting ratio of 60-20-20 respectively. The random splitting seed generator is set to a constant integer of '1' to fix the random splitting of the training and testing datasets whenever the Python code is initiated.

4.2.4 Dropout

A dropout rate of 0.5 is used for all the hidden layers (Hidden_Layer_1, Hidden_Layer_1a to Hidden_Layer_6, Hidden_Layer_6a) according to normal practice, indicating that there is a 50% chance that any nodes presented in the hidden layers will be ignored during the training process. A detailed explanation of dropout can be found in Figure S4 in the supplementary file.

4.2.5 Optimiser and learning rate

Adaptive moment estimation (Adam) algorithm is used to optimise the gradient descent calculation in this work. The Adam optimiser accelerates the gradient descent algorithm via a combination of two gradient descent

Fig. 5 Effect of lean glycol purity on **a** dehydrated gas water content and aromatics content, **b** dehydrated gas hydrate formation temperature water dew point, **c** stripping gas flow rate, and reboiler duty

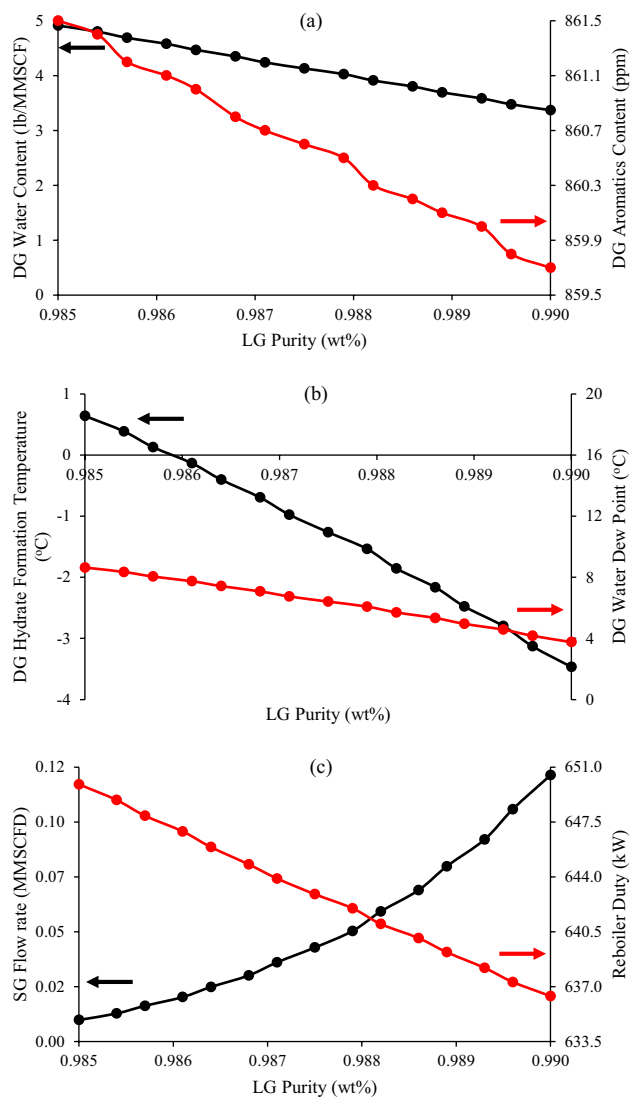
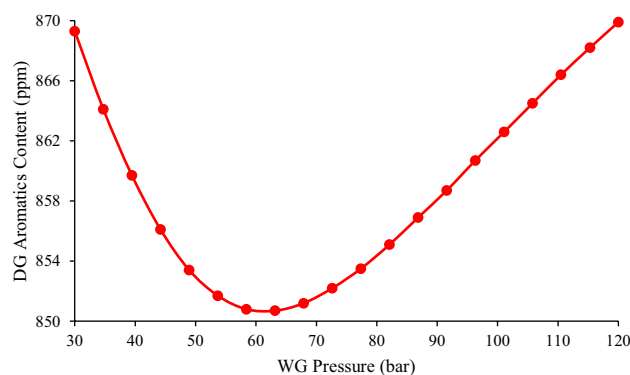


Fig. 6 Effect of wet gas pressure on dehydrated gas aromatics content



methodologies, namely the momentum and root mean square propagation algorithm. The exponential weighted average of the gradients is considered, thus leading to swift convergence of the algorithm. A learning rate of 0.01 is used according to normal practice.

Fig. 7 Effect of wet gas pressure on **a** dehydrated gas water content and aromatics content, **b** dehydrated gas hydrate formation temperature water dew point, **c** stripping gas flow rate, and reboiler duty

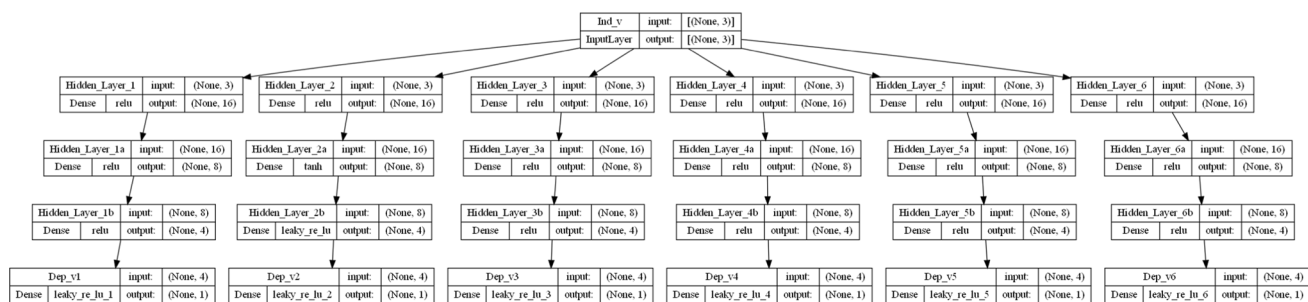
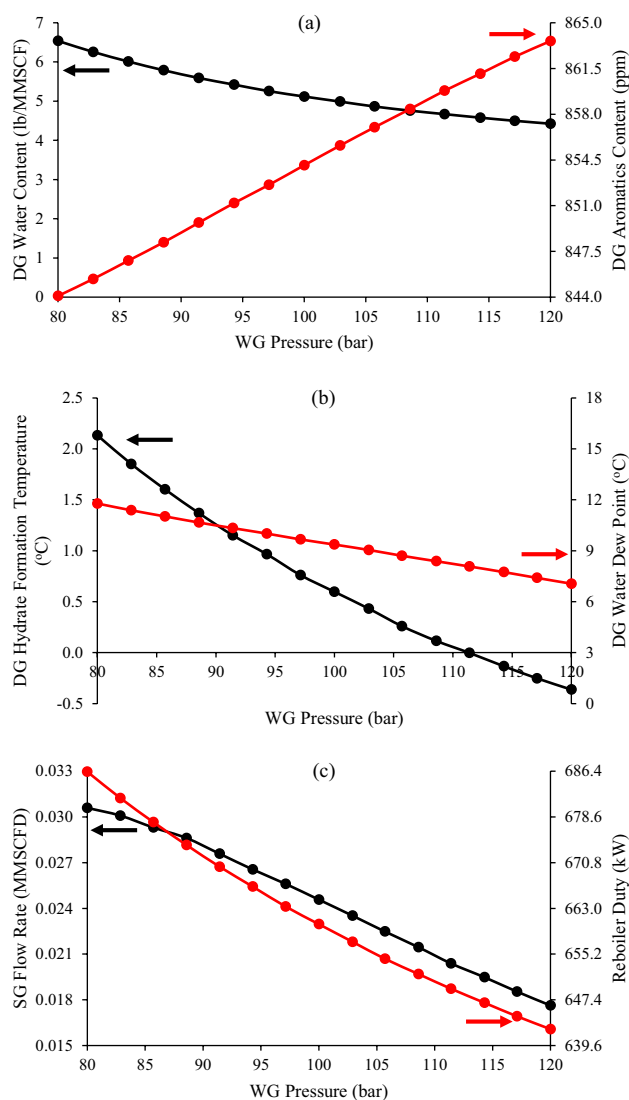


Fig. 8 ANN with MIMO structure conducted for the case study via Keras Python

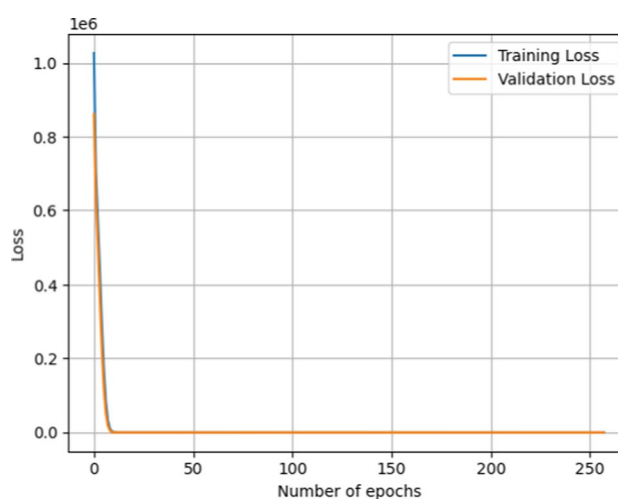
4.2.6 Epochs and early stopping

A total of 300 epochs are used to carry out the ANN training with an early stopping (callback) criteria of minimum model improvement of zero in MSE validation loss with patience of 50 epochs. A detailed explanation of epoch can be found in Figure S3 in the supplementary file.

Table 7 Parameters corresponding to the input and output layers of the MIMO-ANN model

| | Corresponding parameters |
|---------------|---|
| Input layer | |
| Ind_v | LG recirculation rate LG purity WG pressure |
| Output layers | |
| Dep_v1 | DG water content |
| Dep_v2 | DG aromatics content |
| Dep_v3 | DG hydrate formation temperature |
| Dep_v4 | DG water dew point |
| Dep_v5 | SG flowrate |
| Dep_v6 | Reboiler duty |

LG Lean glycol, WG Wet gas, DG Dehydrated gas, SG Stripping gas

Fig. 9 Loss profile (MSE) on training and validation runs of the MIMO-ANN model over the number of epochs

5 Results and discussion

5.1 Loss profile validation, model evaluation via linear regression and error quantification

The overall loss profile (MSE) of the MIMO-ANN model is illustrated in Fig. 9. The training epoch is stopped at 257 initiated by the callback function whereby there is no significant improvement of the MSE for 50 consecutive epochs. As a result, there is a 99% drop in both training and validation loss. The validation loss is slightly higher than the training loss, indicating good training of the MIMO-ANN model and data splitting that does not display overfit results. The overall MSE for the testing dataset is 1.79, implying an outstanding prediction of the MIMO-ANN model on unseen data. The smaller MSE of the testing dataset to that of the training dataset can be explained by the MSE torment due to dropout regularisation during the training phase resulting in a higher training MSE profile.

The predicted outputs by the MIMO-ANN model over actual outputs simulated by Aspen HYSYS for all the studied outputs (DG water content, DG aromatics content, DG hydrate formation temperature, DG water dew point, SG flow rate and Reboiler duty) according to training and testing datasets are shown in Figs. 10, 11, 12, 13, 14, 15. A line of best fit is plotted for all the predicted versus actual outputs on training and testing datasets, and the resulting regression lines are dominantly overlapping with the 45° reference line ($y = x$) with an excellent R^2 score of > 0.999 . This successfully demonstrates the outstanding capability of the MIMO-ANN model to carry out predictions for unseen data.

The mean absolute percentage errors between the actual and predicted outputs for all the training and testing datasets are shown in Fig. 16. The percentage deviation for all the testing datasets is less scattered compared to that of training datasets, which is in agreement with the slightly lower value of overall MSE in the testing dataset

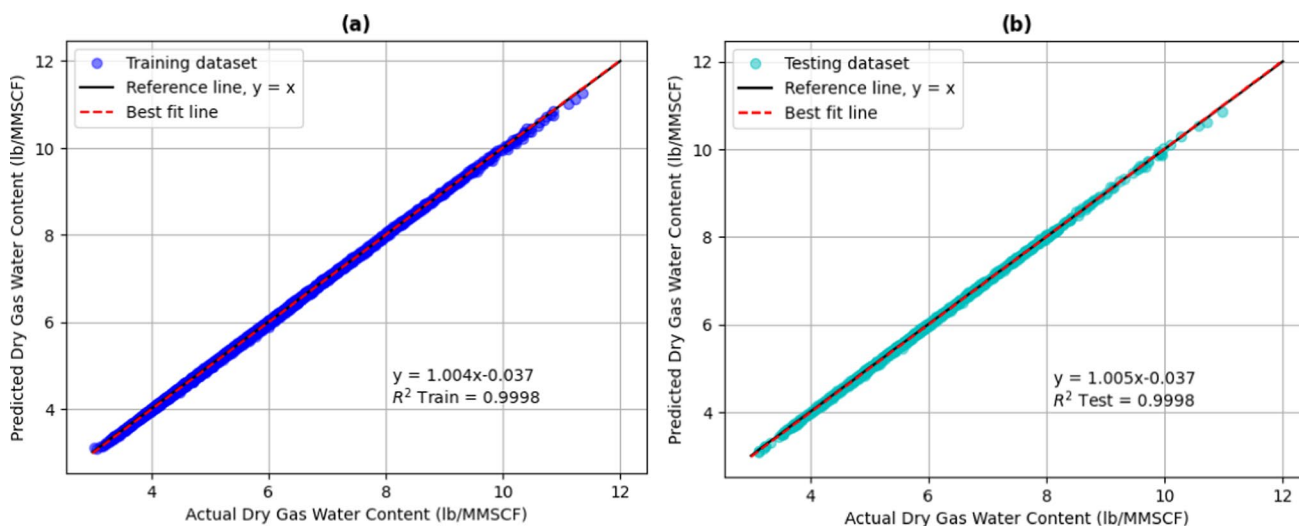


Fig. 10 Predicted versus actual values of dry gas water content for **a** training dataset and **b** testing dataset

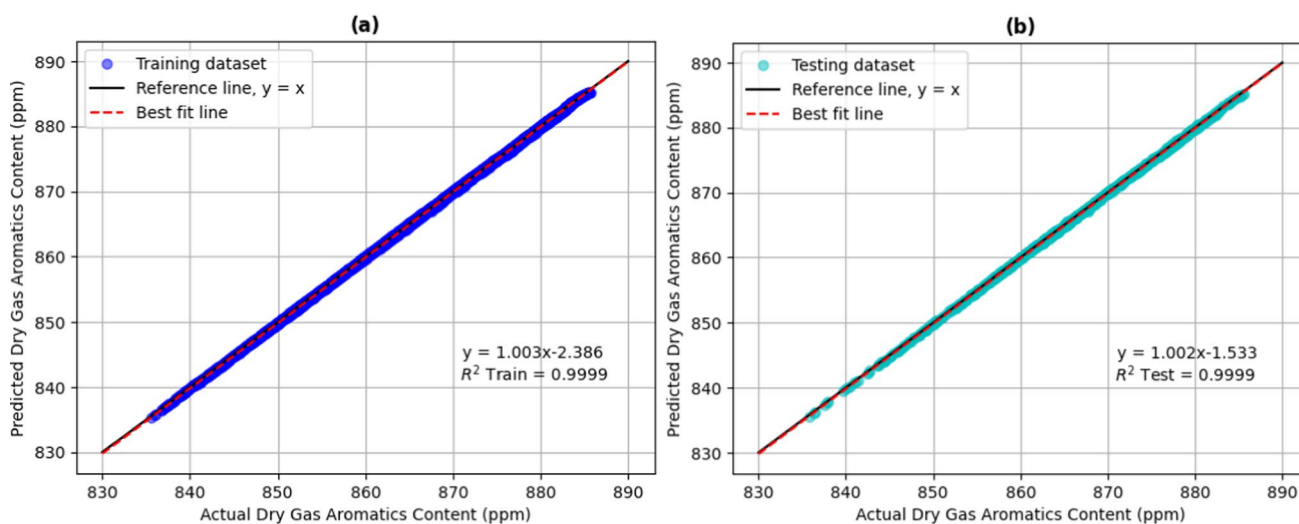


Fig. 11 Predicted versus actual values of dry gas aromatics content for **a** training dataset and **b** testing dataset

compared to the training dataset. In general, the percentage deviation range for all the predicted output ranges is less than 5% with a MAPE value of less than 1%, except for the predicted output values for dehydrated gas hydrate formation temperature and stripping gas flow rate. The percentage error began to surge up to values near -650% and 1000% , resulting in a MAPE value of 16.63% and 11.55% for both the output predictions respectively when the predictions are carried out on pseudo-zero actual values. This is due to the nature of the percentage error formula (shown in Eq. S7) that yields an exceptionally high value by dividing a pseudo-zero value. Therefore, the high MAPE value for both the output predictions is not critical as the prediction performance of the MIMO-ANN model is still highly accurate according to the best-fit line regressed in Figs. 12 and 13.

5.2 Limitations

The MIMO-ANN model developed in this work although demonstrates exceptional predictive performance according to the results shown above, however, is constrained by several limitations which can be further improved in future work. One of the limitations that can be improved in this work is to extend the consideration of additional process inputs. The current process inputs considered are lean glycol recirculation rate, lean glycol purity and wet gas pressure. This may

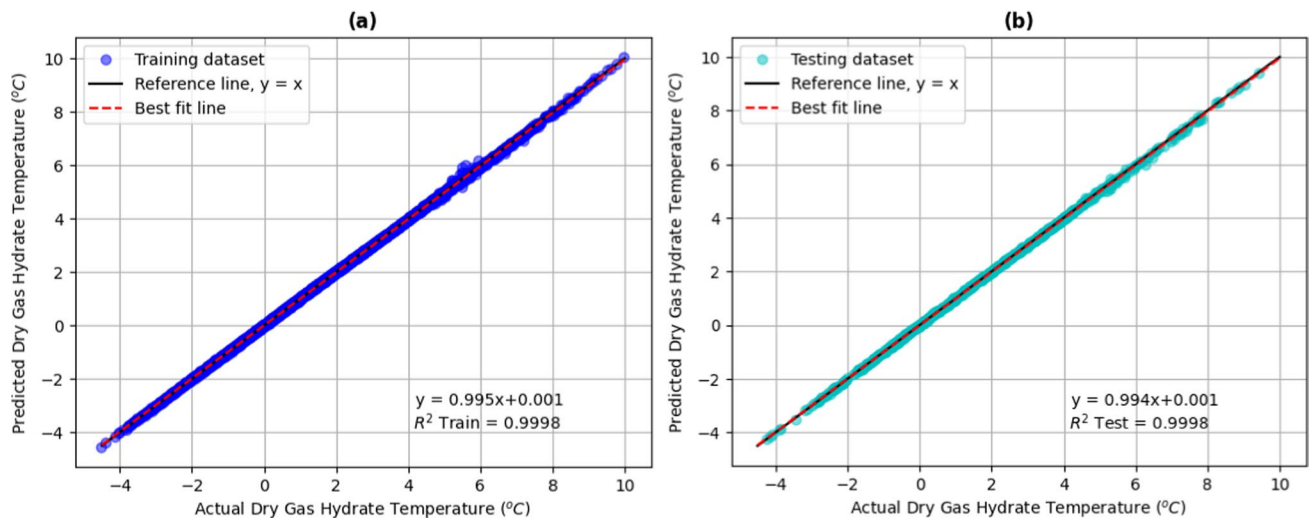


Fig. 12 Predicted versus actual values of dry gas hydrate temperature for **a** training dataset and **b** testing dataset

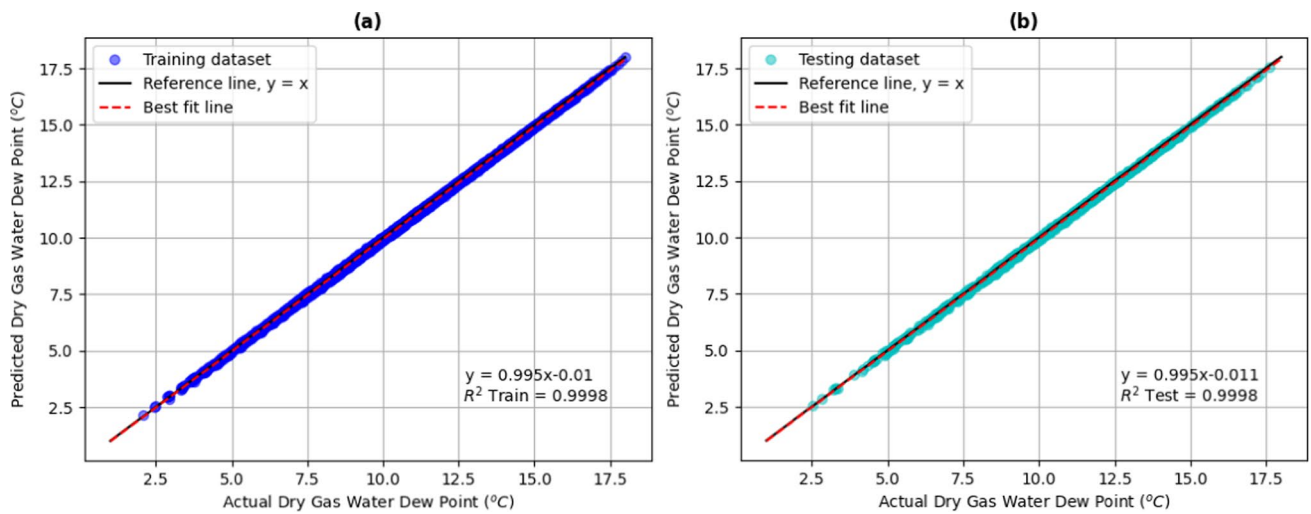


Fig. 13 Predicted versus actual values of dry gas water dew point for **a** training dataset and **b** testing dataset

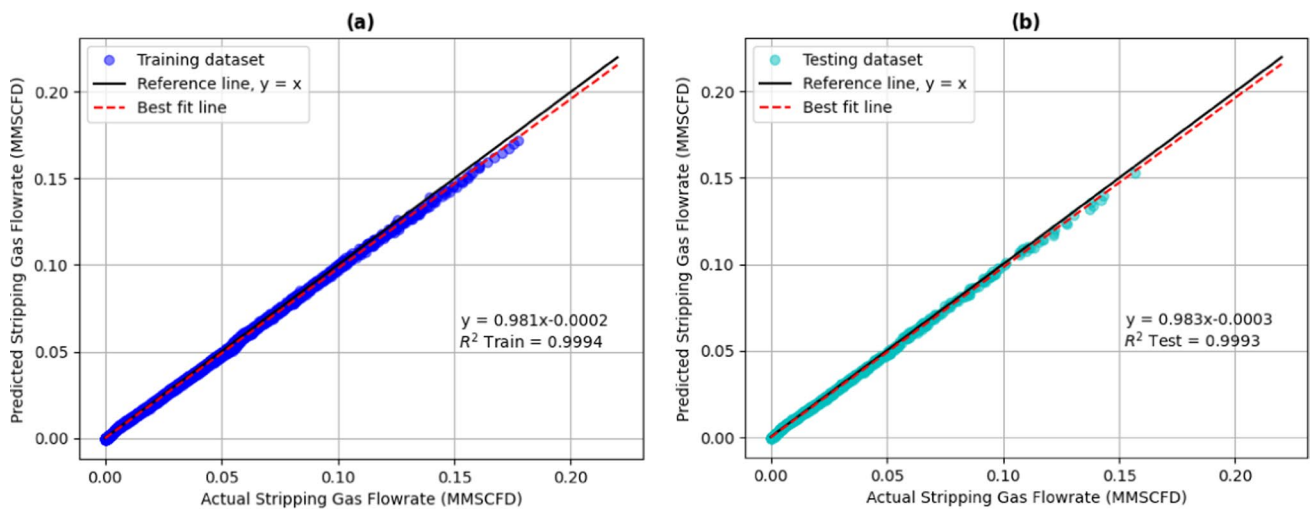


Fig. 14 Predicted versus actual values of stripping gas flow rate for **a** training dataset and **b** testing dataset

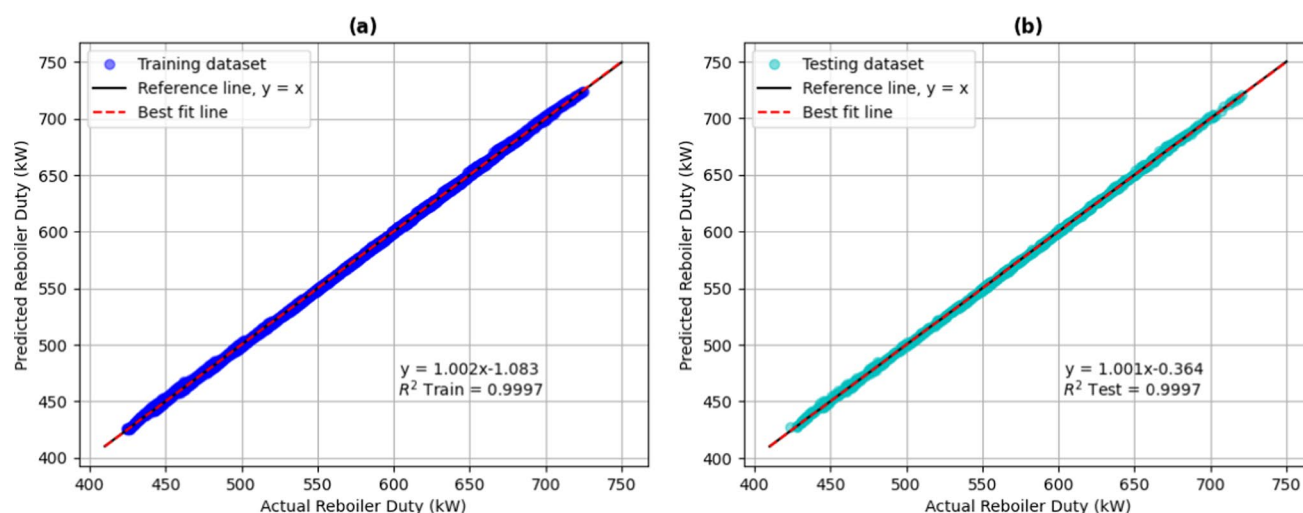


Fig. 15 Predicted versus actual values of reboiler duty for **a** training dataset and **b** testing dataset

be insufficient to generalise the prediction of the process outputs for the entire process as other process inputs such as wet gas composition, wet gas flow rate, wet gas temperature, lean glycol temperature and lean glycol pressure may fluctuate to influence the desired process outputs as well. However, the consideration of additional process inputs will inevitably populate the number of datasets that will be fed into the MIMO-ANN model training drastically, requiring an extensive computational processing capability of the incorporated device upon carrying out the ANN training process.

Another limitation in this work includes a moderate training bound opted for MIMO-ANN training due to the minor lean glycol purity enhancement performed by the stripping column for the given base case scenario. The minute increase in lean glycol purity from 98.28 to 98.56 wt% by the stripping column in the base case narrows down the trainable data range from 98.50 to 99.00 wt% after the consideration of lean glycol recirculation rate and wet gas pressure combinations. This is due to certain combinations of the lean glycol recirculation rate and wet gas pressure resulting in a lean glycol with purity larger than the selected lower bound of the lean glycol purity and therefore fails the stripping column calculation in Aspen HYSYS. This makes lean glycol purity the strictest constraint when selecting the upper and lower bounds of the process inputs.

Furthermore, the above MIMO-ANN model is solely specific to the configuration of the TEG dehydration system. Any modifications to the system especially on the column settings will result in the ANN output to be not reliable and thus the process has to be re-simulated and the ANN model has to be retrained again to obtain new output predictions.

6 Conclusion

This work has successfully developed a MIMO-ANN model with three hidden layers individually to simulate the entire TEG natural gas dehydration process. The overall MSE is 1.79 and all the R^2 scores for the process output predictions are > 0.999 for the testing dataset. The MAPE on all the output predictions is less than 1%, except on dry gas hydrate formation temperature and stripping gas flow predictions that possess a value of 16.63% and 11.55% respectively due to the prediction attempted on pseudo-zero values. The MIMO-ANN model displays an unexceptional performance on unseen data endowed by rigorous ANN training via dropout regularisation. This potentially unravels the application of this MIMO-ANN model in digital twin as a process monitoring and improvement tool.

Various limitations of the MIMO-ANN model such as insufficient process input considerations, moderate ANN training data range and inflexibility are discussed as well, unravelling the potential for model improvement in future work.

Author contributions Goh did the analysis and wrote the main manuscript text, Foo supervised the project, Rafil and Anuar assists with analysis, all authors reviewed the manuscript.

Data availability The authors declare that the data supporting the findings of this study are available within the paper and its supplementary file.

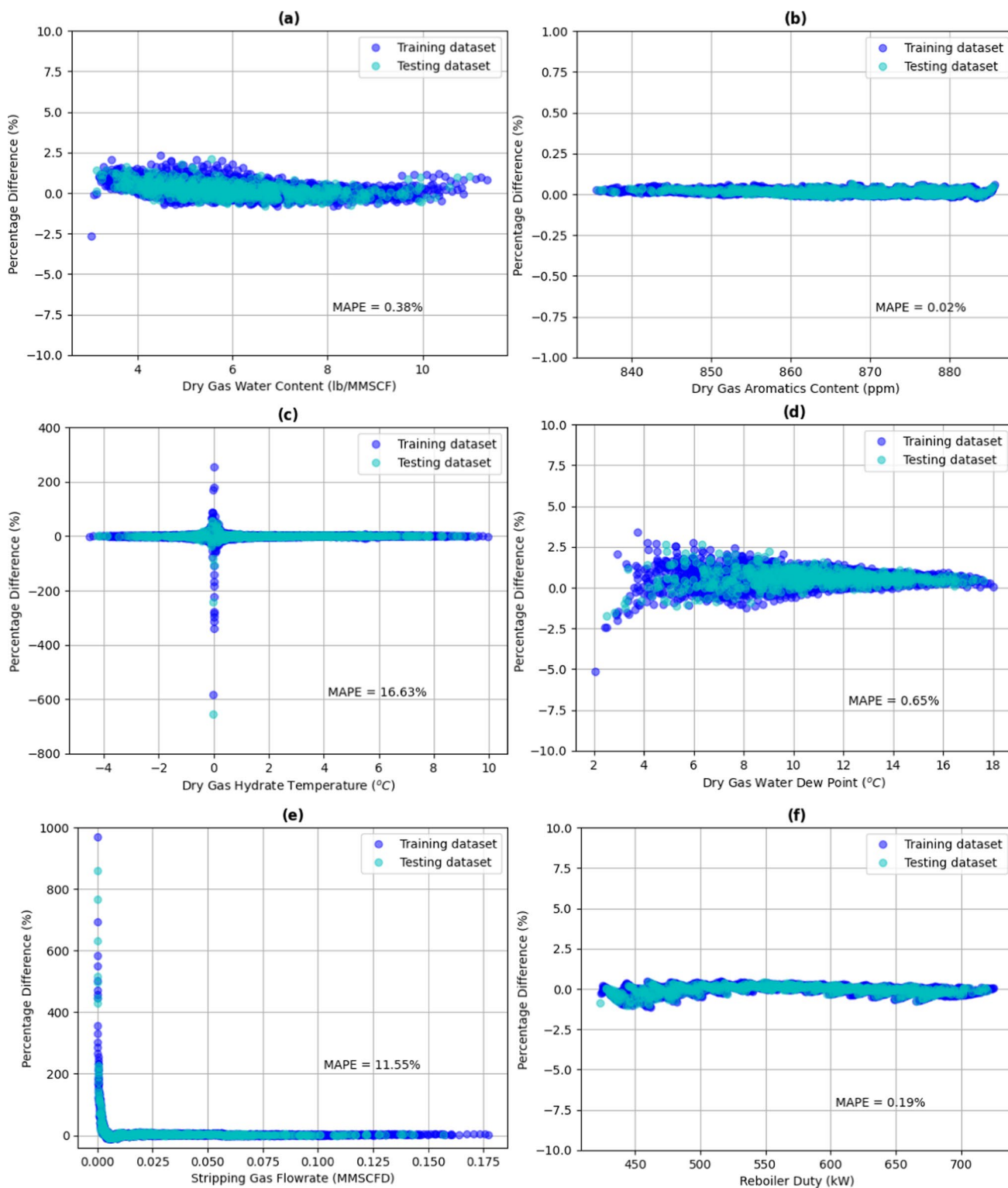


Fig. 16 Percentage error between predicted and actual values for **a** dry gas water content, **b** dry gas aromatics content, **c** dry gas hydrate formation temperature, **d** dry gas water dew point, **e** stripping gas flow rate and **f** reboiler duty based on training and testing dataset

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Al-Haidous S, Al-Ansari T. Sustainable liquefied natural gas supply chain management: a review of quantitative models. *Sustainability*. 2020;12:243.
2. Gerber Machado P, Naoki Akiyoshi Ichige E, Ninni Ramos K, Mouette D. Natural gas vehicles in heavy-duty transportation—a political-economic analysis for Brazil. *Case Stud Transp Policy*. 2021;9(1):22–39.
3. <https://www.iea.org/fuels-and-technologies/gas>. Accessed 29 Nov 2022
4. Slorach PC, Stamford L. Net zero in the heating sector: technological options and environmental sustainability from now to 2050. *Energy Convers Manage*. 2021;230: 113838.
5. Fair C, Cook B, Brighton T, Redman M, Newman S. Gas/condensate and oil well testing—from the surface. 2002.
6. Anyadiiegwu CI, Kerunwa A, Oviawele P. Natural gas dehydration using triethylene glycol (teg). *Petrol Coal*. 2014;56:407–17.
7. Andreeva VR, Ivashov AY, Kotin KA, Voltsov AA. High gas-oil ration separator units. 2016.
8. Bagherisadr M, Bozorgian A. Decomposition of hydrates in the pipeline. *Int J Adv Stud Hum Soc Sci*. 2020;9(4):252–61.
9. Nmegbu G. Optimization of triethylene glycol (teg) dehydration in a natural gas processing plant. *Int J Res Eng Technol*. 2014;03:346–50.
10. Al-Jammali ASI. Study the effect of the main variables on the objectives of the natural gas dehydration plant by aspen-hysys v8.8. *J Petrol Res Stud*. 2022;12(2):47–63.
11. Ahmadi MA, Soleimani R, Bahadori A. A computational intelligence scheme for prediction equilibrium water dew point of natural gas in teg dehydration systems. *Fuel*. 2014;137:145–54.
12. Holder GD. Clathrate hydrates of natural gases, 2nd edn. revised and expanded. Chemical industries series/73 by e. Dendy Sloan Jr. Marcel Dekker, inc.: New york. 1998. 754 pages. \$195.00. Isbn 0-8247-9937-2, *Journal of the American Chemical Society*. 1998; 120 (43): 11212–11212
13. Abu RN, Amah LG, Dulu A, Adeloye OM. Chemical control of natural gas hydrate. *Eur J Eng Technol Res*. 2018;1(6):58–62.
14. Piemonte V, Maschietti M, Gironi F. A triethylene glycol–water system: a study of the teg regeneration processes in natural gas dehydration plants. *Energy Sources, Part A Recov Util Environ Effects*. 2012;34(5):456–64.
15. Michal NI, Pavel D. Natural gas dehydration. In: Sreenath Borra G, editor. *Natural gas*. Rijeka: IntechOpen; 2012. p. Ch. 1.
16. Kao MJ, Tien DC, Jwo CS, Tsung TT. The study of hydrophilic characteristics of ethylene glycol. *J Phys Conf Ser*. 2005;13(1):442.
17. Ahmad Z, Bahadori A, Zhang J. Prediction of equilibrium water dew point of natural gas in teg dehydration systems using bayesian feedforward artificial neural network (fann). *Pet Sci Technol*. 2018;36(20):1620–6.
18. Ghanbari A, Kardani MN, Moazami Goodarzi A, Janghorban Lariche M, Baghban A. Neural computing approach for estimation of natural gas dew point temperature in glycol dehydration plant. *Int J Ambient Energy*. 2020;41(7):775–82.
19. Miracle OK, Koyejo OM. Predicting water content outcomes in natural gas dehydration systems using artificial intelligence. *SSRG-IJCER*; 2021.
20. <https://www.aspentech.com/en/products/engineering/aspen-hysys2022>. Accessed 29 Nov 2022
21. Eldar YC, Ben-Tal A, Nemirovski A. Robust mean-squared error estimation in the presence of model uncertainties. *IEEE Trans Signal Process*. 2005;53(1):168–81.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.