Chemical Engineering Science: X 8 (2020) 100080

Contents lists available at ScienceDirect

Chemical Engineering Science: X

journal homepage: www.elsevier.com/locate/cesx

Surrogate modelling of VLE: Integrating machine learning with thermodynamic constraints

Andres Carranza-Abaid *, Hallvard F. Svendsen, Jana P. Jakobsen

Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 26 June 2020
Received in revised form 26 August 2020
Accepted 12 September 2020

Keywords:
Thermodynamics
Machine learning
Surrogate modelling
CO_2
MEA

ABSTRACT

An easy-to-implement methodology to develop accurate, fast and thermodynamically consistent surrogate machine learning (ML) models for multicomponent phase equilibria is proposed. The methodology is successfully applied to predict the vapour-liquid equilibrium (VLE) behavior of a mixture containing CO_2, monoethanolamine (MEA), and water (H_2O). The accuracy of the surrogate model predictions of VLE for this system is found to be satisfactory as the results provide an average absolute relative difference of 0.50% compared to the estimates obtained with a rigorous thermodynamic model (eNRTL + Peng-Robinson).

It is further demonstrated that the integration of Gibbs phase rule and physical constraints into the development of the ML models is necessary, as it ensures that the models comply with fundamental thermodynamic relationships.

Finally, it is shown that the speed of ML based surrogate models can be ~10 times faster than interpolation methods and ~1000 times faster than rigorous VLE calculations.

1. Introduction

1.1. Motivation and literature review

The emission of anthropogenic greenhouse gases has been one of the main subjects of environmental concerns over the past decades. Development of new, clean and enhanced industrial processes has become a must in order to reach international and national sustainability goals. One of the most promising approaches to operate cleaner industrial processes is the implementation of CO_2 capture and storage. It has been labeled as one of the key technologies that will assist in achieving a global temperature increment of no more than 1.5 °C by the end of 2030 (IPCC, 2018). Albeit there are several other technologies for CO_2 capture, chemical absorption of CO_2 with aqueous amines has been, and seems to be in the foreseeable future, the most commercially ready and competitive technology (Rochelle, 2009; Wu et al., 2020). The high energy demand associated with the CO_2 capture process is its main challenge (Svendsen et al., 2011).

Extensive experimental and modelling research has been conducted aiming to find a system with favorable vapor-liquid equilibrium (VLE) behavior that will lead to a more energetically efficient process. Modelling of VLE is of great importance for these efforts. The knowledge of phase behavior is necessary in order to assess the solvent performance in CO_2 capture processes. Traditionally, the VLE models are divided into two categories: semi-empirical models (also called rigorous models) and empirical models, each with their respective advantages and disadvantages.

In this work, an alternative approach to VLE modeling is proposed which attempts to merge all the main features of both categories. The main objectives of this work are:

- To present an easy-to-implement method based on machine learning technology that combines the robustness of rigorous models with the computational efficiency of an empirical model.
- Demonstrate that machine learning models can properly predict the dependencies between the input and output variables if a proper analysis of the thermodynamic variables is performed.
- Give an insight into how to select properly the input and the output variables of a machine learning model so that it is thermodynamically consistent.

Rigorous models may use different mathematical representations like an equation of state for both phases (phi-phi formulation) or excess Gibbs energy model for liquid phase and equation

* Corresponding author.
E-mail address: andres.c.abaid@ntnu.no (A. Carranza-Abaid).

https://doi.org/10.1016/j.cesx.2020.100080
2590-1400/

of state for the gas phase (gamma-phi formulation). These models make use of governing thermodynamic equations (e.g. Henry's law, Raoult's law), semi-theoretical models (e.g. excess Gibbs energy models for aqueous electrolytic solutions (Chen and Evans, 1986; Lloret et al., 2017; Pitzer, 1973)) together with empirically fitted parameters. Some successful applications of gamma-phi models for the prediction of the CO_2 solubility in several aqueous amines are: 2-aminoethan-1-ol (MEA) (Aronu et al., 2011), piperazine (PZ) and 2-amino-2-methylpropan-1-ol (AMP) (Dash et al., 2011), 2-piperidineethanol (PE) (Sherman et al., 2016), AMP (Hartono et al., 2020) to name a few.

Although the semi-empirical models have proven to give accurate predictions of VLE, the difficulty of developing a thermodynamic model for each CO_2 aqueous amine solvent and the high computational overhead of these computationally complex models have led to the development of empirical models.

A simple empirical VLE model that describes the CO_2 solubility in primary, secondary and tertiary amines was proposed by Gabrielsen et al. (2005). This model was developed by summing up all the reactions into a single overall reaction and lumping the equilibrium constant with the CO_2 physical solubility coefficient to obtain the CO_2 partial pressure. Another empirical approach was taken by Brúder et al. (2012, 2011), Luo et al. (2015), Plesu et al. (2018) where empirical correlations that relate the molar compositions and the temperature with the CO_2 partial pressure were formulated. All these empirical correlations are computationally inexpensive due to their simplicity when compared to gamma-phi models. The main drawback of these empirical models is their low dimensionality, which translates into their limited validity range. Hence, the mathematical functions are different among different systems.

Machine learning has been used as an alternative to create empirical models. Several studies in the literature have used this approach to estimate the VLE of CO_2 capture related thermodynamic systems. These studies have mostly focused on the estimation of CO_2 solubility in different liquid solvents: triisopropanolamine – MEA aqueous solutions (Daneshvar et al., 2004), various alkanol systems (Zarenezhad and Aminian, 2011), several aqueous amine and diamine systems containing MEA, MDEA, PZ, 2-amino-2-methyl-1-propanol (AMP) (Bastani et al., 2013; Hamzehie et al., 2015, 2014), pure water (Ghasemian et al., 2013), mixtures of ethanol and ionic liquids (Mirarab et al., 2014), aqueous PZ solutions (Norouzbahari et al., 2015), piperazine (PZ) and ionic liquids (Golzar et al., 2016), aqueous sodium salt of L-phenylalanine (Garg et al., 2017), aqueous potassium lysinate mixed with MEA (Zhang et al., 2018), among others. In general, the authors of these models claim that their machine learning models have better prediction capabilities than the semi-empirical models. This is expected, since these multidimensional models were fitted to a limited range of operating conditions. However, that is unfortunately also the reason why they cannot be extrapolated. In addition, due to their empirical nature and to the fact that they are constrained to an univariable output, the reported models are not capable of predicting other important thermodynamic quantities such as the CO_2 heat of absorption or speciation in the liquid phase.

Because of these major limitations, a direct application of these models into a process simulation framework is not practical. Besides, in the context of CO_2 absorption into aqueous amine systems, the molar compositions of all the ionic species and the physical solubility of CO_2 are needed for the evaluation of kinetic and mass transfer rate expressions. Even though the CO_2-MEA-H_2O mixture is the most studied amine system, the available data alone is not enough to create an accurate machine learning model that will estimate the VLE behavior of all thermodynamic (e.g. liquid or vapor molar fractions) variables over a broad range of operating conditions.

1.2. What is machine learning?

Machine learning has become popular over the past years in both industrial and research environments. The sudden increase in popularity is due to the fast improvement of the technical capabilities of current-day computers. Machine learning is a computational modelling tool for data management that enables the classification, pattern recognition, clustering and data prediction (Abiodun et al., 2018). Machine learning is closely related to artificial intelligence, hence, it aims at implementing intelligent agents that are capable of mimicking the cognitive functions of a biological brain in order to learn or solve complex problems (Russel and Norvig, 2009).

One of the most prominent and notable methods of machine learning is artificial neural networks (ANN), which were first conceptualized by McCulloch and Pitts several decades ago (McCulloch and Pitts, 1943). An ANN can be defined as a nonlinear vector of functions ($\boldsymbol{\Omega}$) that needs a vector of input variables ($\underline{X}$) and a set of weight parameters ($\underline{\omega}$) in order to estimate a vector of output variables $\underline{\hat{Y}}$ (Bishop, 2006). In general, an ANN has the general form:

$$\underline{\hat{Y}} = \Omega(\underline{X};\underline{\omega}) \tag{1}$$

Due to the high dimensional nature of ANN, any set of experimental or observed values ($\underline{Y}^e$) corresponding to a certain set of experimental input values (X^e) can be approximated using Eq. (1) as long as an appropriate $\underline{\omega}$ is used.

The process of obtaining $\underline{\omega}$ that, together with $\underline{X}^e$, allows an estimation of the predicted output $\underline{\hat{Y}}$ which numerically resembles the experimental values $\underline{Y}^e$ is called ANN training. This process is based on the biological analogy on how a person is trained to perform a task correctly through a feedback procedure. Following this scheme, a person performs a task and receives feedback on the task performance in order to improve the result. This process goes on repeatedly until the person performs the task well enough to meet a standard. Following this analogy, the ANN is trained by utilizing a loss function ($\mathscr{L}$) that evaluates its performance by comparing the experimental values $\underline{Y}^e$ and the predicted output $\underline{\hat{Y}}$:

$$Loss\ function = \mathscr{L}\left(\underline{\hat{Y}},\underline{Y}^e\right) \tag{2}$$

The loss function indicates the total error of the predicted values with respect to the observed values.

Fig. 1 illustrates the training process of a feedforward neural network (FFNN) that has 3 input values, 4 neurons in the hidden layer and that estimates 2 values in the output layer. The FFNN inside the grey square in Fig. 1 and is said to be a feedforward propagation model because the information flows only in one direction. First, the input vector $\underline{X}^e$ is transformed to calculate the hidden layer vector $\underline{Z}$ using an activation function (e.g. an hyperbolic function or sigmoid function) and 2 subsets of the weight parameter vector ($\underline{\omega}^{(0)}$ and $\underline{\omega}^{(1)}$, also note that: $\underline{\omega}^{(0)} \cup \underline{\omega}^{(1)} \cup \underline{\omega}^{(2)} = \underline{\omega}$). Afterwards, vector $\underline{Z}$ uses a linear transformation together with another subset of weight parameters $\underline{\omega}^{(2)}$ to estimate $\underline{\hat{Y}}$. The output vector ($\underline{\hat{Y}}$) is then forwarded to the loss function (marked with a red square in Fig. 1) outside the FFNN. The loss function assesses the ANN performance by comparing $\underline{\hat{Y}}$ and $\underline{Y}^e$ and, depending on this evaluation, a backpropagation signal is sent to modify $\underline{\omega}$ so that the loss function is minimized. This process keeps going until the variation of the parameters stops

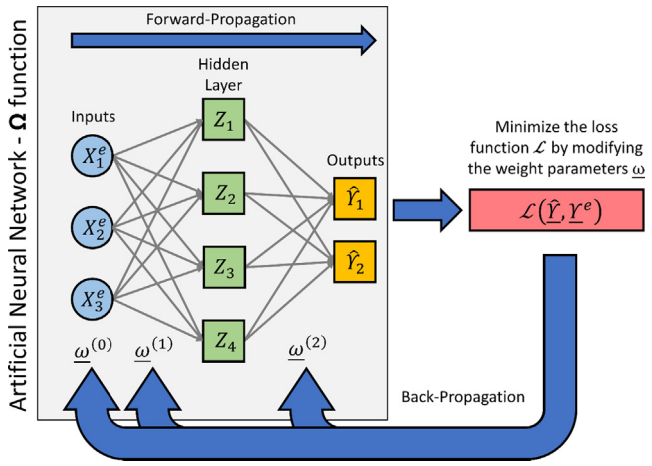

Fig. 1. Diagram of a forward propagation artificial neural network (FFNN) trained by a backpropagation algorithm.

improving the predictions. The loss function can be optimized through different backpropagation training methods such as Levenberg-Marquardt, resilient backpropagation, Bayesian regularization propagation and the scale conjugate gradient among others (Gonzalez Viejo et al., 2019).

In summary, training an ANN is quite similar to the other approaches that have been used for decades or even centuries to develop mathematical models; one of such methods is polynomial fitting. For example, letus consider a case where it is desired to develop a linear model of a process. One must have a set of experimental data that contains the values of the independent variable $\underline{X}^e$ and the dependent variable $\underline{Y}^e$. A linear model contains only the slope and the intersect as the model weight parameters $\underline{\omega}$, so the «training» of a linear model should find the $\underline{\omega}$ that minimize the difference between $\underline{Y}^e$ and the values calculated with the linear model $\underline{\hat{Y}}$. Machine learning models use the same concept as polynomial fitting. The main difference is that the machine learning model equations are more complicated, hence the optimization algorithms are more sophisticated. Despite the similarities between ANN and other mathematical modelling methods, ANN has the advantage of its inherent high dimensionality that allows it to capture non-linear behaviors.

2. Methodology

2.1. The thermodynamic system

The vapor-liquid equilibrium of the reactive mixture of CO_2-MEA-H_2O is a highly non-ideal thermodynamic system due to the distinct chemical nature of the components that constitute the liquid phase. There is a gas (CO_2) dissolved in an already non-ideal mixture of MEA and H_2O. Furthermore, some species are electrolytes and the presence of ions induces highly non-ideal long-distance interactions that contribute to a highly non-linear behavior. The reactions taking place in the liquid phase have been reported to be (Kim et al., 2009):

$$2H_2O \leftrightarrow H_3O^+ + OH^- \quad (3)$$

$$2H_2O + CO_2 \leftrightarrow H_3O^+ + HCO_3^- \quad (4)$$

$$H_2O + HCO_3^- \leftrightarrow H_3O^+ + CO_3^{2-} \quad (5)$$

$$MEAH^+ + H_2O \leftrightarrow H_3O^+ + MEA \quad (6)$$

$$MEACOO^- + H_2O \leftrightarrow MEA + HCO_3^- \quad (7)$$

A sketch of the component distribution in the system of CO_2 chemically absorbed in MEA and H_2O is shown in Fig. 2. It illustrates the presence of 3 volatile components that are distributed between the two phases and the non-volatile components that correspond to the cations and anions formed in the liquid phase due to the chemical reactions. Note that the water is considered as the diluent of the ions and the volatile components in Fig. 2.

2.2. Essential thermodynamic relations

2.2.1. Physical constraints and auxiliary equations

The semi-empirical gamma-phi VLE model that is replicated in this work (also referred to as "base model") considers the phase equilibria of the volatile components and the speciation reactions in the liquid phase. This gamma-phi model is composed of equations for phase equilibria (Henry's law and Raoult's law), chemical equilibria, thermodynamic constraints and complementary auxiliary equations. The complementary auxiliary equations can be either of empirical or semi-empirical nature and have fitted parameters (e.g. an activity coefficient model).

The base model estimates the activity coefficients with the eNRTL model (Chen and Evans, 1986) along with the Peng-Robinson equation of state for the vapor phase (Peng, 1976). The binary parameters of the eNRTL activity coefficient model, the equilibrium constant correlations and the implementation of the semi-empirical model were taken from a previously published paper (Putta et al., 2016).

Since machine learning models use alternate pathways to compute the numerical values of the VLE, there is no certainty that these estimations comply with thermodynamic rules. Therefore, it is mandatory to ensure that the calculated thermodynamic system fulfills physical constraints. The summation constraints for the vapor and liquid phases must be considered in the implementation of machine learning models:

$$\sum_i y_i = 1 \quad (8)$$

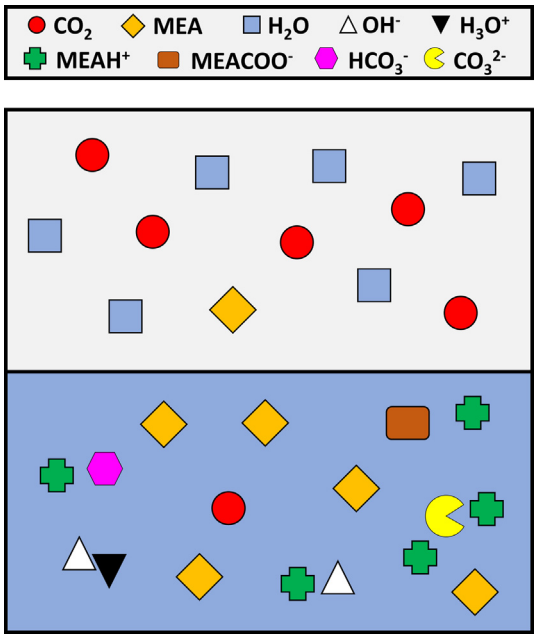

Fig. 2. Sketch of the modelled thermodynamic system with water as the diluting component of the liquid phase.

$$\sum_i x_i = 1 \tag{9}$$

The electro-neutrality constraint is a restriction that only exists in electrolytic systems. This equation arises from the principle that the overall sum of local charges must be 0 in a system at thermodynamic equilibrium (Prausnitz et al., 1999). This relation is:

$$\sum_i n_i \Gamma_i = 0 \tag{10}$$

Here n is the total number of moles of the component and Γ is the ion charge relative to a hydrogen ion. For a more detailed description of the additional fundamental equations and the parameters of the semi-empirical model please refer to the supporting information S1 and S2.

It is customary in the CO_2 capture research field to represent the component concentrations in the liquid phase with non-natural thermodynamic variables: CO_2 loading (α_{CO_2}) and amine weight percent in the liquid solvent on a CO_2 free basis (w_{MEA}). The CO_2 loading is estimated with the following relation:

$$\alpha_{CO_2} = \frac{x_{CO_2,App}}{x_{MEA,App}} \tag{11}$$

where the subscript App refers to apparent and indicates that the molar compositions of CO_2 and MEA are calculated as if they had not reacted in the liquid phase. Therefore, the apparent molar compositions only consider the CO_2, MEA and H_2O molar fractions.

The following equation relates w_{MEA} with the MEA apparent molar fraction on a CO_2 free basis $x^*_{MEA,App}$:

$$x^*_{MEA,App} = \frac{\left(\frac{w_{MEA}}{\bar{M}_{MEA}}\right)}{\left(\frac{w_{MEA}}{\bar{M}_{MEA}} + \frac{100 - w_{MEA}}{\bar{M}_{H_2O}}\right)} \tag{12}$$

where $\bar{M}_{MEA}$ and $\bar{M}_{H_2O}$ are the molecular weights of MEA and H_2O respectively. It is possible to calculate the apparent molar compositions of the liquid phase using Eqs. (11) and (12).

2.2.2. Enthalpy of phase change

The CO_2 enthalpy of phase change (usually referred to as CO_2 enthalpy of absorption) is of fundamental importance in the field of acid–gas treating because it is one of the main characteristics that determines the techno-economic potential of the technology (Oexmann and Kather, 2009; Raksajati et al., 2013). It is this thermodynamic quantity that is needed to estimate the required heat duty of the solvent regeneration system in the process.

As opposed to non-supercritical fluids (MEA-H_2O), the enthalpy of vaporization of permanent gases (CO_2) cannot be measured in pure state. Therefore, it is a function of the solvent into which the CO_2 is absorbed. Moreover, experimentally it is not possible to measure the enthalpy of vaporization and the enthalpies of the individual reactions separately. Therefore, the overall effect is commonly measured and reported as a single CO_2 enthalpy of absorption (Jou et al., 1994; Kim et al., 2014).

An alternative way to obtain the heat of phase change from VLE data is by using an expression derived from the Gibbs-Helmholtz relation (the vant' Hoff equation):

$$\left(\frac{\partial ln(P_{CO_2})}{\partial\left(\frac{1}{T}\right)}\right)_{P,n} = -\frac{\Delta H_{CO_2}}{R} \tag{13}$$

Here P_{CO_2} is the partial pressure of CO_2 in the vapor phase and ΔH_{CO_2} is the CO_2 enthalpy of absorption. For an in-depth discussion of the derivation and the inherent assumptions to obtain Eq. (13) see (Svendsen et al., 2011).

2.2.3. Gibbs phase rule analysis

The Gibbs phase rule for reactive systems provides an information about how many degrees of freedom exist in a closed reactive thermodynamic system. The degrees of freedom refer to the number of independent variables that can be simultaneously set in order to fully specify the state of the thermodynamic system. The equation for the Gibbs phase rule for reactive systems is:

$$F = 2 + N - \pi - r - s \tag{14}$$

where F is the number of degrees of freedom, N is the number of components, π is the number of phases, r is the number of reactions and s is the number of non-summation constraints. It is imperative to follow the Gibbs phase rule in the development of machine learning models. If this rule is broken, thermodynamic quantities cannot be estimated with thermodynamic consistency.

2.3. Developing surrogate thermodynamic models

The principles of the proposed methodology to formulate surrogate machine learning VLE models is summarized in Fig. 3. Fig. 4 presents detailed the algorithms of the steps 4,5 and 7 shown in Fig. 3.

Step 1: The semi-empirical model to be replicated is chosen. As previously mentioned, an eNRTL thermodynamic model for CO_2-MEA-H_2O is utilized in this work as the case study to exemplify the use of the proposed methodology.

Step 2: The number of independent variables that can be selected in the input layer of the ANN is calculated. In the current example of the CO_2-MEA-H_2O mixture, the evaluation of Eq. (14) yields ***F = 3***, which designates that only 3 variables can be selected.

Step 3: The parameters needed to perform the next steps of the algorithm are set. The ANN training parameters are independent variables, dependent variables, limits of the independent variables, architecture of the ANN, simulations to parameters ratio, sampling method, training algorithm, preprocessing function, loss function, accuracy target.

Independent variables: any set of independent variables can be chosen as long as the number of independent variables is equal to the degrees of freedom and the variables are independent of each other (i.e. all the apparent molar fractions of CO_2, MEA and H_2O cannot be chosen simultaneously due to the summation constraint). The fact that ANN can estimate the relationships between different numerical values provide the possibility to select any non-natural thermodynamic variable as independent variable (i.e. CO_2 loading can be selected as independent variable). The independent variables selected in the presented case study are α_{CO_2}, x_{MEA} and temperature (T).

Dependent variables: any number of dependent variables may be chosen. In addition, any variable can be chosen as a dependent variable as long as it is a function of at least one independent variable. In the present case study, there are 13 dependent variables: 9 liquid molar fractions (x_i), 3 vapor molar fractions (y_i) and the total pressure (P).

Limits of the independent variables: the upper and lower boundaries in which the surrogate model is valid are set for each one of the independent variables. The limits set for the surrogate ANN model in this work are presented in Table 1. They were specified so that the machine learning model can be used for applications under common industrial operation conditions in a CO_2 amine-scrubbing plant.

Architecture of the ANN: the number of hidden layers and the number of neurons in the hidden layer must be chosen. In this work a feedforward neural network (FFNN) was chosen as it is the simplest ANN architecture without internal cycles or loops. Therefore, it is expected that the lack of recursive operations will

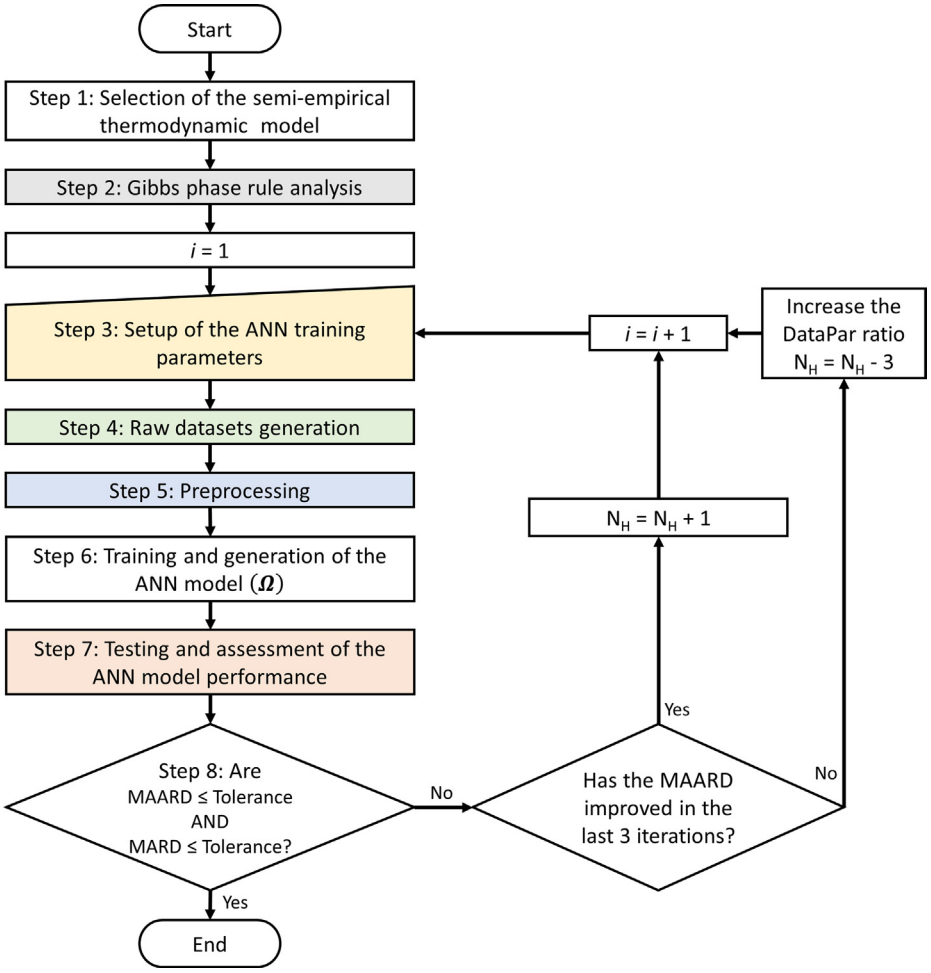

Fig. 3. Algorithm for the development of a surrogate thermodynamic machine learning based model.

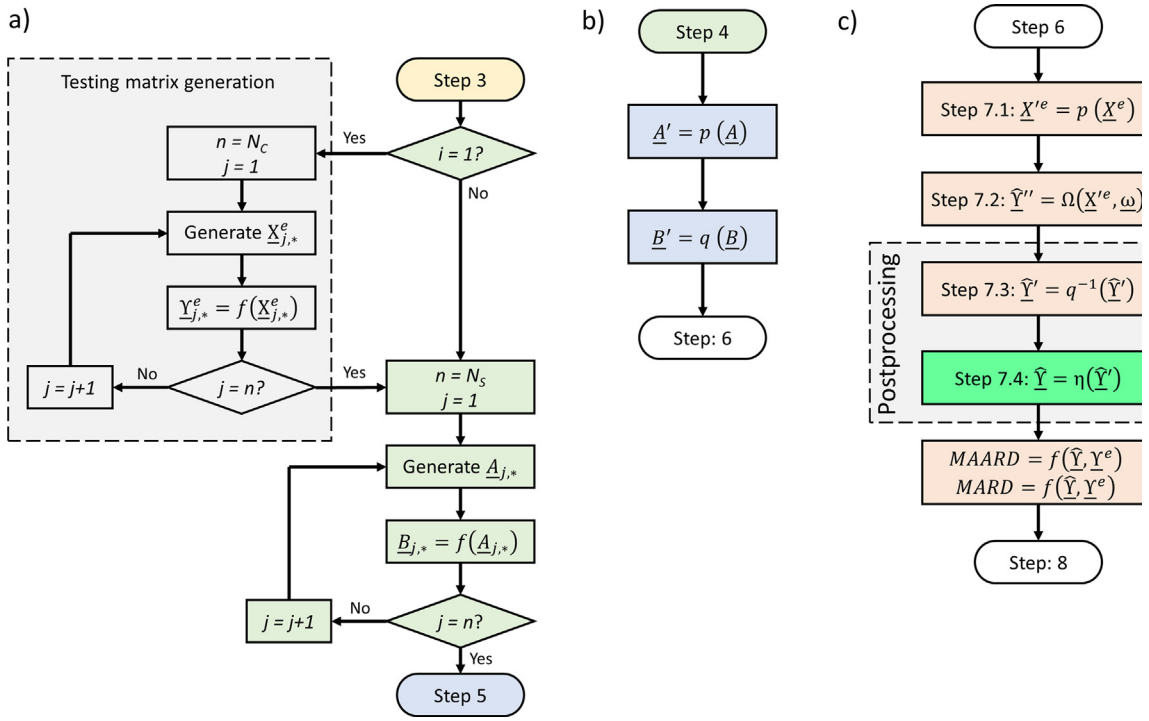

Fig. 4. Expanded algorithms of: a) step 4: raw datasets generation, b) step 5: preprocessing and c) step 7: testing and assessment of the ANN model performance.

Table 1
Limits of the independent variables for the case study.

Variable	Min	Max
$\alpha_{CO_2-MEA}/(molCO_2/molMEA)$	0.001	0.60
$MEAwt/\%$	0.1	0.60
T/K	293.15	393.15

create the most computationally efficient ANN-based surrogate models.

Additionally, the number of hidden layers for all ANN was fixed to 1 because a higher number of hidden layers is used for more complex purposes where deep learning is specifically needed, for example for image pattern recognition (Al-Saffar et al., 2017) or in the development of advanced artificial intelligence for intricate decision-making models for the video-game industry (Vinyals et al., 2019). Furthermore, there is no need to choose more complex architectures according to the universal approximation theorem that states that any continuous function can be approximated using a single hidden layer feed-forward ANN with a finite number of hidden neurons (Cybenko, 1989). The effect of the number of neurons in the hidden layer on the ANN model performance is analyzed and discussed in Section 3.1.

Simulations to parameters ratio: the number of simulations present in the training dataset is selected. Since the overall idea of our proposed methodology is to use data generated by a semi-empirical model to train an ANN, the number of data points generated by the base model is an important training variable. The simulations to parameters ratio is given by:

$$DataPar = \frac{N_S}{N_P} \tag{15}$$

Here *DataPar* is the simulations to parameters ratio, N_S is the number of datapoints in the datasets and N_P is the number of parameters in the ANN. The number of parameters in a single layer FFNN is calculated by:

$$N_P = (N_H N_I + N_H N_O) + (N_H + N_I) \tag{16}$$

where N_H is the number of neurons in the hidden layer, N_I is the number of inputs and N_O is the number of outputs in the FFNN. It is important to remark that *DataPar* must always be greater than 1 so that the ANN training can be an optimization problem. The effect of *DataPar* on the accuracy of the ANN models is studied in Section 3.1.

Sampling method: the values of the independent variables in the raw matrix must be generated with a sampling method. Three different sampling methods are analyzed in this manuscript: random, structured and combined. The random method generates the input raw matrix with a Monte Carlo sampling scheme. The structured method generates permuted vectors that form an evenly distributed grid. The combined scheme divides the number of simulations in two, the first half is generated with the random method while the second half with the structured method. All the values generated with the sampling methods must be within the limits shown in Table 1.

Training algorithm: selection of the optimization method used to fit ANN models. The present work focuses on the most prominent methods to train FFNN: the Levenberg-Marquardt backpropagation method and the Bayesian regularization method.

The Levenberg-Marquardt (LM) backpropagation algorithm is considered because, according to the Matlab documentation (Matlab, 2019), it is the recommended method due to its higher computational speed during the training of feedforward neural networks. The LM method uses a maximum neighborhood search method that is an hybrid between the Taylor series and gradient optimization methods (Marquardt, 1963). This method was later adapted and applied to perform the backpropagation training of ANN (Hagan and Menhaj, 1994). The Bayesian Regularization (BR) backpropagation algorithm combines the Bayesian interpolation method developed by MacKay (1992)) with the Levenberg-Marquadt (LM) optimization method (Dan Foresee and Hagan, 1997). It was chosen because it has been shown in several applications like in biological studies with mice (Okut et al., 2011) or in the cement industry (Garoosiha et al., 2019) that the Bayesian regularization may give better generalization properties than the LM, and thereby, better prediction capabilities. The effect of the training algorithm on the ANN performance is discussed in Section 3.1.

Preprocessing function: any mathematical function can be used to transform the raw data into preprocessed data. Data normalization is a common preprocessing function that scales the data within a certain range usually from 0 to 1. The normalization function is not shown in Fig. 3 or Fig. 4 because Matlab 2019b always includes this step inside the ANN model and therefore it is not necessary to explicitly program it. Caution is advised when using a different programming platform.

Loss function: the loss function assesses the performance of the ANN based on the preprocessed values.

The Matlab framework for performing the ANN training uses the Mean Square Error (MSE) as the optimization function. Nonetheless, in the case of models that are valid over a broad range of molar compositions, using MSE will focus the optimization on the higher values rather than distribute the error evenly throughout the dataset. To avoid this, the optimization function of the model was changed to Mean Relative Square Error (MRSE) instead of the default mean square error (MSE). The optimization function for MRSE has the form:

$$\min \quad \mathscr{L} = \frac{1}{N_S} \sum \left(\frac{\underline{B}' - \underline{\widehat{B}}'}{\underline{B}'} \right)^2 \tag{17}$$

where $\underline{B}'$ is the output matrix generated with the base model and $\underline{\widehat{B}}'$ is the output matrix from the ANN model.

Accuracy target: the degree of desired exactness of the surrogate model with respect to the base model must be specified (see step 8).

Step 4: Here, the datapoints are generated with the semi-empirical model. Fig. 4a shows that in the first iteration of the algorithm, the input matrix $\underline{X}^e$ and the output $\underline{Y}^e$ testing matrices are generated. This testing dataset is used in step 7 to determine if the ANN model can properly predict values for which the ANN model was not trained for. If the predictions of the testing dataset are significantly worse than the ones done with the training dataset, it means that the ANN is overfitted and cannot generalize (this usually happens if the DataPar value is too small). Note that a similar algorithm is used to generate the raw input matrix $\underline{A}$ and the raw output $\underline{B}$ matrices used for the ANN training of each one of the ANN models.

Step 5: The raw input matrix $\underline{A}$ and the raw output matrix $\underline{B}$ are preprocessed using transformation functions (see Fig. 4b). In the present work the matrices are transformed as follows: $\underline{A}' = \underline{A}$ and $\underline{B}' = \ln(\underline{B})$.

Using the natural logarithm of the output variables helps in addressing one of the main challenges in the modelling of VLE with ANN: the big variance between the orders of magnitude in the molar fractions of the individual electrolytes. A large variation in the orders of magnitude of the output variables results in an ill-

conditioned or badly scaled training dataset which reduces the accuracy of the model predictions.

An example of badly scaled data points can be obtained by comparing the orders of magnitude of the unreacted molar fraction of CO_2 (x_{CO_2}) of two different thermodynamic systems. The first case considers a thermodynamic system at $\alpha_{CO_2} = 10^{-5}$, $T = 293.15K$ and $w_{MEA} = 30\%$ which yields molar fraction $x_{CO_2} = 3x10^{-16}$. This small value is caused by the low CO_2 loading and the exothermic nature of the chemical reactions which promote the product formation. Hence, most of the solubilized CO_2 is chemically bound to the amine. On the other hand, the second scenario considers the system to be at $\alpha_{CO_2} = 0.60$, $T = 393.15$ and $w_{MEA} = 30\%$ which leads to an absolutely contrasting value of molar fraction $x_{CO_2} = 3 \times 10^{-3}$. This is because the high CO_2 loading together with the high temperature significantly decreases the molar fraction of CO_2 that can be chemically bound to MEA and, consequently, increases the amount of CO_2 solubilized by van der Waals forces.

In order to overcome this problem, a scaling through a logarithmic transformation can be performed. An additional advantage of this transformation is that it avoids the calculation of negative molar fractions when any of the molar compositions is close to 0.

Step 6: The ANN weight parameters $\underline{\omega}$ are calculated and the ANN function Ω is generated. This step needs the following ANN training parameters: training algorithm, ANN architecture, number of neurons in the hidden layer and the loss function.

In Matlab 2019b, the training datasets are usually divided into 3 parts: the training fraction, the validation fraction and the test ratio. We selected 0.90 for the training fraction, 0.1 for the validation fraction and 0 for the testing ratio. The testing ratio is used to assess the ANN generalization capability of the ANN model to predict values that were not used in the ANN training. Since the ANN model testing in this methodology is performed with an independent dataset, the value was set to be a very low value.

Step 7: The model is tested and assessed by evaluating the ANN model Ω using the testing matrix $\underline{X}^e$ as input and then comparing the output predictions with the testing output matrix $\underline{Y}^e$ (see Fig. 4c).

First, the preprocessing transformation function p is applied on $\underline{X}^e$ to obtain $\underline{X}'^e$. Then, the transformed matrix $\underline{X}'^e$ and $\underline{\omega}$ are used to evaluate Ω in order to get $\underline{\widehat{Y}}''$. The values of $\underline{\widehat{Y}}''$ do not have physical meaning yet because they are in a different mathematical space due to the preprocessing. Hence, a postprocessing procedure must be performed with an anti-transformation function of the form q^{-1} to get physically meaningful values $\underline{\widehat{Y}}'$. In this work, the molar compositions of the liquid and vapor phase in $\underline{\widehat{Y}}'$ are evaluated afterwards with a compositional normalization function (η) in order to comply with the restrictions of Eqs. (8) and (9). This process finally calculates $\underline{\widehat{Y}}$.

After postprocessing, the accuracy of the model can be evaluated by comparing $\underline{\widehat{Y}}$ and the testing output matrix $\underline{Y}^e$. The ANN model capabilities to replicate the base model can be assessed by calculating the "model average absolute relative difference" (MAARD) and "model average relative difference" (MARD) of each one of the variables between the base model and the ANN model. The equations for the MAARD and the MARD are:

$$AARD = \frac{1}{N_C} \sum_{N} \frac{\left|\widehat{Y} - Y^*\right|}{\sqrt{\widehat{Y}Y^*}} * 100\% \tag{18}$$

$$ARD = \frac{1}{N_C} \sum_{N_C} \frac{\widehat{Y} - Y^*}{\sqrt{\widehat{Y}Y^*}} * 100\% \tag{19}$$

where N_C is the total number of datapoints in the testing dataset ($N_C = 250,000$ in the case study).

Step 8: This decision-based step will stop the algorithm if the accuracy target is reached. In the case that the accuracy target is not reached, the algorithm estimates if the number of neurons in the hidden layer or *DataPar* should be increased. It should be remarked that although ANN can replicate the base model, one should specify reasonable values (e.g. do not specify very small MAARD).

3. Results and discussion

3.1. Surrogate model parametrization

Since the number of ANN training parameters is somewhat extensive, a factorial-based study was done to determine the best parameter selection. The analyzed ANN training parameters as well as their factors are presented in Table 2. A total of 54 ANN models were developed considering all permutations. These models were generated by only performing the steps from 1 to 7 of the algorithms shown in Fig. 3 and Fig. 4. Step 8 was not done because this parametric study aims at understanding the effect of the training parameters on the ANN model performance.

Training method: It can be seen from Table 3 that the best models were trained using the BR back propagation method. Moreover, the average MAARD of the models trained with the BR method was 23% better than the models trained with LM.

Sampling method: The results in Table 3 demonstrate that both the combined and random sampling methods provide quite similar results and there is not a clear trend on which one is better. The MAARD of the random sampling ANN models have, on average, a MAARD of 3.78% while for the combined method it is 4.72%.

Architecture of the ANN: The effect of the number of neurons in the hidden layer is clear, as the best models have 50 neurons. This is expected as the number of neurons and number of fitting parameters increase together. The average MAARD of all the models with 10, 30 and 50 neurons are 13.2%, 5.6% and 4.2% respectively, which confirms that a higher number of neurons increase the model accuracy.

Simulations to parameters ratio: As seen in Table 3, a higher DataPar has a higher chance of having higher accuracy. For instance, models #35, #33 and #31 have the same training parameters except for the DataPar and it is seen that the model with the highest DataPar has the best MAARD. It is important to note that the DataPar value does not have a significant impact in the computational performance of the surrogate model as it is only used in the ANN training. However, choosing a large DataPar (e.g. 100) may significantly increase the training time and, as seen in Table 3, the MAARD improvement is so minimal that it may not justify the considerable additional training time.

According to this study, we recommend using the Bayesian Regularization training method and the random sampling method to

Table 2
ANN training parameter values.

Variable	Values
Number of neurons in the hidden layer (N_H)	10/30/50
Simulations to parameters ratio (*DataPar*)	2/3/5
Training method	BR/LM
Sampling method*	R/S/C

* In sampling method R, S and C stand for random, structured and combined respectively

Table 3
ANN model performances using different training parameters.

ANN Model	# Neurons in the hidden layer	DataPar	Training method	Sampling method	Mean MAARD (%)	Mean MARD (%)
35	50	5	BR	Random	0.50	0.03
33	50	3	BR	Random	0.52	0.04
13	50	2	BR	Combined	0.52	−0.01
15	50	3	BR	Combined	0.54	0.00
17	50	5	BR	Combined	0.57	−0.01
31	50	2	BR	Random	0.57	0.09
36	50	5	LM	Random	0.63	0.11
16	50	3	LM	Combined	0.66	−0.02
18	50	5	LM	Combined	0.67	0.00
34	50	3	LM	Random	0.80	0.17

develop thermodynamic ANN surrogate models for new systems. However, the optimal number of neurons in the hidden layer and DataPar may vary from system to system. Therefore, we suggest to start the procedure shown in Fig. 3 with $N_H = 10$ neurons and $DataPar = 2$.

3.2. Surrogate model assessment

3.2.1. Statistical and graphical analysis

In this section, the application and validation of the model #35 is performed (the parameters can be obtained from the authors upon request). This model was chosen because it is ranked as the best model and because it has a MAARD of 0.50%. The MAARD and the experimental AARD (EAARD) of model #35 and the eNRTL model are presented in Table 4. All fitted variables in model #35 show good agreement with the base model, as none surpasses a MAARD value of 1.00%.

The EAARD values were calculated by comparing the model predictions against experimental data: 131 CO_2 partial pressure data points (Aronu et al., 2011), 80 total pressure data points (Aronu et al., 2011), vapor molar fractions data points (Hilliard, 2008) and 16 liquid phase speciation data points (Jakobsen et al., 2005). The operating conditions in the cited references that were outside the validity range of the ANN models were omitted. The MAARD seems to be insignificant as the difference between the EAARD of the base model and model #35 is negligible.

Parity plots between the base model and the ANN model #35 are presented in Fig. 5. The 1,000 datapoints used in Fig. 5 were randomly chosen from the testing dataset. It can be seen in Fig. 5-a-b that the highest residual errors are not the same as the largest relative deviations. The residual errors are higher in the carbamate x_{MEACOO^-} at high loadings because this is when the carbamate concentrations are larger, so even if the relative error is small at these conditions, the residual error will be high.

Fig. 5c-d on the other hand presents the vapor molar fraction of CO_2. The residual plot shows that at low loadings, the residuals are small. This is because at these loadings, the CO_2 is mostly absorbed in the liquid phase by chemical reactions, therefore the free CO_2 in the liquid solution does not exert high pressures of CO_2 in the vapor phase, resulting in a small CO_2 molar fraction. This gives low vapor molar fractions and, therefore, the residuals must be small. In contrast, at higher loadings the CO_2 is absorbed in the liquid phase by van der Waals forces as well, thus the pressure exerted by CO_2 on the vapor phase increases drastically, and the CO_2 vapor molar fraction is higher.

When the relative deviations are compared, the error is well distributed along the entire range of CO_2 loadings and temperatures. There is no trend that suggests that the CO_2 loading and temperature have any effect on the differences in CO_2 molar fractions between the ANN model #35 and the base model.

In order to show the VLE prediction accuracy, the total pressure of the CO_2-MEA-H_2O system was calculated at different conditions and compared against experimental data. The results are presented in Figs. 6–9, where the continuous lines represent the predictions of model #35. The data reported for total pressure in literature (Aronu et al., 2011) are mainly from medium to high pressures, hence there are few data points at low CO_2 loadings or temperatures. Fig. 6 shows that the cited experimental data and the model predictions are in good agreement over a broad range of temperatures and CO_2 loadings.

The speciation predictions of model #35 were calculated and are presented in Fig. 7 and compared against experimental speciation data points (Jakobsen et al., 2005). Even at low loadings, the smooth behavior of the speciation curves was accurately reproduced by the ANN model. It is important to underline that the estimation of the molar compositions need to be reasonably accurate as the liquid molar compositions are often used in kinetic and mass transfer models.

The experimental molar vapor fractions in Fig. 8 were calculated using Dalton's law and the reported partial pressures of CO_2, MEA and H_2O. Fig. 8 shows that the vapor molar compositions are properly predicted by the model #35 at different conditions of CO_2 loading, MEA wt% and temperature.

Although the developed ANN models in this work do not explicitly predict the CO_2 partial pressure P_{CO_2}, it can be calculated by multiplying the total pressure and the CO_2 vapor phase molar fraction. Since the ANN model #35 accurately estimates both variables independently, the calculated P_{CO_2} is in good agreement with the experimental data as shown in Fig. 9. This indicates that the surrogate model is not only accurate, but also represents the thermodynamic system in the same fashion as calculations done with traditional VLE models.

Table 4
Relative deviations between model #35 and the base model and experimental data.

Variable	Model #35 MAARD (%)	Base model EAARD (%)	Model #35 EAARD (%)
x_{H_2O}	0.05	–	–
x_{CO_2}	0.63	–	–
x_{MEA}	0.41	14.6	14.8
$x_{H_3O^+}$	0.86	–	–
x_{MEAH^+}	0.46	28.4	28.6
x_{OH^-}	0.57	–	–
$x_{HCO_3^-}$	0.41	34.6	34.5
$x_{CO_3^{2-}}$	0.67	27.0	26.9
x_{MEACOO^-}	0.35	12.5	12.6
y_{H_2O}	0.39	1.7	1.7
y_{CO_2}	0.79	23.8	23.9
y_{MEA}	0.50	31.5	31.5
P_T	0.41	13.9	14.0
P_{CO_2}	–	20.7	20.8
Overall (%)	0.50	24.4	24.5

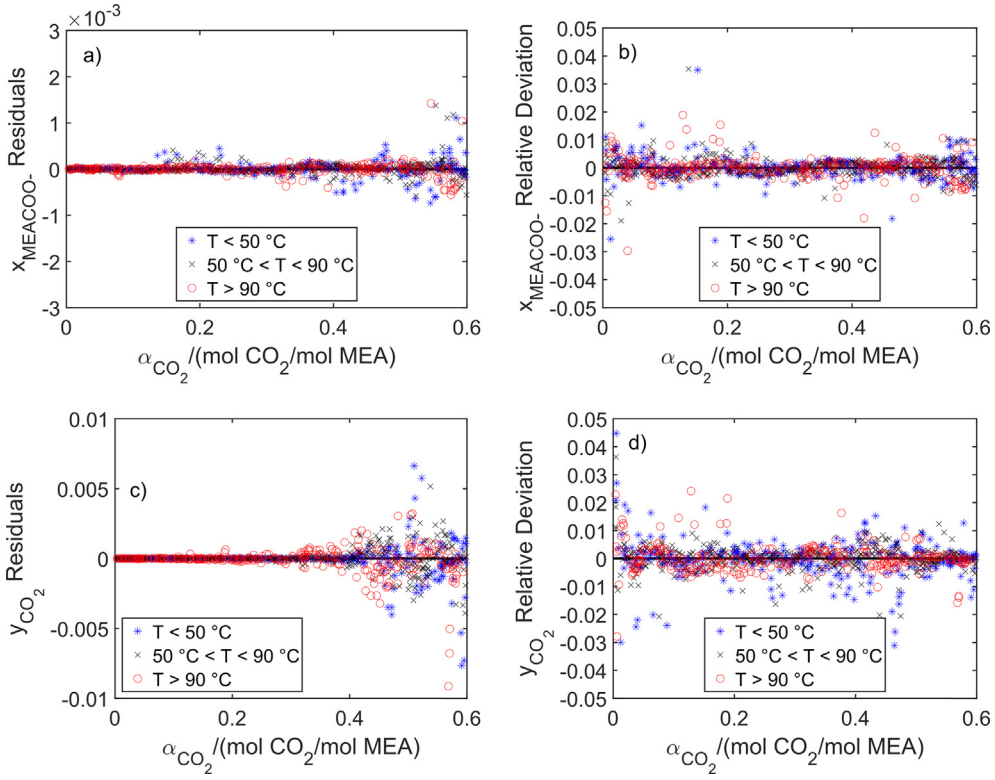

Fig. 5. Parity plots between the base model and the ANN model #35: a) x_{MEACOO^-} residual error parity plot, b) x_{MEACOO^-} relative deviation plot, c) y_{CO_2} residual parity plot and d) y_{CO_2} relative deviation plot.

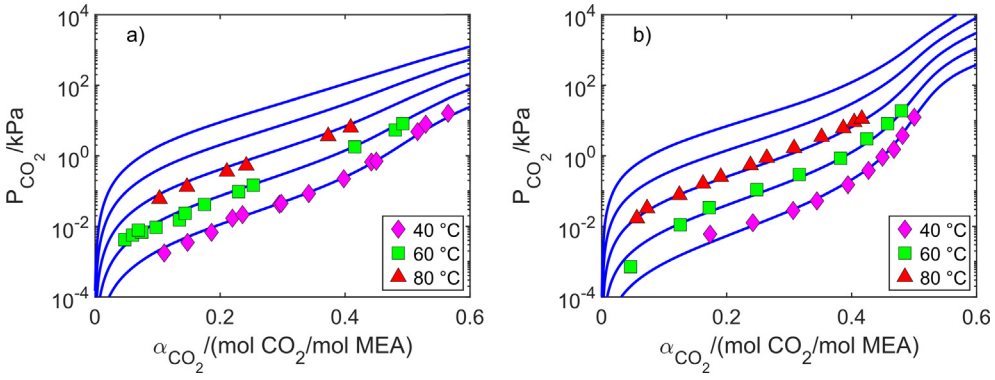

Fig. 6. Total pressure plots at different MEA weight fractions: a) $w_{MEA} = 15\%$ and b) $w_{MEA} = 60\%$. Continuous lines: model prediction with ANN model #35. Experimental data: (Aronu et al., 2011).

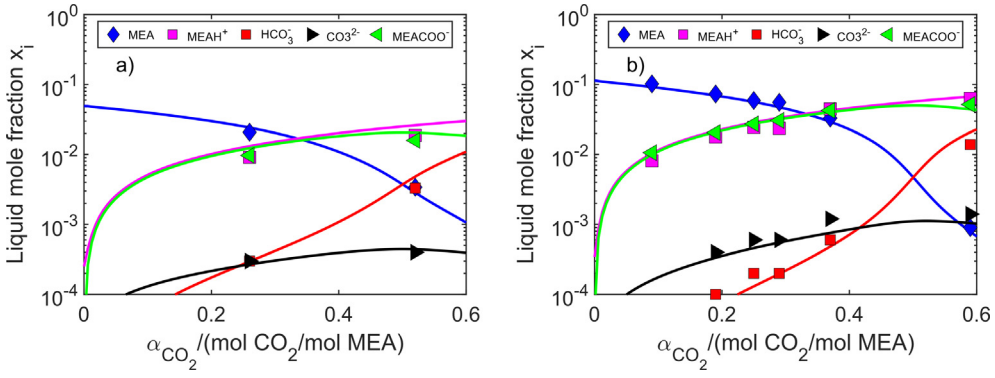

Fig. 7. Speciation plots at different conditions. a) $w_{MEA} = 30\%$ and 40 °C, b) $w_{MEA} = 45\%$ and 20 °C. Continuous lines: model prediction with ANN model #35. Experimental data: (Jakobsen et al., 2005).

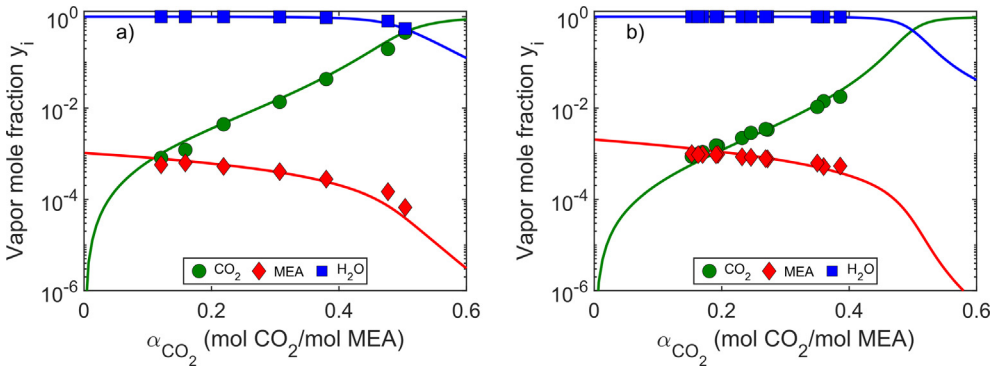

Fig. 8. Molar fraction plots at a) 3.5 MEA molarity and 60 °C and b) 7.0 MEA molarity and 40 °C. Continuous lines: model prediction with ANN model #35. Experimental data: (Hilliard, 2008).

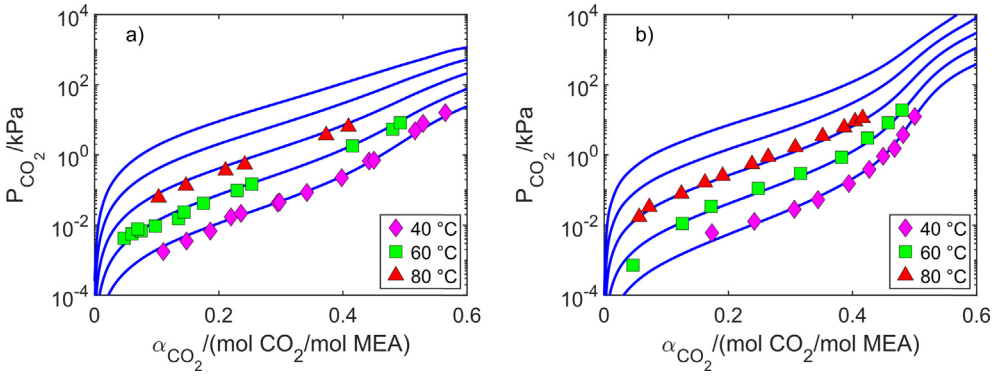

Fig. 9. CO_2 partial pressure plots at different w_{MEA}: a) 15% b) 60%. Continuous lines: model prediction with ANN model #35. Experimental data: (Aronu et al., 2011).

Fig. 10. presents a relative deviation plot for predicted P_{CO_2} and P_T as a function of α_{CO_2} compared with to experimental data. Fig. 10 shows four different sets of relative deviations between the predictions of the base model and model #35 and the experimental values of P_{CO_2} and P_T. It is seen that model #35 and the base model agree very well with each other, as almost all datapoints in Fig. 10 calculated by the base model are covered by the estimations of model #35. Additionally, the relative deviation of model #35 with respect to the experimental data is well distributed and there is no sign of bias.

The predictions of the base model and model #35 were plotted together so their differences could be illustrated. For instance, if only the datapoint of the ANN model #35 is seen, it is an indication that both models predicted the same values. Few datapoints show both model datapoints, whereby it can be concluded that the base model was accurately replicated by model #35.

3.2.2. Thermodynamic constraints

The black-box nature of ANN models may draw skepticism on the physical validity of the model, which in turn, may inhibit their implementation into process engineering frameworks. The ANN models are said to be of black-box nature because, to this day, it is not possible to deduce a physical meaning from the ANN parameters alone. Additionally, the difference between the semi-empirical model and the ANN models is that the first model considers the electronegativity constraint and the summation con-

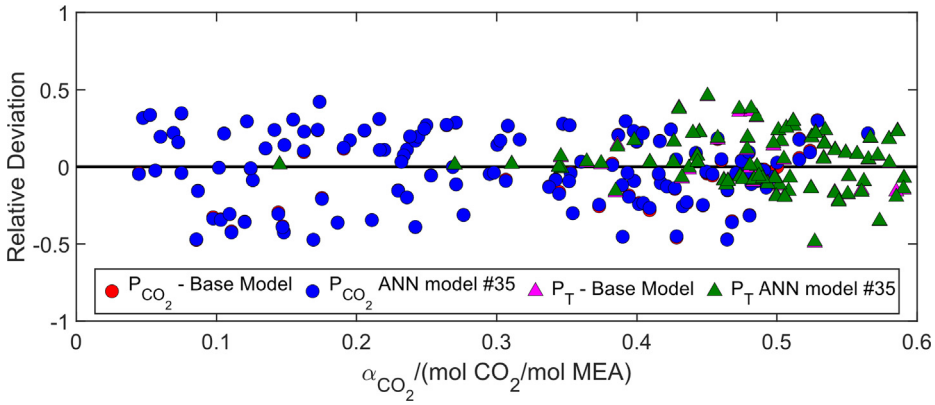

Fig. 10. Relative deviation plot between the experimental values and the predicted values from the ANN model #35 and the base model. Experimental data: (Aronu et al., 2011).

straints in the solution algorithm while the ANN models do not (the summation equation restrictions are used in the postprocessing calculations but not in the neural network itself).

Table 5 presents the mean difference between the values calculated by the ANN model #35 with respect to Eqs. (8)–(10) without performing a compositional normalization. The second column presents the absolute mean error when the molar fractions are compositionally normalized. Note that the absolute error of Eqs. (8) and (9) is 0 on the second column because the normalization redistributes the molar fractions so that their sum is 0.

It can be concluded from Table 5 that although there is a small error in the summation constraints, the fact that the absolute error in Eq. (10) is the same between the second and third column indicates that the error in the molar fractions is well distributed between all the species. The base model fulfills the electronegativity constraint with an error less than 10^{-15} while the error in the summation constraints is 0.

3.2.3. Why does the Gibbs phase rule matter?

Complying with the Gibbs phase rule is crucial, regardless if the model is either semi-empirical or machine learning based. For example, let us consider an over specified system where there are more fixed variables than degrees of freedom. In this case, the thermodynamic system will not have any physical meaning as it cannot exist at the specified conditions. The mathematical effect of overspecification in a thermodynamic system depends on the model type. If one tries to estimate the VLE behavior of a thermodynamic system with a semi-empirical model, a solution cannot be achieved as the mathematical model will be mathematically inconsistent. Unfortunately, an ANN does not have a similar "safe-lock" and it can predict VLE values that may seem reasonable at first glance, but do not have a physical meaning.

Consistent thermodynamic models allow evaluating thermodynamic quantities that are related to the VLE behavior through fundamental equations. In the context of CO_2 capture, an important solvent quantity is the CO_2 enthalpy of absorption which can be estimated using Eq. (13) and the VLE model.

In order to highlight the importance of the Gibbs phase rule and thermodynamic consistency in the ANN model development, an inconsistent model of the VLE system was formulated following the methodology previously presented but omitting the Gibbs phase rule analysis. The inconsistency was imposed on the model by specifying one extra variable and breaking the Gibbs phase rule. The "inconsistent model" was trained with 4 inputs: α_{CO_2}, MEA wt %, T and P_T. The model was developed and compared with the testing dataset, and a MAARD of 0.37% was obtained. The P_T values of the training dataset were given as inputs in the "validation" of the inconsistent model. Hence, the MAARD seems to be low and in agreement with the base model.

Fig. 11 and Fig. 12 show the CO_2 enthalpy of absorption on the left axis and the P_{CO_2} on the right axis, both as functions of α_{CO_2}. This was done demonstrate that a low MAARD is not necessarily a satisfactory indication that the model was "properly understood" by the ANN.

Fig. 11 presents the thermodynamic system at 40 °C, it shows that model #35 accurately reproduced the P_{CO_2} behavior of the base model as the curves of both models are overlapping completely. The heat of absorption predictions between the two models are in excellent agreement (MAARD = 0.30%). The difference between the experimental and the estimated CO_2 heat of absorption is caused by the inherent simplifications to the use of Eq. (13) (Kim et al., 2014; Sherwood and Prausnitz, 1962; Svendsen et al., 2011).

Table 5
Thermodynamic constraints check of ANN model #35.

Eq.	Absolute Error Without Normalization - 10^5	Absolute Error With Normalization - 10^5
(8)	2.34	0
(9)	8.19	0
(10)	12.6	12.6

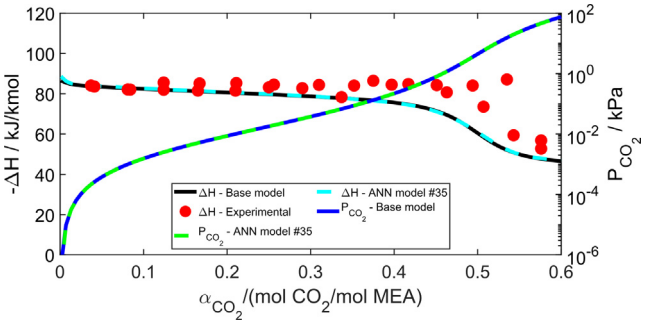

Fig. 11. Enthalpy of absorption calculated as a function of CO_2 loading using 30 MEA wt% at 40 °C. Experimental data: (Kim and Svendsen, 2007).

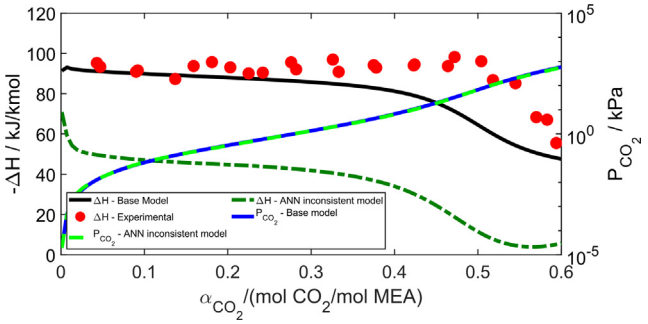

Fig. 12. Enthalpy of absorption calculated as a function of CO_2 loading using 30 MEA wt% at 80 °C. Experimental data: (Kim and Svendsen, 2007).

Fig. 12 on the other hand, presents the predictions of the inconsistent model and the base model at: $w_{MEA} = 30\%$, 120 °C and different loadings and pressures. The heat of absorption predictions with the inconsistent model are inaccurate (relative deviations up to 50%). This is a clear indication of a model inconsistency. An ANN model may look "good" in a statistical analysis even if it was with a wrong conception of the system physics. The inconsistent ANN model created a fallacious dependency of P_{CO_2} on the total pressure even though it is thermodynamically proven that, for this case, it should only be a function of α_{CO_2}, w_{MEA}, and T. This demonstrates that the misuse of machine learning or artificial neural networks can lead to correlation-causation fallacies.

If the P_{CO_2} predictions of the inconsistent model are observed, one would expect that the CO_2 heat of absorption estimations would be accurate as well, but they are not. This might raise the question: why are the heat of absorption values so inaccurate? This can be explained by considering the similitudes between the ANN and biological organisms. Much of the knowledge of animals or humans is based on experience or correlations. They learn to react accordingly to environmental stimuli by creating a cause-effect correlation that does not necessarily indicate that the living being has understood the root cause of the phenomenon. A similar situation occurs when training an ANN. During the process, the ANN learns and finds the correlations between the variables, but that does not mean that it found right dependencies of the system, it just found a mathematical correlation.

Another situation in which thermodynamic inconsistencies in ML models may occur is if there is an interdependence among

the input variables. Let us consider a situation where the ANN model was developed using the CO_2 molarity concentration, T and w_{MEA} as inputs. While the use of three input variables complies with the degrees of freedom of the system, the interdependency between the molar concentration and the temperature will produce thermodynamic inconsistencies if a physically unfeasible combination of input variable values is chosen. Once again, this may not be detectable by the ANN as the VLE predictions do not directly solve the thermodynamic equations. For this reason, concentrations on molarity basis are not recommended and it is preferable to use concentrations on a molal basis or a molar fraction basis to avoid unfeasible scenarios.

3.3. Surrogate model application

Since the goal of developing the surrogate model is to improve the computational time, the computational speed of the different calculation methods was compared. Table 6 presents the comparison of the computational speed of the gamma-phi calculations, different interpolation methods and the ANN surrogate model method proposed in this work. The molar fraction of CO_2 in the vapor phase was chosen as the output variable. A total of 100 batches with 50 simulations each were run in order to estimate the relative computational speed with respect to the gamma-phi calculations shown in Table 6. The values of each batch are different between each other and were generated randomly.

The interpolations were done using the Matlab built-in function interp3. Two interpolating algorithms were chosen: the linear interpolation and the cubic spline interpolation algorithm (de Boor, 1978). The best models with 10, 30 and 50 neurons in the hidden layer are also presented in Table 6. This table shows that the linear interpolation can accelerate the computational speed. However, its intrinsic relative error is high and the savings in computational time may not be high enough to compensate and justify the error.

Cubic spline interpolation has better accuracy when compared to the linear interpolation, but at the cost of lower computational speed. Making a balance between the relative computational speed and the MAARD, cubic spline interpolation appears to outperform the linear interpolation. It is important to underline, that due to the algorithm of the cubic spline interpolation, it may happen that the algorithm estimates values of y_{CO_2} less than 0 whenever there is a steep change in y_{CO_2} when the value is close to infinite dilution.

The machine learning models clearly outperforms the interpolation in both the computational speed and the capabilities for estimation accuracy. The main reason of the computational speed superiority is that the ANN calculations are simple and non-recursive calculations. Another reason for the machine learning outstanding computational characteristics is that the ANN structure allows calculating the entire input matrix in parallel (in this example there were 50 simulations in parallel) as opposed to the interpolation algorithms that must process each one of the inputs individually.

The last three rows of Table 6 suggest that there is a tradeoff between the number of hidden neurons and the computational speed. Fig. 13 presents how the speed of the ANN models is affected by the number of parallel simulations and the number of hidden neurons. The tradeoff between the number of neurons and the computational speed is evident as the model with 10 neurons has the fastest relative speed, however its MAARD is somewhat high, thus the models with 30 and 50 hidden neurons seem to be better. A higher number of neurons implies more operations inside an ANN, therefore it is important to set a reasonable target of the MAARD in order to avoid a large number of neurons that would reduce the computational speed.

From Fig. 13a it is evident that the model with 50 neurons is around 3 orders of magnitude faster than the gamma-phi calculations when performed with more than 100 simulations done in parallel. The computational gain at smaller number of parallel simulations is unclear, hence, an amplification was done and shown in Fig. 13b. The minimum relative speed is reached for all models when there is only one simulation. However, even at this minimum point the relative speed of all 3 models is 40 times the speed of the base model. At 30 parallel simulations the computational gain is around 3 orders of magnitude, while for 400 simultaneous calculations, the speed is around 2,500 times faster if the model with 50 neurons is used.

So, one may wonder: is it really important to speed-up the VLE computation time? If the surrogate model is to be used to model the behavior of a single vapor–liquid separator, the calculations would take 8.0×10^{-5} s instead of 3.2×10^{-3} s. Looking at this scenario, from the end user perspective there is not a noticeable difference that justifies the use of a surrogate model. But what

Table 6
Computational speed comparison between different VLE calculation schemes.

	Elements in the interpolation matrix	Relative computational speed	y_{CO_2} MAARD (%)
Gamma-phi calculations (base model)	–	1	–
Linear interpolation	1000	10	28.17
	8000	7	8.78
	27,000	6	4.67
Cubic spline interpolation	1000	8	9.20
	8000	6	1.74
	27,000	5	0.93
ANN with 10 neurons	–	1202	8.03
ANN with 30 neurons	–	987	1.12
ANN with 50 neurons	–	966	0.49

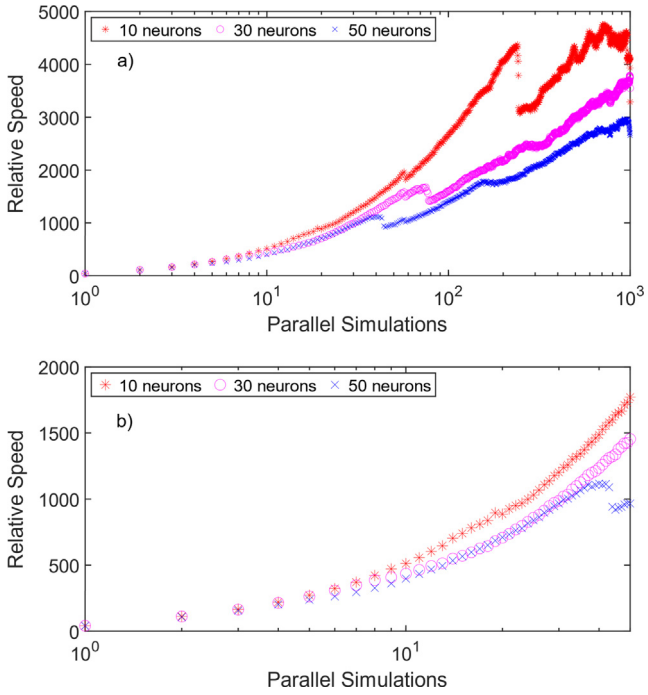

Fig. 13. a) Computational efficiency tradeoff between the number of neurons and the number of parallel simulations. b) Amplification of a).

happens if there is a complex problem that requires a non-steady state 2-D formulation? Let us assume that the geometry uses a 400×400 grid that requires a VLE evaluation in each discretized point. Furthermore, the time is discretized in 400 timesteps and it is assumed that each timestep requires 10 iterations to solve. Considering these conditions, the VLE simulations would take $400 \times 400 \times 100 \times 10 \times 3.2 \times 10^{-3}$ s = 510,000 s (142 h) if the base model is used. In contrast, the simulation would take 0.15 h to solve if the surrogate model with 50 neurons is used (this scenario assumes 400 simultaneous VLE simulations). The use of ANN surrogate models instead of performing the traditional semi-empirical models may thus give extraordinary computational advantages in complex problems that require VLE calculations in multidimensional problems. One of such examples may be the solution of the model needed for the characterization of aerosol emissions from CO_2 capture plants (Majeed and Svendsen, 2018).

4. Conclusions

An easy-to-implement method based on machine learning Artificial Neural Networks (ANN) was developed and proved to be a feasible alternative for the development of accurate, consistent and computationally fast surrogate models. The proposed method was employed to develop a surrogate machine learning thermodynamic model of a ternary system CO_2-MEA-H_2O. The surrogate models were based on a semi-empirical gamma-phi model framework (eNRTL for the liquid phase and Peng Robison for the vapor phase). The advantage of the proposed method is that it can be easily extrapolated to other thermodynamic quantities (e.g. activity coefficients, enthalpy calculations or heat capacities) and systems (e.g. more components or more phases).

A quantitative assessment of the effect of the training parameters on the prediction capabilities of the ANN models was performed. The first conclusion from this study is that a single hidden-layer FFNN architecture is enough to represent the behavior of reactive multiphase systems. Therefore, the need for more complex architectures (e.g. multiple hidden-layers or feedback) may be unnecessary as they may introduce extra training parameters or iteration loops that may make the ML model implementation difficult without significantly improving the prediction capabilities.

Through a statistical analysis, it was found that the accuracy of the surrogate models was improved using Bayesian regularization back-propagation algorithm and a random sampling method to generate the training datasets. The DataPar value is suggested to be as low as possible ($DataPar = 2$ for similar applications). The analysis involved the development of 54 ANN models wherefrom the one with the best prediction capabilities with respect to the semi-empirical model has a $MAARD = 0.50\%$.

It was shown that the integration of the Gibbs phase rule and physical constraints in the ANN framework is of utmost importance as their inclusion allows avoiding thermodynamic inconsistencies that may lead to inaccurate predictions.

It was also demonstrated that good prediction capabilities in ANN models are not necessarily a satisfactory indication of thermodynamic consistency, proper dependency between the variables or that the model complies with the physical constraints. Therefore, the integration of the Gibbs phase rule and physical constraints in the ANN model is a viable method to ensure that the thermodynamic model has the same behavior and dependencies as the semi-empirical models.

The developed surrogate ANN models can be around ~1000 times faster than rigorous calculation methods because the ANN models do not have recursive operations that jeopardize the computational efficiency of the calculations. Additionally, the computational speed of ANN surrogate models outperforms the ones of linear and non-linear interpolation methods.

CRediT authorship contribution statement

Andres Carranza-Abaid: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Hallvard F. Svendsen:** Software, Formal analysis, Resources, Writing - review & editing. **Jana P. Jakobsen:** Conceptualization, Formal analysis, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cesx.2020.100080.

References

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A.E., Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon 4. https://doi.org/10.1016/j.heliyon.2018.e00938.

Al-Saffar, A.A.M., Tao, H., Talab, M.A., 2017. Review of deep convolution neural network in image classification. In: Proceeding - 2017 Int. Conf. Radar, Antenna, Microwave, Electron. Telecommun. ICRAMET 2017 2018-Janua, 26–31. https://doi.org/10.1109/ICRAMET.2017.8253139.

Aronu, U.E., Gondal, S., Hessen, E.T., Haug-Warberg, T., Hartono, A., Hoff, K.A., Svendsen, H.F., 2011. Solubility of CO2 in 15, 30, 45 and 60 mass% MEA from 40 to 120°C and model representation using the extended UNIQUAC framework. Chem. Eng. Sci. 66, 6393–6406. https://doi.org/10.1016/j.ces.2011.08.042.

Bastani, D., Hamzehie, M.E., Davardoost, F., Mazinani, S., Poorbashiri, A., 2013. Prediction of CO2 loading capacity of chemical absorbents using a multi-layer perceptron neural network. Fluid Phase Equilib. 354, 6–11. https://doi.org/10.1016/j.fluid.2013.05.017.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer Science +Business Media LLC..

Brúder, P., Grimstvedt, A., Mejdell, T., Svendsen, H.F., 2011. CO2 capture into aqueous solutions of piperazine activated 2-amino-2-methyl-1-propanol. Chem. Eng. Sci. 66, 6193–6198. https://doi.org/10.1016/j.ces.2011.08.051.

Brúder, P., Lauritsen, K.G., Mejdell, T., Svendsen, H.F., 2012. CO 2 capture into aqueous solutions of 3-methylaminopropylamine activated dimethylmonoethanolamine. Chem. Eng. Sci. 75, 28–37. https://doi.org/10.1016/j.ces.2012.03.005.

Chen, C.-C., Evans, L.B., 1986. A local composition model for the excess Gibbs energy of aqueous electrolyte systems. AIChE J. 32, 444–454. https://doi.org/10.1002/aic.690320311.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control. Signals, Syst. 2, 303–314. https://doi.org/10.1007/BF02551274.

Dan Foresee, F., Hagan, M.T., 1997. Gauss-Newton approximation to bayesian learning. IEEE Int. Conf. Neural Networks - Conf. Proc., vol. 3, pp. 1930–1935. https://doi.org/10.1109/ICNN.1997.614194.

Daneshvar, N., Zaafarani Moattar, M.T., Abedinzadegan Abdi, M., Aber, S., 2004. Carbon dioxide equilibrium absorption in the multi-component systems of CO2 + TIPA + MEA + H2O, CO2 + TIPA + Pz + H2O and CO2 + TIPA + H2O at low CO2 partial pressure: Experimental solubility data, corrosion study and modeling with artificial neural netwo. Sep. Purif. Technol. 37, 135–147. https://doi.org/10.1016/j.seppur.2003.09.004.

Dash, S.K., Samanta, A.N., Bandyopadhyay, S.S., 2011. Solubility of carbon dioxide in aqueous solution of 2-amino-2-methyl-1-propanol and piperazine. Fluid Phase Equilib. 307, 166–174. https://doi.org/10.1016/j.fluid.2011.05.009.

de Boor, C., 1978. A Practical Guide to Splines. Springer-Verlag, New York.

Gabrielsen, J., Michelsen, M.L., Stenby, E.H., Kontogeorgis, G.M., 2005. A model for estimating CO2 solubility in aqueous alkanolamines. Ind. Eng. Chem. Res. 44, 3348–3354. https://doi.org/10.1021/ie048857i.

Garg, S., Shariff, A.M., Shaikh, M.S., Lal, B., Suleman, H., Faiqa, N., 2017. Experimental data, thermodynamic and neural network modeling of CO2 solubility in aqueous sodium salt of l-phenylalanine. J. CO2 Util. 19, 146–156. https://doi.org/10.1016/j.jcou.2017.03.011.

Garoosiha, H., Ahmadi, J., Bayat, H., 2019. The assessment of Levenberg–Marquardt and Bayesian Framework training algorithm for prediction of concrete

shrinkage by the artificial neural network. Cogent Eng. 6, 1–14. https://doi.org/10.1080/23311916.2019.1609179.
Ghasemian, N., Kalbasi, M., Pazuki, G., 2013. Experimental study and mathematical modeling of solubility of CO2 in water: application of artificial neural network and genetic algorithm. J. Dispers. Sci. Technol. 34, 347–355. https://doi.org/10.1080/01932691.2012.667293.
Golzar, K., Modarress, H., Amjad-Iranagh, S., 2016. Evaluation of density, viscosity, surface tension and CO2 solubility for single, binary and ternary aqueous solutions of MDEA, PZ and 12 common ILs by using artificial neural network (ANN) technique. Int. J. Greenh. Gas Control 53, 187–197. https://doi.org/10.1016/j.ijggc.2016.08.008.
Gonzalez Viejo, C., Torrico, D.D., Dunshea, F.R., Fuentes, S., 2019. Development of artificial neural network models to assess beer acceptability based on sensory properties using a robotic pourer: a comparative model approach to achieve an artificial intelligence system. Beverages 5, 33. https://doi.org/10.3390/beverages5020033.
Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the marquardt algorithm. IEEE Trans. Neural Netw. 5, 989–993. https://doi.org/10.1006/brcg.1996.0066.
Hamzehie, M.E., Fattahi, M., Najibi, H., Van der Bruggen, B., Mazinani, S., 2015. Application of artificial neural networks for estimation of solubility of acid gases (H2S and CO2) in 32 commonly ionic liquid and amine solutions. J. Nat. Gas Sci. Eng. 24, 106–114. https://doi.org/10.1016/j.jngse.2015.03.014.
Hamzehie, M.E., Mazinani, S., Davardoost, F., Mokhtare, A., Najibi, H., Van der Bruggen, B., Darvishmanesh, S., 2014. Developing a feed forward multilayer neural network model for prediction of CO2 solubility in blended aqueous amine solutions. J. Nat. Gas Sci. Eng. 21, 19–25. https://doi.org/10.1016/j.jngse.2014.07.022.
Hartono, A., Ahmad, R., Usman, M., Asif, N., Svendsen, H.F., 2020. Solubility of CO2 in 0.1M, 1M and 3M of 2-amino-2-methyl-1-propanol (AMP) from 313 to 393K and model representation using the eNRTL framework. Fluid Phase Equilib. 511. https://doi.org/10.1016/j.fluid.2020.112485.
Hilliard, M.D., 2008. A Predictive Thermodynamic Model for an Aqueous Blend of Potassium Carbonate, Piperazine, and Monoethanolamine for Carbon Dioxide Capture from Flue Gas. Dr. thesis Tech. Univ. Texas Austin 1083.
Jakobsen, J.P., Krane, J., Svendsen, H.F., 2005. Liquid-phase composition determination in CO2-H2O-alkanolamine systems: An NMR study. Ind. Eng. Chem. Res. 44, 9894–9903. https://doi.org/10.1021/ie048813+.
Jou, F.Y., Otto, F.D., Mather, A.E., 1994. Vapor-liquid equilibrium of carbon dioxide in aqueous mixtures of monoethanolamine and methyldiethanolamine. Ind. Eng. Chem. Res. 33, 2002–2005. https://doi.org/10.1021/ie00032a016.
Kim, I., Hoff, K.A., Hessen, E.T., Haug-Warberg, T., Svendsen, H.F., 2009. Enthalpy of absorption of CO2 with alkanolamine solutions predicted from reaction equilibrium constants. Chem. Eng. Sci. 64, 2027–2038. https://doi.org/10.1016/j.ces.2008.12.037.
Kim, I., Hoff, K.A., Mejdell, T., 2014. Heat of absorption of CO2 with aqueous solutions of MEA: new experimental data. Energy Procedia 63, 1446–1455. https://doi.org/10.1016/j.egypro.2014.11.154.
Kim, I., Svendsen, H.F., 2007. Heat of absorption of carbon dioxide (CO2) in monoethanolamine (MEA) and 2-(aminoethyl)ethanolamine (AEEA) solutions. Ind. Eng. Chem. Res. 46, 5803–5809. https://doi.org/10.1021/ie0616489.
Lloret, J.O., Vega, L.F., Llovell, F., 2017. A consistent and transferable thermodynamic model to accurately describe CO2 capture with monoethanolamine. J. CO2 Util. 21, 521–533. https://doi.org/10.1016/j.jcou.2017.08.018.
Luo, X., Hartono, A., Hussain, S., Svendsen, H.F., 2015. Mass transfer and kinetics of carbon dioxide absorption into loaded aqueous monoethanolamine solutions. Chem. Eng. Sci. 123, 57–69. https://doi.org/10.1016/j.ces.2014.10.013.
MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4, 415–447. https://doi.org/10.1162/neco.1992.4.3.415.
Majeed, H., Svendsen, H.F., 2018. Characterization of aerosol emissions from CO2 capture plants treating various power plant and industrial flue gases. Int. J. Greenh. Gas Control 74, 282–295. https://doi.org/10.1016/j.ijggc.2018.04.016.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11, 431–441. 10.1137/0111030.
Matlab, 2019. trainlm: Levenberg-Marquardt backpropagation [WWW Document]. Mathworks Doc. URL https://se.mathworks.com/help/deeplearning/ref/trainlm.html (accessed 12.20.19).
IPCC, 2018. Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. In: Maycock, T., Tignor, M., Waterfield, T. (Eds.), Lonnoy. Press.
McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Phys., 113–133
Mirarab, M., Sharifi, M., Ghayyem, M.A., Mirarab, F., 2014. Prediction of solubility of CO2 in ethanol-[EMIM][Tf2N] ionic liquid mixtures using artificial neural networks based on genetic algorithm. Fluid Phase Equilib. 371, 6–14. https://doi.org/10.1016/j.fluid.2014.02.030.
Norouzbahari, S., Shahhosseini, S., Ghaemi, A., 2015. Modeling of CO2 loading in aqueous solutions of piperazine: Application of an enhanced artificial neural network algorithm. J. Nat. Gas Sci. Eng. 24, 18–25. https://doi.org/10.1016/j.jngse.2015.03.011.
Oexmann, J., Kather, A., 2009. Post-combustion CO2capture in coal-fired power plants: Comparison of integrated chemical absorption processes with piperazine promoted potassium carbonate and MEA. Energy Procedia 1, 799–806. https://doi.org/10.1016/j.egypro.2009.01.106.
Okut, H., Gianola, D., Rosa, G.J.M., Weigel, K.A., 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. Genet. Res. (Camb) 93, 189–201. https://doi.org/10.1017/S0016672310000662.
Peng, R., 1976. P-R.a new equation of state. Proc. Natl. Acad. Sci. USA 15, 11–18.
Pitzer, K.S., 1973. Thermodynamics of electrolytes. I. Theoretical basis and general equations. J. Phys. Chem. 77, 268–277. https://doi.org/10.1021/j100621a026.
Plesu, V., Bonet, J., Bonet-Ruiz, A.E., Chavarria, A., Iancu, P., Llorens, J., 2018. Surrogate model for carbon dioxide equilibrium absorption using aqueous monoethanolamine. Chem. Eng. Trans. 70, 919–924. https://doi.org/10.3303/CET1870154.
Prausnitz, J.M., Lichtenthaler, R.N., Gomes de Azevedo, E., 1999. Molecular Thermodynamics of Fluid-Phase Equilibria. Prentice-Hall.
Putta, K.R., Pinto, D.D.D., Svendsen, H.F., Knuutila, H.K., 2016. CO2 absorption into loaded aqueous MEA solutions: Kinetics assessment using penetration theory. Int. J. Greenh. Gas Control 53, 338–353. https://doi.org/10.1016/j.ijggc.2016.08.009.
Raksajati, A., Ho, M.T., Wiley, D.E., 2013. Reducing the cost of CO2 capture from flue gases using aqueous chemical absorption. Ind. Eng. Chem. Res. 52, 16887–16901. https://doi.org/10.1021/ie402185h.
Rochelle, G.T., 2009. Amine scrubbing for CO2 capture. Science (80-.). 325, 1652–1654.
Russel, S.J., Norvig, P., 2009. Artificial Intelligence: A Modern Approach, third ed. Prentice-Hall, Upper Saddle River, New Jersey.
Sherman, B.J., Ciftja, A.F., Rochelle, G.T., 2016. Thermodynamic and mass transfer modeling of carbon dioxide absorption into aqueous 2-piperidineethanol. Chem. Eng. Sci. 153, 295–307. https://doi.org/10.1016/j.ces.2016.07.019.
Sherwood, A.E., Prausnitz, J.M., 1962. The heat of solution of gases at high pressure. AIChE J. 8, 519–521. https://doi.org/10.1002/aic.690080419.
Svendsen, H.F., Hessen, E.T., Mejdell, T., 2011. Carbon dioxide capture by absorption, challenges and possibilities. Chem. Eng. J. 171, 718–724. https://doi.org/10.1016/j.cej.2011.01.014.
Vinyals, O., Babuschkin, I., Czarnecki, W.M., Mathieu, M., Dudzik, A., Chung, J., Choi, D.H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J.P., Jaderberg, M., Vezhnevets, A.S., Leblond, R., Pohlen, T., Dalibard, V., Budden, D., Sulsky, Y., Molloy, J., Paine, T.L., Gulcehre, C., Wang, Z., Pfaff, T., Wu, Y., Ring, R., Yogatama, D., Wünsch, D., McKinney, K., Smith, O., Schaul, T., Lillicrap, T., Kavukcuoglu, K., Hassabis, D., Apps, C., Silver, D., 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 575, 350–354. https://doi.org/10.1038/s41586-019-1724-z.
Wu, X., Wang, M., Liao, P., Shen, J., Li, Y., 2020. Solvent-based post-combustion CO2 capture for power plants: A critical review and perspective on dynamic modelling, system identification, process control and flexible operation. Appl. Energy 257,. https://doi.org/10.1016/j.apenergy.2019.113941 113941.
Zarenezhad, B., Aminian, A., 2011. Predicting the vapor-liquid equilibrium of carbon dioxide+alkanol systems by using an artificial neural network. Korean J. Chem. Eng. 28, 1286–1292. https://doi.org/10.1007/s11814-010-0492-0.
Zhang, Z., Li, H., Chang, H., Pan, Z., Luo, X., 2018. Machine learning predictive framework for CO2 thermodynamic properties in solution. J. CO2 Util. 26, 152–159. https://doi.org/10.1016/j.jcou.2018.04.025.