Energy and AI 2 (2020) 100037

Contents lists available at ScienceDirect

Energy and AI

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

Prediction of sorption enhanced steam methane reforming products from machine learning based soft-sensor models

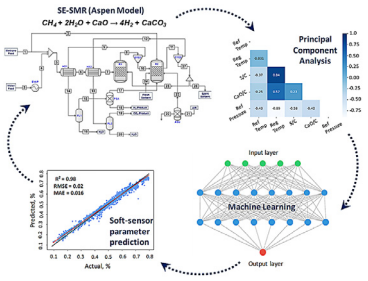
Paula Nkulikiyinka[a], Yongliang Yan[a], Fatih Güleç[a,b], Vasilije Manovic[a], Peter T. Clough[a,*]

[a] Energy and Power Theme, School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK
[b] Advanced Materials Research Group, Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, UK

HIGHLIGHTS

- Random forest and neural network models are applied to the SE-SMR process
- Soft sensor models are developed and offer accurate parameter predictions
- Models can be applied in industrial H_2 production to find unmeasurable parameters

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 September 2020
Received in revised form 6 November 2020
Accepted 6 November 2020
Available online 11 November 2020

Keywords:
Machine learning
Artificial neural network
Soft sensor
Sorption enhanced steam methane reforming
Calcium looping

ABSTRACT

Carbon dioxide-abated hydrogen can be synthesised via various processes, one of which is sorption enhanced steam methane reforming (SE-SMR), which produces separated streams of high purity H_2 and CO_2. Properties of hydrogen and the sorbent material hinder the ability to rapidly upscale SE-SMR, therefore the use of artificial intelligence models is useful in order to assist scale up. Advantages of a data driven soft-sensor model over thermodynamic simulations, is the ability to obtain real time information dependent on actual process conditions. In this study, two soft sensor models have been developed and used to predict and estimate variables that would otherwise be difficult direct measured. Both artificial neural networks and the random forest models were developed as soft sensor prediction models. They were shown to provide good predictions for gas concentrations in the reformer and regenerator reactors of the SE-SMR process using temperature, pressure, steam to carbon ratio and sorbent to carbon ratio as input process features. Both models were very accurate with high R^2 values, all above 98%. However, the random forest model was more precise in the predictions, with consistently higher R^2 values and lower mean absolute error (0.002-0.014) compared to the neural network model (0.005-0.024).

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
CaO/C	Sorbent to Carbon Ratio
CaL	Calcium Looping
MAE	Mean Absolute Error
MSE	Mean Square Error
PC	Principal Component
RF	Random Forest
RMSE	Root Mean Square Error
S/C	Steam to Carbon Ratio
SE-SMR	Sorption-Enhanced Steam Methane Reforming
SMR	Steam Methane Reforming
WGS	Water Gas Shift

1. Introduction

Aside from the imperative need to reduce greenhouse gas emissions from fossil fuel use, there are many other reasons why blue hydrogen, produced from fossil fuels with CO_2 capture, is likely to be at the forefront of future energy systems. For instance, it can be produced in several ways and from numerous resources, both renewable and non-renewable, and upon combustion, no greenhouse gases or other emissions are pro-

* Corresponding author.
E-mail address: p.t.clough@cranfield.ac.uk (P.T. Clough).

https://doi.org/10.1016/j.egyai.2020.100037

duced. Hydrogen can also be used to decarbonise several sectors that are otherwise difficult to reduce greenhouse gas emissions from [1].

Although hydrogen can be produced from various resources, both renewables and non-renewables, the most common method at present for production of H_2 is via steam methane reforming (SMR). In this process, methane is reacted with steam producing carbon monoxide and H_2, then the carbon monoxide is reacted with steam in water gas shift (WGS) reactors to produce CO_2 and more H_2.

The reactions that take place in order to produce H_2 from methane are reforming (reaction 1) and the WGS Reaction (reaction 2), giving the overall SMR reaction (reaction 3):

$$CH_4 + H_2O \leftrightarrow CO + 3H_2 \quad \Delta H_r, 298K = +206\ kJ/mol \tag{1}$$

$$CO + H_2O \leftrightarrow CO_2 + H_2 \quad \Delta H_r, 298K = -41\ kJ/mol \tag{2}$$

$$CH_4 + 2H_2O \leftrightarrow CO_2 + 4H_2 \quad \Delta H_r, 298K = +165\ kJ/mol \tag{3}$$

Sorbent Enhanced Steam Methane Reforming (SE-SMR) is a process which integrates calcium looping (CaL) and SMR. The principal reactions in CaL are carbonation (forward reaction) and calcination (backward reaction) described in equation 4, and the combination of equations 3 (SMR) and 4 (CaL) gives the overall SE-SMR reaction shown in equation 5:

$$CaO + CO_2 \leftrightarrow CaCO_3 \quad \Delta H_r, 298K = -178\ kJ/mol \tag{4}$$

$$CH_4 + 2H_2O + CaO \rightarrow 4H_2 + CaCO_3 \quad \Delta H_r, 298K = -13\ kJ/mol \tag{5}$$

One of the challenges faced in SE-SMR, and in various other thermochemical processes, is the inability to measure and thus control important variables in a reactor and process control systems due to a variety of physical conditional limitations. These variables range from parameters such as product composition, concentration of reactants, spatial-temporal temperature and pressure profiles, and the limitations stem from reasons including financial, physical, inaccuracies, or competing influences [2]. A solution to this challenge is to incorporate artificial intelligence (AI) into the energy sector through the use of soft sensors, which are inferential estimators using predictions based on data to provide real time digital approximations as to the conditions inside the reactor [3,4]. A simpler definition of a soft sensor is that it is a predictive model based on large quantities of data available on an industrial process, which can either be first principle models (white-box models) or data driven models (black-box models). White box models depend on actual mechanical data of the process, whereas the latter uses historically collected process data, which makes black-box models far more practical and readily applicable to process plants [5].

The principle on which soft sensors work is based on the estimation of quality through a mathematical model that uses all available measured process variables [6]. In further detail, the idea of inference estimation (soft sensing) is to approximate the values of primary variables (e.g. gas composition) by the easily measured secondary variables (e.g. temperature), which are correlated to the primary variables.

For example, if Y is the dependent variable (soft sensor output, e.g. concentration) and X_1, $X_2 \ldots X_n$ are the independent inputs (e.g. temperature, flow), the mathematical model for the output prediction can be described with Equation 6, where $\bar{X} = \Sigma(X_1, X_2 \ldots X_n)$ and is the vector of the inputs, B is the vector of the model coefficients, and ε is the output error [6].

$$Y = F(\bar{X}, B) + \varepsilon \tag{6}$$

In recent years, soft sensors have been widely studied and used in industrial process control to improve the quality of product and assure safety in production, and additionally act as backup when hardware sensors are unavailable or unsuitable [7,8]. In addition, the desired measurements which are the key indicators of process performance, are normally ascertained by off-line sample analyses or on-line product quality analysers, which are often costly and require frequent and high cost maintenance, and sometimes add delay to the process as the measurements cannot be used as feedback signals for quality control systems [9].

Table 1
Aspen Plus SE-SMR baseline process model parameters.

Baseline Conditions	
Steam to carbon ratio	4
Sorbent to carbon ratio	1
Pressure	1 bar
Reformer temperature	650 °C
Regenerator temperature	850 °C
Pressure Swing Adsorber (PSA) efficiency	95%
Sorbent carrying capacity	20%
Cyclone separation efficiency	100%

The use of soft sensors is not new and has been extensively implemented in process industries for the past three decades. Recently, machine learning, as a whole, has been used in carbon capture technologies, mainly focused on the process modelling and control of amine-based post-combustion capture systems [10,11]. However, to the best of our knowledge, machine learning algorithms have not been applied to sorption enhanced steam methane reforming to develop a soft sensor model of the process. In this study, artificial neural network (ANN) and random forest (RF) algorithms are employed to construct soft sensor models. A comparison of the models is then conducted against the SE-SMR reactions taking place to evaluate their ability to predict operating parameters including reactor gas concentrations, reformer CH_4 conversion and overall process H_2 purity at specified conditions.

2. Methodology

The development of the data-driven soft sensor followed the typical systematic approach as found in literature, with the following being the main elements of the process [5,12]:

1 Data collection, with the secondary variables selected according to knowledge of the process
2 Data pre-processing
  a Outlier detection using univariate analysis
  b Normalisation where necessary
  c Correlation/redundancy elimination using multivariate principle component analysis (PCA)
3 Regression model development
4 Validation of the soft sensor on independent process data

For the first step, data collection, a database of 13,756 data points was obtained for the machine learning soft sensor, which were produced from an Aspen Plus model that was developed to simulate the SE-SMR process and then applying a sensitivity analysis to the process across a range of commonly employed operating ranges.

2.1. Process Configuration and Simulation

The SE-SMR process was simulated using Aspen Plus V10 (Fig. 1) using the Peng-Robinson property method and the equilibrium calculations were performed, initially using the baseline conditions (Table 1).

The reformer and regenerator reactors were simulated with RGibbs blocks. CO_2-abated hydrogen (stream 18) was obtained from separating hydrogen from the other flue gas components using a pressure swing adsorber with an efficiency of 95%. Water was separated from the system, using flash separators, producing two product streams of water (streams 16 and 20) and the product stream of CO_2 (stream 19). An oxy-fired calciner was added to the process in order to provide heat for the regeneration of the carbonated sorbent. In this case the heat requirement is met by combustion of methane (stream 2) and PSA tail gas (stream 17),

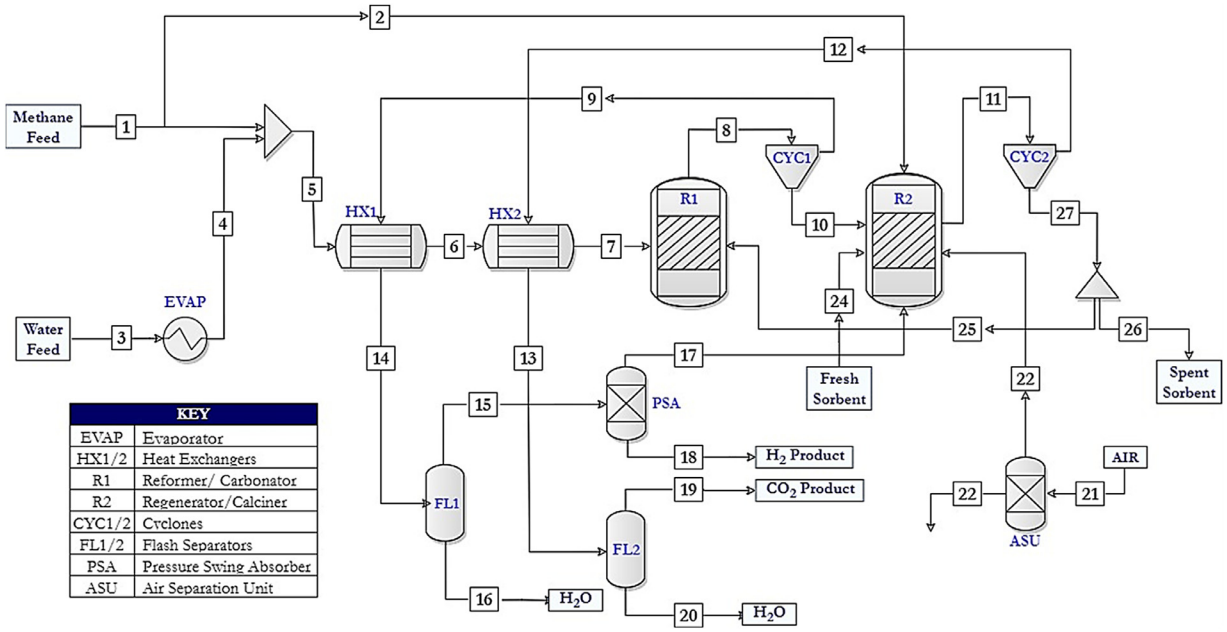


Fig. 1. Process flow diagram of the SE-SMR process simulated in Aspen Plus.

Table 2
Parameter ranges used in the sensitivity analysis to build up the machine learning database. *In order to obtain a ratio, the flowrates for the steam, methane or sorbent, were altered between the given ranges, and a ratio was then calculated

Parameter	Range	Intervals
Steam to carbon ratio	1–7	Various*
Sorbent to carbon ratio	0.1–5	Various*
Reformer pressure	1–10 bar	1, 5, 10 bar
Reformer temperature	500–800 °C	Every 50 °C
Regenerator temperature	700–1000 °C	Every 50 °C

using oxygen (stream 22) from an air separation unit, producing carbon dioxide and steam. To ensure these were the only gases formed, excess oxygen was supplied to ensure complete combustion of methane. Two heat exchangers were used in order to elevate as well as maintain the heat in the streams to and from the reactors, thus achieving an efficient heat integration of the reactors' thermal content. With these baseline conditions, the H_2 purity was found to be 91.65%.

A sensitivity analysis was performed to generate a large database (13,756) that was implemented in the machine learning models. The parameters that were chosen to vary were: steam to carbon ratio (S/C), sorbent to carbon ratio (CaO/C), reformer pressure, as well as reformer and regenerator temperatures. They were varied over the ranges shown in Table 2. These parameters were chosen as they form the independent control/performance variables for the SE-SMR process, and the ranges were chosen as they were typically reported in the available literature [13–16]. The data sampling was arranged with one change per datapoint and the rationale for this was to act as if it were an actual reactor and therefore experiencing slight changes in the conditions in every and any direction. For example, for one datapoint the reformer temperature, regenerator temperature, S/C and reformer pressure were all kept constant with changing CaO/C. Then for the next datapoint, the next reformer temperature was chosen, and again the other three parameters were kept constant, for changing CaO/C, and so forth so that each parameter was altered and was combined with all changes in the other parameters as well. Additionally, as seen in Table 2, whilst the pressures and temperatures had set intervals, the ratios varied between the given ranges as they were calculated from changing flowrates. Again, this was chosen to replicate practical operation as the flowrates would be the parameters altered as opposed to picking specific ratios. What was ensured however was that the ratios were kept within the specified ranges.

Although oxy-fuel combustion usually occurs in 70–80% CO_2 and 30–20% O_2 since burning methane in O_2 releases a lot of heat, instead of recycling CO_2 from stream 12, only CO_2 from the PSA off-gas was recycled. This is because it has been shown in literature, that increasing the oxygen concentration in the calciner reduces the amount of recycled gas (CO_2, H_2O and impurities); thus removing the energy necessary to heat the recycled gas stream [17]. To ensure excess air was supplied to the process at all times, a calculator block was used, which took into account the methane fed to and the CO_2 from the PSA off-gas recycled to the calciner and made appropriate adjustments to the flow from the ASU. In addition, a design specification was used to ensure the quantity of fresh sorbent supplied was equal to the spent sorbent being purged from the calciner. The specification also ensured the relative sorbent make up (Fresh limestone/Sorbent circulation rate) was 0.04 which is in the range of typical rates presented in the literature [18].

2.2. SE-SMR process model validation

In order to validate the SE-SMR process model and thus to ensure that the results from the machine learning models were an accurate representation of the system, the literature data were used, as a comparison with the model predictions is presented in Table 3.

Table 3 shows that the model predictions are in good agreement with the experimental data found in literature. The average percentage error across the H_2 concentration and CO_2 concentration is 2.2% and 42.2%, respectively. The deviation for H_2 concentration is deemed acceptable to validate the model as 2.2% margin of error for experimental results, particularly for this sample size. On the other hand, whilst the CO_2 concentration error is much higher, it should be considered that CO_2 concentrations are relatively low, and a small deviation in process performance can make a significant difference.

2.3. Machine learning model selection

The input variables, or features, chosen were S/C (steam to carbon ratio), CaO/C (sorbent to carbon ratio), reformer pressure and the reformer and regenerator temperatures. The desired soft sensor outputs to be obtained from these features were the gaseous concentrations within

Table 3
Comparison of literature data against Aspen Plus data for validation purposes.

	Li et al. [19]	Johnsen et al. [20]	Arstad et al. [21]
Process parameters			
S/C ratio	5	3	4
CaO/C ratio	0.125	0.18	1.1
Reformer pressure (bar)	1	1	1
Reformer temperature (°C)	630	600	575
Regenerator temperature (°C)	850	850	895
Literature data			
H_2 concentration (%)	95.6	98.4	95.8
CO_2 concentration (%)	2.7	0.4	0.7
Model predictions			
H_2 concentration (%)	97.3	96.9	98.9
CO_2 concentration (%)	1.5	0.5	0.3

the reformer (H_2O, CO, H_2) and regenerator (H_2O, CO_2). Another valued output to be found during process operation was the reformer CH_4 conversion efficiency (Equation 7.) defined as the moles of CH_4 reacted, divided by the moles of CH_4 going entering the reformer [22].

$$CH_4\ conversion\ (\%) = \frac{n_{CH_{4,in}} - n_{CH_{4,out}}}{n_{CH_{4,in}}} \times 100 \tag{7}$$

With the five input features and the desired output performance indicators selected, a machine learning algorithm was developed. As the information gathered was all numerical, very little data pre-processing was required at this stage. One action taken on the data was randomising the order of the data as it had been collected in order of increasing variable e.g. S/C, CaO/C, temperature, pressure. This was an action taken to prevent the models from overfitting the data, by recognising a sequential pattern. Additionally, the data was scaled once it was implemented into Python, more specifically normalised to have a value between zero and one, which prevented the models attributing a higher 'importance' to larger values, and thus giving it higher influence on the overall results.

The next step in the process to develop the soft sensor, was to select appropriate machine learning models that were suitable for the type of data collected. The type of machine learning chosen was 'supervised learning' since the system was presented with labelled input-output data in order to learn from it. Of the two types of supervised learning categories- classification and regression, regression was the type to be used as the data was continuous and numerical. The two models chosen were random forest and artificial neural network.

3. Machine learning theory and calculations

3.1. Random forest

The random forest model is a combination of decision tree predictors, where each tree depends on the values of a random vector sampled independently, as well as with an equal distribution across all trees [23]. They are able to overcome the problem of overfitting as they introduce the randomness into each individual tree by growing each tree on a bootstrap sample of the training data and only using random subsets of the available predictors when creating the split. The formula for the predictions is given in equation 8, where x' is unseen samples, f_b is each tree and B is the number of trees after the spilt.

$$f(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \tag{8}$$

Two principal features of the RF algorithm are out-of-bag (OOB) error estimates and feature importance rankings. Out-of-bag samples refers to a third of samples not being used for fitting a particular regression tree in the forest. The OOB score provides a similar measure of the model's generalization error and is calculated by constructing the predictor for each observation with the trees that apply for that specific observation that was out-of-bag. The overall OOB score is then computed as the average error of all OOB predictions [24]. The feature importance is calculated by recording the improvement in the split-criterion at each split and in each tree. These improvements are then attributed to the feature that were used to split these particular nodes and summed over all trees and for all feature separately [25].

The primary reasons for the random forest algorithm being beneficial for this type of dataset includes that they are quite resistant to overfitting, with the addition of more trees and they do not require complicated or time-consuming processes of variable selection [26]. Additionally, RF can handle missing values and maintain the accuracy of a large proportion of data. An annotation of how this random forest works through the 'trees' is shown in supplementary materials (Figure S.1)

3.2. Artificial Neural Network (ANN)

ANN is a non-parametric, information processing, statistical paradigm, which does not require any pre-assumption of the input–output relationship. ANNs consist of three principal layers: an input layer, several hidden layers and an output layer, which are made up of various numbers of connected 'neurons' or 'nodes' in each layer. The quantity of hidden layers defines the depth of the architecture. The data is transferred from the input node, along to the hidden layer nodes, eventually reaching the output layer to achieve the results or predictions [27]. It has also been shown that ANNs have the ability to approximate any non-linear system with high interpolation capacity. ANNs have a high ability to effectively approximate non-linear systems which is down to the use of one or more hidden layers and non-linear transfer functions in the hidden layer's neurons. A unique attribute of ANNs is that they are adaptive, but with the requirement that the input data is presented to the network in order to recognize the process i.e. it must be trained first, therefore, ANN is a 'supervised learning' algorithm. The training process of the ANNs aims to achieve the highest coefficient of determination (R^2) and the lowest root mean square error (RMSE) and mean absolute error (MAE) equations 9 and (10), where n is the quantity of data required for the network training, y_i is the predicted output and t_i is the target output. This is done by fine-tuning the weights and biases according to the training algorithms [10,28].

$$RMSE = \sqrt{\frac{1}{n}\left[\sum_{i=1}^{n}\left(t_i - y_i\right)^2\right]} \tag{9}$$

$$MAE = \frac{1}{n}\left[\sum_{i=1}^{n}\left|t_i - y_i\right|\right] \tag{10}$$

The use of ANN as a tool for nonlinear soft sensing modelling has been widely used in the recent years, particularly regarding the prediction accuracy and saving of computational costs in many industrial fields [29]. Some examples of the applications include estimation of outputs based on industrial data [11], reduction of costs involved in computational methods [30], and prediction of behaviour for control and

Table 4
Random Forest feature importance for each output parameter. (Ref = Reformer, Reg = Regenerator, T = Temperature, P = Pressure)

Level of importance	Reformer [H_2O]	Reformer [CO]	Reformer [H_2]	Regenerator [H_2O]	Regenerator [CO_2]	[CH_4] reformer conversion
Most	S/C	Ref T	S/C	CaO/C	Reg T	S/C
	Reg T	Ref P	Reg T	S/C	CaO/C	Ref T
	Ref T	S/C	Ref T	Reg T	Ref T	Ref P
	Ref P	Reg T	Ref P	Ref T	Ref P	Reg T
Least	CaO/C	CaO/C	CaO/C	Ref P	S/C	CaO/C

optimization systems [31,32]. Benefits of this model include that it has the ability to model highly non-linear functions and can be trained to accurately generalise when presented with new, unseen data [33].

3.3. Multivariate analysis

As previously mentioned in the Section 3.1, a feature commonly used within the random forest model is the feature importance (Table 4), which is found by the OOB data. When the training set for the current tree is drawn by sampling with replacement, as explained in section 3.1, only a third of the cases are left out of the sample. This OOB data is used to get a running unbiased estimate of the classification error as trees are added to the forest, and also used to get estimates of feature importance, which has proven to be unbiased in many tests [23], (full list of data in supplementary materials Table S.2). In addition, another popular method to quantify the importance of the features on the impact it has on a model, is to permute the values of each feature and measure the effect of permutation on the accuracy of the model. This essentially creates a cut-off point in the feature importance values [34]. An example of this is applied to the CO_2 concentration in the regenerator. Table 4 indicates that regenerator temperature and the sorbent to carbon ratio are the most influential in the CO_2 concentration, and it is clear from permuting the values of the other three features, that the former two are the only important features in predicting CO_2 concentration from the calciner, which is expected. From this it can also be inferred that the value of 0.1 for variable importance acts as a good cut-off point as the R^2 value remained at >0.99 with the permuted features. This was confirmed by testing the other concentrations and permuting values until a significant change in accuracy was found, and across each parameter it was clear that features with an importance greater than 0.1 was considered 'important'.

The equivalence of this application in ANN is backward stepwise elimination to the ANN model (assessing change to RMSE with each input node removal) [35]. However, it has been reported in literature [36,37] that the method of backward stepwise elimination is not always a preferable method of quantifying variable importance, as the method is usually only successful in measuring the most influential variable as opposed to the whole input dataset.

ANN is often seen as a black box, from which it is very difficult to extract useful information for another purpose like feature importance/explanations [38]. Although it is possible, for this reason, to analyse the input features for the ANN model and compare against the random forest variable importance, principal component analysis (PCA) was performed as an effective procedure for the determination of input parameters [39]. PCA is a mathematical model that is used to reduce the dimensionality of a dataset, whilst retaining most of the variation in the dataset, which allows for better interpretation of the data. Some features in a dataset measure related properties and are effectively redundant and therefore should be removed for effective analysis, which is what PCA aims to do. The reduction of features is achieved by identifying the directions (principal components) that the variation is highest [40,41].

The PCA was calculated in Python using 'scikit-learn's PCA package', firstly to check the interaction and redundancies of the five input variables. Fig. 2 shows the correlation plot of the inputs. From this, it is inferred that the reformer pressure and regenerator temperature, are inversely correlated, meaning as one increases the other decreases. It can also be seen that regenerator temperature has little linear correlation with reformer temperature, but is positively correlated to steam to carbon ratio, as the former value is close to zero, and the latter, close to one. Additionally, it is evident that none of the input variables are redundant as none of the variables correlate perfectly throughout, i.e. none of the values are significantly close to one, so they are independent of each other, therefore all inputs are retained for the model development.

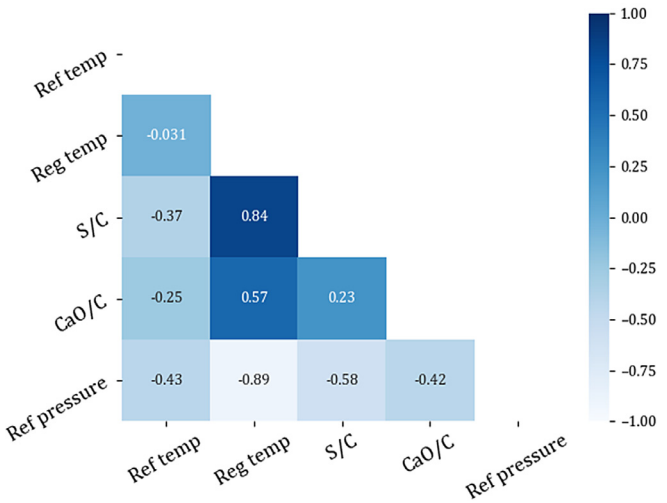


Fig. 2. Correlation heatmap of input features (Ref = Reformer, Reg = Regenerator, temp = Temperature).

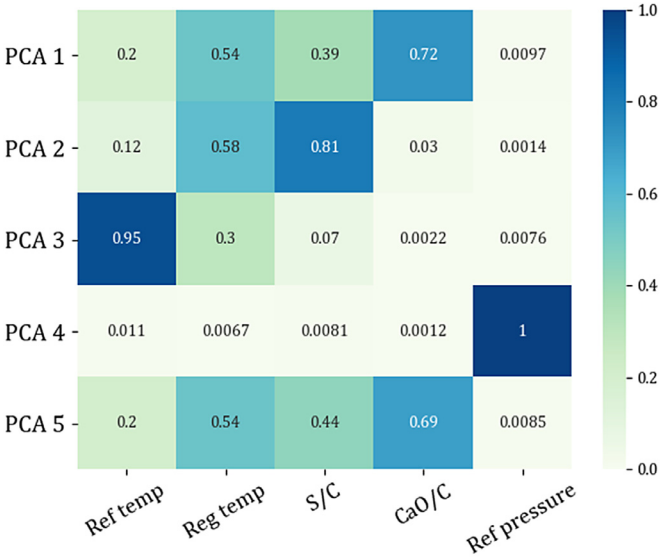


Fig. 3. Heatmap showing dependencies of 5 principle components on input features (Ref = Reformer, Reg = Regenerator, temp = Temperature).

Secondly, Fig. 3 was generated to examine the interrelation among a set of variables in order to identify the underlying structure of those

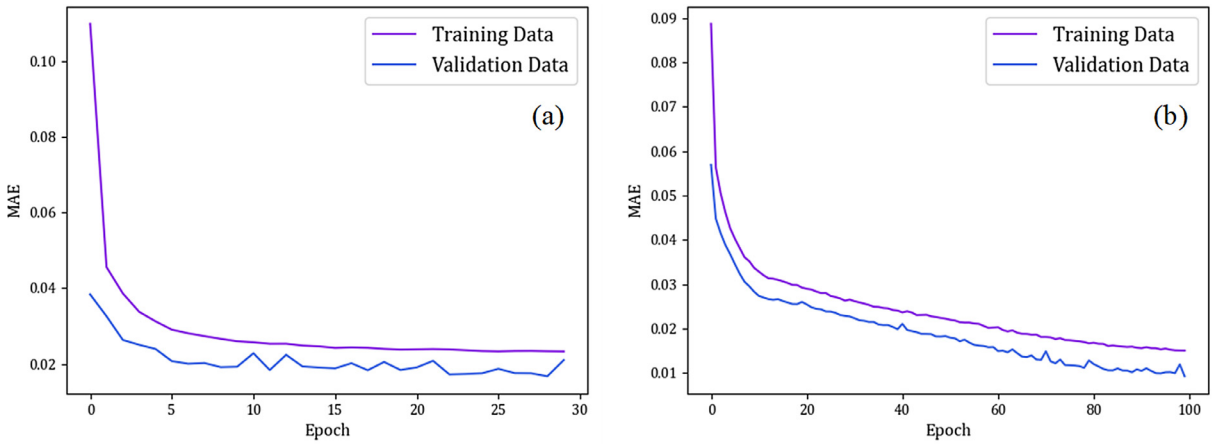


Fig. 4. Number of epochs vs MAE for reformer H_2O concentration: a) 30 epochs, b) 100 epochs.

variables. As mentioned before, PCA allows a reduction in components by looking at the independence (Fig. 2) as well as how they contribute to the variance, i.e. how the variables combine to form the principal components. This can be found by measuring the variance ratio of the principal components.

The results for this dataset was found to be that the first PC explains 32.1% of the variance, PC2 explains, 21.2%, PC3 and PC4 explain 20% each and PC5 explains 0.07%. This indicates a fairly even distribution of variable influence across the first four principal components, and therefore not essentially necessary to remove any parameters. On relatively small data sets, and in examples from literature [40], the first two components normally contribute to around 70–95% of the total variance, hence why PCA is normally depicted with two principal components. PCA becomes a valid method for feature selection, only when the most importance features, are the ones that happen to have to most variation in them, which is usually not true, and therefore features with low variation should not automatically be disregarded [42].

In addition to explaining the variance, Fig. 3 also explains feature importance. From this plot it can be calculated that the top three features that contribute to the five PCs of reformer methane conversion, are the reformer and regenerator temperature, and the steam to carbon ratio, which corresponds to the random forest variable importance results (Table 4).

3.4. Machine learning model setup

The models were coded in Python 3.8 using the PyCharm integrated development environment with the Keras and TensorFlow libraries as the backend. Keras is a library of open sources of the neural network developed in Python, which is focused on minimization, modularity, and scalability. Additionally, TensorFlow is an open-source software library which provides an interface for expressing and executing various machine learning algorithms [43]. Firstly, with the ANN model, various network configurations were tested trialled and compared with their R^2 and MAE values. The elbow-curve method was employed to find the configuration with the least number of nodes per layer that still resulted in high prediction accuracy metrics and minimised computational time.

Alongside the number of nodes per layer, the number of epochs and batch size are two important variables that influence the run time of the model as well as the accuracy of prediction. Simple descriptions of these parameters are:

- Batch size: The number of samples worked through by the model before updating the internal node parameters.
- Epoch: The number of times that the machine learning model works through the full dataset provided whilst training.

The optimal combination was found, by methods recommended in the literature, through multiple iterations and by plotting network accuracy vs epochs [44].

The metrics used to give the accuracy of the models were the coefficient of determination (R^2), the MAE and the root mean square error (RMSE). An indication of a well fitted and highly accurate model was one with low RMSE and MAE values as well as a high R^2. Fig. 4 shows plots of accuracy using the MAE, on the training and validation datasets over training epochs, for CH_4 conversion. Whilst it is clear the variation is lesser with more epochs, the actual MAE at the end of both iterations is very close and the difference is almost negligible after ~10 epochs. Due to this, 10 epochs was tested against the dataset to further reduce computation time, however, it was found that this resulted in error and therefore there is a minimum limit of epochs for the size of dataset used, which in this case was 30.

In literature, it is suggested that a larger number of epochs gives higher accuracy for prediction as the model is able to go through the dataset more times. However, with this dataset and model scenario, it was found that a sufficiently high R^2 and MSE were obtained with low batch sizes and epochs. Reasons for this outcome is thought to be down to the use of a relatively simple and clean dataset obtained from thermodynamic calculations in Aspen Plus, as opposed to using experimental data that will contain experimental error leading to deviations in the dataset. A model with 100 epochs (Fig. 4b) and batch size of 10 gave an R^2 value of 0.95 whereas a model with 30 epochs (Fig. 4a) and batch size of 10, and a lower model run time, gave R^2 value of 0.92. The small difference in error from these two examples hides the computational time required, thus, to minimise computational time (from 907 seconds to 245 seconds), and error a low epoch size (30) was applied.

For the model developed, the batch size chosen was 2 and 30 epochs. To give an insight into what that meant for the data provided:

- With a dataset with 13,756 samples, the dataset was divided into 6,878 batches, each with 2 samples.
- The model weights were updated after each batch of 2 samples. One epoch will involve 6,878 batches i.e. 6,878 updates to the model.
- With 30 epochs, the model will go through the whole dataset 30 times.
- Resulting in a total of 206,340 batches during the entire training process.

An additional factor in the ANN setup to be considered was the activation function used, which defines the output signal of that node, given an input or set of inputs [45]. From literature [11,39], it was found that whilst the Rectified Linear Unit (ReLU), was the most commonly used due to its speed, functions such as hyperbolic tangent (tanh) and sigmoid functions were more appropriate functions to be used with the

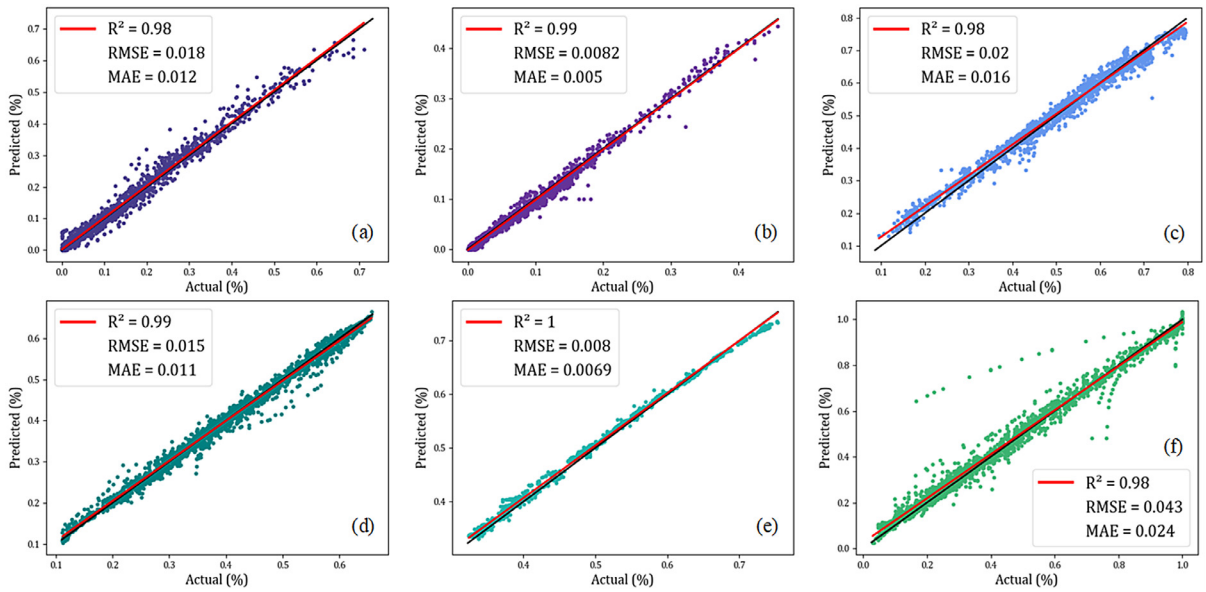


Fig. 5. Actual vs Predicted plots by the ANN algorithm for: a) reformer [H_2O], b) reformer [CO], c) reformer [H_2], d) regenerator [H_2O], e) regenerator [CO_2], f) reformer CH_4 conversion modelled. Black lines in each figure represent $y = x$. Red lines indicate the line of best fit.

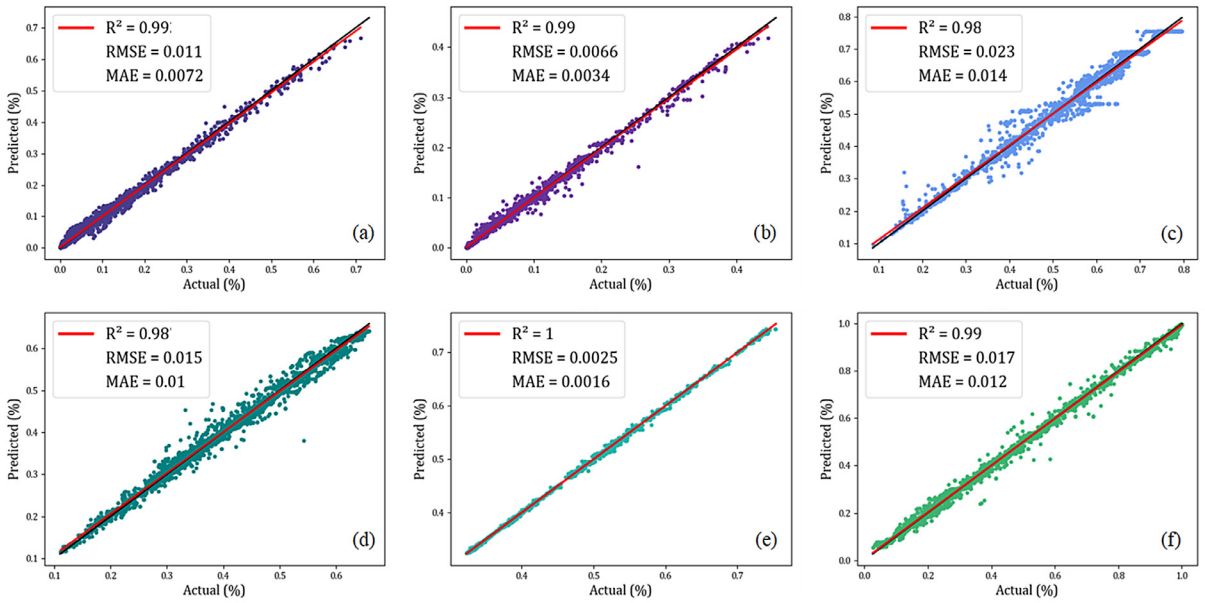


Figure 6. Actual vs Predicted plot by the random forest algorithm for: a) reformer [H_2O], b) reformer [CO], c) reformer [H_2], d) regenerator [H_2O], e) regenerator [CO_2], f) reformer CH_4 conversion modelled. Black lines in each figure represent $y = x$. Red lines indicate the line of best fit.

dataset provided. This was investigated in order to reach the optimal function and it was found that a combination of both the ReLU and sigmoid functions in between the layers of nodes gave the highest accuracy metrics. A comparative view of the functions is shown in supplementary materials (Figure S.2).

The calculations involved in the random forest model are less dependent on the inputs from the user as the model has a more automatic analysis process [46]. What does have to be specified however is the subsample size (the number of preselected directions for splitting the data) the tree depth (which can be specified in different ways) and the number of trees. However, there is a general consensus that implemented default values for these parameters produce good empirical performance in prediction, which is partially why random forests are so popular as a machine learning tool [47].

4. Results and discussion

The results are graphically presented in the form of predicted against actual plots, which shows the accuracy of the model. These plots indicate how well the predicted data fits to with actual data, and whether the patterns in the actual data be reproduced by themodel [48]. A model that performs well would produce a scatter of data that are close to the $y = x$ line. Figs. 5 and 6 depict the calculated Aspen data as 'Actual' data plotted against the 'Predicted' data from the two machine learning models. From both figures and the supporting metrics, it is evident that the random forest performed better than the neural network in the ability to predict the concentration of the gas in each reactor, due to achieving higher R^2 and lower RMSE and MAE values (presented in Table 5). Another comparison of the two models is the prediction of the

Table 5
Neural network vs random forest soft-sensor concentration predictions

	Neural Network		Random Forest	
	R^2	MAE	R^2	MAE
Reformer [H_2O]	0.98	0.012	0.99	0.007
Reformer [CO]	0.99	0.005	0.99	0.003
Reformer [H_2]	0.98	0.016	0.98	0.014
Regenerator [H_2O]	0.99	0.011	0.99	0.010
Regenerator [CO_2]	1.00	0.007	1.00	0.002
CH_4 Conversion	0.98	0.024	0.99	0.012

conversion of CH_4 within the reformer, Fig. 5f (ANN) and Fig. 6f (RF). In this comparison, both the neural network and random forest give a high R^2 value (0.98 and 0.99, respectively), indicating that they produce an accurate prediction of the efficiency of the reaction taking place in the reformer, at any given moment. In practical terms, this translates to allowing an operator to observe if the set conditions (pressure, temperature, flowrates) are giving a sufficient conversion rate and product purity in real time and make appropriate changes to achieve the desired product.

In the plots, the red line represents the line of best fit for the data, with the black line representing $y = x$. For the neural network, the one main difference is that there is a higher spread of data, and this can be explained by MAE value, which gives the magnitude of model prediction error, and is higher for the neural network. For all plots it is clear that that line of best fit almost overlaps the $y = x$ line for most plots, specifically regenerator CO_2 concentration (Fig. 6e, 5e). This is due to the Aspen model calculation, which is able to calculate the most stable species in a gas stream at each temperature, which in this case at the given temperature range, is CO_2. This is translated into the Python models, as the set of five features used from Aspen, allows the model to obtain a more accurate prediction.

Whilst the majority of the graphs are fairly uniform in the points plotted around the line of best fit and $y = x$ line, there are some apparent anomalistic datapoints seen in both Fig. 5f and Fig. 6c. For Fig. 5f, neural network CH_4 conversion, there is a clear curve of data points above, and a few points below the line of best fit and $y = x$ line. To eliminate this being a mistake, this plot was run in Python multiple times, with different epochs and number of nodes in the layers, with the same pattern being detected. An explanation for this is down to the raw data and potentially, that an insufficient set of features was selected to identify all reactions occurring. When analysing the raw data, it is evident that whilst 0.99% was the maximum for the conversion, at the lower end, 0.02% was the minimum however this was for a specific set of inputs, i.e. the results go from 0.99 to 0.02% in a set pattern, with little to no results at 0–0.4% for higher input feature conditions, and the same at 0.6-0.99% for lower input feature data, creating a curve in the overall raw dataset.

For the Fig. 6c, random forest H_2 concentration in the reformer, an explanation for the plateau of datapoints at around 0.5% and again between 0.7–0.8%, is that since H_2 is the final and main product gas from all the reactions taking place in the reformer, it is the essentially least affected by slight changes in input features. This is because it is the agglomeration of 3 different reactions equations 1, 2 and (4), thus a lot of the different combinations of inputs (S/C, CaO/C, temperature, pressure), result in the same H_2 concentrations, which can be seen when analysing the raw data, and can also be seen practically. The data in this study is theoretical and manually adjusted in the Aspen model however, in literature concerning experimental studies on SE-SMR [49–51], there are a number of different conditions that are accepted as 'optimal', and therefore the reason behind different conditions resulting in a high H_2 concentration. This pattern in data can vaguely be seen in the neural network plot as well (Fig. 5c), particularly around 0.7%, but not as clearly. Reasons for this fall under the fact that the ANN is a connected network of neurons, that are grouped in layers and process data in each layer sequentially before passing forward onto next layers, therefore, each message is passed on until the very last end results, and the output is a more holistic representation of the input features. On the other hand, random forests are made up of decision trees, where within each tree, the input is processed and an output is predicted, independent of all the other trees [52]. Additionally, an assumption that RF takes on is that it relies on sampling being representative. For example, if one class consists of two components and in the dataset one component is represented by 100 samples, and another component is represented by one sample - most individual decision trees will see only the first component and RF will misclassify the second one. RF on the whole does a good job at classification but not as much for regression in some cases, as it does not give precise continuous nature prediction. In the case of regression, it does not predict beyond the range in the training data and can sometimes lead to nonsensical predictions if applied to extrapolation domains, as seen in Fig. 6c [53].

Advantages of a data driven soft-sensor model, as opposed to a mathematical modelling program such as Aspen Plus, is that one is able to obtain real time variable measurements dependent on actual process conditions instead of manually inputting each component conditions in order to obtain a parameter output. Despite how thermodynamically accurate an Aspen Plus model is over a soft sensor, in practical terms the use of a representative model is more beneficial. The use of an ANN or RF model acts as the code within a control system which in turn predicts the concentration of the gases, thus bypassing the use of hardware components, and with the metrics shown in Figs. 5 and 6, the predictions are highly accurate. Mathematical programming software such as Aspen Plus, have the feature of performing reactions at equilibrium, whereas actual experimental processes experience various limitations such as diffusional limitations and axial dispersion. For this reason, the use of a soft sensor is highly advantageous over a simulation program, due to the use of actual historical process data. In addition, in practical terms, a soft-sensor can be used as a backup sensor, when the hardware sensor is faulty or is removed for maintenance or replacement, or it can be used in the process of quality control, in place of laboratory analysis, which is often not exhaustive enough [6,54].

5. Conclusion

In this work, a soft sensor model was developed through the application of artificial neural network (ANN) and random forest (RF) algorithms for the hydrogen production process by sorption enhanced steam methane reforming (SE-SMR). The choice and development of the models were discussed as well as parameters that were necessary to produce accurate predictions of the dataset provided. From the results, both models were shown to provide good performance predictions for gas concentrations in the reformer and regenerator reactors using five specified process features, in particular those of high concentrations. Additionally, the CH_4 conversion within the reformer was calculated and predicted with both models able to reflect the efficiency of the reaction very accurately. The random forest model indicated higher levels of accuracy with consistently higher R^2 values and lower mean absolute error (0.002–0.014) compared to the neural network model (0.005–0.024). Overall, this study concludes that ANN and RF algorithms can successfully model a nonlinear process such as SE-SMR. The developed models can be applied as a soft sensor for analysis and evaluation of a SE-SMR based hydrogen production plant and to predict key process performance indicators. The chemical properties of hydrogen and the sorbent material hinder the ability to rapidly upscale this process, therefore the use of machine learning models, that are a realistic representation of experimental data, is very useful, in order to overcome such limitations. With hydrogen fast becoming a leader in clean energy systems, it is essential that all efforts are made to simplify and ensure the safety and feasibility in the process of up-scaling hydrogen production.

The use of a data-driven soft sensor does exactly that and allows for further exploration and incorporation of artificial intelligence into this energy field. Future directions for this study would be to implement the soft sensor experimentally in order to compare the success rate of the soft sensor against hardware sensors, and thus further validating this machine learning application for the process of SE-SMR.

Data Access Statement

Data and codes underlying this work are available via the Cranfield Online Research Data repository: https://doi.org/10.17862/cranfield.rd.13256573 and https://doi.org/10.17862/cranfield.rd.13256633.

Declaration of Competing Interest

None.

Acknowledgments

PN, PTC and VM acknowledge the financial support from the UK Engineering and Physical Sciences Research Council Doctoral Training Partnership (EPSRC DTP) grant No. EP/R513027/1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.egyai.2020.100037.

References

[1] Edwards PP, Kuznetsov VL, David WIF, Brandon NP. Hydrogen and fuel cells: towards a sustainable energy future. Energy Policy 2008;36:4356–62. doi:10.1016/j.enpol.2008.09.036.
[2] Yan W, Tang D, Lin Y. A data-driven soft sensor modeling method based on deep learning and its application. IEEE Trans. Ind. Electron. 2017;64:4237–45. doi:10.1109/TIE.2016.2622668.
[3] Fortuna L, Graziani S, Rizzo A, Xibilia MG. Soft Sensors for Monitoring and Control of Industrial Processes. London: Springer London; 2007. doi:101007/978-1-84628-480-9.
[4] Wang K, Shang C, Liu L, Jiang Y, Huang D, Yang F. Dynamic soft sensor development based on convolutional neural networks. Ind. Eng. Chem. Res. 2019;58:11521–31. doi:10.1021/acs.iecr.9b02513.
[5] Shang C, Yang F, Huang D, Lyu W. Data-driven soft sensor development based on deep learning technique. J. Process Control 2014;24:223–33. doi:10.1016/j.jprocont.2014.01.012.
[6] Torgashov A, Zmeu K. Nonparametric soft sensor evaluation for industrial distillation plant, in: Comput. Aided Chem. Eng., Elsevier 2015:1487–92. doi:10.1016/B978-0-444-63577-8.50093-0.
[7] Yuan X, Huang B, Wang Y, Yang C, Gui W. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. IEEE Trans. Ind. Informatics. 2018;14:3235–43. doi:10.1109/TII.2018.2809730.
[8] Yao L, Ge Z. Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application. IEEE Trans. Ind. Electron. 2018;65:1490–8. doi:10.1109/TIE.2017.2733448.
[9] Yan W, Shao H, Wang X. Soft sensing modeling based on support vector machine and Bayesian model selection. Comput. Chem. Eng. 2004;28:1489–98. doi:10.1016/j.compchemeng.2003.11.004.
[10] Sipöcz N, Tobiesen FA, Assadi M. The use of artificial neural network models for CO_2 capture plants. Appl. Energy. 2011;88:2368–76. doi:10.1016/j.apenergy.2011.01.013.
[11] Zamaniyan A, Joda F, Behroozsarand A, Ebrahimi H. Application of artificial neural networks (ANN) for modeling of industrial hydrogen plant. Int. J. Hydrog. Energy. 2013;38:6289–97. doi:10.1016/j.ijhydene.2013.02.136.
[12] Lin B, Recke B, Knudsen JKH, Jørgensen SB. A systematic approach for soft sensor development. Comput. Chem. Eng. 2007;31:419–25. doi:10.1016/j.compchemeng.2006.05.030.
[13] Barelli L, Bidini G, Gallorini F, Servili S. Hydrogen production through sorption-enhanced steam methane reforming and membrane technology: a review. Energy 2008;33:554–70. doi:10.1016/j.energy.2007.10.018.
[14] Abbas SZ, Dupont V, Mahmud T. Modelling of H_2 production in a packed bed reactor via sorption enhanced steam methane reforming process. Int. J. Hydrog. Energy. 2017;42:18910–21. doi:10.1016/j.ijhydene.2017.05.222.
[15] Antzara A, Heracleous E, Bukur DB, Lemonidou AA. Thermodynamic analysis of hydrogen production via chemical looping steam methane reforming coupled with in Situ CO_2 capture. Energy Procedia 2014;63:6576–89. doi:10.1016/j.egypro.2014.11.694.
[16] García-Lario AL, Aznar M, Grasa GS, García T, Murillo R. Study of nickel catalysts for hydrogen production in sorption enhanced reforming process. J. Power Sources 2013;242:371–9. doi:10.1016/j.jpowsour.2013.05.069.
[17] Erans M, Jeremias M, Manovic V, Anthony EJ. Operation of a 25 kW_{th} calcium looping pilot-plant with high oxygen concentrations in the calciner. J. Vis. Exp. 2017(2017):1–10. doi:10.3791/56112.
[18] Hanak DP, Manovic V. Calcium looping combustion for high-efficiency low-emission power generation. J. Clean. Prod. 2017;161:245–55. doi:10.1016/j.jclepro.2017.05.080.
[19] Li Z, Cai N, Yang J. Continuous Production of hydrogen from sorption-enhanced steam methane reforming in two parallel fixed-bed reactors operated in a cyclic manner. Ind. Eng. Chem. Res. 2006;45:8788–93. doi:10.1021/ie061010x.
[20] Johnsen K, Ryu HJ, Grace JR, Lim CJ. Sorption-enhanced steam reforming of methane in a fluidized bed reactor with dolomite as CO_2-acceptor. Chem. Eng. Sci. 2006;61:1195–202. doi:10.1016/j.ces.2005.08.022.
[21] Arstad B, Prostak J, Blom R. Continuous hydrogen production by sorption enhanced steam methane reforming (SE-SMR) in a circulating fluidized bed reactor: sorbent to catalyst ratio dependencies. Chem. Eng. J. 2012;189–190:413–21. doi:10.1016/j.cej.2012.02.057.
[22] Antzara A, Heracleous E, Bukur DB, Lemonidou AA. Thermodynamic analysis of hydrogen production via chemical looping steam methane reforming coupled with in situ CO_2 capture. Int. J. Greenh. Gas Control. 2015;32:115–28. doi:10.1016/j.ijggc.2014.11.010.
[23] Breiman L. Random Forests. Mach. Learn. 2001;45:5–32. doi:10.1023/A:1010933404324.
[24] James G, Witten D, Hastie T, Tibshirani R. Tree-based methods. An Introd. to Stat. Learn. 2013:303–35. doi:10.1007/978-1-4614-7138-7_8.
[25] Hutengs C, Vohland M. Downscaling land surface temperatures at regional scales with random forest regression. Remote Sens. Environ. 2016;178:127–41. doi:10.1016/j.rse.2016.03.006.
[26] Polishchuk PG, Muratov EN, Artemenko AG, Kolumbin OG, Muratov NN, Kuz'min VE. Application of Random Forest Approach to QSAR prediction of aquatic toxicity. J. Chem. Inf. Model. 2009;49:2481–8. doi:10.1021/ci900203n.
[27] Yan Y, Mattisson T, Moldenhauer P, Anthony EJ, Clough PT. Applying machine learning algorithms in estimating the performance of heterogeneous, multi-component materials as oxygen carriers for chemical-looping processes. Chem. Eng. J. 2020;387:124072. doi:10.1016/j.cej.2020.124072.
[28] Ayodele BV, Cheng CK. Modelling and optimization of syngas production from methane dry reforming over ceria-supported cobalt catalyst using artificial neural networks and Box–Behnken design. J. Ind. Eng. Chem. 2015;32:246–58. doi:10.1016/j.jiec.2015.08.021.
[29] Vo ND, Oh DH, Hong SH, Oh M, Lee CH. Combined approach using mathematical modelling and artificial neural network for chemical industries: Steam methane reformer. Appl. Energy. 2019;255:113809. doi:10.1016/j.apenergy.2019.113809.
[30] Lang Y, Malacina A, Biegler LT, Munteanu S, Madsen JI, Zitney SE. Reduced Order Model Based on Principal Component Analysis for Process Simulation and Optimization. Energy & Fuels 2009;23:1695–706. doi:10.1021/ef800984v.
[31] Ye F, Ma S, Tong L, Xiao J, Bénard P, Chahine R. Artificial neural network based optimization for hydrogen purification performance of pressure swing adsorption. Int. J. Hydrog. Energy. 2019;44:5334–44. doi:10.1016/j.ijhydene.2018.08.104.
[32] Xiao J, Li C, Fang L, Böwer P, Wark M, Bénard P, Chahine R. Machine learning–based optimization for hydrogen purification performance of layered bed pressure swing adsorption. Int. J. Energy Res. 2020;44:4475–92. doi:10.1002/er.5225.
[33] Gardner M, Dorling S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 1998;32:2627–36. doi:10.1016/S1352-2310(97)00447-0.
[34] Matsuo R, Yamazaki T, Suzuki M, Toyama H, Araki K. A random forest algorithm-based approach to capture latent decision variables and their cutoff values. J. Biomed. Inform 2020;110:103548. doi:10.1016/j.jbi.2020.103548.
[35] Şener AN, Günay ME, Leba A, Yıldırım R. Statistical review of dry reforming of methane literature using decision tree and artificial neural network analysis. Catal. Today. 2018;299:289–302. doi:10.1016/j.cattod.2017.05.012.
[36] Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Modell. 2003;160:249–64. doi:10.1016/S0304-3800(02)00257-0.
[37] Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Modell. 2004;178:389–97. doi:10.1016/j.ecolmodel.2004.03.013.
[38] Zhang Z, Beck MW, Winkler DA, Huang B, Sibanda W, Goyal H. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann. Transl. Med. 2018;6:216. doi:10.21037/atm.2018.05.32.
[39] Elmolla ES, Chaudhuri M, Eltoukhy MM. The use of artificial neural network (ANN) for modeling of COD removal from antibiotic aqueous solution by the Fenton process, J. Hazard. Mater. 2010;179:127–34. doi:10.1016/j.jhazmat.2010.02.068.
[40] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 2016;374:20150202. doi:10.1098/rsta.2015.0202.
[41] Ringnér M. What is principal component analysis? Nat. Biotechnol. 2008;26:303–4. doi:10.1038/nbt0308-303.
[42] Song F, Guo Z, Mei D. Feature selection using principal component analysis. In: 2010 Int. Conf. Syst. Sci. Eng. Des. Manuf. Informatiz., IEEE; 2010. p. 27–30. doi:10.1109/ICSEM.2010.14.
[43] Luo XJ, Oyedele LO, Ajayi AO, Akinade OO, Delgado JMD, Owolabi HA, Ahmed A. Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings. Energy AI 2020;2:100015. doi:10.1016/j.egyai.2020.100015.

[44] Brownlee J. What is the Difference between a batch and an epoch in a neural network? Mach. Learn. Mastery. 2018:3–4.
[45] Feng J, Lu S. Performance Analysis of various activation functions in artificial neural networks, J. Phys. Conf. Ser. 2019;1237:022030. doi:10.1088/1742-6596/1237/2/022030.
[46] Gao X, Wen J, Zhang C. An improved random forest algorithm for predicting employee turnover. Math. Probl. Eng. 2019(2019):1–12. doi:10.1155/2019/4140707.
[47] Scornet E. Tuning parameters in random forests. ESAIM Proc. Surv. 2017;60:144–62. doi:10.1051/proc/201760144.
[48] Gotelli NJ. Research frontiers in null model analysis. Glob. Ecol. Biogeogr. 2001;10:337–43. doi:10.1046/j.1466-822X.2001.00249.x.
[49] Ding Y, Alpay E. Adsorption-enhanced steam-methane reforming. Chem. Eng. Sci. 2000;55:3929–40. doi:10.1016/S0009-2509(99)00597-7.
[50] Di Giuliano A, Gallucci K. Sorption enhanced steam methane reforming based on nickel and calcium looping: a review. Chem. Eng. Process. - Process Intensif. 2018;130:240–52. doi:10.1016/j.cep.2018.06.021.
[51] Anderson DM, Kottke PA, Fedorov AG. Thermodynamic analysis of hydrogen production via sorption-enhanced steam methane reforming in a new class of variable volume batch-membrane reactor. Int. J. Hydrogen Energy. 2014;39:17985–97. doi:10.1016/j.ijhydene.2014.03.127.
[52] Liaw A, Wiener M. Classification and regression by random forest. R News 2002;2:18–22.
[53] Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 2018;6:e5518. doi:10.7717/peerj.5518.
[54] Slišković D, Grbić R, Hocenski Ž. Methods for plant data-based process modeling in soft-sensor development. Automatika 2011;52:306–18. doi:10.1080/00051144.2011.11828430.