

Article

# Using Self-Organizing Maps to Elucidate Patterns among Variables in Simulated Syngas Combustion

Dhan Lord B. Fortela <sup>1,2,\*</sup>, Matthew Crawford <sup>1</sup>, Alyssa DeLatre <sup>1</sup>, Spencer Kowalski <sup>1</sup>, Mary Lissard <sup>1</sup>, Ashton Fremin <sup>1</sup>, Wayne Sharp <sup>2,3</sup>, Emmanuel Revellame <sup>2,4</sup>, Rafael Hernandez <sup>1,2</sup> and Mark Zappi <sup>1,2</sup>

<sup>1</sup> Department of Chemical Engineering, University of Louisiana, Lafayette, LA 70504, USA; matthew.crawford1@louisiana.edu (M.C.); alyssa.delatre1@louisiana.edu (A.D.); spencer.kowalski1@louisiana.edu (S.K.); mary.lissard1@louisiana.edu (M.L.); ashton.fremin1@louisiana.edu (A.F.); rafael.hernandez@louisiana.edu (R.H.); mark.zappi@louisiana.edu (M.Z.)

<sup>2</sup> Energy Institute of Louisiana, University of Louisiana, Lafayette, LA 70504, USA; wayne.sharp@louisiana.edu (W.S.); emanuel.revellame@louisiana.edu (E.R.)

<sup>3</sup> Department of Civil Engineering, University of Louisiana, Lafayette, LA 70504, USA

<sup>4</sup> Department of Industrial Technology, University of Louisiana, Lafayette, LA 70504, USA

\* Correspondence: dhanlord.fortela@louisiana.edu; Tel.: +1-337-482-5765

Received: 25 February 2020; Accepted: 17 April 2020; Published: 28 April 2020



**Abstract:** This study focused on demonstrating the use of a self-organizing map (SOM) algorithm to elucidate patterns among variables in simulated syngas combustion. The work was implemented in two stages: (1) modelling and simulation of syngas combustion under various feed composition and reactor temperature implemented in AspenPlus<sup>TM</sup> chemical process simulation software, and (2) pattern recognition among variables using SOM algorithm implemented in MATLAB. The varied levels of feed syngas composition and reactor temperature was randomly sampled from uniform distributions using the Morris screening technique creating four thousand eight hundred simulation conditions implemented in the process simulation which consequently produced a multivariate dataset used in the SOM analysis. Results show that cylindrical SOM topology models the dataset at lower quantization error and topographic error as compared to the rectangular SOM topology indicating suitability of the former for variables pattern elucidation for the simulated combustion. Nonetheless, the variables pattern between component planes from rectangular SOM ( $9 \times 28$  grid) and those from cylindrical SOM ( $9 \times 28$  grid) are almost similar, indicating that either rectangular or cylindrical architectures may be used for variables pattern analysis. The component planes of process variables from trained SOM are a convenient visualization of the trends across all process variables.

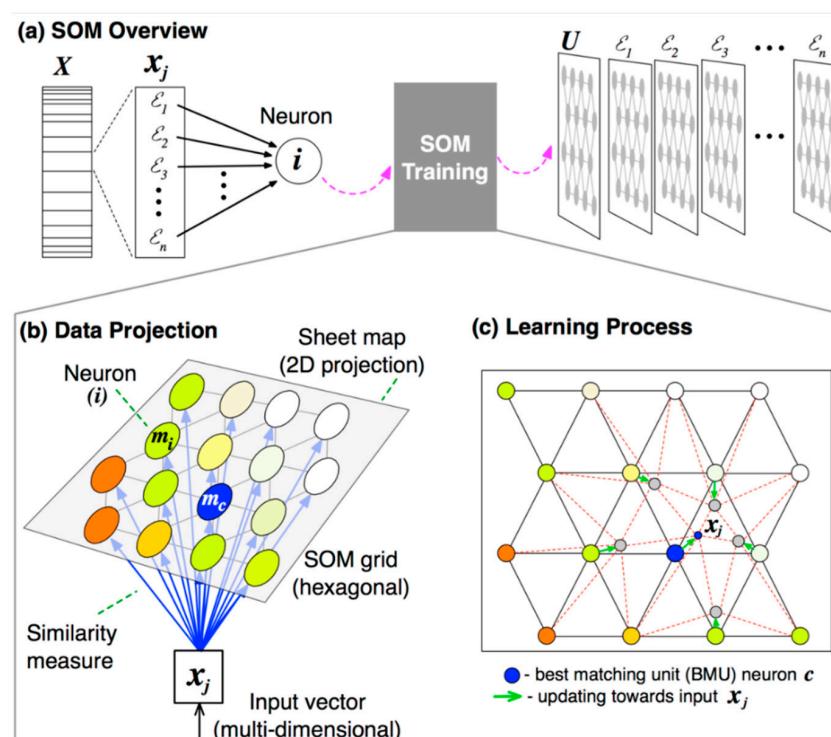
**Keywords:** chemical process simulation; syngas; machine learning; SOM

## 1. Introduction

Like many traditional and emerging technologies on energy and materials use, syngas production and combustion has been among those actively evaluated for cleaner implementation in pilot scale and industrial scale [1]. Syngas (short for ‘synthesis gas’) is typically produced through gasification of many resources such as natural gas, coal and biomass by reaction with steam or oxygen [2]. Syngas have been studied for their chemistry and physics resulting in some established literature on reaction mechanisms, reaction kinetic parameters, etc., of which some, in the form of simulations of system performance, have been used to give insight into the possible design and operation considerations [3,4].

Chemical process simulators (such as AspenPlus, CHEMCAD, and COMSOL Multiphysics) have been instrumental not just in training but also in the practice of chemical process design and operation [5]. Most of these simulators have extensive property databanks and they allow rigorous material and energy balance calculations for modeling complex process plants [6]. These features make process simulators key tools in performing system-oriented analysis, which is the root of the modern process design paradigm [7]. This work aims to demonstrate that systems analysis with process simulators may be enhanced through the use of machine learning algorithms that extract patterns within multivariate datasets of simulations. Specifically, this work implements self-organizing map (SOM) techniques to elucidate patterns among variables in a simulated syngas combustion modeled in AspenPlus™.

SOMs, which are also called Kohonen networks due to the original proponent [8–10], are derivatives of artificial neural network (ANN) algorithms with applications in data visualization, feature extraction, clustering, and supervised classification [11,12]. These applications leverage on SOMs' efficiency in discovering patterns in high-dimensional multivariate data through projections onto low dimensions such as 2D hence their designation as 'maps' [12,13]. In essence, an SOM defines an 'elastic net' of neuron nodes that are fitted to the input data space to approximate its density function in a topologically ordered mapping [8]. The process by which such mappings are formed is the SOM algorithm schematically depicted in Figure 1. So far, SOM is one of two ANNs executing competitive learning, a form of unsupervised learning which allows for specialization of each neuron in the network that ultimately becomes a feature-discovering tool for high-dimensional data [14]. The unsupervised learning nature of SOM also means that it does not require model calibration towards a target response such as those in regression.

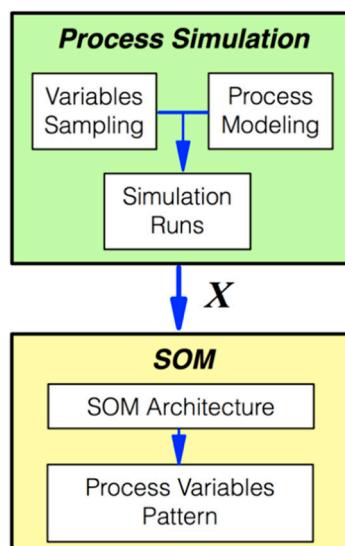


**Figure 1.** Schematic of basic sequential learning SOM applied to a dataset  $X$ : (a) two-dimensional SOM modelling a  $n$ -dimensional ( $\xi_1, \dots, \xi_n$ ) input data vector  $x_j$  onto a hexagonal lattice sheet map of neurons with associated weight vectors such as  $m_i$  in neuron  $i$  including the best-matching unit (BMU)  $c$  with associated weight vector  $m_c$ ; (b)  $x_j$  is projected to all the weight-initialized neurons in the grid to determine the BMU, in this case neuron  $c$ , to initiate the learning process; and (c) the weight vectors  $m_i$  of the BMU and the neighboring neurons are updated to move the active neurons in the BMU neighborhood closer to  $x_j$  in a recursive manner. SOM trains until neuron weight vectors asymptotically converge on all samples in  $X$ .

The following are some of the areas of application for SOM: minmax objective for multirobot-multigoal path planning [15]; graph-mining for economic measures [16]; social interactions [17]; fault diagnosis [8,18]; various clustering tasks [8,19,20]; various classification tasks [8,21]; speech recognition [8]; image simplification of atmospheric data [22] for cloud detection and land cover mapping [23]. The potential of SOM as a tool in chemical process systems has been claimed by Simula and Kangas [24] around two decades ago, and they based their idea on the topology-preserving property of SOM. Liukkonen, Hiltunen [25] used SOM to diagnose the various states of a utility-scale fluidized bed combustion process. Muñoz, Martín-Torre [26] used SOM to extract patterns and cluster dataset on the element release from sediments in contact with acidified seawater conducted via laboratory leaching tests. Ganhadeiro, Christo [27] used SOM to analyze electric energy distribution systems. Li, Du [28] used SOM as a process monitoring computational tool on the Tennessee Eastman chemical process. However, a survey of the current literature indicates that, presently, there is no work on the use of SOM as a data analytics tool in chemical process simulators. Hence, this work aims to demonstrate the potential SOM to elucidate patterns in multivariate datasets of chemical process simulators to aide in the study and clean implementation of alternative technologies such syngas.

## 2. Methodology

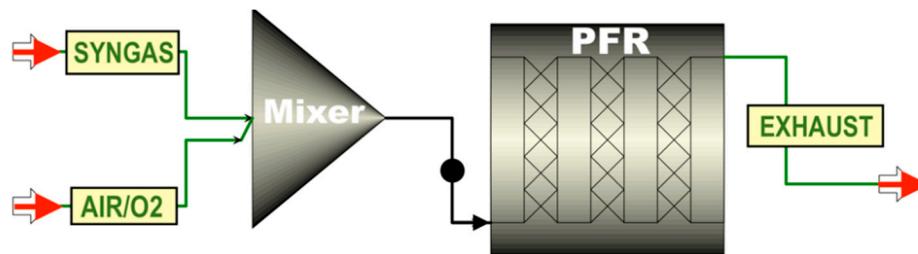
The work was implemented in two stages (Figure 2): (1) modelling and simulation of syngas combustion under various feed composition and reactor temperature implemented in AspenPlus<sup>TM</sup> chemical process simulation software, and (2) pattern recognition among variables using SOM algorithm implemented in MATLAB. For definitions of symbols, refer to the Nomenclature.



**Figure 2.** Schematic overview of using SOM as variables pattern recognition in simulated chemical processes.  $X$  is the set of all simulation variables values ( $\xi_n \ n \in [1 : N]$ ) collected from each simulation run ( $x_j, j \in [1 : J]$ ) and used as input dataset into SOM.

### 2.1. Process Kinetics Simulation

The AspenPlus chemical process kinetics simulation model (Figure 3) consisted of a mixing unit and a combustion reactor modeled as a plug-flow reactor (PFR) under adiabatic condition, which is a common model for syngas combustion [3]. The PFR was sized according to previous works [29] with 0.381 m internal diameter and 5.90 m total length. The thermodynamic properties were calculated through the Soave-Redlich-Kwong (SRK) equation of state as implemented by AspenPlus V10.



**Figure 3.** Process flow diagram of the simulated syngas combustion with AspenPlus using a plug-flow reactor (PFR) under adiabatic condition.

The reaction kinetics were modeled in power-law form and the reaction constant,  $k_b$ , was modeled in Arrhenius law form to account for the effect of reactor temperature on the reaction constant. These reaction mechanism and parameters used in the PFR (Table 1) were based on the works of Khan, DeJong [30] and J. De Kam, V. Morey [3]. The feed gas component mole fraction  $y_{0,a}$  of the syngas, and the reactor temperature, hence the exhaust temperature, were varied according to the Morris sampling technique [31,32]. The Morris sampling technique, which is typically used in global sensitivity analysis, randomly samples  $[r(k+1)]$  combinations of the  $k$  variables with corresponding sets of lower bound and upper bound levels (see Table 2), where  $r$  is the number of increments within the sampling bounds for each variable. In this work, the variables are the six feed gas components and the reactor temperature, which makes  $k = 7$ . Based on previous works implementing the Morris sampling technique, a conservative level of  $r$  was set  $r = 200$ . To ensure the combinations of variable levels randomly covered a wide region of the design space, the Morris sampling was performed using 3 random seed generators. Hence, the total number of simulation runs was  $3 \times [200(7+1)] = 4800$ . The Morris random sampling was implemented in R-statistical software using the 'sensitivity' R-package. See the Supplementary Information for the summary of 4800 simulation conditions, and the accompanying R-script used in implementing Morris sampling on the variables.

**Table 1.** Syngas combustion key reaction steps, rate law expression, and reaction constant parameter as function of temperature T implemented in the PFR.

	Reaction	Rate Expression	Reaction Constant Parameter
1	$CH_4 + 1.5O_2 \rightarrow CO + 2H_2O$	$r_{1,CO} = k_1 C_{CH_4} C_{O_2}^{0.8}$	$k_1 = 1.547 \times 10^8 \exp(-15098/T)$
2	$CO + 0.5O_2 \rightarrow CO_2$	$r_{2,CO_2} = k_2 C_{CO} C_{O_2}^{0.5} C_{H_2O}^{0.5}$	$k_2 = 1.585 \times 10^{10} \exp(-24157/T)$
3	$H_2 + 0.5O_2 \rightarrow H_2O$	$r_{3,H_2O} = k_3 C_{H_2} C_{O_2}^{0.1}$	$k_3 = 1.631 \times 10^9 \exp(-3420/T)$

**Table 2.** Variables sampling bounds (minimum and maximum), median, and average values through the Morris sampling technique.

Variables:	$y_{0,H_2}(\%)$	$y_{0,CO}(\%)$	$y_{0,CH_4}(\%)$	$y_{0,CO_2}(\%)$	$y_{0,N_2}(\%)$	$y_{0,H_2O}(\%)$	$T_{react}(K)$
Minimum:	4.65	12.03	0.00	0.84	0.11	0.05	1000
Maximum:	59.30	67.59	15.10	38.95	54.26	43.43	1500
Median:	25.55	28.02	2.93	11.24	18.10	14.62	1263
Average:	24.91	28.66	2.98	11.33	17.75	14.38	1256

Setting the PFR at a constant temperature allowed for the determination of heat duty supplied by (–) or to (+) the reactor. Table 2 summarizes the sampling bounds (minimum and maximum), median, and average values of the variables. The range (lower-bound and upper-bound) of values used for Morris sampling on the 7 variables in Table 2 were based on the work of McDonell [33] on syngas from coal, and of Martínez, Mahkamov [4] on syngas from biomass. The feed syngas enters the PFR at 486 K and 1 atm at a steady rate total molar flow rate of 5.58 kmol/hr based on previous works [29]. Pure

oxygen was assumed to be mixed with the syngas going to the PFR and was supplied at 5% excess, based on Martínez, Mahkamov [4], calculated for each reactant composition.

## 2.2. Variables Pattern Recognition via SOM

The SOM computational works in the study were implemented through MATLAB (by MathWorks®) with program codes built on functions from the public-domain add-in SOM Toolbox version 2.0 (available for download at <http://www.cis.hut.fi/somtoolbox/>) developed and maintained by the Laboratory of Computer and Information Science at Helsinki University of Technology in Finland, where Kohonen [8] and associates conducted many of the works on SOM. This toolbox has been designed to accompany the SOM literature material by Kohonen [8], and it has been used and tested in several research works. In addition to the script-embedded documentation on the use of the SOM Toolbox, literature explaining some of the main features of code-functions and usage with the Toolbox has been published previously [34,35]. Following are the key components of the SOM implementation in the software. The program scripts used in the study are summarized in the Supplementary Material.

### 2.2.1. Data Pre- and Post-Processing

The simulation variables used for SOM training are the following (16 variables): heat duty ( $Q$ ); mole fractions of syngas components ( $y_{0,a}$ 's): H<sub>2</sub>, CO, CH<sub>4</sub>, CO<sub>2</sub>, N<sub>2</sub>, H<sub>2</sub>O; feed O<sub>2</sub> (at 5% excess); reactor temperature ( $T_{reactor}$ ), total molar flow of outlet stream; mole fractions of outlet stream components ( $y_a$ 's): CO, CH<sub>4</sub>, CO<sub>2</sub>, N<sub>2</sub>, H<sub>2</sub>O, and O<sub>2</sub>. These variables are typical analysis variables in chemical reactors, specifically in syngas combustion. The datasets from the AspenPlus simulation runs were summarized in spreadsheets. Then in MATLAB, the variables were scaled through normalization to treat the variables that were equally important, i.e., eliminate dominance of high-value variables in the mapping [35]. For post-processing after SOM training, the values (as weight vectors at the end of SOM training) were de-normalized to original scale to aid the meaningful interpretation of results [34].

### 2.2.2. Lattice Structure, Map Shape, and Size

The lattice structure may be rectangular or hexagonal. Hexagonal (see Figure 1) was preferred because it does not favor horizontal and vertical directions as much as the rectangular array [8]. The global map shape may be a 2D sheet, a cylinder, or a toroid [34]. Though the projection 2D sheet map has been found to be sufficiently effective in capturing patterns from the original high-dimensional dataset [10], this study evaluated both the 2D sheet and cylinder projections. The sheet map must be elongated, i.e., length > width, to attain a stable orientation as the learning process proceeds [8]. The total number of neurons should be at most  $5\sqrt{n \times S}$ , where  $n$  is the number of variables used for SOM training (i.e.,  $\xi_1, \dots, \xi_n$ ) and  $S$  is the number of samples [35,36] or simulation runs in this work (4800 simulation runs).

### 2.2.3. SOM Initialization and Training

The SOM Toolbox uses Euclidean metric for vectorial distance calculations:  $c = \operatorname{argmin}_i \{x_j - m_i\}$ . The  $c$ -index neuron is the BMU with weight vector  $m_c$ , which will be the first neuron to 'learn'. This step implements the competitive learning nature of SOM. Initialization of weight vectors was linear and training was batch learning. Initialization and training were done in two stages: (1) rough—large radius of neighborhood kernel and fast learning rate, and (2) fine-tuning—small radius of neighborhood kernel and slow learning rate. These were the default parameters in the Toolbox, and the default models were Gaussian model for the neighborhood function  $h_{ci}(t)$ , and reciprocally decreasing learning rate  $\alpha(t)$  (see Supplementary Material for model forms) [34,35].

As an unsupervised learning algorithm, a specific SOM is not amenable to evaluation of the 'best fit' of a particular SOM architecture to the dataset. However, two parameters are typically used to measure the performance of SOM mapping on data—(1) quantization error (Qe), and (2) topographic

error ( $Te$ ) [35].  $Qe$  is the average distance between each data vector and its BMU and it measures map resolution.  $Te$  is the proportion of all data vectors for which first and second BMUs are not adjacent units and it measures topology preservation. Pöhlbauer [37] found that these two measures of quality of SOM were found to be inversely proportional when map sizes are large. To establish a measure of map quality, each  $Qe$  and  $Te$  were normalized within their respective range of values and the normalized  $Qe$  and  $Te$  ( $nQe$  and  $nTe$ ) were added, i.e.,  $\text{sum}(nQe, nTe)$ . Hence, the map sizes that minimize  $\text{sum}(nQe, nTe)$  would indicate a good compromise of the quantization and topographic errors of an SOM model.

### 3. Results

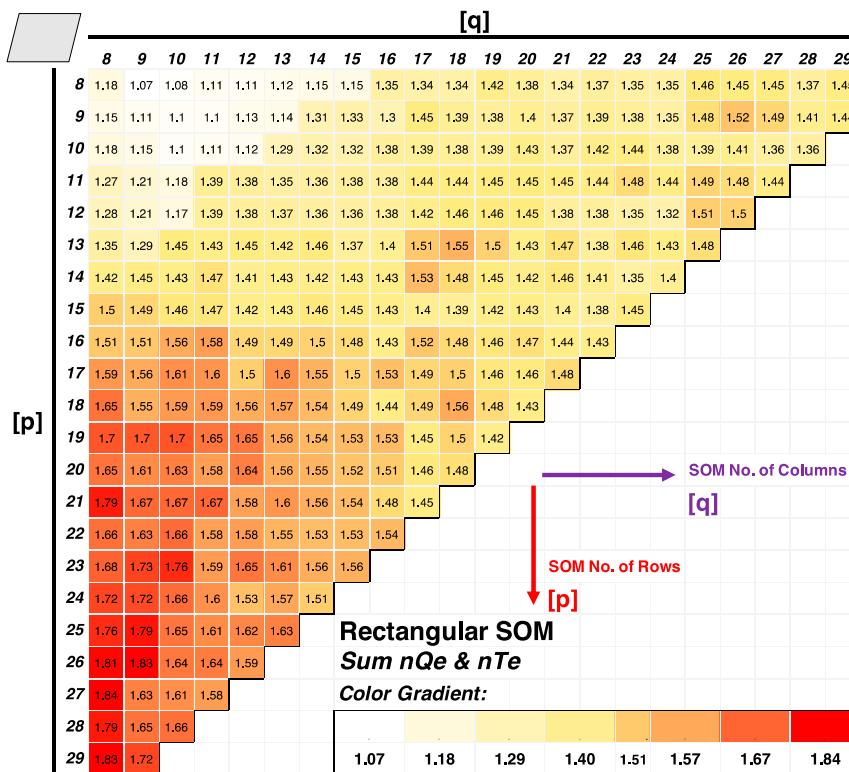
The collected data from process simulations are summarized as Supplementary Information (see the Supplementary Information). Of interest for discussion are the outputs of SOM trained on the process simulation dataset.

#### 3.1. SOM Architecture

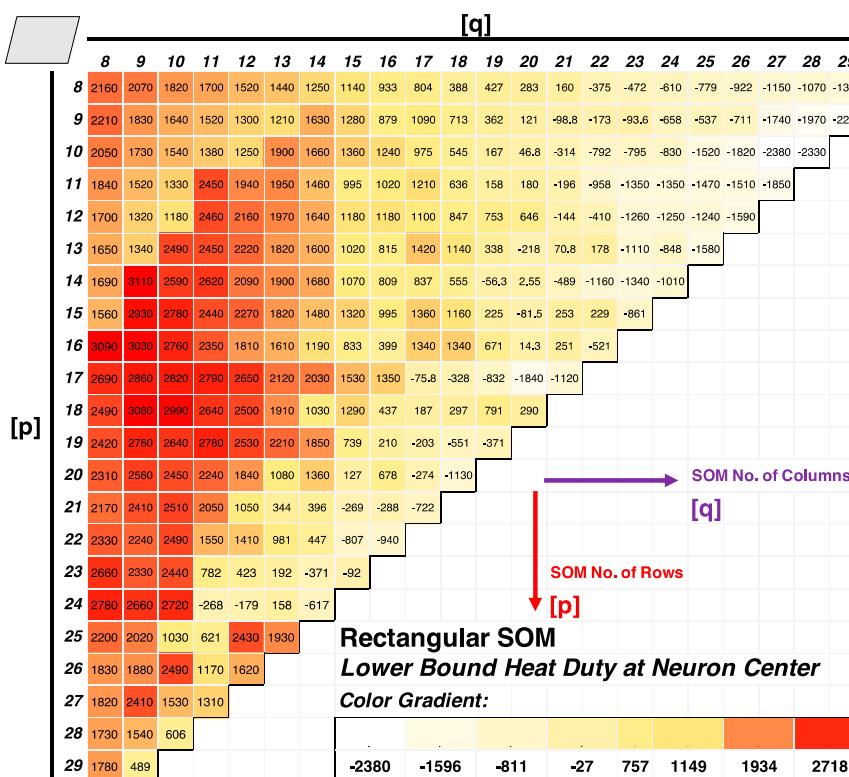
Many possible combinations of size and shape may be specified to create an SOM architecture to map a dataset. When the formula for the number of neurons (size)  $5\sqrt{n \times S}$  is used with  $n = 16$  and  $S = 4800$ , the size of the SOM should be 1386 neurons. With this SOM size, the resulting sets of U-matrix produced random patterns instead of the hypothesized organized patterns on the maps (see Supplementary Material showing  $47 \times 28$  rectangular SOM grid). Based on the findings of Kohonen [10], it is not possible to determine the exact size of the SOM beforehand, and that it must be determined by trial-and-error. Hence, using the U-matrix as indicator for patterns in the SOM, a trial-and-error analysis of the number of neurons was done and it was found that organized patterns emerge if the  $n$  is reduced down to the number of variables originally varied through Morris sampling, i.e.,  $n = 7$ , and if the formula for size is changed to  $2\sqrt{n \times S}$ , which is consistent with the proposed mathematical form to estimate SOM size [38]. Consequently, a working maximum neuron size of  $2\sqrt{7 \times 4800} = 367$  was used (a graphical comparison of U-matrix between a  $9 \times 28$  rectangular grid and a  $47 \times 28$  rectangular grid is shown in the electronic Supplementary Material).

The sum of the normalized  $Qe$  and normalized  $Te$  is summarized in Figure 4 for the rectangular SOM (Figures 4 and 5) and in Figure 6 for the cylindrical SOM (Figures 6 and 7). The sum of errors (Figures 4 and 6) have been compared with one of the response variables—heat duty—to examine the trends of mapping performance with varying SOM size and shape. Frequency distribution plots of the 16 variables are summarized in the supporting information (see Supplementary Material). The range of Heat Duty values in this study is from negative to positive (minimum: -17,497 cal/sec to maximum: 42,498 cal/sec; median: 14,690 cal/sec; mean: 15,138 cal/sec). The lowest nominal value of the head duty at a neuron center was noted for each SOM architecture used for mapping. The closer the heat duty nominal value at a neuron center to the minimum value in the data set (-17,497 cal/sec) would indicate a good mapping of the data. The lower bound heat duty observed on a neuron for each architecture was summarized in Figures 5 and 7. In the chemical system, negative heat duty indicates an exothermic process, while positive heat duty indicates an endothermic process.

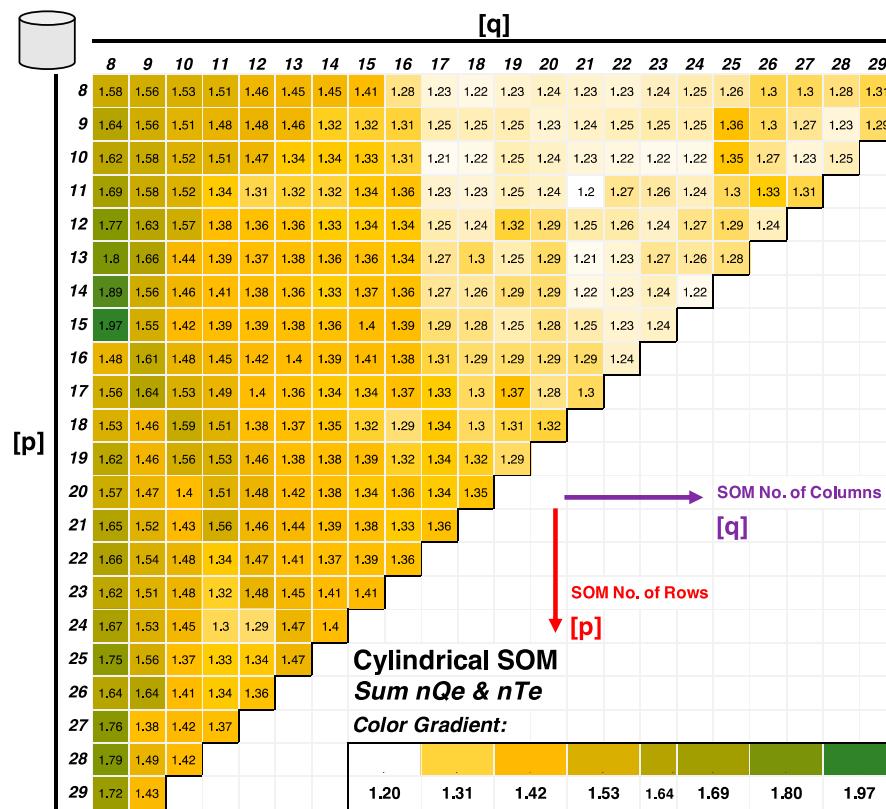
Comparing Figures 4 and 5 for the rectangular SOM, the size and shape of SOM that minimizes the sum of errors ( $8 \times 9$  and  $8 \times 10$  grids) (Figure 4) does not overlay well with the size and shape of SOM that renders heat duty value closest to the lower bound value (example:  $8 \times 22$  to  $8 \times 29$  grids, or  $10 \times 22$  to  $10 \times 28$  grids) (Figure 5). For the cylindrical SOM, on the other hand, the sizes and shapes of SOM that minimizes the sum of errors (Figure 6) overlay well with sizes and shapes that render heat duty value closest to the lower bound value (example:  $9 \times 28$  grid, or  $10 \times 28$  grid) (Figure 7).



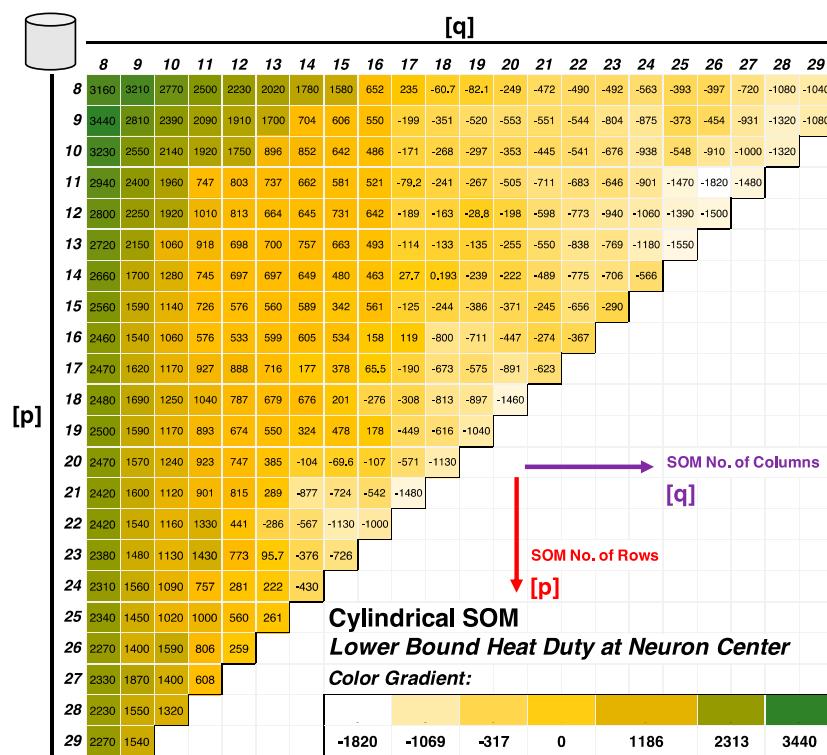
**Figure 4.** Sum of normalized topographic error (Te) and normalized quantization error (Qe) for rectangular SOM in hexagonal lattice at number of rows and number of columns of the SOM lattice.



**Figure 5.** Lowest nominal value of the model variable ‘heat duty’ in a trained rectangular SOM in hexagonal lattice at number of rows and number of columns of the SOM lattice.



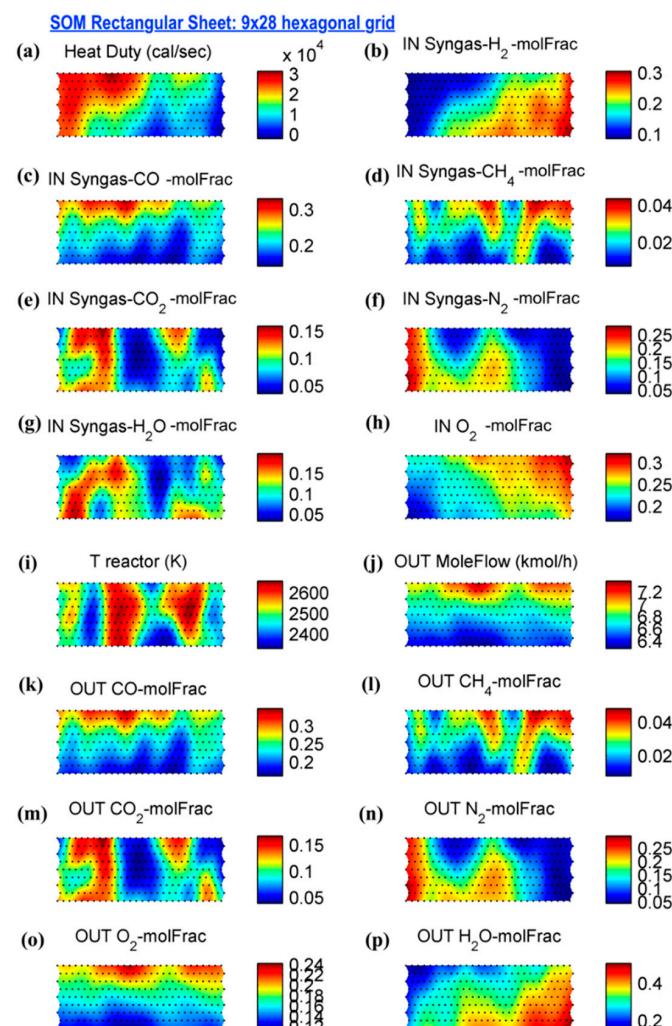
**Figure 6.** Sum of normalized topographic error ( $Te$ ) and normalized quantization error ( $Qe$ ) for cylindrical SOM in hexagonal lattice at number of rows and number of columns of the SOM lattice.



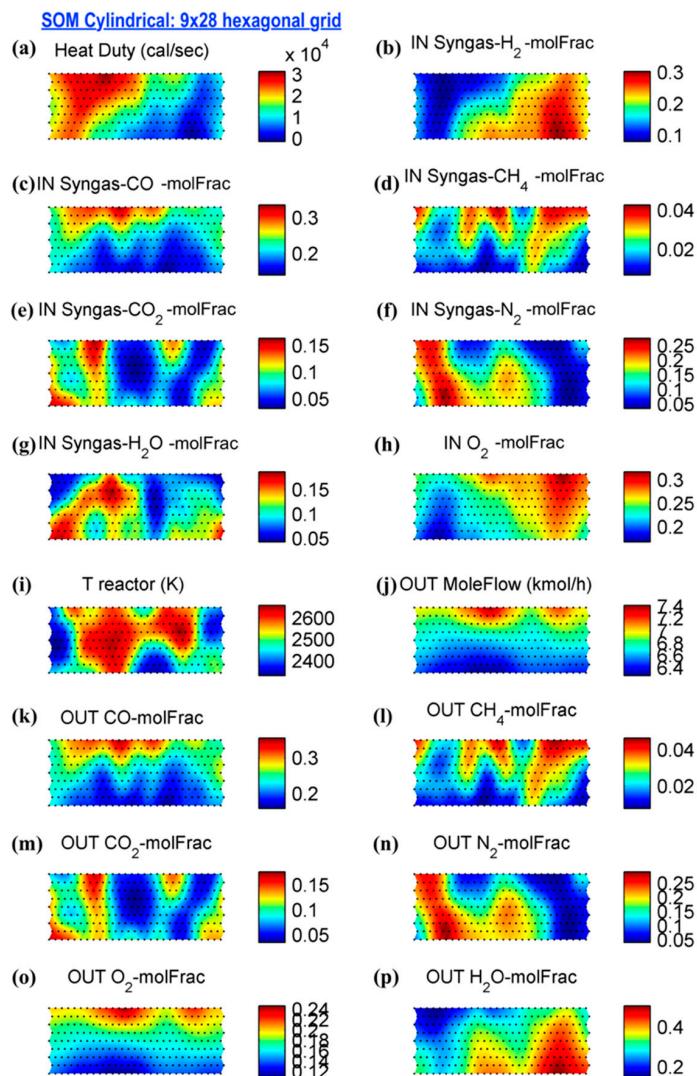
**Figure 7.** Lowest nominal value of the model variable ‘heat duty’ in a trained cylindrical SOM in hexagonal lattice at number of rows and number of columns of the SOM lattice.

### 3.2. Syngas Combustion Variables Pattern

From the evaluation of the SOM architecture's effect on the mapping performance of the simulated syngas combustion dataset (Section 3.1), one SOM architecture in the rectangular ( $9 \times 28$  hexagonal) grid and one architecture in the cylindrical ( $9 \times 28$  hexagonal) grid were used to render the patterns in the simulated variables. Figures 8 and 9 summarize the variables pattern from the trained rectangular SOM and cylindrical SOM, respectively. One of the advantages of using SOM for pattern extraction in multivariate data becomes apparent in these surface renderings. As graphically illustrated in Figure 1, each SOM neuron after training contains the quantized values of all the attributes (variables). When the quantized values are rendered as surface maps with each attribute having its own map (also called component plane), the maps of all the attributes may then be compared for pattern analysis. The black dots in the surface maps in Figures 8 and 9 indicate the neurons of the SOM. A neuron (black dot) on a specific row and column index in one component plane is the same neuron in another component plane of the same row and column index. In other words, component plane rendering is like 'slicing' a trained SOM into its various variables. This then may make the comparison of variables pattern a visual technique, which may be more advantageous over numerical measures such as variables correlation [12].



**Figure 8.** Surface projection of the system variables modelled through rectangular SOM consists of neurons in hexagonal pattern of nine grid rows  $\times$  28 grid columns. (a) process heat duty (cal/sec); (b–h) mole fraction of chemical species in the feed; (i) reactor temperature; (j) total molar out flow from reactor; (k–p) mole fraction of chemical species in the reactor outflow stream.



**Figure 9.** Surface projection of the system variables modelled through cylindrical SOM consists of neurons in hexagonal pattern of nine grid rows  $\times$  28 grid columns. (a) Process heat duty (cal/sec); (b–h) mole fraction of chemical species in the feed; (i) reactor temperature; (j) total molar out flow from reactor; (k–p) mole fraction of chemical species in the reactor outflow stream.

In general, the variable patterns via component planes from the rectangular SOM (Figure 8) are almost of the same variables pattern from the cylindrical SOM (Figure 9). This shows that either rectangular SOM ( $9 \times 28$  grid) or cylindrical SOM ( $9 \times 28$  grid) may be used to analyze the relationships among variables. A variable pattern of interest for clean technologies is the use of syngas for exothermic process (negative heat duty) that produces lesser harmful emissions [1]. In both rectangular and cylindrical SOM results, the heat duty (Figures 8a and 9a) approaches negative values (exothermic) when the mole fraction of  $H_2$  in the syngas (Figures 8b and 9b) increases while at the same time, the mole fraction of CO and  $CH_4$  decreases (Figures 8c,d and 9c,d). Interestingly, the reactor temperature (Figures 8i and 9i) does not seem to show any definite inverse or direct relationship with the heat duty (Figures 8a and 9a). The mole fraction of CO in the outflow stream (Figures 8k and 9k) does not seem to vary so much from the mole fraction of CO in the feed syngas (Figures 8c and 9c). This may be due to the high activation energy for the decomposition of CO to  $CO_2$  (reaction 2 in Table 1), which requires higher reaction temperatures to increase the reaction constant. The  $H_2$  gas in syngas supplied in all the runs was completely consumed; hence, no component plane for  $H_2$  was rendered.

#### 4. Discussion

SOM has been finding applications in areas beyond the original problems it was intended for in the 1970s and 1980s—speech recognition with the inherent tasks of clustering, visualization, and abstraction [8]. The unprecedented efforts in SOM research and applications in recent decades may indicate the growing realization of the potential of SOM as a computational tool in broader spectra. Nonetheless, current literature does not show applications of SOM in the area of computer-based simulations of chemical systems.

Simulation-based analysis of technologies aimed at clean implementation has been gaining interest, owing to the availability of computational tools. Syngas combustion has been a case study of many modelling-simulation analyses due to the rigorous models [39] and availability of empirically estimated model parameter values [40,41]. Inherent to the use of rigorous models is the task of dealing with multivariate datasets arising from chemical simulations of the process. This paper demonstrates the use of a proven machine learning algorithm—SOM—in aiding the analysis of patterns in multivariate dynamical systems such as syngas combustion. This technique is proposed to be a data analytics tool in studies that require evaluation of the effects of varying variables or variables that may naturally fluctuate within certain ranges, and that a set of process conditions must be identified for design and operation of the system. For example, the composition of the syngas was found by Kousheshi, Yari [40] to significantly affect the performance of the reactivity controlled compression ignition (RCCI) syngas/diesel engine. Moreover, the composition of syngas from the gasification of natural gas, coal, or biomass typically varies with H<sub>2</sub> as the main component followed by CO, CO<sub>2</sub> and CH<sub>4</sub> [2].

As shown in Figures 4–7, both rectangular and cylindrical SOMs may be designed to map the original dataset, while both topographic error and quantization error are minimized. The importance of properly designed SOM (size, shape, lattice structure, etc.) is apparent in these results (Figures 5 and 7). The quantization of the dataset via each neuron centroid may result to aggregation of data that may not closely map the original dataset. Selecting the lower-bound value of the heat duty (minimum: −17,497 cal/sec) as an analysis variable on this aspect revealed that certain sizes and shapes of SOM result to neurons centroids that are close to the reference value (−17,497) (upper right portion of Figures 5 and 7), while other SOM architectures produce neuron centroids far from the reference value (lower left portion of Figures 5 and 7). Another SOM design consideration that was already accounted for during the planning of the simulation experiment is the need for a sufficient number of data [8,10], which usually means a large amount of data. This requirement, however, is easily addressed in computer simulation-based works as up-to-date computers are fast enough to run multitudes of chemical process simulations through advanced methods such as parallel computation, for solving large scale equation-oriented models [42]. Further, the 4800 simulation runs in AspenPlus in this work were performed within a couple of days, most of which was spent organizing the data into the data summary spreadsheets and less time on each simulation run, which took less than a minute each on a laptop computer. The results of properly designed and trained SOMs (Figures 8 and 9) may then be used to render the component planes as surface plots. These component plane surfaces (Figures 8 and 9) are 2D visualizations of the originally multi-dimensional dataset. These 2D maps preserve the relationship of all the variables as they are the various attributes of a fixed set of neurons on the SOM. Hence, comparison of the multivariate patterns becomes a task of comparing the surface plot levels of the various variables on a particular area of the maps, since the same portion of each component plane pertains to the same set of neurons ‘sliced’ to render the component planes. This dimension reduction (usually to 2D) while preserving the variables relations in the original dataset is one of the strengths of SOM [10,43,44].

This demonstration of the use SOM for the patterns analysis of multivariate datasets from chemical process simulation aims to lay down the potential of SOM algorithm in improving the data analytics of simulation-based works on clean technologies such as syngas. Numerous possible similar applications are foreseen: integration of SOM algorithm in chemical simulation packages such AspenPlus, COMSOL MultiPhysics, etc.; application in other simulation-heavy areas such as computational fluid dynamics;

mining of material properties suitable for clean technology implementation; the visualization of the variable patterns of traditional systems, such as the operational variables of wastewater treatment facilities. The introduction of machine learning algorithms such as SOM into the methodology of developing and implementing traditional and emerging technologies may be one of the key paths to realizing clean technology implementation.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2571-8797/2/2/11/s1>, Supplementary Materials for the following maybe accessed through the electronic files provided by the authors: (1) AspenPlus model of the syngas combustion, (2) summarized results of the simulation runs, (3) MATLAB file for the SOM computations, (4) R-script on Morris sampling technique for the seven variables ( $y_{0,a}$ 's and  $T_{react}$ ), (5) summary of the 4800 simulation run conditions ( $y_{0,a}$ 's and  $T_{react}$ ), (6) comparison of U-matrix between  $9 \times 28$  rectangular grid and  $47 \times 28$  rectangular grid, and (7) frequency distribution of the 16 variables used for SOM training.

**Author Contributions:** Conceptualization, D.L.B.F., M.C., W.S., and E.R.; methodology, D.L.B.F., M.C., A.D., S.K., M.L., A.F., W.S.; software, D.L.B.F., M.C., A.D., S.K., M.L. and A.F.; validation, D.L.B.F. and M.C.; formal analysis, D.L.B.F., W.S., E.R., R.H. and M.Z.; resources, D.L.B.F., W.S., E.R., R.H. and M.Z.; writing—original draft preparation, D.L.B.F., W.S. and E.R.; writing—review and editing, D.L.B.F., W.S., E.R., R.H. and M.Z.; visualization, D.L.B.F., M.C., A.D., S.K., M.L. and A.F.; supervision, D.L.B.F.; funding acquisition, D.L.B.F., W.S., E.R., R.H. and M.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** Partial funding for the research presented in this paper was provided by NASA EPSCoR, the Louisiana Board of Regents, and Louisiana Space Grant (LaSPACE).

**Acknowledgments:** Appreciation is extended to the students and staff of the Energy Institute of Louisiana (EIL).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Nomenclature

ANN	artificial neural network
$C_a$	chemical species $a$ concentration, mol/m <sup>3</sup>
$r_{b,a}$	rate of generation (+) or consumption (−) of $a$ in reaction step $b$ , mol/m <sup>3</sup> /s
$k_b$	reaction constant parameter in reaction $b$
$T_{react}$	reaction temperature, degree K
$y_a$	gas-phase mole fraction of chemical species $a$ at reactor exhaust, %
$y_{0,a}$	gas-phase mole fraction of chemical species $a$ in the syngas feed, %
BMU	best-matching unit neuron in the SOM grid
X	SOM input dataset of dimension $J \times N$
$\xi_n$	$n$ th feature variable of the SOM input dataset X, $n \in [1 : N]$
$x_j$	$j$ th observation vector in the SOM input dataset X, $j \in [1 : J]$
$m_i$	weight vector of $i$ th neuron in the SOM grid
PFR	plug-flow reactor
SOM	self-organizing map
Te	topographic error of trained SOM
Qe	quantization error of trained SOM

## References

- Whitty, K.J.; Zhang, H.R.; Eddings, E.G. Emissions from Syngas Combustion. *Combust. Sci. Technol.* **2008**, *180*, 1117–1136. [[CrossRef](#)]
- Yang, L.; Ge, X. Chapter Three-Biogas and Syngas Upgrading. In *Advances in Bioenergy*; Li, Y., Ge, X., Eds.; Elsevier: Amsterdam, The Netherlands, 2016; pp. 125–188.
- De Kam, M.J.; Morey, R.V.; Tiffany, D.G. Integrating Biomass to Produce Heat and Power at Ethanol Plants. *Appl. Eng. Agric.* **2009**, *25*, 227–244. [[CrossRef](#)]
- Martinez, J.D.; Mahkamov, K.; Andrade, R.V.; Lora, E.E.S. Syngas production in downdraft biomass gasifiers and its application using internal combustion engines. *Renew. Energy* **2012**, *38*, 1–9. [[CrossRef](#)]

5. Gupta, A. Introduction to Deep Learning: Part 1. In *Chemical Engineering Progress*; AIChE: New York, NY, USA, 2018; pp. 22–29.
6. Dimian, A.C.; Bildea, C.S.; Kiss, A.A. Chapter 2-Introduction in Process Simulation. In *Integrated Design and Simulation of Chemical Processes*; Dimian, A.C., Bildea, C.S., Kiss, A.A., Eds.; Elsevier: Amsterdam, The Netherlands, 2014; pp. 35–71.
7. Dimian, A.C.; Bildea, C.S.; Kiss, A.A. Chapter 1-Integrated Process and Product Design. In *Integrated Design and Simulation of Chemical Processes*; Dimian, A.C., Bildea, C.S., Kiss, A.A., Eds.; Elsevier: Amsterdam, The Netherlands, 2014; pp. 1–33.
8. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Information Sciences; Kohonen, T., Ed.; Springer: New York, NY, USA, 2001.
9. Kohonen, T. Self-Organizing Feature Maps. In *Self-Organization and Associative Memory*; Springer: New York, NY, USA, 1989.
10. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **2013**, *37*, 52–65. [[CrossRef](#)] [[PubMed](#)]
11. Lobo, V.J.A.S. Application of Self-Organizing Maps to the Maritime Environment. In *Information Fusion and Geographic Information Systems*; Springer: Berlin/Heidelberg, Germany, 2009.
12. Vesanto, J. SOM-based data visualization methods. *Intell. Data Anal.* **1999**, *3*, 111–126. [[CrossRef](#)]
13. Kangas, J.; Kohonen, T. Developments and applications of the self-organizing map and related algorithms. *Math. Comput. Simul.* **1996**, *41*, 3–12. [[CrossRef](#)]
14. Rumelhart, D.E.; Zipser, D. Feature Discovery by Competitive Learning. *Cogn. Sci.* **1985**, *9*, 75–112. [[CrossRef](#)]
15. Faigl, J. An Application of Self-Organizing Map for Multirobot Multigoal Path Planning with Minmax Objective. *Comput. Intell. Neurosci.* **2016**, *2016*, 1–15. [[CrossRef](#)] [[PubMed](#)]
16. Resta, M. Graph Mining Based SOM: A Tool to Analyze Economic Stability. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
17. Kamimura, R. Social Interaction and Self-Organizing Maps. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
18. Benítez-Pérez, H.; Ortega-Arjona, J.L.; Pérez, A.B. Using Wavelets for Feature Extraction and Self Organizing Maps for Fault Diagnosis of Nonlinear Dynamic Systems. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
19. Henriques, R.; Lobo, V.; Baçao, F. Spatial Clustering Using Hierarchical SOM. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
20. Zhang, J.; Fang, H. Using Self-Organizing Maps to Visualize, Filter and Cluster Multidimensional Bio-Omics Data. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
21. Kurdtongmee, W. A Self Organizing Map Based Motion Classifier with an Extension to Fall Detection Problem and Its Implementation on a Smartphone. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
22. Skifc, N.; Francis, J. Self-Organizing Maps: A Powerful Tool for the Atmospheric Sciences. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
23. Angeli, S.; Quesney, A.; Gross, L. Image Simplification Using Kohonen Maps: Application to Satellite Data for Cloud Detection and Land Cover Mapping. In *Applications of Self-Organizing Maps*; Johnsson, M., Ed.; IntechOpen: Rijeka, Croatia, 2012.
24. Simula, O.; Kangas, J. Process Monitoring and Visualisation Using Self-Organizing Maps. *Neural Netw. Chem. Eng.* **1995**, *6*, 371–384.
25. Liukkonen, M.; Hiltunen, T.; Hälikkä, E.; Hiltunen, Y. Modeling of the fluidized bed combustion process and NO<sub>x</sub> emissions using self-organizing maps: An application to the diagnosis of process states. *Environ. Model. Softw.* **2011**, *26*, 605–614. [[CrossRef](#)]
26. Munoz, I.; Martín-Torre, M.C.; Galán, B.; Viguri, J. Assessment by self-organizing maps of element release from sediments in contact with acidified seawater in laboratory leaching test conditions. *Environ. Monit. Assess.* **2015**, *187*, 748. [[CrossRef](#)] [[PubMed](#)]
27. Ganhadeiro, T.G.L.; Christo, E.D.S.; Meza, L.A.; Costa, K.A.; Souza, D. Evaluation of Energy Distribution Using Network Data Envelopment Analysis and Kohonen Self Organizing Maps. *Energies* **2018**, *11*, 2677. [[CrossRef](#)]
28. Li, M.; Du, W.; Qian, F.; Zhong, W. Total plant performance evaluation based on big data: Visualization analysis of TE process. *Chin. J. Chem. Eng.* **2018**, *26*, 1736–1749. [[CrossRef](#)]

29. Arena, U.; Di Gregorio, F. Gasification of a solid recovered fuel in a pilot scale fluidized bed reactor. *Fuel* **2014**, *117*, 528–536. [[CrossRef](#)]
30. Khan, A.A.; De Jong, W.; Gort, D.R.; Spliethoff, H. A Fluidized Bed Biomass Combustion Model with Discretized Population Balance. 1. Sensitivity Analysis. *Energy Fuels* **2007**, *21*, 2346–2356. [[CrossRef](#)]
31. Campolongo, F.; Cariboni, J.; Saltelli, A.; Schoutens, W. Enhancing the Morris method. In Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004), Santa Fe, New Mexico, 8–11 March 2005.
32. Morris, M.D. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* **1991**, *33*, 161–174. [[CrossRef](#)]
33. McDonell, V.G. Key Combustion Issues Associated with Syngas and High-Hydrogen Fuels. 2006. Available online: <https://www.netl.doe.gov/research/coal/energy-systems/gasification/gasifipedia/syngas-composition-igcc> (accessed on 5 April 2020).
34. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. *SOM Toolbox for MATLAB* 5; Helsinki University of Technology: Espoo, Finland, 2000.
35. Vesanto, J. SOM Implementation in SOM Toolbox. 2005. Available online: <http://www.cis.hut.fi/somtoolbox/documentation/somalg.shtml> (accessed on 22 October 2019).
36. Tian, J.; Azarian, M.H.; Pecht, M. Anomaly Detection Using Self-Organizing Maps-Based K-Nearest Neighbor Algorithm. In *European Conference of the Prognostics and Health Management Society*; PHM Society: Nantes, France, 2014.
37. Pöhlbauer, G. Survey and Comparison of Quality Measures for Self-Organizing Maps. In *Fifth Workshop on Data Analysis (WDA'04)*; Elfa Academic Press: Cambridge, MA, USA, 2004.
38. Vesanto, J. Using SOM in Data Mining. In *Department of Computer Science and Engineering*; Helsinki University of Tehcnology: Espoo, Firland, 2000.
39. Srinivas, T.; Gupta, A.V.S.S.K.S.; Reddy, B.V. Thermodynamic Equilibrium Model and Exergy Analysis of a Biomass Gasifier. *J. Energy Resour. Technol.* **2009**, *131*, 031801. [[CrossRef](#)]
40. Kousheshi, N.; Yari, M.; Paykani, A.; Mehr, A.S.; De La Fuente, G. Effect of Syngas Composition on the Combustion and Emissions Characteristics of a Syngas/Diesel RCCI Engine. *Energies* **2020**, *13*, 212. [[CrossRef](#)]
41. Kéromnès, A.; Metcalfe, W.; Heufer, K.A.; Donohoe, N.; Das, A.K.; Sung, C.-J.; Herzler, J.; Naumann, C.; Griebel, P.; Mathieu, O.; et al. An experimental and detailed chemical kinetic modeling study of hydrogen and syngas mixture oxidation at elevated pressures. *Combust. Flame* **2013**, *160*, 995–1011. [[CrossRef](#)]
42. Ma, Y.; Weng, J.; Shao, Z.; Chen, X.; Zhu, L.; Zhao, Y. Parallel Computation Method for Solving Large Scale Equation-oriented Models. In *Computer Aided Chemical Engineering*; Gernaey, K.V., Huusom, J.K., Gani, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 239–244.
43. Qian, J.; Nguyen, N.P.; Oya, Y.; Kikugawa, G.; Okabe, T.; Huang, Y.; Ohuchi, F.S. Introducing self-organized maps (SOM) as a visualization tool for materials research and education. *Results Mater.* **2019**, *4*, 100020. [[CrossRef](#)]
44. Kim, K.P.; Yusof, F.; Daud, Z.B.M. Multi-dimensional reduction using self-organizing map. In *21st National Symposium on Mathematical Sciences (SKSM21)*; AIP Publishing: Penang, Malaysia, 2014.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).