CHEMICAL ENGINEERING TRANSACTIONS
VOL. 106, 2023
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Jeng Shiun Lim, Nor Alafiza Yunus, Peck Loo Kiew, Hon Huin Chin

ISBN 979-12-81206-05-2; **ISSN** 2283-9216
DOI: 10.3303/CET23106207

Artificial Neural Network Model for Palm Oil Biomass Gasification Process Prediction

Aisyah Alya Nureena Azhar[a,b], Norhuda Abdul Manaf[a,*], Helmut F. Yabar[b]

[a]Department of Chemical and Environmental Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[b]Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8527, Japan
norhuda.kl@utm.my

Palm oil biomass-based gasification has become a potential technology to overcome anthropogenic environmental challenges. While physical experimentation is time-consuming and expensive, it can be avoided when determining the best settings for a particular gasifier and the behaviour of palm oil biomass. An Artificial Neural Network (ANN) model was developed to estimate syngas composition (CO and H_2) over a wide range of palm oil biomass characteristics and gasifier operating conditions. A vast amount of secondary data comprising both categorical and numerical was gathered for the development of the proposed ANN model. To improve the model's performance, uncorrelated input data were removed using International Business Machines Statistical Package for the Social Sciences (IBM SPSS) Statistics software by utilizing Spearman's Correlation Coefficient (SCC) matrix. Feed-Forward Back Propagation (FFBP) and Levenberg-Marquardt (LM) learning algorithms with one and two hidden layers, as well as a range number of neurons and transfer functions were used to train the network using the ANN toolbox, available in Simulink, MATLAB software. The best-performing network structure was identified based on the lowest Mean-Squared Error (MSE) and highest Regression value, subjected to numbers of network topologies. The developed ANN model is able to accurately predict the output of syngas composition (MSE $\leq$ 0.1 and R^2 > 0.8). The results indicated that the ANN model shows excellent model prediction which can aid in the effective operation of biomass gasification under various operating conditions.

1. Introduction

Malaysia is the second largest producer of crude palm oil in 2021 with approximately 18.11 million metric tons and exports around a third of the world's palm oil (Statista Research Department, 2021). The Malaysian government has announced several incentives to encourage the use of biomass as an alternative source, such as The National Biofuel 2006 (NBP 2006), which acted as the nation's basis to promote the biodiesel industry by incorporating processed palm oil (Rashidi et al., 2022). The Fifth Fuel Policy (5FP2000), introduced by the Malaysian government in 2000, included renewable energy as the fifth fuel in the Eight Malaysian Plan, which ran from 2001 to 2005. Although the strategy has strongly emphasized other renewable energy sources like solar energy and hydropower, biomass, and biofuels were also included. Many attempts are made to turn biomass into high-value products.

A large amount of waste in the form of palm kernel shells (PKS), empty fruit bunch (EFB), mesocarp fibres (MF), palm oil frond (POF), palm oil trunk (POT) and palm kernel cake (PKC) are produced during the extraction of palm oil from palm trees (Parthasarathy et al., 2022). Thermochemical conversion through gasification can be utilized to extract energy from both waste and biomass feedstock to address the problem of excess palm oil waste (Ascher et al., 2022b). The process of gasification involves heating solid biomass or other carbonaceous solids to create synthetic gas (syngas), which are composed of nitrogen gas (N_2), hydrogen gas (H_2), carbon monoxide (CO), carbon dioxide (CO_2), and methane (CH_4) (Zhang et al., 2019). Almost any dry organic material can be gasified through gasification to create a clean burning fuel that can substitute fossil fuels. However, physical experimentation to find output composition, using a given gasifier is time-consuming and expensive.

Paper Received: 05 June 2023; Revised: 15 July 2023; Accepted: 22 September 2023
Please cite this article as: Azhar A.A.N., Manaf N.A., Yabar H.F., 2023, Artificial Neural Network Model for Palm Oil Biomass Gasification Process Prediction, Chemical Engineering Transactions, 106, 1237-1242 DOI:10.3303/CET23106207

As a result, artificial intelligence (AI) like Artificial Neural Network (ANN) comes in handy for outcome prediction when important interactions of complex non-linearities exist in a data set, such as in biomass gasification.
ANN analysis is modelled after how biological neurons communicate with one another, and it consists of an input layer, hidden layers, and output layers (Mohseni-Dargah et al., 2022). ANN is highly capable to model and predict the plant performance between the input and output variables and has outstanding learning capabilities (Serrano et al., 2020). A study by Ascher et al. (2022a) proposed an ANN model to predict ten key measures of gasification performance and the authors displayed a strong predictive performance with coefficient of determination (R^2) of 0.9310.
To the authors' knowledge, very few of the AI-based gasification models can be used to process a variety of palm oil biomass feedstock utilizing both categorical and numerical data using different gasifiers. Hence, this study aims to use ANN to predict gasification outputs across a variety of features of palm oil biomass and gasifier operation condition using ANN. The correlations between the input variables, as well as the best-performing network structures, were also identified and evaluated.

2. Materials and methods

2.1 Data gathering

A total of 334 data sets were gathered from 31 articles. This research only focuses on palm oil biomass, hence only palm oil biomass feedstock information was collected. The data gathered were categorized into input and output layers. For the input layer, the variables are feedstock information, ultimate analysis, proximate analysis, lignocellulosic composition, and gasifier operating conditions. Under each of the variables, there are sub-variables which make a total of 20 input layers. For the output layer, the variables are gas composition, yield and process efficiency. Under each of the variables, there are sub-variables which makes a total of 15 output layers.

2.2 Data pre-processing

2.2.1 Label encoding

Categorical data must be encoded into numbers before using it for model development in the MATLAB ANN toolbox. In label encoding, each data was assigned a value from 1 through N, where N is the number of categories for the variables. There is no relation or order between these categories.

2.2.2 Data analysis

Spearman's Correlation Coefficient (SCC) was calculated using International Business Machines Statistical Package for the Social Sciences (IBM SPSS) Statistics software to find the monotonic relationship between the input variables. Eq (1) was used to calculate SCC:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{1}$$

where r_s is the coefficient, n is the number of points, d^2 is the square of the difference in the ranks of the two coordinates (x, y), and $\sum d^2$ is the sum of each square.

2.2.3 Data normalization

Normalizing data sets transform the data sets to be on a similar scale. Normalization facilitates the training of ANN as the different features are on a similar scale, which helps to stabilize the gradient descent step and helps models converge faster for a given learning rate. Eq (2) was used to normalize the data sets from 0 to 1:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}(u - l) + l \tag{2}$$

where x is the original data, x' is the normalized data, x_{max} is the maximum original value, x_{min} is the minimum original value, u is the upper bound and l is the lower bound of the new range for normalized data. A lower bound of 0 and an upper bound of 1 were applied in this work.

2.3 ANN model development

The data were divided into 80 % training for model development and 20 % testing for performance evaluation. Under the 80% training, the data were trained and validated. Three variables and seven constants were employed to determine the optimum network topologies, as illustrated in Table 1. Feed-Forward Back Propagation (FFBP) were used to reduce error and Levenberg-Marquardt (LM) were used to adjust the weights in accordance with the calculated error (Sushmi & Subbulekshmi, 2022). Their work found satisfactory prediction value with the combined use of FFBP and LM. The transfer function used to develop the output were hyperbolic tangent sigmoid (TANSIG), linear transfer function (PURELIN), and logistic sigmoid (LOGSIG). The number of layers were varied from 1 to 2, whereas the number of neurons were varied from 1 to 10. The performance function was determined by taking the lowest Mean-Squared Error (MSE) value and highest regression value.

Table 1: Fixed and varied parameters during ANN training

	Parameters	Type / Value
Fixed parameters		
	Training data	80 %
	Testing data	20 %
	Network type	FFBP
	Training function	LM
	Performance function	MSE
	Maximum epoch	1,000
	Validation checks	100
Varied parameters		
	Transfer function	TANSIG, PURELIN, LOGSIG
	Number of layers	1 - 2
	Number of neurons	1 - 10

2.4 Performance evaluations

The 20 % testing data were used for the performance evaluations. The Mean Squared Error (MSE) and R^2 were used to assess how well the model performed in making predictions. Eqs(3) and (4) show the formula for the mentioned prediction performance:

$$MSE = \frac{\sum_{i=1}^{n}(y_i{}^a - y_i{}^p)^2}{n} \tag{3}$$

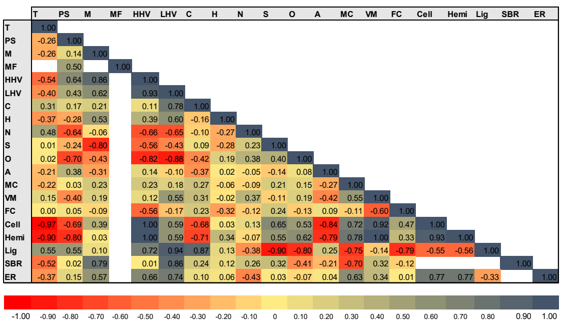
$$R^2 = 1 - \frac{\sum_{I=1}^{n}(y_i{}^a - y_i{}^p)^2}{\sum_{I=1}^{n}(y_i{}^a - \overline{y^a})^2} \tag{4}$$

where n is the number of samples, $y_i{}^a$ is the actual value, $y_i{}^p$ is the predicted value, and $\overline{y^a}$ is the mean of the actual value.

3. Results and discussions

3.1 Input variable analysis

The monotonic relationship between the 20 input variables was analyzed by SCC using IBM SPSS Statistics. An SCC value of 0 denotes a lack of correlation between two variables. The monotonic relationship is strong the closer it is to ± 1. To improve ANN training in this study, variables with no SCC value were eliminated. Based on Figure 1, the input variable mass fraction (MF) was removed as it has no correlation with other input variables. This makes the new input variables to be a total of 19 variables. An SCC ≥ $|0.6|$ was used in a study by Ascher et al. (2022a) to denote a strong correlation, which causes one of the parameters to be removed before model training. SCC was chosen as compared to Pearson correlation, since Pearson correlation only reflects the strength of the linear relationship between two variables, while SCC determines the strength of the monotonic relationship between the two variables (Kim et al., 2023).



	T	PS	M	MF	HHV	LHV	C	H	N	S	O	A	MC	VM	FC	Cell	Hemi	Lig	SBR	ER
T	1.00																			
PS	-0.26	1.00																		
M	-0.26	0.14	1.00																	
MF		0.50		1.00																
HHV	-0.54	0.64	0.86		1.00															
LHV	-0.40	0.43	0.62		0.93	1.00														
C	0.31	0.17	0.21		0.11	0.78	1.00													
H	-0.37	-0.28	0.53		0.39	0.60	-0.16	1.00												
N	0.48	-0.64	-0.06		-0.66	-0.65	-0.10	-0.27	1.00											
S	0.01	-0.24	-0.80		-0.56	-0.43	0.09	-0.28	0.23	1.00										
O	0.02	-0.70	-0.43		-0.82	-0.88	-0.42	0.19	0.38	0.40	1.00									
A	-0.21	0.38	-0.31		0.14	-0.10	-0.37	0.02	-0.05	-0.14	0.08	1.00								
MC	-0.22	0.03	0.23		0.23	0.18	0.27	-0.06	-0.09	0.21	0.15	-0.27	1.00							
VM	0.15	-0.40	0.19		0.12	0.55	0.31	-0.02	0.37	-0.11	0.19	-0.42	0.55	1.00						
FC	0.00	0.05	-0.09		-0.56	-0.17	0.23	-0.32	-0.12	0.24	-0.13	0.09	-0.11	-0.60	1.00					
Cell	-0.97	-0.69	0.39		1.00	0.59	-0.68	0.03	0.13	0.65	0.53	-0.84	0.72	0.92	0.47	1.00				
Hemi	-0.90	-0.80	0.03		1.00	0.59	-0.71	0.34	-0.07	0.55	0.62	-0.79	0.78	1.00	0.33	0.93	1.00			
Lig	0.55	0.55	0.10		0.72	0.94	0.87	0.13	-0.38	-0.90	-0.80	0.25	-0.75	-0.14	-0.79	-0.55	-0.56	1.00		
SBR	-0.52	0.02	0.79		0.01	0.86	0.24	0.12	0.26	0.32	-0.41	-0.21	-0.70	0.32	-0.12				1.00	
ER	-0.37	0.15	0.57		0.66	0.74	0.10	0.06	-0.43	0.03	-0.07	0.04	0.63	0.34	0.01	0.77	0.77	-0.33		1.00

Figure 1: SCC matrix

3.2 Model creation

A total of 93 networks were performed to find the best model. The four best networks were chosen based on their MSE and R^2 values as tabulated in Table 2. All networks have 2 hidden layers with different transfer functions and number of neurons. Different combinations of transfer functions display different MSE and R^2 values. In the case of MSE, Network 13 has the lowest value compared to the others. In the case of R^2, Network 11 has the overall highest R^2 value. The bold values represent the lowest MSE and highest R^2 values.
Figure 2 shows the regression model of all the selected four networks. Network 11 was chosen as the proposed neural network as it has the highest R^2 values compared to the other network. Figure 3 shows the proposed ANN model, having transfer function TANSIG-TANSIG and 2 hidden layers, with 8 and 15 neurons in each layer. A similar result was obtained from the study conducted by Serrano et al. (2020). Their work revealed that the application of FFBP network with 2 hidden layers (21 and 4 neurons) and TANSIG-TANSIG as transfer function exhibited low MSE and high R^2 value. The authors also performed a network comparison with other combinations of transfer functions and same number of hidden layers, but with different number of neurons. The results obtained from this study displayed similarities with the author's work, such that the network with transfer function LOGSIG-TANSIG has higher R^2 value compared to network with transfer function PURELIN-TANSIG.

Table 2: Best-selected network topologies

Network	Transfer function	Topology	MSE	R^2			
				Training	Validation	Testing	All
Network 11	TANSIG-TANSIG	19-8-15-15	0.00852	0.97433	0.92998	**0.97807**	**0.96842**
Network 13	TANSIG-TANSIG	19-10-15-15	**0.00492**	0.96991	0.95156	0.90370	0.95588
Network 21	PURELIN-TANSIG	19-8-15-15	0.00617	0.96992	**0.96102**	0.79732	0.93987
Network 33	LOGSIG-TANSIG	19-10-15-15	0.00805	**0.98063**	0.92915	0.78324	0.94726

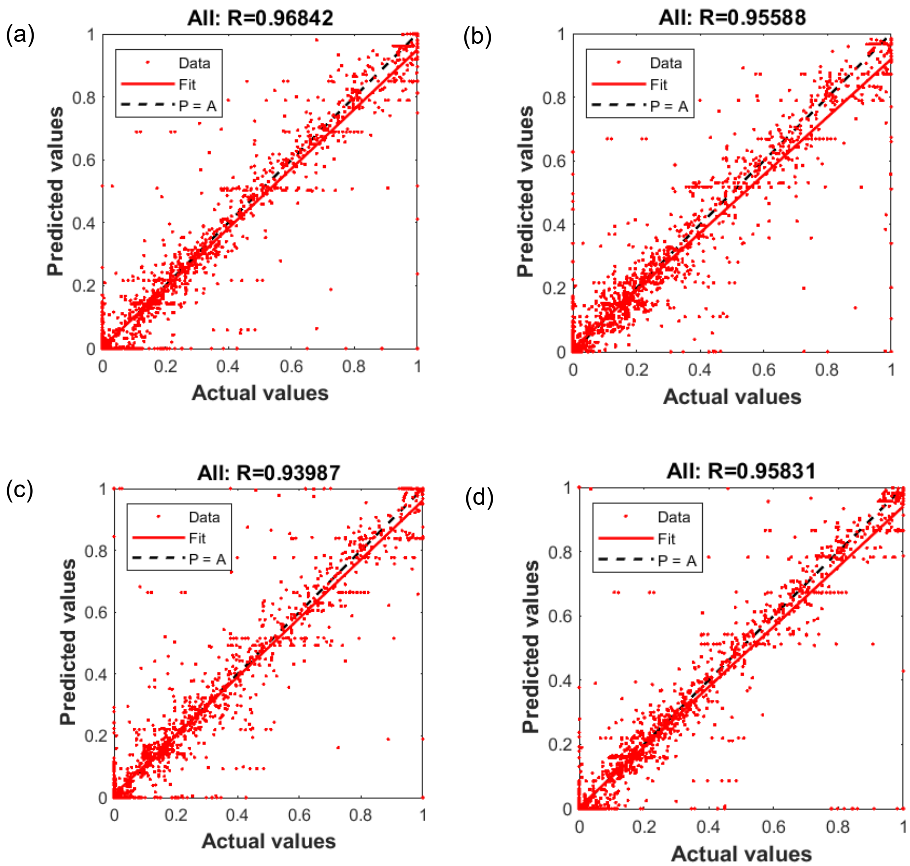


Figure 2: Regression model of the best-selected network topologies (a) Network 11 (b) Network 13 (c) Network 21 (d) Network 33

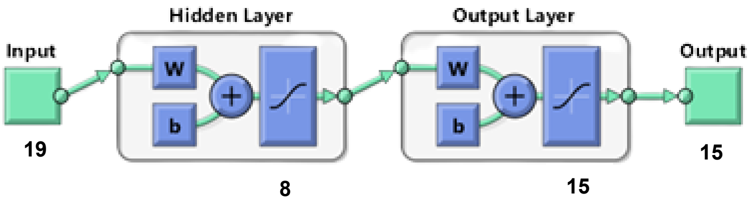


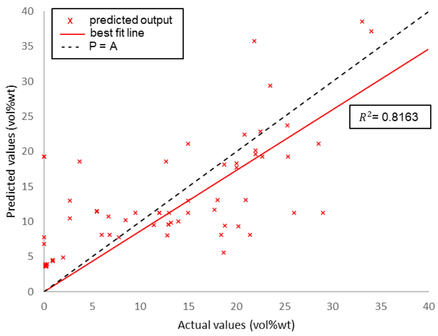
Figure 3: Proposed ANN model

3.3 Performance evaluations

The 20 % test data were introduced in all four networks to assess the ANN prediction output based on their MSE and R^2 values. In this study, the output variables syngas composition CO and H_2 were evaluated for performance evaluations. Table 3 shows the MSE and R^2 values of the predicted output, whereas Figure 4 shows the plotted predicted output of CO and H_2 using the proposed ANN model (Network 11). Both outputs from the proposed ANN display good prediction values with MSE $\leq$ 0.1 and R^2 > 0.8. A similar result was observed in Ascher et al. (2022a), where the prediction of syngas composition exhibited high R^2 value at 0.93 and 0.83. Also, Ren et al. (2023) conducted a study on machine learning methods for biomass gasification modelling by considering a monotonic relationship and recorded a high R^2 value for CO and H_2 with 0.975 and 0.954.

Table 3: MSE and R^2 values of the predicted output

Network	Transfer function	Hidden neurons	Output	MSE	R^2
network11	TANSIG-TANSIG	8	CO	**0.1**	**0.8163**
			H_2	**0.0840**	**0.9037**
network13	TANSIG-TANSIG	10	CO	0.0683	0.8320
			H_2	0.1173	0.9009
network21	PURELIN-TANSIG	8	CO	0.3094	0.2745
			H_2	0.1460	0.1322
network33	LOGSIG-TANSIG	10	CO	0.1634	0.0383
			H_2	0.1646	0

(a)


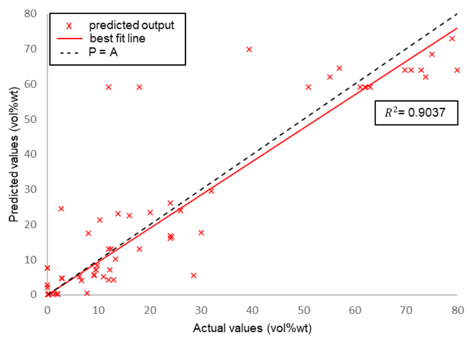
(b)


Figure 4: Predicted output of the proposed ANN (a) CO (b) H_2

4. Conclusions

An ANN model was developed in this study to predict palm oil biomass gasification process. An ANN model with the transfer function TANSIG-TANSIG has been used together with 2 hidden layers and 8 neurons. The predicted output from the proposed ANN model shows good correlations and has been verified with the test data from the literature (MSE $\leq$ 0.1 and R^2 > 0.8). The developed ANN model is able to predict gasification outputs by utilizing both categorical and numerical data from different palm oil biomass feedstocks. This, in turn, aids researchers to present and access system models for industrial plants as well as avoid physical experimentation to study palm oil biomass behaviour in fixed bed, fluidised bed, entrained flow and rotary kiln gasifier operations conditions. The limitations of this study are that the data used to develop the model are only relying on secondary data and palm oil biomass. Hence, potential areas for future research are incorporating data from industrial and a wider range of biomass feedstock. This will make the ANN model more applicable to predict gasification process from different types of biomass feedstock and gasifier operating conditions.

Abbreviations

T – type
PS – particle size (mm)
M – mass (kg)
MF – mass fraction
HHV – higher heating value (MJ/kg)
LHV – lower heating value (MJ/kg)
C – carbon (wt%)
H – hydrogen (wt%)
N –nitrogen (wt%)
S – sulfur (wt%)
O – oxygen (wt%)
A – ash (wt%)
MC – moisture content (wt%)
VM – volatile matter (wt%)
FC – fixed carbon (wt%)
Cell – cellulose (%wt)
Hemi – hemicellulose (%wt)
Lig – lignin (%wt)
SBR – steam biomass ratio (wt/wt)
ER – equivalence ratio

Acknowledgements

The authors wish to acknowledge financial support from Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur.

References

Ascher S., Sloan W., Watson I., You S., 2022a, A comprehensive artificial neural network model for gasification process prediction, Applied Energy, 320, 119289.

Ascher S., Wang X., Watson I., Sloan W., You S., 2022b, Interpretable machine learning to model biomass and waste gasification, Bioresource Technology, 364, 128062.

Kim J.Y., Kim D., Li Z.J., Dariva C., Cao Y., Ellis N., 2023, Predicting and optimizing syngas production from fluidized bed biomass gasifiers: A machine learning approach, Energy, 263, 125900.

Mohseni-Dargah M., Falahati Z., Dabirmanesh B., Nasrollahi P., Khajeh K., 2022, Chapter 12 - Machine learning in surface plasmon resonance for environmental monitoring, Chapter In: M. Asadnia, A. Razmjou, A. Beheshti (Eds.), Artificial intelligence and data science in environmental sensing, Academic Press, United Kingdom, 269-298.

Parthasarathy P., Alherbawi M., Shahbaz M., Mackey H.R., McKay G., Al-Ansari T., 2022, Conversion of oil palm waste into value-added products through pyrolysis: A sensitivity and techno-economic investigation. Biomass Conversion and Biorefinery, DOI: 10.1007/s13399-022-03144-2.

Rashidi N.A., Chai Y.H., Yusup S., 2022, Biomass energy in Malaysia: Current scenario, policies, and implementation challenges, BioEnergy Research, 15(3), 1371-1386.

Ren S., Wu S., Weng Q., 2023, Physics-informed machine learning methods for biomass gasification modeling by considering monotonic relationships, Bioresource Technology, 369, 128472.

Serrano D., Golpour I., Sánchez-Delgado S., 2020, Predicting the effect of bed materials in bubbling fluidized bed gasification using artificial neural networks (ANNs) modeling approach, Fuel, 266, 117021.

Sushmi N.B., Subbulekshmi D., 2022, Performance analysis of FFBP-LM-ANN based hourly GHI prediction using environmental variables: A case study in Chennai. Computational Intelligence and Renewable Energies, 2022, 1713657.

Statista Research Department, 2022, Production of crude palm oil in Malaysia 2012-2021, Statista Research Department<www-statista-com.ezproxy.utm.my/statistics/794709/crude-palm-oil-production-volume-malaysia/> accessed 21.11.2022.

Zhang Y., Cui Y., Chen P., Liu S., Zhou N., Ding K., Fan L., Peng P., Min M., Cheng Y., Wang Y., Wan Y., Liu Y., Li B., Ruan R., 2019, Chapter 14 - Gasification technologies and their energy potentials, Chapter In: M. J. Taherzadeh, K. Bolton, J. Wong, A. Pandey (Eds.), Sustainable resource recovery and zero waste approaches, Elsevier, United States of America, 193-206.