Modelling of Hydrogen production bioprocess using Artificial Neural Networks (ANN)

A dissertation submitted in fulfillment of the academic requirements for the degree of

Master of Science

by

Yeshona Sewsynker

Student number: 210550313

(BSc Hons, UKZN-PMB)

Discipline of Microbiology

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg Campus

South Africa

12 January 2016

PREFACE

The research contained in this dissertation/thesis was completed by the candidate while based in the Discipline of Microbiology, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg Campus, South Africa. The research was financially supported by the National Research Foundation (Grant number: 87661).

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

_____

Signed: Dr. Evariste Bosco Gueguim-Kana

Date: 12 January 2016

DECLARATION 1: PLAGIARISM

I, Yeshona Sewsynker, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

_____

Signed: Yeshona Sewsynker

Date: 12 January 2016

DECLARATION 2: PUBLICATIONS

This thesis consists of one published work. The first author (student) contributed towards experimental work, data collection and manuscript preparation and was guided by the second (supervisor) and/or third author.

Publications (CHAPTER 5)

1. Sewsynker, Y., Gueguim-Kana, E.B. and Lateef, A. 2015. Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs). Biotechnology and Biotechnological Equipment 29(6):1208-1215.

_____

Signed: Yeshona Sewsynker

Date: 12 January 2016

CONFERENCE CONTRIBUTIONS

1. Sewsynker, Y, Gueguim-Kana, EB. Does the Volume Matter? An Insight into Modelling and Optimization of Biohydrogen production across scales. College of Agriculture, Engineering and Science. Research day. 22 September 2015. University Of KwaZulu-Natal, Pietermaritzburg, South Africa.

2. Sewsynker, Y, Gueguim-Kana, EB. Does the Volume Matter? An Insight into Modelling and Optimization of Biohydrogen production across scales. Fourth International Conference on Power Science and Engineering. 15-16 December 2015. Netherlands, Amsterdam.

ABSTRACT

The depletion of fossil fuels and environmental impacts from its combustion are stimulating the development of biofuel production processes. Bioprocess models are required at various stages of this process development to determine the optimum setpoints for maximum yields. In this thesis, a review of the literature on the use of Artificial Neural Networks (ANN) as a tool for modelling and optimization of biofuel production with emphasis on dark fermentative hydrogen production was carried out. Then the impact of culture volume on the accuracy of bioprocess models was studied using Artificial Neural Networks (ANN) and the Response Surface Methodology (RSM). Additionally, ANN was used to develop intelligent bioprocess models to predict hydrogen production based on physicochemical parameters for dark fermentation and Microbial Electrolysis Cells (MECs). The review examined the application of ANN for the modelling and optimization of biohydrogen, biogas, biodiesel, microbial fuel cell technology and bioethanol. The efficiency of ANN in abstracting the non-linear relationship that exists between process inputs and biofuel yield was highlighted. The studies indicated that ANN exhibits superior modelling and optimization ability for biofuel production processes over alternative methods such as the Response Surface Methodology (RSM).

The impact of culture volume on the accuracy of bioprocess models was assessed on ANN and RSM based process models. The process input parameters were hydraulic retention time (10-48 h), inoculum (10-50%) and molasses concentration (100-300 g/L) on the hydrogen yield (mol H_2/ mol sucrose consumed) and two different process scales were considered (80 and 800 mL). The ANN based models gave coefficient of determination (R^2) values of 0.99 and 0.95 whereas the RSM based models gave R^2 values of 0.97 and 0.89 for 80 and 800 mL, respectively. Variations in predictions of optimum setpoints by all four models were negligible. All four optimized conditions were further evaluated at semi-pilot scales (8 L). A comparative assessment of semi-pilot scale and lab scale yields showed a negligible discrepancy. The microbial community responsible for hydrogen production was examined using Next generation sequencing (NGS). Presumptive hydrogen-producing microorganisms present within this system were members of the genus *Clostridia*, *Enterobacter* and *Klebsiella*. This study revealed that volume reduction does not significantly impact on the accuracy of the process model but rather reduces the costs of process development.

The intelligent process models were developed using Multilayer Perceptron (MLP) Neural Networks and trained on bioprocess data available in the public domain from selected studies. The first two models focused on hydrogen production via dark fermentation process with varying yield expression units. The considered input parameters were inoculum type, substrate type, substrate concentration, pH and temperature and the output was the hydrogen yield expressed as mole of hydrogen per mole of substrate (Mol_Model) and cumulative volume (mL) of hydrogen per gram of substrate (Vol_Model). A topology of 5-7-7-1 corresponding to the number of neurons of inputs, hidden (2) and output layers for both models was used with data sizes of 133 (Mol_Model) and 49 (Vol_Model) from 49 and 15 published studies, respectively. For these two models, a high coefficient of determination (R^2) was obtained for the Vol_Model (0.90) compared to the Mol_Model (0.46). Thus, the Vol_Model shows higher predictive accuracy compared to the Mol_Model.

The third model focused on hydrogen production using Microbial Electrolysis Cells (MECs). The considered inputs were substrate type, substrate concentration, pH, temperature, applied voltage, reactor configuration and the output was the hydrogen yield (mol H_2/ mol substrate). A committee of neural networks with a topology of 6-(6, 8, 11, 12, 14)-1 was used. The training data size was 50 from 15 published studies. The coefficients of determination (R^2) for the five models were as follows: 0.90, 0.81, 0.85, 0.70 and 0.80 with an average R^2 value of 0.85 for the five models. Validation on unknown inputs for new MEC processes showed a strong correlation between the observed and predicted hydrogen yields.

The findings from these studies demonstrate that ANN based models are efficient in the development of biofuel processes. Process miniaturization does not impact on the accuracy of ANN and RSM derived process models thus reducing the process development time and costs. Furthermore, ANNs may be used to develop intelligent models to predict hydrogen yield on novel processes based on existing data in public repositories. This will shorten the hydrogen process development time and cost.

Keywords: Modelling and optimization, Biohydrogen production, Artificial Neural Networks, Bioprocess models, Renewable energy

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for his blessings that has helped me attain my goals thus far.

I would like to acknowledge the following people:

To my husband, Avinash Sukai for his support and continuous belief in me;

My supervisor Dr. E.B. Gueguim-Kana for his guidance and brilliant ideas that inspired this research;

The supporting staff at the University of KwaZulu-Natal, Pietermaritzburg campus (Prof. S. Schmidt, Mr. C.H. Hunter, Ms. C. Clark, Dr. S.F. Jamal-Ally and Miss D. Fowlds) for their assistance both academically and professionally;

My family for their well wishes and support;

My friends and peers (Dr. Funmilayo Faloye, Preshanthan Moodley, Mariam Khan, Daneal Rorke, Jade Whiteman, Efenaide Okoh, Nolubabalo Ndaliso, Joelle Penniston and Selisha Naidoo) for their help and encouragement;

The National Research Foundation (NRF) for funding this research.

TABLE OF CONTENTS

Page no.

Note: This thesis consists of a compilation of three manuscripts and one publication whereby each chapter is presented as an individual entity according to the respective journal's rules and regulations. Therefore, there is repetition between chapters.

LIST OF TABLES

CHAPTER 4 CONTINUED…. **Page no.**

CHAPTER 5

LIST OF FIGURES

Figure **Page no.**

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4 **Page no.**

CHAPTER 5

LIST OF ABBREVIATIONS

AA	Ant Algorithm
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANN-GA	Artificial Neural Network coupled Genetic Algorithm
BP	Beyond Petroleum
BPNN	Back Propagation Neural Networks
BOD	Biochemical oxygen demand
CO_2	Carbon Dioxide
COD	Chemical oxygen demand
COD_{in}	Initial chemical oxygen demand
COD_{out}	Final chemical oxygen demand
DNA	Deoxyribonucleic acid
DOE	Factorial Design of Experiment
EA	Ensemble averaging
EIA	Energy Information Administration
FL	Fuzzy Logic
GA	Genetic Algorithm
GRNN	Generalized Regression Neural Networks
H_2	Hydrogen
H_2S	Hydrogen sulfide
HPR	Hydrogen production
HRT	Hydraulic retention time
HY	Hydrogen yield

IPCC	Intergovernmental Panel on Climate Change
ME	Mixture of experts
MEC	Microbial Electrolysis Cells
MFC	Microbial Fuel Cells
MLP	Multilayer perceptron
MSE	Mean Square error
NGS	Next Generation Sequencing
NH_3	Ammonia
OLR	Organic loading rate
ORP	Oxidation-reduction potential
OVAT	One Variable at a time
PHP	Hypertext Preprocessor
PSO	Particle Swarm Optimization
R^2	Coefficient of determination
RBFNN	Radial basis function-based neural networks
RMSE	Root-mean-square error
RSM	Response Surface Methodology
S	Sulfur
S_o	Initial substrate concentration
T°C	Temperature
TOC_{eff}	Effluent total organic carbons
TSS	Total suspended solids
TSS_{in}	Initial total suspended solids
TSS_{out}	Final total suspended solids
USDOE	United States Department of Energy

VFA	Volatile Fatty Acid
VSS	Volatile suspended solids
X_o	Initial biomass concentration

CHAPTER 1

General Introduction

1.1 The need for renewable and sustainable energy sources

Global energy demand is rapidly increasing as populations continue to grow at accelerated levels (Nath and Das, 2011; BP, 2015a). Currently, fossil fuel sources such as coal, petroleum, bitumen, natural gas and tar sand are used as primary sources of energy to meet global energy demand (Das and Veziroglu, 2001). Dependence on fossil fuels has led to the depletion of these energy reserve combined with environmental pollution (Levin *et al*., 2004). This poses significant challenges on a global scale (Levin *et al*., 2004). Among these fossil fuels, oil reserves have shown to be the most exploited energy source globally (BP, 2015a). The annual Beyond Petroleum (BP) statistical review (2015a) reported that the total proven oil reserves reached nearly 1700.1 billion barrels at the end of 2014 which was sufficient to meet approximately 52.5 years of global production. The Middle East is the major oil supplier contributing to a staggering 47.7% of the total world oil reserves (BP, 2015a). However, they are currently experiencing several challenges with regard to government instability, civil unrest and terrorism which pose global energy concerns (Mecad, 2013). Sorrell *et al.* (2009) predicted that a peak in oil production in the Middle East would occur before the year 2020. But due to the high energy demand, these oil reserves are diminishing at an alarming rate (Li, 2007). The global oil reserve to production ratio (Figure 1.1) shows that the Middle East oil reserves, according to the current production rate, would last for approximately 78 years (BP, 2015a).

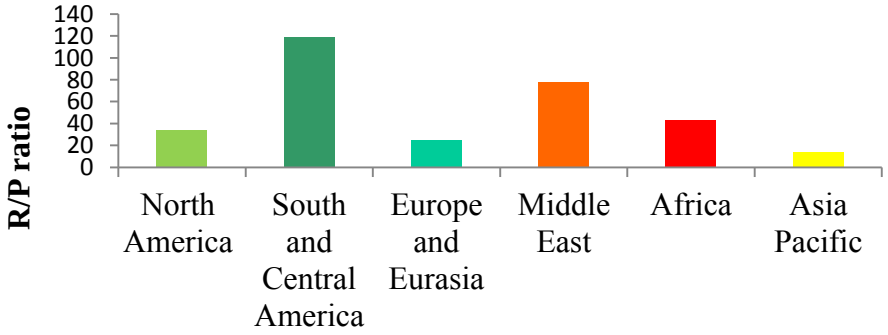


Figure 1.1. Regional oil reserves to production ratio for 2014 (BP, 2015a).

Besides its depletion and non-renewable nature, fossil fuel consumption is detrimental to the environment as well as to human health (Levin *et al*., 2004). Combustion of fossil fuels results in the release of greenhouse gases such as carbon dioxide (CO_2), methane and nitrous oxide (IPPC, 2014). As illustrated in Figure 1.2, the highest carbon emissions were as a result of fossil fuel combustion and industrial application with a total of 65% CO_2 emitted globally (IPPC, 2014).

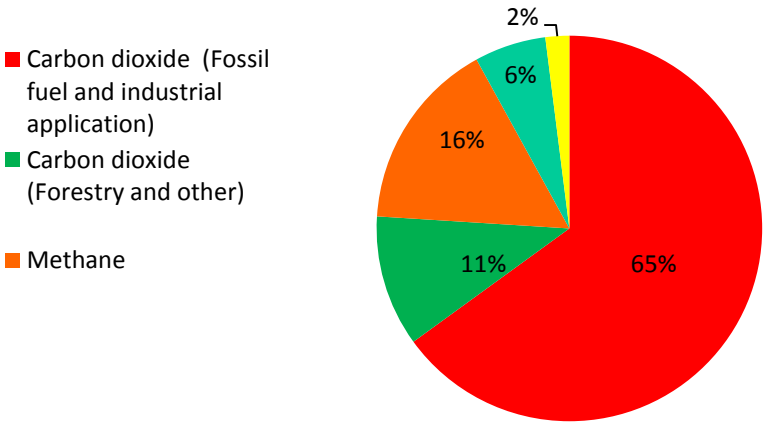


Figure 1.2. Global greenhouse gas emissions for 2010 (IPCC, 2014).

The increased emission of greenhouse gases has been associated with an increase in atmospheric temperature most commonly referred to as global warming in addition to acid rain, ozone depletion, eutrophication and health implications (Smith *et al*., 2009; Hook and Tang, 2013). Global warming will have far-reaching consequences in the near future and therefore cannot be ignored. These include effects on the climate, environment, economic growth and food security (Barbir *et al*., 1990). In a study by Schmidhuber and Tubiello (2007), it was revealed that climate change will have a substantial effect on food security. Droughts will lead to a significant decline in crop yields and livestock. For example, Cooper *et al*. (2008) stated that an estimated 40% of Sub-Saharan farmlands in Africa will not be suitable for farming by the year 2030 as a result of environmental effects such as heat, drought and floods which will contribute to the deterioration in crop yields and livestock production. In order to meet global energy demand without adverse environmental impacts for future generations, energy sources with low carbon emissions should be sought (Asif and Muneer, 2007).

The annual BP Energy Outlook 2035 Report (2015b) projected that world population will reach 8.7 billion by the year 2030, which means an additional 1.6 billion people will require energy.

Renewable energy based technologies have emerged as potential replacements for traditional fossil fuel sources (Levin *et al.*, 2004). These include solar, wind, hydro and geothermal power (Shockey *et al*., 2010). Renewable biofuels such as biohydrogen, biogas, bioethanol, biodiesel (Naik *et al*., 2010) and fuel cell technologies such as Microbial Fuel Cells (MFC) and Microbial Electrolysis Cells (MEC) (Zhou *et al.*, 2012) have shown to be valuable as alternative energy sources. Extensive research is currently focused on the improvement of the low yields observed for these biofuels for commercialization (Levin *et al.*, 2004; Nath and Das, 2011).

1.2. Hydrogen as an alternative

Despite the competitiveness of crude oil due to its low cost compared to biofuels such as biohydrogen, its non-renewable nature in addition to its environmental impact make it unattractive for continuous use in future years. Hydrogen is viewed as an excellent replacement for current energy sources. This is due to its high gravimetric energy density of 122 kJ/g which is approximately 2.9 times higher than petroleum (44 kJ/g), gas (52 kJ/g), coal (40 kJ/g), methane (50.1 kJ/g) and ethanol (26.5 kJ/g). Moreover, the combustion of this fuel results in water as the only by-product (Belafi-Bakó *et al*., 2010). Hydrogen possesses properties that make it compatible with energy technologies such as fuel cells, engines and combustion turbines (Caglar and Ozmen, 2000). The United States Department of Energy (USDOE, 2004) reported that the total contribution of hydrogen to the energy market will reach an estimated 6-10% by the year 2025. Several developed countries have acknowledged the fundamental role of hydrogen as a fuel. It is thus imperative to strive towards a hydrogen-based economy (Turner, 2004).

Hydrogen production may be carried out by both biological and non-biological methods. Non-biological methods for hydrogen generation include electrolysis of water as well as steam reformation of methane, but these however, are expensive and energy intensive (Antoni *et al*., 2007). Biological methods include dark fermentation, photo-fermentation and microbial electrolysis. While photo-fermentation is dependent on light, dark fermentation offers a light independent process with higher yields than photo-fermentation (Das and Veziroglu, 2001).

On the other hand, microbial electrolysis has recently emerged as a method for biohydrogen production and is viewed as a remarkable method for high hydrogen yields by overcoming some of the challenges encountered during dark fermentation (Cheng and Logan, 2007).

In comparison to other technologies, fermentative biohydrogen production via the dark fermentation process is more appealing owing to its high hydrogen production rate, use of low-cost renewable substrates and its low technical requirements (Kotay and Das, 2008; Nandi and Sengputa, 1998; Hawkes *et al*., 2002). Furthermore, the simultaneous reduction of environmental pollutants with combined energy production make it ideal for future use (Van Ginkel and Logan, 2005; Levin *et al*., 2004) . This process has attracted much interest in recent years with government-supported initiatives reaching more than 30 countries thus far. The implementation of hydrogen as an alternative energy source has prompted over 400 projects worldwide. These initiatives are part of a global effort to upsurge energy security, environmental protection, and economic success by means of the industrialization and subsequent commercialization of hydrogen (EIA, 2011). Dark fermentative hydrogen production entails the use of microorganisms under anaerobic conditions to break-down organic matter which results in the production of hydrogen, organic acids (acetic and butyric acid) and alcohols (ethanol and butanol) (Hallenbeck, 2009; Nath and Das, 2011). Additionally, novel technologies such as Microbial Electrolysis Cells (MECs) may potentially overcome some of the challenges encountered during dark fermentation. MECs are based on the commonly known microbial fuel cells (MFCs). MFCs produce electricity from the microbial break-down of organic matter whereas MECs use bacterial metabolism along with the application of a low electric voltage for the production of hydrogen (Logan and Regan, 2006). However the industrialization of this process still faces significant scale up challenges.

Effect of key input parameters on dark fermentative biohydrogen production

Various factors influence the hydrogen production process such as: inoculum type and concentration, substrate type and concentration, pH and temperature (Wang and Wan, 2009a; Wang and Wan, 2009b; Wang and Wan, 2009c). These factors impact on the microbial composition, metabolic fluxes and thus the quantity of hydrogen produced (Wang and Wan, 2009b; Elsharnouby *et al.,* 2013). Reports on the impact of pH on biohydrogen production have revealed that values below 4.5 tend to inhibit the hydrogenase activity and will therefore impact the overall yield (Fang and Liu, 2002; Hawkes *et al.,* 2002; Khanal *et al.,* 2004). With

regards to inoculum type, both pure and mixed culture systems have been used for hydrogen production (Elsharnouby *et al.,* 2013). Studies have indicated that the inoculum concentration, source and microbial community structure influence the hydrogen production process. Available studies on this parameter have indicated that relatively low inoculum concentrations (< 10%) lead to a decrease in the cumulative hydrogen volume (Kotay and Das, 2006; Wang and Jin, 2009; Bakonyi *et al.*, 2011; Veena *et al.*, 2012).

At the laboratory level various feedstocks such as glucose, sucrose and xylose have been experimented with for biohydrogen production (Wang and Wan, 2009c; Mu *et al*., 2006a; Mu *et al*., 2006b; Mu *et al.*, 2009). The synergistic interactions in mixed microbial consortia allow simultaneous substrate degradation and biohydrogen production on complex substrates (Sarkar *et al*., 2013). This is beneficial when resistant lignocellulosic biomass, mainly consisting of xylose, lignin and cellulose is used. Recently, the search for renewable and sustainable substrates for biohydrogen production has gained much attention. These include lignocellulosic biomass, industrial wastes and rich carbohydrates that are produced in large quantities by the sugar refining industry such as molasses (Mafuleka and Gueguim-Kana, 2015; Sekoai and Gueguim-Kana, 2013; Whiteman and Gueguim-Kana, 2014). Despite the advances in biohydrogen process development, its commercialization has been impeded by low yields. Thus, further optimizations through the development of more accurate process models and subsequent scale up on low cost substrates are required.

Biohydrogen process modelling and optimization

Several modelling algorithms have been used for biohydrogen process development. These include One Variable at a time (OVAT), factorial design of Experiment (DOE), Response Surface Methodology (RSM) and Artificial Neural Networks (ANN) (Sekoai and Gueguim Kana, 2013; Venkata-Mohan *et al*., 2009; Nasr *et al*., 2013a; Nasr *et al*., 2013b; Wang and Wan, 2009a; Wang and Wan, 2009c). Although, OVAT has been widely used, it ignores the interactive effects of input parameters on the process output and is not practical to reach a suitable optimum in a low number of experiments (Lotfy *et al*., 2007). Alternatively, the factorial design of experiment (DOE) is tedious, resource-intensive and laborious when the quantity of input parameters is increased (Gueguim-Kana *et al*., 2012). RSM employs polynomial regression analysis to generate a second-order model equation that is used to relate the input parameters to the output. The model equation is used to determine the optimum process setpoints (Mandenius and Brundin, 2008). These models assume that the

polynomial equations can accurately estimate the fermentation dynamics. Nevertheless, RSM disregards the "less important" parameters with a limited understanding of their possible interactive effects on the bioprocess output (Gueguim-Kana *et al.*, 2012).

Artificial Neural Networks (ANN) are a mathematical illustration of the human nervous system. They simulate the learning process of the human brain by mathematically modelling the network structure of interconnected nerve cells (Gueguim-Kana *et al.*, 2012). These systems are totally data-driven and studies existing relationships between input and output parameters in an attempt to identify the effects that govern the process output. One of the most frequently adopted architectures is the multi-layer perceptron (MLP) which is made up of an input layer, one or more hidden layers and the output layer (Gueguim-Kana *et al.*, 2012). These layers comprise neurons, the number of which may differ depending on the intricacy of the process it is being applied to (Whiteman and Gueguim-Kana, 2014). Numerous studies have indicated that ANN is able to abstract relationships from small data sizes, though the data must be statistically well distributed in the input domain (Whiteman and Gueguim-Kana, 2014; Gueguim-Kana *et al*., 2012; Wang and Wan, 2009a; Wang and Wan, 2009c).

1.3 Research Motivation

A hydrogen-based economy has been impeded by high production costs and low yields. This requires further modelling and optimization at lab scale with subsequent scale up. The development of accurate and reliable bioprocess models is imperative for process optimization. Additionally, there is a lack of consensus on the appropriate fermentation volume size for fermentation screening, modelling and optimization at the early stages of process development. Various studies have used a process volume in the range of 100-200 mL for modelling biohydrogen production (Whiteman and Gueguim-Kana, 2014; Sekoai and Gueguim-Kana, 2013; Wang *et al.*, 2005; Wang and Wan, 2009a; Wang and Wan, 2009c; Faloye *et al*., 2013; Faloye *et al*., 2014) and to a lesser extent between 1-6 L (Rosales-Colunga *et al*., 2010; Prakasham *et al*., 2011; Shi *et al*., 2010; Mullai *et al*., 2013). Biohydrogen process development requires extensive knowledge at the lab scale level for efficient scale up (Escamilla-Alvarado *et al*., 2012). Bioprocess modelling inaccuracies introduced at the lab scale have significantly impeded the scale up phase (Schmidt, 2005).

Efforts to overcome these challenges include modelling and optimization of the key input parameters across bioprocess scales.

Conversely, large variations exist between the reported optimum set points of input parameters for fermentative hydrogen production (Wang *et al*., 2009; Wang and Wan, 2008; Wang and Wan, 2009a; Wang and Wan, 2009c; Mu *et al*., 2006a; Mu *et al*., 2009).

In addition, inconsistencies in the reported biohydrogen yield expression units have hampered the process development of hydrogen production. Relating the influence of key input parameters on the corresponding hydrogen yield using a standardized yield expression unit will contribute towards improving the hydrogen development phase. Furthermore, despite the availability of various reports on the influence of key input parameters on biohydrogen production, there is a dearth of knowledge on intelligent models built on pre-existing information which can efficiently predict the hydrogen response on unknown input patterns. The implementation of accurate and reliable process models is necessary for the determination of the optimal set points for biohydrogen production. Thus, the development of efficient ANN models to predict on unknown parameters will contribute significantly towards reducing the biohydrogen development via both dark fermentation and microbial electrolysis.

1.4. Aims

This work aims at developing Artificial Neural Network based process models for biohydrogen production via dark fermentation and microbial electrolysis. Furthermore, it investigates the impact of experimental process volume size on the efficiency of ANN and RSM based process models.

In order to achieve this aim, the following specific objectives were undertaken.

(i) A review of the application of Artificial Neural Networks as a tool for modelling and optimization of biofuel production was carried out.

(ii) Bioprocess models were developed using ANN and RSM for optimization of hydrogen response on operational parameters of inoculum size, substrate concentration and Hydraulic Retention Time (HRT) across two scales (80 and 800 mL).

(iii) Subsequently, a preliminary assessment of biohydrogen production at a semi-pilot scale (8 L) under the optimized conditions determined by the developed models was carried out. The microbial community involved in the hydrogen production process was examined.

(iv) Two ANN based models were developed for prediction of fermentative hydrogen production (mol H_2/ mol substrate and mL H_2/ g substrate) on inputs of inoculum type, substrate type, substrate concentration, pH and temperature.

(v) A committee of Artificial Neural Network models was developed for prediction of hydrogen production (mol H_2/ mol substrate) from Microbial Electrolysis Cells (MEC). The input parameters consisted of substrate type, substrate concentration, pH, temperature, applied voltage and MEC reactor configuration.

1.5. Outline of dissertation/thesis structure

This thesis includes six chapters and conforms to the "research paper format" as outlined in the dissertation/thesis template by the College of Agriculture, Engineering and Science (AES) of the University of KwaZulu-Natal. A literature review of the efficiency of Artificial Neural Networks as a tool for modelling and optimization of biofuel production with emphasis on dark fermentative hydrogen production is presented in Chapter 2.

Chapter 3 focuses on the modelling and optimization of hydrogen response on operational setpoint parameters of inoculum size, sugarcane molasses concentration and Hydraulic Retention Time (HRT) across two bioprocess scales (80 and 800 mL). Response Surface Methodology (RSM) and Artificial Neural Networks (ANN) were used at both scales. A preliminary assessment of biohydrogen production using the optimized set points was carried out at semi-pilot scales. The microbial community involved in this bioprocess was examined.

In Chapter 4, two Artificial Neural Network (ANN) models were developed for the prediction of fermentative hydrogen production using two yield expression units (mol H_2/ mol substrate and mL H_2/ g substrate). Input parameters considered were inoculum type, substrate type, substrate concentration, pH and temperature.

Chapter 5 focuses on the development of a committee of Artificial Neural Networks (ANN) models for prediction of hydrogen production (mol H_2/ mol substrate) from Microbial Electrolysis Cells (MECs). Inputs parameters consisted of substrate type, substrate concentration, pH, temperature, applied voltage and MEC reactor configuration.

The final chapter, Chapter 6, states major conclusions derived from this study, integrates the work and provides recommendations for future research.

References

Antoni D, Zverlov V, Schwarz W. 2007. Biofuels from microbes. Applied Microbiology and Biotechnology 77: 23–35.

Asif M, Muneer T. 2007. Energy supply, its demand and security issues for developed and emerging economies. Renewable and Sustainable Energy Reviews 11: 1388–1413.

Bakonyi P, Nemestóthy, N, Lövitusz E, Bélafi-Bakó, K. 2011. Application of Plackett–Burman experimental design to optimize biohydrogen fermentation by *E. coli* (XL1-BLUE). International Journal of Hydrogen Energy 36: 13949–13954.

Barbir F, Veziroglu TN, Plass HJ. 1990. Environmental damage due to fossil fuels use. International Journal of Hydrogen Energy 10: 739.

Bélafi-Bakó K, Bakonyi P, Nemestóthy N, Pientka Z. 2010. Biohydrogen production in integrated system. Desalination and Water Treatment 14: 116–118.

BP. 2015a. London, United Kingdom, June 2015. Available at https://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2015/bp statistical-review-of-world-energy-2015-full-report.pdf. Accessed 2nd December 2015.

BP. 2015b. London, United Kingdom. Available at http://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2015/bp-energy-outlook-2035-booklet.pdf. Accessed 2nd December 2015.

Caglar A, Ozmen H. 2000. Hydrogen as an attractive fuel in future. Energy Education Science and Technology 6: 1–18.

Cheng SA, Logan BE. 2007. Sustainable and efficient biohydrogen production via electrohydrogenesis, Proceedings of the National Academy of Sciences of the United States of America 104(47): 18871–18873.

Cooper P, Dimes J, Rao K, Shapiro B, Shiferaw B, Twomlow S. 2008. Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: an essential first step in adapting to future climate change. Agriculture, Ecosystems and Environment 126: 24-35.

Das D, Veziroglu TN. 2001. Hydrogen production by biological processes: a survey of literature. International Journal of Hydrogen Energy 26: 13–28.

EIA. 2011. Annual Energy Outlook. 2011, with Projections to 2035. Energy Information Administration. Washington, DC, USA. Available at http://www.eia.gov oiaf aeo aeoref tab.html. Accessed 2nd December 2015.

Elsharnouby O, Hafez H, Nakhla G, El Naggar MH. (2013). A critical literature review on biohydrogen production by pure cultures. International Journal of Hydrogen Energy 38: 4945–4966.

Escamilla-Alvarado C, Rios-Leal E, Ponce-Noyola MT, Poggi-Varaldo HM. 2012. Gas biofuels from solid substrate hydrogenic–methanogenic fermentation of the organic fraction of solid municipal wastes. Process Biochemistry 47(11): 1572–1587.

Faloye FD, Gueguim-Kana EB, Schmidt S. 2014. Optimization of biohydrogen inoculum development via a hybrid pH and microwave treatment technique-Semi pilot scale production assessment. International Journal of Hydrogen Energy 39: 5607–5616.

Faloye FD, Gueguim-Kana EB, Schmidt S. 2013. Optimization of hybrid inoculum development techniques for biohydrogen production and preliminary scale up. International Journal of Hydrogen Energy 38: 11765–11773.

Fang HH, Liu H. 2002. Effect of pH on hydrogen production from glucose by a mixed culture. Bioresource Technology 82: 87–93.

Gueguim-Kana EB, Oloke JK, Lateef A, Adesiyan MO. 2012. Modelling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renewable Energy 46: 276–281.

Hallenbeck PC. 2009. Fermentative hydrogen production: principles, progress, and prognosis. International Journal of Hydrogen Energy 34: 7379–7389.

Hawkes FR, Dinsdale R, Hawkes DL, Hussy I. 2002. Sustainable fermentative hydrogen production: challenges for process optimization. International Journal of Hydrogen Energy 27: 1339–1347.

Hook M, Tang X. 2013. Depletion of fossil fuels and anthropogenic climate change –A review. Energy Policy 52: 797–809.

IPCC. 2014. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, New York, NY, USA. Available at http://www.ipcc.ch/report/ar5/wg3/. Accessed 2nd December 2015.

Khanal SK, Chen WH, Li L, Sung S. 2004. Biological hydrogen production: effects of pH and intermediate products. International Journal of Hydrogen Energy 29: 1123–1131.

Kotay SM, Das D. 2008. Biohydrogen as a renewable energy resource: prospects and potentials. International Journal of Hydrogen Energy 33: 258–263.

Kotay SM, Das D. 2006. Microbial hydrogen production with Bacillus coagulans IIT-BT S1 isolated from anaerobic sewage sludge. Bioresource Technology 98(6): 1183–1190.

Levin DB, Pitt L, Love M. 2004. Biohydrogen production: prospects and limitations to practical application. International Journal of Hydrogen Energy 29: 173–185.

Li R. (2007). Biohydrogen production by photo fermentation from wastewater. PhD Thesis, University of Hong Kong, Pokfulam, China.

Logan BE, Regan JM. 2006. Electricity-producing bacterial communities in microbial fuel cells. Trends in Microbiology 14: 512–518.

Lotfy AW, Khaled M, Ehab R et al. 2007. Citric acid production by a novel *Aspergillus niger* isolate: II. Optimization of process parameters through statistical experimental designs. Bioresource Technology 98: 3470–3477.

Mafuleka S, Gueguim-Kana EB. 2015. Modelling and optimization of xylose and glucose production from napier grass using hybrid pre-treatment techniques. Biomass and Bioenergy 77: 200–208.

Mandenius C, Brundin A. 2008. Review: biocatalysts and bioreactor design–bioprocess optimization using design-of-experiments methodology. Biotechnology Progress 24: 1191–1203.

Mecad. 2013. Region Economic Outlook Update. Middle East and North Africa: Defining the road ahead. Middle East and Central Asia Department. Available at http://www.imf.org/external/pubs/ft/reo/2013/mcd/pdf. Accessed 2nd December 2015.

Mu Y, Wang G, Yu HQ. 2006a. Response surface methodological analysis on biohydrogen production by enriched anaerobic cultures. Enzyme and Microbial Technology 38: 905–913.

Mu Y, Zheng XJ, Yu HQ, Zhu RF. 2006b. Biological hydrogen production by anaerobic sludge at various temperatures. International Journal of Hydrogen Energy 31: 780–785.

Mu Y, Zheng XJ, Yu HQ. 2009. Determining optimum conditions for hydrogen production from glucose by an anaerobic culture using response surface methodology (RSM). International Journal of Hydrogen Energy 34: 7959–7963.

Mullai MK, Yogeswari K, Sridevi K, Ross PR. 2013. Artificial neural network (ANN) modeling for hydrogen production in continuous anaerobic sludge blanket filter (ASBF). Singaporean Journal of Scientific Research 5(1): 1–7.

Naik SN, Goud VV, Rout PK, Dalai AK. 2010. Production of first and second generation biofuels: A comprehensive review. Renewable and Sustainable Energy Reviews 14: 578–597.

Nandi R, Sengupta S. 1998. Microbial production of hydrogen: an overview. Critical Reviews in Microbiology 24: 61–84.

Nasr N, Hafez H, El Naggar MH, Nakhla G. 2013a. Application of artificial neural networks for biohydrogen production. International Journal of Hydrogen Energy 38: 3189–3195.

Nasr N, Tawfik A, Ookawara1 S, Suzuki M. 2013b. Prediction of hydrogen production using Artificial Neural Network. 17th International Water Technology Conference. 5-7 November, Istanbul, Turkey.

Nath K, Das D. 2011. Modeling and optimization of fermentative hydrogen production. Bioresource Technology 102: 8569–8581.

Prakasham RS, Sathish T, Brahmaiah, P. 2011. Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. International Journal of Hydrogen Energy 36: 4332–4339.

Rosales-Colunga LM, Garcia, RG, Rodriguez AD. 2010. Estimation of hydrogen production in genetically modified *E. coli* fermentations using an artificial neural network. International Journal of Hydrogen Energy 35: 13186–13192.

Sarkar O, Kannaiah-Goud R, Venkata-Subhash G, Venkata-Mohan S. 2013. Relative effect of different inorganic acids on selective enrichment of acidogenic biocatalyst for fermentative biohydrogen production from wastewater. Bioresource Technology 147: 321–331.

Schmidhuber J, Tubiello FN. 2007. Global food security under climate change. Proceedings of the National Academy of Sciences. 11 December, Rome, Italy. 104: 19703–19708.

Schmidt FR. 2005. Optimization and scale up of industrial fermentation processes. Applied Microbiology and Biotechnology 68: 425–435.

Sekoai PT, Gueguim-Kana EB. 2013. A two-stage modelling and optimization of biohydrogen production from a mixture of agro-municipal waste. International Journal of Hydrogen Energy. 38: 8657–8663.

Shi Y, Gai G, Zhao X et al. 2010. Back propagation neural network (BPNN) simulation model and influence of operational parameters on hydrogen bio-production through integrative biological reactor (IBR) treating wastewater. Bioinformatics and Biomedical Engineering (iCBBE), 4th International Conference, Chengdu (China)

Shockey I, Kinicki R, Delorey J, Arruda K, Baldwin S, Quinn A. 2010. Feasibility study of alternative energy sources and conservation techniques for implementation in El Yunque National Forest. Available at http://web.cs.wpi.edu/~rek/Projects/USF_Proposal. Accessed 2 December 2015.

Smith KR, Jerrett M, Anderson HR, Burnett RT, Stone V, Derwent R, Atkinson RW, Cohen A, Shonkoff SB, Krewski D, Pope CA, Thun MJ, Thurston G. 2009. Public health benefits of strategies to reduce greenhouse-gas emissions: health implications of short-lived greenhouse pollutants. Health and Climate Change 374: 2091–2103.

Sorrell S, Lehtonen M, Stapleton L, Pujol J, Champion T. 2009. Decomposing road freight energy use in the United Kingdom. Energy Policy 37: 3115–3129.

Turner JA. 2004. Sustainable hydrogen production. Science 305: 972–974.

USDOE. 2004. United States Annual Energy Report. United States Department of Energy. Washington DC.USA. Available at http://www.energy.gov/pdf. Accessed 2nd December 2015.

Van Ginkel SW, Logan BE. 2005. Increased biological hydrogen production with reduced organic loading. Water Research 39: 3819–3826.

Veena T, Tiwari KL, Quraishi A, Jadhav SK. 2012. Biohydrogen production from rice mill effluent. Journal of Applied Science and Environmental Sanitation 7(4): 237–240.

Venkata-Mohan S, Raghavulu SV, Mohanakrishna G, Srikanth S, Sarma PN. 2009. Optimization and evaluation of fermentative hydrogen production and wastewater treatment processes using data enveloping analysis (DEA) and Taguchi design of experimental (DOE) methodology. International Journal of Hydrogen Energy 34(1): 216–226.

Wang B, Wan W, Wang JL. 2009. Effect of ammonia concentration on fermentative hydrogen production by mixed cultures. Bioresource Technology 100: 1211–1213.

Wang G, Mu Y, Yu HQ. 2005. Response surface analysis to evaluate the influence of pH, temperature and substrate concentration on the acidogenesis of sucrose-rich wastewater. Biochemical Engineering Journal 23: 175–184.

Wang JL, Wan W. 2008. Optimization of fermentative hydrogen production process by response surface methodology. International Journal of Hydrogen Energy 33: 6976–6984.

Wang JL, Wan W. 2009a. Application of desirability function based on neural network for optimizing biohydrogen production process. International Journal of Hydrogen Energy 34: 1253–1259.

Wang JL, Wan W. 2009b. Factors influencing fermentative hydrogen production: A review. International Journal of Hydrogen Energy. 34: 799–811.

Wang JL, Wan W. 2009c. Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology. International Journal of Hydrogen Energy 34: 255–261.

Wang X, Jin B. 2009. Process optimization of biological hydrogen production from molasses by a newly isolated *Clostridium butyricum* W5. Journal of Bioscience and Bioengineering. 107 (2): 138–144.

Whiteman JK, Gueguim-Kana EB. 2014. Comparative Assessment of the Artificial Neural Network and Response Surface Modelling Efficiencies for Biohydrogen Production on Sugar Cane Molasses. Bioenergy Research 7(1): 295–305.

Zhou M, Wang H, Hassett DJ, Gu T. 2012. Recent advances in microbial fuel cells (MFCs) and microbial electrolysis cells (MECs) for wastewater treatment, bioenergy and bioproducts. Journal of Chemical Technology and Biotechnology 88: 508–518.

CHAPTER 2

Artificial Neural Networks: An efficient tool for Modelling and Optimization of Biofuel production

This chapter has been submitted to Biofuel Research Journal with the title: Artificial Neural Networks: An efficient tool for Modelling and Optimization of Biofuel production.

The manuscript is presented in the following pages.

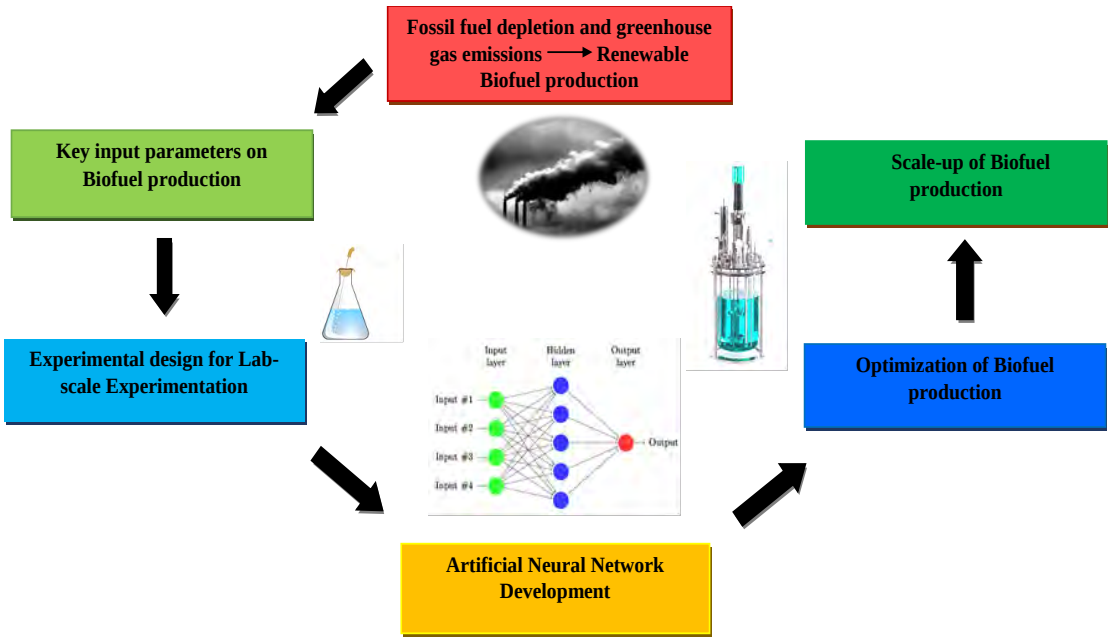
Artificial Neural Networks: An efficient tool for Modelling and Optimization of Biofuel production

Yeshona Sewsynker, Funmilayo Faloye, Evariste Bosco Gueguim-Kana
Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209
Pietermaritzburg, South Africa
Correspondence: kanag@ukzn.ac.za
Phone: +27332605527

Highlights

- Artificial Neural Networks (ANNs) have emerged as a tool for modelling complex non-linear bioprocesses.
- The efficiency of ANN to capture the non-linear interactions in biofuel production.
- Application of Artificial Neural Networks in various biofuels optimization was reviewed.

Graphical Abstract



Abstract

In view of the looming energy crisis as a result of depleting fossil fuel resources and environmental concerns from greenhouse gas emissions, the need for sustainable energy sources have secured global attention. Research is currently focused on renewable sources of energy and biofuel due to availability and their environmental friendliness. Biofuel production like other bioprocesses is controlled by several process parameters including pH, temperature and substrate concentration. However, the improvement of biofuel production requires a robust process model that accurately relates the effect of input variables on the process output. Artificial Neural Networks (ANNs) have emerged as a tool for modelling complex, non-linear processes. ANNs are applied in the prediction of various process outcomes and its use in biofuel production is currently in the early phase of development. This review highlights the efficiency of Artificial Neural Networks (ANNs) as a tool for modelling and optimization of biofuel production with emphasis on dark fermentative hydrogen production and its potential for future application. Recent findings on the application of ANN for the optimization of biohydrogen, biogas, biodiesel, microbial fuel cell technology and bioethanol were reviewed. In addition, comparative studies on ANN and other modelling techniques such as the Response Surface Methodology (RSM) on the optimization of biofuel production were evaluated.

Keywords: Modelling, Optimization, Biofuel production, Artificial Neural Networks, Bioprocess

1. Introduction

Bioprocesses are described as biological systems that are non-linear, complex and unsteady, thus presenting challenges in developing a precise physical-based formula to characterize its physical performance. In addition, the development of accurate bioprocess models continue to baffle experts as a result of the non-linear nature of the biochemical network interactions that occur during fermentation processes (Franco-Lara *et al*., 2006). Bioprocesses are influenced by several parameters which include pH, temperature, hydraulic retention time, and substrate concentration. The determination of the optimum setpoints of these parameters is therefore crucial for bioprocess development and scale up (Wang and Wan, 2009a; Wang and Wan, 2009b; Wang and Wan, 2009c).

Mathematical and statistical based models can provide vital information for the understanding, analysis and prediction of biological processes (Nath and Das, 2011) and they are required for the optimization of the key parameters in order to improve the process output (Escamilla-Alvarado *et al*., 2012). These bioprocess models can provide insight into the individual as well as the interactive effect of the various input parameters on the target output. Nevertheless, the non-linearities associated with microbial fermentations have limited the use of these bioprocess models. Non-linear systems, as opposed to linear systems, are not standardized which results in deviations between the results obtained. The implementation of bioprocess models that are able to efficiently encapsulate these non-linearities are of paramount importance for optimization and scale up of the bioprocess (Almeida, 2002).

Biofuel production has emerged as a promising alternative to fossil fuel sources (Levin *et al.*, 2004), the development of which may help overcome the current energy crisis and also provide a clean source of energy to combat the phenomenon of global warming (Levin *et al.*, 2004; Nath and Das, 2011). Current biofuels include bioethanol, biodiesel, biohydrogen, biogas (Naik *et al*., 2010) and fuel cell technologies such as Microbial Fuel Cells (MFC) and Microbial Electrolysis Cells (MEC) (Zhou *et al.*, 2012). The major limitation of these biofuels may be attributed to their low yields and production rates observed (Nath and Das, 2011).

Modelling and optimization of biofuel production processes will contribute to increased understanding of the process inputs for optimum yield and production rate. The main goal of modelling is to optimize the processes involved in producing these biofuels in order to improve the yields. Various modelling algorithms have been applied in biofuel production

processes (Nath and Das, 2011; Wang and Wan, 2009a; Wang and Wan, 2009b; Gueguim-Kana *et al*., 2012a; Abu-Qdais *et al.*, 2010; Saraphirom and Reungsang, 2010; Mohamed *et al.*, 2013; Sewsynker *et al*., 2015), and results have shown that modelling and optimization can enhance biofuel yields (Ghosh *et al*., 2012, Gueguim-Kana *et al*., 2012a).

For instance, Ghosh *et al*. (2012) used the Response Surface Methodology (RSM) to optimize biohydrogen production on inputs of glucose concentration, fixed nitrogen, and light intensity in a single-stage photo-fermentation with the photosynthetic bacterium *Rhodobacter capsulatus*. Their results showed that these parameters had a significant interactive effect on the biohydrogen yield and nitrogenase activity. The optimized biohydrogen yield (5.5 mol H_2/mol glucose) was 85% higher than previously achieved.

Traditionally, modelling and optimization of bioprocesses has been carried out using the One Variable At a Time approach (OVAT), factorial Design Of Experiment (DOE) and Response Surface Methodology (RSM) (Nath and Das, 2011; Kyazze *et al*., 2010; Selembo *et al.*, 2009; Lu *et al*., 2011; Tartakovsky *et al*., 2009; Guo *et al.*, 2010; Yahya *et al.*, 2013, Cheng *et al.*, 2011; Whiteman and Gueguim-Kana, 2014). These approaches have been extensively used and their concepts as well as limitations are well known. For example, OVAT does not consider the interactive effect of parameters on the process and therefore the optimum setpoints may be completely ignored (Wang and Lu, 2005; Gueguim-Kana *et al*., 2012a). Moreover, it is unfeasible for the search to accomplish an appropriate optimum in a restricted amount of experimental setups (Lotfy *et al*., 2007). The factorial Design of Experiment (DOE) has shown to be unappealing since it is time-consuming, resource demanding and labour intensive when the numbers of input factors are increased (Wang and Wan, 2009c). On the other hand, the Response Surface Methodology (RSM) disregards the "less important" parameters with a limited understanding of their possible interactive effects on the bioprocess output (Desai *et al*., 2008; Gueguim-Kana *et al.*, 2012a).

Artificial Intelligence (AI) tools have emerged as a promising method for the modelling and optimization of bioprocesses. Some of these include Artificial Neural Networks (ANN) and Genetic Algorithm (GA) (Zhang *et al*., 2010; Prakasham *et al*., 2011; Abu-Qdais *et al.*, 2010), Fuzzy Logic (FL), Ant Algorithm (AA) and Particle Swarm Optimization (PSO) all of which are considered suitable for the design of bioprocesses for research and development (Haider *et al*., 2008; Garlapati and Banerjee, 2010). In the last decade, ANN has been applied in

multivariate non-linear bioprocess research and development. They are efficient for the development of bioprocess models devoid of previous information regarding the kinetics and metabolic fluxes that occur within the cells and cell surroundings (Gueguim-Kana *et al.*, 2012a). ANN models simulate the linkage that exists in biological neurons with extraordinary capability for learning, analysis, association and adaptation (Nagata and Chu, 2003).

ANNs can be described as a mathematical understanding of the neurological functioning of the human brain. They emulate the brain's learning process by arithmetically modelling the network structure of interconnected nerve cells (Nagata and Chu, 2003). Furthermore, ANNs are entirely data-based with no previous knowledge of the events that govern the process (Shi *et al*., 2010). They consist of an input layer, one or more hidden layers, and an output layer (shown in Figure 1). The neurons of the hidden layer assist the network in establishing the complex associations that subsist between the input and output parameters (Nagata and Chu, 2003).

The appeal of ANNs as a modelling tool stems from their extraordinary information processing features which are attributed primarily to non-linearity, high parallelism, fault and noise acceptance, as well as their learning and generalization abilities. In contrast to traditional modelling tools, ANNs offer a model-free, adaptive, parallel-processing, and vigorous elucidation with error and failure tolerance. Moreover, its learning capability for processing inaccurate and fuzzy information and its ability to generalize unseen patterns is impeccable (Levstek and Lakota, 2010). ANN possesses the ability to sketch process input and outputs devoid of causal assumption regarding the division of data. ANNs have gained much attention as significant soft computing tools not limited to data processing and analysis but can also be applied to solve difficulties in multifaceted and non-linear processes (Almeida, 2002).

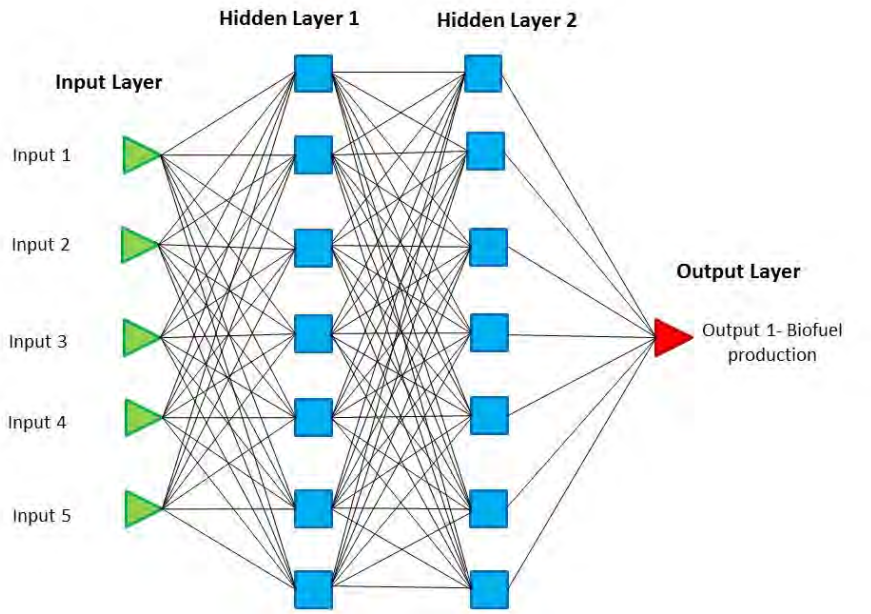


Figure 1: General Topology of a multilayer structure of an Artificial Neural Network

The rapid development of algorithms and information technology is the major motivation behind the broad application of ANNs in research and development (Huang *et al*., 2007). Currently, ANNs are employed in the prediction of various outcomes including process control, medicine, forensic science, biotechnology, weather forecasting, finance and investment and food science. However, it is noteworthy to state that the use of ANNs in biofuel production is currently in the early phases of development. This review therefore highlights the efficiency of Artificial Neural Networks (ANNs) as a tool for the modelling and optimization of biofuel production. This paper summarizes various studies on the application of ANN in biofuel production including biohydrogen, biogas, microbial fuel cell technology and bioethanol. The biohydrogen production process is discussed in detail in terms of the production process, the effects of process parameters and challenges associated with its modelling and optimization. In addition, the comparison of ANN to commonly used modelling techniques such as response surface methodology (RSM) for biofuel production is also highlighted.

2. Principles of Artificial Neural Networks

ANN involves the interconnection of a structure known as artificial neurons similar to biological neurons (Levstek and Lakota, 2010; Graupe, 2007). The principle behind ANNs is

to mimic the functioning and learning process of the human brain using an artificial neuron. An artificial neuron is a computational model that is inspired by biological neurons. Biological neurons consist of dendrites, soma, axon and synapses. The dendrites are used for receiving signals from other neurons and can also be referred to as chemical receptors. Additionally, the soma makes up the cell body of a neuron and is involved in processing the input signals. This is followed by the emission of the processed signals to neurons that are in close proximity to the axon. Finally, the neurons are linked via the synapses which also control the transmission of signals among the neurons. The actual structure and functioning of a biological neuron is far more intricate as compared to the simple design of an artificial neuron (Huang *et al*., 2007, Levstek and Lakota, 2010).

An artificial neural network is composed of groups of interconnected processing elements known as neurons and the links between these neurons are known as weights and biases (Gurney, 1997). Furthermore, in contrast to a biological neuron, an artificial neuron receives a sequence of input information (*xi*) linked to a weight factor (*wi*-). Basically, the neuron adds the weighed inputs and forwards the outcome to a transfer function to produce an output. The output information is thereafter transmitted to an alternate neuron as an input or may be employed directly as a network result. The weights are referred to as the attachment strength linking the neurons. As a result of some input signals being more significant compared to others, the utilization of weights as equivalent to the significance of each input signal provides a well-organized process to create an ideal output. Weights are changeable for the duration of network training and there are various algorithms available for the adjustment of weights during network training (Graupe, 2007).

The network architecture or topology refers to the pattern of interconnections among the neurons that makes up a network (Marchitan *et al*., 2010). Artificial neurons develop layers with different types of connections between them i.e. a neuron of one layer can be linked with neurons of at least one other layer. There are different types of connections used between layers and these are referred to as inter-layer connections. With regard to inter-layer connections, a neuron in one layer is linked with all the neurons in the subsequent layer, thus resulting in a completely connected network. However, if the neurons are connected to only some of the neurons in the next layer then the network is only partially connected. Usually, neurons in one layer send output information to the next layer, and they may (feedback

networks) or may not obtain information back from the next layer. Also, these neurons may or may not be linked with each other in the same layer (Huang *et al*., 2007).

Alternatively, in more complex structures the neurons communicate among themselves within a layer called intra-layer connections. Regarding these intra-layer connections, once the input information has been obtained from the previous layer, neurons within one layer converse with each other several times prior to transmitting their output to another layer (Yu *et al*., 2007). ANNs are occasionally referred to as machine learning algorithms, since changing its connection weights (training) causes the network to learn the solution to a problem. The strength of connection among the neurons is stored as a weight-value for the specific connection. The system is able to learn new knowledge by adjusting these connection weights. The learning ability of an ANN is determined by its design and by the algorithmic method selected for training. This algorithm attempts to reduce the error that is computed by various methods depending on the specific technique used to adjust the connections (i.e. the learning algorithm) (Levstek and Lakota, 2010).

Generally, learning can be done by (i) supervised and (ii) unsupervised training. During supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its subsequent outputs against the desired outputs. Errors are then computed, causing the system to adjust the weights which control the network. This process is repeated over and over again as the weights are constantly adjusted. On the contrary, with unsupervised training, the network is provided with inputs but without the desired outputs. The neural network system on its own then selects what characteristics it will use to classify the input data (Bishop, 1995, Levstek and Lakota, 2010).

3. ANN types and training algorithms

ANNs are characterized according to their functions. Common ANNs described in studies include Hopfield (Hopfield and Tank, 1986), Kohonen (Zupan and Gasteiger, 1999; Huang *et al*., 2007), Recurrent (Pham, 1994), Counter propagation (Zupan and Gasteiger, 1999), Radial basis function (RBF) networks (Schalkoff, 1997) and Back propagation (Desai *et al*., 2008; Whiteman and Gueguim-Kana, 2014; Rosales-Colunga *et al.,* 2010; Nikhil *et al.*, 2008). The Back propagation neural network which employs a supervised learning process has been frequently reported in biofuel process modelling as shown in Table 1-5 and is discussed in detail.

3.1. Back Propagation Neural Networks (BPNN)

This type of network is the most extensively studied and involves the minimization of a performance function (Sadrzadeh *et al*., 2008). In general, this network is a multilayer perceptron (MLP) architecture which is mostly used to solve non-linear regression problems (Marchitan *et al*., 2010). The multilayer perceptron includes an input layer with nodes that embody the input variable to the problem, the output layer with nodes that signify the dependent variable (what is modelled), and one or more hidden layers consisting of nodes to facilitate the encapsulation of non-linearity in data. The back propagation is usually used for training of feed forward networks and has been extensively studied (Figure 2) (Almeida, 2002; Marchitan *et al*., 2010). By means of supervised learning, this network is able to learn the mapping from one data set to another by exploiting the examples. Back propagation describes the manner in which the error computed at the output side is propagated backward from the output layer to the hidden layer and lastly to the input layer. In these networks, the data are fed forward directly into the network with no feedback (Huang *et al*., 2007; Levstek and Lakota, 2010) and the neurons can be completely or partially interconnected. During training, the weight and biases are adjusted with the goal of fitting the predicted response closer to the experimental response (Hagan *et al*., 1996). BPNNs are versatile and may be employed for data modelling and process control in medicine, forensic science, biotechnology, weather forecasting, finance and investment and food science (Hassoun, 1995; Lou and Nakai, 2001; Mjalli, 2005; Messikh *et al*., 2007; Desai *et al*., 2008).

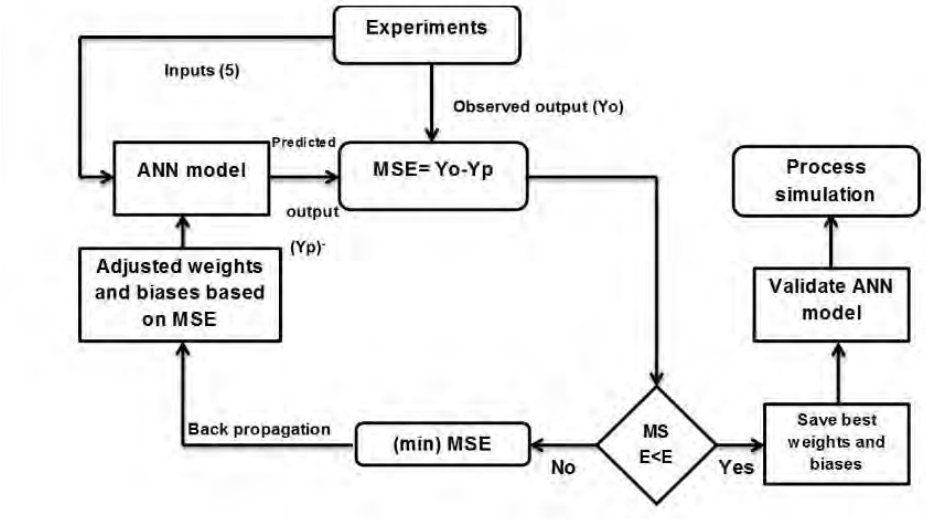


Figure 2: The Back propagation training flowchart for Artificial Neural Networks

4. Biofuel Production

4.1. Fermentative hydrogen production by dark fermentation

The dark fermentation process involves the break-down of carbohydrates by anaerobic bacteria to produce hydrogen (Hallenbeck and Ghosh, 2009). This method is viewed as a promising system for practical application in the near future. The benefits of this process over light dependent processes include cheaper process with lower energy requirements, higher hydrogen production and the application of low-value waste materials as feedstock (Levin *et al*., 2004). Although the dark fermentation process is widely accepted as a potential system to be implemented at an industrial scale, there are several limitations to its commercialization (Nath and Das, 2011). The amount of hydrogen that may be produced by the dark fermentation process is dependent on the metabolic pathways adopted by the bacteria (Hallenbeck and Ghosh, 2009). These reactions result in the production of hydrogen in addition to other products such as carbon dioxide and simple organic compounds such as volatile fatty acids (VFA). Some VFAs produced during hydrogen production include acetate and butyrate. The maximum theoretical value of hydrogen that may be produced under ideal conditions when glucose is used as the substrate is 4 moles of hydrogen per mole of glucose which occur via the acetate pathway (Hallenbeck and Ghosh, 2009). On the other hand, only 2 moles of hydrogen per mole of glucose is produced via the butyrate pathway (Vazquez *et al*., 2009). Current efforts are directed towards the search for renewable, cheap waste material that is rich in carbohydrates and can be readily utilized by the hydrogen-producing microorganisms. These include agricultural residues (Mafuleka and Gueguim-Kana, 2015) and industrial wastewater such as those from the sugar refining industry (Whiteman and Gueguim-Kana, 2014; Wang and Jin, 2009).

Biohydrogen production can be operated in either batch or continuous mode but the majority of studies have employed batch mode which is simpler to operate and has shown to be more feasible for research (Wang and Wan, 2009a; Wang and Wan, 2009c). Alternatively, various studies have reported semi-pilot scale and pilot scale reactor systems with continuous mode of production for biohydrogen production (Ren *et al*., 2006; Chang *et al*., 2011). Although both batch and continuous systems present several benefits during the developmental phase, their practical application for commercialization is limited by the low yields observed.

4.2. Metabolic pathways for dark fermentative hydrogen production

The major species involved in hydrogen production are members of the genus *Clostridia* (Nandi and Sengupta, 1998). *Clostridium* spp. are rod-shaped, Gram-positive and are endospore-forming bacteria (Holt *et al*., 1994). Their resistant endospores allow them to survive extreme conditions above or below their optimum (Holt *et al*., 1994). The production of hydrogen by these species is generally associated with two pathways (Cai *et al*., 2011). The first pathway involves the conversion of pyruvate to acetyl-CoA and CO_2 by the enzyme pyruvate ferredoxin oxidoreductase which leads to the generation of reduced ferredoxin (Fd). Hydrogen is then generated from the reduced ferredoxin by the enzyme hydrogenase. The second pathway entails re-oxidizing a part of the NADH that was produced from glycolysis by the NADH- ferredoxin oxidoreductase to generate reduced ferredoxin (Vardar-Schara *et al*., 2008), which in turn is re-oxidized by hydrogenase to produce hydrogen. *Clostridium* spp. can stoichiometrically produce 2 and 4 moles of hydrogen per mole of glucose by the butyrate and acetate pathways, respectively (Hallenbeck and Ghosh, 2009). However the practical yields achieved are lower due to formation of other fermentative by-products. Studies have reported that the butyrate pathway produces lower yields due to inhibitory effects on hydrogen production and cell growth (Chin *et al*., 2003; Berrios-Rivera *et al*., 2000). Furthermore, this pathway has been documented as the major conflicting pathway during hydrogen production since it consumes more NADH than the acetate pathway which results in a decrease in the hydrogen yield (Kumar *et al*., 2001).

4.3. Key parameters that influence the biohydrogen production process

Several factors have been shown to impact the hydrogen production process. These include: inoculum type, substrate type and concentration, temperature, and pH (Wang and Wan, 2009b). These factors affect the microbial community composition, the metabolic fluxes and ultimately the amount of hydrogen produced in the system (Wang and Wan, 2009b; Elsharnouby *et al.,* 2013). A slight change from the optimum setpoint may have a significant impact on the process yield (Fang and Liu, 2002; Fan *et al.,* 2004; Mu *et al*., 2006a; Wang *et al*., 2007). The influences of these parameters are briefly discussed below.

4.3.1. Inoculum type and concentration

The production of hydrogen is a specific mechanism to dispose of excess electrons through the activity of hydrogenases in bacteria. Specific types of bacteria that possess such capability include strict anaerobes which could be a single species (pure culture) or a mixture of two known species (co-culture) (Valdez -Vazquez *et al*., 2005). Examples of these are members of the genus *Clostridium* and *Enterobacter* (Hung *et al*., 2011). On the other hand a mixture of different types of microbes (mixed culture) may be used (Li and Fang, 2007). Mixed culture communities that are capable of producing hydrogen are ubiquitous in natural environments such as soil, wastewater, sewage sludge, compost and animal dung. (Sivagurunathan *et al*., 2014; Wang and Wan, 2009c; Cheong and Hansen, 2006; Hu and Chen, 2007; An *et al.*, 2014). Within these mixed consortia, a synergistic interaction occurs whereby other microbes that are not involved in the hydrogen production process create favourable conditions for the hydrogen-producing microorganisms (Sarkar *et al*., 2013; Yasin *et al*., 2013, Chen *et al*., 2015). Besides *Clostridium* and *Enterobacter* spp., other microbes capable of producing hydrogen such as *Klebsiella* spp. (Niu *et al*., 2010), *Bacillus* spp. (Kotay and Das, 2006), *Pseudomonas* spp. (Guo *et al*., 2008), *Escherichia coli (*Bisaillon *et al*., 2006; Turcot *et al*., 2008*)*, *Ethanoligenens* spp. (Xing *et al*., 2008), *Citrobacter* spp. (Oh *et al*., 2008), *Ruminococcus* spp. (Ntaikou *et al*., 2008) have been reported.

The type and characteristics of the microbial inoculum employed plays a significant role in the hydrogen production process. Pure cultures have shown to produce higher yields in comparison to mixed culture consortia (Masset *et al*., 2012). Even though pure cultures prove efficient for studying the mechanisms within the fermentation process, they still pose many challenges during operation. For instance, major hydrogen-producers such as *Clostridium* spp. require strictly anaerobic conditions for growth, therefore the addition of reducing agents in the culture medium is crucial to maintain low redox potentials in order to eliminate oxygen from the system. Alternatively, mixed cultures comprise a vast community of microorganisms that synergistically interact for growth and development. Mixed cultures are simpler to operate, cheaper to use since they do not require the addition of expensive reducing agents and are able to metabolize a wide range of substrates compared to pure cultures (Masset *et al*., 2012). Nevertheless, the presence of hydrogen-consuming microorganisms such as methanogens poses a huge challenge in the case of mixed cultures for biohydrogen production. Energy intensive and costly pretreatment methods are required for the inhibition

of hydrogen-consuming bacteria in mixed microbial communities. Several studies have reported the different inoculum pretreatment techniques that lead to higher hydrogen yields (Faloye *et al*., 2013; Faloye *et al*., 2014; Ren *et al*., 2008).

Regarding inoculum size, studies have shown that the initial cell concentration significantly influences the lag phase of cell growth, product formation and overall productivity (Wang and Jin, 2009). This may be attributed to the adjustment of the cells to fresh medium. A short lag phase may occur due to the cells' rapid adaptation to the conditions applied, whereas a long lag phase may be attributed to slow microbial growth and adjustment in the new system to establish biomass production and product formation (Kotay and Das, 2006; Wang and Jin, 2009). Studies have indicated that optimum initial inoculum concentration for biohydrogen production is influenced by the inoculum source used. In a study by Prakasham *et al.* (2011) it was shown that increasing both pH and inoculum concentration (anaerobic digested sludge) simultaneously for biohydrogen production would favour the fermentation process. The aforementioned authors stated that the optimum inoculum concentration was between 65-75%. On the other hand, Whiteman and Gueguim-Kana (2014) reported that the optimum inoculum size for hydrogen production was 15% with anaerobic digested sludge.

4.3.2. Substrate type and concentration

The substrate type can affect the hydrogen yield by selecting the metabolic pathway within microorganisms. Various types of substrates have been used for fermentative hydrogen production (Wang and Wan, 2009b). These include simple carbohydrates such as glucose and sucrose which are easily utilized by microorganisms; however these substrates are costly. Therefore, current research on fermentative hydrogen production is driven towards the use of waste material such as lignocellulosic biomass and industrial effluents that are both abundant and cost-effective (Mafuleka and Gueguim-Kana, 2015). Lignocellulosic materials require pretreatments prior to use for fermentation due to their complex structures that cannot be degraded by bacteria. Both physical (e.g. milling, extrusion and microwave) and chemical pretreatment methods (e.g. acids or bases) have been reported (Ramadoss *et al*., 2014). Industrial effluents such as food waste, dairy waste and sugar refining waste are much more attractive since they do not require expensive pretreatments and are readily accessible to the microorganisms compared to lignocellulosic materials (Whiteman and Gueguim-Kana, 2013).

Optimum substrate concentration for maximum hydrogen yield has been reported by several studies (Whiteman and Gueguim-Kana, 2014; Sekoai and Gueguim-Kana, 2013; Wang and Jin, 2009). Usually, an increase in the substrate concentration results in a higher hydrogen yield up to a certain level. Glucose and sucrose concentrations have been reported to be optimum within the range of 10-30 g/L (Wang and Wan, 2009a; Wang and Wan, 2009c; Mu *et al.*, 2009). Wang and Wan (2009c) reported a maximum yield of 305.3 mL H_2/g glucose. This result was consistent with Wang and Wan (2009b). The search for cheap and renewable substrates for use in biohydrogen production is currently underway (Lay *et al*., 2010; Whiteman and Gueguim-Kana, 2014). Renewable and abundant feedstock such as sugar cane molasses may be valuable for biohydrogen production. Molasses are a by-product of the sugarcane refining industry. It is a thick, dark syrup that results from the crystallization and extraction of the majority of sucrose from sugar cane. Current uses of molasses include animal feed additives, sweeteners or feedstock for renewable energy production such as bioethanol. Molasses consist of 50% sugar by dry weight and is primarily made of sucrose and is a much cheaper alternative to pure glucose (Wang and Jin, 2009). This substrate contains essential vitamins and minerals that are required by the microbes involved in biohydrogen production (Beshay and Moreira, 2005). Therefore, it does not require the supplementation of expensive essential vitamins and minerals (iron, nitrogen, phosphorus) that are fundamental for the bioprocess, thereby reducing production costs (He *et al*., 2007). Whiteman and Gueguim-Kana (2014) investigated the influence of substrate concentration on biohydrogen production and reported a maximum cumulative hydrogen volume (84.33 mL) at an optimum molasses concentration of 150 g/L. Conversely, Wang and Jin (2009) assessed the influence of carbon and nitrogen concentrations present in molasses for biohydrogen production and observed a maximum hydrogen yield (1.85 mol hydrogen/mol hexose) when the molasses concentration was 100 g/L.

4.3.3. Hydraulic retention time

Generally, HRT is considered an important operational parameter affecting continuous production of biohydrogen (Zhang *et al*., 2006). Nevertheless, several batch studies have considered HRT as an input parameter (Sekoai and Gueguim-Kana, 2013; Kim *et al.,* 2004; Lay, 2001). Studies that have considered HRT as a parameter have indicated that the control of HRT is essential for the inhibition of the hydrogen-consuming microbes such as methanogens (Chen *et al.,* 2001).

Moreover, an optimum HRT is significantly influenced by the substrate type used. Numerous studies have reported an optimum HRT for maximum biohydrogen production of 1-6 days (Thanwised *et al*., 2012; Zhang *et al.*, 2006). Generally, short HRTs are beneficial for hydrogen production due to costs associated with longer times. In addition, hydrogen-producers are able to grow and reproduce rapidly, whereas the methanogens require longer HRTs to proliferate (Liu *et al.,* 2008). In a study by Kim *et al.* (2004), it was found that short HRTs (<3 days) increased biohydrogen production. This result was consistent with Kim *et al*. (2010) and Tawfik *et al.* (2012). On the other hand, Sekoai and Gueguim-Kana (2013) modeled and optimized biohydrogen production and considered HRT as one of the inputs. Optimum biohydrogen yield was reported at an HRT of approximately 3.5 days. Also, Jayalakshmi *et al.* (2009) reported an optimum HRT of 7 days for biohydrogen production. Various studies have indicated that shorter HRTs can lead to a low pH (Liu *et al.,* 2008; Shin and Youn, 2005; Chang and Lin, 2004). This interaction has shown to enhance this system since it is a biological method for eliminating methanogens at mesophilic and thermophilic conditions (Oh *et al.,* 2004).

4.3.4. pH

The pH parameter is considered crucial for the hydrogen production process since it influences the hydrogenase system, substrate utilization and the metabolic activity of the hydrogen-producing microorganisms (Kothari *et al*., 2012). Studies have indicated that pH affects various activities within the bacterial cells including nutrient uptake due to cell membrane sensitivity (Li and Fang, 2007; Khanal *et al*., 2004). An increase in pH could enhance the activity of the hydrogen-producing bacteria up to a certain point beyond which it will adversely affect hydrogen production. Variations in pH can modify important processes such as metabolic activity, protein synthesis, and adaptation to extreme conditions by the bacteria (Kothari *et al*., 2012). Since the majority of studies were conducted in batch mode without pH control, whereby only the initial pH was investigated, the optimum initial pH for hydrogen production reported has shown to vary between studies. Initial pH values may influence the duration of the lag phase during fermentation. An optimum initial pH within the range of 6-7.5 was reported in numerous studies (Hawkes *et al*., 2002; Khanal *et al*., 2004). Generally, an initial pH between 4-4.5 may lead to an extended lag phase by inhibiting the hydrogenase activity which in turn affects the hydrogen production process (Fang and Liu, 2002; Hawkes *et al.,* 2002; Khanal *et al.,* 2004). At a lower initial pH,

hydrogen production occurs gradually over a longer time period (Sinha and Pandey, 2011). Conversely, a higher initial pH results in faster rate of hydrogen and acid production which eventually affects the buffering ability of the system.

4.3.5. Temperature

Temperature has been shown to have significant effects on hydrogen production processes. This parameter influences the growth rate and metabolic pathways of hydrogen-producing bacteria which in turn affects the activity of hydrogen-producing enzymes such as hydrogenases (Elsharnouby *et al*., 2013). It also influences substrate degradation efficiency, volatile fatty acid production, microbial communities and overall hydrogen yields (Fang and Liu, 2002). Various temperature ranges exist for carrying out biohydrogen production. Commonly used temperature conditions for hydrogen production are at mesophilic (20-40 °C), thermophilic (40-65 °C) or hyperthermophilic conditions (>80 °C) (Sinha and Pandey, 2011). The majority of studies have been carried out under mesophilic conditions (Elsharnouby *et al*., 2013). These temperature ranges are beneficial as a result of low-costs. High temperatures may cause protein denaturation within the hydrogen-producing bacteria which result in a decline in hydrogen production (Sinha and Pandey, 2011). Other studies have recommended higher temperatures for hydrogen production as this can eliminate non-spore forming methanogens and may improve the process yields (Lay *et al.,* 1999).

4.3.6. Reactor configuration

The reactor configuration used for biohydrogen production may vary with regard to vessel size. Bioreactors range from laboratory scale reactors (100-500 mL), semi-pilot scale (2-10 L) and pilot scale (10-400 L). These vessels may be operated in batch, fed batch and continuous mode (Show *et al*., 2011; De Gioannis *et al*., 2013). For industrial purposes, the continuous mode of hydrogen production is more feasible due to its many advantages. This includes monitoring and regulation of process parameters at their optimum (Ismail *et al*., 2009). Types of bioreactors reported by previous studies include Continuous stirred tank reactors (CSTR) (Chen and Lin, 2003; Zhang *et al*., 2007; Kim *et al*., 2011); anaerobic fluidized bed reactors (AFBR) (Lee *et al*., 2004; Zhang *et al*., 2008); upflow anaerobic sludge blanket reactors (UABR) (Chang and Lin, 2004; Gavala *et al*., 2006), anaerobic sequencing batch reactors (Vijaya-Bhaskar *et al*., 2008) and membrane bioreactors (Oh *et*

al., 2004).. The most frequently used for hydrogen fermentation processes are CSTRs. These bioreactors have been well-recognized for efficient homogenous mixing of the fermentation medium that results in high mass transfer (Show *et al*., 2011).

4.4. Challenges associated with Biohydrogen Modelling and Optimization

Bioprocess development is carried out during the initial stages of the fermentation process to achieve maximum hydrogen yields. The application of accurate and reliable bioprocess models is therefore imperative for bioprocess optimization. Even though several attempts have been made to abstract the relationships between the key inputs and the corresponding hydrogen output, significant variations exist between the reported optimum setpoints of input parameters for fermentative hydrogen production (Wang *et al*., 2005; Wang and Wan, 2009a; Wang and Wan, 2009c; Mu *et al*., 2006a; Mu *et al*., 2009).

Biohydrogen process development requires extensive knowledge at the lab scale level for efficient scale up (Escamilla-Alvarado *et al*., 2012). Previous reports have been limited due to modelling inaccuracies at the lab scale level which have significantly hindered the scale up phase (Schmidt, 2005). For instance, there is a lack of uniformity with regard to the most suitable fermentation volume size that should be adopted for screening, modelling and optimization during the initial stages of process development. Previous studies have often reported a volume size in the range of 100-200 mL (Whiteman and Gueguim-Kana, 2014; Sekoai and Gueguim-Kana, 2013; Wang *et al.*, 2005; Wang and Wan, 2009a; Wang and Wan, 2009c; Faloye *et al*., 2013; Faloye *et al*., 2014) and to a lesser extent between 1-6 L for modelling biohydrogen research (Rosales-Colunga *et al*., 2010; Prakasham *et al*., 2011; Shi *et al*., 2010; Mullai *et al*., 2013). Despite the availability of these studies, there is no reasonable scientific explanation for the selected volume used during the modelling and optimization process. Besides the inconsistency in the process volume size used for the development of bioprocess models, there is a gap of knowledge of the potential impact of the volume size on the model accuracy.

Lower volume sizes would accomplish a higher mixing efficiency and mass transfer compared to larger vessels (Schmidt, 2005). As opposed to chemical reactors, the scale up of microbial fermentation processes is drastically challenged by reproducibility in yield as the scale is increased. This could be as a result of the physiological characteristics (growth of the microorganisms) as well as the output (product formation) that occur within the reactor

(Votruba and Sobotka, 1992). A frequent catastrophic occurrence is the inability to sustain physiological conditions experienced at lab scale during the scale up process and it continues to pose a huge challenge. Scale-up of microbial fermentations presents several challenges since large vessels are considerably more heterogeneous in contrast to smaller vessels (Shuler and Kargi, 2002). The scaling process does not function in a linear manner, therefore even when geometrically similar vessels are employed, it may be impossible to achieve a similar rate of shear, mixing time, and mass transfer previously observed from the small vessel to the larger vessel (Shuler and Kargi, 2002). Efforts to overcome these challenges include modelling and optimization across bioprocess scales in order to determine the impact of the volume on the model accuracy and output yield.

Furthermore, significant discrepancies exist in the methods used for the different yield expression units by several studies on the key parameters that influence fermentative hydrogen production and have impeded the biohydrogen process development phase. Relating the key input parameters to the corresponding hydrogen output by using a standardized yield expression unit will contribute towards improving the hydrogen production process. Despite the availability of dispersed reports on the influence of key input parameters on biohydrogen production using different yield expression units, there is a lack of knowledge of intelligent models that have been built on pre-existing information which can efficiently predict the hydrogen response on unknown input patterns from the available public repositories. The development of accurate and reliable models will assist in determining the optimum setpoints for hydrogen production and could shorten the bioprocess development stage.

5. Application of ANNs in biofuel production

The efficiency of ANNs in bioprocess modelling has been well documented (Bourquin *et al*., 1998; Desai *et al*., 2008; Levstek and Lakota, 2010). More importantly, its use for modelling and optimization of biofuel production has proven valuable (Gueguim-Kana *et al.*, 2012a; Whiteman and Gueguim-Kana, 2014; Tardast *et al*., 2014; Ahmadian-Moghadam *et al.*, 2013). The superiority of ANN as a modelling tool essentially lies in its ability to represent the non-linearities in bioprocesses efficiently coupled with the capability of learning from historical data (Nath and Das, 2011). Other merits include the ability to approximate different forms of non-linear functions as well as the non-requirement of a prior specification of a suitable fitting function (Desai *et al*., 2008).

The development of biofuel production, like many other bioprocesses, requires the development of an accurate model to achieve process optimization and subsequent scale up towards industrialization. Several studies have reported the application of ANN for modelling and optimization of the key parameters associated with microbial fermentation in biofuel production (Wang and Wan, 2009a; Wang and Wan, 2009c; Gueguim-Kana *et al*., 2012a; Gueguim-Kana *et al*., 2012b ; Whiteman and Gueguim-Kana, 2014). In the same vein, some of these studies are further discussed below.

5.1. Biohydrogen production

The production of biohydrogen via the dark fermentation process entails the use of microorganisms under anaerobic conditions to degrade organic matter. Biohydrogen is viewed as an excellent potential replacement for conventional fossil fuels due to its high energy density (122 kJ/g) and its combustion which results in water as the only by-product. However, the commercialization of this process has been limited due to the low yields observed (Nath and Das, 2011). The use of ANNs for modelling and optimization of biohydrogen production has been widely reported. For instance, Wang and Wan (2009c) established the influence of temperature, initial pH and glucose concentration on hydrogen production output using BPNN. The prediction accuracy and optimization abilities of the response surface methodology and artificial neural network model were compared. The results showed that the root mean square error and the prediction error for the neural network model (17.80 and 7.70%) was much lower than that of the RSM model (38.40 and 16.60%), indicating the efficiency of ANN. In another study, the optimization of biohydrogen production on the input parameter of pH, glucose to xylose ratio, inoculum age and concentration as inputs resulted in a 14.25% improvement in the hydrogen yield, which further emphasized the efficiency of ANN for process optimization (Prakasham *et al*., 2011).

In addition, ANN models have been successfully used for the realtime monitoring and prediction of biohydrogen production. For example, Rosales-Colunga *et al*. (2010) estimated hydrogen production on inputs of oxidation reduction potential (ORP), dissolved CO_2 and pH during hydrogen fermentation. A coefficient of determination (R^2) value of 0.95 was observed indicating that the model had a good fitness (Rosales-Colunga *et al*., 2010). The authors reported that ANN models successfully estimated the hydrogen production using only on-line parameters, suggesting that this software sensor was a low-cost efficient tool for the monitoring of the biohydrogen process. Other studies reported on optimization of

biohydrogen production using ANN are summarized in Table 1. Important aspects of the developed neural network models such as the type of ANN used, ANN structure and coefficient of determination (R^2) are presented.

Table 1: Summary of Biohydrogen production modelling studies using ANN

Input Parameter	**Output parameters**	**Type of ANN**	**ANN Structure**	**R^2 value**	**References**
pH, glucose: xylose ratio, inoculum size, inoculum age	Cumulative H_2	BPNN	4-10-1	0.99	Prakasham *et al.* (2011)
T°C, pH, S_o	SE (%), HPR, HY	BPNN	3-5-1	-	Wang and Wan (2009a)
T°C, pH, S_o	HY	BPNN	3-4-1	-	Wang and Wan (2009c)
S_o, Inoculum %, T°C	Cumulative H_2	BPNN	4-(6-10)-1	0.91	Whiteman and Gueguim-Kana (2014)
ORP, pH, dissolved CO_2	HPR	BPNN	-	0.96	Rosales-Colunga *et al*. (2010)
HRT, So, ORP, pH, recycle ratio, alkalinity	HPR	BPNN	12-20-1	0.80	Nikhil *et al*. (2008)
OLR, ORP, pH, alkalinity	HPR	BPNN	4-3-1	-	Shi *et al*. (2010)

Table 1: Continued…

Input Parameters	Output parameters	Type of ANN	ANN Structure	R^2 value	References
pH, S_o, X_o, T°C, time	HPR	BPNN	5-6-4-1	0.98	Nasr *et al.* (2013a)
OLR, pH, VSS yield	HPR	BPNN	3-8-4-1	0.85	Nasr *et al.* (2013b)
OLR, HRT, influent alkalinity	HY,HPR, TOC eff, products conc.	BPNN	-	-	Mu and Yu (2007)
pH, Temperature, So and HRT	HY	BPNN	4-12-4-1	0.99	Mullai *et al*. (2013)

ORP: Oxidation-reduction potential; CO_2: Carbon dioxide; HPR: Hydrogen production; HRT: Hydraulic retention time; So: Initial substrate concentration, Xo= Initial biomass concentration; T°C: Temperature; SE (%): Substrate degradation efficiency; OLR: Organic loading rate; H_2: Hydrogen; TOC_{eff} : Effluent total organic carbons; VSS yield: Volatile suspended solids yield; BPNN: Back propagation neural network; HY: Hydrogen yield; R^2: Coefficient of determination.

5.2. Biogas production

The production of biogas involves the anaerobic digestion of organic materials. Biogas mainly comprises methane (55% to 70%), carbon dioxide (30% to 45%) and hydrogen (less than 10%) (Jönsson *et al*., 2003). The methane upgraded from biogas, may be used for heat and electricity generation or as a fuel for vehicles (Wellinger and Linberg 2000). Optimization of this process may improve the production and application of biogas as an alternative fuel to conventional fossil fuel sources. The use of ANNs for biogas production has been widely studied. Levstek and Lakota (2010) reviewed the use of ANNs for compounds prediction in biogas from anaerobic digestion. These authors summarized some of the most significant studies of the assessment and prediction of biogas constituents during production using ANNs.

Similarly, Ozkaya *et al*. (2007) studied the effect of leachate, pH, alkalinity, chemical oxygen demand (COD), sulphate, conductivity, chloride, temperature (°C) and refuse age on methane fraction (%) in biogas. The ANN model was developed to capture the effect of the inputs on methane fraction using field-scale bioreactors. These models were shown to be versatile and may be applied at large scale production. In another study, a multilayer back propagation ANN with two hidden layers and sigmoid function was trained to simulate the digestion process during biogas production. The ANN model successfully captured the underlying patterns in the training data set with input parameters of temperature, total solids, total volatile solids and pH. The performance of the ANN model demonstrated its efficiency with a correlation coefficient (R^2) of 0.87 (Abu Qdais *et al*., 2010).

In a study by Elnekave *et al.* (2012), three different ANNs, viz. back propagation (BPNN), radial basis function-based neural networks (RBF) and generalized regression neural networks (GRNN) were used to model the effect of flow rate, volumetric load, initial chemical oxygen demand (COD_{in}) and initial total suspended solids (TSS_{in}) on final chemical oxygen demand (COD_{out}), and final total suspended solids (TSS_{out}) for biogas production. The results indicated that the BPNN gave the best predictions with an average deviation in the range of 6.4-15.6% from the experimental values. These authors successfully developed an ANN model that was able to achieve a relatively high COD removal efficiency (77-79%) with simultaneous biogas production of 880-11000 m^3/day). Other selected studies reporting the use of ANN for optimization of biogas production are presented in Table 2. Moreover,

noteworthy characteristics pertaining to the developed neural network models such as the type of ANN used, input and output parameters, ANN structure and coefficient of determination (R^2) obtained are presented.

Table 2: Summary of Biogas production modelling studies using ANN

Input Parameters	Output parameters	Type of ANN	ANN Structure	R^2 value	References
Flow rate, Volumetric load, COD_{in}, TSS_{in}	COD_{out}, TSS_{out}, Biogas production	GRNN RBF BPNN	-	-	Elnekave *et al*. (2012)
OLR, VFA, influent-effluent alkalinity, influent-effluent pH, T°C	Biogas production	BPNN	-	0.93	Kanat and Saral (2009)
Sludge concentrations	Methane production	BPNN	5-7-1	0.99	Mahanty *et al*. (2013)
Co-substrates concentration	Biogas production	BPNN	5-2-1	-	Gueguim-Kana *et al*. (2012a)
Leachate (pH, Alkalinity, COD, sulphate, conductivity, chloride, waste T°C) and Refuse age	Methane fraction (%) in biogas	BPNN	-	0.96	Ozkaya *et al.* (2007)

Table 2: Continued…

Input Parameters	Output parameters	Type of ANN	ANN Structure	R^2 value	References
Peak current , Pre peak slope	COD removal efficiency (%), Methane production	BPNN	2-3-1	-	Harper Jr. and Taewoo (2013)
T°C, pH, TS, TVS	Biogas yield	BPNN	-	0.87	Abu-Qdais *et al*. (2010)
H_2S:S LR, H_2S in biogas, total sulphides, , pH, OLR	H_2S and NH_3 in biogas	BPNN	4-3-1	0.91, 0.83	Strik *et al.* (2005)

COD_{in}: Chemical oxgen demand (initial); TSS_{in}: Total suspended solids (initial) ; COD_{out} (final): Chemical oxygen demand (final) , TSS_{out}: Total suspended solids (final); OLR: Organic loading rate; VFA: Volatile fatty acids ; T°C: Temperature; H_2S:Hydrogen sulfide; S:Sulfur; NH_3: Ammonia; GRNN: Generalized regression neural neworks; RBF: Radial basis function-based neural network; BPNN: Back propagation neural network; LR: Loading rate; R^2: Coefficient of determination

5.3. Microbial Fuel Cell Technology

Microbial Fuel Cells (MFC) and Microbial Electrolysis Cells (MEC) make up the microbial fuel technology. While MFCs produce an electric current from the microbial decomposition of organic compounds, MECs partially reverse the process by using bacterial metabolism to generate hydrogen from organic material with an electric current (Cheng and Logan, 2007). MFC technology has been shown to be efficient for energy generation with simultaneous wastewater treatment (Logan and Regan, 2006). Although these systems prove useful, they are still limited by the low yields and lack of information pertaining to the influence of the interactive effect of key parameters on the process output. Thus there is a need to optimize the process parameters to enhance hydrogen and electricity production as well as improving its efficiency for wastewater treatment.

Although the field of mathematical models is highly advanced and has been extensively used for bioprocess modelling (Logan *et al.*, 2006; Kinoshita *et al*., 1988; Wang, 2004), its application for MEC and MFC has been scarcely reported. Mathematical models may assist in the development of these systems with regard to design, power (MFC) and hydrogen (MEC) output. These models can be used to test the hypothesis regarding microbial community composition, microbial activity and mode of electron transfer in these systems.

In a study by Tardast *et al.* (2014), an ANN model was applied for the prediction of power density on inputs of pH, temperature and electron acceptor concentration. The ANN model had a low mean square error (MSE) and R^2 value of 0.0023 and 0.99 respectively suggesting high prediction accuracy. The low MSE and high R^2 value showed that the ANN was able to accurately model the considered inputs with the corresponding output. A similar result was observed in a previous study by the same authors (Tardast *et al*., 2012). The concept of MEC technology for hydrogen production is a relatively new research area (Cheng and Logan, 2007). Hence, the use of conventional modelling approaches for optimization of hydrogen production in MECs has been scantily reported (Gil-Carerra *et al*., 2013; Tartakovsky *et al*., 2011; Yahya *et al*., 2015).

In our previous study, we developed a committee of ANN models on hydrogen production using microbial electrolysis cells (MEC) with inputs of substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration (Sewsynker *et al*.,

2015). The coefficients of determination for the five models were 0.90, 0.81, 0.85, 0.70 and 0.80, respectively. The results showed that the ANN committee was able to efficiently extract the non-linear behavior between the inputs and the target output (Sewsynker *et al*., 2015) The use of accurate and reliable models such as ANNs will help broaden the knowledge of both MFC and MEC systems and will contribute to increased yield and wastewater treatment efficiency.

As shown in Table 3, few studies from literature have modelled and optimized electricity and biohydrogen production from MFC technologies using ANN. A summary of the various studies is presented in Table 3.

Table 3: Summary of Microbial Fuel Cell (MFC) technology modelling studies using ANN

Input Parameters	Output parameters	Type of ANN	ANN Structure	R^2 value	References
pH, BOD, COD, TSS	Current generation	BPNN	4-4-1	-	Tardast *et al.* (2012)
T°C, pH, Electron acceptor concentration	Power density , Current density	BPNN	3-3-1	0.98	Tardast *et al.* (2014)
T°C, ferrous sulphate concentration	Voltage	BPNN	3-9-1	0.90	Garg *et al.* (2014)
pH, T°C, S_o, substrate type, applied voltage, MEC configuration	HY	BPNN	6-(6, 8, 11, 12, 14)-1	0.90, 0.81, 0.85, 0.70 and 0.80	Sewsynker *et al*. (2015)

BOD: Biochemical oxygen demand; COD: Chemical oxygen demand ; TSS: Total suspended solids ; T°C: Temperature ; S_o: Initial substrate concentration; MEC: Microbial Electrolysis Cell ; ; HY: Hydrogen yield ; BPNN: Back propagation neural network; R^2: Coefficient of determination

5.4. Biodiesel production

Biodiesel will play a major role in providing an alternative fuel for automobiles in the near future. The use of microalgae for biodiesel production represents a renewable and sustainable energy source due to their high biomass productivity and ability to treat both air and wastewater sources (Christenson and Sims, 2011). The advantages of using microalgae as opposed to oil crops (e.g. soybeans) are that microalgae have simple structures and high photosynthetic efficiency. Additionally, microalgae can be produced throughout the year since its growth conditions can be controlled compared to plant sources that only grow seasonally (Wu *et al*., 2012).

Nonetheless, the commercialization of microalgae biomass for biofuel production is still facing significant difficulties. These include high production costs and low yields. In light of these challenges, there is a need to model and optimize the biomass production and lipid profile during biodiesel production. The utilization of ANNs for the prediction of chemical compositions of lipids for biodiesel production has been well established (Jahirul *et al*.; 2014; Baroutian *et al*., 2008; Kumar *et al*., 2000). Few studies have reported the use of ANN for optimization of biodiesel production from microalgae. Mohamed *et al.* (2013) comparatively assessed ANN and RSM models for determining the effect of glucose concentration, yeast extract and sodium nitrate on the lipid productivity of *Tetraselmis* sp. FTC 209. Their findings revealed that even though both ANN and RSM efficiently modelled the considered inputs on the output, the ANN model was more robust for prediction in non-linear systems (Mohamed *et al.*, 2013).

Similarly, Wu and Shi (2006) investigated the effect of glucose concentration on biomass concentration (*Chlorella pyrenoidosa* 15-2070) with the use of a hybrid ANN model and a deterministic kinetic model. Optimized biomass concentrations and maximum productivity for the hybrid ANN was 10 and 40% higher than that predicted by the deterministic kinetic. Other reports on the use of ANNs for optimization of biomass concentration of microalgae for biodiesel production are summarized in Table 4. The reviewed literature on optimization of biodiesel production elucidates the efficiency of the development of process models such as ANNs. As shown in Table 4, modelling and optimization of biodiesel production using ANN has been scantily reported. Major elements of the developed neural network models such as

the type of ANN used, ANN structure and coefficient of determination (R^2) are presented in Table 4.

Table 4: Summary of Biodiesel production modelling studies using ANN

Input parameter	Output parameter	Type of ANN	ANN structure	R^2-value	References
S_o, yeast extract concentration and sodium nitrate concentration	Lipid productivity, Biomass Concentration	BPNN	3-10-1	0.99	Mohamed *et al.* (2013)
Time	Biomass Concentration	-	-	0.95	Furlong *et al.* (2013)
pH	Biomass Concentration	-	1-2-1	-	Galvão *et al.* (2013)
Glucose Concentration	Biomass Concentration	HNN	1-3-1	-	Wu and Shi (2006)

BPNN: Back propagation neural network; R^2: Coefficient of determination

5.5. Bioethanol production

Another renewable and sustainable fuel alternative to the depleting petroleum sources is bioethanol. The production of this fuel occurs via the microbial fermentation of organic matter. The most commonly used substrates for bioethanol production are corn, sugar cane and wheat (Sarkar *et al*., 2012). However, its competitiveness and market value with fossil fuels has limited its implementation. The main goal of bioethanol optimization is to increase yields while reducing costs.

As shown in Table 5, the application of ANNs for modelling and optimization of bioethanol production is still limited. Ahmadian-Moghadam *et al.* (2013) assessed the effect of initial substrate (molasses) concentration, live yeast cells and dead yeast cells as input process parameters on bioethanol production using *Saccharomyces cerevisae*. An R^2 value of 0.93 was obtained which shows that the model was suitable for recognizing patterns in the data and accurately predicted the bioethanol yield. In a more recent study by Betiku and Taiwo (2015), the effect of breadfruit hydrolysate concentration, hydraulic retention time and pH on bioethanol production was evaluated using ANN and RSM. The ANN model had a prediction error of 0.24% compared to 1.67% by RSM. These results further confirm ANNs ability to model non-linear processes compared to other modelling techniques such as RSM. Table 5 shows the studies on the modelling and optimization of bioethanol production with the corresponding bioethanol output. Key features of the developed neural network models such as the type of ANN used, ANN structure and R^2 value are presented in Table 5.

Table 5: Summary of Bioethanol production modelling studies using ANN

Input parameter	Output parameter	Type of ANN	ANN structure	R^2-value	References
S_o, Live Yeast Cells, Dead Yeast Cells	Bioethanol production	-	-	0.93	Ahmadian-Moghadam *et al.* (2013)
S_o , HRT and pH	Bioethanol production	BPNN	3-3-1	1	Betiku and Taiwo (2015)

S_o: Initial substrate concentration; HRT: Hydraulic retention time; BPNN: Back propagation neural network; R^2: Coefficient of determination

6. Genetic Algorithms (GA) coupled with ANN for optimization

Genetic Algorithm is an artificial intelligence based stochastic non-linear optimization technique (Goldberg, 1989). This class of Algorithm was based on the evolutionary process of natural selection and genetics in nature (Renner and Ekárt, 2003; Shopova and Bancheva, 2006). While ANNs are typically used for modelling non-linear associations between the process input variables and the target output, Genetic algorithm (GA) is an optimization algorithm that determines the optimum input setpoints for the maximum process output (Davis, 1991). Genetic Algorithm has proven to be effective in solving various optimization problems in bioprocess development (Sarkar and Modak, 2003). Once the ANN model is developed and validated, it is deemed an objective function for optimization by the GA module. During the optimization process, the first generation which comprises chromosomes that are made up of genes (i.e. the inputs being investigated) is assessed using the ANN model. Subsequently, the best solutions are chosen for breeding purposes in order to obtain the second generation (Sexton *et al*., 1999; Desai *et al*., 2008; Whiteman and Gueguim-Kana, 2014).

These individuals are then combined arbitrarily for 'crossing over' to take place, thereby imitating the biological phenomenon of natural selection. The parent chromosomes will pair and thereafter exchange genes at randomly spread out points to produce the next generation. In order to improve this process, mutations are added and genes on specific chromosomes are arbitrarily substituted with values that occur within the search range. Once this occurs, a new assortment transpires for the generation of new individuals and possible solutions. This process is repeated several times till an optimum threshold is met, thereby generating a potential global optimal solution. (Sexton *et al*., 1999; Whiteman and Gueguim-Kana, 2014). Several studies have employed ANN models coupled with GA (ANN-GA) for optimization (Pansandideh and Niaki, 2006; Sexton *et al*., 1999; Desai *et al*., 2008). In particular, the application of ANN-GA for biofuel production has been extensively studied (Wang and Wan, 2009c; Gueguim-Kana *et al*., 2012a; Gueguim-Kana *et al*., 2012b; Whiteman and Gueguim-Kana, 2014; Betiku and Taiwo, 2015; Abu-Qdais *et al*., 2010; Ahmadian-Moghadam *et al.*, 2013). For instance, Gueguim-Kana *et al.* (2012a) reported the modelling and

optimization of biogas production on mixed substrates of sawdust, cow dung, banana stem, rice bran and paper waste using ANN coupled with Genetic Algorithm (GA). The optimized substrate profile predicted biogas production of 10.14 L. Assessment of the optimal profile gave a biogas production of 10.28 L, which shows an 8.64% improvement in biogas yield coupled with a reduction in the lag phase with the onset of production from day 3 compared to day 8 (Gueguim-Kana *et al.*, 2012a). These results demonstrate the high modelling ability of ANN for non-linear processes such as biogas production. Application of such tools would provide much more insight into the optimum conditions required for maximum biofuel production.

7. Comparative assessment of ANN and RSM for modelling and optimization of biofuel production

Several studies have comparatively examined the use of ANN and RSM for bioprocess modelling and optimization (Desai *et al*., 2008; Giordano *et al*., 2010; Gueguim-Kana *et al*., 2012b). More specifically, the comparative assessment of ANN and RSM for biofuel production is currently increasing (Whiteman and Gueguim-Kana, 2014; Wang and Wan, 2009c; Mohamed *et al*., 2013; Betiku and Taiwo, 2015). In a study by Wang and Wan (2009c), RSM and ANN efficiency were compared for modelling biohydrogen production. The findings revealed that the RSM model had a much higher prediction error (16.60%) in contrast to the ANN model (7.70%) indicating the efficiency of ANN over RSM for predicting non-linear systems. This result was also in accordance with Whiteman and Gueguim-Kana (2014). Similarly, Mohamed *et al.* (2013) comparatively used ANN and RSM models for modelling and optimizing biodiesel production. These authors reported that although RSM was able to relate the considered process inputs to the output, the ANN model was more robust for predicting the non-linear systems. Betiku and Taiwo (2015) investigated bioethanol production using ANN and RSM. The abovementioned authors indicated that the ANN model had a prediction error of 0.24% compared to 3.41% by RSM which further confirms the superiority of ANN over RSM.

A summary of these comparative studies is presented in Table 6. Although both ANN and RSM have been reported to be suitable in modelling and optimization of bioprocesses, ANN models have proven to be more efficient for non-linear processes such as microbial fermentations.

Table 6: Summary of Comparative modelling studies using ANN and RSM for biofuel production

Input Parameters	Output parameters	ANN R^2	RSM R^2	ANN Prediction error (%)	RSM Prediction error (%)	References
S_o, Inoculum %, T°C	Cumulative H_2	0.91	0.75	15.12	119.08	Whiteman and Gueguim-Kana (2014)
T°C, pH, S_o	HY	-	-	7.70	16.60	Wang and Wan (2009c)
S_o, yeast extract concentration and sodium nitrate concentration	Lipid productivity, Biomass Concentration	0.99	0.99	4.18	5.72	Mohamed *et al.* (2013)
S_o, HRT and pH	Bioethanol production	1	0.99	0.24	3.41	Betiku and Taiwo (2015)

S_o: Initial substrate concentration, T°C: Temperature; H_2: Hydrogen; HY: Hydrogen yield; HRT: Hydraulic retention time; R^2: Coefficient of determination.

Conclusion

Regardless of the complex biological systems associated with bioprocesses, ANNs have shown to efficiently encapsulate the non-linear behavior of various fermentation processes in biofuel production. The studies highlighted in this review show the high prediction accuracy of ANNs. It is apparent that ANNs are becoming a powerful tool in modelling biofuel production due to its flexible learning algorithm, diverse network topology, fast learning algorithm, and high error tolerance for non-linear processes such as those associated with microbial fermentations. Particularly, the use of ANN for biohydrogen production has shown to be valuable. However, bioprocess experimentation with very small data sizes may be problematic in several instances and may be unlikely to provide sufficient information for network training. The use of virtual experimentation by employing ANN and GA in bioprocess development can also reduce costs and process development time.

References

1. Abu-Qdais, H., Bani-Hani, K., Shatnawi, N., 2010. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resour. Conserv. Recy. 54,359-363.

2. Ahmadian-Moghadam, H., Elegado, F.B., Nayve, R., 2013. Prediction of ethanol concentration in biofuel production using Artificial Neural Networks. Am. J. Model. Optim. 1(3), 31-35.

3. Almeida, J.S., 2002. Predictive non-linear modeling of complex data by artificial neural networks. Curr. Opin. Biotechnol. 3, 72-76.

4. An, D., Li, Q., Wang, X.,Yang, H., Guo, L., 2014. Characterization on hydrogen production performance of a newly isolated *Clostridium beijerinckii* YA001 using xylose. Int. J. Hydrogen Energy 39, 19928-19936.

5. Baroutian, S., Aroua, M. K., 2008. Estimation of Vegetable Oil-Based Ethyl Esters Biodiesel Densities using Artificial Neural Networks. J. Appl. Sci. 8, 3005-3011.

6. Berrios-Rivera, S.J., Yang, Y.T., Bennett, G.N., San, K.Y., 2000. Effect of glucose analog supplementation on metabolic flux distribution in anaerobic chemostat cultures of *Escherichia coli*. Metab. Eng. 2, 149-154.

7. Beshay, U., Moreira, M., 2005. Production of alkaline protease with Teredinobacter turnirae in controlled fed-batch fermentation, Biotechnol. Lett. 27, 1457-1460.

8. Betiku, E., Taiwo, A.E. 2015., Modeling and optimization of bioethanol production from breadfruit starch hydrolyzate vis- a-vis response surface methodology and artificial neural network. Renew. Energ. 74, 87-94.

9. Bisaillon, A., Turcot, J., Hallenbeck, P.C., 2006. The effect of nutrient limitation on hydrogen production by batch cultures of Escherichia coli. Int. J. Hydrogen Energy 31, 1504-1508.

10. Bishop, C. 1995., Neural networks for pattern recognition. Oxford University Press.

11. Bourquin, J., Schmidli, H., Hoogevest, P.V., Leuenberger, H., 1998. Advantages of artificial neural networks (ANNs) as alternative modeling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form. Eur. J. Pharm. Sci. 7, 5-16.

12. Cai, G., Jin, B., Saint, C., Monis, P., 2011. Genetic manipulation of butyrate formation pathways in *Clostridium butyricum*. J. Biotechnol. 155, 269-274.

13. Chang, F.Y., Lin, C.Y., 2004. Biohydrogen production using an up-flow anaerobic sludge blanket reactor. Int. J. Hydrogen Energy 29, 33-39.

14. Chang, S., Li, J., Liu, F., 2011. Continuous biohydrogen production from diluted molasses in an anaerobic contact reactor. Front Environ. Sci. Eng. China 5(1), 140-148.

15. Chen, C.C., Lin, C.Y., 2003. Using sucrose as a substrate in an anaerobic hydrogen producing reactor. Adv. Environ. Res. 7, 695-699.

16. Chen, C.C, Lin, C.Y., Chang, J., 2001. Kinetics of hydrogen production with continuous anaerobic cultures utilizing sucrose as the limiting substrate. Appl. Microbiol. Biotechnol. 57, 56-64.

17. Chen, P., Wang, Y., Yan, L., Wang, Y., Li, S, Yan, X., Wang, N., Liang, N., Li, H., 2015. Feasibility of biohydrogen production from industrial wastes using defined microbial co-culture. Biol. Res. 48, 24-31.

18. Cheng, S.A., Logan, B.E., 2007. Sustainable and efficient biohydrogen production via electrohydrogenesis. Proc Natl Acad Sci USA. 104(47), 18871-18873

19. Cheng, S., Xing, D., Logan, B.E., 2011. Electricity generation of single-chamber microbial fuel cells at low temperatures. Biosens. Bioelectron. 26, 1913-1917.

20. Cheong, D.Y., Hansen, C.L., 2006. Bacterial stress enrichment enhances anaerobic hydrogen production in cattle manure sludge. Appl. Microbiol. Biotechnol. 72, 635-643.

21. Chin, H., Chen, Z., Chou, C.P. (2003). Fedbatch operation using Clostridium acetobutylicum suspension culture as biocatalyst for enhancing hydrogen production. Biotechnol. Progr. 19, 383–388.

22. Christenson, L., Sims, R., 2011. Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. Biotechnol. Adv. 29:686-702.

23. Davis, L., 1991. Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York.

24. De Gioannis, G., Muntoni, A., Polettini, A., Pomi, R., 2013. A review of dark fermentative hydrogen production from biodegradable municipal waste fractions. Waste Manage. 33, 1345-1361.

25. Desai, K.M., Survase, S.A., Saudagar, P.S., Lele, S.S., Singhal, R.S., 2008. Comparison of artificial neural network (ANN) and response surface

methodology (RSM) in fermentation media optimization: Case study of fermentative production of scleroglucan. J. Biochem. Eng. 41, 266-273.

26. Elnekave, M., Celik, S.O., Tatlier, M., Tufekci, N., 2012. Artificial Neural Network predictions of upflow anaerobic sludge blanket (UASB) reactor performance in the treatment of citrus juice wastewater. Pol. J. of Environ. stud. 21(1), 49-56.

27. Elsharnouby, O, Hafez, H, Nakhla, G., Naggar, M.H.E., 2013. A critical literature review on biohydrogen production by pure cultures. Int. J. Hydrogen Energy 38, 4945-4966.

28. Escamilla-Alvarado, C., Rios-Leal, E., Ponce-Noyola, M.T., Poggi-Varaldo, H.M., 2012. Gas biofuels from solid substrate hydrogenic–methanogenic fermentation of the organic fraction of solid municipal wastes. Process Biochem. 47(11), 1572-1587.

29. Faloye, F.D., Gueguim-Kana, E.B., Schmidt, S., 2013. Optimization of hybrid inoculum development techniques for biohydrogen production and preliminary scale up. Int. J. Hydrogen Energy 38, 11765-11773.

30. Faloye, F.D., Gueguim-Kana, E.B., Schmidt, S., 2014. Optimization of biohydrogen inoculum development via a hybrid pH and microwave treatment technique-Semi pilot scale production assessment. Int. J. Hydrogen Energy 39, 5607-5616.

31. Fan, Y., Li, C., Lay, J.J., Hou, H., Zhang, G., 2004. Optimization of initial substrate and pH levels for germination of sporing hydrogen-producing anaerobes in cow dung compost. Bioresour. Technol. 91, 189-193.

32. Fang, H.H.P., Liu, H., 2002. Effect of pH on hydrogen production from glucose by mixed culture. Bioresour. Technol. 82, 87-93.

33. Franco-Lara, E., Link, H., Weuster-Botz, D., 2006. Evaluation of artificial neural networks for modeling and optimization of medium composition with a genetic algorithm. Process Biochem. 41, 2200-2206.

34. Furlong, V.B., Pereira-Filho, R.D., Margarites, A.C., Goularte, P.G., Costa, J.A.V., 2013. Estimating microalgae Synechococcus nidulans daily biomass concentration using neuro-fuzzy network Ciência e Tecnologia de Alimentos 33(1), 142-147.

35. Galvão, R.M., Santana, T.S., Fontes, C.H.O., Sales, E.A., 2013. Modeling of Biomass Production of Haematococcus pluvialis. Appl. Math. 4, 50-56.

36. Garg, A., Vijayaraghavan, V., Mahapatra, S.S., Tai, K., Wong, C.H., 2014. Performance evaluation of microbial fuel cell by artificial intelligence methods. Expert Syst. Appl. 41:1389-1399.

37. Garlapati, V.K., Banerjee, R., 2010. Evolutionary and swarm intelligence-based approaches for optimization of lipase extraction from fermented broth. Engineering in Life Sciences 10, 265-273.

38. Gavala, H.N., Skiadas, I.V., Ahring, B.K., 2006. Biological hydrogen production in suspended and attached growth anaerobic reactor systems. Int. J. Hydrogen Energy 31, 1164-1175.

39. Ghosh, D., Sobro, I.F., and Hallenbeck, P.C., 2012. Optimization of the Hydrogen Yield from Single-Stage Photofermentation of Glucose by Rhodobacter capsulatus JP91 using Response Surface Methodology. Bioresource Technol. 123, 199-206.

40. Gil-Carrera, L., Escapa, A., Carracedo, B., Morán, A., Gómez, X., 2013. Performance of a semi-pilot tubular microbial electrolysis cell (MEC) under several hydraulic retention times and applied voltages. Bioresource Technol. 146, 63-69.

41. Giordano, P.C., Martínez, H.D., Iglesias, A.A., Beccaria, A.J., Goicoech, H.C., 2010. Application of response surface methodology and artificial neural networks for optimization of recombinant *Oryza sativa* non-symbiotic hemoglobin 1 production by *Escherichia coli* in medium containing by product glycerol. Bioresource Technol. 101, 7537-7544.

42. Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, New York.

43. Graupe, D., 2007. Principles of artificial neural networks. Vol. 6. World Scientific.

44. Gueguim-Kana, E.B.G., Oloke, J.K., Lateef, A., Adesiyan, M.O., 2012a. Modelling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renew. Energ. 46, 276-281.

45. Gueguim-Kana, E.B., Oloke, J.K., Lateef, A., Oyebanji, A., 2012b. Comparative evaluation of artificial neural network coupled genetic algorithm and response surface methodology for modelling and optimization of citric acid production by Aspergillus niger MCBN297. Chem. Eng. Trans. 27, 397-402.

46. Guo, L., Li, X.M., Bo, X., Yang, Q., Zeng, G.M., Liao, D.X., et al., 2008. Impacts of sterilization, microwave and ultrasonication pretreatment on hydrogen producing using waste sludge. Bioresource Technol. 99, 3651-3658.

47. Guo, K., Tang, X., Du, Z., Li, H., 2010. Hydrogen production from acetate in a cathode-on-top single-chamber microbial electrolysis cell with a mipor cathode. Biochem. Eng. J. 51, 48-52.

48. Gurney, K., 1997. An Introduction to Neural Networks. CRC Press

49. Hagan, M.T., Demuth, H.B., Beale, M., 1996. Neural Network Design. PWS Publishing Co, Boston.

50. Haider, M.A., Pakshirajan, K., Singh, A., Chaudhry, S., 2008. Artificial neural network genetic algorithm approach to optimize media constituents for enhancing lipase production by a soil microorganism. Appl. Biochem. Biotech. 144, 225-235.

51. Hallenbeck, P.C. and Ghosh, D., 2009. Advances in fermentation biohydrogen production: the way forward? Trends Biotechnol. 27(5), 287-297.

52. Harper Jr., W.F., Taewoo, Y., 2013. Using electronic signals and neural networks to monitor the performance of an anaerobic bioreactor. Int. J. Water Res. Environ. Eng. 5(9), 521-532.

53. Hassoun, M.H., 1995. Fundamentals of Artificial Neural Networks. MIT Press, Cambridge.

54. Hawkes, F., Dinsdale, R., Hawkes, D., Hussy, I., 2002. Sustainable fermentative hydrogen production: challenges for process optimisation. Int. J. Hydrogen Energy 27(11-12), 1339-1347.

55. He, X., Guo, X., Liu, N., and Zhang, B., 2007. Ergosterol production from molasses by genetically modified Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 75, 55-60.

56. Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T., Williams, S.T., 1994. Bergey's Manual of Determinative Bacteriology, 9th ed. Williams & Willkins, Baltimore, M. D.

57. Hopfield, J.J., Tank, D.W., 1986. Computing with neural circuits: a model. Science. 233, 625–633.

58. Hu, B., Chen, S.L., 2007. Pretreatment of methanogenic granules for immobilized hydrogen fermentation. Int. J. Hydrogen Energy 32, 3266-3273.

59. Huang, Y., Kangas, L.J., Rasco, B.A., 2007. Applications of Artificial Neural Networks (ANNs) in Food Science. Crit. Rev. Food Sci. 47, 113-126.

60. Hung, C.H., Chang, Y.T., Chang, Y.J., 2011. Roles of microorganisms other than Clostridium and Enterobacter in anaerobic fermentative biohydrogen production systems–A review. Bioresource Technol. 102(18), 8437-8444.

61. Ismail, F., Abd-Aziz, S., Mei-Ling, C., and Hassan, M.A., 2009. Statistical optimization of biohydrogen production using food waste under thermophilic conditions. Open Ren. Energy. J. 2, 124-131.

62. Jahirul, M.I., Brown, R.J., Senadeera, W., Moghaddam, L., 2014. Estimation of biodiesel properties from chemical composition – An artificial neural network (ANN) approach. In Ringo, Elena (Ed.) International Conference on Environment and Renewable Energy, City Internationale Universitaire de Paris, France.

63. Jayalakshmi, S., Joseph, K., Sukumaran, V., 2009. Bio hydrogen generation from kitchen waste in an inclined plug flow reactor. Int. J. Hydrogen Energy 34, 8854–8858.

64. Jönsson, O., Polman E, Jensen JK, Eklund R, Schyl H, Ivarsson S., 2003. Sustainable gas enters the European gas distribution system. Danish Gas Technology Center.

65. Kanat, G., Saral, A., 2009. Estimation of biogas production rate in a thermophilic UASB reactor using Artificial Neural Networks. Environ. Model. Assess. 14, 607-614.

66. Khanal, S.K., Chen, W.-H., Li, L., and Sung, S., 2004. Biological hydrogen production: effects of pH and intermediate products. Int. J. Hydrogen Energy 29, 1123-1131.

67. Kim, S.H., Han, S.K., Shin, H.S., 2004. Feasibility of biohydrogen production by anaerobic co-digestion of food waste and sewage sludge. Int. J. Hydrogen Energy 29, 1607-1616.

68. Kim, D.H., Kim, S.H., Kim, K.Y., Shin, H.S., 2010. Experience of a pilot-scale hydrogenproducing anaerobic sequencing batch reactor (ASBR) treating food waste. Int. J. Hydrogen Energy 35, 1590-1594.

69. Kim, D.H., Wu, J., Jeong, K.W., Kim, M.S., Shin, H.S., 2011. Natural inducement of hydrogen from food waste by temperature control. Int. J. Hydrogen Energy 36, 1066-1067.

70. Kinoshita, K., McLarnon, F., Cairns, E., 1988. Fuel cell handbook: prepared by Lawrence Berkeley Laboratory. 2-4.

71. Kotay, S.M., Das, D., 2006. Microbial hydrogen production with Bacillus coagulans IIT-BT S1 isolated from anaerobic sewage sludge. Bioresource Technol. 98(6), 1183-1190.

72. Kothari, R., Singh, D.P., Tyagi, V.V., Tyagi, S.K., 2012. Fermentative hydrogen production- an alternative clean energy source. Renew. Sust. Energ. Rev. 16, 2337-2346.

73. Kumar, J., Bansal A., and Jha, M.K., 2000. Comparison of Statistical and Neural Network Techniques in Predicting Physical Properties of Various Mixtures of Diesel and Biodiesel. Proceedings of the World Congress on Engineering and Computer Science. San Francisco, USA.

74. Kumar, N., Ghosh, A., Das, D., 2001. Redirection of biochemical pathways for the enhancement of H_2 production by *Enterobacter cloacae*. Biotechnol. Lett. 23, 537-541.

75. Kyazze, G., Popov, A., Dinsdale, R., Esteves, S., Hawkes, F., Premier, G., Guwy, A., 2010. Influence of catholyte pH and temperature on hydrogen production from acetate using a two chamber concentric tubular microbial electrolysis cell. Int. J. of Hydrogen Energ. 35, 7716-7722.

76. Lay, J.J., 2001. Biohydrogen generation by mesophilic anaerobic fermentation of microcrystalline cellulose. Biotechnol. Bioeng. 74(4), 280-287.

77. Lay, J.J., Lee, Y.J., Noike, T., 1999. Feasibility of biological hydrogen production from organic fraction of municipal solid waste. Water Res. 33, 2579-2586.

78. Lay, C.H., Wu, J.H., Hsiao, C.L., Chang, J.J., Chen, C.C., Lin, C.Y., 2010. Biohydrogen production from soluble condensed molasses fermentation using anaerobic fermentation. Int. J. Hydrogen Energy 35, 13445-13451.

79. Lee, K.S., Wu, J.F., Lo, Y.S., Lo, Y.C., Lin, P.J. Chang, J.S., 2004. Anaerobic hydrogen production with an efficient carrier-induced granular sludge bed bioreactor. Biotechnol. Bioeng. 87, 648-657.

80. Levin, D.B., Pitt, L., Love, M., 2004. Biohydrogen production: prospects and limitations to practical application. Int. J. of Hydrogen Energy 29, 173-185.

81. Levstek, T., Lakota, M. 2010., The use of artificial neural networks for compounds prediction in biogas from anaerobic digestion – A review. Agricultura 7, 15-22.

82. Li, C.L, Fang, H.H.P., 2007. Fermentative hydrogen production from waste water and solid wastes by mixed culture. Crit. Rev. Env. Sci. Technol. 37, 1-39.

83. Liu, D., Angelidaki, I., Zeng, R. J., & Min, B. (2008). Bio-hydrogen production by dark fermentation from organic wastes and residues. Kgs. Lyngby: DTU Environment.

84. Logan, B.E., Hamelers, S. Rozendal, R., Schroder, U., Keller, J., Freguia, S., Aelterman, P. Verstraete, W., Rabaey, K., 2006. Microbial Fuel Cells: Methodology and Technology. Environ. Sci. Technol. 40(17), 5181-5192.

85. Logan, B.E. and Regan, J.M., 2006. Electricity-producing bacterial communities in microbial fuel cells. Trends in Microbiol. 14, 512-518.

86. Lotfy, W.A., Ghanem, K.M., El-Helow, E.R., 2007. Citric acid production by a novel *Aspergillus niger* isolate:II. Optimization of process parameters through statistical experimental designs. Bioresource Technol. 98, 3470-3477.

87. Lou, W., Nakai, S., 2001. Application of artificial neural networks for predicting the thermal inactivation of bacteria: a combined effect of temperature, pH and water activity. Food Res. Int. 34, 573-591

88. Lu, L., Ren, N., Zhao, X., Wang, H., Wu, D., Xing, D., 2011. Hydrogen production, methanogen inhibition and microbial community structures in psychrophilic single chamber microbial electrolysis cells. Energ. Environ. Sci. 4, 1329-1336.

89. Mafuleka, S., Gueguim Kana, E.B. (2015). Modelling and optimization of xylose and glucose production from napier grass using hybrid pre-treatment techniques. Biomass Bioenergy 77, 200-208.

90. Mahanty, B., Zafar, M., Park, H.S., 2013. Characterization of co-digestion of industrial sludges for biogas production by artificial neural network and statistical regression models. Environm. Technol. 34(13–14), 2145-2153.

91. Marchitan, N., Cojocaru, C., Mereuta, A., Ducac, G.H., Cretescud, I., Gonta, M., 2010. Modeling and optimization of tartaric acid reactive extraction from aqueous solutions: A comparison between response surface methodology and artificial neural network. Sep. Purif. Technol. 75, 273–285.

92. Masset, J., Calusinska, M., Hamilton, C., Hiligsmann, S., Joris, B., Wilmotte, A. et al., 2012. Fermentative hydrogen production from glucose and starch using pure strains and artificial co cultures of Clostridium spp. Biotechnol. Biofuels 5(1), 1-15.

93. Messikh, N., Samar, M.H., Messikh, L., 2007. Neural network analysis of liquid-liquid extraction of phenol from wastewater using TBP solvent. Desalination 208, 42–48.

94. Mjalli, F.S., 2005. Neural network model-based predictive control of liquid-liquid extraction contactors. Chem. Eng. Sci. 60, 239-253.

95. Mohamed, M.S., Tan, J.S., Mohamad, R., Mokhtar, M.N., Ariff, A.B., 2013. Comparative Analyses of Response Surface Methodology and Artificial Neural Network on Medium Optimization for Tetraselmissp. FTC209 Grown under Mixotrophic Condition. The Sci. World J. 1-14. Article ID 948940.

96. Mu, Y., Yu, H.Q., 2007. Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. Int. J. of Hydrogen Energ. 32, 3308-3314.

97. Mu, Y., Zheng X.J., Yu, H.Q., Zhu, R.F., 2006a. Biological hydrogen production by anaerobic sludge at various temperatures. Int. J. Hydrogen Energy 31, 780-785.

98. Mu, Y., Wang, G., Yu, H.Q., 2006b. Response surface methodological analysis on biohydrogen production by enriched anaerobic cultures. Enzyme Microb. Technol. 38, 905-913.

99. Mu, Y., Zheng, X.J., Yu, H.Q., 2009. Determining optimum conditions for hydrogen production from glucose by an anaerobic culture using response surface methodology (RSM). Int. J. Hydrogen Energy 34, 7959-7963.

100. Mullai, P., Yogeswari, M.K., Sridevi, K., Ross, P.R., 2013. Artificial neural network (ANN) modeling for hydrogen production in continuous anaerobic sludge blanket filter (ASBF). Singaporean J. Sci. Res. 5(1), 1-7.

101. Naik, S.N., Goud, V.V., Rout, P.K., Dalai, A.K., 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sust. Energ. Rev. 14, 578-597.

102. Nagata, U., Chu, K.H., 2003. Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnol. Lett. 25, 1837-1842.

103. Nandi, R., Sengupta, S., 1998. Microbial production of hydrogen— An overview. Crit. Rev. Microbiol. 24(1), 61-84.

104. Nasr, N., Hafez, H., El Naggar, M.H., Nakhla, G., 2013a. Application of artificial neural networks for biohydrogen production. Int. J. of Hydrogen Energy 38, 3189-3195.

105. Nasr, M., Tawfik, A., Ookawara1, S., Suzuki, M., 2013b. Prediction of hydrogen production using Artificial Neural Network. International Water Technology Conference, Istanbul.

106. Nath, K., Das, D., 2011. Modeling and optimization of fermentative hydrogen production. Bioresource Technol. 102, 8569-8581.

107. Nikhil, B.O., Visa, A., Lin, C.Y., Puhakka, J.A., Yli-Harja, O., 2008. An artificial neural network based model for predicting H_2 production rates in a sucrose-based bioreactor system. World Acad. Sci. Eng. Technol. Eng. 37, 20-25.

108. Niu, K., Zhang, X., Tan, W.S., Zhu, M.L., 2010. Characteristics of fermentative hydrogen production with *Klebsiella pneumoniae* ECU-15 isolated from anaerobic sewage sludge. Int. J. Hydrogen Energy 35(1), 71-80.

109. Ntaikou, I., Gavala, H.N., Kornaros, M., Lyberatos, G., 2008. Hydrogen production from sugars and sweet sorghum biomass using *Ruminococcus albus*. Int. J. Hydrogen Energy 33, 1153-1163.

110. Oh, S.E., Iyer, P., Bruns, M.A., Logan, B.E., 2004. Biological hydrogen production using a membrane bioreactor. Biotechnol. Bioeng. 87, 119-127.

111. Oh, Y.K., Kim, H.J., Park, S., Kim, M.S., Ryu, D.D.Y., 2008. Metabolic-flux analysis of hydrogen production pathway in *Citrobacter amalonaticus* Y19. Int. J. Hydrogen Energy 33, 1471-1482.

112. Ozkaya, B., Demir, A., Bilgili, M.S., 2007. Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environ. Modell. Softw. 22, 815-822.

113. Pasandideh, S.H., Niaki, S.T., 2006. Multi-response simulation optimization using genetic algorithm within desirability function framework. Appl. Math. Comput. 175, 366-382.

114. Pham, D.T., 1994. Neural networks in engineering. In: Rzevski G. *et al*. Applications of Artificial Intelligence in Engineering IX, AIENG/94,. Comput. Mech. 3-36.

115. Prakasham, R.S., Sathish, T., Brahmaiah, P., 2011. Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. Int. J. Hydrogen Energy 36, 4332-4339.

116. Ramadoss, G., Muthukumar, K., 2014. Ultrasound assisted ammonia pretreatment of sugarcane bagasse for fermentable sugar production. Biochem. Eng. J. 83, 33-41.

117. Ren, N.Q., Guo, W.Q., Wang, X.J., Xiang, W.S., Liu, B.F., Wang, X.Z., 2008. Effects of different pretreatment methods on fermentation types and dominant bacteria for hydrogen production. Int. J. Hydrogen Energy 33, 4318-4324.

118. Ren, N.Q., Li, J., Li, B., Wang, Y., Liu, S., 2006. Biohydrogen production from molasses by anaerobic fermentation with a pilot-scale bioreactor system. Int. J. Hydrogen Energy 31, 2147-2157.

119. Renner, G., Ekárt, A., 2003. Genetic algorithms in computer aided design. Computer Aided Design. 35, 709-726.

120. Rosales-Colunga, L.M., Garcia, R.G., Rodriguez, A.D., 2010. Estimation of hydrogen production in genetically modified *E. coli* fermentations using an artificial neural network. Int. J. of Hydrogen Energy 35, 13186-13192.

121. Sadrzadeh, M., Mohammadi, T., Ivakpour, J., Kasiri, N., 2008. Separation of lead ions from wastewater using electrodialysis: comparing mathematical and neural network modeling. Chem. Eng. J. 144, 431-441.

122. Saraphirom P., Reungsang, A., 2010. Optimization of biohydrogen production from sweet sorghum syrup using statistical methods. Int. J. of Hydrogen Energy 35, 13435-13444.

123. Sarkar, D., Modak, J.M., 2003. Optimization of fed-batch bioreactor using genetic algorithm. Chem. Eng. Sci. 58, 2283-2296.

124. Sarkar, N., Ghosh, S.K., Bannerjee, S., Aikat, K., 2012. Bioethanol production from agricultural wastes: An overview. Renew. Energ. 37, 19-27.

125. Sarkar, O., Kannaiah, G.R., Venkata, S.G, Venkata-Mohan, S., 2013. Relative effect of different inorganic acids on selective enrichment of acidogenic biocatalyst for fermentative biohydrogen production from wastewater. Bioresource Technol. 147, 321-331.

126. Schalkoff, R.J., 1997. Artificial Neural Networks. McGraw-Hill, New York.

127. Schmidt, F.R., 2005. Optimization and scale up of industrial fermentation processes. Appl. Microbiol. Biotechnol. 68, 425-435.

128. Sekoai, P.T., Gueguim-Kana, E.B., 2013. A two-stage modelling and optimization of biohydrogen production from a mixture of agro-municipal waste. Int. J. Hydrogen Energy 38, 8657-8663.

129. Selembo, P.A., Perez, J.M., Lloyd, W.A., Logan, B.E., 2009. High hydrogen production from glycerol or glucose by electrohydrogenesis using microbial electrolysis cells. Int. J. of Hydrogen Energ. 34, 5373-5381.

130. Sexton, R.S., Dorsey, R.E., Johnson, J.D., 1999. Optimization of neural networks: a comparative analysis of the genetic algorithm and simulated annealing. Eur. J .Oper. Res. 114, 589-601.

131. Sewsynker, Y., Gueguim-Kana, E.B., Lateef, A., 2015. Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs). Biotechnol. Biotec. Eq. 29(6), 1208-1215.

132. Shi, Y., Gai, G., Zhao, X., Zhu, J., Zhang, P., 2010. Back propagation neural network (BPNN) simulation model and influence of operational parameters on hydrogen bio-production through integrative biological reactor (IBR) treating wastewater. Hei Long-jiang, China: College of Power and Energ. Eng. Harbin Eng. Uni.

133. Shin, H.S., Youn, J.H., 2005. Conversion of food waste into hydrogen by thermophilic acidogenesis. Biodegradation 16, 33-44.

134. Shopova, E.G., Vaklieva-Bancheva, N.G., 2006. BASIC- A genetic algorithm for engineering problems solution. Comput. Chem. Eng. 30, 1293–1309.

135. Show, K.Y., Lee, D.J., Chang, J.S., 2011. Bioreactor and process design for biohydrogen production. Bioresource Technol. 102, 8524-8533.

136. Shuler, M.L., Kargi, F., 2002. Bioprocess engineering: Basic concepts. 2. Upper Saddle River, NJ: Prentice Hall PTR.

137. Sinha, P., Pandey, A., 2011. An evaluative report and challenges for fermentative biohydrogen production. Int. J. Hydrogen Energy 36,7460-7468.

138. Sivagurunathan, P., Sen, B., Lin, C.Y., 2014. Batch fermentative hydrogen production by enriched mixed culture: Combination strategy and their microbial composition. J. Biosci. Bioeng. 117(2), 222-228.

139. Strik, D., Domnanovich, A.M., Zani, L., Braun, R., Holubar, P., 2005. Prediction of trace compounds in biogas from anaerobic digestion using the Matlab Neural Network Toolbox. Environ. Modell. Softw. 20, 803-810.

140. Tardast, A., Rahimnejad, M., Najafpour, G., Ghoreyshi, A., Premier, G.C., Bakeri, G. and Oh, S.E., 2014. Use of artificial neural network for the prediction of bioelectricity production in a membrane less microbial fuel cell. Fuel.117, 697-703.

141. Tardast, A., Rahimnejad, M., Najafpour, G., Pirzade, K., Mokhtarian, N., 2012. Prediction of bioelectricity production by neural network. J. Biotechnol. Pharm. Res. 3(3), 62-68.

142. Tartakovsky, B., Manuel, M.F., Wang, H., Guiot, S.R., 2009. High rate membrane-less microbial electrolysis cell for continuous hydrogen production. Int. J. of Hydrogen Energy 34, 672-677.

143. Tartakovsky, B., Mehta, P., Santoyo, G., Guiot, S.R., 2011. Maximizing hydrogen production in a microbial electrolysis cell by real time optimization of applied voltage. Int. J. Hydrogen Energy 36, 10557-10564.

144. Tawfik, A., El-Qelish, M., 2012. Continuous hydrogen production from co-digestion of municipal food waste and kitchen wastewater in mesophilic anaerobic baffled reactor. Bioresource Technol. 114, 270-274.

145. Thanwised, P., Wirojanagud, W., Reungsang, A., 2012. Effect of hydraulic retention time on hydrogen production and chemical oxygen demand removal from tapioca wastewater using anaerobic mixed cultures in anaerobic baffled reactor (ABR). Int. J. Hydrogen Energy 37, 15503-15510.

146. Turcot, J., Bisaillon, A., Hallenbeck, P.C., 2008. Hydrogen production by continuous cultures of Escherchia coli under different nutrient regimes. Int. J. Hydrogen Energy 33, 1465-1470.

147. Valdez-Vazquez, I., Ríos-Leal, E., Esparza-García, F., Cecchi, F., Poggi-Varaldo, H.M., 2005. Semi-continuous solid substrate anaerobic reactors for H_2 production from organic waste: Mesophilic versus thermophilic regime. Int. J. Hydrogen Energy 30, 1383-1391.

148. Vardar-Schara, G., Maeda, T., Wood, T.K., 2008. Metabolically engineered bacteria for producing hydrogen via fermentation. Microb. Biotechnol. 1,107-125.

149. Vazquez, G.D., Navarro, C.B.C., Colunga, L.M.R., Rodriguez, A.D., Flores, E.A., 2009. Continuous biohydrogen product using cheese whey: Improving the hydrogen production rate. Int. J. Hydrogen Energy 34, 4296-4304.

150. Vijaya-Bhaskar, Y., Venkata-Mohan, S. Sarma, P., 2008. Effect of substrate loading rate of chemical wastewater on fermentative biohydrogen production in biofilm configured sequencing batch reactor. Bioresource Technol. 99, 6941-6948.

151. Votruba, J., Sobotka, M., 1992. Physiological similarity and bioreactor scale-up. Folia Microbiol. 37, 331-345.

152. Wang, C.Y., 2004. Fundamental models for fuel cell engineering. Chemical. Rev. 104(10), 4727-4766.

153. Wang, C.H., Lu, W.B., Chang, J.S., 2007. Feasibility study on fermentative conversion of raw and hydrolyzed starch to hydrogen using anaerobic mixed microflora. Int. J. Hydrogen Energy 32, 3849-59.

154. Wang, G., Mu, Y., Yu, H.Q., 2005. Response surface analysis to evaluate the influence of pH, temperature and substrate concentration on the acidogenesis of sucrose-rich wastewater. Biochem. Eng. J. 23, 175–184.

155. Wang, J.L., Wan, W., 2009a. Application of desirability function based on neural network for optimizing biohydrogen production process. Int. J. Hydrogen Energy 34, 1253-1259.

156. Wang, J.L., Wan, W., 2009b. Factors influencing fermentative hydrogen production: a review. Int J Hydrogen Energy 34, 799-811.

157. Wang, J.L., Wan, W., 2009c. Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology. Int. J. Hydrogen Energy 34, 255-261.

158. Wang, X., Jin, B., 2009. Process optimization of biological hydrogen production from molasses by a newly isolated *Clostridium butyricum* W5. J. Biosci. Bioeng. 107(2), 138-144.

159. Wang, Y.X., Lu, Z.X., 2005. Optimization of processing parameters for the mycelial growth and extracellular polysaccharide production by *Boletus* sp. ACCC50328. Process Biochem. 40, 1043–1051.

160. Wellinger, A., Linberg, A., 2000. Biogas upgrading and utilization-IEA Bioenergy Task. 24.

161. Whiteman, J.K., Gueguim-Kana, E.B., 2014. Comparative assessment of the Artificial Neural Network and Response Surface Modelling Efficiencies for Biohydrogen Production on Sugar Cane Molasses. Bioenerg. Res. 7(1), 295-305.

162. Wu, Z., Shi, X., 2006. Optimization for high-density cultivation of heterotrophic Chlorella based on a hybrid neural network model. Lett. Appl. Microbiol. 44, 13-18.

163. Wu, X., Ruan, R., Du, Z., Liu, Y., 2012. Current Status and Prospects of Biodiesel Production from Microalgae. Energies. 5, 2667-2682.

164. Xing, D.F., Ren, N.Q., Wang, A.J., Li, Q.B., Feng, Y.J., Ma, F., 2008. Continuous hydrogen production of auto-aggregative Ethanoligenens harbinense YUAN-3 under non-sterile condition. Int. J. Hydrogen Energy 33, 1489-1495.

165. Yahya, A.Z., Wahab, A.K.A, Hussain, M.A., 2013. Optimal Production of Biohydrogen Gas via Microbial Electrolysis Cells (MEC) in a Controlled Batch Reactor System. Chem. Eng. J. 32, 727-732.

166. Yahya, A.Z., Hussain, M.A., Wahab, A.K.A., 2015. Modeling, optimization, and control of microbial electrolysis cells in a fed-batch reactor for production of renewable biohydrogen gas. Int. J. Energ. Res. 39(4), 557-572.

167. Yasin, N.H., Mumtaz, T., Hassan, M.A., Abd, R.N., 2013. Food waste and food processing waste for biohydrogen production: A review. J. Environ. Manage. 130C, 375-385.

168. Yu, L., Shouyang W., Kin, K.L., 2007. Basic Learning Principles of Artificial Neural Networks. Foreign-Exchange-Rate Forecasting With Artificial Neural Networks. 27-37.

169. Zhang, H., Bruns, M.A., Logan, B.E., 2006. Biological hydrogen production by *Clostridium acetobutylicum* in an unsaturated flow reactor. Water Res. 40, 728-734.

170. Zhang, Z.P., Show, K.Y., Tay, J.H., Liang, D.T., Lee, D.J., Jiang, W.J., 2007. Rapid formation of hydrogen-producing granules in an anaerobic continuous stirred tank reactor induced by acid incubation. Biotechnol. Bioeng. 96, 1040-1050.

171. Zhang, Z.P., Show, K.Y., Tay, J.H., Liang, D.T. Lee, D.J., 2008. Biohydrogen production with anaerobic fluidized bed reactors- a comparison of biofilm-based and granule-based systems. Int. J. Hydrogen Energy 33, 1559-1564.

172. Zhang, Y., Teng, L., Quan, Y., Tian, H., Dong, Y., Meng, Q. et al., 2010. Artificial intelligence based optimization of fermentation medium for b-glucosidase production from newly isolated strain Tolypocladium cylindrosporum. Lecture Notes in Computer Science. 6330, 325-332.

173. Zhou, M., Wang, H., Hassett, D.J. and Gu, T., 2012. Recent advances in microbial fuel cells (MFCs) and microbial electrolysis cells (MECs) for wastewater treatment, bioenergy and bioproducts. J. Chem. Technol. Biotechnol. 88, 508-518.

174. Zupan, J., Gasteiger, J., 1999. Neural networks in chemistry and drug design. Wiley-VCH, Weinheim. 9-36.

CHAPTER 3

Does the Volume Matter? An Insight into Modelling and Optimization of Biohydrogen production across scales

This chapter has been submitted to the journal Bioenergy Research with the title: Does the Volume Matter? An Insight into Modelling and Optimization of Biohydrogen production across scales

The manuscript is presented in the following pages.

Does the Volume Matter? An Insight into Modelling and Optimization of Biohydrogen production Across Scales

Yeshona Sewsynker and Evariste Bosco Gueguim Kana*

Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

*Corresponding author. E-mail: kanag@ukzn.ac.za

Abstract

There is a renewed interest in biohydrogen production as a potential alternative to depleting fossil fuels. Its scale up requires the availability of accurate and reliable process models that relate the key operational parameters to hydrogen yields at various scales of the process development. In this paper, the Response Surface Methodology (RSM) and Artificial Neural Networks (ANN) were used to model and optimize biohydrogen production at two different process scales. The input variables consisted of inoculum size (10-50%), molasses concentration (100-300 g/L) and Hydraulic Retention Time (10-48 h) and the output was the hydrogen yield (mol H_2/ mol sucrose consumed). The considered process scales were the culture volumes of 80 and 800 mL. Seventeen experimental data were generated at each scale and used for model development and process optimization, thus a total of two models at each scale. ANN based models gave R^2 values of 0.99 and 0.95 whereas RSM based models gave R^2 values of 0.97 and 0.89 for 80 and 800 mL, respectively. Process optimization with these models gave predicted yields of 0.87 and 0.73 mol H_2/mol sucrose consumed (ANN based models) and 1.09 and 0.72 mol H_2/mol sucrose consumed (RSM based models) for 80 and 800 mL, respectively. Models validation on ANN models gave experimental yields of 0.89 and 0.71 mol H_2/mol sucrose consumed compared to 0.99 and 0.70 mol H_2/mol sucrose consumed (RSM models) for 80 and 800 mL, respectively. These models showed relatively negligible deviations from their predicted values across scales. The RSM model at 80 mL (RSM_Model_{80}) predicted the highest yield compared to the other three models. A comparative analysis of the prediction errors indicated that the ANN model at 80 mL (ANN_Model_{80}) displayed a higher accuracy for prediction on unknown data. Semi-pilot scale (8 L) process assessments under optimized conditions showed negligible yield discrepancies from the predictive values of the models at lab scale. Microbial community analysis using Next Generation Sequencing (NGS) revealed the presence of presumptive hydrogen-producing microorganisms which were members within the genus *Clostridia*, *Enterbacter* and *Klebsiella*. These findings suggested that miniaturization of experiments for biohydrogen model development do not significantly impact on the model accuracy, thus reducing costs during the process developmental stage.

Keywords: Modelling, Optimization, Bioprocess scales, Biohydrogen production, Artificial Neural Networks, Response Surface Methodology

Introduction

The global energy crisis combined with environmental impact from fossil fuel consumption has skewed research towards alternative renewable sources [1]. Currently, the fermentative production of hydrogen is gaining significant interest. This is due to its high energy yield (122 kJ/g) that is about 2.9 times greater than fossil fuels [2]. Additionally, the only by-product formed from its combustion is water [3]. Various biological methods exist for hydrogen production and include photo-fermentation, dark fermentation and microbial electrolysis. However, at present the most energy efficient method is via dark fermentation [2]. During this process hydrogen is generated by microbial degradation of organic matter under anaerobic conditions [4].

The commercialization of hydrogen has been impeded by its low yields [5]. Current efforts are being channelled towards modelling and optimization of the physicochemical parameters that have a significant impact on the production process. Some of these parameters include pH, temperature, hydraulic retention time, agitation, substrate type and concentration and inoculum type and concentration [6–8]. Optimum input ranges for temperature, pH, substrate concentration and hydraulic retention time have been reported between 25-40°C, 6-9, 10-30 g/L and less than 3 days depending on the substrate type, respectively [1, 6–10]. The types of microbes involved in hydrogen production include both pure and mixed cultures. Pure cultures mainly comprise of *Clostridium* and *Enterobacter* spp. [11], whereas mixed cultures consist of a range of microbes that display synergistic interactions for metabolic functioning and survival [12–17].

The search for cheap and renewable substrates for biohydrogen production is currently underway [4, 18, 19]. Renewable and sustainable feedstocks such as sugar cane molasses may be valuable potential feedstocks for biohydrogen production. Molasses are by-products from the crystallization and extraction of the majority of sucrose from sugar cane. Cheeseman [20] reported that the production of sugar cane molasses in South Africa approximates to 850 000 tons per year. This feedstock is a much cheaper alternative to glucose and contains essential vitamins and minerals that are required by the microbes involved in biohydrogen production [21]. Therefore, it does not necessitate the supplementation of expensive essential vitamins and minerals (iron, nitrogen, phosphorus) that are fundamental for the bioprocess, thereby reducing production costs [22].

Biohydrogen production process scale up requires the development of process models that relate the abovementioned key input parameters to the hydrogen yields at the lab scale level [23]. The use of models that are accurate and reliable at various scales of the process development phase is of paramount importance. Modelling and optimization of bioprocesses have been carried out using the One Variable at a Time approach (OVAT) and statistical methods. However, limitations of OVAT are

that (1) it does not consider the interactive effects of parameters on the process output and (2) it is impractical to obtain a suitable optimum with few experiments [24]. Factorial design of experiment (DOE) is tedious, resource-intensive and laborious when the quantity of inputs is increased.

Conversely, multivariate methods such as the Response Surface Methodogy (RSM) and Artificial Neural Networks (ANN) have proven to be more efficient with regard to their predictive accuracy on complex non-linear bioprocesses. RSM is a statistical modelling system which employs a polynomial regression analysis to produce a second-order model equation thereby relating the process inputs and output. Hence, the optimum process operational set points are obtained by solving the model equation [25]. RSM assumes that the model equation can estimate the fermentation dynamics quite accurately whereas ANN is completely data-driven and studies the relationship between input and output variables in an attempt to understand the underlying effects that govern the process, similar to the human brain. The most common ANN architecture is the multi-layered perceptron (MLP) which consists of an input, one or more hidden layers and the output, comprising of neurons which may differ in amount subject to the complexity of the process it is being applied to [19].

ANN has proven to be more suitable for modeling bioprocesses compared to RSM [6, 19]. This is owing to the fact that ANN does not require a prior knowledge of the process kinetics. Studies have indicated that ANN can work well even with relatively less data. However, the data must be statistically well distributed in the input domain [19, 26, 27]. Therefore, experimental data of RSM should be adequate to build an effective ANN model. The use of ANN and RSM as modelling tools for biohydrogen production has been reported [6, 7, 19, 26, 28–30], as well as a comparative assessment of both tools [6, 19].

There is a lack of consensus on the appropriate fermentation process volume size for model development, process optimization, and substrate screening required at an early stage of process development. The most commonly reported volume for modelling biohydrogen research has been in the range of 100-200 mL [6, 7, 19, 28, 29, 31–33] and to a lesser extent between 1-6 L [34–37]. There is a dearth of studies on the scientific rationale of the choice of fermentation volume size for process modelling and optimization. Biohydrogen process development requires extensive process knowledge from laboratory scale for efficient scale up [23]. Modelling inaccuracies at the laboratory scale significantly impact the scale up phase [38]. Several studies have reported on modelling and optimization of biohydrogen production [6, 7, 19, 28, 29, 31–37]. In addition to a lack of uniformity in the process volume size used for the development of the above models, there is a gap of knowledge on the potential impact of the volume size on the model accuracy.

In this study, the Response Surface Methodology (RSM) and Artificial Neural Networks (ANN) were comparatively used to model and optimize biohydrogen production at two different process scales. This was performed to determine the impact of process scale (80 and 800 mL) on each model's efficiency. In addition, the sensitivity of each input parameter was examined across both model types (RSM and ANN) and scales (80 and 800 mL). Furthermore, the optimized models were comparatively assessed at semi-pilot scale.

Materials and Methods

Experimental setup

Substrate and Inoculum Pretreatment

The inoculum source used in this study was the anaerobic digested sludge obtained from Darville wastewater treatment plant, Pietermaritzburg, South Africa. The sludge was autoclaved at 121 °C for 10 min to deactivate the hydrogen-consuming methanogens [18]. The substrate used in this study was sugar cane molasses (C-Molasses), a by-product from the Illovo Sugar Mill, Eston, South Africa with the composition as shown in Table 1. The molasses were heated at 60 °C for 30 min to decrease the vegetative microbial cells.

Experimental design

The Box-Behnken response design was used to generate seventeen experimental runs for the development of each model. The input parameters consisted of inoculum size (10-50 %), molasses concentration (100-300 g/L) and hydraulic retention time (HRT) (10-48 h). The parameters, with their coded values and search ranges are shown in Table 2.

Table 1. Composition of the sugarcane molasses (on 100% dark matter basis)

Component	Content
Moisture	26.67 %
Non-Structural Carbohydrates	16.69 %
Crude Protein	4.76 %
Nitrogen	0.76 %
Calcium	0.90 %
Magnesium	0.50 %
Potassium	4.70 %
Sodium	0.11 %
Potassium/Calcium+Magnesium	1.39 %
Phosphorus	0.12%
Zinc	8 mg/kg
Copper	2 mg/kg
Manganese	94 mg/kg
Iron	166 mg/kg

Batch Fermentation Experiments

Bath fermentations were carried out at two scales of 80 and 800 mL to generate data for modelling and optimization at these scales. The bioreactors used were modified Erlenmeyer flasks. Thirty four batch experiments (seventeen batches per scale) were carried out with pretreated molasses and inoculated with the treated sludge. No mineral salts were added since the molasses naturally contained the essential minerals and vitamins required by the microbes for biohydrogen production (shown in Table 1). The reactors were thereafter flushed with nitrogen gas for 2 min to create anaerobic conditions. The input parameters, namely inoculum size, molasses concentration and hydraulic retention time (HRT) were maintained according to the design in Table 3 and the volumes were made up to each respective scale (80 and 800 mL) using autoclaved water. Operational temperature, initial pH and agitation were maintained at 37.5 °C, 6.5 and 120 rpm, respectively. All thirty four experiments were carried out in duplicate.

Table 2. Input variables and their ranges used by Box–Behnken for design generation

Variable	Coded Factor	Input	Coded Values			Unit
			-1	0	1	
Inoculum size	A	10-50	10	30	50	%
Molasses Concentration	B	100-300	100	200	300	g/L
Hydraulic Retention Time	C	10-48	10	29	48	hours

Analytical Procedure

The hydrogen fraction of mixed biogas was determined using the hydrogen sensor BCP-H_2 (Bluesens, Germany) with an operational range of 0-100% and a measuring principle based on a thermal conductivity detector. The gas volume was measured using the water displacement method. The cumulative volume of biohydrogen produced was computed according to Equation (1).

$$V_{H,i} = V_{H,i-1} + C_{H,I} (V_{G,i,}-V_{G,i-1}) + V_H (C_{H,i} - C_{H,i-1}) \quad (1)$$

$V_{H,i}$ and $V_{H,i-1}$ are cumulative hydrogen gas volumes at the current (i) and previous (i−1) time intervals, $V_{G,i,}$ and $V_{G,i-1}$ the total biogas volumes in the current and previous time intervals, $C_{H,i}$ and $C_{H,i-1}$ the fraction of hydrogen gas in the headspace of the reactor in the current and previous time intervals, and V_H the total volume of headspace in the reactor [39].

The hydrogen output was computed in terms of hydrogen yield (mol hydrogen/mol sucrose consumed) at STP (Standard Temperature and Pressure). For substrate consumption, the sucrose content was determined using a Biochemistry Analyzer (Model 2700 select-dual configuration, YSI, USA).

Semi-pilot scale process with optimized models

Optimized conditions from the four models were further evaluated at semi-pilot scale in a 10 L bioreactor (Labfors INFORS HT, Switzerland). The reactor was heat sterilized and four processes were set up according to the optimized setpoints of HRT, molasses concentration and inoculum size, determined from ANN and RSM based models, with reactor working volumes of 8 L. The control setpoints of initial pH, temperature and agitation were maintained at 6.5, 37.5°C and 180 rpm, respectively. These batch processes were designated RSM_$Model_{80}$_8L, RSM_$Model_{800}$_8L, ANN_$Model_{80}$_8L and ANN_$Model_{800}$_8L. All experiments were carried out in duplicate. Hydrogen

fraction of the biogas was monitored using a hydrogen sensor BCP-H_2 (Bluesens, Germany) with a detection range of 0-100% and a measuring principle based on thermal conductivity detector. Gas volume was measured using a milligas counter (MGC, Bluesens, Germany). The hydrogen sensor was interfaced to the F-lab Biogas software described by Faloye *et al.* [32] and the sampling interval was set to 1 min. The cumulative volume of hydrogen produced was calculated according to Equation (1).

DNA extraction and Next Generation Sequencing (NGS)

DNA extraction

DNA extraction was carried out according to the modified method of Orsini and Romano-Spica [40]. A 1 mL sample was extracted during peak hydrogen production from the bioreactor using the optimized conditions for the most accurate model (shown under results section) and was suspended in 1 ml of extraction buffer (50 mM Tris-HCl, 25 mM EDTA, 0.1% (w/v) SDS, 0.1% (w/v) PVP, pH 8.0). The sample was thereafter centrifuged at 8000 rpm for 1 min. The supernatant was discarded and the pellet was suspended in 500 µl of lysis buffer (50 mM Tris-HCl, 25 mM EDTA, 3.0% (w/v) SDS, 1.0% (w/v) PVP, pH 8.0). The sample was thereafter heated at 90 ^oC for 10 min and rapidly cooled in liquid nitrogen. A pre-warmed (65 ^oC) extraction solution (500 µl; 10 mM Tris-HCl, 1 mM EDTA, 300 mM sodium acetate, 1.0% (w/v) PVP) was added to the sample. Phenol:choloroform:isoamyalcohol (25:24:1) was added to the tube and mixed by inversion. Isopropanol was used to precipitate the resulting DNA. The DNA pellet was subsequently washed with 70% ethanol and thereafter re-suspended in 100 µl TE buffer (pH 8.0). The DNA extract was quantified using the Nanodrop 2000 spectrophotometer.

Next Generation Sequencing (NGS) and analysis

The 16S rRNA gene fragments of extracted DNA were amplified by PCR using the universal bacterial primer 907R (5'-CCGTCAATTCMTTTGAGTTT-3') [41]. Next generation sequencing was performed on Illumina MiSeq platform (Inqaba Biotec, South Africa). The raw reads obtained for the metagenome were used (high quality reads, q>30 were only selected) for taxonomic profiling and was carried out using CLC Genomics Workbench 8.5.1 [42], with an e-value less than 5×10^{-3}.

RSM Model development and Validation

The experimental data obtained from batch fermentations for 80 and 800 mL were used to develop the polynomial equations that relate hydrogen production to the process input parameters. The general form of the polynomial model is shown in Equation 2:

$$Y= \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3 + \alpha_{11}x_1^2 + \alpha_{22}x_2^2 + \alpha_{33}x_3^2 + \alpha_{12}x_1\,x_2 + \alpha_{13}x_1\,x_3 + \alpha_{23}x_2\,x_3 \quad (2)$$

where Y represents the response output, α_0 is the intercept, α_1x_1,α_2x_2, α_3x_3 are the linear coefficients, $\alpha_{11}x_1^2$, $\alpha_{22}x_2^2$, $\alpha_{33}x_3^2$ are the quadratic coefficients and $\alpha_{12}x_1\,x_2$, $\alpha_{13}x_1\,x_3$, $\alpha_{23}x_2\,x_3$ represent the interaction of coefficients. The significance of these models was assessed using the Analysis of variance (ANOVA) (Design Expert software, Stat Ease, Inc.). Optimum input set points for biohydrogen yield were obtained by solving the equation using the methods of Myers and Montgomery [43]. These set points were thereafter validated experimentally in duplicate. R^2 values for the RSM models were calculated as the Proportional Reduction of Error (PRE) according to Equation 3 [44].

$$\mathbf{PRE}= \frac{E_1 - E_2}{E_2} \qquad (3)$$

where E_1 (or total sum of squares, SST) is the prediction errors made when excluding the independent variables. E_2 (residual sum of squares, SSE) measures the prediction errors made when the prediction is based on the independent variables.

Artificial Neural Network Modelling

ANN structure

Two individual neural networks built on multilayer perceptrons were structured and used for model development. Each neural network had a topology of 3-5-5-1, corresponding to the number of neurons of input, hidden (two) and output layers (Figure 1).

The input vector comprised hydraulic retention time (HRT), sugarcane molasses concentration, and inoculum size on the corresponding output (hydrogen yield). The feed forward architecture was adopted, whereby the input layer neurons transmitted signals to the hidden layer neurons [45]. For the hidden layer, a sigmoid transfer function was implemented. This hidden layer had two main purposes: (1) the addition of the weighted inputs together with the linked bias; (2) then, to change the input data to a non-linear form, as shown in the following Equations 4 and 5 [45]:

$$\text{sum} = \sum_i^n = 1^{xiwi} + \theta \qquad (4)$$

where, w_i (i = 1, n) are the connection weights, θ is the bias and x_i is the input variable (Desai *et al*., 2008)

$$f(sum) = \frac{1}{1+\exp(-sum)} \qquad (5)$$

The learning patterns were randomly selected during the learning process. The Mean Square Error (MSE) between predicted and observed data was calculated according to Equation 6.

$$\text{RMSE}=\sqrt{\frac{\sum_{i}^{N}\sum_{n=1}^{M}\left(y^{i}n-\hat{y}^{i}n\right)^{2}}{NM}} \qquad (6)$$

where, N refers to the number of patterns used in the training; M denotes the number of output nodes; i denotes the index of the input pattern (vector) and $y^{i}n$ and $\hat{y}^{i}n$ are the actual and predicted outputs, respectively.

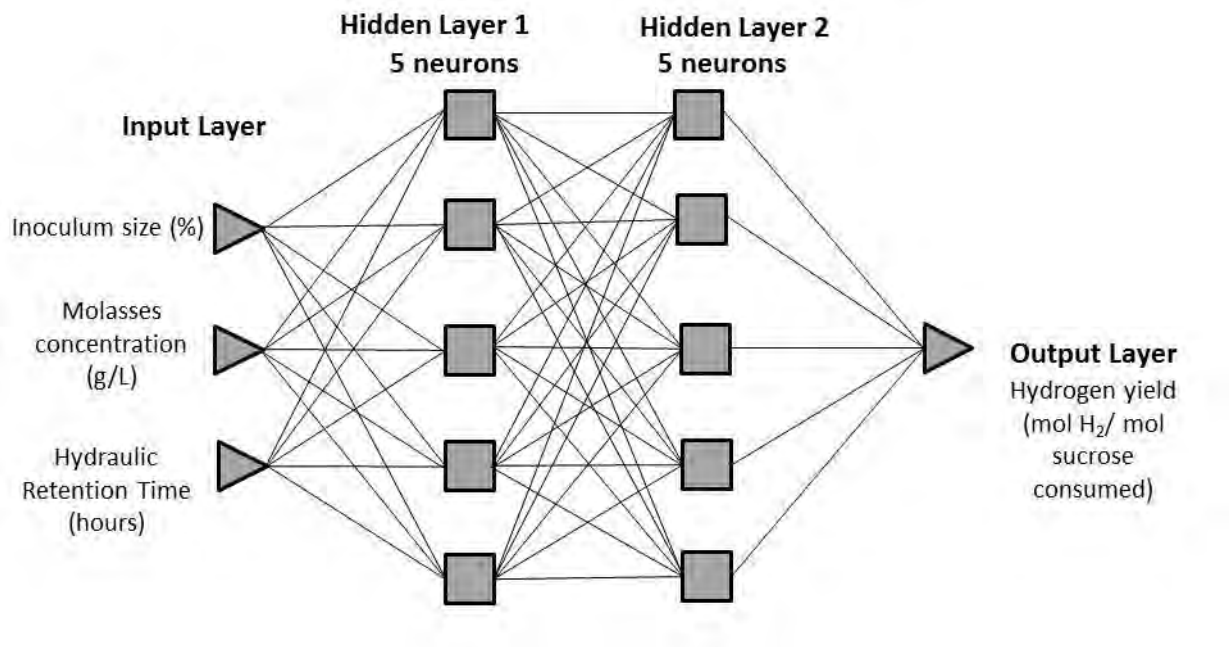


Figure 1. Topology of Neural Networks used for ANN_Model$_{80}$ and ANN_Model$_{800}$. It consists of one input layer (three neurons), two hidden layers (five neurons each) and one output layer (one neuron)

Experimental data from the Box-Behnken design were divided into training and validation sets and used to train the ANN models at 80 (ANN_Model$_{80}$) and 800 mL (ANN_Model$_{800}$). The back propagation (BP) algorithm was used for the training process (Figure 2). The Mean Square Error (MSE) for training and validation was 0.007 and 0.025, respectively for the (ANN_Model$_{80}$), 0.08 and 0.005 for training and validation respectively for (ANN_Model$_{800}$). The training was completed after 14000 and 5000 epochs for the ANN_Model$_{80}$ and ANN_Model$_{800}$, respectively.

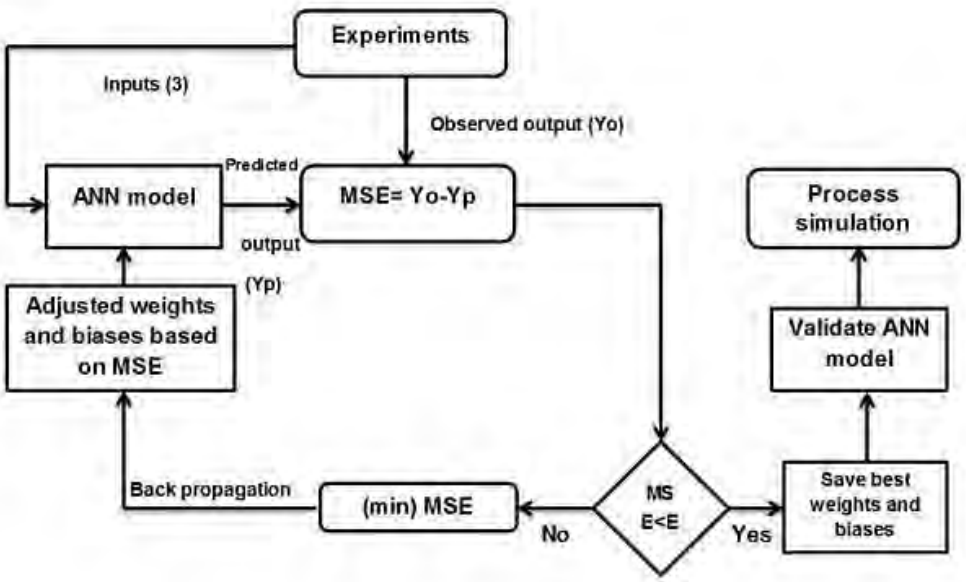


Figure 2. The back propagation training flowchart for artificial neural network. Note: Mean square error (MSE)

Sensitivity Analysis

To determine the effect of fractional changes of each input parameter on the hydrogen output, a sensitivity analysis was performed. A fractional change on each parameter was carried out in the ranges of -100 to 100% in increments of 5 while keeping other parameters at their midpoint values. The midrange value was assigned a value of 0%.

Table 3. Experimental fermentation batches with the observed and predicted hydrogen yields from RSM and ANN at 80 and 800 mL culture volumes

	Input parameters			Hydrogen yield (mol H_2/mol sucrose consumed)					
Run	Inoculum size (%)	Molasses concentration (g/L)	Hydraulic Retention Time (hours)	$Observed_{80}$	RSM_Model_{80}	ANN_Model_{80}	$Observed_{800}$	RSM_Model_{800}	ANN_Model_{800}
1	30	200	29	0.966	0.899	0,900	0.494	0.508	0,510
2	50	300	29	0.204	0.150	0,282	0.253	0.201	0,255
3	10	200	10	0	-0.041	0,081	0.047	0.042	0,050
4	50	200	48	0.492	0.533	0,493	0.335	0.340	0,338
5	30	200	29	0.832	0.899	0,900	0.528	0.508	0,510
6	10	300	29	0.371	0.294	0,374	0.220	0.108	0,221
7	10	100	29	0.442	0.496	0,444	0.414	0.467	0,415
8	30	100	48	1.15	1.037	0,879	0.808	0.691	0,782
9	30	200	29	0.897	0.899	0,900	0.503	0.508	0,510
10	50	200	10	0	-0.06	0,001	0.223	0.158	0,225
11	30	200	29	0.915	0.899	0,900	0.513	0.508	0,510
12	30	100	10	0	-0.012	0,00171	0.471	0.424	0,473
13	10	200	48	0.354	0.418	0,358	0.195	0.260	0,198
14	30	300	10	0	0.116	0,0823	0.0125	0.129	0,0180
15	50	100	29	0.656	0.732	0,658	0.457	0.569	0,300
16	30	200	29	0.883	0.899	0,900	0.501	0.508	0,510
17	30	300	48	0.113	0.125	0,114	0.212	0.260	0,121

Results and Discussion

Assessment of the Significance of RSM Models

The fitness of the RSM models was assessed using Analysis of Variance (ANOVA). Results are presented in Table 4 and 5. The coefficient of determination (R^2) is illustrative of the fraction of total deviation in the dependent variable that can be accounted for by the independent variables of the model. It is used as a measure of variance and ranges between 0 and 1. R^2 values greater than 0.70 are indicative of a good model [45]. The coefficients of determination (R^2) were 0.97 (RSM_Model_{80}) and 0.89 (RSM_Model_{800}), indicating that these models could account for 97% and 89% of variations in the observed data. The relatively low P-values of <0.0001 (RSM_Model_{80}) and 0.0120 (RSM_Model_{800}) and the high F values of 29.04 (RSM_Model_{80}) and 6.31 (RSM_Model_{800}) further elucidate the significance of these models. Moreover, an F value this high has only a 0.01 % chance that it is due to noise. The lack of fit, P value for the two models at 80 and 800 mL were 0.0365 and 0.0002 and indicated that the lack of fit was not significant in relation to the pure error. Regarding the ANOVA of coefficient of estimates, generally "Prob>F" less than 0.05 is suggestive of the significance of the model terms. From Table 4 and 5, it can be seen that the most significant of these variables for both RSM_Model_{80} and RSM_Model_{800} were Hydraulic Retention Time (C) with a P value of <0.0001 and 0.0120 followed by Molasses concentration (B) and finally the mutual interaction of Molasses concentration and hydraulic retention time (BC). The polynomial models are shown in Equations 7 and 8.

$$\text{Hydrogen yield at 80 mL} = 0.90 + 0.023A - 0.20B + 0.26C - 0.095AB + 0.034AC - 0.26BC - 0.29A^2 - 0.19B^2 - 0.39C^2 \quad (7)$$

$$\text{Hydrogen yield at 800 mL} = 0.51 + 0.049A - 0.18B + 0.100C - 2.42X10^{-3}AB - 8.989X10^{-3}AC - 0.035BC - 0.17A^2 - 2.027X10^{-3}B^2 - 0.13C^2 \quad (8)$$

where, A is the inoculum size, B is the molasses concentration and C is the hydraulic retention time (HRT).

The optimum conditions predicted by the RSM_Model_{80} for maximum hydrogen production were 34.84% inoculum size, 100g/L molasses and 41.84 hours HRT compared to 32.71% inoculum size, 100g/L molasses and 38.44 hours HRT by the RSM_Model_{800}. These conditions were determined using equation (7) and (8) for the RSM_Model_{80} and RSM_Model_{800}, respectively.

Table 4. Analysis of variance (ANOVA) of RSM_Model_{80}

Factor	Coefficient	Sum of Squares	Degrees of freedom (*df*)	Standard Error	*F* value	*p* value (probability>*F*)
Intercept or model	0.90	2.47	9	1	29.04	<0.001
A-Inoculum size (%)	0.023	0.004257	1	1	0.45	0.5235
B-Molasses concentration (g/L)	-0.20	0.31	1	1	32.45	0.0007
C-Hydraulic Retention Time (hours)	0.26	0.56	1	1	59.16	0.0001
AB	-0.095	0.036	1	0.049	3.83	0.0912
AC	0.034	0.004744	1	0.049	0.50	0.5014
BC	-0.26	0.27	1	0.049	28.72	0.0011
A^2	-0.29	0.36	1	0.047	38.28	0.0005
B^2	-0.19	0.15	1	0.047	15.71	0.0054
C^2	-0.39	0.65	1	0.047	69.31	<0.0001
Residual Error	-	0.066	7	-	-	-
Lack of fit	-	0.057	3	-	7.99	0.365
Pure Error	-	0.009452	4	-	-	-

Table 5. Analysis of variance (ANOVA) of RSM_Model_{800}

Factor	Coefficient	Sum of Squares	Degrees of freedom (*df*)	Standard Error	*F* value	*p* value (probability>*F*)
Intercept or model	0.51	0.58	9	1	6.31	0.0120
A-Inoculum size (%)	0.049	0.019	1	1	1.87	0.2133
B-Molasses concentration (g/L)	-0.18	0.26	1	1	25.75	0.0041
C-Hydraulic Retention Time (hours)	0.100	0.079	1	1	7.76	0.0271
AB	-0.002426	0.00002354	1	0.051	0.002299	0.9631
AC	-0.008989	0.0003232	1	0.051	0.032	0.8640
BC	-0.035	0.004779	1	0.051	0.47	0.5164
A^2	-0.17	0.13	1	0.049	12.39	0.0097
B^2	-0.002027	0.00001731	1	0.049	0.001691	0.9684
C^2	-0.13	0.075	1	0.049	7.37	0.0300
Residual Error	-	0.072	7	-	-	-
Lack of fit	-	0.071	3	-	134.02	0.0002
Pure Error	-	0.0007059	4	-	-	-

Assessment of the Significance of ANN Models

The Analysis of Variance on ANN models gave R^2 values of 0.99 (ANN_Model_{80}) and 0.95 (ANN_Model_{800}) as shown in Table 6. Therefore, these models were able to account for 99 and 95% of the variability in the observed data. The relatively low P-values of 0.046 and 0.324 and the high F values of 326.73 and 36.14 further elucidate the significance of these models at 80 and 800 mL respectively. The high R^2 values indicated that both models were able to abstract the relationships between the input and corresponding output. Optimized conditions for ANN models were derived from the sensitivity analysis and are shown below.

Table 6. Comparison of ANOVA for the developed RSM and ANN models

Source	Sum of Squares	df	Mean Squares	F-value	P-value	R^2
RSM_Model_{80}	2.47	9	0.27	29.04	<0.0001	0.97
RSM_Model_{800}	0.58	9	0.065	6.31	0.0120	0.89
ANN_Model_{80}	0.42	1	0.42	326.73	0.046	0.99
ANN_Model_{800}	0.23	1	0.23	36.14	0.324	0.95

df: degrees of freedom; F-value: Fisher-Snedecor distribution value; P-value: Probability value; R^2: Coefficient of determination

Assessment of the developed models on experimental data

Figure 3 (a-d) shows the one-to-one relationship between the experimental and predicted yields for both ANN and RSM models at 80 and 800 mL. The data points were scattered along or near the diagonal line thus illustrating the closeness between the observed and model predicted yields. This was more pronounced for the ANN models as compared to the RSM models. Hence, the ANN models showed a higher correlation between the observed and model predicted experimental data.

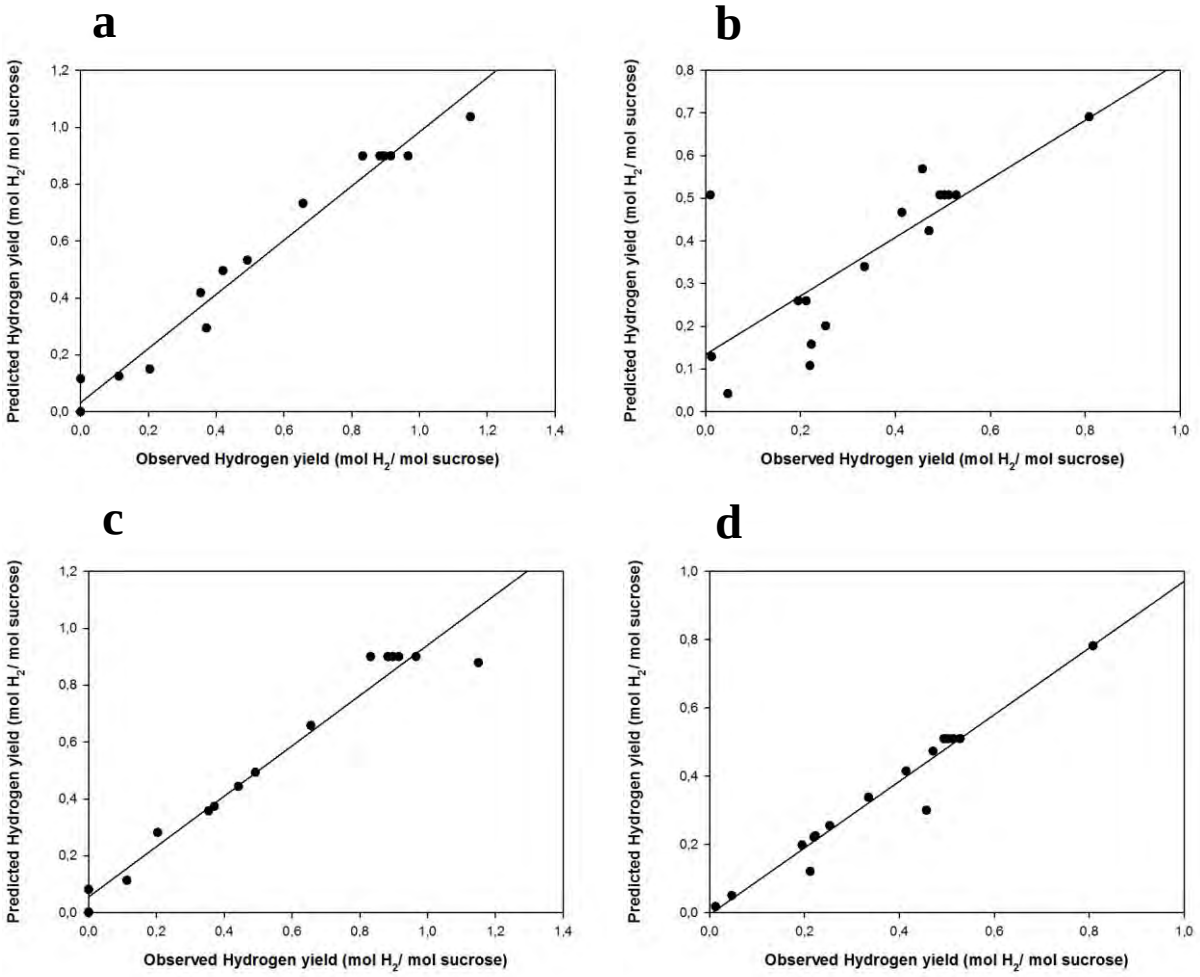


Figure 3. Predicted versus observed biohydrogen yields (mol H_2/mol sucrose consumed) (a) RSM_Model$_{80}$ (R^2=0.97), (b) RSM_Model$_{800}$ (R^2= 0.89), (c) ANN_Model$_{80}$ (R^2=0.99) and (d) ANN_Model$_{800}$ (R^2=0.95), respectively. Note: The diagonal line illustrates expectations under a one-to-one relationship between predicted and observed values

Sensitivity Analysis on RSM and ANN models

Sensitivity values display the change in the systems' output relative to change in the input. Typically, a large sensitivity to a variable suggests that the output can vary substantially with a small variation in the input parameter [46, 47]. On the other hand, a low sensitivity implies that a small variation in the output occurs even if there is a large variation in the input variable.

For the RSM models, this was done by assessing the coefficients in the polynomial equations. The polynomial coefficients showed that HRT had the most significant influence on the hydrogen yield, followed by inoculum size and molasses concentration in decreasing order. RSM_Model$_{80}$ had coefficients of 0.26 (HRT), 0.023 (Inoculum size) and 0.20 (molasses concentration) compared to

0.100 (HRT), 0.049 (inoculum size) and -0.18 (molasses concentration) for the RSM_Model$_{800}$. In addition, the interactive effect of inoculum size and HRT had the highest coefficients of 0.034 (RSM_Model$_{80}$) and -0.008989 (RSM_Model$_{800}$) for both models. These data revealed that the order of parameter sensitivity was similar at both scales (80 ml and 800ml). The RSM models showed that a slight change in HRT will significantly influence the hydrogen yield. On the other hand, the low sensitivity of molasses concentration suggests that even if a large variation occurs in this parameter within the range studied, little change would occur in the hydrogen output. Additionally, the interactive effect of inoculum size and HRT has a significant influence the hydrogen yield.

Several techniques exist for performing sensitivity analysis on ANN models. In this study, a fractional reduction analysis was used. The variations in hydrogen outputs as a response to fractional change on process inputs for the ANN_Model$_{80}$ and ANN_Model$_{800}$ is illustrated in Figure 4a and b.

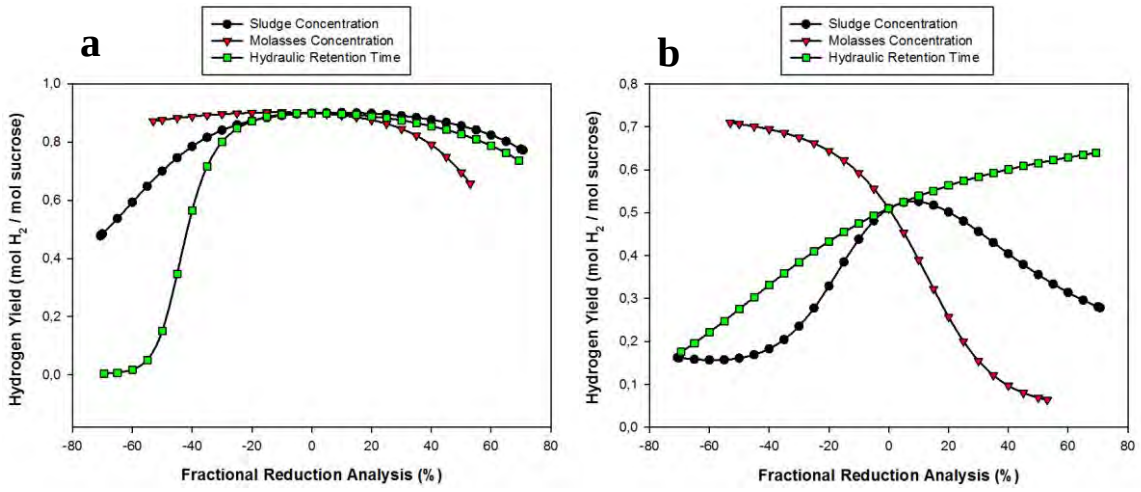


Figure 4. Fractional change of input parameters of Inoculum size, Molasses concentration and HRT (a) ANN_Model$_{80}$ and (b) ANN_Model$_{800}$ on hydrogen yield

Sensitivity analysis with the ANN_Model$_{80}$ indicated that a fractional increase in HRT from -55 to -20% (13.05 to 23.20 h) significantly enhanced hydrogen yield. However, a fractional change in the same parameter between -15 and 45% (24.65 to 42.05 h) did not significantly impact the hydrogen yield. With regards to inoculum size, it can be seen that a fractional increase from -70.7 to 10% (inoculum percentage of 8.79 to 33%) significantly increased hydrogen yield, but further fractional increases beyond 30% (39% inoculum concentration) led to a decline hydrogen yield. When considering molasses concentration, a fractional increase above 15% (operational molasses concentration of 230 g/L) resulted in a significant decline in hydrogen yield. Optimum hydrogen yield was observed between a fractional change from -53 to 10% (94-220 g/L molasses) from its midpoint

value. The impact of fractional variations of input parameters on hydrogen yield can be further observed in Figure 4a.

With regard to sensitivity analysis for the ANN_Model$_{800}$, a fractional change in HRT from -69.4 to 69.4% (0-49.13 hours) led to a significant increase in hydrogen yield. Inoculum size, on the other hand, showed that a fractional reduction from its midpoint value to 70.7% (inoculum percentage of 8.79%) resulted in a decrease in hydrogen yield. Likewise, a fractional increase from 20 to 70.7% (36-51.21% inoculum concentration) from its midpoint value negatively impacted the hydrogen yield. Optimum hydrogen yield was observed between 31.5-34.5% inoculum concentration corresponding to a fractional increase of 5 to 15% from its baseline value. With the molasses substrate, it was observed that a fractional increase from -15 to 53% (170-306 g/L of molasses) resulted in a significant decline in hydrogen yield whereas a fractional reduction from -15 to -50% (170 to 100 g/L of molasses) led to a substantial increase in hydrogen yield. These fractional changes can be observed in Figure 4b.

To compare the relative sensitivity of input parameters, the gradient of each series was obtained. Generally, the higher the slope, the larger the effect of the specific input on the corresponding output. The gradient of each series for the ANN_Model$_{80}$ were 0.0052 (HRT), 0.017 and (inoculum size) and -0.0014 and (molasses concentration) compared to 0.0034 (HRT), 0.0016 (inoculum size) and -0.0072 and (molasses concentration) obtained for the ANN_Model$_{800}$. A comparative assessment of these gradients for both ANN_Model$_{80}$ and ANN_Model$_{80}$ showed that HRT had the greatest influence on the hydrogen output followed by inoculum size and molasses concentration in decreasing order. The high slope obtained for HRT indicated that a slight change in this parameter led to a large variation in the hydrogen yield. Alternatively, the low sensitivity observed for molasses concentration implies that even if a large variation occurred in this parameter, a slight change would be observed in the hydrogen output. It is interesting to note that the sensitivities of the input parameters for both ANN models developed at the two different process scales exhibited the same order of sensitivity with HRT followed by inoculum size and molasses concentration in decreasing order. A striking observation is the consensus on the relative importance of the input parameters as shown by the sensitivity analysis with the RSM and ANN models at both scales. Both modelling algorithms revealed that HRT, inoculum size and molasses concentration in decreasing order affected biohydrogen production. This pattern was observed for both 80 and 800 ml process scales.

Previous studies have indicated that HRT significantly influenced the biohydrogen production process and the optimum reported HRT value is within the ranges of 1-6 days depending on the substrate source [10, 28, 48–52]. Generally, short HRTs are advantageous for hydrogen production, since the hydrogen-producers are able to grow and reproduce rapidly, whereas the methanogens which are hydrogen consumers require longer HRTs to proliferate [53]. The inoculum concentration has shown to impact the hydrogen yield and reports on the optimum inoculum concentration have indicated that

this parameter is dependent on the inoculum source and community adopted. A low inoculum concentration of 15% (anaerobic sludge as inoculum) was found to be optimum in a study by Whiteman and Gueguim-Kana [19] in contrast to a high concentration of between 65-75% by Prakasham *et al.* [35] using the same inoculum type. Several studies have indicated that moderately low inoculum concentrations (<10%) resulted in a decrease in the cumulative hydrogen volume [54–57]. The optimum substrate concentration for maximum hydrogen yield has been reported by several studies [6, 7, 19, 28, 55, 58]. Low substrate levels would result in a decline in the hydrogen yields observed due to rapid substrate degradation, whereas elevated substrate concentrations result in longer lag phases during fermentation [55, 58]. Studies on molasses have revealed that optimum concentration for maximum biohydrogen production ranges between 100-150 g/L [19, 55]. Optimum conditions predicted by the ANN models in this study are shown in Table 7.

Table 7. Optimized input variables with their hydrogen output from the RSM and ANN Models

Model	Inoculum size (%)	Molasses concentration (g/L)	Hydraulic Retention Time (hours)	Predicted H_2 yield (mol H_2/mol sucrose consumed)	Observed H_2 yield (mol H_2/mol sucrose consumed)	Prediction error (%)
RSM_Model$_{80}$	34.84	100	41.84	1.09	0.990	10.10
RSM_Model$_{800}$	32.71	100	38.44	0.720	0.700	2.86
ANN_Model$_{80}$	33	100	29	0.870	0.890	2.25
ANN_Model$_{800}$	33	100	40	0.730	0.710	2.82

Comparison of RSM and ANN for prediction accuracy and optimization efficiency

The percentage error difference (between the experimental yield and the predicted yields) and the coefficients of determination (R^2 values) for the RSM and ANN models are presented in Table 7. Of the four models developed, the RSM_Model$_{80}$ and ANN_Model$_{80}$ predicted the maximum hydrogen yield of 1.09 and 0.870 mol H_2/ mol sucrose consumed whereas RSM_Model$_{800}$ and ANN_Model$_{800}$ predicted hydrogen yields of 0.720 and 0.730 mol H_2/ mol sucrose consumed, respectively. Experimental validations of RSM_Model$_{80}$ and ANN_Model$_{80}$ gave hydrogen yields of 0.990 and 0.890 mol H_2/ mol sucrose consumed. On the other hand, the RSM_Model$_{800}$ and ANN_Model$_{800}$ gave experimental hydrogen yields of 0.700 and 0.710, respectively. The lower hydrogen yields obtained at 800 mL compared to the 80 mL process scale may be due to poor mass transfer that occurs within larger vessels (Schmidt, 2005). Slight variations were observed between the optimized conditions for the RSM_Model$_{800}$ and ANN_Model$_{800}$ with predicted hydrogen yields of 0.720 and 0.730 mol H_2/

mol sucrose, respectively. The high level of similarity between the hydrogen yields predicted by the RSM_Model$_{800}$ and the ANN_Model$_{800}$ demonstrates the modelling efficiency of the developed models.

Although the RSM_Model$_{80}$ gave a higher hydrogen yield compared to the other three models, its prediction error (10.10%) was relatively higher compared to these models (2.25, 2.82% and 2.86 for the ANN_Model$_{80}$, ANN_Model$_{800}$ and RSM_Model$_{800}$, respectively). These results suggest that the ANN models were much more accurate for prediction on unseen data compared to the RSM models. A slight variation was observed among the models on their predicted optimum operational parameters. For example, the RSM_Model$_{80}$ predicted an optimum inoculum size, molasses concentration and HRT of 34.84%, 100 g/L and 41.84 h whereas the RSM_Model$_{800}$ predicted a 32.41% inoculum, 100 g/L molasses concentration and 38.44 h HRT. The slightly higher inoculum size with a longer HRT for the RSM_Model$_{80}$ may account for the increase in observed yield. On the other hand, the ANN_Model$_{80}$ and ANN_Model$_{800}$ predicted the same optimum inoculum size, molasses concentration of 33%, and 100 g/L with a difference in the HRT. ANN_Model$_{80}$ predicted an HRT of 29 h compare to 40 h predicted by the ANN_Model$_{800}$. With regards to these models, a lower mass transfer observed at the larger process volume (800 mL) compared to the lower process volume (80 mL) may have contributed to the slightly higher yield for the ANN_Model$_{80}$ compared to the ANN_Model$_{800}$. These results suggest that although the RSM_Model$_{80}$ displayed the highest predicted hydrogen yield, the ANN_Model$_{80}$ exhibited a higher prediction accuracy on unknown data.

The comparative predictive superiority of ANN over RSM has been reported in various studies [6, 19, 45]. Generally, ANN models exhibit higher modelling and optimization abilities. Additionally, ANN has a greater generalization capability whereby it can approximate the majority of non-linear and quadratic functions whereas RSM is mostly suitable for quadratic estimations. Desai *et al.* [45] compared ANN and RSM for fermentation medium optimization for scleroglucan production and showed that ANN had a greater generalization ability than RSM. Similarly, Whiteman and Gueguim-Kana [19] comparatively evaluated RSM and ANN for biohydrogen production and revealed ANNs' superiority over RSM.

Potential Impact of process scale on biohydrogen yield

The observed hydrogen yields at both scales are shown in Table 3. Maximum hydrogen yield obtained at the 80 mL process volume was 1.15 mol H_2/mol sucrose consumed on inputs of 48 hours (HRT), 30% inoculum size and 100 g/L molasses concentration. This result was comparable to the yield at 800 mL (0.808 mol H_2/mol sucrose consumed) under similar conditions. No hydrogen production was observed at 80 mL process volume when the HRT was 10 hours. On the other hand, low hydrogen

yields (0.050, 0.225, 0.473 and 0.0180 mol H_2/mol sucrose consumed) were observed at a process volume of 800 mL with an HRT of 10 hours.

A paired sample *t-test* was performed on the experimentally observed hydrogen yields at both process scales (80 and 800 mL). The average mean and *t-test* statistic was calculated using MS excel 2010 (Microsoft, Inc, USA). The significance was noted when $p < 0.05$. Results showed that the average mean between the hydrogen yields at 80 and 800 mL was 0.123 ± 0.062 mol H_2/mol sucrose consumed ($t_{16,17} = 1.99$, $p = 0.064$). These results indicate that there was no significant difference ($p>0.05$) between biohydrogen yields across scales. However the 80 mL process volume exhibited a slightly higher yield compared to the 800 mL process volume. This result may be attributed to the mixing efficiency and thus mass transfer within the reactors. Generally, lower scales would achieve higher mixing efficiency and mass transfer as opposed to larger vessels [38].

In contrast to chemical reactors, the scale up of microbial fermentation processes is significantly challenged by the reproducibility in yield as the scale increases. This is due to the physiology of growth and thus product formation within the reactor [59]. A frequent catastrophic scenario is the inability to maintain physiological conditions from lab scale to a larger scale. Variations from physiological uniformity that is initiated by environmental changes may induce stress on the microorganisms. These stress conditions can reduce the cells' physiological functions, thereby resulting in a lower product yield [59]. The lower yield observed at 800 mL compared to the lower process volume of 80 mL may be as result of reduced mass transfer which was previously described by Formenti *et al*. [60]. Scale up poses various challenges since large vessels are substantially more heterogeneous compared to smaller vessels [38, 61]. Shuler and Kargi [61] stated that even when geometrically similar vessels are employed, it appears impossible to retain the same level of shear, mixing time, and mass transfer from the small vessel to the larger vessel because power and mixing constraints generally fail to scale in a linear manner. The approaches to address such challenges include using multiple small reactors as opposed to one large reactor. Rouf *et al.* [62] compared a 6000 L vessel to six 1000 L bioreactors of equal size and revealed that although the production costs for the 6000 L were lower than that of using multiple reactors, the downstream processing of using multiple reactors was much cheaper. Efficient scale up requires the application of deterministic models such as computational fluid dynamics to achieve similar mixing efficiencies between different scales. Statistical data from this study showed no significant yield difference across both scales. Thus, a scale down from 800 to 80 ml could reduce the bioprocess screening time and model development cost.

Comparative Assessment of the Optimized Models for Biohydrogen production at Semi-pilot scale

As shown in Figure 5(a-d), short lag phases were observed for all optimized conditions. The observed lag phases for the semi-pilot bioprocesses were 2 h, 5 h, 6 h and 4 h for RSM_Model$_{80}$_8L, RSM_Model$_{800}$_8L, ANN_Model$_{80}$_8L and ANN_Model$_{800}$_8L, respectively. Short lag phases are desirable. Generally, short lag phases indicate that the microorganisms adapted well to the medium. Lab scale studies on biohydrogen production have shown lag phase times of 10, 11 and 20 h [55, 63, 64] compared to semi-pilot and pilot scale experiments with higher lag phases of 19 and 24 h [33, 65].

With regards to the RSM_Model$_{80}$_8L, the exponential phase lasted from 2 to 19 h. The maximum hydrogen fraction and cumulative volume of hydrogen were 46.59% and 3180.48 mL, respectively. Conversely, the RSM_Model$_{800}$_8L had an exponential phase of 5 to 18 h with a maximum hydrogen fraction and cumulative volume of hydrogen of 38.96% and 2618.12 mL. An exponential phase of 6 to 20 h was observed for the ANN_Model$_{80}$_8L with a maximum hydrogen fraction and cumulative volume of 45.04% and 2929.40 mL. Similarly, the ANN_Model$_{800}$_8L had an exponential phase that lasted from 4 to 19 h with a maximum hydrogen fraction and cumulative volume of 44.01% and 2876.93 mL, respectively. Zhou *et al*. [63] indicated that the exponential growth phase for hydrogen production lasted approximately 21.2 h in lab scale experiments.

Hydrogen production generally takes place during the exponential phase of growth in microorganisms [66]. Studies have shown that peak hydrogen fraction may differ depending on the process time, substrate type and vessel size in semi-pilot and pilot scale experiments [67]. In a study by Ren *et al.* [68], peak hydrogen fraction of 52% was observed when using a 2000 L pilot-scale bioreactor fed with molasses and operated for 200 days. Similarly, in a study by Chang *et al.* [69], a 12 L bioreactor that was operated for 95 days gave a peak hydrogen fraction of 40.4%. Likewise, Lin *et al.* [67] investigated hydrogen production using a 400 L bioreactor that was operated for 65 days and gave a peak hydrogen fraction of 37.8% using sucrose as the substrate.

Maximum hydrogen yields for the semi-pilot processes obtained in this study were 0.89, 0.76, 0.81 and 0.78 mol H_2/ mol sucrose consumed for RSM_Model$_{80}$_8L, RSM_Model$_{800}$_8L, ANN_Model$_{80}$_8L and ANN_Model$_{800}$_8L, respectively. Although, these data exhibit high similarity to their corresponding lab scale, slight variations between the semi-pilot scale and lab scale is observed. These data may suggest that hydrogen production is influenced by the process scale, in line with previous reported studies. For example, Chang *et al.* [69] obtained a hydrogen yield of 1.40 mol H_2 mol/ mol glucose when a 12 L bioreactor was used compared to 1.04 mol H_2 mol/ mol sucrose by Lin *et al.* [67] when a 400 L bioreactor was used. In addition, Faloye *et al*. [33] reported a yield of 2.07 mol H_2 mol/ mol glucose using a 7 L bioreactor compared to 2.91 mol H_2 mol H_2/ mol by Masset *et al*. [70] when a 20 L bioreactor was used. These studies suggest that hydrogen production is

dependent on scale which may be attributed to conventional scale-up challenges encountered during fermentation process development. After 18 h HRT, the pH values were recorded and were shown to decrease from an initial value of 6.50 to 5.82, 5.94, 5.84, and 5.89 for the RSM_Model$_{80}$_8L, RSM_Model$_{800}$_8L, ANN_Model$_{80}$_8L and ANN_Model$_{800}$_8L, respectively. This decrease may be due to the metabolic processes that resulted in acid formation [71]. Optimum pH for biohydrogen production has been reported in the range of 5.5-6 [54, 71, 72].

Hydrogen production in the RSM_Model$_{80}$_8L batch lasted for 19 h and corresponded to a substrate degradation efficiency and final pH of 55% and 4.39, respectively, whereas the RSM_Model$_{800}$_8L showed a hydrogen production that lasted 18 h with a substrate degradation of 44% and a final pH of 5.30. Similar to the RSM based semi-pilot experiments, the ANN derived semi-pilot experiments showed a relatively short hydrogen production phase and low substrate degradation efficiencies. For example ANN_Model$_{80}$_8L had a hydrogen phase that lasted 20 h with a substrate degradation efficiency and final pH of 43% and 4.48, respectively, and ANN_Model$_{800}$_8L showed a hydrogen phase of 19 h and a corresponding substrate degradation efficiency and final pH of 47% and 4.35, respectively. The slight deviations in terms of substrate degradation efficiency between the two ANN optimized experiments at semi-pilot scale are due to the longer HRT predicted by the ANN_Model$_{800}$. Nonetheless, the slightly higher HRT predicted by the ANN_Model$_{800}$_8L did not significantly impact on the maximum hydrogen yields observed for both models (0.81 and 0.78 mol H_2/ mol sucrose consumed for the ANN_Model$_{80}$_8L and ANN_Model$_{800}$_8L, respectively). Large similarities were observed across all four optimized conditions. For instance, the various phases of hydrogen production display similarity in terms of fermentation time and hydrogen fractions produced. Other similarities include the substrate degradation efficiency and final pH values observed. This highlights that all four models developed in this study were efficient for predicting biohydrogen production on inputs of inoculum size, molasses concentration and HRT.

Slight changes in operational parameters such as pH may have adverse effects on the hydrogen-producers [71]. Faloye *et al.* [33] reported a peak hydrogen production of 56.8% accompanied by a shorter lag phase when the pH was controlled compared to uncontrolled pH (49%) in a semi-pilot reactor (7 L). The decline phases observed for all four batches in this study may be due to the shift in metabolic process from acidogenic to solventogenic fermentation [71]. Solventogenic fermentation results in the production of VFAs such as acetate, butyrate and ethanol which leads to a decrease in the pH, thus resulting in a shift in the cells' metabolism. Several attempts have been made to reduce the production of VFAs [32, 73, 74]. These methods include regulating the pH during the process or the addition of buffers at the start of the process thereby maintaining the pH which promotes growth of hydrogen-producing bacteria [33, 53, 73, 74]. Changes in the operational setpoints promotes the growth of hydrogen-consuming bacteria such as homoacetogens which are chemolithoautotrophic and

use hydrogen and carbon dioxide for growth thereby producing acetate when the peak hydrogen production occurs [71].

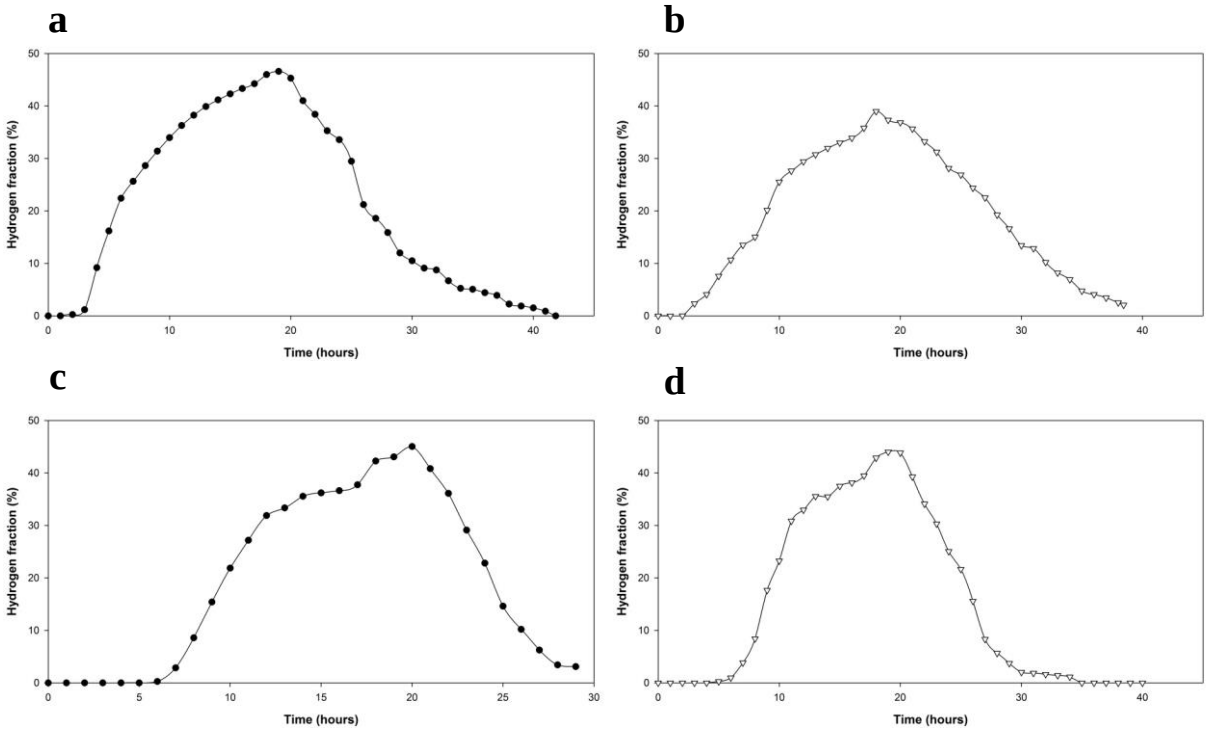


Figure 5. Evolution of hydrogen fraction for the optimized runs: (a) RSM_Model$_{80}$_8L, (b) RSM_Model$_{800}$_8L , (c) ANN_Model$_{80}$_8L and (d)ANN_Model$_{800}$_8L

Microbial community analysis using Next Generation Sequencing

To determine the microbial community involved within the hydrogen production process, next generation sequencing (NGS) was performed on the ANN_Model$_{80}$_8L. Results based on phylum classification are shown in Figure 6.

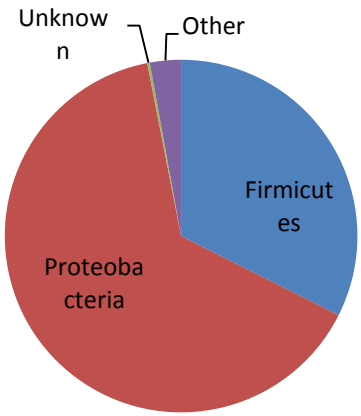


Figure 6. Taxonomic assignment of bacteria based on Phylum classification

As shown in Figure 6, bacteria in the phylum Proteobacteria comprise the majority of microorganisms present (64.51%), followed by the Firmicutes (32.41%), other (2.85%) and unknown microbes (0.23%). Previous studies on hydrogen production have shown that Firmicutes dominate hydrogen-producing communities followed by Proteobacteria. This result indicated that some microbes were able to survive the inoculum heat pretreatment. Heat treatment does not completely select for hydrogen-producing bacteria. For instance, hydrogen-producers that do not form endospores include bacteria such as *Enterobacter* spp., *Klebsiella* spp. and *Citrobacter* spp. and were shown to survive pretreatments [75–77]. Kraemer and Bagley [78] stated that vegetative cells are not completely inhibited by heat treatment which is highly dependent on whether the inoculum source is dry or wet in addition to the time and temperature of exposure. Studies on inoculum pretreatment have reported the presence of non-spore-forming bacteria [79–81].

Presumptive hydrogen-producing microorganisms detected in this study were members of the genus *Clostridium*, *Enterobacter* and *Klebsiella*. At species level it was shown that *Clostridium bifermentans*, *Clostridium butyricum*, *Enterobacter clocae* and *Klebsiella pneumonia* were present in this system. Generally, major hydrogen-producers are found within the genus *Clostridium*. These microorganisms are Gram-positive, rod-shaped, strictly anaerobic and form endospores that allow them to survive extreme conditions [82, 83]. The majority of studies on biohydrogen production have revealed the presence of microorganisms within this genus [84]. In a study by Wang *et al.* [85], a pure culture of *C. bifermentans* was used as the inoculum to digest wastewater sludge and gave a hydrogen yield of 0.9 mmol-H_2/g-dried solids. Likewise, *C. butyricum* has shown to be excellent for hydrogen production with reported yields of 0.22 and 2.9 mol H_2/ mol hexose using pure cultures [66, 86].

On the other hand, *Enterobacter* spp. are Gram-negative, rod-shaped and facultative anaerobes [83]. *Enterobacter cloacae* was detected within this community. Members of the genus *Enterobacter* are known for their hydrogen-producing capabilities [87–89]. Other hydrogen-producers found in this study were members within the genus *Klebsiella*. These microbes are Gram-negative, facultative anaerobes and are rod-shaped. The presence of *Klebsiella* spp. has been shown in several hydrogen-producing reactors [76, 90, 91]. Numerous studies have reported that facultative anaerobes such as *Enterobacter* spp. and *Klebsiella* spp. play an important role in the utilization of excess oxygen present within the bioreactor, thus creating anaerobic conditions required for hydrogen production [76, 87, 88].

Conclusion

In this study, RSM and ANN models were implemented for fermentative hydrogen processes across two scales. Process scales of 80 and 800 mL were considered with inputs of inoculum size, molasses concentration and HRT with hydrogen yield as the corresponding output. Results showed that the R^2 value across all scales were relatively high. ANN based models gave R^2 values of 0.99 (ANN_Model_{80}) and 0.95 (ANN_Model_{800}) whereas RSM based models gave R^2 values of 0.97 (RSM_Model_{80}) and 0.89 (RSM_Model_{800}). ANN_Model_{80} displayed a higher accuracy for prediction on unknown data compared to the RSM_Model_{80}, RSM_Model_{800} and ANN_Model_{800}, respectively. ANN models are known for its higher generalization in addition to its modelling ability. Furthermore, a sensitivity analysis was performed on both the RSM and ANN based models. Results revealed that HRT, inoculum size and molasses concentration influenced the biohydrogen production process in decreasing order. The obtained data revealed that variation in process scale within the studied window did not impact on the efficiency of ANN or RSM derived process models. An assessment of the optimized models at semi-pilot scale gave a relatively similar hydrogen production profile with peak fractions within the range of 38.96% and 46.59%. Next Generation Sequencing (NGS) revealed the presence of presumptive hydrogen-producing microorganisms which were members within the genus *Clostridia*, *Enterobacter* and *Klebsiella*. These findings suggested that miniaturization of experiments for biohydrogen model development does not significantly impact on the model accuracy. This is of paramount importance as it reduces the process developmental time and resources towards the commercialization of biohydrogen.

References

1. Levin DB, Pitt L, Love M (2004) Biohydrogen production: prospects and limitations to practical application. Int J Hydrogen Energy 29:173–185
2. Nath K, Das D (2011) Modeling and optimization of fermentative hydrogen production. Bioresour Technol 102:8569–8581
3. Belafi-Bakó K, Bakonyi P, Nemesto´thy N, Pientka Z (2010) Biohydrogen production in integrated system Desalination and Water Treat 14:116–118
4. Lay CH, Wu JH, Hsiao CL, Chang JJ, Chen CC, Lin CY (2010) Biohydrogen production from soluble condensed molasses fermentation using anaerobic fermentation. Int J Hydrogen Energy 35:13445–13451
5. Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H (2006) Enhanced hydrogen production from glucose using ldh-and frd-inactivated *Escherichia coli* strains. App Microbiol Biotechnol 73:67–72

6. Wang JL, Wan W (2009a) Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology. Int J Hydrogen Energy 34:255–261
7. Wang JL, Wan W (2009b) Application of desirability function based on neural network for optimizing biohydrogen production process. Int J Hydrogen Energy 34:1253–1259
8. Wang JL, Wan W (2009c) Factors influencing fermentative hydrogen production:A review. Int J Hydrogen Energy 34:799– 811
9. Wang JL, Wan W (2008a) The effect of substrate concentration on biohydrogen production by using kinetic models. Sci China Ser B 51(11):1110–1117
10. Kim SH, Han SK, Shin HS (2004) Feasibility of biohydrogen production by anaerobic co-digestion of food waste and sewage sludge. Int J Hydrogen Energy 29:1607–1616
11. Hung CH, Chang YT, Chang YJ (2011) Roles of microorganisms other than Clostridium and *Enterobacter* in anaerobic fermentative biohydrogen production systems–A review. Bioresour Technol 102(18):8437–8444
12. Li CL, Fang HHP (2007) Fermentative hydrogen production from wastewater and solid wastes by mixed cultures. Crit Rev Env Sci Technol 37:1–39
13. Wang JL, Wan W (2008b) Comparison of different pretreatment methods for enriching hydrogen-producing cultures from digested sludge. Int J Hydrogen Energy 33:2934–2941
14. Cheong DY, Hansen CL (2006) Bacterial stress enrichment enhances anaerobic hydrogen production in cattle manure sludge. App Microbiol Biotechnol 72:635–643
15. Hu B, Chen SL (2007) Pretreatment of methanogenic granules for immobilized hydrogen fermentation. Int J Hydrogen Energy 32:3266–3273
16. Zhu H, Beland M (2006) Evaluation of alternative methods of preparing hydrogen producing seeds from digested wastewater sludge. Int J Hydrogen Energy 31:1980–1988
17. An D, Li Q, Wang X, Yang H, Guo L (2014) Characterization on hydrogen production performance of a newly isolated *Clostridium beijerinckii* YA001 using xylose. Int J Hydrogen Energy 39:19928–19936
18. Moodley P, Gueguim-Kana EB (2015) Optimization of xylose and glucose production from sugarcane leaves (*Saccharum officinarum*) using hybrid pretreatment techniques and assessment for hydrogen generation at semi-pilot scale. Int J Hydrogen Energy 40:3859–3867
19. Whiteman JK, Gueguim-Kana EB (2014) Comparative Assessment of the Artificial Neural Network and Response Surface Modelling Efficiencies for Biohydrogen Production on Sugar Cane Molasses. Bioenergy Res 7(1):295–305
20. Cheeseman, O (2006) Environmental Impacts of Sugar Production: The Cultivation and Processing of Sugarcane and Sugar Beet, 1st edn. London, UK, pp 7–8
21. Beshay U, Moreira M (2005) Production of alkaline protease with *Teredinobacter turnirae* in controlled fed-batch fermentation. Biotechnol Lett 27:1457–1460

22. He X, Guo X, Liu N, Zhang B (2007) Ergosterol production from molasses by genetically modified *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 75:55–60
23. Escamilla-Alvarado C, Rios-Leal E, Ponce-Noyola MT, Poggi-Varaldo HM (2012) Gas biofuels from solid substrate hydrogenic–methanogenic fermentation of the organic fraction of solid municipal wastes. Process Biochem 47(11):1572–1587
24. Lotfy, WA, Khaled M, Ehab R et al (2007) Citric acid production by a novel *Aspergillus niger* isolate: II. Optimization of process parameters through statistical experimental designs. Bioresour Technol 98:3470–3477
25. Mandenius C, Brundin A (2008) Review: biocatalysts and bioreactor design–bioprocess optimization using design-of-experiments methodology. Biotechnol Progress 24:1191–1203
26. Nasr N, Hafez H, El Naggar MH, Nakhla G (2013) Application of artificial neural networks for biohydrogen production. Int J Hydrogen Energy 38:3189–3195
27. Sewsynker Y, Gueguim-Kana EB, Lateef A (2015) Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs). Biotechnol Biotec Eq 29(6):1208–1215
28. Sekoai PT, Gueguim-Kana EB (2013) A two-stage modelling and optimization of biohydrogen production from a mixture of agro-municipal waste. Int J Hydrogen Energy 38:8657–8663
29. Wang G, Mu Y, Yu HQ (2005) Response surface analysis to evaluate the influence of pH, temperature and substrate concentration on the acidogenesis of sucrose-rich wastewater. Biochem Eng J 23:175–184
30. Mu Y, Zheng XJ, Yu HQ (2009) Determining optimum conditions for hydrogen production from glucose by an anaerobic culture using response surface methodology (RSM). Int J Hydrogen Energy 34:7959–7963
31. Wang JL, Wan W (2008c) Optimization of fermentative hydrogen production process by response surface methodology. Int J Hydrogen Energy 33:6976–6984
32. Faloye FD, Gueguim-Kana EB, Schmidt S (2013) Optimization of hybrid inoculum development techniques for biohydrogen production and preliminary scale up. Int J Hydrogen Energy 38:11765–11773
33. Faloye FD, Gueguim-Kana EB, Schmidt S (2014) Optimization of biohydrogen inoculum development via a hybrid pH and microwave treatment technique-Semi pilot scale production assessment. Int J Hydrogen Energy 39:5607–5616
34. Rosales-Colunga LM, Garcia RG, Rodriguez AD (2010) Estimation of hydrogen production in genetically modified *E.coli* fermentations using an artificial neural network. Int J Hydrogen Energy 35:13186–13192

35. Prakasham RS, Sathish T, Brahmaiah P (2011) Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. Int J Hydrogen Energy 36:4332–4339
36. Shi Y, Gai G, Zhao X et al (2010) Back propagation neural network (BPNN) simulation model and influence of operational parameters on hydrogen bio-production through integrative biological reactor (IBR) treating wastewater. Bioinformatics and Biomedical Engineering (iCBBE), 4th International Conference, Chengdu (China)
37. Mullai P, Yogeswari MK, Sridevi K, Ross PR (2013) Artificial neural network (ANN) modeling for hydrogen production in continuous anaerobic sludge blanket filter (ASBF). Singaporean J Sci Res 5(1):1–7
38. Schmidt, FR (2005) Optimization and scale up of industrial fermentation processes. Appl. Microbiol Biotechnol 68:425–435
39. Chong ML, Rahim RA, Shirai Y, Hassan MA (2009a) Biohydrogen production by Clostridium butyricum EB6 from palm oil mill effluent. Int J Hydrogen Energy 34:764–771
40. Orsini M, Romano-Spica V (2001) A microwave-based method for nucleic acid isolation from environmental samples. Lett App Microbiol 33:17–20
41. Muyzer G, Teske A, Wirsen CO, Jannasch HW (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. Arch Microbiol 164:165–172
42. CLC Genomics Workbench 8.5.1 http://www.clcbio.com/products/clc-genomics-workbench
43. Myers RH, Montgomery DC (1995) Response surface methodology:process and product optimization using designed experiments, 1st edn. Wiley-Interscience, USA
44. Frankfort-Nachmias C, Leon-Guerrero A (2011) Social statistics for a diverse society, 6th edn. Pine Forge, USA, pp 1–537
45. Desai KM, Survase SA, Saudagar PS et al (2008) Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. J Biochem Eng 41:266–273
46. Gunawan R, Cao Y, Petzold L et al (2005) Sensitivity analysis of discrete stochastic systems. Biophys J 288(4):2530–2540
47. Shojaeefard MH, Akbari M, Tahani M et al (2013) Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminum to brass. Adv Mater Sci Eng 2013:1–7
48. Chen C, Lin C, Chang J (2001) Kinetics of hydrogen production with continuous anaerobic cultures utilizing sucrose as the limiting substrate. App Microbiol Biotechnol 57:56–64
49. Thanwised P, Wirojanagud W, Reungsang A (2012). Effect of hydraulic retention time on hydrogen production and chemical oxygen demand removal from tapioca wastewater using

anaerobic mixed cultures in anaerobic baffled reactor (ABR). Int J Hydrogen Energy 37:15503–15510

50. Zhang H, Bruns MA, Logan BE (2006) Biological hydrogen production by *Clostridium acetobutylicum* in an unsaturated flow reactor. Water Res 40:728–734
51. Kim DH, Kim SH, Kim KY, Shin HS (2010) Experience of a pilot-scale hydrogen producing anaerobic sequencing batch reactor (ASBR) treating food waste. Int J Hydrogen Energy 35:1590–1594
52. Tawfik A, El-Qelish M (2012) Continuous hydrogen production from co-digestion of municipal food waste and kitchen wastewater in mesophilic anaerobic baffled reactor. Bioresour Technol 114:270–274
53. Liu D, Zeng RJ, Angelidaki I (2008) Effects of pH and hydraulic retention time on hydrogen production versus methanogenesis during anaerobic fermentation of organic household solid waste under extreme-thermophilic temperature (70 °C). Biotechnol Bioeng 100:1108–1114
54. Kotay SM, Das D (2006) Microbial hydrogen production with Bacillus coagulans IIT-BT S1 isolated from anaerobic sewage sludge. Bioresour. Technol. 98(6), 1183–1190
55. Wang X, Jin B (2009) Process optimization of biological hydrogen production from molasses by a newly isolated *Clostridium butyricum* W5. J Biosci Bioeng 107(2):138–144
56. Bakonyi P, Nemestóthy N, Lövitusz , Bélafi-Bakó K (2011) Application of Plackett–Burman experimental design to optimize biohydrogen fermentation by E. coli (XL1-BLUE) Int J Hydrogen Energy 36:13949–13954
57. Veena T, Tiwari KL, Quraishi A, Jadhav SK (2012) Biohydrogen production from rice mill effluent. J Appl Sci Environ Sanit 7(4): 237–240
58. Yossan S, O-Thong S, Prasertsan P (2012) Effect of initial pH, nutrients and temperature on hydrogen production from palm oil mill effluent using thermotolerant consortia and corresponding microbial communities. Int J Hydrog Energy 37(18):13807–13814
59. Votruba J, Sobotka M (1992) Physiological similarity and bioreactor scale-up. Folia Microbiol 37:331–345
60. Formenti LR, Nørregaard A, Bolic A, Hernandez DQ, Hagemann T, Heins AL, Larsson H, Mears L, Mauricio-Iglesias M, Krühne U, Gernaey KV (2014) Challenges in industrial fermentation technology research. Biotechnol J 9:727–738
61. Shuler ML, Kargi F (2002) Bioprocess engineering: Basic concepts. Upper Saddle River, NJ: Prentice Hall PTR
62. Rouf SA, Moo-Young M, Scharer JM, Douglas PL (2000) Single versus multiple bioreactor scaleup: economy for high-value products. Biochem Eng J 6:25–31

63. Zhou P, Elbeshbishy E, Nakhla G (2013) Optimization of biological hydrogen production for anaerobic co-digestion of food waste and wastewater biosolids. Bioresour Technol 130:710–718
64. Logan B, Oh SE, Kim IK, Van Ginkel SW (2002) Biological hydrogen production measured in batch anaerobic respirometers. Environ Sci Technol 36:2530–2535
65. Lee YW, Chung J (2010) Bioproduction of hydrogen from food waste by pilot-scale combined hydrogen/methane fermentation. Int J Hydrogen Energy 35(21):11746–11755
66. Chong ML, Sabaratnam V, Shirai Y, Hassan MA (2009b) Biohydrogen production from biomass and industrial wastes by dark fermentation. Int J Hydrogen Energy 34(8):3277–3287
67. Lin CY, Wu SY, Lin PJ, Chang JS, Hung CH, Lee KS et al (2011) A pilot-scale high-rate biohydrogen production system with mixed microflora. Int J Hydrogen Energy 36:8758–8794
68. Ren N, Li J, Li B, Wang Y, Liu S (2006) Biohydrogen production from molasses by anaerobic fermentation with a pilot-scale bioreactor system. Int J Hydrogen Energy 31:2147–2157
69. Chang S, Li J, Liu F (2011) Continuous biohydrogen production from diluted molasses in an anaerobic contact reactor. Front Environ Sci Eng China 5(1):140–148
70. Masset J, Calusinska M, Hamilton C, Hiligsmann S, Joris B, Wilmotte A et al (2012) Fermentative hydrogen production from glucose and starch using pure strains and artificial co-cultures of *Clostridium* spp. Biotechnol Biofuels 5(1):1–15
71. Khanal SK, Chen WS, Li L, Sung S (2004) Biological hydrogen production: Effects of pH and intermediate products. Int J Hydrogen Energy 29(11):1123–1131
72. Lin CY, Lin CY, Wu JH, Chen CC (2007) Effect of a thermal pretreatment of influent on the fermentative hydrogen production from molasses. J Environ Eng Manage 17(2):117–122
73. Zagrodnik R, Laniecki M (2015) The role of pH control on biohydrogen production by single stage hybrid dark- and photo-fermentation. Bioresour Technol 194:187–195
74. Ciranna A, Santala V, Karp M (2012) Enhancing biohydrogen production of the alkalithermophile *Thermobrachium celere*. Int J Hydrogen Energy 37:5550–5558
75. Nakashimada Y, Rachman MA, Kakizono T, Nishio N (2002) Hydrogen production of *Enterobacter aerogenes* altered by extracellular and intracellular redox state. Int J Hydrogen Energy 27(11-12):1399–1405
76. Maintinguer SI, Fernandes BS, Duarte ICS, Saavedra NK, Adorno MT, Varesche MBA (2011) Fermentative hydrogen production with xylose by *Clostridium* and *Klebsiella* species in anaerobic batch reactors. Int J Hydrogen Energy 36:13508–13517
77. Oh YK, Seol EH, Kim JR, Park S (2003) Fermentative biohydrogen production by a new chemoheterotrophic bacterium *Citrobacter* sp. Y19. Int J Hydrogen Energy 28(12):1353–1359

78. Kraemer JT, Bagley DM (2007) Improving the yield from fermentative hydrogen production. Biotechnol Lett 29:685–695
79. Ahn Y, Park EJ, Oh YK, Park S, Webster G, Weightman AJ (2005) Biofilm microbial community of a thermophilic trickling biofilter used for continuous biohydrogen production. FEMS Microbiol Lett 249(1):31–38
80. Iyer P, Bruns MA, Zhang HS, Van Ginkel S, Logan BE (2004) H2-producing bacterial communities from a heat-treated soil inoculum. Appl Microbiol Biotechnol 66(2):166–173
81. Kim DH, Jang S, Yun YM, Lee MK, Moon C, Kang WS, Kwak SS, Kim MS (2014) Effect of acid-pretreatment on hydrogen fermentation of food waste: Microbial community analysis by next generation sequencing. Int. J Hydrogen Energy 39:16302–16309
82. Nandi R, Sengupta S (1998) Microbial production of hydrogen— An overview, Crit Rev Microbiol 24(1): 61–84
83. Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST (1994) Bergey's Manual of Determinative Bacteriology, Williams and Willkins, Baltimore
84. Elsharnouby O, Hafez H, Nakhla G, Naggar MHE (2013) A critical literature review on biohydrogen production by pure cultures. Int J Hydrogen Energy 38: 4945–4966
85. Wang CC, Chang CW, Chu CP, Lee DJ, Chang BV, Liao CS (2003) Producing hydrogen from wastewater sludge by *Clostridium bifermentans*. J Biotechnol 102(1):83–92
86. Chen WM, Tseng ZJ, Lee KS, Chang JS (2005) Fermentative hydrogen production with *Clostridium butyricum* CGS5 isolated from anaerobic sewage sludge. Int J Hydrogen Energy 30:1063–1070
87. Khanna N, Kotay SM, Gilbert JJ, Das D (2011) Improvement of biohydrogen production by *Enterobacter cloacae* IIT-BT 08 under regulated pH. J Biotechnol 152:9–15
88. Jo JH, Lee DS, Park D, Choe W, Park JM (2008) Optimization of key process variables for enhanced hydrogen production by *Enterobacter aerogenes* using statistical methods. Bioresour Technol 99:2061–2066
89. Kumar N, Das D (2000) Enhancement of hydrogen production by Enterobacter cloacae IIT-BT 08. Process Biochem 35:589–593
90. Abreu AA, Karakashev D, Angelidaki I, Sousa DZ, Alves MM (2012) Biohydrogen production from arabinose and glucose using extreme thermophilic anaerobic mixed cultures. Biotechnol Biofuels 5:1–12
91. Koskinen PEP, Lay CH, Puhakka JA, Lin PJ, Wu SY, Orlygsson J, Lin CY (2008) High-efficiency hydrogen production by an anaerobic, thermophilic enrichment culture from an Icelandic hot spring. Biotechnol Bioeng 101:665–678

CHAPTER 4

Intelligent Models to Predict Hydrogen Yield in Dark Microbial Fermentations using Existing Knowledge

This chapter has been submitted to International Journal of Hydrogen Energy with the title: Intelligent Models to Predict Hydrogen Yield in Dark Microbial Fermentations using Existing Knowledge.

The manuscript is presented in the following pages.

Intelligent Models to Predict Hydrogen Yield in Dark Microbial Fermentations using Existing Knowledge

Yeshona Sewsynker and Evariste Bosco Gueguim Kana

Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209

Pietermaritzburg, South Africa

Correspondence: kanag@ukzn.ac.za

Phone: +27332605527

Abstract

Bioprocess development for hydrogen production requires an excellent understanding of the influence of key operational parameters on hydrogen yields at early stages of the innovation chain. Knowledge on the impact of inoculum type, substrate type, substrate concentration, pH and temperature on fermentative hydrogen yields exist in the public domain. This study builds on this knowledge to implement intelligent models that could predict the hydrogen response on new physicochemical input values. Two Artificial Neural Network (ANN) models for hydrogen production were implemented and assessed using published data from 64 selected studies. For both models the multilayer perceptron (MLP) class of neural network was used with a topology of 5-7-7-1 corresponding to the number of neurons of inputs, hidden (2) and output layers. The input variables consisted of inoculum type (mixed and pure cultures), substrate type (xylose, glucose and sucrose), substrate concentration, pH and temperature. The output was the hydrogen yield expressed as mole of hydrogen per mole of substrate (Mol_Model) or as cumulative volume of hydrogen per gram substrate (Vol_Model). These models were validated by predicting the yields on experimental studies not previously used for model training. A high coefficient of determination (R^2) was obtained for Vol_Model (0.90) whereas a low value was observed with Mol_Model (0.46). Sensitivity analysis revealed that the most significant inputs on the process yield were temperature, pH, substrate type, inoculum type and substrate concentration (Mol_Model) and temperature, inoculum type, substrate type, pH and substrate concentration (Vol_Model) in decreasing order. These findings showed that the Vol_Model efficiently abstracted the non-linear relationship between the considered inputs and biohydrogen yield with a higher prediction accuracy on new physicochemical parameters. Thus, these ANN derived models could be used to navigate the optimization space and shorten the biohydrogen process development time.

Keywords: Modelling, Biohydrogen production, Artificial Neural Network, Sensitivity Analysis

1. Introduction

Fossil fuel depletion and the steep increase in greenhouse gas emissions have driven research towards renewable energy methods [1]. Biohydrogen has proven to be an excellent potential alternative to fossil fuel sources [2], due to its high gravimetric energy density of 122 kJ/g, which is approximately 2.9 times higher than conventional fossil fuels. Additionally, the combustion of hydrogen results in water as the only by-product [3]. Various biological methods exist for hydrogen production and include: photo-fermentation, dark fermentation and microbial electrolysis. Dark fermentation has shown to generate superior hydrogen production rates compared to other processes in terms of energy efficiency, and can exploit a wide array of renewable organic matter [4]. It proceeds via the butyrate (2 mol H_2/ mol hexose) or acetate pathway (4 mol H_2/ mol hexose) however; practical yields do not reach theoretical values due to metabolic limitations [5].

Bioprocess development and scale up requires modelling and optimization of the key parameters that impact the output at the initial stages of process development [6]. Key parameters for biohydrogen production include pH, temperature, inoculum source, substrate type and substrate concentration and may affect the microbes that are involved in the process. A bioprocess model provides insight into the individual as well as the interactive effects of the various input parameters on the corresponding output. Nevertheless, the non-linearities associated with microbial fermentations increase the complexity in model development. Non-linear systems as opposed to linear systems are not standardized thus resulting in deviations between results obtained [7]. The implementation of bioprocess models that efficiently encapsulate these non-linearities are of paramount importance for optimization and scale up of the process [7]. Numerous studies have attempted to provide models that relate these physicochemical inputs to the hydrogen yields [8─12].

Several factors have shown to impact the hydrogen production process and include: inoculum type, substrate type and concentration, temperature, and pH [8─13]. These factors affect the microbial community composition, impact the metabolic fluxes and ultimately the amount of hydrogen produced in the system, thus selecting the metabolic pathway for biohydrogen production [13,14]. A slight change from the optimum set point may have a significant impact

on the process yield [13,15─27]. Studies have revealed that pH values below 4.5 inhibit the hydrogenase activity and thus will influence the overall yield [15,22,28]. Both pure and mixed cultures may be used for hydrogen production [14]. The latter are cheaper to operate, simpler to control at large scale without contamination and have a broader choice of substrate [29─31].

With regards to substrate type, pure glucose has been mostly used for biohydrogen research [10,14]. It is easily metabolized by most microorganisms. However, the availability and costs associated with glucose as a substrate have restricted its potential use for biofuel research. Alternatively, sucrose and xylose have been used to a lesser extent [17,19]. Sucrose, a disaccharide is more resistant for microbes to degrade; however, the more versatile the culture, as in the case of a mixed consortium, the less challenging it becomes. Synergistic interactions between microbial communities permit simultaneous carbohydrate degradation and biohydrogen production using complex substrates. This is advantageous when considering substrates such as lignocellulosic biomass that is mainly comprised of xylose, lignin and cellulose. Reports on pure xylose as a substrate are scarce since it is commonly accessible from waste plant matter that may be pretreated for the fermentation process. Optimum substrate concentration has been reported within the range of 10-30 g/L [10,11,19]. Wang and Wan [10] reported a maximum yield of 305.3 mL H_2/g glucose. This result was consistent with Wang and Wan [11]. Contrariwise, Mu *et al*. [19] obtained a maximum yield of 252 mL H_2/g sucrose.

Significant variations exist between the reported optimum set points of input parameters for fermentative hydrogen production [10,11,17,19,32─34]. The development of bioprocess models at the initial stages of the optimization process is of paramount importance for efficiently relating the key parameters on the hydrogen output [2,35]. The implementation of accurate and reliable process models is necessary for the determination of the optimal set points for biohydrogen production.

Different bioprocess modelling algorithms have been employed in biohydrogen research. These include the Response Surface Methodology (RSM), fractional factorial design and Artificial Neural Networks (ANN) [8,10,11,36─38]. Sekoai and Gueguim Kana [8] used RSM to model the effect of substrate concentration, pH, temperature and hydraulic retention time on the hydrogen production process and indicated that this model was able to adequately relate the inputs to the hydrogen output. Likewise, Venkata-Mohan *et al*. [36] modelled the

effect of inoculum type and pretreatment, inlet pH and feed composition on the hydrogen production and substrate degradation efficiency using a fractional factorial design (Taguchi method). Results showed that the developed model was able to determine the optimum conditions for hydrogen production and substrate degradation.

ANNs are described as mathematical representations of the neurological functioning of the human brain. They imitate the brain's learning process by mathematically modelling the network structure of interconnected nerve cells [39] and can be used for bioprocess model development without prior knowledge of the kinetics of metabolic fluxes within the cell and the cultural environment [40]. The effectiveness of ANN in bioprocess development has been reported in various studies [9─12,35,37,38,40─45]. The ability of ANN models to accurately capture the non-linear relationships in hydrogen fermentation processes were illustrated by the high correlation between the observed and predicted data in the above-mentioned studies.

For instance, Prakasham *et al.* [35] developed an ANN model on hydrogen production with inputs of pH, glucose to xylose ratio, inoculum size and inoculum age. Whiteman and Gueguim-Kana [9] implemented an ANN model to determine the effect of temperature, initial pH, substrate concentration and inoculum size on hydrogen yield. Both models showed a high level of correlation between the predicted and observed with R^2 values above 0.90 [9,35]. Similarly, Nikhil *et al.* [42] investigated the influence of hydraulic retention time (HRT), recycle ratio, sucrose concentration and degradation, biomass concentration, pH, alkalinity, oxidation-reduction potential (ORP), and acid and alcohol concentrations on hydrogen production rate and acquired a coefficient of determination of 0.90. These models were implemented with data sample sizes below 50, as large numbers of bioprocess experimentations are costly and time consuming. The predictive accuracy of ANN may be enhanced with an increase in data size [46]. With the exception of the study by Nasr *et al* [37], the application of ANN on biofuel bioprocess modelling with a data set beyond 30 has been scantily reported [43─45].

Biohydrogen yields have been typically expressed using the specific hydrogen production potential (mL H_2/ g substrate) and the number of moles of hydrogen per mole of substrate consumed (mol H_2/mol substrate). These different units in hydrogen yields may be defined as the cumulative volume of hydrogen produced with regard to the total substrate consumed (mL H_2/g substrate) [10,18] and the number of moles of hydrogen per mole of substrate (hexose sugar) consumed (mol H_2/mol substrate) [19,24]. The latter is more often used since it

accounts for the stoichiometric yield and can be associated with the metabolic pathway adopted by the microorganisms involved in the fermentation process. It is thus believed to be more suitable for yield comparison [17─19,24,47].

To the best of our knowledge, there has been no comparative scientific study on reproducibility of the above-mentioned yield expressions in the public domain. Additionally, despite the availability of scattered reports on the effects of individual as well as the interactive effect of input parameters on biohydrogen response, there is a dearth of knowledge of intelligent models built on existing information which can efficiently predict the hydrogen response on unknown input patterns in the available public repositories. This study aims at using Artificial Intelligence to implement intelligent models from existing knowledge that could predict the hydrogen response on new physicochemical input values. Two Artificial Neural Network models for hydrogen production, based on yield expression type, were developed with input variables of (pure and mixed), substrate type (xylose, glucose and sucrose), substrate concentration, temperature and pH. The developed models were thereafter assessed on new input patterns for hydrogen production.

2. Materials and Methods

2.1. Data collection

Following an extensive survey of the published literature on the effect of various physicochemical parameters on biohydrogen production, 64 studies were selected to generate 182 data points for this study. These were divided into 133 data points (Mol_Model) from 49 published studies and 49 data points (Vol_Model) from 15 published studies under varied input conditions.

The selected input variables consisted of inoculum type, temperature, pH, substrate type and concentration. The model output was the hydrogen yield as mol H_2/ mol substrate (Mol_Model) or mL H_2/ g substrate (Vol_Model). For the Mol_Model the input parameter types and ranges were inoculum type (pure or mixed), substrate type (xylose, glucose and sucrose), substrate concentration (10-40g/L), temperature (25-40°C) and pH (4.5-9) (Table 1 and 2). With regard to the Vol_Model, the ranges for the input parameters were inoculum type (pure or mixed culture), temperature (25-40°C), pH (5-9), substrate types (xylose, glucose and sucrose) and substrate concentration (10-35 g/l) (Table 3 and 4).

Table 1. Database used for the development of the ANN Mol_Model

Carbon source	No. of data points	Inoculum type		Temperature (°C)	pH	Substrate concentration (g/L)	Source
		Mixed culture	Pure culture				
Xylose (1)	27	12	15	35-40	5.5-9	10-40	[18,20,23,48–53]
Glucose (2)	83	43	40	25-40	4.5-7.5	10-20	[17,19,26,27,47–50,54–79]
Sucrose (3)	23	17	6	30-40	4.7-8.5	10-30	[23–25,68,80–86]

Note: Numbers next to substrates were assigned to distinguish between the various types. This was performed based on the molecular weights of each substrate, thus xylose (150.13 g/mol), glucose (180.16 g/mol) and sucrose (342.2965 g/mol) were ranked as numerical values 1, 2 and 3, respectively. Inoculum type was designated as (1) for mixed cultures and (2) for pure cultures.

Table 2. Ranges for input parameters used in the Mol_Model development

Parameter	Minimum	Maximum	Unit
Inoculum type	1	2	-
Temperature	25	40	°C
pH	4.5	9	-
Substrate type	1	3	-
Substrate concentration	10	40	g/L
Hydrogen yield	0.5	2	mol H_2/mol substrate

Note: Inoculum type: mixed culture (1); pure culture (2); Substrate type: xylose (1); glucose (2); sucrose (3). This was performed based on the molecular weights, as described in Table 1.

Table 3. Database used for the development of the ANN Vol_Model

Carbon source	No. of data points	Inoculum type		Temperature (°C)	pH	Substrate concentration (g/L)	Source
		Mixed culture	Pure culture				
Xylose (1)	3	2	1	37	6-6.8	10-20	[53,87]
Glucose (2)	39	32	7	30-40	5-9	10-40	[10,11,32–34,60,61,88–91]
Sucrose (3)	7	7	0	30-37	5-9	10-30	[18,92]

Note: Inoculum type: mixed culture (1); pure culture (2); Substrate type: xylose (1); glucose (2); sucrose (3). This was performed based on the molecular weights, as described in Table 1.

Table 4. Ranges for input and output parameters used in the Vol_Model development

Parameter	Minimum	Maximum	Unit
Inoculum type	1	2	-
Temperature	25	40	°C
pH	5	9	-
Substrate type	1	3	-
Substrate concentration	10	35	g/L
Hydrogen yield	101	305.3	mL H_2/g substrate

Note: Inoculum type: mixed culture (1); pure culture (2); Substrate type: xylose (1); glucose (2); sucrose (3). This was performed based on the molecular weights, as described in Table 1.

2.2. Neural network development

For model development, two separate Artificial Neural Networks (ANN) built on multilayer perceptrons were structured. Each Neural network had a topology of 5-7-7-1, corresponding to the number of neurons of input, hidden (two) and output layers (Figure 1).

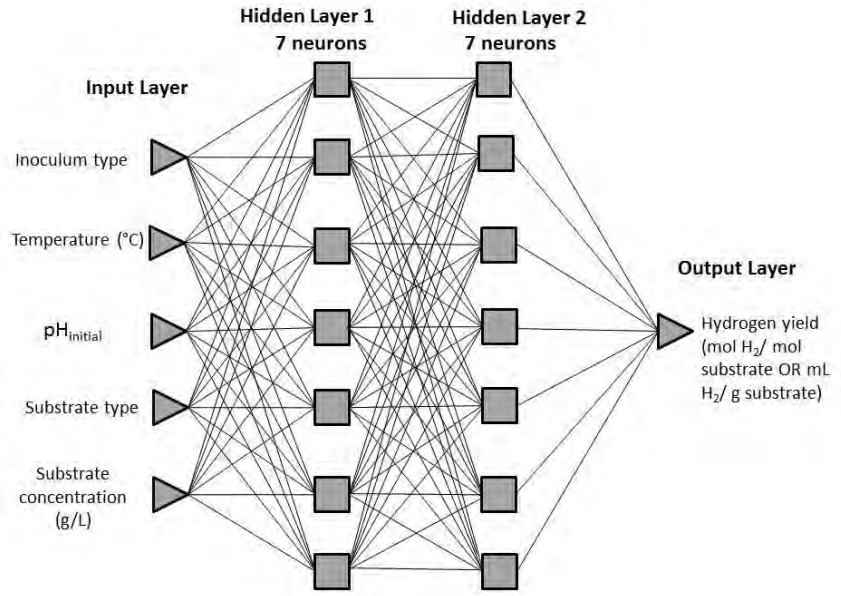


Figure 1. Topology of Neural Networks used for Mol_Model and Vol_Model. It consists of one input layer (five neurons), two hidden layers (seven neurons each) and one output layer (one neuron).

The feed forward architecture was adopted, whereby the input layer neurons transmitted signals to the hidden layer neurons [93]. For the hidden layer, a sigmoid transfer function was implemented. This hidden layer had two main purposes: (i) the addition of the weighted inputs together with the linked bias; (ii) then, to change the input data to a non-linear form, as shown in Equations 1 and 2 [93]:

$$\text{sum} = \sum_{i}^{n} = 1^{xiwi} + \theta \qquad (1)$$

where w_i (i = 1, n) are the connection weights, θ is the bias and x_i is the input variable (Desai *et al*., 2008)

$$f(sum) = \frac{1}{1+\exp(-sum)} \qquad (2)$$

The learning patterns were randomly selected during the learning process. The Mean Square Error (MSE) between predicted and observed for the cross-validating data for both models was calculated according to Equation 3.

$$\text{RMSE}= \sqrt{\frac{\sum_i^N \sum_{n=1}^M (y^i n - \hat{y}^i n)^2}{NM}} \qquad (3)$$

where N refers to the number of patterns used in the training; M denotes the number of output nodes; i denotes the index of the input pattern (vector) and $y^i n$ and $\hat{y}^i n$ are the actual and predicted outputs, respectively.

2.3. Data pretreatment, ANN training and validation

Prior to using the data, noise reduction was achieved by discarding the outliers. Data normalization was carried out according to Equation 4. Data on substrate and inoculum type were transformed to coded values. This was done based on the molecular weights for each substrate. Thus, xylose (150.13 g/mol), glucose (180.16 g/mol) and sucrose (342.2965 g/mol) were coded as 1, 2 and 3, respectively. For inoculum type, mixed cultures and pure culture were coded as 1 and 2, respectively.

$$\text{Normalized } (e_i) = \frac{e_i - E_{min}}{E_{max} - E_{min}} \qquad (4)$$

where e_i is the normalized data and E_{min} and E_{max} denote the minimum and maximum values set at -0.9 and 0.9, respectively.

Both ANN models were trained using the back propagation (BP) algorithm [9] with a momentum and learning rate of 0.05 for both models, respectively. In this process, the error between the experimental (observed) and predicted data was propagated backward through the network and used to adjust the neurons' connections. This process was repeated until the MSE between the experimental and predicted data was reduced below an acceptable threshold (Figure 2).

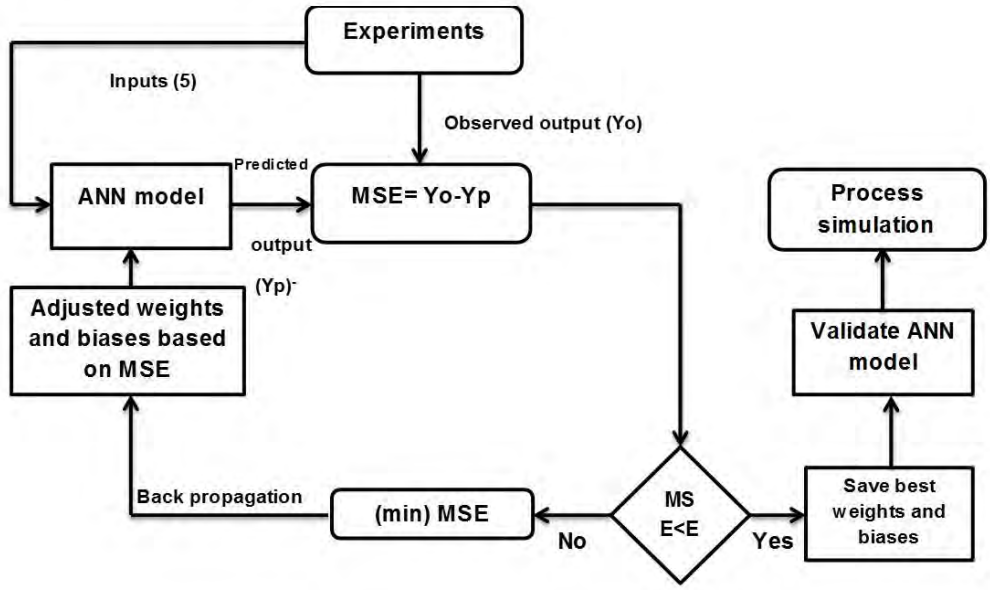


Figure 2. The back propagation training flowchart for artificial neural network.

2.4. Sensitivity Analysis

Sensitivity analysis was carried out to assess the impact of fractional changes of each input parameter on hydrogen response. A fractional change on each parameter was carried out in the ranges of -100 to 100% while keeping other parameters at their midpoint values (0%). The midrange values for substrate and inoculum type were assigned based on the most commonly reported substrate and inoculum type. For substrate type, xylose, glucose and sucrose were designated -50, 0 and 50%, respectively. Regarding inoculum type, mixed and pure cultures were designated -50 and 50%, respectively.

3. Results and discussion

3.1. Experimental Data Overview

An examination of the database indicated that glucose, a monosaccharide that is easily broken down, has most commonly been used as a substrate for biohydrogen research [10,11,17,19,47]. Xylose and sucrose have also been used but to a lesser extent [18,87]. Biohydrogen yields on these substrates vary from 0.5 to 2.35 mol H_2/mol glucose and 101 to 305.3 mL H_2/g glucose, depending on substrate type and the substrate concentration which ranges from 10-40g/L [10,76].

Inoculum type has shown to play a major role on hydrogen yields. Mixed cultures have been more commonly used compared to pure cultures since the former are simpler to handle, cheaper to operate and metabolize a wide range of substrate types. Pure cultures include microbes from the genus *Clostridium*. Mixed culture communities for hydrogen production are present in natural environments such as soil, wastewater, sewage sludge, compost and animal dung [47,53,60,91,94]. Within these mixed consortia, a synergistic interaction occurs whereby the non-hydrogen producing microbes create favourable conditions for the hydrogen-producing microorganisms [95─97].

The pH parameter has been suggested as one of the most critical variables in bioprocesses. Surprisingly, most reported studies on biohydrogen have been carried out without pH regulation throughout the duration of the process and the stated initial values ranged from 4.5-9 [98─100]. Unlike the pH parameter, temperature has been regulated in most studies on biohydrogen in the range 25-40°C [1,19,24,53,66,68,69,85,101].

3.2. Challenges associated with non-uniformity for expression of hydrogen yields

Studies on biohydrogen production have commonly reported yields as mL H_2/g substrate and mol H_2/mol substrate. The non-uniformity in unit expression for biohydrogen research has significantly impeded the process development [1]. The need for common expression in hydrogen data has previously been expressed by several authors [1,102,103]. Additionally, the non-uniformity in biohydrogen yield expression poses significant challenges in the reproducibility of experiments.

3.3. *Assessment of the Models' Significance*

During the training process, the MSE between the predicted and the observed data reduced to 0.004 and 0.42 for training and cross validation for the Mol_Model, and to 0.006 and 0.08 for training and cross validation for the Vol_Model. The fitness of the two models was assessed using Analysis of Variance (ANOVA) (Table 5). The coefficients of determination (R^2) for Mol_Model and Vol_Model were 0.46 and 0.90 respectively, thus indicating that these models could account for 46% and 90% of variations in the observed data. R^2 values > 0.70 are regarded as good models [93]. Low P-values of < 0.019162 and 0.008908 and high F-values of 26.71 and 72.89 were observed for the Mol_Model and Vol_Model, respectively. These statistical indices point to a relative predictive superiority of the Vol_Model over the Mol_Model.

Logan *et al*. [103] stated that the amount of hydrogen produced from a substrate is generally calculated for specific carbohydrates on a molar basis. Usually, the number of moles of hydrogen produced in an experiment is calculated from the volume of hydrogen produced and the ideal gas law as $n_{H2}=V_{H2}P/(RT)$, where P (bar) is the atmospheric pressure measured in the laboratory and R is 0.08314 L bar/K mol. However various researchers have substituted the standard temperature and pressure in the same equation for computing the hydrogen yield [62,104] and this can account for great variation in the yields observed among other factors.

The limitations of using the ideal gas law are that: (1) it only works well at low pressures and high temperatures; (2) most gases do not behave ideally above a pressure of 1 atm; (3) it does not work well near the condensation conditions of a gas. Thus, using the ideal gas law at standard conditions poses additional limitations since the pressure differs from place to place and will affect the volume of gas produced since pressure is inversely proportional to the volume of gas [105,106]. Conversely, the mL H_2/ g substrate unit of expression presents different challenges. For example, the cumulative volume of hydrogen gas is influenced by the environmental pressure and temperature at which the experiment is conducted. These are not taken into account with this unit of expression.

The mol H_2/ mol substrate unit of expression is a stoichiometric yield that may be used for determining the metabolic pathways adopted by the microbes involved in the fermentation process. However, this may be much more complex when considering mixed microbial consortia. This is a result of the various microbes present within mixed culture systems that follow different metabolic pathways in the same system. Therefore, it is difficult to determine the metabolic pathways based on the stoichiometric yield [107]. In such circumstances, microbial community analysis needs to be performed using high throughput methods such as next generation sequencing for establishing the microbes involved and ultimately the major metabolic pathways in the fermentation process [108]. Another investigative method for confirmation of the metabolic pathway followed is volatile fatty acid (VFA) analysis [109].

Therefore, standardization of reporting hydrogen yields is crucial for overcoming the discrepancies in hydrogen output under similar conditions. Optimization of biohydrogen production requires an in-depth knowledge of the key parameters that drive the fermentation process. The availability of an enhanced biohydrogen process model that accurately predicts process output over a wide range of input conditions will reduce the experimental burden. In

this study, the Vol_Model showed a relative superiority over the Mol_Model. These results indicate that biohydrogen research should apply this unit of expression for reporting yields in future studies for a more accurate comparison.

The developed models in this study were thereafter assessed for their predictive accuracy on hydrogen yields from 33 and 10 data points for the Mol_Model and Vol_Model, respectively. These data points were obtained from studies by Mu *et al*. [19], Lin and Cheng [23], Wang *et al.* [24], Wang *et al*. [25], Xing *et al*. [27], Sivagurunathan *et al.* [47], Lo *et al.* [50], An *et al.* [53], Kawagoshi *et al.* [55], Baghchehsaraee *et al.* [59], Wang and Wan [60], Zhao *et al.* [66], Junghare *et al.* [68], Liu *et al*. [69], Tang *et al*. [70], Kurokawa and Tanisho [74] and Mei *et al.* [85] for Mol_Model and Wang and Wan [10], Mu *et al.* [18], Wang and Wan [60], Qian *et al.* [61], Cheng *et al.* [87] and for Vol_Model. These data were not previously used in the development of the models.

Table 5. Analysis of Variance (ANOVA) for Mol_Model and Vol_Model

Source	Sum of Squares	df	Mean Squares	F-value	P-value	R^2
Mol_Model	4.089058	1	4.089058	26.71	0.019162	0.46
Vol_Model	21572.04	1	21572.04	72.89	0.008908	0.90

The plots of predicted versus observed hydrogen yield values are depicted in Figure 3(a) and 3(b) for the Mol_Model and Vol_Model respectively. Figure 3(a) showed that the data points are scattered on either side of the diagonal, thereby illustrating a weak relationship between the predicted and the observed hydrogen yields for the Mol_model. On the other hand, Figure 3(b) showed that most data points were aligned near the diagonal, thus illustrating the closeness between the predicted and observed yields.

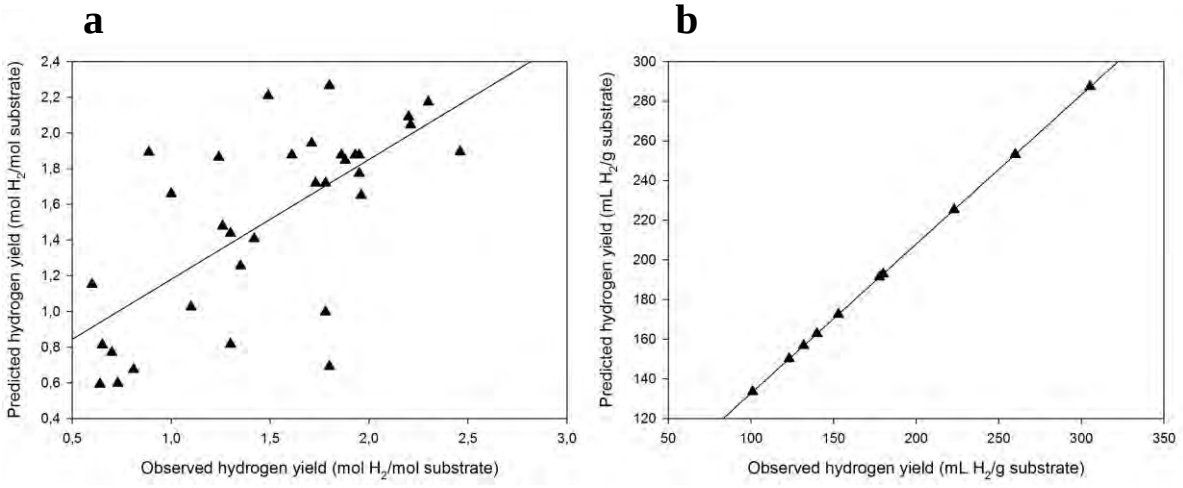


Figure 3. Predicted versus observed biohydrogen yields (mol H_2/mol substrate and mL H_2/g substrate) values for 10 and 33 experimental data sets for (a) Mol_Model (R^2=0.46) and (b) Vol_Model (R^2= 0.90), respectively. Note: The diagonal line illustrates expectations under a one-to-one relationship between predicted and observed values.

3.4. Mol_Model prediction on new experiments not used for training

The Mol_Model was further assessed using 33 data points from studies carried out by Mu *et al*. [19], Lin and Cheng [23], Wang *et al.* [24], Wang *et al*. [25], Xing *et al*. [27], Sivagurunathan *et al.* [47], Lo *et al.* [50], An *et al.* [53], Kawagoshi *et al.* [55], Baghchehsaraee *et al.* [59], Wang and Wan [60], Zhao *et al.* [66], Junghare *et al.* [68], Liu *et al*. [69], Tang *et al*. [70], Kurokawa and Tanisho [74] and Mei *et al.* [85] (Figure 4a).

For instance, the observed hydrogen yields in the study by Mu *et al*. [19] with input variables of temperature, pH and substrate concentration were 1.78 and 1.73 mol H_2/mol glucose against 1.72 mol H_2/mol glucose predicted by the Mol_Model. Wang *et al.* [24] studied the influence of pH, temperature and substrate concentration on hydrogen production. Observed hydrogen yields were 0.65, 2.23 and 0.887 mol H_2/mol sucrose against the Mol_Model predicted of 0.81, 2.17 and 1.89 mol H_2/mol sucrose, respectively.

Similarly, Sivagurunathan *et al.* [47] evaluated the effect of individual and combined mixed culture inoculum sources on biohydrogen production. Observed hydrogen yields were 1.93 (cow dung+anaerobic sludge+pig slurry), 1.86 (anaerobic sludge), 1.95 (anaerobic sludge + pig slurry) and 1.61 (cow dung+pig slurry) mol H_2/mol glucose. The Mol_Model under similar conditions predicted a hydrogen yields of 1.88 mol H_2/mol glucose for all the above-

mentioned inoculum combination types. Likewise, Kawagoshi *et al.* [55] investigated the influence of pH on hydrogen production using anaerobic digested sludge as an inoculum. The observed hydrogen yield was 1.80 mol H_2/mol glucose against 0.69 mol H_2/mol glucose predicted by the Mol_Model. In addition, Baghchehsaraee *et al.* [59] obtained a hydrogen yield of 1.95 mol H_2/mol glucose compared to 1.77 mol H_2/mol glucose predicted by the Mol_Model. Wang and Wan [60] observed a hydrogen yield of 1.78 mol H_2/mol glucose compared to 1.00 mol H_2/mol glucose predicted by the Mol_Model. Alternatively, the observed yield of 1.96 mol H_2/mol glucose by Zhao *et al.* [66] was slightly higher than that predicted by the present Mol_Model under similar conditions (1.65 mol H_2/mol glucose). In a study by Junghare *et al.* [68], an experimental hydrogen yield of 1.49 mol H_2/mol sucrose was achieved which was significantly lower than that predicted by the Mol_Model (2.21 mol H_2/mol sucrose) developed in this study. A similar correlation pattern between the predicted and the observed yields using the Mol_Model was observed by Liu *et al*. [69], Tang *et al*. [70], Kurokawa and Tanisho [74] and Lo *et al.* [50] and the predicted yields by the implemented Mol_Model.

On the other hand, Wang *et al*. [25] obtained hydrogen yields of 2.46 mol H_2/mol sucrose, 2.21 mol H_2/mol sucrose and 1.35 mol H_2/mol sucrose against predicted yields of 1.89, 2.04 and 1.26 mol H_2/mol sucrose by the Mol_Model, respectively. A comparable result was observed between the experimental and predicted yields by Lin and Cheng [23], Xing *et al*. [27], An *et al.* [53] and Mei *et al.* [85]. Although this model exhibited a low R^2 value (0.46), it is interesting to note that the trends between the predicted and observed values in Figure 4(a) display a high level of parallelization.

The low correlation observed within the Mol_Model does not demonstrate any weakness of ANN as a modelling tool but is attributable to the non-uniformity in reported hydrogen yields initially used for model development. Non-uniformities that have led to the variations observed may be related to the inconsistencies between methods used for yield computation in terms of the pressure and temperature used. Despite the low R^2 value obtained with the Mol_Model, an acceptable match pattern was observed between the predicted and experimental hydrogen output (Figure 4a). This result shows the robustness of ANN models for bioprocess development.

3.5. Vol_Model prediction on new experimental studies

The developed Vol_Model was further assessed by predicting the hydrogen response on 10 data points from studies carried out by Wang and Wan [10], Mu *et al.* [18], Wang and Wan [60], Qian *et al.* [61], Cheng *et al.* [87] and Wang and Wan [89] (Figure 4b). A high correlation was obtained between the predicted and the observed hydrogen response using this model. For example, the observed hydrogen yield in the study of Wang and Wan [10] under the input variables of temperature, pH and substrate concentration was 131.9 mL H_2/g glucose. In this study, the Vol_Model predicted 156.7 mL H_2/g glucose. Additionally, in the same study by Wang and Wan [10], hydrogen yields of 123.1 mL H_2/g glucose and 305.3 mL H_2/g glucose were observed. The implemented Vol_Model predicted 150.1 mL H_2/g glucose and 287.3 mL H_2/g glucose under similar conditions.

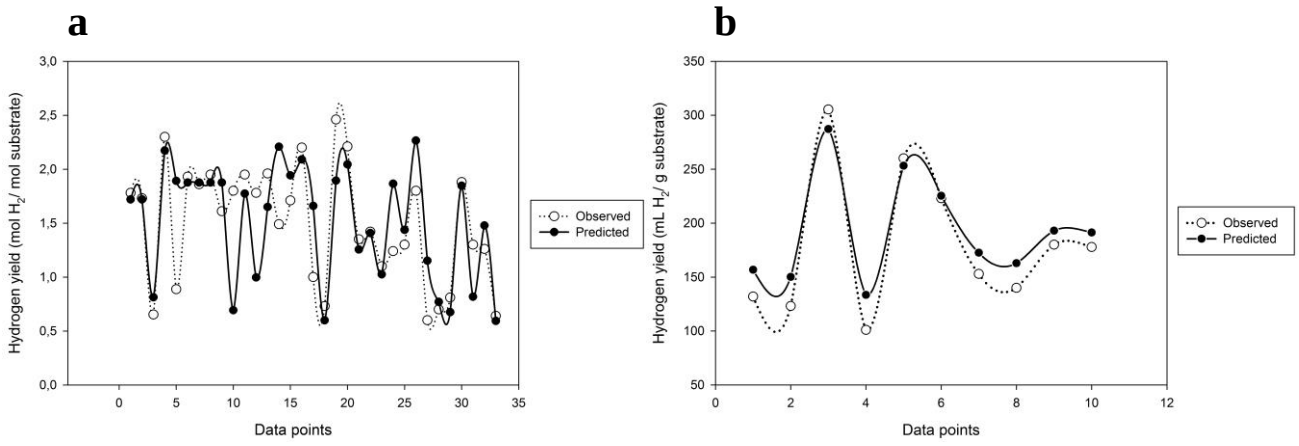


Figure 4. Observed hydrogen yield compared to the predicted for (a) Mol_Model (33 data points) and (b) Vol_Model (10 data points)

The Vol_Model predicted hydrogen yields of 133.5 and 253.2 mL H_2/g sucrose obtained with the experimental inputs from the study of Mu *et al.* [18]. The observed yields were 101 and 255 mL H_2/g sucrose. The considered input pattern was pH (6.0 and 5.5), temperature (30 and 35°C) and substrate concentration (20 and 25g/L). Likewise, Wang and Wan [60] achieved a hydrogen yield of 223 mL H_2/g glucose against 225.3 mL H_2/g glucose predicted by the Vol_Model. A similar correlation pattern between the predicted and the observed yields using the Vol_Model was observed by Wang and Wan [60], Qian *et al.* [61], Cheng *et al*. [87] and Wang and Wan [89]. In the study by Qian *et al.* [61] observed yields were 152.9 mL H_2/g glucose and 140 mL H_2/g glucose compared to 172.6 mL H_2/g glucose and 162.8 mL H_2/g

glucose predicted by the Vol_Model. Cheng *et al.* [87] observed a hydrogen yield of 177.8 mL H_2/g xylose compared to 191.3 mL H_2/g xylose (predicted by Vol_Model). Furthermore, the observed yield by Wang and Wan [89] was 180 mL H_2/g glucose vs 192.9 mL H_2/g glucose predicted by the Vol_Model. The high correlation observed between the experimental and predicted hydrogen yields can be linked to the models' high generalization ability to predict on novel input parameters. The slight discrepancies between the observed and model predicted yields may be accounted for by the ineluctable mismatch in experimental conditions encountered in microbial bioprocesses.

Nevertheless, ANN models have proven to be valuable especially for their robustness in biological systems prediction [9,10]. These observations demonstrate the Vol_Models' high predictive ability in novel virtual experimentations.

3.6. Sensitivity analysis of input parameters on ANN models

Sensitivity analysis was carried out on the developed models to determine the relative sensitivity of hydrogen yields on input parameters of inoculum type, substrate type, substrate concentration, pH and temperature. A sensitivity indicator represents the adjustment in the systems' outputs attributable to variations in the process input parameters. A large sensitivity to a parameter implies that the process output can change considerably with slight variation in the input parameter [110,111]. Conversely, a low sensitivity indicates a little change will occur in the output of the system even if a large variation occurs in the input parameter. The sensitivity analysis for the Mol_Model (Figure 5a and c) and Vol_Model (Figure 5b and d) is shown below. The lines on each graph represent the rate of change of the output with respect to a change in each input.

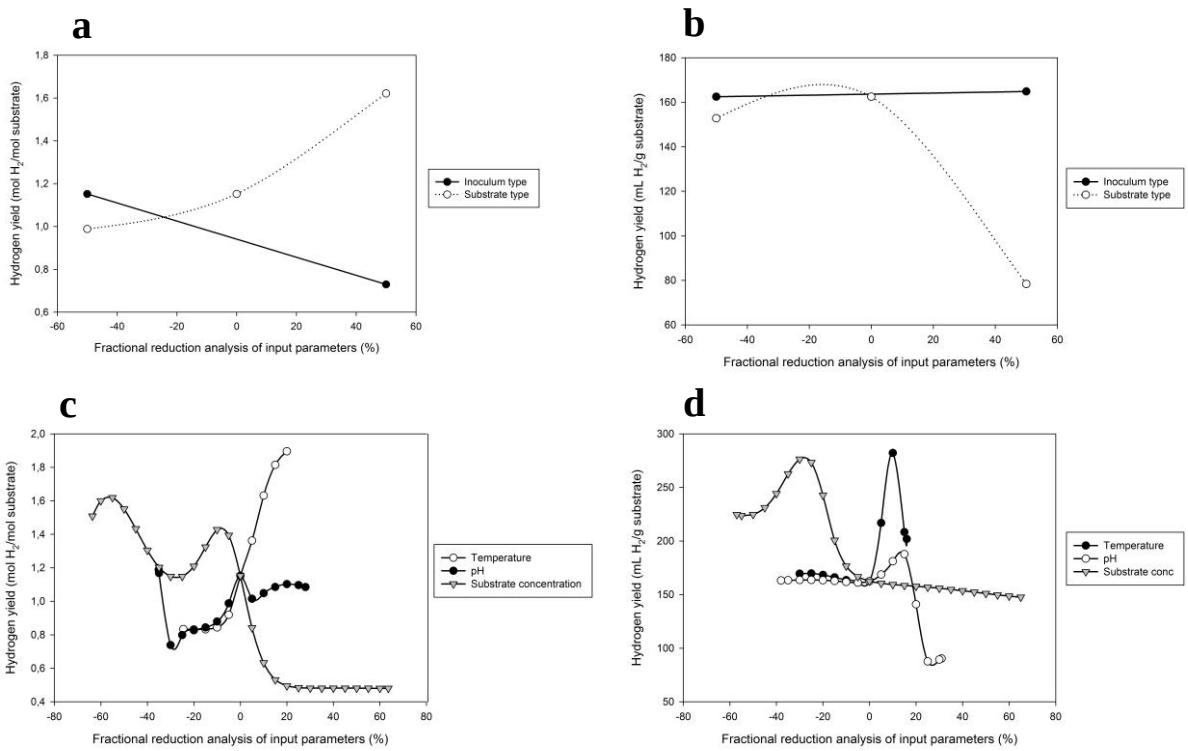


Figure 5 (a-d). Impact of fractional change of input parameters on hydrogen output (a) Inoculum type and Substrate type (Mol_Model), (b) Inoculum type and substrate type (Vol_Model), (c) Temperature, pH and substrate concentration (Mol_Model) and (d) Temperature, pH and substrate concentration (Vol_Model) on hydrogen yield

Sensitivity analysis on the Mol_Model indicated that using mixed cultures had a relatively higher hydrogen yield (1.15 mol H_2/ mol glucose) compared to pure cultures (0.73 mol H_2/ mol glucose). Whereas, the Vol_Model showed that for inoculum type, both pure (50%) and mixed cultures (-50%) were efficient for hydrogen production with similar yields of 162.5 mL H_2/g substrate and 164.8 mL H_2/g glucose, respectively. The slight increase in hydrogen yield for mixed cultures for the Vol_Model may be accounted for by the synergistic interactions between mixed microflora [61].

The diverse microorganisms exhibit different metabolic pathways thus they are able to produce hydrogen from a wide range of substrates. Mixed cultures have shown to be effective for hydrogen production since the microbial community displays synergistic interactions for biohydrogen production. However, mixed culture communities may require a pretreatment to deactivate the hydrogen-consuming microbes such as the methanogens and encourage hydrogen producers such as *Clostridium* spp. Several studies have shown that different

pretreatment methods of the mixed inoculum source may lead to improved yields [62,112,113]. Nevertheless, other studies have demonstrated that pure cultures result in higher yields compared to mixed culture systems [61,73,114─117].

With reference to substrate type, a fractional increase of 50% from the baseline which corresponded to sucrose, resulted in the maximum predicted hydrogen yield of 1.62 mol H_2/mol sucrose followed by glucose (1.15 mol H_2/mol glucose) and xylose (0.99 mol H_2/mol xylose) for the Mol_Model. On the other hand, the Vol_Model showed that for substrate type, glucose (162.5 mL H_2/g glucose) had the greatest influence on the hydrogen yield followed by xylose (152.8 mL H_2/g xylose) and sucrose (78.4 mL H_2/g sucrose).

With the substrate concentration, it was shown that a 10% reduction from its midpoint value (22.5 g/L) resulted in a maximum predicted hydrogen yield of 1.42 mol H_2/mol glucose for the Mol_Model. The Vol_Model however showed that a fractional reduction of 30% from its base line value (15.19 g/L) resulted in the maximum predicted yield of 276,4 mL H_2/g glucose. Interestingly, the optimum substrate concentration predicted by both models was within the range previously stated by [10,11,17─19,60].

The Mol_Model showed that a fractional change in temperature of 24.5% increase from the base line (thus an operating temperature of 40.45°C) resulted in a maximum predicted hydrogen yield of 1.94 mol H_2/mol glucose. On the other hand, a fractional change in temperature of 24.5% reduction from its midpoint value (24.55°C) decreased the hydrogen yield to 0.83 mol H_2/mol glucose. For the Vol_Model, a fractional increase in temperature of 10% from its base line (38.5°C) resulted in a maximum predicted hydrogen yield of 282.1 mL H_2/g glucose. The higher temperature values predicted by the Mol_Model (40.45°C) and Vol_Model (38.5°C) were consistent with previous studies on hydrogen production [10,11,13,18,19,60], since the hydrogen-producers generally grow under mesophilic (20-40°C) and thermophilic conditions (40-60°C).

A fractional change in the pH value has shown to significantly influence hydrogen yield. A fractional reduction in pH of 35% from its base line (4.39) resulted in a maximum hydrogen yield of 1.17 mol H_2/mol glucose. However, when the pH was fixed at its midpoint value (6.75) a slightly lower hydrogen yield of 1.15 mol H_2/mol glucose was observed. Regarding the Vol_Model, optimum pH was predicted at a fractional increase of 15% from its midpoint value (operational pH of 8.05), corresponding to a predicted hydrogen yield of 187,7 mL H_2/g

glucose. The optimum pH range for hydrogen production has been reported to be between 5-9 [13,98–100]. Previous studies have shown that pH values below 4.5 inhibit the hydrogenase activity during the fermentation process and are unfavorable for the hydrogen production process [15,22,28]. The predicted pH by the Vol_Model indicates that an initial pH of approximately 8 is required for maximum hydrogen yield. This value was within the range (5-9) previously reported for optimum hydrogen yield [15,12].

A comparison of the relative sensitivity of process inputs was computed based on the rate of change on hydrogen output for the Mol_Model and the Vol_Model. Slope values showed that for the Mol_Model, the relative sensitivity increased from substrate concentration (-0.0104), inoculum type (-0.0042), substrate type (0.0025), pH (0.0063) to temperature (0.0271) whereas for the Vol Model, the relative sensitivity increased from substrate concentration (-0.9648), pH (-0.837), substrate type (-0.7441), inoculum type (0.0235) to temperature (1.4567).

3.7. Limitations of the developed models

Additional parameters not considered in this study during model development, but which may impact on biohydrogen production include the inoculum pretreatment temperature and time since many microbes survive certain pretreatments and may impact on the overall yield [62]. Different microbes exist in the various sources and have shown to influence the process yield. Additionally, the quantity of essential nutrients for hydrogen production within fermentation medium (such as iron, nitrogen and phosphorus) have not been considered [13,14,35,62].

4. Conclusion

In this study, two intelligent bioprocess models have been implemented using ANN on 64 reported studies on fermentative hydrogen processes for the Mol_Model and Vol_Model, based on yield expression units (mL $H_{2/}$ g substrate and mol H_2/ mol substrate). A significant discrepancy was observed in the size of explainable variations of the two models with coefficient of determinations (R^2) of 0.46 and 0.90 for the Mol_Model and Vol_Model, respectively. Assessment of these models indicated that they were able to efficiently encapsulate the highly non-linear associations between the inputs and corresponding hydrogen yield pattern within the design space. The Vol_Model showed a superior biohydrogen predictive accuracy on various novel experimental data not used for training.

Thus, the implementation of intelligent bioprocess models with aggregated knowledge from several laboratories will significantly shorten the process development cost and time.

References

[1] Levin, D.B., Pitt, L. and Love, M. 2004. Biohydrogen production: prospects and limitations to practical application. International Journal of Hydrogen Energy 29:173-185.

[2] Nath, K. and Das, D. 2011. Modeling and optimization of fermentative hydrogen production. Bioresource Technology 102:8569-8581.

[3] Bélafi-Bakó, K., Bakonyi, P., Nemestóthy, N. and Pientka, Z. 2010. Biohydrogen production in integrated system. Desalination and Water Treatment 14:116-118.

[4] Mizuno, O., Dinsdale, R., Hawkes, F.R. and Hawkes, D.L. 2010. Enhancement of hydrogen production from glucose by nitrogen gas sparging. Bioresource Technology 73:59-65.

[5] Thauer, R.K., Jungermann, K. and Decker, K. 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriology Reviews 41:100-180.

[6] Escamilla-Alvarado, C., Rios-Leal, E., Ponce-Noyola, M.T. and Poggi-Varaldo, H.M. 2012. Gas biofuels from solid substrate hydrogenic–methanogenic fermentation of the organic fraction of solid municipal wastes. Process Biochemistry 47(11):1572-1587.

[7] Almeida, J.S. 2002. Predictive non-linear modeling of complex data by artificial neural networks. Current Opinions in Biotechnology 3:72-76.

[8] Sekoai, P.T. and Gueguim-Kana, E.B. 2013. A two-stage modelling and optimization of biohydrogen production from a mixture of agro-municipal waste. International Journal of Hydrogen Energy. 38:8657-8663.

[9] Whiteman, J.K. and Gueguim-Kana, E.B. 2014. Comparative assessment of the Artificial Neural Network and Response Surface Modelling Efficiencies for Biohydrogen Production on Sugar Cane Molasses. Bioenergy Research 7(1):295-305.

[10] Wang, J.L. and Wan, W. 2009a. Application of desirability function based on neural network for optimizing biohydrogen production process. International Journal of Hydrogen Energy 34:1253-1259.

[11] Wang, J.L. and Wan, W. 2009b. Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology. International Journal of Hydrogen Energy 34:255-261.

[12] Shi, Y., Gai, G., Zhao, X. *et al.* 2010. Back propagation neural network (BPNN) simulation model and influence of operational parameters on hydrogen bio-production through integrative biological reactor (IBR) treating wastewater. Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference, 18-20 June, Chengdu, China.

[13] Wang, J.L. and Wan, W. 2009c. Factors influencing fermentative hydrogen production: A review. International Journal of Hydrogen Energy. 34:799-811.

[14] Elsharnouby, O., Hafez, H., Nakhla, G., and El Naggar, M.H. 2013. A critical literature review on biohydrogen production by pure cultures. International Journal of Hydrogen Energy 38:4945-4966.

[15] Fang, H.H. and Liu, H. 2002. Effect of pH on hydrogen production from glucose by a mixed culture. Bioresource Technology 82:87-93.

[16] Fan, Y., Li, C., Lay, J.J., Hou, H. and Zhang, G. 2004. Optimization of initial substrate and pH levels for germination of sporing hydrogen-producing anaerobes in cow dung compost. Bioresource Technology 91:189-193.

[17] Mu, Y., Zheng X.J., Yu, H.Q. and Zhu, R.F. 2006a. Biological hydrogen production by anaerobic sludge at various temperatures. International Journal of Hydrogen Energy 31:780-785.

[18] Mu, Y., Wang, G. and Yu, H.Q. 2006b. Response surface methodological analysis on biohydrogen production by enriched anaerobic cultures. Enzyme and Microbial Technology 38:905-913.

[19] Mu, Y., Zheng, X.J. and Yu, H.Q. 2009. Determining optimum conditions for hydrogen production from glucose by an anaerobic culture using response surface methodology (RSM). International Journal of Hydrogen Energy 34:7959-7963.

[20] Lin, C.Y., Wu, C.C. and Hung, C.H. 2008a. Temperature effects on fermentative hydrogen production from xylose using mixed anaerobic cultures. International Journal of Hydrogen Energy 33:43-50.

[21] Lin, C.Y., Chang, C.C. and Hung, C.H. 2008b. Fermentative hydrogen production from starch using natural mixed cultures. International Journal of Hydrogen Energy 33:2445-2453.

[22] Khanal, S.K., Chen, W.H., Li, L. and Sung, S. 2004. Biological hydrogen production: effects of pH and intermediate products. International Journal of Hydrogen Energy 29:1123-1131.

[23] Lin, C.Y. and Cheng, C.H. 2006. Fermentative hydrogen production from xylose using anaerobic mixed microflora. International Journal of Hydrogen Energy 31:832-840.

[24] Wang, G., Mu, Y. and Yu, H.Q. 2005. Response surface analysis to evaluate the influence of pH, temperature and substrate concentration on the acidogenesis of sucrose-rich wastewater. Biochemical Engineering Journal 23:175-184.

[25] Wang, C.H., Lin, P.J., Chang, J.S. 2006. Fermentative conversion of sucrose and pineapple waste into hydrogen gas in phosphate-buffered culture seeded with municipal sewage sludge. Process Biochemistry 41:1353-1358.

[26] Wang, C.H., Lu, W.B. and Chang, J.S. 2007. Feasibility study on fermentative conversion of raw and hydrolyzed starch to hydrogen using anaerobic mixed microflora. International Journal of Hydrogen Energy 32:3849-3859.

[27] Xing, D., Ren, N., Wang, A., Li, Q., Feng, Y. and Ma, F. 2008. Continuous hydrogen production of auto-aggregative *Ethanoligenens harbinense* YUAN-3 under non-sterile condition. International Journal of Hydrogen Energy 33:1489-1495.

[28] Hawkes, F., Dinsdale, R., Hawkes, D. and Hussy, I. 2002. Sustainable fermentative hydrogen production: challenges for process optimisation. International Journal of Hydrogen Energy 27:1339-1347.

[29] Li, C.L. and Fang, H.H.P. 2007. Fermentative hydrogen production from wastewater and solid wastes by mixed cultures. Critical reviews in Environmental Science and Technology 37:1-39.

[30] Kleerebezem, R. and van Loosdrecht, M. 2007. Mixed culture biotechnology for bioenergy production. Current Opinion in Biotechnology 18:207-212.

[31] Temudo, M.F., Kleerebezem, R., and van Loosdrecht, M. (2007). Influence of the pH on (open) mixed culture fermentation of glucose: A chemostat study. Biotechnology and Bioengineering 98:69-79.

[32] Wang, B., Wan, W. and Wang, J.L. 2009. Effect of ammonia concentration on fermentative hydrogen production by mixed cultures. Bioresource Technology 100:1211-1213.

[33] Wang, J.L. and Wan, W. 2008a. Effect of temperature on fermentative hydrogen production by mixed cultures. International Journal of Hydrogen Energy 33:5392-5397.

[34] Wang, J.L. and Wan, W. 2008b. The effect of substrate concentration on biohydrogen production by using kinetic models. Science in China Series B: Chemistry 51(11):110-1117.

[35] Prakasham, R.S., Sathish, T. and Brahmaiah, P. 2011. Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. International Journal of Hydrogen Energy 36:4332-4339.

[36] Venkata-Mohan, S., Raghavulu, S.V., Mohanakrishna, G., Srikanth, S. and Sarma, P.N. 2009a. Optimization and evaluation of fermentative hydrogen production and wastewater treatment processes using data enveloping analysis (DEA) and Taguchi design of experimental (DOE) methodology. International Journal of Hydrogen Energy. 34(1):216-226.

[37] Nasr, N., Hafez, H., El Naggar, M.H. and Nakhla, G. 2013a. Application of artificial neural networks for biohydrogen production. International Journal of Hydrogen Energy 38:3189-3195.

[38] Nasr, M., Tawfik, A., Ookawara1, S. and Suzuki, M. 2013b. Prediction of hydrogen production using Artificial Neural Network. International Water Technology Conference, IWTC17, 5-7 November, Istanbul.

[39] Nagata, U. and Chu, K.H. 2003. Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnology Letters 25:1837-1842.

[40] Gueguim-Kana, E.B., Oloke, J.K., Lateef, A. *et al.* 2012. Modelling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renewable Energy 46:276-281.

[41] Sewsynker, Y., Gueguim-Kana, E.B. and Lateef, A. 2015. Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs). Biotechnology and Biotechnological Equipment 29(6):1208-1215.

[42] Nikhil, B.O., Visa, A., Lin, C.Y., Puhakka, J.A. and Yli-Harja, O. 2008. An artificial neural network based model for predicting H_2 production rates in a sucrose-based bioreactor system. World Academy of Science, Engineering and Technology 2(1):20-25.

[43] Rosales-Colunga, L.M., Garcia, R.G. and Rodriguez, A.D. 2010. Estimation of hydrogen production in genetically modified *Escherichia coli* fermentations using an artificial neural network. International Journal of Hydrogen Energy 35:13186-13192.

[44] Mu, Y. and Yu, H.Q. 2007. Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. International Journal of Hydrogen Energy 32:3308-3314.

[45] Mullai, P., Yogeswari, M.K., Sridevi, K. and Ross, P.R. 2013. Artificial neural network (ANN) modeling for hydrogen production in continuous anaerobic sludge blanket filter (ASBF). Singaporean Journal of Scientific Research 5(1):1-7.

[46] Huang, Y., Kangas, L.J. and Rasco, B.A. 2007. Applications of Artificial Neural Networks (ANNs) in Food Science. Critical Reviews in Food Science 47:113-126.

[47] Sivagurunathan, P., Sen, B. and Lin, C.Y. 2014. Batch fermentative hydrogen production by enriched mixed culture: Combination strategy and their microbial composition. Journal of Bioscience and Bioengineering. 117(2):222-228.

[48] de Sa, L.R.B., Cammarota, M.C., de Oliveira, T.C., Oliveir, E.M.M., Matos, A. and Ferreira-Leitaõ, V.S. 2013. Pentoses, hexoses and glycerin as substrates for biohydrogen production: An approach for Brazilian biofuel integration. International Journal of Hydrogen Energy 38:2986-2997.

[49] Kumar, N. and Das, D. 2000. Enhancement of hydrogen production by *Enterobacter cloacae* IIT-BT 08. Process Biochemistry 35:589-593.

[50] Lo, Y.C., Chen, W.M., Hung, C.H., Chen, S.D. and Chang, J.S. 2008. Dark H_2 fermentation from sucrose and xylose using H_2-producing indigenous bacteria: Feasibility and kinetic studies. Water Research 42:827-842.

[51] Mäkinen, A.E., Nissilä, M.E. and Puhakka, J.A. 2012. Dark fermentative hydrogen production from xylose by a hot spring enrichment culture. International Journal of Hydrogen Energy 37:12234-12240.

[52] Long, C., Cui, J., Liu, Z., Liu, Y., Long, M. and Hu, Z. 2010. Statistical optimization of fermentative hydrogen production from xylose by newly isolated *Enterobacter* sp. CN1. International Journal of Hydrogen Energy 35:6657-6664.

[53] An, D., Li, Q., Wang, X., Yang, H. and Guo, L. 2014. Characterization on hydrogen production performance of a newly isolated *Clostridium beijerinckii* YA001 using xylose. International Journal of Hydrogen Energy. 39:19928-19936.

[54] Zheng, X.J., Zheng, Y.M. and Yu, H.Q. 2010. Influence of NaCl on Hydrogen Production from Glucose by Anaerobic Cultures. Environmental Technology 26:1073-1080.

[55] Kawagoshi, Y., Hino, N., Fujimoto, A., Nakao, M., Fujita, Y., Sugimura, S. and Furukawa, K. 2005. Effect of Inoculum Conditioning on Hydrogen Fermentation and pH Effect on Bacterial Community Relevant to Hydrogen Production. Journal of Bioscience and Bioengineering 100(5): 524-530.

[56] Zheng, X.J. and Yu, H.Q. 2004a. Inhibitory effects of butyrate on biological hydrogen production with mixed anaerobic cultures. Journal of Environmental Management 74:65-70.

[57] Chang, S., Li, J.Z. and Liu, F. 2011. Evaluation of different pretreatment methods for preparing hydrogen-producing seed inocula from waste activated sludge. Renewable Energy 36:1517-1522.

[58] Ning, Y.Y., Jin, D.W., Sheng, G.P., Harada, H. and Shi, X.Y. 2012. Evaluation of the stability of hydrogen production and microbial diversity by anaerobic sludge with chloroform treatment. Renewable Energy 38:253-257.

[59] Baghchehsaraee, B., Nakhla, G., Karamaneva, D., Margaritisa, A. and Reid, G. 2008. The effect of heat pretreatment temperature on fermentative hydrogen production using mixed cultures. International Journal of Hydrogen Energy 33:4064-4073.

[60] Wang, J.L. and Wan, W. 2008c. Comparison of different pretreatment methods for enriching hydrogen-producing cultures from digested sludge. International Journal of Hydrogen Energy 33:2934-2941.

[61] Qian, C.X., Chen, L.Y., Rong, H. and Yuan, X.M. 2011. Hydrogen production by mixed culture of several facultative bacteria and anaerobic bacteria. Progress in Natural Science: Materials International 21:506-511.

[62] Faloye, F.D., Gueguim-Kana, E.B. and Schmidt, S. 2013. Optimization of hybrid inoculum development techniques for biohydrogen production and preliminary scale up. International Journal of Hydrogen Energy 38:11765-11773.

[63] Woo, J.H. and Song, Y.C. 2009. Influence of Temperature and Duration of Heat Treatment Used for Anaerobic Seed Sludge on Biohydrogen Fermentation. Korean Society of Civil Engineers (KSCE) Journal of Civil Engineering 14(2):141-147.

[64] Khanna, N., Kotay, S.M., Gilbert, J.J. and Das, D. 2011. Improvement of biohydrogen production by *Enterobacter cloacae* IIT-BT 08 under regulated pH. Journal of Biotechnology 152:9-15.

[65] Taguchi, F., Hang, J.D., Takiguchi, S. and Morimoto, M. 1992. Efficient hydrogen production from starch by a bacterium isolated from termites. Journal of Fermentation and Bioengineering 73:244-245.

[66] Zhao, X., Xing, D., Liu, B., Lu, L., Zhao, J. and Ren, N. 2012. The effects of metal ions and l-cysteine on hydA gene expression and hydrogen production by *Clostridium beijerinckii* RZF-1108. International Journal of Hydrogen Energy 37:13711-13717.

[67] Chong, M., Abdul, R.N., Yee, P.L., Aziz, S.A., Rahim, R.A., Shirai, Y. *et al.* 2009a. Effects of pH, glucose and iron sulfate concentration on the yield of biohydrogen by *Clostridium butyricum* EB6. International Journal of Hydrogen Energy 34:8859-8865.

[68] Junghare, M., Subudhi, S. and Lal, B. 2012. Improvement of hydrogen production under decreased partial pressure by newly isolated alkaline tolerant anaerobe, *Clostridium butyricum* TM-9A: optimization of process parameters. International Journal of Hydrogen Energy 37:3160-3168.

[69] Liu, B., Ren, N., Xing, D., Ding, J., Zheng, G., Guo, W. et al. 2009. Hydrogen production by immobilized *R. faecalis* RLD-53 using soluble metabolites from ethanol fermentation bacteria *E. harbinense* B49. Bioresource Technology 100:2719-2723.

[70] Tang, J., Yuan, Y., Guo, W. and Ren, N. 2012. Inhibitory effects of acetate and ethanol on biohydrogen production of *Ethanoligenens harbinense* B49. International Journal of Hydrogen Energy 37:741-747.

[71] Guo, W., Ren, N., Wang, X., Xiang, W., Ding, J., You, Y. *et al*. 2009. Optimization of culture conditions for hydrogen production by Ethanoligenens harbinense B49 using response surface methodology. Bioresource Technology 100:1192-1196.

[72] Xu, L., Ren, N., Wang, X. and Jia, Y. 2008. Biohydrogen production by *Ethanoligenens harbinense* B49: nutrient optimization. International Journal of Hydrogen Energy 33:6962-6967.

[73] Yokoi, H., Ohkawara, T., Hirose, J., Hayashi, S. and Takasaki, Y. 1995. Characteristics of hydrogen production by aciduric *Enterobacter aerogenesstrain* HO-39. Journal of Fermentation and Bioengineering 80:571-574.

[74] Kurokawa, T. and Tanisho, S. 2005. Effects of formate on fermentative hydrogen production by *Enterobacter aerogenes*. Marine Biotechnology 7(2):112-118.

[75] Zheng, X.J. and Yu, H.Q. 2004b. Biological Hydrogen Production by Enriched Anaerobic Cultures in the Presence of Copper and Zinc. Journal of Environmental Science and Health A39(1):89-101.

[76] Ding, J., Liu, B.F., Ren, N.Q., Xing, D.F., Guo, W.Q., Xu, J.F. and Xie, G.J. 2009. Hydrogen production from glucose by co-culture of *Clostridium butyricum* and immobilized *Rhodopseudomonas faecalis* RLD-53. International Journal of Hydrogen Energy 34:3647-3652.

[77] Sharma, Y. and Li, B. 2008. Optimizing hydrogen production from organic wastewater treatment in batch reactors through experimental and kinetic analysis. International Journal of Hydrogen Energy 34:6171-6180.

[78] Xu, J.F., Ren, N.Q., Wang, A.J., Qiu, J., Zhao, Q.L., Feng, Y.J. and Liu, B.F. 2010. Cell growth and hydrogen production on the mixture of xylose and glucose using a novel strain of *Clostridium* sp. HR-1 isolated from cow dung compost. International Journal of Hydrogen Energy 35(24):13467-13474.

[79] Ghosh, D. and Hallenbeck, P.C. 2009. Fermentative hydrogen yields from different sugars by batch cultures of metabolically engineered Escherichia coli DJT135. International Journal of Hydrogen Energy 34:7979-7982.

[80] Gadhamshetty, V., Johnson, D.C., Nirmalakhandan, N., Smith, G.B. and Deng, S. 2009. Feasibility of biohydrogen production at low temperatures in unbuffered reactors. International Journal of Hydrogen Energy 34:1233-1243.

[81] Chen, W., Tseng, Z., Lee, K. and Chang, J. 2005. Fermentative hydrogen production with *Clostridium butyricum* CGS5 isolated from anaerobic sewage sludge. International Journal of Hydrogen Energy 30:1063-1070.

[82] Plangklang, P., Reungsang, A. and Pattra, S. 2012. Enhanced biohydrogen production from sugarcane juice by immobilized Clostridium butyricum on sugarcane bagasse. International Journal of Hydrogen Energy 37:15525-15532.

[83] Niu, K., Zhang, X., Tan, W. and Zhu, M. 2010. Characteristics of fermentative hydrogen production with Klebsiella pneumonia ECU-15 isolated from anaerobic sewage sludge. International Journal of Hydrogen Energy 35:71-80.

[84] Lin, C.Y. and Lay, C.H. 2004. Carbon/nitrogen-ratio effect on fermentative hydrogen production by mixed microflora. International Journal of Hydrogen Energy 29:41-45.

[85] Mei, N., Zergane, N., Postec, A., Erauso, G., Ollier, A., Payri, C., Pelletier, B., Fardeau, M.L., Ollivier, B. and Quéméneur, M. 2014. Fermentative hydrogen production by a new alkaliphilic *Clostridium* sp. (strain PROH2) isolated from a shallow submarine hydrothermal chimney in Prony Bay, New Caledonia. International Journal of Hydrogen Energy. 39:19465-19473.

[86] Sun, Q., Xiao, W., Xi, D., Shi, J., Yan, X. and Zhou, Z. 2010. Statistical optimization of biohydrogen production from sucrose by a co-culture of *Clostridium acidisoli* and *Rhodobacter sphaeroides*. International Journal of Hydrogen Energy 35:4076-4084.

[87] Cheng, J., Song, W., Xia, A., Su, H., Zhou, J. and Cen, K. 2012. Sequential generation of hydrogen and methane from xylose by two-stage anaerobic fermentation. International Journal of Hydrogen Energy 37:13323-13329.

[88] Wang, J.L. and Wan, W. 2008d. Effect of Fe^{2+} concentration on fermentative hydrogen production by mixed cultures. International Journal of Hydrogen Energy 33:1215-1220.

[89] Wang, J.L. and Wan, W. 2008e. Influence of Ni^{2+} concentration on biohydrogen production. Bioresource Technology 99:8864-8868.

[90] Wang, J.L. and Wan, W. 2008f. Optimization of fermentative hydrogen production process by response surface methodology. International Journal of Hydrogen Energy 33:6976-6984.

[91] Cheong, D.Y. and Hansen, C.L. 2006. Bacterial stress enrichment enhances anaerobic hydrogen production in cattle manure sludge. Applied Microbiology and Biotechnology 72:635-643.

[92] Lee, Y.J., Miyahara, T. and Noike, T. 2002. Effect of pH on microbial hydrogen fermentation. Journal of Chemical Technology and Biotechnology 77:694-698.

[93] Desai, K.M., Survase, S.A., Saudagar, P.S. *et al.* 2008. Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. Journal of Biochemical Engineering 41:266-273.

[94] Hu, B. and Chen, S.L. 2007. Pretreatment of methanogenic granules for immobilized hydrogen fermentation. International Journal of Hydrogen Energy 32:3266–3273.

[95] Sarkar, O., Kannaiah-Goud, R., Venkata-Subhash, G. and Venkata-Mohan, S. 2013. Relative effect of different inorganic acids on selective enrichment of acidogenic biocatalyst for fermentative biohydrogen production from wastewater. Bioresource Technology 147:321-331.

[96] Yasin, N.H., Mumtaz, T., Hassan, M.A., Rahman, N.A. 2013. Food waste and food processing waste for biohydrogen production: A review Journal for Environmental Management 130:375-385

[97] Chen, P., Wang, Y., Yan, L., Wang, Y., Li, S., Yan, X., Wang, N., Liang, N. and Li, H. 2015. Feasibility of biohydrogen production from industrial wastes using defined microbial co-culture. Biological Research 48:24-31.

[98] Oh, S.E., Iyer, P., Bruns, M.A. and Logan, B.E. 2004. Biological hydrogen production using a membrane bioreactor. Biotechnology and Bioengineering 87:119-127.

[99] Zhang, Z.P., Show, K.Y., Tay, J.H., Liang, D.T., and Lee, D.J. 2008. Biohydrogen production with anaerobic fluidized bed reactors-A comparison of biofilm-based and granule-based systems. International Journal of Hydrogen Energy 33:1559-1564.

[100] Venkata-Mohan, S., Mohanakrishna, G., Goud, R.K., and Sarma, P. 2009b. Acidogenic fermentation of vegetable based market waste to harness biohydrogen with simultaneous stabilization. Bioresource Technology 100:3061-3068.

[101] Sinha, P. and Pandey, A. 2011. An evaluative report and challenges for fermentative biohydrogen production. International Journal of Hydrogen Energy 36:7460-7478.

[102] Ivanova, G., Rakhely, G. and Kovacs, K.L. 2009. Thermophilic biohydrogen production from energy plants by *Caldicellulosiruptor saccharolyticus* and comparison with related studies. International Journal of Hydrogen Energy 34:3659-3670.

[103] Logan, B.E., Call, D., Cheng, S. et al. 2008. Microbial electrolysis cells for high yield hydrogen gas production from organic matter. Environmental Science and Technology 42(23):8630-8640.

[104] Romão, B.B., Batista, F. R. X., Ferreira, J. S., Costa, H. C. B., Resende, M. M. and Cardoso, V. L. 2014. Biohydrogen Production Through Dark Fermentation by a Microbial Consortium Using Whey Permeate as Substrate. Applied Biochemistry and Biotechnology 172:3670-3685.

[105] Levine, S. 1985. Derivation of the Ideal Gas Law. Journal of Chemical Education 62(5):399.

[106] Laugier, A. and Garai, J. 2007. Derivation of the Ideal Gas Law. Journal of Chemical Education. 84(11):1832-1833.

[107] Ioaniss, A.P. 2015. Dark fermentative hydrogen production from lignocellulosic biomass. Fang, Z., Smith Jr., R.L., Qi, X. (eds), Production of Hydrogen from Renewable sources. Biofuels and Biorefineries, Dordrecht Heidelberg, New York, London. pp. 24.

[108] Wirth, R., Kovács, E., Maróti, G., Bagi, Z., Rákhely, G. and Kovács, K.L. 2012. Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. Biotechnology for Biofuels 5:41-56.

[109] Zhang, L, Li, J., Ban, Q., He, J. and Jha, A.K. 2012. Metabolic Pathways of Hydrogen Production in Fermentative Acidogenic Microflora. Journal of Microbiology and Biotechnology 22(5): 668–673.

[110] Gunawan, R., Cao, Y., Petzold, L. *et al.* 2005. Sensitivity analysis of discrete stochastic systems. Biophysical Journal 288(4):2530-2540.

[111] Shojaeefard, M.H., Akbari, M., Tahani, M. *et al.* 2013. Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminum to brass. Advances in Material Science and Engineering 2013:1-7.

[112] Ren, N.Q., Guo, W.Q., Wang, X.J., Xiang, W.S., Liu, B.F. and Wang, X.Z. Effects of different pretreatment methods on fermentation types and dominant bacteria for hydrogen production International Journal of Hydrogen Energy 33:4318-4324.

[113] Oh, S.E., van Ginkel, S. and Logan, B.E. 2003. The relative effectiveness of pH control and heat treatment for enhancing biohydrogen gas production. Environmental Science and Technology 37:5186-5190.

[114] Masset, J., Calusinska, M., Hamilton, C., Hiligsmann, S., Joris, B., Wilmotte, A. and Thonart, P. 2012. Fermentative hydrogen production from glucose and starch using pure strains and artificial co-cultures of *Clostridium* spp. Biotechnology for Biofuels 5(35):1-15.

[115] Bisaillon, A., Turcot, J. and Hallenbeck, P.C. 2006. The effect of nutrient limitation on hydrogen production by batch cultures of *Escherichia coli*. International Journal of Hydrogen Energy. 31:1504-1508.

[116] Seppälä, J.J., Puhakka, J.A., Yli-Harja, O., Karp, M.T. and Santala, V. 2011. Fermentative hydrogen production by *Clostridium butyricum* and *Escherichia coli* in pure and cocultures. International Journal of Hydrogen Energy 36:10701-10708.

[117] Chong, M., Rahim, R.A., Shirai, Y. and Hassan, M.A. 2009b. Biohydrogen production by *Clostridium butyricum* EB6 from palm oil mill effluent. International Journal of Hydrogen Energy 34:764-771.

CHAPTER 5

Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs)

This chapter has been published with the title: Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs) in Biotechnology and Biotechnological Equipment (2015, 29(6):1208-1215)

The published paper is presented in the following pages.

Biotechnology & Biotechnological Equipment, 2015
Vol. 29, No. 6, 1208–1215, http://dx.doi.org/10.1080/13102818.2015.1062732

Taylor & Francis
Taylor & Francis Group

ARTICLE; BIOINFORMATICS

Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs)

Yeshona Sewsynker[a], Evariste Bosco Gueguim Kana[b]* and Agbaje Lateef[b]

[a]*Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa;* [b]*Microbiology Unit, Department of Pure and Applied Biology, Ladoke Akintola University of Technology, Ogbomoso, Nigeria*

(*Received 16 March 2015; accepted 12 June 2015*)

The enhancement of hydrogen yield in microbial electrolysis cells (MECs) requires a robust process model that accurately relates the effect of anodic physicochemical input variables to the process output. Artificial neural networks (ANNs) have been used for the modelling of complex and non-linear processes. This paper reports the modelling of biohydrogen yield in MECs by using a committee of five ANNs. A topology of 6–(6, 8, 11, 12, 14)–1 was adopted, corresponding to the number of neurons of inputs, hidden (varied) and output layers. The ANN inputs were substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration. Model development was carried out with 50 data points from 15 published studies. The coefficients of determination (R^2) between the experimental and predicted hydrogen yields for the five models were as follows: 0.90, 0.81, 0.85, 0.70 and 0.80. Model validation on new MEC processes showed a strong correlation between the observed and predicted hydrogen yields. Sensitivity analysis revealed that the performance of MEC was highly affected by variations in the substrate type, followed by applied voltage, substrate concentration, pH, MEC configuration and temperature in decreasing order. This study showed that the committee model accurately modelled the non-linear relationship between the considered physicochemical parameters of MEC and hydrogen yield, and thus could be used to navigate the optimization window in MEC scale-up processes.

Keywords: biohydrogen production; microbial electrolysis cell; modelling; artificial neural network committee

Introduction

The reliance on conventional fossil fuels has resulted in the imminent energy catastrophe with the combined challenge of global warming and the depletion of these energy reserves.[1] Research on renewable methods for producing energy has received utmost attention in the last few years.[2] Hydrogen is a potential source of energy and thus has gained prominence over contending technologies like ethanol and alternative biofuels.[2] Hydrogen has a high gravimetric energy density of 122 kJ/g, which is roughly 2.9 times greater than that of conventional fossil fuels, such as petroleum (44 kJ/g), gas (52 kJ/g), coal (40 kJ/g), methane (50.1kJ/g) and ethanol (26.5 kJ/g). In addition, water is the only by-product from its combustion.[3]

Presently, biological production of hydrogen is carried out using photo-fermentation, dark fermentation and microbial electrolysis. Photo-fermentation employs both algae and photosynthetic bacteria for hydrogen production and light serves as the energy source. This process is limited by its dependence on light energy.[4] Dark fermentation generates hydrogen via microbial breakdown of organic materials under anaerobic conditions in the absence of light, yielding only 2–3 mol H_2/mol glucose against the expected theoretical yield of 12 mol H_2/mol glucose. This low yield of H_2 is referred to as the 'fermentation barrier.'[5] Microbial electrolysis cells (MECs) are devices, capable of generating hydrogen from waste biomass.[6] MEC systems are associated with the well-known microbial fuel cells (MFC). While MFCs generate electricity from the microbial degradation of organic compounds, MECs, to a certain degree, reverse the process by using bacterial metabolism to generate hydrogen from organic material, by applying an electric voltage.[7]

These systems operate by a process, referred to as electrohydrogenesis.[8] The MEC reactor is primarily made of anode chamber, cathode chamber, membrane and direct current (DC) power supply. During microbial electrolysis, organic matter is employed as an electron donor and is consumed by the exoelectrogenic bacteria at the anodic compartment with simultaneous release of electrons and protons.[9] The electrons pass through the circuit and coalesce with protons at the cathode resulting in the formation of hydrogen. Owing to the applied voltage in MEC systems, volatile fatty acids (VFAs) can be further broken down; thus the hydrogen yield is greatly improved.[5]

*Corresponding author. Email: kanag@ukzn.ac.za

Microbial metabolism in the anodic chamber is greatly influenced by process parameters, such as temperature, substrate type, concentration, pH and MEC configuration. These anodic input parameters have been investigated by several authors in various studies by using the One Variable At a Time (OVAT) approach, or statistical methods. [10–15] These approaches, although widely used, have substantial challenges.[16] For instance, OVAT approach suffers at least two major weaknesses: (1) the interactive effect of parameters on the process are completely ignored; (2) it is impractical for the search of achieving a suitable optimum in a limited amount of experimentations [17] and the factorial design of experiment (DOE) has shown to be time-consuming, resource demanding and laborious when the number of input factors is increased.

MECs, like most bioprocesses, are non-linear, complex and unsteady, so it is difficult to derive an accurate physical-based formula to represent its physical behaviour. Besides, the development of precise bioprocess models remains a crucial challenge for experts, owing to the non-linear nature of the biochemical network interactions. [18] Artificial neural networks (ANNs) are a mathematical interpretation of the neurological functioning of the human brain. ANNs mimic the brain's learning procedure by arithmetically modelling the network structure of interconnected nerve cells.[19] Gueguim Kana et al. [16] indicated that ANNs are appropriate for the development of bioprocess models devoid of previous knowledge relating to the kinetics of metabolic instabilities within the cell and the cultural background. Furthermore, ANNs are completely data-based, without demanding a widespread systematic description of the occurrences that are overriding the process.[20] A classic neural network consists of an input layer, one or more hidden layers and an output layer. Neurons of the hidden layer facilitate the network in determining the intricate relations that exist among the input and output parameters.[19] The efficiency of ANN in bioprocess modelling has been reported in various studies. Nasr et al. [21] developed an ANN model of dark fermentation on inputs of pH, initial substrate and biomass concentrations, temperature and time with a coefficient of determination of 0.98. Wang and Wan [22] studied the effect of temperature, initial pH and substrate concentration on hydrogen yield and substrate degradation efficiency by using an ANN model. Likewise, Mu and Yu [23] generated an ANN model on the continuous flow system functioning and indicated that the model successfully gave an account of the daily disparities of the reactor operation.

Reports of the use of a Committee of ANNs in bioprocess modelling are scanty. To the best of our knowledge, the application of a committee of ANNs in the modelling of biohydrogen yield by using MECs on input parameters of substrate type, substrate concentration, pH, temperature (°C), applied voltage and reactor configuration, has not yet been reported.

In this study, no novel investigation of individual input parameters on the performance of MECs was carried out, but a committee of five ANNs was structured on multilayer perceptron topology and trained on 41 data points from 15 published papers, in order to abstract the pattern relating the substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration to hydrogen yield in MEC systems. Additionally, the relative importance of these inputs on MEC performance was assessed using the ANN committee.

Materials and methods

Experimental data and model development

After an extensive survey of literature on the performance of MECs under different input parameter types and control set points, 50 data points (41 data points for training and 9 for validation) were selected from 15 published works with varied input conditions. The selected input variables consisted of substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration.

A committee made up of five ANNs, as depicted in Figure 1, was implemented using Hypertext Preprocessor (PHP) programming environment. The idea of using a committee machine was that the strategy could lead to a significant improvement of the network prediction performance of new MEC process conditions with little computational efforts. This could be achieved by using ensemble averaging (EA), where the MEC hydrogen outputs are linearly combined, or by using a mixture of experts (ME), where the MEC hydrogen outputs are non-linearly combined. A topology of 6–(6, 8, 11, 12, 14)–1 was adopted,

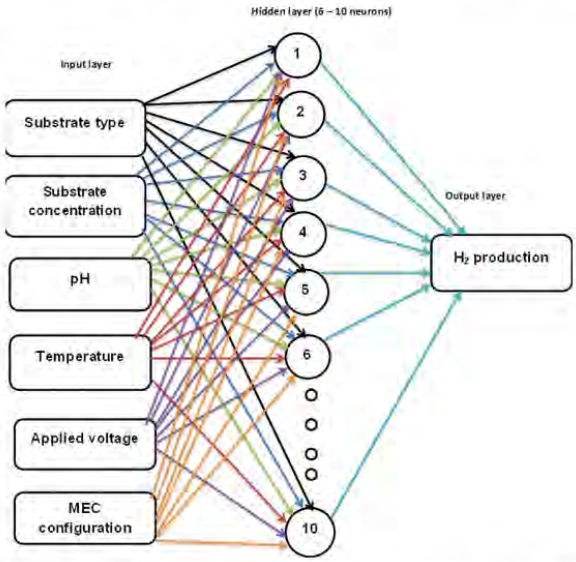


Figure 1. Topology of neural network committee with one input layer (six neurons), one hidden layer (six to ten neurons) and one output layer (one neuron).

corresponding to the number of neurons of inputs, hidden (varied) and output layers (Figure 1). This architecture had a feed forward nature, whereby the input layer neurons transmit signals to the hidden layer neurons.[24]

For the hidden layer, a sigmoid transfer function was implemented. This hidden layer had two main purposes: (1) the addition of the weighted inputs together with the linked bias; (2) subsequently, to shift the input data to a non-linear form, as shown in the following equations [24]:

$$\text{sum} = \sum_i^n = 1^{xiwi} + \theta, \tag{1}$$

where w_i $(i = 1, n)$ are the connection weights, θ is the bias and x_i is the input variable,[24]

$$f(\text{sum}) = \frac{1}{1 + \exp(-\text{sum})}. \tag{2}$$

ANN training and validation

Prior to training the network, the data on substrate type input were transformed to numerical values. This was done based on the molecular weights for each substrate. Thus, acetate (59.04 g/mol), propionic acid (74.08 g/mol), butyric acid (88.11 g/mol), lactic acid (90.08 g/mol), glycerol (92.09 g/mol), valeric acid (102.13 g/mol), cellulose (176.15 g/mol) and glucose (180.16 g/mol) were ranked with the numerical values 1, 2, 3, 4, 5, 6, 7 and 8, respectively. In addition, the training data set (41 data points) was normalized within the range of −0.9 to 0.9 by using the following equation:

$$\text{normalized}\,(e_i) = \frac{e_i - E_{\min}}{E_{\max} - E_{\min}}, \tag{3}$$

where e_i is the normalized data and $E_{\min}$ and $E_{\max}$ are denoting the minimum and maximum values set at −0.9 and 0.9, respectively.

The back propagation (BP) algorithm [25] was used for training. Tables 1 and 2 show the database and model input ranges, respectively. During this process, the error between the experimental (observed) and predicted data was calculated and propagated backward through the network. Consequently, the algorithm adjusted the weights in each consecutive layer to decrease the error. This process was repeated until the error between the experimental and predicted data was reduced below an acceptable threshold (Figure 2).[16,19,25] With this algorithm, individual members of the committee machine learnt the non-linear relationship between considered MEC input parameters and hydrogen yield (mol H_2/mol substrate) by adjusting their synaptic weights in order to reduce the error difference (root-mean-square error (RMSE)) between the predicted output and actual experimental output values (Equation (4)). The committee was trained for 700 epochs to lower the RMSE to an acceptable threshold. The epoch referred to the number of times that all of the training vectors were used once to update the weights:

$$\text{RMSE} = \sqrt{\frac{\sum_i^N \sum_{n=1}^M (y^i n - \hat{y}^i n)^2}{NM}}, \tag{4}$$

Table 2. Ranges for input and output parameters used in the committee model development.

Parameter	Minimum	Maximum	Unit
Substrate type	1	8	–
Substrate concentration	0.1	91	g/L
pH	4.5	9	–
Temperature	4	30	°C
Applied voltage	0.2	1.2	V
*MEC configuration	1	2	–
H_2	0	9	mol H_2/mol substrate

Note: *MEC configuration: single chamber (1); two chamber (2); hydrogen yield (H_2). This was done based on the molecular weights, as described in Table 1.

Table 1. Database used for the development of ANN committee model.

Carbon source	No. of data points	pH	T (°C)	S (g/L)	E_{app} (V)	MEC configuration	Source
Glycerol (5)	2	7	30	1	0.9	1	[11]
Acetate (1)	41	4.9–9	4–30	0.118–90.7	0.2–1.2	1–2	[5,10–14,26–34]
Glucose (8)	2	7	30	1	0.6–0.9	1	[11,5]
Cellulose (7)	1	7	30	1	0.6	2	[5]
Butyric acid (3)	1	7	30	1	0.6	1	[5]
Lactic acid (4)	1	7	30	1	0.6	1	[5]
Propionic acid (2)	1	7	30	1	0.6	1	[5]
Valeric acid (6)	1	7	30	1	0.6	1	[5]

Note: Temperature (T); substrate concentration (S); applied voltage (E_{app}); MEC configuration was designated as (1) for single chamber and (2) for two chambers. Numbers next to substrates were allocated to differentiate between the various types. This was done based on the molecular weights of each substrate thus acetate (59.04 g/mol), propionic acid (74.08 g/mol), butyric acid (88.11 g/mol), lactic acid (90.08 g/mol), glycerol (92.09 g/mol), valeric acid (102.13 g/mol), cellulose (176.15 g/mol) and glucose (180.16 g/mol) were ranked as numerical values 1, 2, 3, 4, 5, 6, 7 and 8, respectively.

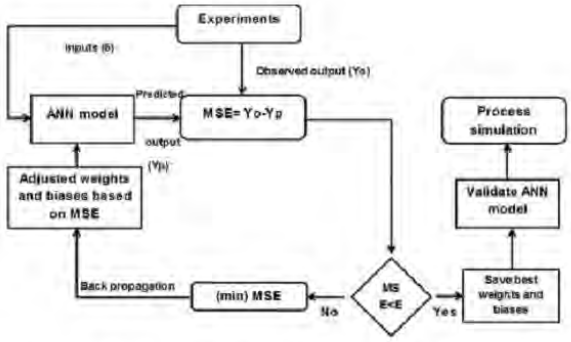


Figure 2. The back propagation training flowchart for artificial neural network. Note: Mean square error (MSE).

where N refers to the number of patterns used in the training, M denotes the number of output nodes, i denotes the index of the input pattern (vector) and y^in and $\hat{y}^in$ are the actual and predicted output values, respectively.

The coefficient of determination (R^2) was calculated as the quotient of the variances of the model predicted values and observed values of the dependent variable. To identify the critical parameters and their degree of importance on the model output, a sensitivity analysis was carried out. This was achieved by calculating the output percent difference when each input parameter varied from its minimum value to its maximum value. This provided information about the more sensitive parameters.

Results and discussion

Effect of anodic process inputs on MEC output

An overview of the database, developed for this study, indicated that acetate, which is an intermediate metabolic product, has been used as a substrate for most of the reported MEC investigations.[5,10−14,26−34] Glucose, glycerol and other organic acids have been used to a lesser extent.[11,5] The maximum H_2 yield obtained by using acetate was 3.8 mol H_2/mol acetate, reported by Tartakovsky et al.[13] Furthermore, it was observed that the substrate concentration in various MEC studies has been relatively maintained in the range of 0.5 and 5 g/L of substrate. Most reports showed that the pH within the anodic chamber was not controlled [9] and the initial reported values were within the range of 6.5−7.0.[35] With regard to the used reactor configuration, the single and double chamber are regularly adopted [5] and the voltage values most commonly applied were within the range of 0.5 V−0.8 V.[10,13,14] Additionally, these processes have been frequently operated with controlled temperature in the range of 30 to 40 °C.[10]

Assessment of the developed ANN committee model

For the initial assessment, the developed ANN committee was used to predict the hydrogen yield on nine data points from the published works of Cheng and Logan,[5] Tartakovsky et al.,[13] Guo et al.,[14] Jia et al. [33] and Lee and Rittmann.[34] These data were not previously used in the development of the committee model.

The second assessment was carried out on 41 data points and the predicted and observed hydrogen yields were compared. The obtained coefficients of determination (R^2) values were 0.90, 0.81, 0.85, 0.70, 0.80 and 0.85 for committee member 1, 2, 3, 4, 5 and the mean/average of the five committee members, respectively. An $R^2 >$ 0.70 is generally regarded as an indication of a good model.[24] The plots of predicted versus observed hydrogen yield values are depicted in Figure 3(a), 3(b), 3(c), 3(d), 3(e) and 3(f), showing most data points aligned near the diagonal, thus illustrating the closeness between the predicted and observed yields.

The relative difference in the obtained coefficients of determination among the various committee members can be accounted for by the difference in the number of neurons in the hidden layer of each committee member. The R^2 values ranged from 0.70 to 0.90, illustrating the susceptibility of the learning efficiency of ANN models of complex non-linear microbial processes on the number of hidden neurons. However, it has not been clearly established whether a higher number of neurons on the hidden layer of the network will enhance or impede its predictive ability on novel bioprocesses. Very often a single neural network member with a fixed number of hidden neurons is structured, trained, validated and used as a model for novel bioprocesses.[16,20−23,36−38]

This approach negates the contribution of poor members and gives greater generalization to the whole model in the prediction of hydrogen yield as a function of substrate type and concentration, pH, temperature, applied voltage and reactor configuration on new MECs.

These findings elucidate that the ANN committee, with several sigmoid transfer functions, is efficiently able to simulate the highly non-linear relationship between the input parameters and biohydrogen yield pattern within the design window. The relative predictive superiority of ANN over alternative bioprocess modelling technique tools has been described by Gueguim Kana et al.,[16] Desai et al.,[24] Rosales-Colunga et al.,[36] Nikhil et al. [37] and Whiteman and Gueguim Kana.[25]

Guo et al. [14] investigated the effect of varying applied voltage on biohydrogen yield by using MECs and achieved a maximum H_2 yield of 3.52 mol H_2/mol acetate. The average value, predicted by the developed committee model for the same experimental conditions, was 4.48 mol H_2/mol substrate. Jia et al. [33] studied the effect of varying applied voltage on biohydrogen yield by using MECs and reported a maximum H_2 yield of 1.64 mol H_2/mol acetate at an applied voltage of 1.2 V, whilst at 1.0 V and 1.15 V the H_2 yield was 1.4 mol H_2 /mol acetate and 1.3 mol H_2 /mol acetate, respectively. The present model under similar conditions gave H_2 yields of

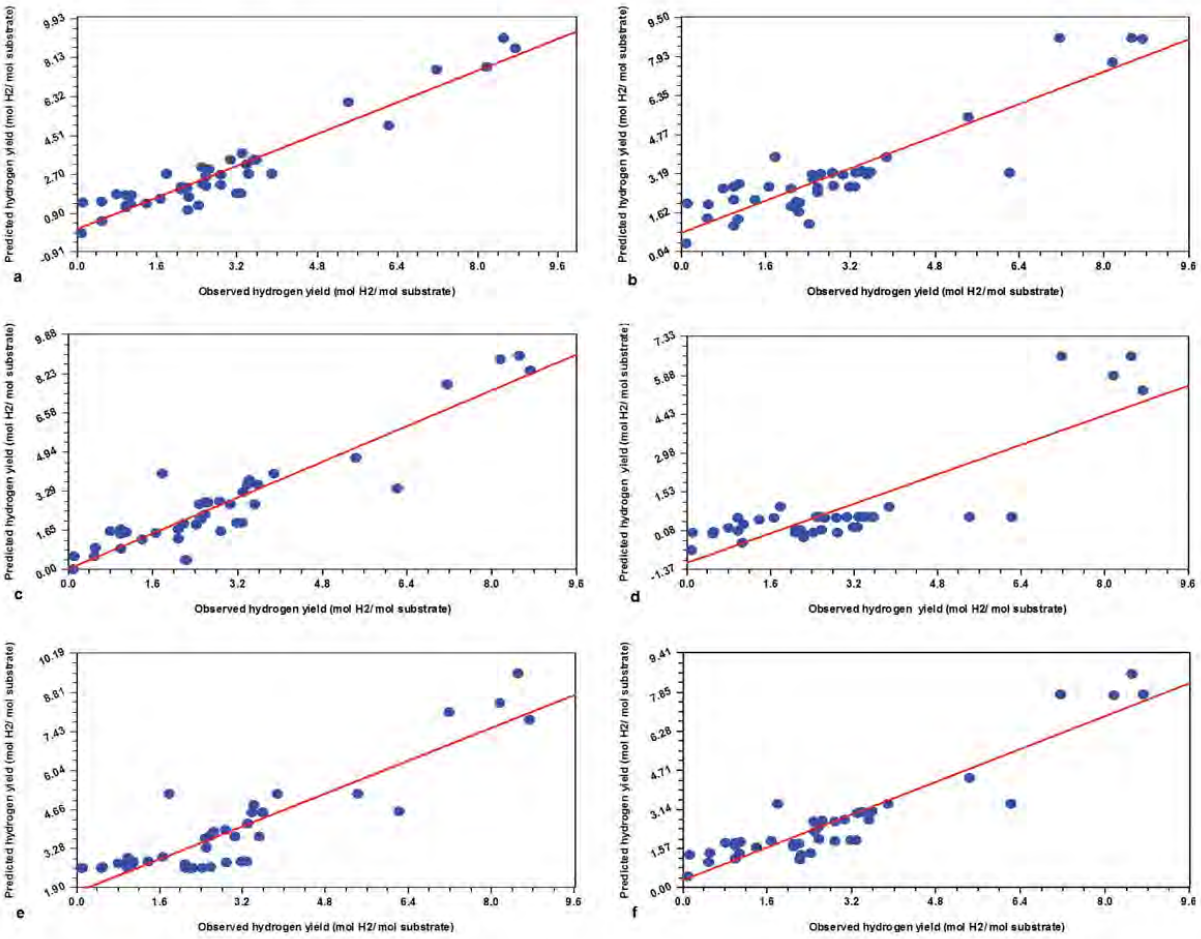


Figure 3. Committee member model 1−5 (a−e) and the average (f) predicted versus observed biohydrogen yields (mol H_2/mol substrate) values for 41 experimental data sets. Note: The diagonal line illustrates expectations under a one-to-one relationship between predicted and observed values. $R^2 = 0.90$ (a); $R^2 = 0.81$ (b); $R^2 = 0.85$ (c); $R^2 = 0.70$ (d); $R^2 = 0.80$ (e); $R^2 = 0.85$ (f).

4.04 mol H_2/mol acetate, 3.71 mol H_2/mol acetate and 3.97 mol H_2/mol acetate for applied voltage values of 1.2 V, 1.0 V and 1.15 V, respectively. The discrepancy between the predicted model and experimental hydrogen yields obtained by Jia et al. [33] could be accounted for by the inoculum type (synthetic wastewater) and the anodic electrode material type of graphite granules used in the experiment, which were not considered in the committed model input space.

A similar study was carried out by Tartakovsky et al. [13] to investigate the effect of applied voltage of 1 V and 1.5 V on MEC hydrogen yield. The present committee model predicted 3.36 mol H_2/mol acetate and 3.20 mol H_2/mol acetate, respectively, against the values of 3.7 mol H_2/mol acetate and 3.8 mol H_2/mol acetate, respectively, obtained by Tartakovsky et al.[13] The difference between the model predicted and experimental hydrogen yields in the study by Tartakovsky et al. [13] could be accounted for by the inoculum type (anaerobic sludge) and the anodic electrode material type of carbon felt used in the experiment, which were not considered in the committee model input space. The apparent discrepancy between the observed hydrogen yield and the predicted values by the committee model in the work of Cheng and Logan [5] (8.01 mol H_2/mol acetate against 5.20 mol H_2/mol acetate) and the study of Lee and Rittmann [34] (4 mol H_2/mol acetate against 4.44 mol H_2/mol acetate) suggest that, in addition to the considered input space, other variables, such as the type of electrodes, inoculum source, among others, impact the overall hydrogen yield in MEC processes.

Nevertheless, a common pattern emerged in the hydrogen yield trend between the predicted and observed data point, as illustrated in Figure 4, but with the average predicted data consistently above the observed values. These observations illustrated the efficiency of the

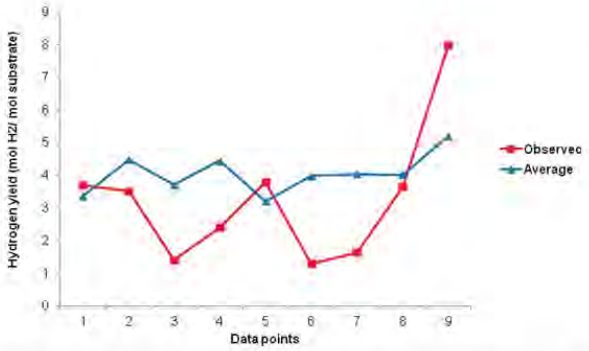


Figure 4. Observed H_2 yield (mol H_2/mol substrate) compared to the average predicted (of the five models) H_2 yield (mol H_2/mol substrate) on the nine data points used for testing of the model.

developed committee model to capture the pattern of hydrogen yield in MEC within the input space of substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration.

Sensitivity analysis

Sensitivity analysis is used to determine how 'sensitive' a model is to the changes in the value of the input parameters or to the changes in the structure of the model. The sensitivity indicators describe the change in the systems' outputs due to variations in the process input parameters. A large sensitivity to a parameter suggests that the process performance can drastically change with little variation in the parameter.[39,40] Vice versa, a small sensitivity suggests little change in the performance of the system even if a large variation occurs in the input parameter.

This analysis was carried out on the developed committee network to determine the relative sensitivity of the hydrogen yield of MEC on the input parameters of substrate type, substrate concentration, pH, temperature, applied voltage and reactor configuration on hydrogen yield in MECs. The data revealed that the hydrogen yield was highly affected by substrate type, followed by applied voltage, substrate concentration, pH, reactor configuration and temperature with a sensitivity of 0.68, 0.37, 0.30, 0.17, 0.15 and 0.032 in decreasing order, respectively. This implied that slight changes in the organic load of the feed stream with regard to its concentration or type, will have a high impact on the hydrogen yield in MEC processes. These findings are in line with the study undertaken by Selembo et al.,[11] who reported a significant difference in MEC yield by using glucose, pure glycerol and glycerol by-products and achieving yields of 7.2 mol H_2/mol glucose, 3.9 mol H_2/mol glycerol and 1.8 mol H_2/mol glycerol, respectively. Similar observations on substrate type and concentration have been reported by Cheng and Logan,[5] Liu et al.,[32] Jia et al. [41] and Tartakovsky et al.[42]

In the same vein, slight changes in the applied voltage have shown to considerably effect the H_2 yield.[31] Hu et al. [31] demonstrated that when maintaining the other parameters at constant set points and varying the applied voltage values between 0.4 **V** and 0.6 **V**, the corresponding H_2 yield was 1.0 mol H_2/mol acetate and 2.5 mol H_2/mol acetate, respectively. Similar trends have been reported by Tartakovsky et al.,[13] Guo et al.,[14] Jia et al.,[33] Call and Logan [43] and Escapa et al.[44]

Conclusions

The development of a reliable bioprocess model, capable of accurately navigating the optimization search space, is highly challenging owing to the non-linearities, associated with these processes. In this work, a committee of ANN was developed by using 15 reported investigations on MEC processes. An average coefficient of determination of 0.85 was achieved on novel MEC processes, not used in model training. Furthermore, the ANN committee revealed that H_2 yield performance of MEC was greatly influenced by variations in substrate type, followed by applied voltage, substrate concentration, pH, MEC configuration and temperature in decreasing order 0.68, 0.37, 0.30, 0.17, 0.15 and 0.032, respectively. This study demonstrated that the ANN committee model efficiently encapsulated the non-linear relationship between the input parameters and the hydrogen yield and thus may serve as a useful device to navigate the optimization search space in MEC scale-up and industrial development in addition to other bioprocesses. As the pressure to transit from fossil fuel sources to renewable ones steadily builds up and various research groups are streaming high value data on novel renewable fuel processes, it is highly likely that bioprocess models, capable of synergistically integrating these data streams, would shorten the energy transition curve.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Research Foundation of South Africa [grant number CPRR 905044]. Opinions, findings and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Research Foundation.

References

[1] Levin DB, Pitt L, Love M. Biohydrogen production: prospects and limitations to practical application. Int J Hydrogen Energy. 2004;29:173–185.
[2] Nath K, Das D. Modeling and optimization of fermentative hydrogen production. Bioresour Technol. 2011;102:8569–8581.

[3] Belafi-Bakó K, Bakonyi P, Nemestóthy N, et al. Biohydrogen production in integrated system. Desalin Water Treat. 2010;14:116−118.
[4] Das D, Veziroglu TN. Advances in biological hydrogen production processes. Int J Hydrogen Energy. 2008;33: 6046−6057.
[5] Cheng SA, Logan BE. Sustainable and efficient biohydrogen production via electrohydrogenesis. Proc Natl Acad Sci USA. 2007;104(47):18871−18873.
[6] Ambler JR, Logan BE. Evaluation of stainless steel cathodes and a bicarbonate buffer for hydrogen production in microbial electrolysis cells using a new method for measuring gas production. Int J Hydrogen Energy. 2011;36: 160−166.
[7] Logan BE, Regan JM. Electricity-producing bacterial communities in microbial fuel cells. Trends Microbiol. 2006;14:512−518.
[8] Logan BE, Call D, Cheng S, et al. Microbial electrolysis cells for high yield hydrogen gas production from organic matter. Environ Sci Technol. 2008;42(23):8630−8640.
[9] Zhu G, Wu T, Jha AK, et al. Review of biohydrogen production and new application in the pollution control via microbial electrolysis cell. Desalin Water Treat. 2014;52 (28−30):5413−5421.
[10] Kyazze G, Popov A, Dinsdale R, et al. Influence of catholyte pH and temperature on hydrogen production from acetate using a two chamber concentric tubular microbial electrolysis cell. Int J Hydrogen Energy. 2010;35: 7716−7722.
[11] Selembo PA, Perez JM, Lloyd WA, et al. High hydrogen production from glycerol or glucose by electrohydrogenesis using microbial electrolysis cells. Int J Hydrogen Energy. 2009;34:5373−5381.
[12] Lu L, Ren N, Zhao X, et al. Hydrogen production, methanogen inhibition and microbial community structures in psychrophilic single chamber microbial electrolysis cells. Energy Environ Sci. 2011;4:1329−1336.
[13] Tartakovsky B, Manuel MF, Wang H, Guiot SR. High rate membrane-less microbial electrolysis cell for continuous hydrogen production. Int J Hydrogen Energy. 2009;34:672−677.
[14] Guo K, Tang X, Du Z, et al. Hydrogen production from acetate in a cathode-on-top single-chamber microbial electrolysis cell with a mipor cathode. Biochem Eng J. 2010;51:48−52.
[15] Azwar Y, Abdul-Wahab AK, Mohamed Azlan Hussain MA. Optimal production of biohydrogen gas via microbial electrolysis cells (MEC) in a controlled batch reactor system. Chem Eng J. 2013;32:727−732.
[16] Gueguim Kana EB, Oloke JK, Lateef A, et al. Modelling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renew Energy. 2012;46:276−281.
[17] LOTFY AL, Khaled M, Ehab R, et al. Citric acid production by a novel *Aspergillus niger* isolate: II. Optimization of process parameters through statistical experimental designs. Bioresour Technol. 2007;98:3470−3477.
[18] Franco-Lara E, Link H, Weuster-Botz D. Evaluation of artificial neural networks for modeling and optimization of medium composition with a genetic algorithm. Process Biochem. 2006;41:2200−2206.
[19] Nagata U, Chu KH. Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnol Lett. 2003;25:1837−1842.
[20] Shi Y, Gai G, Zhao X, et al. Back propagation neural network (BPNN) simulation model and influence of operational parameters on hydrogen bio-production through integrative biological reactor (IBR) treating wastewater. Paper presented at: Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference. 2010 June 18−20; Chengdu (China).
[21] Nasr N, Hafez H, El Naggar MH, et al. Application of artificial neural networks for biohydrogen production. Int J Hydrogen Energy. 2013;38:3189−3195.
[22] Wang J, Wan W. Application of desirability function based on neural network for optimizing biohydrogen production process. Int J Hydrogen Energy. 2009;34:1253−1259.
[23] Mu Y, Yu HQ. Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. Int J Hydrogen Energy. 2007;32:3308−3314.
[24] Desai KM, Survase SA, Saudagar PS, et al. Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. J Biochem Eng. 2008;41:266−273.
[25] Whiteman JK, Gueguim Kana EB. Comparative assessment of the artificial neural network and response surface modeling efficiencies for biohydrogen production on sugar cane molasses. Bioenergy Res. 2014;7(1):295−305.
[26] Liu H, Grot S, Logan BE. Electrochemically assisted microbial production of hydrogen from acetate. Environ Sci Technol. 2005;39:4317−4320.
[27] Chae KJ, Choi MJ, Lee J, et al. Biohydrogen production via biocatalyzed electrolysis in acetate-fed bioelectrochemical cells and microbial community analysis. Int J Hydrogen Energy. 2008;33:5184−5192.
[28] Chae KJ, Choi MJ, Kim KY, et al. Selective inhibition of methanogens for the improvement of biohydrogen production in microbial electrolysis cells. Int J Hydrogen Energy. 2010;35:13379−13386.
[29] Nam JY, Logan BE. Enhanced hydrogen generation using a saline catholyte in a two chamber microbial electrolysis cell. Int J Hydrogen Energy. 2011;36:15105−15110.
[30] Nam JY, Logan BE. Optimization of catholyte concentration and anolyte pHs in two chamber microbial electrolysis cells. Int J Hydrogen Energy. 2012;37:18622−18628.
[31] Hu H, Fan F, Liu H. Hydrogen production using single-chamber membrane-free microbial electrolysis cells. Water Res. 2008;42:4172−4178.
[32] Liu W, Huang S, Zhou A, et al. Hydrogen generation in microbial electrolysis cell feeding with fermentation liquid of waste activated sludge. Int J Hydrogen Energy. 2012;37:13859−13864.
[33] Jia YH, Ryu JH, Kim CH, et al. Enhancing hydrogen production efficiency in microbial electrolysis cell with membrane electrode assembly cathode. J Ind Eng Chem. 2012;18:715−719.
[34] Lee HS, Rittmann BE. Characterization of energy losses in an upflow single-chamber microbial electrolysis cell. Int J Hydrogen Energy. 2010;35:920−927.
[35] Wang AJ, Liu WZ, Ren NQ, Zhou JZ, Cheng SA. Key factors affecting microbial anode potential in a microbial electrolysis cell for H_2 production. Int J Hydrogen Energy. 2010;35:13481−13487.
[36] Rosales-Colunga LM, Garcia RG, Rodriguez AD. Estimation of hydrogen production in genetically modified *Escherichia coli* fermentations using an artificial neural network. Int J Hydrogen Energy. 2010;35:13186−13192.

[37] Nikhil BO, Visa A, Lin CY, et al. An artificial neural network based model for predicting H2 production rates in a sucrose-based bioreactor system. W.A.S.E.T. 2008;37:20−25.
[38] Prakasham RS, Sathish T, Brahmaiah P. Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. Int J Hydrogen energy. 2011;36: 4332−4339.
[39] Gunawan R, Cao Y, Petzold L, et al. Sensitivity analysis of discrete stochastic systems. Biophys J. 2005;88(4): 2530−2540.
[40] Shojaeefard MH, Akbari M, Tahani M, et al. Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminum to brass. Adv Mater Sci Eng. 2013;2013:1−7.
[41] Jia YH, Choi JY, Ryu JH, et al. Hydrogen production from wastewater using a microbial electrolysis cell. Korean J Chem Eng. 2010;27(6):1854−1859.
[42] Tartakovsky B, Mehta P, Santoyo G, et al. Maximizing hydrogen production in a microbial electrolysis cell by real time optimization of applied voltage. Int J Hydrogen Energy. 2011;36:10557−10564.
[43] Call D, Logan BE. Hydrogen production in a single chamber microbial electrolysis cell lacking a membrane. Environ Sci Technol. 2008;42:3401−3406.
[44] Escapa A, Manuel MF, Moran A, et al. Hydrogen production from glycerol in a membraneless microbial electrolysis cell. Energy Fuels. 2009;23:4612−4618.

CHAPTER 6

Conclusions and Recommendations for Further Research

6.1. Conclusions

In this study, the impact of experimental process volume size on the efficiency of Artificial Neural Networks (ANN) and Response Surface Methodology was investigated at process scales of 80 and 800 mL. In addition, ANN models were developed for biohydrogen prediction from dark fermentation and microbial electrolysis using existing knowledge in public repositories. Major findings and their significance are summarized as follows:

6.1.1. The impact of process volume size on bioprocess modelling efficiency was assessed using RSM and ANN for hydrogen production. This was carried out across two process scales (80 and 800 mL). Results showed that there was no significant difference ($p>0.05$) between modelling at two different bioprocess scales (80 and 800 mL) for biohydrogen production. ANN based models achieved higher coefficient of determination (R^2) values (0.99 and 0.95) compared to RSM based models (0.97 and 0.89) for 80 and 800 mL, respectively. In addition, the lowest prediction error (2.25 %) was observed for the ANN model at a process volume of 80 mL. Thus, ANN had a higher modelling and optimization efficiency compared to RSM for complex, non-linear systems.

6.1.1.1. Semi-pilot scales at 8 L process volume for all four optimized conditions showed negligible deviations from their corresponding flask volumes. Therefore process miniaturization does not impact on the accuracy of ANN and RSM derived process models thus, this reduces the process development time and costs.

6.1.1.2. Microbial community analysis of the semi-pilot scale process carried out at peak hydrogen production phase revealed that presumptive hydrogen-producing microorganisms within this system were members of the genus *Clostridia*, *Enterobacter* and *Klebsiella*.

6.1.2. Two intelligent bioprocess models were implemented using ANN on 64 selected reported studies for fermentative hydrogen processes based on two yield expression units (mL H_2/ g substrate and mol H_2/ mol substrate). A high coefficient of determination (R^2) was

obtained for the cumulative volume of hydrogen per gram substrate model (Vol_Model) (mL $H_{2/}$ g substrate) (0.90) whereas a low value was observed with the mole of hydrogen per mole of substrate model (Mol_Model) (0.46). These findings showed that the Vol_Model efficiently abstracted the non-linear relationship between the considered inputs and hydrogen yield with a higher prediction accuracy on novel biohydrogen experiments. Thus, these ANN derived models could be used to predict hydrogen yields on novel experimental inputs or to navigate the optimization space and shorten the biohydrogen process development time.

6.1.3. A committee of ANN models was developed using 15 selected reported investigations on MEC processes. The coefficient of determination (R^2) between the experimental and predicted hydrogen yields for the five models were as follows: 0.90, 0.81, 0.85, 0.70 and 0.80. An average R^2 value of 0.85 was obtained for the five models. Validation on new MEC processes showed a strong correlation between the observed and predicted hydrogen yields. Findings showed that the committee of networks accurately modelled the non-linear relationship between the considered physicochemical parameters of MEC and hydrogen yield, and thus could be used to determine the optimum set points in MEC scale-up processes.

6.2. Recommendations for future studies

Based on the findings derived from this study, the following recommendations can be made for future research on biohydrogen process development:

6.2.1. Dark fermentative hydrogen production may be integrated with other processes such as biodiesel, biogas, microbial fuel cell technology (Microbial Fuel Cells and Microbial Electrolysis Cells) and bioethanol by the use of a two-stage system. This may assist in the achievement of higher substrate conversion and energy efficiency when using substrates such as sugarcane molasses.

6.2.2. A standard unit of expression for reporting hydrogen yields should be used to enable the inter-laboratory reproducibility within the research community. This will enhance hydrogen process development towards commercialization.

6.2.3. The application of Artificial Intelligence (AI) tools such as ANN on existing process data provide virtual experimentations for dark fermentative hydrogen production and

Microbial Electrolysis Cells (MECs). This could significantly lower the time and costs of process development.

6.2.4. Improvement in the capability of the hydrogen-producing microorganisms by using metabolic engineering and immobilization techniques for higher hydrogen yields in addition to the utilization of low-cost materials with regard to substrate type, reactor configurations and modes of production. This will enhance the industrial feasibility of the biohydrogen production process as it will significantly reduce costs associated with upstream, production and downstream processes.