PREDICTING CROP YIELD AND FIELD ENERGY OUTPUT FOR OIL PALM USING GENETIC ALGORITHM AND NEURAL NETWORK MODELS

YOUSIF YAKOUB HILAL ALHILAL

FK 2019 36
PREDICTING CROP YIELD AND FIELD ENERGY OUTPUT FOR OIL PALM USING GENETIC ALGORITHM AND NEURAL NETWORK MODELS

By

YOUSIF YAKOUB HILAL ALHILAL

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

January 2019
COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs, and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia
Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

PREDICTING CROP YIELD AND FIELD ENERGY OUTPUT FOR OIL PALM USING GENETIC ALGORITHM AND NEURAL NETWORK MODELS

By

YOUSIF YAKOUB HILAL ALHILAL

January 2019

Chairman : Professor Ir. Azmi b. Dato’ Yahya, PhD
Faculty : Engineering

For many years, the Malaysian oil palm industry has been facing the challenge of reduced rate of palm oil yield due to sizeable difference between the crop’s actual yield and the crop’s genetic yield potential. This gap has grown wider over time and has been of great concern since oil palm is a very important commodity that contributes significantly to the country’s GDP. Currently, Malaysia has devoted a high percentage of the land resource and material inputs to agriculture, whereby a large proportion of them are used for oil palm cultivation. However, the typical yields are only 50–60% of the potential, and artificial intelligence research on modelling of the crop yield and energy consumption is still at its infancy.

Forecasting oil palm production and selecting significant variables that effects production are complex activities. Accurate prediction results are required for this type of analysis and can provide the basis for the decisions and plans for the management of agricultural crops in the local, regional, and global scale. In the field of agricultural engineering, artificial intelligence has helped to reduced operational periods and costs. There was not enough information available on the implementation of neural networks and genetic algorithm for the prediction and selecting input variables in oil palm yield and output energy.

This research presents the development of a GA and SW as a variables selection method in ANN and NARX models for predicting oil palm yield and output energy. Data were collected from 11 districts in 11 states in Malaysia for FFB and PO models, which includes Kedah, Kelantan, Johor, Melaka, Penang, Pahang, Perak, Selangor, Terengganu, Sabah, and Sarawak. The study is based on monthly data from 2005 to 2015. In FFB and PO models, the data used 15 variables, namely: percentage of mature area and percentage of immature area, rainfall, rainy days, humidity, radiation,
temperature, surface wind speed, evaporation and cloud cover, O3, CO, NO2, SO2, and PM10. The study used input energy data from 8 variables for developing energy models. These data included human power, electricity, fuel, water, fertilizers, and seed. Data were collected from Peninsular Malaysia, Sabah and Sarawak over a period of 11 years (annual data from 2005 to 2015).

Results showed that GA was able to select the variables correctly, while also being an easy-to-use variable selection tool. It proved to be more effective than the Stepwise. The findings of this research, using 11 years of climate change and air pollution, have significantly affected the oil palm production. Surface wind speed and humidity were recorded at an impact ratio of up to 100%, which correlated negatively on the productivity of oil palm plantations. Surface wind speed and humidity reduced the productivity of oil palm FFB plantations for 5.12 and 4.61 ton/ha/11year in the Sabah and Sarawak respectively. Additionally, the surface wind speed is considered the most essential variable recorded with an impact ratio of up to 100% on FFB in Selangor, Terengganu, and Kelantan while the cloud cover, average NO2 in the air, average PM10 in the air, humidity, radiation, and O3 recorded the most significant impact up to 100% on FFB in Perak, Melaka, Johor, Kedah, Penang, and Pahang respectively.

Fuel consumption, water, and P-fertilizer consumption are considered the most important variables in oil palm plantation operations, its importance being the relative values of 45%, 34.3 %, and 23 %. These variables impacted oil palm operation during the 11 years at 67.764, 45.38, 16.24 GJ /ha for Peninsular Malaysia, Sabah and Sarawak, respectively. In this study, the performances of six models (namely, ANN, GA-ANN, SW-ANN, NARX, GA-NARX and SW-NARX) are compared with one another as well as with multiple linear models. The GA-NARX was chosen as the best yield model in 9 states (Perak, Sabah, Sarawak, Selangor, Terengganu, Pahang, Kedah, Kelantan and Penang), while the GA-ANN was considered the best yield model recorded in Melaka and Johor. Additionally the GA-NARX was chosen as the best energy model in Peninsular Malaysia, Sarawak and Sabah, with the average accuracy percentage simulation being 0.95.07, 95.55 and 87.43 % respectively.

Finally, this research concluded that a genetic algorithm is useful for selecting input variables in oil palm production. It is a user-friendly variable selection tool with excellent results compared to Stepwise, especially in a large search space. The GA-ANN and GA-NARX models perform markedly better than the other models in the most training algorithms with different numbers of hidden layers.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MERAMALKAN HASIL TANAMAN DAN TENAGA OUTPUT LAPANGAN UNTUK MINYAK KELAPA SAWIT MENGGUNAKAN ALGORITMA GENETIK DAN MODEL RANGKAIAN NEURAL

Oleh

YOUSIF YAKOUB HILAL

Januari 2019

Pengerusi : Profesor Ir. Azmi b. Dato’ Yahya, PhD
Fakulti : Kejuruteraan

Selama beberapa tahun industri kelapa sawit Malaysia menghadapi cabaran penurunan kadar hasil kelapa sawit disebabkan oleh perbezaan ketara antara hasil sebenar tanaman dengan kemampuan hasil genetik tanaman. Perbezaan ini telah membesar dengan masa dan telah menjadi satu kebimbangan disebabkan kelapa sawit merupakan komoditi yang terpenting yang menyumbang secara signifikan kepada KDNK negara. Buat masa ini, Malaysia telah memperuntukkan peratusan yang tinggi dalam sumber tanah dan input bahan untuk pertanian dimana kadar yang tersangat besar digunakan untuk penanaman kelapa sawit. Walau bagaimanapun, hasil lazimnya kelapa adalah hanya 50–60% daripada hasil potensinya manakala penyelidikan kecerdasan buatan dalam pemodelan hasil tanaman dan penggunaan tenaga adalah masih pada peringkat awal.

Peramalan pengeluaran kelapa sawit dan pemilihan pemboleh ubah yang signifikan dalam memberi kesan pada pengeluaran adalah akiviti yang kompleks. Dapat ramalan yang tepat diperlukan untuk analisis dan yang dapat memberikan asas bagi keputusan dan perancangan bagi pengurusan tanaman pertanian pada skala tempatan, regional, dan global. Dalam bidang kejuruteraan pertanian, kemajuan dalam kecerdasan buatan telah dapat membantu dalam mengurangkan jatul masa kendalian dan kos. Tiada terdapat pengetahuan yang mendalam mengenai menggunakan rangkaian neural dan algoritma genetik dalam ramalan dan pemilihan pemboleh ubah input bagi hasil kelapa sawit dan keluaran tenaga.

Penyelidikan ini membentangkan pembangunan GA dan SW sebagai kaedah pemilihan pembolehuhubah dalam ANN dan NARX model untuk meramalkan hasil kelapa sawit dan keluaran tenaga. Data dikumpulkan dari 11 daerah di 11 negeri di Malaysia bagi model FFB dan PO, termasuk Kedah, Kelantan, Johor, Melaka, Pulau Pinang, Pahang, Perak,

Keputusan menunjukkan bahawa GA dapat memilih pembolehubah dengan betul, sementara ia juga merupakan alat pemilihan pembolehubah yang mudah digunakan. Ia terbukti lebih efektif daripada Stepwise. Dapatan kajian ini yang menggunakan 11 tahun perubahan cuaca dan pencemaran udara yang mempengaruhi pengeluaran kelapa sawit telah dilaporkan adalah signifikan. Kelajuan dan kelembapan permukaan angin direkodkan pada kesan nisbah sehingga 100%, yang berkorlasi negatif terhadap produktiviti ladang kelapa sawit. Kelajuan dan kelembapan permukaan angin telah mengurangkan produktiviti ladang FFB kelapa sawit masing-masing 5.12 dan 4.61 ton/ha/11 tahun di Sabah dan Sarawak. Tambahan pula, kelajuan permukaan angin direkodkan sebagai pembolehubah yang paling penting yang direkodkan dengan kesan nisbah sehingga 100% pada FFB di Selangor, Terengganu, dan Kelantan manakala penutup awan, purata NO₂ di udara, purata PM₁₀ di udara, kelembapan, radiasi, dan O₃ mencatat kesan yang paling signifikan sehingga 100% pada FFB masing-masing di Perak, Melaka, Johor, Kedah, Pulau Pinang dan Pahang.

Penggunaan bahan api, air, dan penggunaan P-baja dianggap pembolehubah yang paling penting dalam operasi ladang kelapa sawit, kepentingannya adalah nilai relatif 45%, 34.3%, dan 23%. Pembolehubah ini telah mempengaruhi operasi kelapa sawit selama 11 tahun di 67.764, 45.38, 16.24 GJ / ha masing-masing di Semenanjung Malaysia, Sabah dan Sarawak. Dalam kajian ini, prestasi enam model (iaitu, ANN, GA-ANN, SW-ANN, NARX, GA-NARX dan SW-NARX) dapat dibandingkan antara satu sama lain dan dengan model berganda. GA-NARX dipilih sebagai model hasil terbaik di 9 buah negeri (Perak, Sabah, Sarawak, Selangor, Terengganu, Pahang, Kedah, Kelantan dan Pulau Pinang) manakala GA-ANN dianggap model hasil terbaik yang direkodkan di Melaka dan Johor. Selain itu, GA-NARX telah dipilih sebagai model tenaga terbaik di Semenanjung Malaysia, Sarawak dan Sabah dengan purata peratusan ketepatan simulasi masing-masing iaitu 0.95.07, 95.55 dan 87.43%.

Akhirnya, kajian ini telah menyimpulkan bahawa algoritma genetik adalah berguna untuk memilih pembolehubah input dalam pengeluaran kelapa sawit. Ia adalah pilihan alat pembolehubah yang mesra pengguna dengan keputusan yang cemerlang berbanding dengan Stepwise terutamanya dalam ruang carian yang besar. Model GA-ANN dan GA-NARX adalah ketara lebih baik daripada model lain dalam kebanyakan algoritma latihan dengan bilangan lapisan tersembunyi yang berbeza.
ACKNOWLEDGEMENTS

In the Name of Allah, the Merciful, the Most Merciful

Praise be to Allah, Lord of the Worlds, and Blessings be upon the Noblest of Prophets and Messengers, our Prophet Mohammad. I thank the God for giving me strength, patience, motivation and special people that aided me in the completion of this research work.

Firstly, I wish to acknowledge with the high level of appreciation the unquantifiable, timely and valuable contributions of the chairman of my supervisory committee, Prof. Dr. Azmi bin Dato Yahya for his guidance, advice, and support during my research and writing of the thesis. The knowledge and experience that I have gained working under his supervision will be of great value to me throughout the rest of life. Similarly, the valuable contributions of the members of my supervisory committee in persons of Prof. Wan Ismail Wan Ishak and Dr. Zulfa Hanan Ash'ari are highly acknowledged.

My appreciation also goes to my country Iraq and the Universiti Putra Malaysia for providing funds for the research project under the Research University Grant GP-IPS/2017/9516000. A great thanks to the Faculty of Engineering, the Department of Biological and Agricultural Engineering, Malaysian Palm Oil Board, Meteorological Department, Department of Statistics Malaysian Prime Minister’s Department, Department of Environment Ministry of Natural Resources & Environment and International Fertilizer Association that provided the working environment for performing this work.

I thank the assistant director in Department of Statistics Malaysian Prime Minister’s Department, Nurul Aini Binti Abdul Wahab for extended immediate helping hand and enormous support. I also want to appreciate the kind assistance from Dr. Jean-Marc Roda of CIRAD, Prof. Dr. Shamsuddin Jusop of Department of Land Management, Dr. Noor Azura Bt Ismail for director-general Malaysian Meteorological Department and Dr. Zulkifli Hashim, Head agronomy and Geospatial Technology Unit, MPOB.

I also wish to appreciate the support from my friends, Dr. Mohammad Jawaid, and Dr. Khalid Hakeem. Last but not least to my parents and my wife and children for their continued emotional support, prayers and advice and also for their help in various ways.
I certify that a Thesis Examination Committee has met on 30 January 2019 to conduct the final examination of Yousif Yakoub Hilal Alhilal on his thesis entitled "Predicting Crop Yield and Field Energy Output for Oil Palm Using Genetic Algorithm and Neural Network Models" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**Desa bin Ahmad, PhD**
Professor Ir.
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

**Samsuzana binti Abd Aziz, PhD**
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

**Rimfiel b Janius, PhD**
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

**Lilik Sutiarso, PhD**
Professor
Universitas Gadjah Mada
Indonesia
(External Examiner)

\[Signature\]

**RUSLI HAJI ABDULLAH, PhD**
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 April 2019
This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

**Azmi Yahya, PhD**
Professor, Ir
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

**Wan Ishak Wan Ismail, PhD**
Professor, Ir
Faculty of Engineering
Universiti Putra Malaysia
(Member)

**Zulfa Hanan Asha'ari, PhD**
Senior Lecturer
Faculty of Environmental Studies
Universiti Putra Malaysia
(Member)

---

**ROBIAH BINTI YUNUS, PhD**
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:
Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: _______________________  Date: __________________

Name and Matric No.: Yousif Yakoub Hilal, GS41852
Declaration by Members of Supervisory Committee

This is to confirm that:
• the research conducted and the writing of this thesis was under our supervision;
• supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) were adhered to.

Signature: ____________________________
Name of Chairman of Supervisory Committee: Professor Ir Dr. Azmi Yahya

Signature: ____________________________
Name of Member of Supervisory Committee: Professor Ir Dr. Wan Ishak Wan Ismail

Signature: ____________________________
Name of Member of Supervisory Committee: Dr. Zulfa Hanan Asha'ari
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRAK</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>APPROVAL</td>
<td>vi</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xxii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xxix</td>
</tr>
</tbody>
</table>
2.7.1 Structure and Operation of a Basic Genetic Algorithm

2.7.2 Genetic Algorithm Applications

2.8 Neural Networks

2.8.1 The Neuron and Transfer Function

2.8.2 Network Architecture and Number of Layers

2.8.3 Learning Rules and Training Function

2.8.4 Neural Network Types

2.8.5 Classical Artificial Neural Networks

2.8.6 Time-Series NARX Neural Networks

2.8.7 Applications of ANN and NARX

2.9 Modelling

2.9.1 Oil Palm Yield Modeling

2.9.2 Oil Palm Energy Modeling

2.10 Summary

3 MATERIALS AND METHODS

3.1 Introduction

3.2 General Information about the Sites Analysis

3.3 Collection of Input and Output Data

3.4 Data Validation and Exploration

3.5 Development of the Model Selection Based on GA for Crop Yield and Energy

3.6 Development of the Model Selection Based on Stepwise for Crop Yield and Energy

3.7 Development of ANN Models for Crop Yield and Energy

3.8 Development of NARX Models for Crop Yield and Energy

3.9 Development of Linear Models (Mathematical Models) for Crop Yield and Energy

3.10 Forecasting Simulation Models

3.11 Sensitivity Test for the Optimal Model

4 RESULTS AND DISCUSSION

4.1 Introduction

4.2 Development of Oil Palm Yield Models

4.2.1 Kedah

4.2.1.1 Mathematical Model

4.2.1.2 Selection Variables

4.2.1.3 ANN and NARX Models

4.2.1.4 GA with ANN and NARX Models

4.2.1.5 Stepwise with ANN and NARX Models

4.2.1.6 Simulation and Selecting Model

4.2.1.7 Sensitivity Analysis

4.2.2 Kelantan

4.2.2.1 Mathematical Model

4.2.2.2 Selection Variables

4.2.2.3 ANN and NARX Models

4.2.2.4 GA with ANN and NARX Models
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.2.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>108</td>
</tr>
<tr>
<td>4.2.2.6</td>
<td>Simulation and Selecting Model</td>
<td>111</td>
</tr>
<tr>
<td>4.2.2.7</td>
<td>Sensitivity Analysis</td>
<td>112</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Johor</td>
<td>113</td>
</tr>
<tr>
<td>4.2.3.1</td>
<td>Mathematical Models</td>
<td>114</td>
</tr>
<tr>
<td>4.2.3.2</td>
<td>Selection Variables</td>
<td>116</td>
</tr>
<tr>
<td>4.2.3.3</td>
<td>ANN and NARX Models</td>
<td>119</td>
</tr>
<tr>
<td>4.2.3.4</td>
<td>GA with ANN and NARX Models</td>
<td>121</td>
</tr>
<tr>
<td>4.2.3.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>124</td>
</tr>
<tr>
<td>4.2.3.6</td>
<td>Simulation and Selecting Model</td>
<td>127</td>
</tr>
<tr>
<td>4.2.3.7</td>
<td>Sensitivity Analysis</td>
<td>128</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Melaka</td>
<td>130</td>
</tr>
<tr>
<td>4.2.4.1</td>
<td>Mathematical Model</td>
<td>130</td>
</tr>
<tr>
<td>4.2.4.2</td>
<td>Selection Variables</td>
<td>133</td>
</tr>
<tr>
<td>4.2.4.3</td>
<td>ANN and NARX Models</td>
<td>136</td>
</tr>
<tr>
<td>4.2.4.4</td>
<td>GA with ANN and NARX Models</td>
<td>138</td>
</tr>
<tr>
<td>4.2.4.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>140</td>
</tr>
<tr>
<td>4.2.4.6</td>
<td>Simulation and Selecting Model</td>
<td>143</td>
</tr>
<tr>
<td>4.2.4.7</td>
<td>Sensitivity Analysis</td>
<td>144</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Penang</td>
<td>146</td>
</tr>
<tr>
<td>4.2.5.1</td>
<td>Mathematical Model</td>
<td>146</td>
</tr>
<tr>
<td>4.2.5.2</td>
<td>Selection Variables</td>
<td>148</td>
</tr>
<tr>
<td>4.2.5.3</td>
<td>ANN and NARX Models</td>
<td>150</td>
</tr>
<tr>
<td>4.2.5.4</td>
<td>GA with ANN and NARX Models</td>
<td>152</td>
</tr>
<tr>
<td>4.2.5.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>154</td>
</tr>
<tr>
<td>4.2.5.6</td>
<td>Simulation and Selecting Model</td>
<td>156</td>
</tr>
<tr>
<td>4.2.5.7</td>
<td>Sensitivity Analysis</td>
<td>158</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Perak</td>
<td>159</td>
</tr>
<tr>
<td>4.2.6.1</td>
<td>Mathematical Model</td>
<td>160</td>
</tr>
<tr>
<td>4.2.6.2</td>
<td>Selection Variables</td>
<td>162</td>
</tr>
<tr>
<td>4.2.6.3</td>
<td>ANN and NARX Models</td>
<td>165</td>
</tr>
<tr>
<td>4.2.6.4</td>
<td>GA with ANN and NARX Models</td>
<td>167</td>
</tr>
<tr>
<td>4.2.6.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>169</td>
</tr>
<tr>
<td>4.2.6.6</td>
<td>Simulation and Selecting Model</td>
<td>171</td>
</tr>
<tr>
<td>4.2.6.7</td>
<td>Sensitivity Analysis</td>
<td>173</td>
</tr>
<tr>
<td>4.2.7</td>
<td>Pahang</td>
<td>174</td>
</tr>
<tr>
<td>4.2.7.1</td>
<td>Mathematical Model</td>
<td>174</td>
</tr>
<tr>
<td>4.2.7.2</td>
<td>Selection Variables</td>
<td>177</td>
</tr>
<tr>
<td>4.2.7.3</td>
<td>ANN and NARX Models</td>
<td>179</td>
</tr>
<tr>
<td>4.2.7.4</td>
<td>GA with ANN and NARX Models</td>
<td>181</td>
</tr>
<tr>
<td>4.2.7.5</td>
<td>Stepwise with ANN and NARX Models</td>
<td>183</td>
</tr>
<tr>
<td>4.2.7.6</td>
<td>Simulation and Selecting Model</td>
<td>185</td>
</tr>
<tr>
<td>4.2.7.7</td>
<td>Sensitivity Analysis</td>
<td>187</td>
</tr>
</tbody>
</table>
4.2.8 Selangor
4.2.8.1 Mathematical Model
4.2.8.2 Selection Variables
4.2.8.3 ANN and NARX Models
4.2.8.4 GA with ANN and NARX Models
4.2.8.5 Stepwise with ANN and NARX Models
4.2.8.6 Simulation and Selecting Model
4.2.8.7 Sensitivity Analysis

4.2.9 Terengganu
4.2.9.1 Mathematical Model
4.2.9.2 Selection Variables
4.2.9.3 ANN and NARX Models
4.2.9.4 GA with ANN and NARX Models
4.2.9.5 Stepwise with ANN and NARX Models
4.2.9.6 Simulation and Selecting Model
4.2.9.7 Sensitivity Analysis

4.2.10 Borneo (Sabah and Sarawak) Malaysia’s
4.2.10.1 Mathematical Model
4.2.10.2 Selection Variables
4.2.10.3 ANN and NARX Models
4.2.10.4 GA with ANN and NARX Models
4.2.10.5 Stepwise with ANN and NARX Models
4.2.10.6 Simulation and Selecting Model
4.2.10.7 Sensitivity Analysis

4.3 Development of Energy Models
4.3.1 Mathematical Model
4.3.2 Selection Variables
4.3.3 ANN and NARX Models
4.3.4 GA with ANN and NARX Models
4.3.5 Stepwise with ANN and NARX Models
4.3.6 Simulation and Selecting Model
4.3.7 Sensitivity Analysis

4.4 Summary of results

5 CONCLUSIONS AND RECOMMENDATIONS
5.1 Conclusions
5.2 Recommendations
5.3 Research Contribution

REFERENCES
APPENDICES
BIODATA OF STUDENT
LIST OF PUBLICATIONS
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Wind properties in relation to suitability for oil palm growing</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Human energy in agricultural operation</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>Equivalent water energy in oil palm plantation</td>
<td>26</td>
</tr>
<tr>
<td>3.1</td>
<td>Description of the study sites</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Input and output data for FFB and PO models</td>
<td>52</td>
</tr>
<tr>
<td>3.3</td>
<td>Input and output data for output energy models</td>
<td>54</td>
</tr>
<tr>
<td>3.4</td>
<td>Energy coefficients of different inputs and outputs used in oil palm production</td>
<td>55</td>
</tr>
<tr>
<td>3.5</td>
<td>LM training parameters with their default values</td>
<td>68</td>
</tr>
<tr>
<td>3.6</td>
<td>RP training parameters with their default values</td>
<td>69</td>
</tr>
<tr>
<td>3.7</td>
<td>GDX training parameters with their default values</td>
<td>70</td>
</tr>
<tr>
<td>4.1</td>
<td>ANOVA for response surface linear model FFB in Kedah.</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>ANOVA for response surface linear model PO in Kedah</td>
<td>81</td>
</tr>
<tr>
<td>4.3</td>
<td>The performance of the mathematical model in Kedah</td>
<td>82</td>
</tr>
<tr>
<td>4.4</td>
<td>The model selection for Kedah</td>
<td>86</td>
</tr>
<tr>
<td>4.5</td>
<td>Performance of model selection by Stepwise in Kedah</td>
<td>86</td>
</tr>
<tr>
<td>4.6</td>
<td>The performance of the ANN models in Kedah</td>
<td>87</td>
</tr>
<tr>
<td>4.7</td>
<td>The performance of the NARX models in Kedah</td>
<td>88</td>
</tr>
<tr>
<td>4.8</td>
<td>The performance of the GA-ANN models in Kedah</td>
<td>90</td>
</tr>
<tr>
<td>4.9</td>
<td>The performance of the GA-NARX models in Kedah</td>
<td>90</td>
</tr>
<tr>
<td>4.10</td>
<td>The performance of the SW-ANN models in Kedah</td>
<td>93</td>
</tr>
<tr>
<td>4.11</td>
<td>The performance of the SW-NARX models in Kedah</td>
<td>93</td>
</tr>
</tbody>
</table>
4.12 Simulation models of FFB in Kedah
4.13 Simulation models of PO in Kedah
4.14 ANOVA for Response Surface Linear model FFB in Kelantan
4.15 ANOVA for Response Surface Linear model PO in Kelantan
4.16 The performance of the mathematical model in Kelantan
4.17 The model selection for Kelantan
4.18 Performance of model selection by Stepwise in Kelantan
4.19 The performance of the ANN models in Kelantan
4.20 The performance of the NARX models in Kelantan
4.21 The performance of the GA-ANN models in Kelantan
4.22 The performance of the GA-NARX models in Kelantan
4.23 The performance of the SW-ANN models in Kelantan
4.24 The performance of the SW-NARX models in Kelantan
4.25 Simulation models of FFB in Kelantan
4.26 Simulation models of PO in Kelantan
4.27 ANOVA for Response Surface Linear model FFB in Johor
4.28 ANOVA for Response Surface Linear model PO in Johor
4.29 The performance of the mathematical model in Johor
4.30 The model selection for Johor
4.31 Performance of model selection by Stepwise in Johor
4.32 The performance of the ANN models in Johor
4.33 The performance of the NARX models in Johor
4.34 The performance of the GA-ANN models in Johor
4.35 The performance of the GA-NARX models in Johor
4.36 The performance of the SW-ANN models in Johor 125
4.37 The performance of the SW-NARX models in Johor 125
4.38 Simulation models of FFB in Johor 127
4.39 Simulation models of PO in Johor 128
4.40 ANOVA for Response Surface Linear model FFB in Melaka 131
4.41 ANOVA for Response Surface Linear model PO in Melaka 131
4.42 The performance of the mathematical model in Melaka 133
4.43 The model selection for Melaka 135
4.44 Performance of model selection by Stepwise in Melaka 135
4.45 The performance of the ANN models in Melaka 136
4.46 The performance of the NARX models in Melaka 137
4.47 The performance of the GA-ANN models in Melaka 138
4.48 The performance of the GA-NARX models in Melaka 139
4.49 The performance of the SW-ANN models in Melaka 141
4.50 The performance of the SW-NARX models in Melaka 142
4.51 Simulation models of FFB in Melaka 143
4.52 Simulation models of PO in Melaka 144
4.53 ANOVA for Response Surface Linear model FFB in Penang 146
4.54 ANOVA for Response Surface Linear model PO in Penang 147
4.55 The performance of the mathematical model in Penang 147
4.56 The model selection for Penang 150
4.57 Performance of model selection by Stepwise in Penang 150
4.58 The performance of the ANN models in Penang 151
4.59 The performance of the NARX models in Penang 152
4.60  The performance of the GA-ANN models in Penang 153
4.61  The performance of the GA-NARX models in Penang 154
4.62  The performance of the SW-ANN models in Penang 155
4.63  The performance of the SW-NARX models in Penang 156
4.64  Simulation models of FFB in Penang 157
4.65  Simulation models of PO in Penang 157
4.66  ANOVA for Response Surface Linear model FFB in Perak 160
4.67  ANOVA for Response Surface Linear model PO in Perak 161
4.68  The performance of the mathematical model in Perak 161
4.69  The model selection for Perak 165
4.70  Performance of model selection by Stepwise in Perak 165
4.71  The performance of the ANN models in Perak 166
4.72  The performance of the NARX models in Perak 166
4.73  The performance of the GA-ANN models in Perak 168
4.74  The performance of the GA-NARX models in Perak 168
4.75  The performance of the SW-ANN models in Perak 170
4.76  The performance of the SW-NARX models in Perak 170
4.77  Simulation models of FFB in Perak 172
4.78  Simulation models of PO in Perak 172
4.79  ANOVA for Response Surface Linear model FFB in Pahang 175
4.80  ANOVA for Response Surface Linear model PO in Pahang 175
4.81  The performance of the mathematical model in Pahang 176
4.82  The model selection for Pahang 179
4.83  Performance of model selection by Stepwise in Pahang 179
4.84 The performance of the ANN models in Pahang 180
4.85 The performance of the NARX models in Pahang 181
4.86 The performance of the GA-ANN models in Pahang 182
4.87 The performance of the GA-NARX models in Pahang 182
4.88 The performance of the SW-ANN models in Pahang 184
4.89 The performance of the SW-NARX models in Pahang 184
4.90 Simulation models of FFB in Pahang 186
4.91 Simulation models of PO in Pahang 186
4.92 ANOVA for Response Surface Linear model FFB in Selangor 189
4.93 ANOVA for Response Surface Linear model PO in Selangor 190
4.94 The performance of the mathematical model in Selangor 190
4.95 The model selection for Selangor 193
4.96 Performance of model selection by Stepwise in Selangor 193
4.97 The performance of the ANN models in Selangor 194
4.98 The performance of the NARX models in Selangor 194
4.99 The performance of the GA-ANN models in Selangor 196
4.100 The performance of the GA-NARX models in Selangor 196
4.101 The performance of the SW-ANN models in Selangor 197
4.102 The performance of the SW-NARX models in Selangor 198
4.103 Simulation models of FFB in Selangor 199
4.104 Simulation models of PO in Selangor 200
4.105 ANOVA for Response Surface Linear model FFB in Terengganu 202
4.106 ANOVA for Response Surface Linear model PO in Terengganu 203
4.107 The performance of the mathematical model in Terengganu 203
<table>
<thead>
<tr>
<th>Section Number</th>
<th>Section Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.108</td>
<td>The model selection for Terengganu</td>
<td>206</td>
</tr>
<tr>
<td>4.109</td>
<td>Performance of model selection by Stepwise in Terengganu</td>
<td>206</td>
</tr>
<tr>
<td>4.110</td>
<td>The performance of the ANN models in Terengganu</td>
<td>207</td>
</tr>
<tr>
<td>4.111</td>
<td>The performance of the NARX models in Terengganu</td>
<td>208</td>
</tr>
<tr>
<td>4.112</td>
<td>The performance of the GA-ANN models in Terengganu</td>
<td>209</td>
</tr>
<tr>
<td>4.113</td>
<td>The performance of the GA-NARX models in Terengganu</td>
<td>209</td>
</tr>
<tr>
<td>4.114</td>
<td>The performance of the SW-ANN models in Terengganu</td>
<td>211</td>
</tr>
<tr>
<td>4.115</td>
<td>The performance of the SW-NARX models in Terengganu</td>
<td>211</td>
</tr>
<tr>
<td>4.116</td>
<td>Simulation models of FFB in Terengganu</td>
<td>212</td>
</tr>
<tr>
<td>4.117</td>
<td>Simulation models of PO in Terengganu</td>
<td>213</td>
</tr>
<tr>
<td>4.118</td>
<td>ANOVA for Response Surface Linear model FFB in Sabah</td>
<td>215</td>
</tr>
<tr>
<td>4.119</td>
<td>ANOVA for Response Surface Linear model PO in Sabah</td>
<td>216</td>
</tr>
<tr>
<td>4.120</td>
<td>ANOVA for Response Surface Linear model FFB in Sarawak</td>
<td>216</td>
</tr>
<tr>
<td>4.121</td>
<td>ANOVA for Response Surface Linear model PO in Sarawak</td>
<td>217</td>
</tr>
<tr>
<td>4.122</td>
<td>The performance of the mathematical model in Sabah</td>
<td>219</td>
</tr>
<tr>
<td>4.123</td>
<td>The performance of the mathematical model in Sarawak</td>
<td>219</td>
</tr>
<tr>
<td>4.124</td>
<td>The model selection for Sabah and Sarawak</td>
<td>222</td>
</tr>
<tr>
<td>4.125</td>
<td>Performance of model selection by Stepwise in Sabah and Sarawak</td>
<td>222</td>
</tr>
<tr>
<td>4.126</td>
<td>The performance of the ANN models in Sabah</td>
<td>223</td>
</tr>
<tr>
<td>4.127</td>
<td>The performance of the ANN models in Sarawak.</td>
<td>224</td>
</tr>
<tr>
<td>4.128</td>
<td>The performance of the NARX models in Sabah</td>
<td>225</td>
</tr>
<tr>
<td>4.129</td>
<td>The performance of the NARX models in Sarawak</td>
<td>225</td>
</tr>
<tr>
<td>4.130</td>
<td>The performance of the GA-ANN models in Sabah</td>
<td>227</td>
</tr>
<tr>
<td>4.131</td>
<td>The performance of the GA-ANN models in Sarawak</td>
<td>227</td>
</tr>
</tbody>
</table>
4.132 The performance of the GA-NARX models in Sabah 228
4.133 The performance of the GA-NARX models in Sarawak 228
4.134 The performance of the SW-ANN models in Sabah 231
4.135 The performance of the SW-ANN models in Sarawak 231
4.136 The performance of the SW-NARX models in Sabah 231
4.137 The performance of the SW-NARX models in Sarawak 232
4.138 Simulation models of FFB in Sabah 233
4.139 Simulation models of PO in Sabah 234
4.140 Simulation models of FFB in Sarawak 234
4.141 Simulation models of PO in Sarawak 235
4.142 ANOVA for Response Surface Linear model energy in Peninsular Malaysia 238
4.143 ANOVA for Response Surface Linear model energy in Sabah 239
4.144 ANOVA for Response Surface Linear model energy in Sarawak 239
4.145 The performance of the mathematical model 240
4.146 The model selection 244
4.147 Performance of energy model selection by Stepwise 244
4.148 The performance of the ANN models in Peninsular Malaysia 246
4.149 The performance of the ANN models in Sabah. 246
4.150 The performance of the ANN models at Sarawak 247
4.151 The performance of the NARX models 248
4.152 The performance of the GA-ANN models in Peninsular Malaysia 250
4.153 The performance of the GA-ANN models in Sabah 250
4.154 The performance of the GA-ANN models in Sarawak 251
4.155 The performance of the GA-NARX models 251
<p>| 4.156 | The performance of the SW-ANN models in Peninsular Malaysia | 254 |
| 4.157 | The performance of the SW-ANN models in Sabah | 254 |
| 4.158 | The performance of the SW-ANN models at Sarawak | 255 |
| 4.159 | The performance of the SW-NARX models | 255 |
| 4.160 | Final prediction models for FFB, PO and output energy | 264 |</p>
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Global oils and fats in 2015</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Oil palm efficiency vs. other major oil crops</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Global palm oil production from 2005 to 2015</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>World consumption of major oils and fats in 2016</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>PO production in Malaysia and Indonesia</td>
<td>14</td>
</tr>
<tr>
<td>2.6</td>
<td>Global palm oil production in 2015</td>
<td>15</td>
</tr>
<tr>
<td>2.7</td>
<td>Distribution of the yield of FFB and PO during 41 years</td>
<td>16</td>
</tr>
<tr>
<td>2.8</td>
<td>Global price for FFB and PO during 11 years</td>
<td>16</td>
</tr>
<tr>
<td>2.9</td>
<td>The curve of yield phases</td>
<td>18</td>
</tr>
<tr>
<td>2.10</td>
<td>Fertilizer nutrient (NPK) consumption</td>
<td>26</td>
</tr>
<tr>
<td>2.11</td>
<td>Monthly demand for germination seeds 2014/2015</td>
<td>28</td>
</tr>
<tr>
<td>2.12</td>
<td>Taxonomy of IVS</td>
<td>30</td>
</tr>
<tr>
<td>2.13</td>
<td>The flowchart of the genetic algorithm</td>
<td>33</td>
</tr>
<tr>
<td>2.14</td>
<td>Neural Network Architectures</td>
<td>36</td>
</tr>
<tr>
<td>2.15</td>
<td>Main architectures of neural networks</td>
<td>37</td>
</tr>
<tr>
<td>2.16</td>
<td>Block diagram of NARX</td>
<td>40</td>
</tr>
<tr>
<td>2.17</td>
<td>Sample NARX network series-parallel architecture (open - loop) structure</td>
<td>41</td>
</tr>
<tr>
<td>2.18</td>
<td>Sample NARX network parallel network (closed - loop)</td>
<td>41</td>
</tr>
<tr>
<td>2.19</td>
<td>Analysis energy consumption in agricultural production</td>
<td>46</td>
</tr>
<tr>
<td>3.1</td>
<td>Overall methodology of the work</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Oil palm planted area by states</td>
<td>51</td>
</tr>
</tbody>
</table>
3.3 Steps of the approach to fill the missing values
3.4 The flow chart of the GA-CA programmer
3.5 The flow chart of the Stepwise programmer
3.6 Input and output variables of the ANN architecture
3.7 Input and output variables by GA and Stepwise in the ANN architecture
3.8 Log and Tan - sigmoid transfer function
3.9 Snapshot of NARX model
3.10 Input and output variables of the NARX architecture
3.11 Input and output variables by GA and Stepwise in the NARX architecture
3.12 Analysis of the Variables Importance
4.1 The scatter plot of the predicted vs. actual for FFB model in Kedah
4.2 The scatter plot of the predicted vs. actual for PO model in Kedah
4.3 The best fitness value of genetic algorithm performance in Kedah
4.4 The scatter plots of ANN and NARX model predicted vs. actual values in Kedah
4.5 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Kedah
4.6 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Kedah
4.7 The most significant independent variables for Kedah
4.8 The scatter plot of the predicted vs. actual for FFB model in Kelantan
4.9 The scatter plot of the predicted vs. actual for PO model in Kelantan
4.10 The best fitness value of genetic algorithm performance in Kelantan
4.11 The scatter plots of ANN and NARX models predicted vs. actual values in Kelantan
4.12 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Kelantan

4.13 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Kelantan

4.14 The most significant independent variables for Kelantan

4.15 The scatter plot of the predicted vs. actual for FFB model in Johor

4.16 The scatter plot of the predicted vs. actual for PO model in Johor

4.17 The best fitness value of genetic algorithm performance in Johor

4.18 The scatter plots of ANN and NARX models predicted vs. actual values in Johor

4.19 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Johor

4.20 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Johor

4.21 The most significant independent variables for Johor

4.22 The scatter plot of the predicted vs. actual for FFB model in Melaka

4.23 The scatter plot of the predicted vs. actual for PO model in Melaka

4.24 The best fitness value of genetic algorithm performance in Melaka

4.25 The scatter plots of ANN and NARX models predicted vs. actual values in Melaka

4.26 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Melaka

4.27 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Melaka

4.28 The most significant independent variables for Melaka

4.29 The scatter plot of the predicted vs. actual for FFB model in Penang

4.30 The scatter plot of the predicted vs. actual for PO model in Penang

4.31 The best fitness value of genetic algorithm performance in Penang
4.32 The scatter plots of ANN and NARX models predicted vs. actual values in Penang

4.33 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Penang

4.34 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Penang

4.35 The most significant independent variables for Penang

4.36 The scatter plot of the predicted vs. actual for FFB model in Perak.

4.37 The scatter plot of the predicted vs. actual for PO model in Perak

4.38 The best fitness value of genetic algorithm performance in Perak

4.39 The scatter plots of ANN and NARX models predicted vs. actual values in Perak

4.40 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Perak

4.41 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Perak

4.42 The most significant independent variables for Perak

4.43 The scatter plot of the predicted vs. actual for FFB model in Pahang

4.44 The scatter plot of the predicted vs. actual for PO model in Pahang

4.45 The best fitness value of genetic algorithm performance in Pahang

4.46 The scatter plots of ANN and NARX models predicted vs. actual values in Pahang

4.47 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Pahang

4.48 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Pahang

4.49 The most significant independent variables for Pahang

4.50 The scatter plot of the predicted vs. actual for FFB model in Selangor

4.51 The scatter plot of the predicted vs. actual for PO model in Selangor.
4.52 The best fitness value of genetic algorithm performance in Selangor

4.53 The scatter plots of ANN and NARX models predicted vs. actual values in Selangor

4.54 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Selangor

4.55 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Selangor

4.56 The most significant independent variables for Selangor

4.57 The scatter plot of the predicted vs. actual for FFB model in Terengganu

4.58 The scatter plot of the predicted vs. actual for PO model in Terengganu

4.59 The best fitness value of genetic algorithm performance in Terengganu

4.60 The scatter plots of ANN and NARX models predicted vs. actual values in Terengganu

4.61 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Terengganu

4.62 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Terengganu

4.63 The most significant independent variables for Terengganu

4.64 The scatter plot of the predicted vs. actual for FFB model in Sabah

4.65 The scatter plot of the predicted vs. actual for PO model in Sabah

4.66 The scatter plot of the predicted vs. actual for FFB model in Sarawak

4.67 The scatter plot of the predicted vs. actual for PO model in Sarawak

4.68 The best fitness value of genetic algorithm performance in Sabah

4.69 The best fitness value of genetic algorithm performance in Sarawak

4.70 The scatter plots of ANN and NARX models predicted vs. actual values in Sabah
4.71 The scatter plots of ANN and NARX models predicted vs. actual values in Sarawak
4.72 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Sabah
4.73 The scatter plots of GA-ANN and GA-NARX models predicted vs. actual values in Sarawak
4.74 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Sabah
4.75 The scatter plots of SW-ANN and SW-NARX models predicted vs. actual values in Sarawak
4.76 The most significant independent variables for Sabah
4.77 The most significant independent variables for Sarawak
4.78 The scatter plot of the predicted vs. actual for energy model in Peninsular Malaysia
4.79 The scatter plot of the predicted vs. actual for energy model in Sabah
4.80 The scatter plot of the predicted vs. actual for energy model in Sarawak
4.81 The best fitness value of genetic algorithm performance in Peninsular Malaysia
4.82 The best fitness value of genetic algorithm performance in Sabah
4.83 The best fitness value of genetic algorithm performance in Sarawak
4.84 The scatter plots of ANN models predicted vs. actual energy values
4.85 The scatter plots of NARX models predicted vs. actual energy values
4.86 The scatter plots of GA-ANN models predicted vs. actual energy values
4.87 The scatter plots of GA-NARX models predicted vs. actual energy values
4.88 The scatter plots of SW-ANN models predicted vs. actual energy values

xxvii
4.89 The scatter plots of SW-NARX models predicted vs. actual energy values

4.90 Simulation models of output energy in Peninsular Malaysia

4.91 Simulation models of output energy in Sabah

4.92 Simulation models of output energy in Sarawak

4.93 The most significant independent variables for Peninsular Malaysia

4.94 The most significant independent variables for Sabah

4.95 The most significant independent variables for Sarawak
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAR</td>
<td>Applied Agricultural Resources</td>
</tr>
<tr>
<td>AAP</td>
<td>Average Accuracy percentage</td>
</tr>
<tr>
<td>Af</td>
<td>Worldwide zones of Tropical rainforest climate</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>API</td>
<td>Air Pollutants Included</td>
</tr>
<tr>
<td>APSIM</td>
<td>Agricultural Production Systems Simulator</td>
</tr>
<tr>
<td>ASAE</td>
<td>American Society of Association Executives</td>
</tr>
<tr>
<td>AR</td>
<td>Autoregressive model</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive–Moving-Average models</td>
</tr>
<tr>
<td>ARX</td>
<td>Autoregressive with exogenous terms model</td>
</tr>
<tr>
<td>BP</td>
<td>Back Propagation Algorithm</td>
</tr>
<tr>
<td>CA</td>
<td>Correlation Analysis</td>
</tr>
<tr>
<td>R</td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>CPO</td>
<td>Crude Palm Oil</td>
</tr>
<tr>
<td>FFB</td>
<td>Fresh Fruit Bunch</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GHG</td>
<td>Greenhouse Gas</td>
</tr>
<tr>
<td>GDX</td>
<td>Gradient Descent with momentum and adaptive learning rate algorithm</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>H</td>
<td>Number of neurons at hidden layers</td>
</tr>
<tr>
<td>IFA</td>
<td>International Fertilizer Association</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IVI</td>
<td>Independent Variable Importance</td>
</tr>
<tr>
<td>IVS</td>
<td>input variable selection</td>
</tr>
<tr>
<td>LM</td>
<td>Levenberg-Marquardt backpropagation algorithm</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>MPOB</td>
<td>Malaysian Palm Oil Board</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>NARX</td>
<td>Nonlinear Autoregressive Exogenous Neural Network</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>OPSIM</td>
<td>Operations Simulator</td>
</tr>
<tr>
<td>PO</td>
<td>Palm Oil</td>
</tr>
<tr>
<td>RP</td>
<td>Resilient backpropagation algorithm</td>
</tr>
<tr>
<td>SW</td>
<td>Stepwise</td>
</tr>
<tr>
<td>TDL</td>
<td>Tapped Delay Line</td>
</tr>
<tr>
<td>Te.</td>
<td>Testing Phase</td>
</tr>
<tr>
<td>Tr.</td>
<td>Training Phase</td>
</tr>
<tr>
<td>USDA</td>
<td>United States Department of Agriculture</td>
</tr>
<tr>
<td>Val.</td>
<td>Validation Phase</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background

Oil palm is one of the most significant agricultural products in southeast Asia, Africa and South America. The oil palm tree (*Elaeis guineensis jacq*) is a monocotyledonous perennial plant indigenous to West Africa. The consumption of its product, palm oil, goes back as far as 5000 years to ancient Egypt. Oil palms are widely grown in more than 43 countries located mainly between 10° N and 10° S of the equator (Corley & Tinker, 2016). Today, products of oil palm feed are consumed in excess by three billion people in more than 150 countries worldwide. It is interesting to note that palm oil and its many derivative products are available in more than 40% of the packaged products in the world, and palm kernel meal is popularly used as fertilizer and livestock feed. Lately, there has been a growing demand for renewable energy; consequently, palm oil is widely used in the production of biodiesel (Fairhurst & Griffiths, 2014).

Oil palm allocations in Indonesia (10 million hectares), Malaysia (5 million hectares) and Nigeria (3 million hectares) comprise approximately 10% of the world’s permanent cropland. Malaysia and Indonesia have been the hubs of this vibrant development, with both countries increasing oil palm land usage by 40% and 150%, respectively, over the last 10 years. Together, they now supply more than 80% of the world’s palm oil production (FAO, 2016).

The local demand for vegetable oils is predicted to grow by 36% within the next 10 years, with biofuels making up one-third of the increase. As worldwide requirements continue to increase, available land decreases in the traditional production centres (USDA, 2015). The worldwide need for palm oil for consumption purposes is predicted to continue to increase as a result of world population growth, higher per capita consumption, and the developed world’s significant shift away from unhealthy animal fats to healthy vegetable oils. In the 2014 to 2015 period, for both the EU and the US, per capita consumption of oils and fats was 60.5 kg and 59 kg, respectively. This represents a large number of consumption in comparison to developing countries such as India, Pakistan, and Nigeria, which possesses per capita consumption for oil and fats of 16.3 kg, 21.7 kg and 14 kg, respectively. With increasing income levels in the developing world, there will be a need for greater production levels of vegetable oil to satisfy the increased demand. The oil palm fruits are considered the most efficient oilseed to meet the increased need as they are of relatively high productivity. Average annual growth in world vegetable oil production between 1990 and 2015 was palm oil (7.03%), Rapeseed oil (5.2%), Soybean oil (4.6%) and Sunflower oil (4.4%). Palm oil production was 11 and 63.5 million tons in 1990 and 2015, respectively. Its share in food use grew from 32.5% in 2013 and surpassed 34% by 2015 (Oil World, 2016).
The Malaysian agricultural sector has become one of the pillars of the national economy of Malaysia. The increase in demand for palm oil has resulted in the rapid growth of the agro-industry within Southeast Asian countries, particularly Malaysia who leads the way in production. The Malaysian government has emphasized the development and further expansion of oil palm plantations in its five-year plans in order to make the industry a leading contributor to the growth of the national economy (Otieno et al., 2016).

Another important reason for the Malaysian government’s emphasis on the oil palm industry is to use it as an opportunity to reduce the poverty level in rural communities. Towards this end, the rural community has been encouraged to actively participate in the palm oil production activities for additional income and for a better life. The aim of the five-year plans is to increase the industry’s gross national income contribution from the present RM 52.7 billion to RM178.0 billion by 2020. On the other hand, the oil palm plantations are currently facing a productivity gap due to various farmers possessing varying farming backgrounds apart from having to face the challenge of change and environmental pollution (Barcelos et al., 2015; MPC, 2017).

Oil palm is widely cultivated in several parts of the country (Otieno et al., 2015). Currently, oil palm is planted on 5.23 million hectares, constituting 15.8% of the total land area and more than 70% of agricultural land in the country (Otieno et al., 2016). Malaysia produces up to 19 million tons of palm oil, with an export of nearly 89% (USDA, 2015). It currently contributes 39% of global palm oil production and 44% of world exports. In terms of total oils and fats, Malaysia’s contribution to the global total is 12% of production and 27% of exports of oils and fats. Ranked among the largest producers and exporters of palm oil and its derived products, Malaysia plays a crucial role in meeting the increasing universal requirement for oils and fats sustainably (Shanmuganathan et al., 2014).

1.2 Oil Palm Yield and Energy

Oil palm provides the highest potential yield per hectare of all sources of vegetable oil. It is capable of producing double the amount of oil compared to rapeseed and almost four times more than soybeans, groundnut and sunflower per hectare per year. As of 2014, the United Nations have estimated that, under the ideal management of high-yield breeding programmers, different varieties of oil palm can produce more than 20 tons of FFBs per ha per year which translates to over five tons of oil per ha per year. About 10% of the dry biomass of the crop comprises of the oils while 90% comprises of cellulosic material and fibre which can be used as second-generation materials for the production of biofuel (Barcelos et al., 2015). In Malaysia, there were 18 biodiesel plants producing a total of 2.34 million ton/yr. Both Johor and Selangor recorded five biodiesel plants operating in 2015, with a total production capacity of 0.99 million tons/yr. for Johor, followed by Selangor with 0.42 million ton/yr. Remaining eight plants with total annual production capacity of 0.92 million tons are located in Pahang, Perak, Sabah and Sarawak (MPOB, 2017).
The 2015 average FFB yield for the estates sector was reported at 18.48 ton/ha, a marginal drop of 0.8% or 0.15 ton/ha from 18.63 ton/ha registered in 2014. On the other hand, palm oil opening stocks were 2.02 million tons, higher by 1.4% in comparison with the previous year’s opening level at 1.99 million tons. Stocks for the first half of 2015 were lower compared to stocks in the second half of 2015, due to lower supply caused by low production from July 2015 onwards (MPOB, 2016).

For the last few years, environmental issues have become increasingly relevant in relation to economic activities and public health, both in Malaysia and globally. A specific concern is the atmospheric environmental problems, which in the past has been ignored in Malaysia but have now emerged as a significant national concern in recent years (Dislich et al., 2017). Globally, air pollution has become a major threat to the health and well-being of humans as well as plant life. Based on the severity of the pollution and also the duration of exposure, air pollution can potentially be significantly unhealthy for humans add in the case of oil palms, negatively affect growth and yield (Kusin et al., 2015).

Energy is the key to agro-processing development in Malaysia. Energy and environment are two sides of the same coin; increasing energy consumption anywhere will be accompanied by increased negative effects on the environment. It is accepted that air pollution, acid rain, and, particularly, global climate change are the inevitable consequences of greenhouse gas emissions from the burning of fossil fuels. Agriculture both produces and subsequently consumes energy. It uses huge supplies of locally available energy, namely, seed, manure and animate energy as direct and indirect energy sources (Eksioglu et al., 2015; Meijide et al., 2017). The direct energy includes diesel or gasoline fuels, human power, animal energy, fertilisers, and chemicals. Indirect energy is released directly into the source of energy that is not directly put into agricultural activity but released through a conversion process. An example of an indirect source of energy is machinery. Energy input for machinery refers to the energy requirement in producing the machine instead of the energy required to operate the machinery. Energy to operate the machinery can be classified as energy input from human labour (Michaelides, 2012).

It is a priority among designers and planners to develop energy-efficient agricultural systems that require low energy input in comparison with the output of food. This will lower the greenhouse gas emissions from agricultural production systems (Begum & Nazri, 2013; Nabavi-Pelesaraei et al., 2013a).

1.3 Genetic Algorithm and Neural Network Hybrid

In the oil palm industry, modelling and selecting variables play a significant role in the effort to understand different problems. Modelling is employed in decision making, and advances in computer technology have made available novel approaches for studying modelling. While choosing variables is for the purpose of selecting the “best” subset of
predictors or is defined as “finding a set of predictor variables which gives a good fit, predicts the dependent value well and is as small as possible.” It is used to determine the most significant factor affecting agricultural production.

Modelling and selecting variables can be grouped into statistical and heuristic methods. The statistical method is defined as “the analysis of the relationship between multiple measurements made on groups of subjects or objects, with the model usually containing systematic elements and random effects.” Mathematically, statistical modelling can be defined as “a set of probability distributions on the sample space.” Modelling includes the proper application of statistical analysis approaches with specific assumptions on testing the hypothesis, interpreting the data, and drawing a conclusion that can be applied (Kodratoff, 2014). Selection of input variables is an essential and important consideration in determining the ideal functional form of statistical models. The selection of input variables is normal in developing all statistical models and is highly dependent on the discovering relationships within the available data for the identification of appropriate predictors of the model output. Traditionally, logistic regression models are used as the variables selection methods (Sun et al., 2016 a).

The heuristic approach is defined as “pertaining to the use of general knowledge based on experimentation, evaluating possible answers or solutions, or trial-and-error methods relating to solving problems by experience rather than theory and optimisation solving by finding values of the variables that minimise or maximise the objective function while satisfying the constraints. Heuristic also refers to the problem-solving method that requires the conception of a hypothetical answer to a problem at the beginning of an inquiry to provide guidance to the inquiry. The most important types of the heuristic approaches comprise the neural network (NN) model and genetic algorithm (GA), which are based on the rules of thumb and extensively employed in different fields. A very significant feature of neural networks is their adaptive nature where earning by example: substitutes for “programming” in problem-solving. This feature makes these computational approaches very attractive as application domains, where one possesses little or inadequate comprehension of the issue to be addressed, but where training data or examples exist (Asta, 2015).

The notion of a neural network hinges on the human brain, which is made up of billions of neurons interconnected by synapses. In the same manner, NN is composed of many computational units which are also called neurons. The interconnections of the neurons dictate the characteristics of both a brain and a neural network (Da Silva et al., 2017). The feed forward back propagation ANN is a popular method employed to train neural networks. ANN has been widely applied to predict yield, energy consumption, energy demand, environmental problems and solve different types of issues (Chang et al., 2012). Currently, a nonlinear autoregressive NN with exogenous inputs NARX Time series prediction algorithms has been frequently used in several areas, e.g., predicting financial markets, weather forecasting, and complex dynamical system analysis (Khamis & Abdullah, 2014).
GAs are stochastic search approaches that provide guidance to a population of solutions towards an optimum employing the principles of evolution and natural genetics. Recently, GAs have become a popular optimisation tool for several research areas, such as system control, control design, and science and engineering. GAs draw inspiration from the evolution of populations (Mohanta & Sethi, 2011). Algorithms, which combine GAs and NN, have exhibited enhanced convergence properties compared to pure backpropagation. Such hybrid systems can locate the weights and also the architecture of NN, such as a number of layers, the processing elements per layer and the manner in which processing elements are connected. To summarize, GAs has been applied in NNs for three main functions: (i) train the weights of the connections, (ii) design the structure of the network, and (iii) locate an optimal learning rule (Chang et al., 2012; Karimi and Yousefi, 2012).

1.4 Problem Statement

Over the last decade, the Malaysian oil palm industry has been facing the challenge of the reduced rate of palm oil yield, caused by the sizeable difference between the actual production of palm oil and the crop’s genetic potential with high land usage. The gap has grown wider over time. The oil palm yield varies in various areas of Malaysia and is distributed between high yield, medium and worst, which have significantly affected the efficiency of production. In Malaysia, current planting materials are capable of producing 40 tons of FFBs /ha/yr, yielding 6–7 ton of oil. However, the reality shows average yields to be only between 50 to 60% of this potential. In addition to other challenges include, the labour shortage is the most severe constraint, and presently the industry is highly dependent on foreign workers. Furthermore, available land for expansion is limited, particularly in Peninsular Malaysia where land cost is also significantly higher.

The demand for palm oil continues to increase, merely expanding the oil palm plantations is not an advisable response. A viable response would be by way of increasing the output of existing plantations. Modelling in various aspects related to agriculture is important, given the dynamic conditions of oil palm production. Despite the strong need for accurate forecasts, the current status of these predictions is far from satisfactory. No well-defined forecasting method exists that takes into account most of the factors that drive yield. Although there are established models, they tend to be “one size fits all,” and are linear.

The challenge in modelling oil palm yield is due to the fact that it does not follow a linear model. It typically takes a nonlinear growth curve. The function of a growth curve and production have a sigmoid form. In modelling a non-linear curve, the problem becomes more complex when there are additional independent variables. The major hindrance in modelling the behavior of yield and energy consumption are the challenge of extracting the constants of the mathematical models. In light of the complexity of these relationships, traditional data-processing techniques are unable to satisfactorily investigate the process and product parameters because of non-linear relationships.
among the variables. Non-linear methods can be used to address this issue as they are powerful predictive tools. One method for modelling non-linear (accommodating multivariate) and non-parametric data is Neural Networks (NNs), which is a model-free estimation. Exploratory research on artificial intelligence has revealed that little has been done on oil palm yield and energy consumption.

The immediate and obvious effect of adding a high number of input variables is that the size of an NN increase, which raises the computational burden related to querying the network - a significant effect in the determination of the training speed. Climatic phenomena, air pollutants, and energy consumption have a direct impact on oil palm production or yield. Environmental change is the most common stressful condition oil-palm faces, so monitoring these related factors is beneficial for the prediction of oil palm yields (Corley and Tinker, 2016; Saadon et al., 2014). Thus, a lot of variables that can lead to maximum oil palm production should be identified. Few studies have focused on this differentiation with focused on a limited number of variables. As such, any of the factors mentioned cannot be randomly ignored as it could have an effect on the prediction accuracy. Selection of the most informative variables or elimination of the uninformative ones could enhance the performance of multivariate calibration models.

Thus, the manner in which variables are selected is deemed an important area of agricultural research. One of the most common methods uses traditional statistics. Although generally understood and easily computed, these methods involve the addition/removal of one variable at a time, based on section. According to Ficken (2015), establishing specific variables that exert considerable effect collectively would be difficult because linear relationships or linear correlations consider only one parameter at a time. The challenges of input variables selection emerge because of (i) the number of available data is huge; (ii) this huge data creates redundancies due to high correlations between potential input variables; and (iii) some variables have slight or no predictive capability. Genetic algorithm which is a tool for computational optimization can be combined with NNs. Specifically, in cases where a large number of potential inputs are available, they can be used to select optimal subsets of inputs for model development.

1.5 Research Objectives

The main objectives of the research are to develop of a hybrid neural networks model with GA-selected inputs for predicting oil palm yield and energy consumption within the multiple areas in Malaysia, based on large-scale climate indices, types of oil palm areas, air pollution and energy consumption. The specific objectives and of the study are as follows:

1- To develop a genetic algorithm and Stepwise as input variable selection models to identify the most significant variables that affect the FFB production, palm oil yield and output energy.
2- To develop and propose the preferred ANN and NARX architectures, and their hybrids with a genetic algorithm and Stepwise from earlier input variable selection model determinations for the predictions of FFB, palm oil yield and output energy.

3- To develop multiple linear regression models (Mathematical Model) to predict FFB, palm oil yield and output energy and compare it with the final NN system.

4- To explore the effects of the selected significant variables in the production of FFB, palm oil, and energy that will lead to the maximum the oil palm production in Malaysia.

1.6 Scope of the Study

This study focuses on the interactions between climate data including rainfall, rainy days, humidity, radiation, temperature, surface wind speed, evaporation, cloud cover and air pollution data as well as types of oil palm areas data, namely: O3, CO, NO2, SO2, PM10, the percentage of mature area and percentage of the immature area. Additionally, the interactions between input energy data including eight variables: human power, electricity, fuel, water, NPK fertilizers, and seed. This research will cover the years 2005 to 2015 to gain more accurate results on the impact of environmental sources and input energy over the Malaysian states.

The selections of data which need to be used and input for the study will be based on the availability of data from reliable sources. The effect of some important parameters in oil palm yield such as in-situ soil classification, soil moisture content, etc., was not considered in this investigation because these data are not available in a time series format and furthermore the related information may vary from plantation to plantation even though plantations are within the same district. As well as, sometimes the accessibility to the data is almost impossible because these data were classified as confidential.

The establishment of an integrated FFB, PO, and output energy models for the various areas of Malaysia that contain: (i) Determine model inputs that are extremely important via development of two methods, namely GA and Stepwise methods, to improve the intelligent prediction models. (ii) Investigate the effectiveness of two architectures in NNs namely multi-layer feedforward backpropagation (ANN) and nonlinear autoregressive exogenous neural network (NARX). A comparative study will also be performed between the results obtained from the NNs and the results obtained from the multi-regression technique in a statistical approach. A comparative study will be performed between the results obtained from the NNs and the results obtained from the multi-regression technique in a statistical approach. (iii) Select hybrid model criteria by which the performance is evaluated as they can have a significant effect on the model architecture and weight optimization techniques. (iv) Determine a hybrid model architecture including a number of hidden nodes in hidden layers, training algorithms, transfer function, and a number of delays.
Oil palm production models have tested based on the monthly data according to the data classification from the reliable source (MPOB). Finally, the results and finding of this survey will inform the key players in Malaysian oil palm production to know some information about the software's which predicts the oil palm production and output energy. Also, let the are investors in palm oil and smallholders, research centers, industrial manufacturers of oil palm products, distributors and traders of Malaysian oil palm products know that different yields and output energy due to different types of oil palm areas, environmental variables, energy consumption variables, and disparities.

1.7 Thesis Layout

The thesis systematically consists of five main parts. Brief descriptions of the content of these chapters are presented below:

Chapter 1 contains the background of the research, formulation of the problem and problem definition, research objectives, thesis layout has been highlighted and systematics writing.

Chapter 2 reviews the various literature on the topic. This chapter contains the information necessary to understand the issues discussed in this study. These reviews related to global Importance and Oil palm Production in Malaysia, general view of the environment change and types of oil palm area, the concept of energy and energy sources in Malaysia oil palm. History and background of variables selection and genetic algorithm with stepwise methods, general view of ANN and NARX to problem, prediction and oil palm models.

Chapter 3 discusses the Methodology, principal and understanding concept of GA, Stepwise, ANN, and NARX. The chapter also explains the process details of the selected variables and application design predictions.

Chapter 4 contains a discussion on the implementation of the structured analysis and design in Chapter 3. In addition, it discusses the results from the running of the models.

Finally, Chapter 5 contains the conclusions of the work which has been discussed in previous chapters, especially in Chapter 3 and Chapter 4. The final part of this chapter contains suggestions that were put forward for the continuation of further research.
REFERENCES


294


BIODATA OF STUDENT

Yousif Yakoub Hilal was born in Basrah, Iraq on August 30th, 1974. He received the Bachelor and the Master degrees in Agricultural Mechanization from the University of Basrah, Iraq. In 2001, he joined the Department of the Agricultural Mechanization, University of Basrah, Iraq as a lecture since 2007 until now, Yousif Yakoub is attached to the Department of the Agricultural Machinery and Equipment, University of Mosul, Iraq.

Currently, he is pursuing a Ph.D. degree at the Biological and Agricultural Engineering Department, Universiti Putra Malaysia under the supervision of Prof. Ir. Dr. Azmi b. Dato’ Yahya. His research interest include artificial intelligence and applied agricultural engineering. He is happily married and blessed with five children.
LIST OF PUBLICATIONS

Journal Articles


Submitted to: Journal of Agricultural Sciences and Technology (ISI Q2)


Submitted to: Field Crops Research (ISI Q1)


Submitted to: The Journal of Animal and Plant Sciences (ISI Q3).


Submitted to: Agricultural Engineering International: CIGR Journal (Q2)
STATUS CONFIRMATION FOR THESIS / PROJECT REPORT AND COPYRIGHT

ACADEMIC SESSION: ________________

TITLE OF THESIS / PROJECT REPORT:
PREDICTING CROP YIELD AND FIELD ENERGY OUTPUT FOR OIL PALM USING GENETIC ALGORITHM AND NEURAL NETWORK MODELS

NAME OF STUDENT: YOUSIF YAKOUB HILAL

I acknowledge that the copyright and other intellectual property in the thesis/project report belonged to Universiti Putra Malaysia and I agree to allow this thesis/project report to be placed at the library under the following terms:

1. This thesis/project report is the property of Universiti Putra Malaysia.

2. The library of Universiti Putra Malaysia has the right to make copies for educational purposes only.

3. The library of Universiti Putra Malaysia is allowed to make copies of this thesis for academic exchange.

I declare that this thesis is classified as:

*Please tick (V)

☐ CONFIDENTIAL (Contain confidential information under Official Secret Act 1972).

☐ RESTRICTED (Contains restricted information as specified by the organization/institution where research was done).

☐ OPEN ACCESS I agree that my thesis/project report to be published as hard copy or online open access.

This thesis is submitted for:

☐ PATENT Embargo from ______ until ______

(date) (date)

Approved by:

(Signature of Student) (Signature of Chairman of Supervisory Committee)
New IC No/ Passport No.: Name:

Date: Date:

[Note: If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization/institution with period and reasons for confidentially or restricted.]