UNIVERSITI PUTRA MALAYSIA

STATISTICAL MODELING OF RIVER WATER QUALITY INDEX

ZALINA BINTI MOHD ALI

IPM 2019 23

STATISTICAL MODELING OF RIVER WATER QUALITY INDEX

By

ZALINA BINTI MOHD ALI

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

May 2019

DEDICATIONS

To my dearest families :
my husband, Khalid
my children; Hilmi, Hazim, Humam and Hakiim
my mum and dad, Kamariah and Mohd Ali
my grandma, Minah
my brother, Isham
my spouse's parent and sisters.

I dedicate my dissertation work and everything I do to all of you. Thank you for always been there during my hard and demanding time. Your kindness, encouragement, and effort will always be appreciated

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

STATISTICAL MODELING OF RIVER WATER QUALITY INDEX

By

ZALINA BINTI MOHD ALI

May 2019

Chairman : **Noor Akma bt Ibrahim, PhD**
Faculty : **Institute For Mathematical Research**

Water quality index (WQI) is a unitless number that indicates overall water quality at a specific time and location using several important water quality variables. In Malaysia, general WQI which is based on water quality expert opinion has been introduced by the Department of Environment (DOE) to describe the status of a specific location at identified rivers. The accuracy of DOE-WQI measured by the experts is assessed using four important phases namely variables selection, weights determination, variables transformation and variables aggregation, i.e. WQI calculation. However, the experts opinion approach is found to be the most subjective in nature.

On determining weights in WQI development, new statistical methodologies were introduced in this study to enhance the current Malaysian DOE-WQI. The Langat River in Selangor was chosen as the location is significantly altered due to several environmental issues. In this study, several descriptive models as well as Bayesian models have been introduced based on a data-driven approach. To enable comparison between models, water quality variables in the Malaysian DOE-WQI calculation was employed.

For the descriptive models, 17 Principal Component Analysis (PCA) models had been applied to five selected monitoring stations and four PCA models were found to have similar patterns with the existing DOE-WQI. The models are the PC Standardization, i.e. based on Minimum-Maximum approach known as D1 and D5 as well as the Model E1 and E2 which are based on the re-weighting of the eigenvector elements from two different approaches. The findings also showed that using the relative importance based on relative rank of the first PCA eigenvector elements provided an alternative way to calculate the PCA-WQI, as described in Model E2. Similar approach was conducted

using standard deviation of the data for all stations as described in Model E3 and each station separately in Model E4. The results showed that the new weights based on the relative ranks of the standard deviation for Model E3 and Model E4 had contributed well in the new WQI calculations. Both new models are simpler, consistent, stable, comparable and reliable.

Furthermore, new Simultaneous-WQI (S-WQI) model in Bayesian approach for each station was introduced to improve WQI estimates. Several potential Bayesian models were considered and the best Bayesian model was selected based on two certain criteria, i.e. Deviance Information Criteria (DIC) values and monitoring convergence. S-WQI can be estimated accurately using a general form of Bayesian model with further constraint in the variance of sub-index pH, SIpH, i.e. the natural water quality characteristic for each station. New parameters from the best Bayesian model were used to re-calculate weights in WQI calculation. Based on the new WQI calculation using the Bayesian knowledge, narrower ranges for each individual observation at all stations were found, indicating better estimates of DOE-WQI. Several interesting further research were also discussed in order to provide WQI researchers better understanding on the benefits and limitations of the different indices.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PEMODELAN STATISTIK INDEKS KUALITI AIR SUNGAI

Oleh

ZALINA BINTI MOHD ALI

Mei 2019

Pengerusi : **Noor Akma bt. Ibrahim, PhD**
Fakulti : **Institut Penyelidikan Matematik**

Indeks Kualiti Air (IKA) merupakan satu angka tanpa unit yang menentukan kualiti air secara keseluruhan pada masa dan tempat tertentu menggunakan beberapa pemboleh ubah penting kualiti air. Di Malaysia, IKA umum yang berdasarkan kepada pandangan pakar kualiti air diperkenalkan oleh Jabatan Alam Sekitar (JAS) untuk menerangkan status bagi lokasi tertentu di sungai-sungai yang telah dikenal pasti. Ketepatan IKA-JAS yang diukur daripada pakar dinilai menggunakan empat fasa utama iaitu pemilihan pemboleh ubah, penentuan pemberat, penjelmaan pemboleh ubah dan pengagregatan pemboleh ubah. Walau bagaimanapun, pendekatan pandangan pakar telah didapati sebagai subjektif secara semulajadi.

Dengan memberi tumpuan ke atas penentuan pemberat-pemberat dalam pembangunan IKA, kaedah-kaedah statistik baru diperkenalkan dalam kajian ini untuk meningkatkan penggunaan IKA-JAS Malaysia sedia ada. Tumpuan diberikan ke atas Sungai Langat di Selangor kerana lokasi ini mengalami perubahan yang ketara disebabkan oleh pelbagai isu persekitaran. Dalam kajian ini, model deskriptif dan model Bayesian baru diperkenalkan berdasarkan kepada pendekatan berpandukan-data. Untuk membolehkan perbandingan di antara model dilakukan, pemboleh ubah kualiti air dalam pengiraan IKA-JAS telah digunakan.

Bagi model-model deskriptif, 17 model Analisis Komponen Utama (AKU) digunakan terhadap lima stesen pemantauan terpilih dan didapati, empat model AKU tersebut mempunyai bentuk yang sama seperti IKA-JAS sedia ada. Model-model tersebut ialah Pempiawaian KU, i.e. berdasarkan Pendekatan Minimum-Maksimum dikenali sebagai D1 dan D5 serta Model E1 dan Model E2 yang berdasarkan kepada pengubahsuaian semula pemberat bagi eigen vektor daripada dua pendekatan yang berbeza. Hasil

penemuan turut menunjukkan potensi penggunaan kepentingan relatif berdasarkan kepada pangkat relatif bagi eigen vektor pertama AKU sebagai salah satu cara alternatif dalam pengiraan IKA-AKU seperti yang diterangkan oleh Model E2. Pendekatan yang sama turut dilakukan menggunakan sisihan piawai data bagi semua stesen seperti yang diterangkan dalam Model E3 dan setiap stesen berasingan dalam Model E4. Keputusan menunjukkan pemberat baru yang berdasarkan kepada pangkat relatif sisihan piawai bagi Model E3 dan E4 memberi sumbangan yang baik dalam pengiraan IKA baru. Kedua-dua model tersebut adalah ringkas, konsisten, stabil, boleh dibandingkan dan dipercayai.

Seterusnya, model IKA-Serentak (IKA-S) baru dalam pendekatan Bayesian bagi setiap stesen diperkenalkan untuk menambah baik anggaran-anggaran IKA. Beberapa model Bayesian yang berpotensi telah dipertimbangkan dan model Bayesian terbaik telah dipilih berdasarkan dua kriteria tertentu iaitu nilai-nilai Kriteria Informasi Devians (KID) dan penumpuan pemantauan. IKA-S boleh dianggarkan dengan baik menggunakan model Bayesian umum dengan lanjutan kekangan dalam varians SIpH iaitu kriteria kualiti air secara semulajadi bagi setiap stesen. Berdasarkan kepada pengiraan IKA baru menggunakan pengetahuan Bayesian, julat yang kecil didapati bagi setiap cerapan individu pada semua stesen yang menunjukkan anggaran yang lebih baik bagi IKA-JAS. Beberapa penyelidikan berterusan yang menarik turut dibincangkan untuk mendapatkan kefahaman yang lebih baik dalam kalangan penyelidik-penyelidik IKA terhadap kelebihan dan kekurangan bagi setiap indeks yang digunakan.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Prof. Dr. Noor Akma Ibrahim and Prof. Dr. Kerrie Mengersen. My sincere gratitude goes to them for their continuous support along my Ph.D journey. Their great knowledge, support, patience and motivation helped me a lot of doing this interesting research as well as giving momentum in writing this tesis successfully.

Besides my knowledgeable supervisors, I would like to thank the rest of my PhD and thesis committee: Associate Prof. Dr. Mahendran Shitan and Associcate Prof. Dr. Hafizan Juahir, for their knowledgeable opinion and inspiration. Their supervision with all the difficult questions had motivated and inspired me to expand this research from different perspectives.

My sincere thanks also goes to Department of Environment, Department of Statistics Malaysia and National Hydraulic Research Institute of Malaysia for providing the data and information which related and definitely benefited to my PhD works.

I also wish to express my gratitude to Universiti Kebangsaan Malaysia and the Ministry of Higher Education as well as Universiti Putra Malaysia for providing fully or partially financial sponsorship during my research. I truly believe that without this funding aid, this research will not succeed.

My thanks also go to the following university staff : Associate Prof Dr. Nur Riza Mohd Suradi, Dr. Zamira Hasanah Zamzuri and Prof. Dr. Kamarulzaman Ibrahim for their interesting sharing and support. In addition, my thanks definitely must go to my fellow friends, support staffs, librarians, my teachers, my lecturers and everyone that had really influenced my life along this journey.

I also really appreciate the thoughtful comments and constructive suggestions from the thesis examination committee and I would like thank all of them for careful and thorough reading of this thesis which really help me to improve the quality of this thesis.

Last but not least, I would like to thank the very close people in my life; my family, my parents and my relatives for supporting me physically, emotionally and spiritually throughout this research and writing this thesis.

Thanks for all your help and support!

This thesis was submitted to the Senate of 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:

Noor Akma bt Ibrahim, PhD
Professor
Institute for Mathematical Sciences
Universiti Putra Malaysia
(Chairperson)

Kerrie Mengersen, PhD
Professor
Centre for Dynamic Systems
Queensland University of Technology
(Member)

Mahendran a/l Shitan, PhD
Assosiate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Hafizan bin Juahir, PhD
Assosiate Professor
School 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 other 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: Zalina Mohd Ali, GS28904

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) are adhered to.

Signature: ______________________
Name of Chairman of Supervisory Committee: **Noor Akma bt Ibrahim**

Signature: ______________________
Name of Member of Supervisory Committee: **Kerrie Mengersen**

Signature: ______________________
Name of Member of Supervisory Committee: **Mahendran a/l Shitan**

Signature: ______________________
Name of Member of Supervisory Committee: **Hafizan bin Juahir**

TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	iii
ACKNOWLEDGEMENTS	v
APPROVAL	vi
LIST OF TABLES	xiii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xix

CHAPTER

1 **INTRODUCTION** 1
- 1.1 Statistical Methods and River Water Quality Index 1
- 1.2 Motivation of the Study 2
- 1.3 Research Aims and Objectives 3
- 1.4 Study Method 4
- 1.5 Scope of the Study 5
- 1.6 Contribution of the Study 5
- 1.7 Outline of Thesis 6

2 **A REVIEW OF STATISTICAL RIVER WATER QUALITY INDEX MODELS** 9
- 2.1 Introduction 9
- 2.2 The Importance of River Water Quality Index for Accurate Water Quality Assessment 9
- 2.3 Weights Determination in River WQI Model and The Need for Modeling Tools 11
- 2.4 River WQI Model Classification 12
  - 2.4.1 Deterministic Models 12
  - 2.4.2 Statistical Models 14
- 2.5 Parametric Statistical Models 15
  - 2.5.1 Multivariate Models 15
  - 2.5.2 Regression Models 17
  - 2.5.3 Time Series Models 18
  - 2.5.4 Stochastic or Probability Models 18
  - 2.5.5 Distributional Models 19
- 2.6 Non-Parametric Statistical Models 20
  - 2.6.1 Artificial Neural Network Models 20
  - 2.6.2 Relative Ranking Models 21
- 2.7 Bayesian Statistical Models 22
- 2.8 Discussion 23
  - 2.8.1 Statistical Models Comparison 24

2.8.2 Bootstrap Uncertainties Analysis 25
2.8.3 Relative Ranking Weights Determination 26
2.8.4 Development of Within-Station WQI 26
2.8.5 Bayesian Models Selection 28
2.9 Conclusion 28

3 **PRELIMINARY DATA ANALYSIS OF RIVER WATER QUALITY** 30
3.1 Introduction 30
3.2 Water Quality Index Calculation and Interpretation 30
3.3 Case Study 31
3.3.1 Study Area 32
3.3.2 Data and Monitoring Stations 34
3.4 Analysis of Water Quality Data 38
3.4.1 Summary of Statistical Distribution Measures 38
3.4.2 Visual Interpretation of Water Quality Data Distribution 39
3.4.3 Normality Test 39
3.4.4 Status of River Water Quality Index 50
3.5 Conclusion 50

4 **DESCRIPTIVE MODEL-BASED APPROACH IN WQI DEVELOPMENT** 52
4.1 Introduction 52
4.2 WQI Overall-Station Model 52
4.2.1 PCA Models of River WQI 52
4.2.2 Five Basic PCA Models of River WQI 56
4.2.3 PCA Models of River WQI based on Rescaling PCA Scores 58
4.2.4 PCA Models of River WQI based on Renormalized Weights of Eigenvector Elements 59
4.2.5 New Relative Ranking Model of River WQI 60
4.2.6 Comparative Performance of WQI Models 62
4.3 New WQI Within-Station Model 83
4.3.1 New Weights Calculation for Within-Station WQI 83
4.3.2 New Weights Calculation for Overall-Station WQI 86
4.3.3 Comparative Performance of New WQI Models at Different Stations 86
4.4 Discussion 96

5 **BAYESIAN MODEL-BASED APPROACH IN SIMULTANEOUS-WQI DEVELOPMENT** 100
5.1 Introduction 100
5.2 Water Quality Model 100
5.3 Development of Simultaneous-WQI (S-WQI) Model 101
5.3.1 Model 1 : General S-WQI Model 101
5.3.2 Model 2 : Distributed-lag Model in S-WQI for Each Water Quality Variable 102
5.3.3 Model 3 : Distributed-lag Model for S-WQI 103
5.4 Bayesian Models for Simultaneous-WQI (S-WQI) 103
5.4.1 Case I : S-WQI Model with Station Specific Variance 103

5.4.2 Case 2 : WQI Model with Common Variance over Time 106
5.4.3 Case 3 : WQI Model with Further Constraint 107
5.5 Diagnostic Checking 109
5.5.1 Monitoring Convergence 110
5.5.2 DIC Value 123
5.6 Sensitivity Analysis of the Best Model 125
5.7 Discussion 125

6 **NEW WQI DEVELOPMENT USING THE BEST BAYESIAN SIMULTANEOUS-WQI MODEL-BASED DECISION** 127
6.1 Introduction 127
6.2 Relationship between Bayesian Simultaneous-WQI (S-WQI) and Natural Water Quality Variable, pH 127
6.3 Parameter Estimation for Model Coefficient 130
6.4 New WQI Development 131
6.4.1 New Relative Weights 131
6.4.2 New Weights for WQI Calculation 132
6.5 Comparative Performance of WQI 134
6.5.1 Temporal Assesment 134
6.5.2 Comparability with Dissolved Oxygen 138
6.5.3 Comparability of Distribution 138
6.5.4 Reliability Analysis 140
6.6 Discussion 140

7 **CONCLUSION AND FUTURE WORK** 144
7.1 Conclusion 144
7.2 Future Work 146

BIBLIOGRAPHY 147
APPENDICES 160
BIODATA OF STUDENT 232
LIST OF PUBLICATIONS 233
INDEX 234

LIST OF TABLES

Table		Page
2.1	Significance Ratings and Weights in NSF-WQI	14
3.1	The Relative Importance by Malaysian DOE	30
3.2	Classification of Water Quality by Malaysian DOE	32
3.3	DOE Selected Monitoring Stations at Langat River	34
3.4	Median, Mean, SD and CV for all Variables at each Station from 1995-2007.	40
3.5	Shapiro-Wilk Test for Selected Variables based on Different Transformation Categories	50
4.1	Pearson Correlation Matrix for Sub-Indices	54
4.2	Spearman Correlation Matrix for Sub-Indices	54
4.3	Results of the Eigenvector Elements and Eigenvalues of the Sub-Indices Water Quality	55
4.4	Description of the Five Basic PCA Models	56
4.5	Description of PCA Models based on Rescaling PCA Scores	58
4.6	Description of PCA Models based on Renormalized Weights of Eigenvector Elements	59
4.7	New Weights based on the Eigenvector Elements (Model E1)	60
4.8	New Relative Weights based on Rankings of the First Eigenvector Elements (Model E2)	61
4.9	New Relative Weights based on the Ranks of the Overall Standard Deviation (Model E3)	62
4.10	Average of Absolute Difference in Ranks for Selected Models	71
4.11	Pearson Correlation Matrix for all Models using Training Data	72
4.12	Spearman Correlation Matrix for all Models using Training Data	72
4.13	WQI Mean and Standard Deviation (in parentheses) for Original and Bootstrap Data	76

4.14 New Relative Weights for Station 1L01 84
4.15 New Relative Weights for Station 1L02 84
4.16 New Relative Weights for Station 1L03 84
4.17 New Relative Weights for Station 1L05 85
4.18 New Relative Weights for Station 1L07 85
4.19 New Weights Calculation for Overall-WQI 86

5.1 DIC Values For Selected Cases 124
5.2 DIC Values For Model 1 Case 3 at Different Precision Values 125
5.3 DIC Values For Selected Cases 126

6.1 Selected Statistical Measures' Values for Parameter in Models for Data with No Missing Values 130
6.2 Selected Statistical Measures' Values for Parameter in Models for Data with Missing Values 131
6.3 Temporary Weights of WQI at Selected Sampling Sites using Method 1 133
6.4 Temporary Weights of WQI at Selected Sampling Sites using Method 2 133
6.5 New Weights of WQI at Selected Sampling Sites 133
6.6 Summary of Pearson and Spearman Correlation Coefficients between Indices and DO 138

LIST OF FIGURES

Figure		Page
1.1	Study Flow of the Thesis	8
2.1	Basic Procedures in WQI Development	11
3.1	Population Number by Districts in Langat River Basin	33
3.2	Locations of the Selected Monitoring Stations	35
3.3	WQI Values versus Sampling Time at Station 1L01	36
3.4	WQI Values versus Sampling Time at Station 1L02	36
3.5	WQI Values versus Sampling Time at Station 1L03	37
3.6	WQI Values versus Sampling Time at Station 1L05	37
3.7	WQI Values versus Sampling Time at Station 1L07	38
3.8	Distributions of Water Quality Variables	41
3.9	Distributions of Log Water Quality Variables	41
3.10	Distributions of Inverse Water Quality Variables	42
3.11	Distributions of Square Root of Water Quality Variables	42
3.12	Distributions of SIDO at Selected Stations from 1995-2007	43
3.13	Distributions of SIBOD at Selected Stations from 1995-2007	43
3.14	Distributions of SICOD at Selected Stations from 1995-2007	44
3.15	Distributions of SIAN at Selected Stations from 1995-2007	44
3.16	Distributions of SISS at Selected Stations from 1995-2007	45
3.17	Distributions of SIpH at Selected Stations from 1995-2007	45
3.18	Q-Q Plot of (a) SIDO (b) SIBOD (c) SICOD (d) SIAN (e) SISS and (f) SIpH from 1995-2007	46
3.19	Q-Q Plot of Log (a) SIAN and SIBOD (b) SICOD and SIDO (c) SISS and SIpH from 1995-2007	47

3.20 Q-Q Plot of Inverse (a) SIAN and SIBOD (b) SICOD and SIDO (c) SISS and SIpH from 1995-2007 48

3.21 Q-Q Plot of Square Root (a) SIAN and SIBOD (b) SICOD and SIDO (c) SISS and SIpH from 1995-2007 49

3.22 Average Value of WQI by Year for DOE-WQI (Model A) 51

4.1 Average Values of Indices by Year for Model A and Model (a) B1 (b) B2 (c) B3 (d) B4 and (e) B5 64

4.2 Average Values of Indices by Year for Model A and Model (a) C1 (b) C2 (c) C3 (d) C4 and (e) C5 65

4.3 Average Values of Indices by Year for Model A and Model (a) D1 (b) D2 (c) D3 (d) D4 and (e) D5 66

4.4 Average Values of Indices by Year for Model A and Model (a) E1 (b) E2 (c) E3 67

4.5 Monthly Average Indices for Selected Models 68

4.6 Monthly Average Indices for Selected Models versus A 69

4.7 Ratios for Selected Models versus A 70

4.8 Distribution of Selected Models for Training Data, 1997-2007 73

4.9 Distribution of Selected Models for Test Data, 2008-2009 74

4.10 Mean Values of WQI Based on Bootstrap Sampling of Model A 77

4.11 Mean Values of WQI Based on Bootstrap Sampling of Model D1 77

4.12 Mean Values of WQI Based on Bootstrap Sampling of Model D5 78

4.13 Mean Values of WQI Based on Bootstrap Sampling of Model E1 78

4.14 Mean Values of WQI Based on Bootstrap Sampling of Model E2 79

4.15 Mean Values of WQI Based on Bootstrap Sampling of Model E3 79

4.16 Overall Performance Model A by Year 80

4.17 Overall Performance Model D1 by Year 80

4.18 Overall Performance Model D5 by Year 81

4.19 Overall Performance Model E1 by Year 81

4.20 Overall Performance of Model E2 by Year 82

4.21 Overall Performance of Model E3 by Year 82
4.22 Box Plot for the Bootstrap Means of WQI by Type of Models 83
4.23 Box Plot for WQI Grouped by Type of Model Used 87
4.24 Temporal Trend of DOE-WQI and Model E4 at Station 1L01 88
4.25 Temporal Trend of DOE-WQI and Model E4 at Station 1L02 88
4.26 Temporal Trend of DOE-WQI and Model E4 at Station 1L03 89
4.27 Temporal Trend of DOE-WQI and Model E4 at Station 1L05 89
4.28 Temporal Trend of DOE-WQI and Model E4 at Station 1L07 90
4.29 Correlation and Dispersion of DO versus Selected Models at Station 1L01 91
4.30 Correlation and Dispersion of DO versus Selected Models at Station 1L02 91
4.31 Correlation and Dispersion of DO versus Selected Models at Station 1L03 92
4.32 Correlation and Dispersion of DO versus Selected Models at Station 1L05 92
4.33 Correlation and Dispersion of DO versus Selected Models at Station 1L07 93
4.34 Distribution of Selected Model for Training Data 94
4.35 Distribution of Selected Model for Test Data 95
4.36 Bootstrap Confidence Interval of Model E4 by Year for (a) Station 1L01 (b) Station 1L02 (c) Station 1L03 (d) Station 1L05 (e) Station 1L07 98
4.37 Bootstrap Confidence Interval of Model E4 by Sampling Time for (a) Station 1L01 (b) Station 1L02 (c) Station 1L03 (d) Station 1L05 (e) Station 1L07 99

5.1 Trace Plot and Kernel Density Plot for Model 1 (Case 1) at Station 1L03 111
5.2 Trace Plot and Kernel Density Plot of 400,000 Iterations with Thinning each 100th for Model 2 (Case 2) at Station 1L01 112
5.3 Trace Plot and Kernel Density Plot of 400,000 Iterations with Thinning each 100th for Model 2 (Case 3) at Station 1L01 113
5.4 Trace Plot and Kernel Density Plot of 400,000 Iterations with Thinning each 100th for Model 2 (Case 3) at Station 1L03 114
5.5 Autocorrelation Plot for Model 2 (Case 3) of 400,000 Iterations with Thinning each 100th at Station 1L01 115

5.6 Autocorrelation Plot for Model 2 (Case 3) of 400,000 Iterations with Thinning each 100th at Station 1L03 116

5.7 Trace Plot and Kernel Density Plot of 400,000 Iterations with Thinning each 100th for Model 1 (Case 3) at Station 1L01 117

5.8 Traceplot and Kernel Density Plot of 400,000 iterations with Thinning each 100th for Model 1 (Case 3) at Station 1L03 118

5.9 Autocorrelation Plot for Model 1 (Case 3) of 400,000 Iterations with Thinning each 100th at Station 1L01 119

5.10 Autocorrelation Plot for Model 1 (Case 3) of 400,000 Iterations with Thinning each 100th at Station 1L03 120

5.11 Model 1 (Case 3) Trace Plot and Kernel Density Plot for Station 1L01 121

5.12 Model 1 (Case 3) Autocorrelation Plot of 400,000 Iterations with Thinning 100th at Station 1L01 122

6.1 Mean of Posterior Mean for S-WQI (a) versus pH (b) versus SIpH at Station 1L01 128

6.2 Mean of Posterior Mean for S-WQI (a) versus pH (b) versus SIpH at Station 1L02 128

6.3 Mean of Posterior Mean for S-WQI (a) versus pH (b) versus SIpH at Station 1L03 128

6.4 Mean of Posterior Mean for S-WQI (a) versus pH (b) versus SIpH at Station 1L05 129

6.5 Mean of Posterior Mean for S-WQI (a) versus pH (b) versus SIpH at Station 1L07 129

6.6 Temporal Trend of WQI for Model A and Model BWQI at Station 1L01 135

6.7 Temporal Trend of WQI for Model A and Model BWQI at Station 1L02 135

6.8 Temporal Trend of WQI for Model A and Model BWQI at Station 1L03 136

6.9 Temporal Trend of WQI for Model A and Model BWQI at Station 1L05 136

6.10 Temporal Trend of WQI for Model A and Model BWQI at Station 1L07 137

6.11 Temporal Trend of WQI for Model A, Model Within-Station BWQI and Model Overall-Station BWQI at Langat River 137

6.12 Histogram of Indices for (a) Training Data and (b) Test Data 139

6.13 Confidence Interval of BWQI at Station 1L01 141

6.14 Confidence Interval of BWQI at Station 1L02 141

6.15 Confidence Interval of BWQI at Station 1L03 142

6.16 Confidence Interval of BWQI at Station 1L05 142

6.17 Confidence Interval of BWQI at Station 1L07 143

LIST OF ABBREVIATIONS

AN	Ammoniacal Nitrogen
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BOD	Biochemical Oxygen Demand
BWQI	Bayesian WQI
CI	Confidence Interval
COD	Chemical Oxygen Demand
CPI	Composite Pollution Index
CV	Coefficient of Variation
DO	Dissolved Oxygen
DOE	Department of Environment
DOE-WQI	Department of Environment-Water Quality Index
FA	Factor Analysis
HDI	Human Development Index
IBI	Index of Biotic Integrity
JAGS	Just Another Gibbs Sampling
MCDM	Multi-Criteria Decision Making
MCMC	Markov Chain Monte Carlo
MM	Multivariate Methods
NSF	National Sanitation Foundation
NSF-WQI	National Sanitation Foundation-Water Quality Index
PC	Principal Component
PCA	Principal Component Analysis
pH	potential of Hydrogen
SAS	Statistical Analysis Software
SD	Standard Deviation
SI	Sub-Index
SIAN	Sub-Index of Ammoniacal Nitrogen
SIBOD	Sub-Index of Biochemical Oxygen Demand
SICOD	Sub-Index of Chemical Oxygen Demand
SIDO	Sub-Index of Dissolved Oxygen
SIpH	Sub-Index of potential of Hydrogen
SISS	Sub-Index of Total Suspended Solid
SS	Suspended Solid
S-W	Stock-Watson Index
S-WQI	Simultaneous-Water Quality Index
WQI	Water Quality Index

CHAPTER 1

INTRODUCTION

This chapter provides the background for this thesis. It introduces the particular statistical methods for water quality index development that will be discussed later in the thesis. It also provides an outline of the research aims and objectives as well as the focus of subsequent chapters.

1.1 Statistical Methods and River Water Quality Index

Studies of river water quality often involve investigation on water quality variables such as water quality sub-indices and water quality index (WQI). These variables vary with respect to different time and locations due to natural and environmental factors. The WQI is useful in providing an overall measure of river water quality that can help regulatory bodies determine the status of the water. Various techniques for the development of river WQI have been introduced and gained great interest among researchers in various parts of the world especially if the particular WQI needs to be developed at the same location. However, the use of different techniques for WQI development may lead to different values and interpretations. This scenario has improved research interest in WQI study and the beneficial results are expected to enhance the importance of water quality (Abbasi and Abbasi 2012).

In this study, river WQI is chosen since the river forms a large majority of water surface studies (Alves et al., 2014), has the most concern of the communities involved (Kotti et al., 2005) as well as being the most common practice used by the Department of Environment (DOE) to determine the quality of water in many countries as mentioned in Brown et al. (1970), Cude (2001), Lohani and Mustapha (1982a), CCME (2001) and Thi Minh Hanh et al. (2010). The best known river WQI is the National Sanitation Foundation (NSF)-WQI which was based on the expert opinion approach (Brown et al., 1970). This approach has been employed in many countries in the world including Malaysia (DOE, 1997). In the NSF-WQI development, different weights were rated by various panels of experts for the same water quality variables. However, according to Harkins (1974), this approach was found to be subjective in nature and had destroyed the objectivity and comparability of the index developed.

To reduce the subjectivity, application of statistical methods for determining WQI has been increasing, as described by many researchers, since the water quality variables that are important can vary with respect to time and locations. Hence, statistical methods in the development of WQI is important to provide tools for better estimation of water quality. The statistical methods also allow calculation, alteration and understanding of the river water quality. Although many works have been done in WQI are able to explain the quality of water and status of the river, there are still many room for improvement on WQI development particularly with regards to the weights associated with the important variable on water quality. In normal practice, the different weights rated for the same water quality variables are subjective in nature and may extinguish the objectivity and comparability of the WQI developed. A possible solution to the subjective weights derived in the development of WQI lies in formulating and implementing accurate variable weights in river WQI models. However, no detailed study has been found among the plethora literature to determine the stability, comparability, accuracy, reliability, and flexibility of the river WQI comprehensively. To achieve this, detailed analyses should be done in the index comparison as suggested by Lumb et al. (2011a). Hence, performance of various statistical models of river WQI are essential to provide better explanation in the estimation and prediction of WQI. This comprehensive model comparison need efficient data-driven approach with appropriate statistical methods. All the consideration stated provide the basis and aim of this thesis.

1.2 Motivation of the Study

In WQI development, weights determination for the selected water quality variables were assigned with regards to the differences in variable importance (Dunnette, 1979). Some of the variables would be of greater importance than others (Abbasi and Abbasi, 2012). According to Sutadian et al. (2016), the weights of water quality variables can be assigned with equal weights or unequal weights. However, establishing the variables' weights is still a matter of judgment, therefore it is subjective in nature (Abbasi and Abbasi, 2012). In Brown et al. (1970), the importance of water quality variables were assigned using a value of 1 (most important) to 5 (less important). The weights obtained for the respective water quality variable in Brown et al. (1970) were summed up to the final WQI as shown in Equation 1.1.

$$WQI = \sum_{i=1}^{p} w_i Y_i \tag{1.1}$$

where WQI is the water quality index, w_i is the weight or the relative importance of the ith water quality variable, the $\sum_{i=1}^{p} w_i = 1$, p is the number of water quality variables and Y_i refers to the respective sub-indices of water quality variables, i.e. the quality of the ith water quality variables, a number between 0 and 100.

In previous WQI development based on statistical methods, the most commonly used is principal component analysis (PCA) as supported in Whittaker et al. (2012). The relative weights from the PCA are determined by the elements of eigenvector (Chow-Fraser, 2006) or the normalized of the elements of the eigenvector (Coletti et al., 2010). Several PCA models have been introduced in the literature to calculate the WQI using the first principal component (Coletti et al., 2010) and weighted average of principal component (Chow-Fraser, 2006). However, the use of PCA in the WQI development was limited to the water quality assessment (Lohani and Todino, 1984) and classification (Thi Minh Hanh et al., 2010) at particular stations and rivers. The PCA models of river WQI also provides several way of interpretations based on the WQI values obtained at different stations and rivers. For instance, the PCA used in Lohani and Mustapha (1982a) represent good river water quality for the smallest index values but in Chow-Fraser (2006), the largest index values represents good river water quality. Different index values obtained are caused by the various transformation method of raw water quality data in the PCA models that were considered in river WQI determination. Hence, it is important to explore on the behavior of different PCA models in influencing the WQI values and this may assist the better decision making in river WQI determination (Alves et al., 2014). Despite the difficulties of various PCA model to determine accurate results for river WQI, data-driven methods in PCA models of river WQI are still well-received among researchers due to the ability to simultaneously analyze the water quality data (Wu and Kuo, 2012).

Based on PCA models, large amount of datasets are needed to allow variables or weights determination in WQI development. However, the requirement of sufficient data will not be a problem if simple statistical model such as relative ranking model is considered. The relative ranking model is capable of handling small dataset and determining the relative weights for selected variables without difficulty. Hence, the use of relative weights based on selected statistics derived from the relative ranking model as proposed in this study can provide an alternative to the WQI estimation and prediction. On the other hand, knowledge-based methods such as Bayesian modeling is also considerably important (Goethals, 2005) especially when missing values in dataset is needed in the development of WQI (Goethals and Niels, 2010). The use of Bayesian approaches give benefit and strong evidence of better parameter estimations especially if over-parameterized models are assumed. Benefits of using Bayesian model and uncertainty estimation have been discussed by Reichert and Omlin (1997). Therefore, the new relative weights proposed in the Bayesian model-based approach in this study could give better information on the uncertainty of the WQI estimates. In addition, a good WQI estimate could be determined based on statistical model comparison of stability, comparability, accuracy, reliability and flexibility. Hence, the construction of various WQI models using statistical methods in this study is as crucial as discussing WQI model fitting.

1.3 Research Aims and Objectives

The current variable weights practices in most of the countries are based on the expert opinion approach. Hence, the particular focus of this study is to develop these variable weights in enhancing the current expert opinion of WQI. The overall aim of this research is to propose statistical methodology in weights determination for river WQI development. A range of statistical models are considered and compared with the current WQI used by Department of Environment (DOE).

In order to address the aim of the thesis, the following objectives are established:

1. To develop WQI overall-station models based on PCA and to compare with the DOE-WQI.

2. To develop WQI overall-station and within each station models based on relative ranking weights and to compare with the DOE-WQI.

3. To construct and compare several Bayesian WQI within-station models.

4. To develop a within-station WQI based on the best Bayesian model and to compare with the DOE-WQI.

1.4 Study Method

The main focus in this study is to determine new weights of available water quality variables defined by the experts using the appropriate statistical methods. The new variable weights are based on descriptive and Bayesian model-based approaches . These weights are then used to develop the new WQI calculation which were compared with the existing DOE-WQI. The model selected describes the weights of water quality variables best. With respect to the four objectives stated in Section 1.3, this study is carried out in four phases. The phases are:

1. Examination of the internal structure of water quality data for across sampling sites in order to determine weights using PCA. The weights were used to calculate the PCA scores. In addition, several models based on PCA scores in this descriptive model-based approach were considered.

2. Next, new weights based on relative importance of their respective weight ranks which were obtained from the eigenvector elements in PCA approach were introduced for the first time in this study. Then, the weight ranks in PCA approach were simplified by using relative ranks for a selected descriptive statistic, i.e. standard deviation of the water quality data. The ranks of the standard deviation were used to recalculate new weights. Based on the new weights, new WQI model was computed for overall-station and within each station WQI using Equation 1.1.

3. The models performance were also discussed based on the models that preserve the same temporal with the DOE-WQI. The best model was selected based on stability, comparability, accuracy, reliability and flexibility.

4. Based on the reliability results obtained in Step 3, this study continued to identify several models that fit water quality data within-station under construction of several Bayesian models. Three basic Bayesian models were considered i.e. general Simultaneous-WQI (S-WQI) model without lag, distributed-lag model in S-WQI for each water quality variable and distributed-lag model for S-WQI. Each model was tested in different situations, i.e. S-WQI model with station specific variance, S-WQI model with common variance over time and S-WQI model with further constraint. The most appropriate Bayesian model was selected based on Bayesian diagnostic checking and sensitivity analysis.

5. Coefficient values of the best Bayesian model in Step 4 were used to determine new relative weights of water quality variables. Based on the relative weights, new WQI values were determined using Equation 1.1. The index values were compared with the DOE-WQI for their comparability, reliability and flexibility. The results obtained from the Bayesian model provided an alternative for river WQI estimation especially on the intervals of confidence from the Bayesian perspective.

1.5 Scope of the Study

A case study was presented in this study to focus and emphasize weights determination from statistical models. The weights were determined by statistical descriptive model-based approach and Bayesian model-based approach. In the descriptive model, several PCA and new relative ranking weights models were introduced. Furthermore, several Bayesian models were proposed for the first time in this study and the best model was used to determine weights that may define the importance of water quality variables within-station. The development and application of the statistical models were illustrated from several sampling sites that have been identified. Six water quality variables considered in the DOE-WQI determination : Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Suspended Solids (SS), Chemical Oxygen Demand (COD), Ammoniacal Nitrogen (AN) and potential of Hydrogen (pH) were selected to enable comparison. The statistical methods introduced in this study involved water quality data from Langat River, Selangor since the report by Juahir et al. (2011), Lim et al. (2013a) and Lim et al. (2013b) showed that the Langat river water quality has dynamically changed due to land use transfer to be overload and other environmental exchanges. However, only water quality data from 1995-2009 in the Langat River were considered. The limitations of data selection were based on the availability of water quality data in the Langat River with the permission from Malaysian DOE. Although this may appear to limit a full discussion of the results, the use of proposed statistical methodology in this study may be applicable for other sampling sites and rivers. Since data-driven based approach was applied, modification in the statistical models for other sampling sites and rivers are slightly required depending on how the statistical models are able to validate the data.

1.6 Contribution of the Study

There are several contributions to the existing body of knowledge on WQI in this study. The contributions include development of new descriptive and Bayesian model-based approaches to represent values of selected water quality variables weights accurately. Both approaches are also applicable to identify weights within each sampling sites which are rarely considered by previous researchers. In addition, novelty of validations in PCA models allows the assumption that the test data shares distribution from the training data with the same mean, μ and covariance matrix, Σ. New PCA models which produce a temporal pattern which is close to DOE-WQI is also the main contribution to the literature of statistical methods in WQI development. Furthermore, several criteria introduced in model comparison will also give benefits in determining the best model of river WQI. Also, the Bayesian model-based approach gives additional information on the probability and uncertainty for random quantities of WQI within each sampling site. Besides, similar water quality variables in DOE-WQI have been maintained in statistical models of WQI to enable comparisons between the expert opinion and statistical method approaches. The results obtained from both approaches allow comparison to be made about the water quality variable that give more weights in WQI determination.

1.7 Outline of Thesis

In this thesis, several statistical models have been introduced to gain better knowledge in water quality variable weights determination and WQI estimation. The more accurate the WQI estimation is the better it indicates the quality of water (Lumb et al., 2011a).

Discussion in Chapter 2 starts with reviewing statistical models of river WQI. Apart from that, several potential statistical methods to the proposed for new weights determination and WQI estimation will be discussed in this chapter. The current state of index estimation in other areas using various statistical methods will be explained further. Moreover, an overview of the economic index, i.e. Stock-Watson (S-W) coincident index will also be provided and the suitability of using this method will be discussed. Likewise, suitability of the economic index from Bayesian perspective in WQI estimation will be explained.

Then, preliminary analysis of water quality data will be presented in Chapter 3. Malaysian DOE-WQI calculation and the interpretation of the WQI values will also be provided. Details of sampling sites that have been chosen as a case studies are also discussed. Water quality data analysis including summary of distribution measures, normality test and status of river water quality will be further discussed in this chapter.

Next, several WQI models will be introduced and new weights of water quality variables based on PCA will be developed in Chapter 4. The assumption of multivariate normality in PCA is relaxed due to descriptive approach applied in the WQI development. After the identification and estimation stages in PCA-WQI development, the next step in the validation stage is implemented. New weights calculation based on ranks of the first PCA eigenvector elements and standard deviation are introduced. Several comparative performance of WQI model will also be discussed. Specific statistical analyses and computation using Statistical Analysis Software (SAS) 9.4 will also be carried out.

Then, several potential Bayesian models based on the coincident index approach which was modified from the S-W coincident index will be presented in Chapter 5. The proposed models mentioned earlier are developed under the Bayesian environment due to the ability of the Bayesian approach in incorporating with non-normality data distribution (Brooks et al., 2004), small time series data sets (van de Schoot and Depaoli, 2014), irregular observations (He, 2003) and unbalanced data (Hensman et al., 2013). The potential of new Simultaneous-WQI (S-WQI) model is investigated and the best model that may well describe the water quality data at the selected stations is chosen. The Bayesian models using available software, i.e. Just Another Gibbs Sampler (JAGS) is implemented.

The new Bayesian-WQI for each selected sampling site will be further calculated in Chapter 6 using modification of variables weights obtained from the best model in Chapter 5. New weights based on relative importance ranks of the coefficient parameters in the selected Bayesian model will be used in this chapter. The same process as discussed in Chapter 4 will be repeated with novelty in validation stages. Then, the WQI credible interval will be illustrated to provide better information on the WQI estimation. All computations in this chapter are performed using SAS 9.4.

Finally in Chapter 7, the results of this thesis are summarized and the contributions made to the area of WQI development are highlighted. Several potential areas that would be very useful for future work including new water quality variables to be included, new promising sub-indices data-based to re-calculate new WQI, a potential of nonlinear relationship in Bayesian S-WQI model as well as the use of Bayesian S-WQI model in sufficient regular time series of new experimental data are suggested. A validation of WQI using new experimental data collected at the same selected sampling sites in Langat River are also recommended for future work. The flow of the tesis from Chapter 1 to Chapter 7 is shown in Figure 1.1.

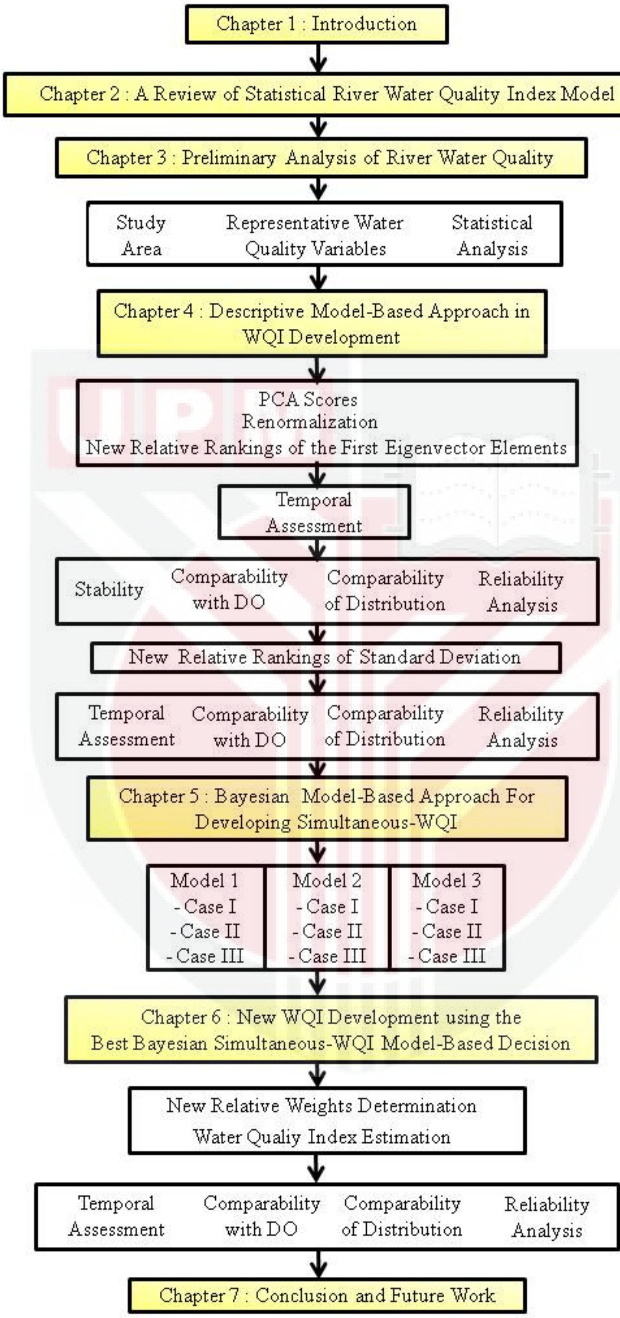

Figure 1.1: Study Flow of the Thesis

BIBLIOGRAPHY

Abba, S., Hadi, S. J., and Abdullahi, J. (2017). River Water Modelling Prediction Using Multi-Linear Regression, Artificial Neural Network, and Adaptive Neuro-Fuzzy Inference System Techniques. *Procedia Computer Science*, 120:75–82.

Abbasi, T. and Abbasi, S. A. (2012). *Water Quality Indices*. Elsevier, Amsterdam.

Abd Majid, N., Rainis, R., and Mohamed, A. F. (2018). Spatial Analysis of Development Pressure in the Langat Basin, Selangor, Malaysia. In *IOP Conference Series: Earth and Environmental Science*, volume 169, pages 1–9. IOP Publishing.

Abyaneh, H. Z. (2014). Evaluation of Multivariate Linear Regression and Artificial Neural Networks in Prediction of Water Quality Parameters. *Journal of Environmental Health Science and Engineering*, 12(1):40.

Ahmad, S., Khan, I. H., and Parida, B. (2001). Performance of Stochastic Approaches for Forecasting River Water Quality. *Water Research*, 35(18):4261–4266.

Alexakis, D., Gotsis, D., and Giakoumakis, S. (2015). Evaluation of Soil Salinization in a Mediterranean Site (Agoulinitsa district–West Greece). *Arabian Journal of Geosciences*, 8(3):1373–1383.

Ali, M. Z. (2007). *The Application of the Artificial Neural Network Model for River Water Quality Classification with Emphasis on the Impact of Land Use Activities : A Case Study from Several Catchments in Malaysia*. PhD thesis, The University of Nottingham.

Alves, M. T. R., Teresa, F. B., and Nabout, J. C. (2014). A Global Scientific Literature of Research on Water Quality Indices: Trends, Biases and Future Directions. *Acta Limnologica Brasiliensia*, 26(3):245–253.

Amini, A., Ali, T. M., Ghazali, A. H., Huat, B. K., et al. (2009). Adjustment of Peak Streamflows of a Tropical River for Urbanization. *American Journal of Environmental Sciences*, 5(3):285–294.

Apha, A. (1995). WPCF, Standard Methods for the Examination of Water and Wastewater. *American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA*.

Ayanso, A., Chatterjee, D., and Cho, D. I. (2011). E-Government Readiness Index: A Methodology and Analysis. *Government Information Quarterly*, 28(4):522–532.

Babamoradi, H., van de Berg, F., and Rinnan, Å. (2013a). Bootstrap Based Confidence Limits in Principal Component Analysis - A Case Study. *Chemometrics and Intelligent Laboratory Systems*, 120:97–105.

Babamoradi, H., van de Berg, F., and Rinnan, Å. (2013b). Comparison of Bootstrap and Asymptotic Confidence Limits for Control Charts in Batch MSPC Strategies. *Chemometrics and Intelligent Laboratory Systems*, 127:102–111.

Bai, V., Bouwmeester, R., and Mohan, S. (2009). Fuzzy Logic Water Quality Index and Importance of Water Quality Parameters. *Air, Soil & Water Research)*, 2:51–59.

Bartram, J. and Ballance, R. (1996). *Water Quality Monitoring : A Practical Guide to the Design and Implementation of Freshwater Quality and Monitoring Programs*. CRC Press.

Baxter, M. (1995). Standardization and Transformation in Principal Component Analysis, with Applications to Archaeometry. *Applied Statistics*, 44(4):513–527.

Beamonte, E., Bermúdez, J. D., Casino, A., and Veres, E. (2005). A Global Stochastic Index for Water Quality: The Case of the River Turia in Spain. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(4):424.

Beck, M. B. and van Straten, G. (2012). *Uncertainty and Forecasting of Water Quality*. Springer Science & Business Media.

Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B., and Bobée, B. (2007). A Review of Statistical Water Temperature Models. *Canadian Water Resources Journal*, 32(3):179–192.

Blanchard, J. L., Coll, M., Trenkel, V. M., Vergnon, R., Yemane, D., Jouffre, D., Link, J. S., and Shin, Y. (2010). Trend Analysis of Indicators: A Comparison of Recent Changes in the Status of Marine Ecosystems Around the World. *ICES Journal of Marine Science*, 67(4):732–744.

Bowerman, B. L., O'Connell, R. T., and Koehler, A. B. (2005). *Forecasting, Time Series, and Regression: An Applied Approach*. Thomson Brooks/Cole.

Box, G. E. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.

Brooks, S., King, R., and Morgan, B. (2004). Bayesian Approach to Combining Animal Abundance and Demographic Data. *Animal Biodiversity and Conservation)*, 27(1):515–529.

Brown, R. M., McClelland, N. I., Deininger, R. A., and Tozer, R. G. (1970). A Water Quality Index - Do We Dare. *Water and Sewage Works*, 117(10):339–343.

Cailas, M. D. and Gehr, R. (1987). Discussion of "Water Quality Index for Chao Phraya River" by Bindu N. Lohani and G. Todino (December, 1984, Vol. 110, No. 6). *Journal of Environmental Engineering*, 113(6):1390–1392.

Camara, M., Jamil, N. R., and Abdullah, A. F. B. (2019). Impact of Land Uses on Water Quality in Malaysia: A Review. *Ecological Processes*, 8(1):10.

Cao, Y., Williams, D. D., and Williams, N. E. (1999). Data Transformation and Standardization in the Multivariate Analysis of River Water Quality. *Ecological Applications*, 9(2):669–677.

Carlie, S. (2010). Assessment Of Water Quality Characteristics During Base and Storm Flow Events On Sungai Langat Basin. Master's thesis, Universiti Teknologi Malaysia.

Carpenter, J. and Bithell, J. (2000). Bootstrap Confidence Intervals : When, Which, What? A Practical Guide for Medical Statisticians. *Statistics in Medicine*, 19(9):1141–1164.

Castillo, L. E. and Nieto, F. H. (2008). Using the Gibbs Sampler for Obtaining a Coincident Economic Index: A Model-Based Approach. *Estadística*, 60(174-175):19–42.

CCME (2001). CCME Water Quality Index 1.0 User's Manual. Technical report, Canadian Council of Ministers of the Environment,.

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis*, 1(4):651–673.

Changyong, F., Hongyue, W., Naiji, L., Tian, C., Hua, H., Ying, L., and Xin, M. T. (2014). Log-Transformation and Its Implication for Data Analysis. *Shanghai Archives of Psychiatry*, 26(2):105.

Chow-Fraser, P. (2006). Development of the Water Quality Index (WQI) to Assess Effects of Basin-Wide Land-Use Alteration on Coastal Marshes of the Laurentian Great Lakes. *Coastal wetlands of the Laurentian Great Lakes: Health, Habitat and Indicators*, pages 137–185.

Chowdhury, M. S. U., Othman, F., Jaafar, W. Z. W., Mood, N. C., and Adham, M. I. (2018). Assessment of Pollution and Improvement Measure of Water Quality Parameters using Scenarios Modeling for Sungai Selangor Basin. *Sains Malaysiana*, 47(3):457–469.

Coletti, C., Testezlaf, R., Ribeiro, T. A., de Souza, R. T., and Pereira, D. d. A. (2010). Water Quality Index using Multivariate Factorial Analysis. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 14(5):517–522.

Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons.

Cooke, S. E., Ahmed, S. M., and MacAlpine, N. (2000). *Introductory Guide to Surface Water Quality Monitoring in Agriculture*. Alberta Agriculture, Food, and Rural Development, Conservation and Development Branch.

Cordoba, E. B., Martinez, A. C., and Ferrer, E. V. (2010). Water Quality Indicators: Comparison of a Probabilistic Index and a General Quality Index. The Case of the Confederación Hidrográfica del Júcar (Spain). *Ecological Indicators*, 10(5):1049–1054.

Crone, T. M. (1994). New Indexes Track the State of the States. *Business Review*, January/February:19–31.

Cude, C. G. (2001). Oregon Water Quality Index a Tool for Evaluating Water Quality Management Effectiveness. *JAWRA Journal of the American Water Resources Association*, 37(1):125–137.

Debels, P., Figueroa, R., Urrutia, R., Barra, R., and Niell, X. (2005). Evaluation of Water Quality in the Chillán River (Central Chile) using Physicochemical Parameters and A Modified Water Quality Index. *Environmental Monitoring and Assessment*, 110(1-3):301–322.

DID (2009). Study on the River Water Quality Trends and Indexes in Peninsular Malaysia. Technical report, Ministry of Natural Resources and Environment.

Dilks, D. W., Canale, R. P., and Meier, P. G. (1992). Development of Bayesian Monte Carlo Techniques for Water Quality Model Uncertainty. *Ecological Modelling*, 62:1–3.

Dinius, S. (1987). Design of An Index of Water Quality. *JAWRA Journal of the American Water Resources Association)*, 23(5):833–843.

Dixon, P. (1992). The Bootstrap and Jacknife: Describing the Precision of Ecological Indices. In Scheiner, S. and Gurevitch, J., editors, *Design and Analysis of Ecological Experiments*, pages 290–318. Chapman and Hall, London, England.

DOE (1997). Malaysian Environmental Quality Report 1997. Technical report, Ministry of Science, Technology and the Environment, Malaysia.

DOE (2007). Malaysian Environmental Quality Report 2007. Technical report, Ministry of Nature Resources and Environment, Malaysia.

DOE (2009). Malaysian Environmental Quality Report 2009. Technical report, Ministry of Nature Resources and Environment, Malaysia.

Dogan, E., Sengorur, B., and Koklu, R. (2009). Modeling Biological Oxygen Demand of the Melen River in Turkey Using an Artificial Neural Network Technique. *Journal of Environmental Management*, 90(2):1229–1235.

Dolph, C. L., Sheshukov, A. Y., Chizinski, C. J., Vondracek, B., and Wilson, B. (2010). The Index of Biological Integrity and the Bootstrap: Can Random Sampling Error Affect Stream Impairment Decisions? *Ecological Indicators*, 10(2):527–537.

Dominic, J. A., Aris, A. Z., and Sulaiman, W. N. A. (2015). Factors Controlling the Suspended Sediment Yield During Rainfall Events of Dry and Wet Weather Conditions in a Tropical Urban Catchment. *Water Resources Management*, 29(12):4519–4538.

dos Santos Simões, F., Moreira, A. B., Bisinoti, M. C., Gimenez, S. M. N., and Yabe, M. J. S. (2008). Water Quality Index as a Simple Indicator of Aquaculture Effects on Aquatic Bodies. *Ecological Indicators*, 8(5):476–484.

DOSM, J. P. M. (1983). *Laporan Am Banci Penduduk Malaysia 1980*. Jabatan Perangkaan Malaysia.

DOSM, J. P. M. (1995). *Laporan Am Banci Penduduk Malaysia 1991*. Jabatan Perangkaan Malaysia.

DOSM, J. P. M. (2001). *Taburan Penduduk mengikut Pihak Berkuasa Tempatan dan Mukim*. Jabatan Perangkaan Malaysia.

DOSM, J. P. M. (2018). *Perangkaan Penting Malaysia 2018*. Jabatan Perangkaan Malaysia.

Dunnette, D. (1979). A Geographically Variable Water Quality Index Used in Oregon. *Journal (Water Pollution Control Federation)*, 51(1):53–61.

Efron, B. (1979). Bootstrap Methods : Another Look at the Jacknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (1981). Nonparametric Estimates of Standard Error: the Jackknife, the Bootstrap and Other Methods. *Biometrika*, 68:589–599.

Efron, B. and Tibshirani, R. (1991). Statistical Data Analysis in the Computer Age. *Science*, 253:390–395.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Erondu, E. and Nduka, E. (1993). A Model for Determining the Water Quality Index (WQI) for the Classification of the New Calabar River at Aluu-Port Harcourt, Nigeria. *International Journal of Environmental Studies*, 44(2-3):131–134.

Ewaid, S. H. and Abed, S. A. (2017). Water Quality Index for Al-Gharraf River, Southern Iraq. *The Egyptian Journal of Aquatic Research*, 43(2):117–122.

Fatimah, M. Y. (1996). Environmental Management of Lake Kenyir, Malaysia: Prevention of Eutrophication. *J. Ensearch*, 9:43–48.

Fernández, N., Ramírez, A., and Solano, F. (2004). Physico-chemical Water Quality Indices - A Comparative Review. *Bistua: Revista de la Facultad de Ciencias Básicas*, 2(1):4211.

Fore, L. S., Karr, J. W., and Conquest, L. L. (1994). Statistical Properties of an Index of Biological Integrity Used to Evaluate Water Resources. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(5):1077–1087.

Freni, G. and Mannina, G. (2010). Bayesian Approach for Uncertainty Quantification in Water Quality Modelling : The Influence of Prior Distribution. *Journal of Hydrology*, 392(1):31–39.

Garson, G. D. (2012). *Testing Statistical Assumptions*. Statistical Associates Publishing.

Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., and Ramli, M. F. (2012). Artificial Neural Network Modeling of the Water Quality Index for Kinta River (Malaysia) using Water Quality Variables as Predictors. *Marine Pollution Bulletin*, 64(11):2409–2420.

Gburek, W. and Folmar, G. (1999). Flow and Chemical Contributions to Stream Flow in an Upland Watershed : A Baseflow Survey. *Journal of Hydrology*, 217(1):1–18.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis (Vol 2)*. CRC press.

Gelman, A. and Shirley, K. (2011). Inference from Simulations and Monitoring Convergence. In *Handbook of Markov Chain Monte Carlo*, pages 163–174, Boca Raton, FL. CRC Press.

Goethals, P. (2005). *Data Driven Development of Predictive Ecological Models for Benthic Macroinvertebrates in Rivers*. Phd dissertation, Ghent University.

Goethals, P. and Niels, D. (2010). Development of a Concept for Integrated Ecological River Assessment in Flanders, Belgium. *Journal of Limnology*, 60(1s):7–15.

HACH (2003). *Water Analysis Hand Book. Fourth Ed.* Hach Company.

Harkins, R. D. (1974). An Objective Water Quality Index. *Journal (Water Pollution Control Federation*, 46(3):588–591.

He, C. Z. (2003). Bayesian Modeling of Age-Specific Survival in Bird Nesting Studies under Irregular Visits. *Biometrics*, 59(4):962–973.

Helsel, D. R. (1987). Advantages of Nonparametric Procedures for Analysis of Water Quality Data. *Hydrological Sciences Journal*, 32(2):179–190.

Hensman, J., Lawrence, N. D., and Rattray, M. (2013). Hierarchical Bayesian Modelling of Gene Expression Time Series Across Irregularly Sampled Replicates and Clusters. *BMC Bioinformatics)*, 14(1):252.

Hoseinzadeh, E., Khorsandi, H., Wei, C., and Alipour, M. (2015). Evaluation of Aydughmush River Water Quality Using the National Sanitation Foundation Water Quality Index (NSFWQI), River Pollution Index (RPI), and Forestry Water Quality Index (FWQI). *Desalination and Water Treatment*, 54(11):2994–3002.

House, M. A. (1986). *Water Quality Indices*. PhD thesis, Middlesex University London.

Idrus, S. and Samad, A. H. (2004). Urban Land Use Change and the Langat Basin Ecosystem Health. *Journal of the Malaysian Institute of Planners*, 2(1):51–68.

Idrus, S., Shah, A., and Mohamed, A. (2003). Analysis of Land Use and Land Cover Changes 1974-2001 in the Langat Basin, Malaysia using Geographic Information System (GIS). In *Ecosystem Health of the Langat Basin, Proceedings of the 2003 Research Symposium on Ecosystem of the Langat Basin*, pages 209–225.

Jackson, D. A. (1993). Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology*, 74(8):2204–2214.

Jeffers, J. (1962). Principal Component Analysis of Designed Experiment. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 12(3):230–242.

Johnson, J. B. and Omland, K. S. (2004). Model Selection in Ecology and Evolution. *Trends in Ecology & Evolution*, 19(2):101–108.

Johnson, R. A. and Wichern, D. W. (2014). *Applied Multivariate Statistical Analysis*. Prentice-Hall.

Jollands, N., Lermit, J., and Patterson, M. (2003). The Usefullness of Aggregate Indicators in Policy Making and Evaluation : A Discussion with Application to Eco-Efficiency Indicators in New Zealand. Technical report, Australian National University.

Joung, H., Miller, W., Mahannah, C., and Guitjens2, J. (1979). A Generalized Water Quality Index Based on Multivariate Factor Analysis. *Journal of Environmental Quality*, 8(1):95–100.

Juahir, H. (2009). *Water Quality Data Analysis and Modeling of the Langat River Basin*. PhD thesis, University of Malaya.

Juahir, H., Zain, S. M., Aris, A. Z., Yusof, M. K., Samah, M. A. A., and Mokhtar, M. (2010a). Hydrological Trend Analysis due to Land Use Changes at Langat River Basin. *Environment Asia*, 3:20–31.

Juahir, H., Zain, S. M., Aris, A. Z., Yusoff, M. K., and Mokhtar, M. B. (2010b). Spatial Assessment of Langat River Water Quality Using Chemometrics. *Journal of Environmental Monitoring*, 12(1):287–295.

Juahir, H., Zain, S. M., Toriman, M. E., Mokhtar, M., and Man, H. C. (2004). Application of Artificial Neural Network Models for Predicting Water Quality Index. *Malaysian Journal of Civil Engineering*, 16(2).

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., and Mokhtar, M. (2011). Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. *Environmental Monitoring and Assessment*, 173(1):625–641.

Kambe, J., Aoyama, T., Yamauchi, A., and Nagashima, U. (2007). Extraction of a Parameter as an Index to Assess Water Quality of the Tamagawa, Tokyo, Japan, by Using Neural Networks and Multivariate Analysis. *Journal of Computer Chemistry, Japan*, 6(1):19–26.

Kannel, P. R., Lee, S., Lee, Y.-S., Kanel, S. R., and Khan, S. P. (2007). Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. *Environmental Monitoring and Assessment*, 132(1):93–110.

Khairuddin, M. and Abd Malek, A. (2001). Program Pencegahan Pencemaran dan Peningkatan Kualiti Air Sungai Langat. *Proceeding Simposium Penyelidikan Lembangan Langat*, pages 183–189.

Kotti, M. E., Vlessidis, A. G., Thanasoulias, N. C., and Evmiridis, N. P. (2005). Assessment of River Water Quality in Northwestern Greece. *Water Resources Management*, 19(1):77–94.

Kumar, D. and Alappat, B. J. (2009). NSF-Water Quality Index : Does It Represent the Experts' Opinion? *Practical Periodical of Hazardous, Toxic an Radioactive Waste Management*, 13(1):75–79.

Kumar, D. and Allapat, B. J. (2004). Selection of the Appropriate Aggregation Function for Calculating Leachate Pollution Index. *Practical Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 8(4):253–264.

Landwehr, J. M. (1974). *Water Quality Indices - Construction and Analysis*. PhD thesis, University of Michigan.

Landwehr, J. M. (1979). A Statistical View of a Class of Water Quality Indices. *Water Resources*, 15(2):460–468.

Landwehr, J. M. and Deininger, R. (1976). A Comparison of Several Water Quality Indexes. *Journal (Water Pollution Control Federal)*, 48(5):954–958.

Lek, S. and Park, Y. S. (2008). Artificial Neural Networks. In *Encyclopedia of Ecology*. Elsevier.

Lewis, F. and Torgerson, P. (2012). A Tutorial in Estimating the Prevalence of Disease in Humans and Animals in the Absence of a Gold Standard Diagnostic. *Emerging Themes in Epidemiology*, 9(1):9.

Liebman, H. (1969). *Atlas of Water Quality: Methods and Practical Conditions*. Oldenbourgh, Munich.

Lim, W. Y., Aris, A. Z., and Praveena, S. M. (2013a). Application of the Chemometric Approach to Evaluate the Spatial Variation of Water Chemistry and the Identification of the Sources of Pollution in Langat River, Malaysia. *Arabian Journal of Geosciences*, 6(12):4891–4901.

Lim, W. Y., Aris, A. Z., and Tengku Ismail, T. H. (2013b). Spatial Geochemical Distribution and Sources of Heavy Metals in the Sediment of Langat River, Western Peninsular Malaysia. *Environmental Forensics*, 14(2):133–145.

Linstone, H. A. and Turoff, M. (1975). *The Delphi Method : Techniques and Application*. Addison-Wesley Publishing Company.

Lobato, T., Hauser-Davis, R., Oliviera, T., Silveira, A., Silva, H., Tavares, M., and Saraiva, A. (2015). Construction of a Novel Water Quality Index and Quality Indicator for Reservoir Water Quality Evaluation : A Case Study in the Amazon Region. *Journal of Hydrology*, 522:674–683.

Lockwood, J., Schervish, M. J., Gurian, P., and Small, M. J. (2001). Characterization of Arsenic Occurrence in Source Waters of US Community Water System. *Journal of the American Statistical Association)*, 96(456):1184–1193.

Lohani, B. and Mustapha, N. (1982a). Indices for Water Quality Assessment in Rivers: A Case Study of the Linggi River in Malaysia. *Water Supply & Management*, 6(6):545–555.

Lohani, B. and Mustapha, N. (1982b). Stochastic Water Quality Index. *Environmental Technology*, 3(1-11):521–528.

Lohani, B. N. and Todino, G. (1984). Water Quality Index for Chao Phraya River. *Journal of Environmental Engineering*, 110(6):1163–1176.

Lumb, A., Sharma, T., and Bibeault, J.-F. (2011a). A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. *Water Quality, Exposure and Health*, 3(1):11–24.

Lumb, A., Sharma, T., Bibeault, J.-F., and Klawunn, P. (2011b). A Comparative Study of USA and Canadian Water Quality Index Models. *Water Quality, Exposure and Health*, 3(3-4):203–216.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS Book: a Practical Introduction to Bayesian Analysis*. CRC Press.

Lystrom, D. J. (1978). *Multiple Regression Modeling Approach for Regional Water Quality Management*. Environmental Research Laboratory, Athens.

Ma, F., Li, C., Wang, X., Yang, Z., Sun, C., and Liang, P. (2014). A Bayesian Method for Comparison Water Quality Evaluation of the Danjiangkou Reservoir Water Source Area, for the Middle Route of the South-to-North Water Diversion Project in China. *Frontiers of Earth Science*, 8(2):242.

Mahmood, R., Messer, J. J., Nemanich, F. J., Liff, C. I., and George, D. B. (1982). A Multivariate Water Quality Index for Use in Management of a Wildland Watershed. Technical report, Utah State University.

Maier, H. R. and Dandy, G. C. (1996). The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. *Water resources research*, 32(4):1013–1022.

Maier, H. R. and Dandy, G. C. (1999). Empirical Comparison of Various Methods for Training Feed-Forward Neural Networks for Salinity Forecasting. *Water Resources Research*, 35(8):2591–2596.

Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P. (2010). Methods Used for the Development of Neural Networks for the Prediction of Water Resource Variables in River Systems: Current Status and Future Directions. *Environmental modelling & software*, 25(8):891–909.

Malakoff, D. (1999). Bayes Offers A 'New' Way to Make Sense of Numbers. *Science*, 286(5444):1460–1464.

McCarthy, M. A. (2007). *Bayesian Methods for Ecology*. Cambridge University Press.

McClelland, N. I. (1974). *Water Quality Index Application in the Kansas River Basin*. US Environmental Protection Agency-Region VII.

McKenna, J. (2003). An Enhanced Cluster Analysis Program with Boostrap Significance Testing for Ecological Community Analysis. *Environmental Modelling & Software*, 18(3):205–220.

Mehlman, D. W., S. U. L. and Kelt, D. A. (1995). Bootstrapping Principal Component Analysis — A Comment. *Ecology*, 76(2):640–643.

Memarian, H., Balasundram, S. K., Talib, J. B., Teh Boon Sung, C., Mohd Sood, A., and Abbaspour, K. C. (2013). KINEROS2 Application for Land Use/Cover Change Impact Analysis at the Hulu Langat Basin, Malaysia. *Water and Environment Journal*, 27(4):549–560.

Mohd Nasir, M. F., Samsudin, M. S., Mohamad, I., Awaluddin, M. R. A., Mansor, M. A., Juahir, H., and Ramli, N. (2011). River Water Quality Modeling using Combined Principal Component Analysis (PCA) and Multiple Linear Regression (MLR): A Case Study at Klang River. *World Applied Sciences Journal)*, 14:73–82.

Monteiro, M. and Costa, M. (2018). A Time Series Model Comparison for Monitoring and Forecasting Water Quality Variables. *Hydrology*, 5(3):37.

Mustapha, A., Aris, A. Z., Ramli, M. F., and Juahir, H. (2012). Temporal Aspects of Surface Water Quality Variation Using Robust Statistical Tools. *The Scientific World Journal*, 2012.

Myung, J. I. and Pitt, M. A. (2016). Model Comparison in Psychology. *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, 5:1–34.

Nasir, M. F. M., Juahir, H., Roslan, N., Mohd, I., Shafie, N. A., and Ramli, N. (2011). Artificial Neural Networks Combined with Sensitivity Analysis as a Prediction Model for Water Quality Index in Juru River, Malaysia. *International Journal of Environmental protection*, 1(3):1–8.

Nguefack-Tsague, G., Klasen, S., and Zucchini, W. (2011). On Weighting the Components of the Human Development Index : A Statistical Justification. *Journal of Human Development and Capabilities)*, 12(2):183–202.

Noorazuan, M., Raini, R., Juahir, H., Zain, S., and Jaafar, N. (2003). GIS Application in Evaluating Land Use-Land Cover Change and It's Impact On Hydrological Regime in Langat River Basin, Malaysia. In *2nd Annual Asian Conference of Map Asia*, pages 14–15.

Ntzoufras, I. (2011). *Bayesian Modeling Using WinBUGS*. John Wiley & Sons.

OECD. (2008). *Handbook on Constructing Composite Indicators : Methodology and User Guide*. Joint Research Centre-European Commission.

Othman, F., ME, A. E., and Mohamed, I. (2012). Trend Analysis of a Tropical Urban River Water Quality in Malaysia. *Journal of Environmental Monitoring*, 14(12):3164–3173.

Ott, W. (1978). *Water Quality Indices: a Survey of Indices Used in the United States*. U.S. Environmental Protection Agency Office of Research and Development.

Parmar, K. S. and Bhardwaj, R. (2014). Water Quality Management Using Statistical Analysis and Time-Series Prediction Model. *Applied Water Science*, 4(4):425–434.

Peng, G. et al. (2004). Testing Normality of Data using SAS. Paper presented at PharmaSUG 2004. Retrieved Jul 4th, 2017.

Pesce, S. F. and Wunderlin, D. A. (2000). Use of Water Quality Indices to Verify the Impact of Cóordoba City (Argentina) on Suguía River. *Water Research*, 34(11):2915–2926.

Petersen, W., Bertino, L., Callies, U., and Zorita, E. (2001). Process Identification by Principal Component Analysis of River Water-Quality Data. *Ecological Modelling*, 138(1):193–213.

Plummer, M. et al. (2003). JAGS: A program for Analysis of Bayesian Graphical Models using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, volume 124, page 125. Vienna.

Plummer, M., Stukalov, A., Denwood, M., and Plummer, M. M. (2016). Package 'rjags'. The Comprehensive R Archive Network. Retrieved March 29th, 2018.

Primpas, I., Tsirtsis, G., Karydis, M., and Kokkoris, G. D. (2010). Principal Component Analysis: Development of a Multivariate Index for Assessing Eutrophication According to the European Water Framework Directive. *Ecological Indicators*, 10(2):178–183.

Qian, S. S. and Reckhow, Kenneth, H. (2007). Combining Model Results and Monitoring Data for Water Quality Assessment. *Environmental Science & Technology)*, 41(14):5008–5013.

Qian, S. S., Schulman, A., Koplos, J., Kotros, A., and Kellar, P. (2004). A Hierarchical Modelling Approach for Estimating National Distribution of Chemicals in Public Drinking Water Systems. *Environmental Science & Technology)*, 38(4):1176–1182.

Qian, S. S., Stow, C. A., and Borsuk, M. E. (2003). On Monte Carlo Methods for Bayesian Inference. *Ecological Modelling*, 159(2):269–277.

Rahaman, Z. A., Che Rus, S. F., Omar, M. A., and Ismail, W. R. (2016). Rivers and Lakes as Natural Heritage: Water Quality Status in The Northern States of Peninsular Malaysia. *KEMANUSIAAN: The Asian Journal of Humanities*, 23(1):109–128.

Razali, N. M., Wah, Y. B., et al. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2(1):21–33.

Reichert, P. and Omlin, M. (1997). On the Usefulness of Overparameterized Ecological Models. *Ecological Modelling*, 95(2-2):289–299.

Reid, M. and Spencer, K. (2009). Use of Principal Component Analysis (PCA) on Estuarine Sediment Datasets : The Effect of Data Pre-treatment. *Environmental Pollution*, 157(8):2275–2281.

Reklaitė, A. (2011). Coincident, Leading and Recession Indexes for the Lithuanian Economy. *Baltic Journal of Economics*, 11(1):91–107.

Roszkowska, E. (2013). Rank Ordering Criteria Weighting Methods – A Comparative Overview. *Optimum Studia Ekonomiczne*, 65(5):14–33.

Rudolph, A., Ahumada, R., and Pérez, C. (2002). Dissolved Oxygen Content as An Index of Water Quality in San Vicente Bay, Chile (36° S). *Environmental Monitoring and Assessment*, 78(1):89–100.

Sahoo, M. M., Patra, K., Swain, J., and Khatua, K. (2017). Evaluation of Water Quality with Application of Bayes' Rule and Entropy Weight Method. *European Journal of Environmental and Civil Engineering*, 21(6):730–752.

Salami, E. and Ehteshami, M. (2015). Simulation, Evaluation and Prediction Modeling of River Water Quality Properties (Case Study:Ireland Rivers). *International journal of environmental science and technology*, 12(10):3235–3242.

Salvati, L. and Carlucci, M. (2014). A Composite Index of Sustainable Development at the Local Scale: Italy as a Case Study. *Ecological Indicators*, 43:162–171.

Sapari, N. and Adlan, M. N. (2009). Non-Point Sources (NPS) Pollution Modelling and Water Quality Improvement Using Wetland. In *International Engineering Convention, Danascus, Syria*. Federation of Engineering Institutions of Islamic Countries.

Sarkar, A. and Pandey, P. (2015). River Water Quality Modelling Using Artificial Neural Network Technique. *Aquatic Procedia*, 4:1070–1077.

Sarkar, C. and Abbasi, S. A. (2006). QUALIDEX – A New Software for Generating Water Quality Indices. *Environmental Monitoring and Assessment*, 119(1):201–231.

Schaeffer, D. J. and Janardan, K. G. (1977). Communicating Environmental Information to the Public : A New Water Quality Index. *The Journal of Environmental Education*, 8(4):18–26.

Shah, K. A. and Joshi, G. S. (2017). Evaluation of Water Quality Index for River Sabarmati, Gujarat, India. *Applied Water Science*, 7(3):1349–1358.

Shoji, H., Yamanoto, T., and Nakamura, T. (1966). Factor Analysis on Stream Pollution of the Yodo River Systems. *Air and Water Pollution*, 10:291–299.

Shrestha, S. and Kazama, F. (2007). Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan. *Environmental Modelling & Software*, 22(4):464–475.

Simeonova, P., Simeonov, V., and Andreev, G. (2003). Water Quality Study of the Struma River Basin, Bulgaria (1989–1998). *Open Chemistry*, 1(2):121–136.

Singh, K. P., Basant, A., Malik, A., and Jain, G. (2009). Artificial Neural Network Modeling of the River Water Quality – A Case Study. *Ecological Modelling*, 220(6):888–895.

Singh, R. P., Nath, S., Prasad, S. C., and Nema, A. K. (2008). Selection of Suitable Aggregation Function for Estimation of Aggregate Pollution Index for River Ganges in India. *Journal of Environmental Engineering*, 134(8):689–701.

Singkran, N., Yenpiem, A., and Sasitorn, P. (2010). Determining Water Conditions in the Northeastern Rivers of Thailand Using Time Series and Water Quality Index Models. *Journal of Sustainable Energy & Environment)*, 1(2):47–58.

Sinha, K. and Das, P. (2015). Assessment of Water Quality Index Using Cluster Analysis and Artificial Neural Network Modeling: A Case Study of the Hooghly River Basin, West Bengal, India. *Desalination and Water Treatment*, 54(1):28–36.

Sládeček, V. (1973). The Reality of Three British Biotic Indices. *Water Research*, 7(7):995–1002.

Song, T. and Kim, K. (2009). Development of A Water Quality Loading Index Based on Water Quality Modeling. *Journal of Environmental Management*, 90(3):1534–1543.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS User Manual. MRC Biostatistics Unit. Retrieved Oct 27th, 2011.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The Deviance Information Criterion: 12 Years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stapleton, E. (1976). The Normal Distribution As a Model of Technological Substitution. *Technological Forecasting and Social Change*, 8(3):325–334.

Stock, J. H. and Watson, M. W. (1988). *A Probability Model of the Coincident Economic Indicators*. Cambridge University Press.

Stock, J. H. and Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators. *NBER Macroeconomics Annual*, 4:351–394.

Sun, W., Xia, C., Xu, M., Guo, J., and Sun, G. (2016). Application of Modified Water Quality Indices as Indicators to Assess the Spatial and Temporal Trends of Water Quality in the Dongjiang River. *Ecological Indicators*, 66:306–312.

Suratmam, S., Sailan, M., Hee, Y., Bedurus, E., and Latif, M. T. (2015). A Preliminary Study of Water Quality Index in Terengganu River Basin, Malaysia. *Sains Malaysiana*, 44(1):67–73.

Sutadian, A. D., Muttil, N., Yilmaz, A. G., and Perera, B. (2016). Development of River Water Quality Indices - A Review. *Environmental Monitoring and Assessment*, 188(1):58.

Terrado, M., Barceló, D., Tauler, R., Borrell, E., and de Campos, S. (2010). Surface-Water-Quality Indices for the Analysis of Data Generated by Automated Sampling Networks. *TrAC Trends in Analytical Chemistry*, 29(1):40–52.

Thi Minh Hanh, P., Sthiannopkao, S., The Ba, D., and Kim, K.-W. (2010). Development of Water Quality Indexes to Identify Pollutants in Vietnam's Surface Water. *Journal of Environmental Engineering*, 137(4):273–283.

Tong, S. T. and Chen, W. (2002). Modelling the Relationship Between Land Use and Surface Water Quality. *Journal of Environmental Management*, 66(4):377–393.

Tonidandel, S., LeBreton, J. M., and Johnson, J. W. (2009). Determining the Statistical Significance of Relative Weights. *Psychological Methods*, 14(4):387.

Uusitalo, L., Lehikoinen, A., Helle, I., and Myrberg, K. (2015). An Overview of Methods to Evaluate Uncertainty of Deterministic Models in Decision Support. *Environmental Modelling & Software*, 63:24–31.

van de Schoot, R. and Depaoli, S. (2014). Bayesian Analysis : Where to Start and What to Report. *European Health Psychologist)*, 16(2):75–64.

Vandekerckhove, J. (2014). A Cognitive Latent Variable Model for the Simultaneous Analysis of Behavioral and Personality Data. *Journal of Mathematical Psychology*, 60:58–71.

Vignolo, A., Pochettino, A., and Cicerone, D. (2006). Water Quality Assessment Using Remote Sensing Techniques: Medrano Creek, Argentina. *Journal of Environmental Management*, 81(4):429–433.

Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., Mustavich, L. F., and Li, B.-L. (2013). Assessment of Surface Water Quality via Multivariate Statistical Techniques : A Case Study of the Songhua River Harbin Region, China. *Journal of Hydro-Environment Research*, 7(1):30–40.

Waugh, F. V. (1962). Factor Analysis : Some Basic Principles and an Application. *Agricultural Economics Research*, 14:77–80.

Whitehead, P., Wilby, R., Battarbee, R., Kernan, M., and Wade, A. J. (2009). A Review of the Potential Impacts of Climate Change on Surface Water Quality. *Hydrological Sciences Journal*, 54(1):101–123.

Whittaker, G., Lautenbach, S., and Volk, M. (2012). What is a Good Index? Problems with Statistically Based Indicators and the Malmquist Index as Alternative. In *International Environmental Modelling and Software Society*. Leipzig, Germany.

Wilhm, J. L. and Dorris, T. C. (1968). Biological Parameters for Water Quality Criteria. *Bioscience*, 18(6):477–481.

Wolock, D. M., Winter, T. C., and McMahon, G. (2004). Delineation and Evaluation of Hydrologic-Landscape Regions in the United States Using Geographic Information System Tools and Multivariate Statistical Analyses. *Environmental Management*, 34(1):S71–S88.

Wu, E. M.-Y. and Kuo, S.-L. (2012). Applying a Multivariate Statistical Analysis Model to Evaluate the Water Quality of a Watershed. *Water Environment Research*, 84(12):2075–2085.

Wu, R., Qian, S. S., Hao, F., Cheng, H., Zhi, D., and Zhang, J. (2011). Modeling Contaminant Concentration Distributions in China's Centralized Source Waters. *Environmental Science & Technology*, 45(14):6041–6048.

Yang, L., Linyu, X., and Shun, L. (2009). Water Quality Analysis of the Songhua River Basin using Multivariate Techniques. *Journal of Water Resources and Protection*, 1(2):110.

Yap, B. W. and Sim, C. H. (2011). Comparisons of Various Types of Normality Tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155.

Yu, C., Quinn, J. T., Dufournaud, C. M., Harrington, J. J., Rogers, P. P., and Lohani, B. N. (1998). Effective Dimensionality of Environment Indicators : A Principal Component Analysis with Bootstrap Confidence Intervals. *Journal of Environmental Management*, 53(1):101–119.

Zainal Abidin, M., Kutty, A. A., Lihan, T., and Zakaria, N. A. (2018). Hydrological Change Effects on Sungai Langat Water Quality. *Sains Malaysiana*, 47(7):1401–1411.

Zainudin, Z. (2010). Benchmarking River Water Quality in Malaysia. *Jurutera*, February:12–15.

BIODATA OF STUDENT

The student, Zalina Mohd Ali, was born in Malacca on 9th December 1976. She studied at Sekolah Kebangsaan Bendahara Seri Maharaja from 1983 to 1988 and Sekolah Menengah Sharifah Rodziah from 1989 - 1993. She attended Sekolah Seri Puteri for matriculation session from 1994-1995 and Science and Mathematics Faculty from 1995-1997 for her B.A. in Statistics. After that, she spent one year during her MSc at the same faculty. She has joined the same faculty as a lecturer a year after. She has started her PhD journey at INSPEM, UPM since December 2010 due to her passion for interdisciplinary approaches in various areas. As part of her Ph.D studies, she focuses on water quality index model comparison based on data-driven approaches from statistical perspectives. She is also interested in analysing environmental, educational and customer satisfaction measurement data. She is currently working as a Senior Lecturer at Statistics Programme in Faculty of Science and Technology, Universiti Kebangsaan Malaysia. The student can be reached via email address; zalina99@ukm.edu.my.

LIST OF PUBLICATIONS

The following are the list of publications that arise from this study.

Journal articles:

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2013). Statistical perspective of river surface water quality index, *Malaysia Water Research Journal*.

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2013). Comparison analysis of classical PCA and robust PCA in WQI development, *Malaysia Water Research Journal*.

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2013). New relative importance of water quality variables in Langat River, *International Journal of Chemical & Environmental Engineering*.

Books/Chapter in Books :

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan, Hafizan Juahir and Faridatul Azna Ahmad Shahabuddin (2012). Temporal Water Quality Assessment of Langat River from 1995-2006, *chapter in the book of Water Quality Monitoring and Assessment*, pp. 321-346. InTech.

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2014). Discriminant Analysis of Water Quality Data in Langat River, *chapter in the book of From Sources to Solution*, pp. 597-601. Springer.

Proceedings:

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2013). The selection of the Bayesian coincident-index models using model comparison criterion with application in Langat river water quality data. *In the PROCEEDINGS OF THE 20TH NATIONAL SYMPOSIUM ON MATHEMATICAL SCIENCES 2013*, 18 - 20 December 2012, Putrajaya, pp. 1283-1292.

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2014). Bayesian model-based approach for developing a river water quality index. *In the PROCEEDINGS OF THE STATISTICS AND OPERATIONAL RESEARCH INTERNATIONAL CONFERENCE (SORIC 2013)*, 3 - 5 December 2013, Sarawak, pp. 45-54.

Zalina Mohd Ali, Noor Akma Ibrahim, Kerrie Mengersen, Mahendran Shitan and Hafizan Juahir (2014). Robust principal component analysis in water quality index development. *In the THE 3RD INTERNATIONAL CONFERENCE ON MATHEMATICAL SCIENCES (ICMS2013)*, 17 - 19 December 2013, Kuala Lumpur, pp. 1091-1097.

INDEX

accuracy, 2–4, 13, 20, 24, 25, 62, 75, 76, 100, 131, 140
artificial neural network, 14, 20
autocorrelation, 14, 110, 115–117, 119–122, 125, 126

Bayesian models, 4–6, 14, 28, 100, 101, 103, 130
bootstrap, 25, 26, 54, 75–79, 83, 97–99, 140, 145
box plots, 39, 83, 86, 87

coefficient of variation, 38, 107
coefficient values, 4, 54, 131, 132
comparability, 1–4, 12, 13, 21, 23, 24, 29, 71, 72, 90, 93, 134, 138
comparison, 2, 3, 5, 10, 11, 13, 17, 19, 21, 23, 24, 26, 28, 29, 58, 59, 62, 72, 90, 101, 109, 138, 144–146
confidence interval, 25, 54, 75, 76, 98, 99, 140–143
consistent, 25, 27, 32, 54, 63, 71, 72, 83, 88, 90, 93, 96, 108, 134, 138, 140, 145
convergence, 11, 28, 100, 101, 107–110, 117, 123–126, 145
correlation, 16, 18, 21, 26, 52, 54, 55, 71, 72, 90–93, 115
credible interval, 7, 101, 130, 140

data-driven approach, 2, 5, 14, 20, 100, 126, 140, 144–146
descriptive model, 4, 5, 52, 76, 100
deterministic models, 9, 12–14
Deviance Information Criteria (DIC), 28, 100, 107, 108, 123–126, 145
diagnostic checking, 4, 100, 109, 123, 126, 145
dispersion, 61, 71, 76, 90–93, 96
distributed-lag model, 4, 102–109

eigenvalues, 16, 53, 55, 57
eigenvector elements, 2, 4, 6, 15, 24, 25, 53, 55–57, 59–61, 96, 145

expert opinion, 1, 3, 5, 9, 10, 12, 16, 21, 59, 61, 144
exponential model, 19

flexibility, 2–4, 24, 96

histogram, 39, 41, 42, 72, 139

inverse, 39, 42, 48, 50

Just Another Gibbs Sampler (JAGS), 6, 28, 101, 109, 110, 123

kernel density plot, 110–114, 117, 118, 121

log transformation, 15, 39, 41, 47, 50

multivariate, 6, 14–17, 21, 52

non-normality, 6, 51, 52, 145
non-parametric, 9, 14, 20, 21, 25, 26, 52
normality, 6, 15, 20, 39, 47, 49, 52, 100
normalization, 15, 21, 24, 58

overall-station, 3, 4, 52, 61, 86, 134, 137

parametric, 9, 14, 15, 20, 21, 25, 39, 51
Pearson correlation, 52–54, 72, 90, 138
performance, 2, 4, 6, 19–21, 23, 24, 62, 76, 80–82, 86, 123, 134, 144
posterior, 23, 28, 109, 110, 121, 123, 126–130, 132
precision, 25, 76, 93, 108, 110, 123, 125, 140
principal component analysis (PCA), 2–6, 14, 15, 23–26, 28, 52–56, 58–62, 71, 75, 96, 145
prior, 22, 54, 100, 103–105, 109, 110, 125, 126
probability, 5, 14, 18, 19, 23, 100, 140

Q-Q plot, 39, 46–49

regression, 14, 17, 18, 26, 102, 132
relative ranking, 3, 5, 19–21, 26, 52, 59, 60
reliability, 2–4, 10, 17, 20, 22–25, 75, 93, 96, 110, 140, 145

robust, 21, 54, 61, 71, 96, 123, 126, 145

sampling, 18, 25, 27, 38, 75, 77–79
sampling sites, 4–7, 18, 26, 27, 29, 30, 75, 100, 101, 123, 125, 127, 133, 134, 144
sampling time, 17, 28, 36–38, 76, 93, 96, 99–104, 108, 127, 140
sensitivity, 4, 13, 16, 19–21, 25, 96, 100, 109, 110, 125, 126, 140, 146
Shapiro-Wilk, 39, 50
Simultaneous-WQI (S-WQI), 4, 6, 7, 97, 100, 101, 103, 126, 127, 144, 145
Spearman correlation, 52, 54, 72, 90, 138
square root, 39, 42, 50
stability, 2–4, 13, 24, 38, 71, 75, 96, 107, 123, 125–127, 130, 140, 145
standard deviation, 4, 6, 18, 19, 38, 61, 62, 72, 76, 83, 86, 96, 130, 145
standardization, 15, 16, 24, 53, 58, 96
Statistical Analysis Software (SAS), 6
statistical models, 2, 3, 5, 6, 9, 12, 14, 15, 20, 22, 24, 26, 29, 62, 100
stochastic, 14, 18, 19
Stock-Watson (S-W) index, 6, 27, 101
sub-indices, 1, 2, 7, 24, 30, 38, 39, 52–57, 59–61, 71, 100–102, 132, 133, 145, 146

temporal, 4, 5, 27, 29, 53, 62, 63, 72, 75, 88–90, 101, 107, 134–138, 146
time series, 6, 7, 14, 17, 18, 27, 97, 101
trace plot, 110–114, 117, 118, 121
transformation, 3, 10, 11, 15, 24, 28, 39, 41, 42, 47, 50, 53

uncertainty, 3, 5, 13, 18, 22, 25, 27, 76, 100

weights, 2–6, 10–21, 23, 24, 26, 27, 29–31, 56, 57, 59–62, 71, 83–86, 96, 102, 126, 131–135, 138, 140, 144, 145
within-station, 3–5, 26, 83, 134, 137

UNIVERSITI PUTRA MALAYSIA
STATUS CONFIRMATION FOR THESIS/PROJECT REPORT AND COPYRIGHT
ACADEMIC SESSION: SECOND SEMESTER 2018/2019

TITLE OF THE THESIS/PROJECT REPORT:
STATISTICAL MODELING OF RIVER WATER QUALITY INDEX

NAME OF STUDENT: ZALINA BINTI MOHD ALI

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(✓)

☐	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 acces.

This thesis is submitted for:

☐	PATENT	Embargo from ___________ until ___________. (date) (date)

Approved by:

____
(Signature of Student)
New IC No/Passport No.:761209-04-5354

Date:

____
(Signature of Chairman of Supervisory Committee)
Name: **Noor Akma bt Ibrahim**

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.]