UNIVERSITI PUTRA MALAYSIA

PATTERN RECOGNITION APPROACH FOR SPATIAL AND TEMPORAL VARIATION ANALYSIS OF SURFACE WATER QUALITY OF KLANG RIVER BASIN, MALAYSIA

MOHD FAHMI BIN MOHD NASIR

FPAS 2013 20
PATTERN RECOGNITION APPROACH FOR SPATIAL AND TEMPORAL VARIATION ANALYSIS OF SURFACE WATER QUALITY OF KLANG RIVER BASIN, MALAYSIA

MOHD FAHMI BIN MOHD NASIR

MASTER OF SCIENCE
UNIVERSITI PUTRA MALAYSIA

2013
PATTERN RECOGNITION APPROACH FOR SPATIAL AND TEMPORAL VARIATION ANALYSIS OF SURFACE WATER QUALITY OF KLANG RIVER BASIN, MALAYSIA

By

MOHD FAHMI BIN MOHD NASIR

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

February 2013
COPYRIGHT

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

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

PATTERN RECOGNITION APPROACH FOR SPATIAL AND TEMPORAL VARIATION ANALYSIS OF SURFACE WATER QUALITY OF KLANG RIVER BASIN, MALAYSIA

By

MOHD FAHMI BIN MOHD NASIR

February 2013

Chairman: Hafizan Juahir, PhD

Faculty: Environmental Studies

In Malaysia, the hydrology and surface water quality of a river are often discussed among the city dwellers especially in the urban area as a consequence from the flash flood scenario during the rain event seasons. This study was conducted in Klang River in attempt to interpret the relationship between hydrological and surface water quality parameters, to estimate the pollution loading in the river with and without the utilization of the hydrological data and to ascertain the input and output parameters of the water quality data using the artificial neural network (ANN). The data was collected from the Department of Environment (DOE) and Department of Irrigation and Drainage (DID), Malaysia. Approximately 30 water quality parameters, rainfall, discharge and direct runoff data were considered for data analysis. This study integrates statistical tools such as non-parametric analysis, multivariate methods and ANN modeling in order to
describe the spatial and temporal variation of water quality at Klang River Basin. This study revealed that among the 24 rainfall stations, Rs11 and Rs14 indicate the highest variations of rainfall with Rs02 exhibiting a downward trend with rainfall while the other stations indicate no trend. Among all the variables, temperature (TEMP) and mercury (Hg) were indicate the highest correlation value at \( p<0.05 \) (\( R=-0.245 \) and \( R=0.295 \)) with rainfall while other stations have no significant relationship with rainfall. Despite that, river discharge illustrate Ds03 and Ds04 having an upward trend with only dissolved oxygen (DO), ammoniacal nitrogen (NH\textsubscript{3}N), TEMP, conductivity (COND), salinity (SAL), total solid (TS), chloride (Cl\textsuperscript{-}), phosphate (PO\textsuperscript{4-3}), magnesium (Mg) and methylene blue active substances (MBAS) significantly (\( p<0.05 \)) correlated with discharge and direct runoff. Source identification using principal component analysis (PCA) confirmed that an anthropogenic activity does influence the river water quality. The clustering of 30 monitoring stations exhibit 4 clusters categorizes as Clean (C), Moderately Polluted (MP), Polluted (P) and Highly Polluted (HP) as the spatial factor meanwhile seasonal Wet, Dry and Transitional (Trans) and water level Low (L), Normal (N) and High (H) as temporal factors. The spatial (92%), seasonal (93%) and water level (99%) are correctly assigned using standard mode, forward stepwise and backward stepwise mode of discriminant analysis (DA). PCA recognized 22 parameters in the C groups. Meanwhile, all the parameters were identified in MP and P except for arsenic (As), zinc (Zn), iron (Fe), oil and grease (OG)) and (NH\textsubscript{3}N, PO\textsuperscript{4-3} and MBAS). These pollutants are mainly comes from soil erosion, anthropogenic, household, domestic and industrial wastewater and sullage. Based on the source identification and absolute principal component score-ANN (APCS-ANN) model, industrial and domestic wastewaters effluent are categories as pollution source 1 (PS1) and notified as the
highest contributor (33%) while flood mitigation activities and seasonal effect (PS5) as the lowest contributor (1%). The ANN-DA models revealed better prediction performance using ANN-DA Seasonal (ADSe) model due to strong correlation ($R=0.9871$) with low root mean square error (RMSE) and sum of squares error (SSE) values compared to the other 2 models. This study caters an integrated picture for government agencies to solve the large complex datasets into a more comprehensive and systematic approach. This multivariate statistical technique provides substantial information and knowledge on the identification of pollution sources for improvement and maintenance of the watershed management practices in Malaysia. Hence, assist the Klang River Basin management to meet the criteria required by the authorities, especially DOE and DID.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Master Sains

PENDEKATAN MENGENALPASTI CORAK ANALISIS VARIASI RUANG DAN MASA KUALITI PERMUKAAN AIR DI LEMBANGAN SUNGAI KLANG, MALAYSIA

Oleh

MOHD FAHMI BIN MOHD NASIR

Februari 2013

Pengerusi: Hafizan Juahir, PhD

Fakulti: Pengajian Alam Sekitar

Di Malaysia, hidrologi dan kualiti air permukaan sungai telah menjadi salah satu topik yang sering dibincangkan oleh penduduk bandar oleh disebabkan berkaitan dengan senario banjir kilat semasa musim hujan. Kajian ini adalah untuk mengkaji hubungan antara hidrologi dan data kualiti air permukaan, menganggarkan beban pencemaran dengan penggunaan dan tanpa penggunaan data hidrologi dan menentukan parameter input dan output data kualiti air dengan mengaplikasikan rangkaian tiruan neural (ANN). Semua pangkalan data telah diperolehi daripada Jabatan Alam Sekitar (JAS) dan Jabatan Pengaliran dan Saliran (JPS) Malaysia. Kira-kira 30 parameter kualiti air dipilih untuk analisis data termasuk data hujan, air pelepasan dan air larian langsung. Kajian ini mengintegrasikan alat statistik seperti analisis bukan parametrik, kaedah multivariat dan model ANN untuk menggambarkan variasi ruang dan masa kualiti air di Lembangan
Sungai Klang. Kajian ini mendedahkan bahawa di antara 24 stesen hujan, Rs11 dan Rs14 menunjukkan variasi hujan tertinggi pada tahun 1997-2010 manakala Rs02 menunjukkan trend (menurun) dan stesen hujan lain (tiada trend). Sebaliknya, di antara semua pembolehubah suhu (TEMP) dan merkuri (Hg) telah menunjukkan nilai korelasi tertinggi pada $p<0.05$ ($R=-0.245$ dan $R=0.295$) untuk hujan manakala stesen lain tidak menimbulkan apa-apa hubungan yang signifikan dengan hujan. Walaupun demikian, air pelepasan sungai telah menunjukkan trend menaik di Ds03 dan Ds04 dengan hanya oksigen terlarut (DO), ammonia nitrogen (NH$_3$N), TEMP, kekonduksian (COND), kemasinan (SAL), jumlah pepejal (TS), klorin (Cl$^-$), fosfat (PO$_4^{3-}$), magnesium (Mg) dan bahan aktif metilena biru (MBAS) telah dikenalpasti dengan ketara ($p<0.05$) berkait rapat dengan air pelepasan dan air aliran langsung. Pengenalpastian sumber pencemar menggunakan analisis komponen prinsipal (PCA) menunjukkan bahawa aktiviti antropogenik mempengaruhi kualiti air sungai di Kuala Lumpur (KL). Kelompok 30 stesen pemantauan mempamerkan 4 kelompok dikategorikan sebagai Bersih (C), Sederhana Tercemar (MP), Tercemar (P) dan Sangat Tercemar (HP) yang mewakili faktor ruang manakala faktor musim dikategorikan sebagai Basah (W), Kering (D) dan Peralihan (Trans) dan paras air yang Rendah (L), Normal (N) dan Tinggi (H) sebagai faktor masa. Ruang (92%), musim (93%) dan paras air (99%) ditandakan betul menggunakan analysis diskriminan (DA) mod biasa, mod ke hadapan langkah demi langkah dan mod mundur langkah demi langkah. PCA mengenal pasti 22 parameter dalam kumpulan bersih (C). Sementara itu, semua parameter telah dikenal pasti dalam MP kecuali arsenik (As), zink (Zn), besi (Fe) dan OG sama juga dengan P dan MP (NH$_3$N, PO$_4^{3-}$ dan MBAS). Pencemaran ini terutamanya daripada hakisan tanah, antropogenik, isi rumah, air kumbuhan domestik dan industri dan air basuhan.
Berdasarkan pencarian sumber dan model perincian komponen utama mutlak-ANN (APCS-ANN), perindustrian dan air sisa domestik adalah dikategorikan sebagai sumber pencemar 1 (PS1) dikenalpasti sebagai penyumbang tertinggi (33%) manakala aktiviti tebatan banjir dan kesan musim (PS5) menunjukkan sebagai penyumbang terendah iaitu 1%. Model ANN-DA mendedahkan prestasi ramalan yang lebih baik menggunakan model musim ANN-DA (ADSe) di mana \( R=0.9871 \) menunjukkan korelasi yang kukuh dengan nilai ralat purata punca kuasa dua (RMSE) dan jumlah ralat punca kuasa dua (SSE) rendah berbanding dengan 2 model yang lain. Kajian ini memberikan gambaran yang bersepadu untuk agensi kerajaan menyelesaikan masalah berkaitan dengan timbunan data yang komplek kepada pendekatan yang lebih menyeluruh dan sistematik. Teknik statistik multivariat menyediakan lebih banyak maklumat dan pengetahuan dalam mengenal pasti sumber pencemaran bagi perbaikan dan penyelenggaraan amalan pengurusan kawasan tadahan air di Malaysia. Oleh itu, membantu pengurusan Lembangan Sungai Klang untuk memenuhi kriteria yang dikehendaki oleh pihak berkuasa, terutama JAS dan JPS.
ACKNOWLEDGEMENTS

Alhamdulillah, I am grateful and thankful to the Almighty Allah S.W.T for giving me strength, courage, patience and determination in preparing and completing this research successfully.

I would like to express my deepest gratitude to my supervisor, Dr Hafizan Juahir for his continuous supervision, assistance, guidance, opportunity and trust for the accomplishment of this research. Other than that, I would also like to thank my co-supervisor, Assoc. Prof. Dr. Mohammad Firuz Ramli and Prof. Dr. Sharifuddin Md Zain for their assistance, guidance and ideas in order for me to improved and beautify the research.

In addition, I wish to thank all the government agencies and staffs especially from Faculty of Environmental Studies (UPM), DOE and DID in favor of providing the database for this study. Special thanks to Alam Sekitar Malaysia Sdn Bhd (ASMA) that have been assisting with the site visits and also providing substantial information in order to strengthen this research. My deepest appreciation to the Ministry of Higher Education (MOHE) and Graduate Research Fellowship (GRF) from Universiti Putra Malaysia for the financial support in order for me to complete this research.

I would like to be grateful to my parents Mohd Nasir Said and Aeshah Mahmod for their constant support, encouragement, financial support, prayers and for being the most understanding parents during my ups and downs. Not forgotten to my family member Hathib Al-Anshaary, Mohd Fadli, Norhafizah and Mohd Fazil Hilmi for their supports and continuously encouraging me to keep on striving for excellence. My special thanks
to Nur Aliaa for always being there in supporting, encouraging, helping, reminding and motivates me constantly.

I wish to be thankful to all my friends for their continuous support, reminders, encouragement, advice, ideas, assistance and for lending their hands throughout my studies. I dedicate this appreciation to Isahak, Munirah, Amar, Saiful, Roshide, Hazzeman, Hajar, Norazida, Nurul Afiqah, Norliza, Nur Hazirah, Shah, Adamu and Nura. There are no words that I can give other than thankful to all the assistance that I have gained in making this research to live. I believe only Allah can repay the kindness that each one have give me in order for my dreams to come true. Honestly, these experiences have been nothing but the best in my life. Last but not least, thank you to all that have significantly or insignificantly contribute to the completion of this research. Thank You.
I certify that a Thesis Examination Committee has met on 11 March 2013 to conduct the final examination of Mohd Fahmi bin Mohd Nasir on his thesis entitled "Pattern Recognition Approach for Spatial and Temporal Variation Analysis of Surface Water Quality of Klang River Basin, Malaysia" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

**Puziah bt Abdul Latif, PhD**
Associate Professor
Faculty of Environmental Studies
Universiti Putra Malaysia
(Chairman)

**Ahmad Zaharin Aris, PhD**
Senior Lecturer
Faculty of Environmental Studies
Universiti Putra Malaysia
(Internal examiner)

**Ahmad Makmom bin Abdullah, PhD**
Associate Professor
Faculty of Environmental Studies
Universiti Putra Malaysia
(Internal examiner)

**Mohd Talib Latif, PhD**
Associate Professor
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
(External examiner)

_NORITAH OMAR, PhD_
Professor and Deputy Dean
School of Graduates Studies
Universiti Putra Malaysia

_Date: 23 May 2013_
This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory committee were as follows:

**Hafizan Juahir, PhD**  
Senior Lecturer  
Faculty of Environmental Studies  
Universiti Putra Malaysia  
(Chairman)

**Mohammad Firuz Ramli, PhD**  
Associate Professor  
Faculty of Environmental Studies  
Universiti Putra Malaysia  
(Committee)

__________________________________________  
**BUJANG BIN KIM HUAT, PhD**  
Professor and Dean  
School of Graduates Studies  
Universiti Putra Malaysia

Date:
DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and its not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

___________________________
MOHD FAHMI MOHD NASIR
Date: 28 February 2013
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRAK</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>APPROVAL</td>
<td>vii</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xvii</td>
</tr>
</tbody>
</table>

CHAPTER
1  INTRODUCTION
   1.1 General introduction  1
   1.2 Problem statement     6
   1.3 Significant of study  7
   1.4 Research objective    8
   1.5 Research hypothesis   8

2  LITERATURE REVIEW
   2.1 Water                  9
   2.2 Water pollution        11
   2.3 River                  12
   2.4 River pollution        15
   2.5 Effects of climate change and hydrological on water quality 16
   2.6 River water quality parameter 19
      2.6.1 Physicochemical parameter 22
      2.6.2 Metal 27
      2.6.3 Biological parameter 31
   2.7 River water quality analysis and modeling 32
      2.7.1 Trend analysis 32
      2.7.2 Multivariate statistical analysis 33
      2.7.3 Application of Artificial neural network (ANN) 35

3  METHODOLOGY
   3.1 Background of study area 38
   3.2 Data collection 40
      3.2.1 Water sample collection 41
      3.2.2 Preprocessing data 44
   3.4 Rainfall analysis 45
   3.5 Discharge and direct runoff analysis 47
   3.6 Trend analysis 48
   3.7 Spatial and temporal analysis 50
3.7.1 Hierarchical agglomerative cluster analysis 50
3.7.2 Discriminant analysis 51
3.7.3 Principal component analysis 52
3.7.4 Receptor modeling based on absolute principal component scores 54
3.8 Artificial neural network 57
  3.8.1 Evaluation of Model Performance 59
3.9 Sensitivity analysis 60

4 RESULTS AND DISCUSSION 61
4.1 Rainfall analysis 61
4.2 Discharge and direct runoff analysis 67
4.3 Spatial and temporal variation of river water quality 80
  4.3.1 Source identification using PCA/FA at Klang River, KL including hydrological trend 81
  4.3.2 Spatial classification of surface water quality data 85
  4.3.3 Spatial discrimination of river water quality pattern 88
  4.3.4 Temporal variation in river water quality 92
4.4 Data structure determination and source identification 99
4.5 Source apportionment using APCS-ANN models 119
4.6 Prediction of water quality using DA-ANN models 121

5 CONCLUSION 127
REFERENCES 132
APPENDICES 155
BIODATA OF STUDENT 168
LIST OF PUBLICATION 169
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Description of thirty water quality stations along Klang River Basin</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Pair of water quality stations and discharge stations</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>The selected rainfall station along Klang River</td>
<td>46</td>
</tr>
<tr>
<td>3.4</td>
<td>Discharge stations at Wilayah Persekutuan, Kuala Lumpur</td>
<td>48</td>
</tr>
<tr>
<td>4.1</td>
<td>Descriptive statistic rainfall</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Class of rainfall (%) for selected station within Klang River</td>
<td>64</td>
</tr>
<tr>
<td>4.3</td>
<td>Mann-Kendall test of trend for rainfall during 1997-2010</td>
<td>66</td>
</tr>
<tr>
<td>4.4</td>
<td>Descriptive statistics of four discharge stations (m$^3$/d)</td>
<td>68</td>
</tr>
<tr>
<td>4.5</td>
<td>Parametric tests for discharge at the four sampling stations</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>Mann-Kendall test for discharge</td>
<td>76</td>
</tr>
<tr>
<td>4.7</td>
<td>Mann-Kendall test for direct runoff</td>
<td>79</td>
</tr>
<tr>
<td>4.8</td>
<td>Factor loadings after varimax rotation</td>
<td>82</td>
</tr>
<tr>
<td>4.9</td>
<td>Classification matrix for DA of regions variations at Klang River</td>
<td>90</td>
</tr>
<tr>
<td>4.10</td>
<td>Classification matrix for DA of seasonal variations at Klang River</td>
<td>94</td>
</tr>
<tr>
<td>4.11</td>
<td>Classification matrix for DA of water level variations at Klang River</td>
<td>97</td>
</tr>
<tr>
<td>4.12</td>
<td>Factor loadings after varimax rotation of physico-chemical parameters (30) on significant PCs for C, MP, P and HP data sets</td>
<td>106</td>
</tr>
<tr>
<td>4.13</td>
<td>Factor loadings after varimax rotation of physico-chemical parameters (30) on significant PCs for Wet, Dry and Transitional data sets</td>
<td>116</td>
</tr>
<tr>
<td>4.14</td>
<td>Factor loadings after varimax rotation of physico-chemical parameters (30) on significant PCs for Low, Normal and High data sets</td>
<td>117</td>
</tr>
<tr>
<td>4.15</td>
<td>Contribution for each pollution sources</td>
<td>121</td>
</tr>
<tr>
<td>4.16</td>
<td>Evaluation of model performance of three ANN-DA models</td>
<td>123</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Hydrologic cycle (USGS, 2012)</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>The river flow (Geomorphology, 2012)</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Land uses at Klang Valley (CHKL, 2012e)</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>Study area and geographical location of water quality stations at Klang River Basin</td>
<td>40</td>
</tr>
<tr>
<td>3.3</td>
<td>The location of pair of water quality stations and discharge stations at Kuala Lumpur</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>Flow diagram of overall methodology</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>The location of selected rainfall station along Klang River Basin</td>
<td>46</td>
</tr>
<tr>
<td>3.6</td>
<td>Example of network structure used in possible sources recognition by APCS-ANN model</td>
<td>58</td>
</tr>
<tr>
<td>3.7</td>
<td>Example of network structure used in spatial and temporal recognition by ANN-DA model</td>
<td>58</td>
</tr>
<tr>
<td>4.1</td>
<td>Box-and-whisker plots for mean, minimum, maximum and maximum-minimum ratio of discharge for all stations</td>
<td>69</td>
</tr>
<tr>
<td>4.2</td>
<td>Mean discharges at four discharge stations: (a) Ds01, (b) Ds02, (c) Ds03 and (d) Ds03</td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>Max discharges at four discharge stations: (a) Ds01, (b) Ds02, (c) Ds03 and (d) Ds03</td>
<td>72</td>
</tr>
<tr>
<td>4.4</td>
<td>Min discharge at four discharge stations: (a) Ds01, (b) Ds02, (c) Ds03 and (d) Ds03</td>
<td>73</td>
</tr>
<tr>
<td>4.5</td>
<td>Max/min discharge at four discharge stations: (a) Ds01, (b) Ds02, (c) Ds03 and (d) Ds03</td>
<td>74</td>
</tr>
<tr>
<td>4.6</td>
<td>Direct runoff at four discharge stations</td>
<td>77</td>
</tr>
<tr>
<td>4.7</td>
<td>Dendrogram showing clustering of sampling stations according to the surface water quality characteristics of Klang River Basin</td>
<td>86</td>
</tr>
</tbody>
</table>
4.8 Spatial distribution of sampling stations into 4 clusters (Clean, Moderately polluted, Polluted and Clean) 86
4.9 The chosen sampling station for river water quality monitoring 87
4.10 Box and whisker plots of some parameters discriminated by spatial DA at Klang River Basin 90
4.11 Box and whisker plots of some parameters recognized by seasonal DA at Klang River Basin 94
4.12 Box and whisker plots of some parameters water level DA at Klang River Basin 98
4.13 (a) Prediction and (b) %RE of ADSp model 124
4.14 (a) Prediction and (b) %RE of ADSe model 125
4.15 (a) Prediction and (b) %RE of ADWi model 126
LIST OF ABBREVIATIONS

ADSe  ANN-DA Season
ADSp  ANN-DA Spatial
ADWL  ANN-DA Water level
ANN   Artificial neural network
ANN-DA Artificial neural network-discriminant analysis
APCS  Absolute principal component scores
APCS-ANN Absolute principal component scores -artificial neural network
APHA  American Public Health Association
As    Arsenic
ASMA  Alam Sekitar Malaysia Sdn Bhd
BOD   Biochemical oxygen demand
C    Clean
CA    Cluster analysis
Ca    Calcium
CEEP  Centre Europeen d'Etudes des Polyphosphates
CHKL  Kuala Lumpur City Hall
Cl    Chloride
CM    Classification matrice
CO2   Carbon dioxide
COD   Chemical oxygen demand
COND  Conductivity
Cr    Chromium
cumecs Cubic meter per second
D    Dry
DA    Discriminant analysis
DF    Discriminant function
DID   Department of Irrigation and Drainage
DO    Dissolved oxygen
DOE  Department of Environment
Dr   Direct runoff
DS   Dissolved solid
Ds   Discharge station
E    East
E. coli  Escherichia coli
Eq.  Equation
FA   Factor analysis
Fe   Iron
H    High
HACA Hierarchical agglomerative cluster analysis
NH₃N  Ammoniacal nitrogen
HP   Highly polluted
K    Potassium
kg   Kilogram
KL   Kuala Lumpur
km   Kilometre
km²  Square kilometre
L    Low
L-PS Leave-one-out pollution source
m³/s Cubic metre per second
max  Maximum
MBAS Methylene blue active substance
MeHg Methylmercury
Mg   Magnesium
min  Minimum
mm   Millimetre
MP   Moderately polluted
N    Normal
Na   Sodium
no.  Number
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO$_3^-$</td>
<td>Nitrate</td>
</tr>
<tr>
<td>NT</td>
<td>No trend</td>
</tr>
<tr>
<td>Obs.</td>
<td>Observation</td>
</tr>
<tr>
<td>OG</td>
<td>Oil and grease</td>
</tr>
<tr>
<td>P</td>
<td>Polluted</td>
</tr>
<tr>
<td>Pb</td>
<td>Lead</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCs</td>
<td>Principal components</td>
</tr>
<tr>
<td>PO$_4^{3-}$</td>
<td>Phosphate</td>
</tr>
<tr>
<td>$R$</td>
<td>Coefficient of correlation</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>Rs</td>
<td>Rainfall station</td>
</tr>
<tr>
<td>S</td>
<td>Mann Kendall Statistic</td>
</tr>
<tr>
<td>SAL</td>
<td>Salinity</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SMIs</td>
<td>Small and Medium Industries</td>
</tr>
<tr>
<td>SS</td>
<td>Suspended solids</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum of squares error</td>
</tr>
<tr>
<td>T</td>
<td>Transitional</td>
</tr>
<tr>
<td>TDS</td>
<td>Total dissolved solid</td>
</tr>
<tr>
<td>TEMP</td>
<td>Temperature</td>
</tr>
<tr>
<td>TS</td>
<td>Total solid</td>
</tr>
<tr>
<td>TUR</td>
<td>Turbidity</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>USGS</td>
<td>United States Geological Survey</td>
</tr>
<tr>
<td>VF</td>
<td>Varifactor</td>
</tr>
<tr>
<td>W</td>
<td>Wet</td>
</tr>
<tr>
<td>WHAT</td>
<td>Web Hydrograph Analysis Tool</td>
</tr>
<tr>
<td>WP</td>
<td>Wilayah Persekutuan</td>
</tr>
<tr>
<td>WQI</td>
<td>Water quality index</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>WQPs</td>
<td>Water quality parameters</td>
</tr>
<tr>
<td>Z</td>
<td>Normalized Test Statistic</td>
</tr>
<tr>
<td>Zn</td>
<td>Zinc</td>
</tr>
<tr>
<td>% RE</td>
<td>Percent residual error</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 General introduction

All living organisms are inherently dependent on the accessibility towards water in order for them to exist and live. During the pre-historical years, all access of water in the world was used directly from the natural ecosystem. However in the new millennium, the consumerisms of the water resources were much more different and it has attracted numerous attentions among mankind. The dramatic increase in public awareness and concern about water resources has promptly caused severe water sources degradation. Nowadays, the increment amounts of water have conveyed to the degradation of the natural freshwater ecosystem. In fact, the degradation of freshwater quality has now reached a point where shortage of pure water is likely limited to food production, ecosystem function, and urban water supply system (Jury and Vaux, 2007). It has also contributes to numerous global awareness, including human drinking water supply and species survival (Christer and Birgitta, 2008).

Due to the alteration of our natural water quality it has generated alarming issues concerning to the level of pollutants loading from anthropogenic input in the aquatic environment especially in the surface water, this problems are mostly consequences by the rapid industrialization, uncontrolled urbanization and rapid economic development surrounding the major cities (Naji et al., 2010). According to Tong and Chen (2002), land use types correlate with most water quality characteristic. Despite that Juahir et al. (2010a) also suggested that the changes of land use type in Langat River Basin is also
contribute to the decreasing of water quality in Langat River, Malaysia. Indeed, Klang River flows through major towns in Kuala Lumpur and Selangor which undergo various developments with explicit land use types. Therefore, the pollution in the Klang River is mainly introduced by the municipal infrastructure and intensive human activities resulted to the substantial hydrological deformation and environmental deterioration. The expansion of urban areas in Klang River Basin affects the river by the increase of pollution load into river system and changes of the surface water quality. The natural process factor such as weathering, precipitation, soil erosion and anthropogenic influence such as urbanization, agricultural and industrial activities often associated to the changes in the surface water quality (Vega et al., 1998; Singh et al., 2004; Kazi et al., 2009). Nonetheless, industrial and household waste which are discharge directly or through leakages in the sewage systems causes excessive pollution of surface and underground water (Akcay et al., 2003). Based on high concentration of biochemical oxygen demand (BOD), NH₃N and suspended solids (SS), polluted rivers was contributed by sewage and discharge from agro-based and manufacturing industries, livestock farming and earthworks and land clearing activities (DOE, 2006).

Indeed, climate corresponded with the river water as it gives significant effect to the surface runoff that enters into the river bodies which known as seasonal variation (Singh et al., 2004). Furthermore, high flows are also vital for reducing level of various elements in rivers, by downstream transport, retention or emission for example de-nitrification that requires inundation of soils and oxygen depletion to occur (Forshay and Stanley, 2005; Pinay et al., 2007). Nevertheless, urban development also substantially owing to the modification on flood runoff and water quality (Tong and Chen, 2002).
Other than that Costa et al. (2003), also found that conversion of vegetation will disrupt the hydrological cycle of a drainage basin as it will altered the balance between rainfall and evaporation of the area.

Urban rivers are also polluted with discharge from sewage treatment plants where the overflowing sewage due to the rainfall (Nix and Merry, 1990) causing fecal contamination which is a major concern in the river especially in the city (Miyabara et al., 1994). However, lack of studies prone to the aspect of flow variation interaction with the physicochemical characteristics of the water, especially when these are affected by waste discharge and non-point sources pollution (Palmer et al., 2005).

In order to sustain the rivers water in Malaysia, monitoring programs that consist of frequent collection of water and determination of physiochemical parameters will representatively provide the status of the surface water quality. Since 1978, DOE Malaysia has performed continuous monitoring action however in 1995 the responsibility have been given to Alam Sekitar Malaysia Sdn Bhd (ASMA) under a privatisation arrangement which is resulting to a larger data matrices. In order to interpret the huge and complex data matrices of physico-chemical parameters, application of accurate method to conduct the data analysis need to be used (Chapman, 1992; Dixson and Chiswell, 1996).

Initially, the program covered all the river basin in Malaysia, involving mainly manual sampling and in-situ measurements of the river water samples. Sampling networks for monitoring the water quality of rivers constitute a great source of data for obtaining a local and temporal vision of the river state. Besides that, a wide network gives an insight into the evolution of the river ecosystem over time (Bouza-Deano et al., 2008). According to the DOE (1997), 143 river basins in Malaysia involved in this program in
order to monitor the river water quality changes on a daily basis involved a total of 1,064 stations. DOE found that 60% from the total stations were clean, 35% are slightly polluted and 5% are polluted. DOE has established thirty water quality stations along Klang River Basin and extended their jurisdiction in order to identify the pollution sources. The low dissolved oxygen levels were recognized in Klang River about 47.3% saturation and high ammonium levels were 7.8 mg/L. The report also stated that Klang River and its tributaries' were categorized as polluted and slightly polluted.

There are a lot of advance statistical tools for trend, temporal and spatial variation investigation (Vega *et al*., 1998; Bordalo *et al*., 2001; Chang *et al*., 2001; Liou *et al*., 2003; Juahir *et al*., 2010c). The rank-based non-parametric Mann-Kendall statistical test (Kendall, 1975; Mann, 1945) has been widely used to assess the significant of trends in hydro-meteorological time series such as streamflow, rainfall, and temperature and water quality. The non-parametric statistical tests are more suitable for non-normally distributed data or random and identically distributed data which is frequently used in hydro-meteorological time series. In this study, the changes of discharge, direct runoff and rainfall intensity were measured based on the changes in trend. Non-parametric trend analysis was carried out to investigate if there is any significant relation between the changes hydrology variables and water quality parameters. In addition, this assessment also requires application of multivariate analysis and artificial intelligence for exceptional data illustration.

Therefore, the application of different multivariate statistical technique such as cluster analysis (CA), DA, PCA and factor analysis (FA) can assists in interpretation of complex data matrices to in depth understand the river water quality and its ecosystem status. In fact, multivariate statistical analysis have been effectively employed to
evaluate the temporal and spatial characteristics of surface and freshwater quality that affected by seasonality (Yeung, 1999; Helena et al., 2000; Singh et al., 2004; Singh et al., 2005; Kuppusamy and Grirdhar, 2006; Zhou et al., 2007). Other than that, ANN have been widely used in water resources management problem and modeling (Garcia and Shigidi, 2006; Nikolos et al., 2008). Therefore, this study implement ANN to determine the most significant hydrological and water quality variable that can give scientific effect to the surface water quality system. ANN is universal estimators of non-linear mapping which able to learn and generalize relationship between input and output data from examples (training data) (Dayoff, 1990). Moreover, with the addition of pattern recognition and pattern classification this program also can be use in predictive purposes (Garcia and Shigidi, 2006).

Generally, water quality refers to the characteristics of water whether physical, chemical or biological characteristic. Based on the water quality data, the water quality index (WQI) was developed to evaluate the water quality status and river classification in Malaysia. WQI provides prediction in changes and trends in the water quality by considering multiple parameters. WQI is formed by six selected water quality variables namely as DO, BOD, chemical oxygen demand (COD), SS, NH\textsubscript{3}N and pH (DOE, 1997). WQI values ranges from 0-100. The values in the range of 81-100 are considered as clean. Whereas, the value ranges from 60-80 and 0-59 are classified as slightly polluted and polluted area respectively. In this study a total number of thirty water quality parameters (WQPs) used to define water quality namely DO, BOD, COD, SS, NH\textsubscript{3}N, pH, TEMP, COND, SAL, turbidity (TUR), dissolved solid (DS), TS, nitrate (NO\textsubscript{3}⁻), Cl⁻, PO\textsubscript{4}³⁻, As, Zn, chromium (Cr), lead (Pb), calcium (Ca), Fe, potassium (K), Mg, sodium
(Na), OG, MBAS, *Escherichia coli* (*E. coli*), coliform and two hydrological parameter namely as discharge and direct runoff.

### 1.2 Problem statements

The Klang Valley conurbation is considered the most developed and fastest growing region in the country with the highest rate of urban growth. Relative to the rapid development and expansion of city limits have affected environmental quality especially water pollution. The increment amounts of water usage have endured serious transformation to the water management that brings degradation of the natural freshwater ecosystem. The water pollution becomes a critical issue in due to the natural quality of the water courses that has been altered by the impact of various human activities and water uses. Particularly, the runoff from land carries various degree of residues into the river system as a cause of the rainfall which known as non-point sources pollution (Tong and Chen, 2002). Periods of extreme flow by the rainfall, may induce serious hazards to the water quality as it will increase the entry of various pollutants and the collection of effluent maybe targeted to larger areas (Chang and Carlson, 2005). The DOE report also stated that Klang River and its tributaries’ were categorized as polluted and slightly polluted. Even though DOE/DID have regular monitoring programs and historical hydrological/WQP data contains plentiful amount of data however there is still lack of data analysis which requires advance tools to extract all possible information from the river due to the challenge that need to be face in interpreting the data.
1.3 Significant of study

Assessment of these trend changes is crucial not only to understand and review past water pollution scenarios, but also to predict trend changes that will occur, formulate policies and strategies, as well as control the water management system. Observations of historical environmental and hydrological data are generally used for planning and designing the environmental study. Other than that, it provide the identification of possible sources that influence water quality and offers a useful tool for decent management of water resources and facilitate to a better solution in finding the pollution sources (Vega et al., 1998; Lee et al., 2001; Wunderlin et al., 2001; Simeonov et al., 2003; Simeonov et al., 2004; Juahir et al., 2010c). In spite that, mankind also encounter series of hardship in manipulating, redesigning channel structures and flows as the impact will eventually lead to the changes in fluctuation of the physicochemical variables. Therefore in this study, association with the hydrological data such as rainfall and water discharge (stream flow) provided by the DID, will facilitate in the investigation of hydrological variables that are significantly affect the river water. This study will be as a baseline data and information for future reference to investigate the pollution trends.
1.4 Research objective

The objectives of this study are:

1. To interpret the relationship between hydrological trend and surface water quality in Klang River system.
2. To develop an application with and without hydrological trend in estimating the pollution loading using multivariate analysis
3. To ascertain the better set of input and output parameters of the water quality at Klang River using ANN.

1.5 Research hypothesis

The hypotheses of this study are:

1. There are strong relationship between hydrological trend and surface water quality.
2. Hydrological trend influence in estimating pollution loading using multivariate analysis
3. ANN come out with better set of input and output parameters of the water quality at Klang River.
REFERENCES


Naji, A., Ismail, A. and Ismail, A. R. 2010. Chemical speciation and contaminatio
assessment of Zn and Cd sequential extraction in surface sediment of Klang

Nakhla, G., Al-Sabawi, M., Bassi, A. and, Liu, V. 2003. Anaerobic treatability of high oil

Neal, C., Christophersen, N., Neale, R., Smith, C. J., Whitehead, P. G. and Reynolds, B.
1988. Chloride in precipitation and streamwater for the upland catchment of

2003. An inventory of heavy metals inputs to agricultural soils in England and

neural networks as an alternative approach to groundwater numerical modelling

challenge to adaptive catchment management*. 13(2): 18 [online] URL:

Nix, P. G. and Merry, C. H. 1990. Use of sediment bags as a monitor of fecal pollution in

biochemical mechanisms. *Toxicology* 73: 127-146.

framework development for predicting the longitudinal dispersion coefficient in
natural streams using artificial neural network. *Environmental Progress and

Noori, R., Karbassi, A., Farokhnia, A. and Dehghani, M. 2009. Predicting the
longitudinal dispersion coefficient using support vector machine and adaptive
neuro-fuzzy inference system techniques. *Environmental Engineering Science*
26(10): 1503-1510.

consequences from environmental restoration at the Iron Mountain Super fund
site, California*. Proceedings of the National Academy of Sciences: Geology,
mineralogy and human welfare, Arnold and Mabel Beckman Center in Irvine,
CA, pp. 3455-3462.


