# MARKOV CHAIN ANALYSIS TO DETECT WATER QUALITY LEVEL

NURUL NABIHAH BINTI RAHMAN

UNIVERSITI TEKNOLOGI MALAYSIA

MARKOV CHAIN ANALYSIS TO DETECT
WATER QUALITY LEVEL

NURUL NABIHAH BINTI RAHMAN

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Engineering Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2013

Specially dedicated to

my beloved mother, Pn. Hjh. Aminah Adam, my siblings and my dearest one.

Millions of thankful to my supervisor, Dr. Fadhilah Yusof and those

who supported me in any respect during the completion of the thesis.

## ACKNOWLEDGEMENT

First and foremost, I would like to grasp hold of this golden opportunity to express my very deepest gratitude and warmest appreciation to my supervisor, Dr. Fadhilah Yusof for her valuable and constructive suggestions during the planning and development of this thesis. All of her encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of this thesis. Her willingness to give her time so generously has been very much appreciated.

On top of that, I would like to express my appreciation to the parties concerned from IPASA, Faculty of Civil Engineering, UTM for their cooperation and kind assistance by providing the data needed in my project. Without the data given, I might not complete my thesis.

Not forgotten to my family and my friends for supported me and encouraged me to complete this thesis so that I will not procrastinate in doing it.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of this thesis and above all, to the Almighty God, who never cease in loving me and for the continued guidance and protection.

## ABSTRACT

The quantificational analysis on progress of river water quality is important in knowing the dynamic change of water quality level. A mathematical model based on Markov chain is established in order to detect the water quality level in rivers. In this study, the level changes of water quality of River A, River B, River C and River D in 2010 based on water quality parameters of DO, BOD, COD, SS, pH and AN will be determined using Markov chain model. In order to determine the water quality level in rivers, the developing of the Markov chain model has to be conducted. There are three main steps in developing this model. The steps are establishing the transition probability matrix, calculate degree of absolute progress (DAP) and degree of relative progress (DRP). Based on the results of Markov chain model, the DRP value of River A, River B, River C and River D are: $\text{DRP}_A$= 0.041, $\text{DRP}_B$= -0.031, $\text{DRP}_C$= 0.047 and $\text{DRP}_D$= -0.014, respectively. The DRP values for the rivers are in the sequence of ascending value of $\text{DRP}_B < \text{DRP}_D < \text{DRP}_A < \text{DRP}_C$. If the orders of water quality in rivers are arranged from the most deteriorated to the most improved, it will start from River B followed by River D, River A and River C. In other words, River B has the least improvement of changes among the rivers meanwhile River C has the most improvement of changes. After the Markov chain analysis to detect the water quality level has been done, the Water Quality Index (WQI) method is applied next in order to do justification of the Markov chain results. Surprisingly, the Markov chain model results match very well with the WQI method results when comparisons for both methods are made. To sum up, the results from Markov chain analysis can be justified by using the WQI method.

# ABSTRAK

Analisis quantifikasional mengenai kemajuan kualiti air sungai adalah penting dalam mengetahui perubahan dinamik tahap kualiti air. Satu model matematik berdasarkan rantai Markov ditubuhkan untuk mengesan tahap kualiti air di sungai. Dalam kajian ini, perubahan tahap kualiti air di Sungai A, Sungai B, Sungai C dan Sungai D pada tahun 2010 berdasarkan parameter kualiti air DO, BOD, COD, SS, pH dan AN akan ditentukan dengan menggunakan model rantaian Markov. Dalam usaha untuk menentukan tahap kualiti air di sungai-sungai, pemodelan rantai Markov perlu dilaksanakan. Terdapat tiga langkah utama dalam membangunkan model ini. Langkah-langkah tersebut adalah mewujudkan matriks kebarangkalian peralihan, mengira tahap kemajuan mutlak (DAP) dan ijazah kemajuan relatif (DRP). Berdasarkan keputusan model rantai Markov, nilai DRP Sungai A, Sungai B, Sungai C dan Sungai D adalah seperti berikut: $\text{DRP}_A$= 0.041, $\text{DRP}_B$= -0.031, $\text{DRP}_C$= 0.047 and $\text{DRP}_D$= -0.01. Dalam turutan menaik, nilai DRP utk setiap sungai adalah $\text{DRP}_B < \text{DRP}_D < \text{DRP}_A < \text{DRP}_C$. Jika kita mengatur susunan kualiti air di sungai-sungai dari yang paling merosot kepada yang paling baik, ia akan bermula dari Sungai B diikuti oleh Sungai D, Sungai A dan Sungai C. Dalam erti kata lain, kita boleh mengatakan bahawa Sungai B mempunyai peningkatan yang paling sedikit berbanding sungai-sungai yang lain sementara Sungai C mempunyai peningkatan yang paling ketara. Selepas analisis rantai Markov untuk mengesan tahap kualiti air telah dilakukan, kaedah Indeks Kualiti Air (WQI) digunakan seterusnya untuk melakukan justifikasi keputusan rantai Markov. Apabila perbandingan bagi kedua-dua kaedah dibuat, keputusan model rantaian Markov didapati sangat sepadan dengan keputusan daripada kaedah WQI. Kesimpulannya, keputusan daripada analisis rantai Markov telah berjaya dijustifikasikan dengan menggunakan kaedah WQI.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| DOE | - | Department of Environment |
| WQI | - | Water Quality Index |
| DO | - | Dissolved Oxygen |
| BOD | - | Biochemical Oxygen Demand |
| COD | - | Chemical Oxygen Demand |
| SS | - | Suspended Solids |
| AN | - | Ammoniacal Nitrogen |
| DAP | - | Degree of Absolute Progress |
| DRP | - | Degree of Relative Progress |
| SI | - | Sub-index |
| SIDO | - | Sub-index of Dissolved Oxygen |
| SIBOD | - | Sub-index of Biochemical Oxygen Demand |
| SICOD | - | Sub-index of Chemical Oxygen Demand |
| SIAN | - | Sub-index of Ammoniacal Nitrogen |
| SISS | - | Sub-index of Suspended Solids |
| SIpH | - | Sub-index of pH |

## LIST OF SYMBOLS

| | | |
|---|---|---|
| *mg/l* | - | milligrams per liter |
| $\mu$ | - | mean |
| $t$ | - | time |
| $x$ | - | random variable |
| $t$ | - | index set |
| $Y_0$ ,$Y_1$, …, $Y_{t-2}$ | - | past states |
| $Y_{t-1}$ | - | present state |
| $Y_t$ | - | state of the system at time *t* |
| $i$ | - | present level of water quality |
| $j$ | - | future level water quality |
| $p_{ij}$ | - | transition probability of Markov chain |
| $e$ | - | set of water quality level |
| $k$ | - | water quality parameters in rivers |
| $n$ | - | water quality parameters samples |
| $M_i$ | - | sum of samples that have level *i water* quality |
| $m_i$ | - | number of samples that have level *i water* quality |
| $n_{ij}$ | - | sum of the changes from *i water* quality level at present state to change to *j* water quality level at future state |
| $n_{ij}$ /$M_i$ | - | transition probability of which the water quality changed from *i* level to *j* level in rivers |
| $p$ | - | transition probability matrix of the rivers |
| $w$ | - | weight |
| $S$ | - | advancement matrix of transition probability matrix |

## LIST OF APPENDICES

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

The water quality status of rivers in Malaysia becomes a major issue nowadays. Rivers play an important role in human lives. Unfortunately, many of them have forgotten about the importance of rivers and this has lead to serious degradation in the rivers today. When a substance which degrades the water quality enters the waterway, it will result a river to be polluted since the substance will alters its natural functions.

Rivers are crucial as they serve many benefits to the human being and aquatic life. Here are the importances of rivers. First is agriculture. Rivers act as a backbone in supporting agricultural development and productivity in Malaysia. Irrigation is employed primarily to cater for the double cropping of paddy. Next, domestic and industrial water supply. In Malaysia, majority the water supply comes from rivers. The country's economic shift towards industrialization in the 1980s, coupled with an increase in population and urban growth has boost up the demands for residential and industrial water. Then, sanctuary for biodiversity. Malaysian rivers are endowed with rich plant and animal life. They manage in support and sustaining biodiversity in the surrounding environment.

Moreover, rivers might serve as transportation. Some remote villages in Sabah and Sarawak for instance, are still depending on rivers for transportation. Besides that, rivers can be a place for recreation. Malaysian rivers have high recreation value, which provide many opportunities for outdoor activities. They provide travel routes for exploration, commerce and recreation. Finally, rivers are also crucial for generating hydropower. Hydroelectric power output constituted between 7% -12% of the total national energy generated annually for the period from 1996 to 2000. It was estimated that 85% of the total hydropower potential of the country is found in the states of Sabah and Sarawak.

The qualities of rivers are very important since water play a major role in supporting all aspects of human, ecosystem survival and health. According to Heathcote (1998), almost all water uses can be affected by water quality directly. All the physical, chemical, biological, and microbiological conditions that exist in the watercourse and the subsurface aquifer maybe affected by activities such as fish survival, diversification and growth, recreational, agricultural, industrial and commercial sectors. Conflict in the watershed is often triggered by deteriorating water quality problems. This is due to doubt whether the quality of water used are safe or not to use.

In Malaysia, The Department of Environment (DOE ) has been conducting monitoring of river since 1978, primarily to establish baselines and to detect water quality changes in river water quality and has since been extended to identifying of pollution sources as well.

A total of 1064 water quality monitoring stations located in the 143 river basins were monitored in 2007. Out of this monitoring station 1064, 638 (60%) were found to be clean, 376 (35%) slightly polluted and 50 (5%) contamination. Compared with 2006, the year 2007 has recorded a significant improvement in water quality. Number of clean river basin is 91 (64%) compared with 80 in 2006 and number of slightly polluted river basin decreased from 59 in 2006 to 45 (31%) in 2007. However, the number of polluted river basins remains in 7 (5%). Biochemical Oxygen Demand (BOD), Ammoniacal Nitrogen (AN) and Suspended Solids (SS) are the main pollutants that have been identified as the year before. Sewage discharges of untreated or partially treated from agricultural and manufacturing industry has been attributed to the high BOD. Domestic sewage and livestock are the main source for AN while most of earthworks and land clearing activities come from SS, DOE (2010).

Malaysia is a developing country that moves towards vision 2020. Unfortunately the development that had been carried throughout the country also contributes bad impact to the environment especially water quality. This scenario has become extremely bad that will not only affects human health but also the entire environment. Due to the unpleasant development, the aquatic communities in river also live in danger situation. Therefore, the authorities are responsible to highlight this issue of environmental impact caused by development in a serious manner. By doing that, any negative impact to the environment can be minimized.

## 1.2 Background of the Study

Detailed studies related to the level of water quality is important for detect the changes in the river. Changes in water quality can provide guideline in determine whether the water quality improved or deteriorated. River is one of the important elements in human life in order to carry out their daily activities such as agriculture, domestic and industrial supply and many more. Hence, their increase or decrease of the water quality level should be given full attention.

Water quality data of four rivers in Johor in 2010 are analyzed to detect the water quality level. Since the data are private and confidential thus the name of the rivers involve in this study cannot be reveal. In this study, four rivers namely as River A, River B, River C and River D in 2010 are studied to detect the water quality level. Generally River A is situated nearby residential and squatter areas meanwhile River B involves more in urbanization and uncontrolled discharge waste. On top of that, manufacturing industry and dumping rubbish are being focused at River C while River D is relying more on silt from earthworks and land and agriculture sector. Determining of water quality level changes in these rivers need to be monitored continuously since the rivers contribute huge impact to the society.

In order to better understand and manage the river, it is often helpful to use a water quality model. In this study, a special kind of stochastic process, called a Markov chain, where the outcome of an experiment depends only on the outcome of the previous experiment is studied. In other words, the next state of the system depends only on the present state, not on preceding states. The aim of this study is to detect the water quality level in River A, River B, River C and River D in 2010 by using Markov chain with justification with Water Quality Index (WQI) method provided by Department of Environment (DOE), Malaysia.

The Water Quality Index (WQI) is a method employed to classify the river water quality monitored. Currently, in Malaysia, water quality data was used to classify the rivers into Class I, II, III, IV or V based on Water Quality Index (WQI). The WQI defined five classes (I, II, III, IV and V) for classification of rivers based on the descending order of water quality vis-a-vis Class I being the “best” and Class V being the “worst”. Water Quality Index (WQI) is computed based on water quality parameters. There are Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH, Suspended Solids (SS) and Ammoniacal Nitrogen (AN).

From a scientific perspective though, it is compulsory to know the dynamic change of water quality level. Therefore, to avoid this problem to become worst in the future, the quantificational analysis on progress of river water quality must be conducted so that a mathematical model based on Markov chain to detect the river water quality level of rivers can be developed.

## 1.3 Statement of the Problem

Nowadays, Malaysia is facing a lot of environmental issues where the issue that has been most concern is the river water quality level. Therefore, the move was essential to prevent the rivers from being further polluted as well as to tackle the water quality issues in rivers in the best way.

Stochastic process based on Markov chain is simply a probability process that is any process in nature whose evolution can be analyze successfully in terms of probability. On the basis of the results of river water quality evaluation, a quantitative value is needed to figure out how the river water quality level changed, to check whether it is improved and to compare the level changes in different rivers. Thus, a mathematical model based on Markov chain is established to detect the water quality level in rivers.

In this study, the water quality level of River A, River B, River C and River D in 2010 are going to be determined using Markov chain model. This is followed by comparing the level changes of water quality between the rivers. The comparisons between each river are crucial in order to know which of the river that has greatest improvement and least improvement. Thus, the deteriorated rivers can be focused more in future. Even the Markov chain analysis to detect changes in water quality level has been done, the results of Markov chain model are needed to be justified in order for the model to be a robust model to be implemented in future study. Thus, Water Quality Index (WQI) method is chosen for justifying the Markov chain results. From this view, the results from both methods are expecting to match very well with each other.

## 1.4 Objectives of the Study

The main objectives of this study are:

i. To determine the water quality level of rivers in 2010 using Markov chain model.
ii. To compare the level changes of water quality between rivers in 2010.
iii. To justify the Markov chain result analysis of rivers in 2010 using Water Quality Index (WQI) method.

## 1.5 Scope of the Study

In this study, the water quality level for the four rivers will be detected using Markov chain model. The data were obtained from IPASA, Faculty of Civil

Engineering, UTM. Since the data taken are strictly confidential, in this study rivers involved are namely as River A, River B, River C and River D.

The data that are going to be used in this study are water quality data of River A, River B, River C and River D in 2010 based on water quality parameters involve. The water quality parameters are Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), pH and Ammoniacal Nitrogen (AN) with a scalar value. The water qualities data obtained are having same duration of time but in different types of river (River A, River B, River C and River D).

In detecting the water quality level of River A, River B, River C and River D in 2010 using Markov chain model, two concepts were used: degree of absolute progress (DAP) and degree of relative progress (DRP). For justification of Markov chain result, the Water Quality Index (WQI) provided by Department of Environment (DOE), Malaysia has been used.

## 1.6 Significance of the Study

River plays a crucial role in human life. This is because the river is the main source of water supply in their daily activities for example; source for drinking, watering plants, agriculture and so on. On top of that, river also can be a place for recreational and tourism area where it helps to boost the economy growth of the country. All parties should be responsible on how to give more attention in order to protect the rivers from being further polluted.

The changes in water quality can provide guideline in determine whether the water quality improved or deteriorated. If the level of water quality is at a satisfactory level, periodical monitoring should be carried out so that the water quality can be maintained. However, more frequent monitoring should be carried out if the level of water quality in the river has become more deteriorated. The move was essential to prevent the rivers from being further polluted as well as to tackle the water quality issues in rivers in the best way.

In this study, the Markov chain model is studied in order to detect the water quality level of River A, River B, River C and River D in 2010. Since some of the rivers are having tendency to become more polluted in future, thus the authorities are responsible to highlight this issue of environmental impact caused by development in a serious manner.

To preserve and protect the river and its environment from further deterioration, it may become necessary to raise awareness in the community about the importance of rivers and the role it played. Various ways can be done to increase awareness of the importance of rivers in daily life. Introducing to the public measures and practices that may contribute to the preservation of the river can be implemented by organizing the awareness campaigns through print and electronic media. By doing that, any negative impact to the environment can be minimized.

## 1.7 Organization of the Study

In this study, the work will be organized into five main chapters. Chapter 1 presents the introduction of the study. It includes the background, statement of the problem, objectives, scope, significance and organization of the study.

Chapter 2 discusses the literature that has been reviewed for the current study. Some studies that used Markov Chain models are presented here and followed by studies done using Water Quality Index.

The research methodology is detailed out in Chapter 3. This chapter will begin with the introduction of the water quality parameters used, which are Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), pH and Ammoniacal Nitrogen (AN). The descriptive statistics formula is to be introduced next. The chapter then continues by introducing the descriptive statistics formula, stochastic process and Markov chain. This is followed by data screening procedure. The calculation of the transition probability matrix, degree of absolute progress (DAP) and degree of relative progress (DRP) for developing Markov chain model will also be presented. Finally, the Water Quality Index (WQI) method is also to be discussed in order to justify the Markov chain analysis result.

Results and discussions have been discussed in details in Chapter 4. In this chapter, the descriptive statistics of the water quality data based on water quality parameters are to be discussed. This is followed by data screening before classifying the water quality parameters of River A, River B, River C and River D in 2010 from the scalar value to water quality class of Class I, II, III, IV and IV by referring to "Class Based of DOE Water Quality Index Classification, DOE (2005)". The calculation of the transition probability matrix, degree of absolute progress (DAP) and degree of relative progress (DRP) for developing the Markov chain model will then be done. Finally, the Markov chain results analysis are justified further by the Water Quality Index (WQI) method.

Finally, conclusion and recommendation for the future researchers in Chapter 5 are made. This chapter summarizes the study and makes thoroughly conclusions based on the results and discussions from previous chapter. Suggestions for further research are also recommended in this section.

# REFERENCES


Adam. M. J. Markov Chain Monte Carlo. In E. Baker, P. Peterson and B. McGraw, editors, *International Encyclopedia of Education*. Elsevier, 3rd edition, 2010.

Ali H. Ahmed Suliman. (2010).*Using Hec-Ras and Qual2e to Assess Johor River Water Quality*. Faculty of Civil Engineering, Universiti Teknologi Malaysia. Project report for Master degree of Engineering.

Avvannavar, S. M., and Shrihari, S. (2007). *Evaluation of water quality index for drinking purposes for river Netravathi*, Mangalore, South India. Environmental Monitoring and Assessment.

Classes in Malaysian Water Quality Index (2005). Department of Environment Malaysia (DOE). National Water Quality Standards for Malaysia. Retrieved on September 27, 2012.

Davis, A. P. & McCuen, R. H. (2005). *Storm water management for smart growth*. 1st edition. Springer Science and Business Media.

Feng W., and Zou Z. (2007). Dynamic evaluation for water quality of rivers based on Markov process. Techniques and Equipment for Environmental Pollution. 1 (8):132-135.

Han P., Wang P. J., Sun W. and Zhu, D. (2008). Drought forecasting based on the vegetation temperature condition index using weighted Markov model. *Agricultural research in the arid areas,* vol.26, pp.196-206.

Heathcote, I.W. (1998). *Integrated Watershed Management: Principles and Practice.* John Wiley & Sons, Inc. New York.

He B., Chen C., and Gao D. (2003). Markov Method of Dynamic assessment on Water quality Environment Engineering, pp.60-62.

Jun-e L., Fengping A., and Zhanglin G.(2009). The Water Quality Evaluation based on the unascertained Markov Forecast Model. *Chinese Control and Decision Conference.*

Khuan, L. Y., Noraliza Hamzah, and Rozita Jailani. (2002). *Prediction of water quality index (WQI) based on artificial neutral network.* Student Conference on Research and Development Proceedings, Shah Alam, Malaysia.

Lohani, V. K. and Loganathan, G. V. (1997). An early warning system for drought management.

Ma H., L Liu. and T Chen. (2008). Assessment Model Based on Markov Chain. *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, 602-605.

Norris, J. R. (1997). *Cambridge Series In Statistical and Probabilistic Mathematics: Markov Chains.* United Kingdom: Cambridge University Press.

River Water Quality Status (2010). Department of Environment Malaysia (DOE). Water Marine River Water. Retrieved on June 22, 2012.

Sarkar, C., and Abbasi, S. A. (2006). *Qualidex – A new software for generating water quality indice*. Environmental Monitoring and Assessment. 119, 201–231.

Sarukkai, R. R. (2001). *Link Prediction and Path Analysis Using Markov Chains.* 9th International World Wide Web Conference (WWW9) Refereed Papers.

Sato, R. C. and  Zouain, D. M. (2010). Markov Models in health care. *Einstein*, 8(3 Pt 1):376-9.

Sun C., Zhang G. and Lin X. (2003). Model of Markov chain with weights and its application in predicting the precipitation state. *Theory and practice of system engineering*, vol. 4, pp.100-105.

Swaminathan (2011). *Practitioner's Handbook on the Modelling of Dynamic Change in Ecosystems.* Chennai, India: M S Swaminathan Research Foundation. using the Palmer drought index. *Journal of the American Water Resources Association,* vol. 33, pp. 1375-1386.

Washington State Department of Ecology (2002). Introduction to water quality index. Retrieved on September 27, 2012.

Yun Y., Zou Z., Feng W., and Ru M. (2009). *Quantificational Analysis On Progress Of River Water Quality In China.* Journal of Environmental Sciences. 21, 770–773.