

Prediction Model of Water Quality in Chaophraya River using Artificial Neural Network

Mr. Weeris Treeratanajaru



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science and
Information Technology

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2017

Copyright of Chulalongkorn University



5573106823

แบบจำลองการทำนายคุณภาพน้ำในแม่น้ำเจ้าพระยา โดยใช้โครงข่ายประสาทเทียม


นายวีริศ ตริรัตน์จาร์




วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการ
คอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2560
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

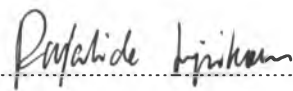
Thesis Title Prediction Model of Water Quality in Chaophraya
River using Artificial Neural Network
By Mr. Weeris Treeratanajaru
Field of Study Computer Science and Information Technology
Thesis Advisor Associate Professor Rajalida Lipikorn, Ph.D.
Thesis Co-Advisor Supawin Watcharamul, Ph.D.

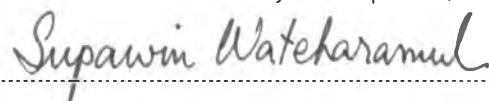
Accepted by the Faculty of Science, Chulalongkorn University in Partial
Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Faculty of Science
(Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

.....Chairman
(Associate Professor Nagul Cooharajanone, Ph.D.)

.....Thesis Advisor
(Associate Professor Rajalida Lipikorn, Ph.D.)

.....Thesis Co-Advisor
(Supawin Watcharamul, Ph.D.)

.....Examiner
(Assistant Professor Pattarasinee Bhattarakosol, Ph.D.)

.....Examiner
(Assistant Professor Roongkan Nuisin, Ph.D.)

.....External Examiner
(Chao Nokyoo, Ph.D.)

วีริศ ตรีรัตนจารุ : แบบจำลองการทำนายคุณภาพน้ำในแม่น้ำเจ้าพระยา โดยใช้โครงข่ายประสาทเทียม (Prediction Model of Water Quality in Chaophraya River using Artificial Neural Network) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. รัชลิดา ลิปิกรณ์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: อ. ดร. ศุภวิน วัชรมูล, 158 หน้า.

แบบจำลองเพื่อทำนายคุณภาพน้ำได้รับการวิจัย พัฒนาและประยุกต์ใช้กับแหล่งน้ำต่างๆ ทั่วโลก ลักษณะของแหล่งน้ำที่แตกต่างกันทำให้แบบจำลองที่เหมาะสมสำหรับแหล่งน้ำแต่ละแห่งมีความเฉพาะเจาะจง วิทยานิพนธ์ฉบับนี้นำเสนอแบบจำลองใหม่เพื่อพยากรณ์พารามิเตอร์คุณภาพน้ำของแม่น้ำเจ้าพระยาในอนาคต แบบจำลองใหม่นี้พัฒนาขึ้นจากแบบจำลองโครงข่ายประสาทเทียม โดยพัฒนาให้สามารถรับข้อมูลแบบหลายมิติได้ (พารามิเตอร์ที่แสดงคุณภาพน้ำที่ย้อนไปในอดีตและต้นน้ำ) ที่เรียกว่าโครงข่ายประสาทเทียมสถานที่และเวลา ซึ่งแตกต่างจากโครงข่ายประสาทเทียมแบบดั้งเดิม ทั้งนี้แบบจำลองที่ทันสมัยที่สุดในขณะนี้จำนวนหนึ่งถูกนำมาทดสอบและเปรียบเทียบกับแบบจำลองที่ถูกรับเสนอ โดยใช้ข้อมูลคุณภาพน้ำของแม่น้ำเจ้าพระยาในช่วง 17 ปี เมื่อเปรียบเทียบกับค่าสัมประสิทธิ์สหสัมพันธ์ของการทำนายค่าพารามิเตอร์พบว่า แบบจำลองใหม่ทำนายได้แม่นยำกว่า (Spearman's rho = 0.73 ± 0.06) เมื่อเปรียบเทียบกับแบบจำลองอื่นๆ (Spearman's rho = 0.67 ± 0.08 และ Spearman's rho = 0.57 ± 0.15) นอกจากนี้แบบจำลองใหม่นี้ถูกพัฒนาขึ้นเป็นกรอบกว้างๆ เพื่อให้สามารถนำไปประยุกต์กับแม่น้ำสายอื่นๆได้อีกด้วย



| | | |
|----------|---|----------------------------------|
| ภาควิชา | คณิตศาสตร์และวิทยาการคอมพิวเตอร์ | ลายมือชื่อ นิสิต |
| | | ลายมือชื่อ อ.ที่ปรึกษาหลัก |
| สาขาวิชา | วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ | ลายมือชื่อ อ.ที่ปรึกษาร่วม |

ปีการศึกษา 2560

5573106823 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

KEYWORDS: WATER QUALITY PREDICTION MODEL / ARTIFICIAL NEURAL NETWORK / SPACE AND TIME NEURAL NETWORK

WEERIS TREERATANAJARU: Prediction Model of Water Quality in Chaophraya River using Artificial Neural Network. ADVISOR: ASSOC. PROF. RAJALIDA LIPIKORN, Ph.D., CO-ADVISOR: SUPAWIN WATCHARAMUL, Ph.D., 158 pp.

Water quality prediction models have been researched, developed and applied to various water resources around the world. The characteristics of different water resources make the model to be suitable for each specific resource. This study proposes the new model to forecast the water quality parameters of Chaophraya River in the future that was developed from artificial neural network (ANN). Unlike the traditional ANN, the proposed model called space and time neural network (STNN) is able to accept input multi-dimensional data (historical and upstream water quality records). The state-of-the-art models and the purposed model were tested and compared using the Chaophraya River's water quality measured over a period of 17 years. The STNN model outperforms the others in term of water quality prediction correlation coefficient (Spearman's rho = 0.73 ± 0.06) compared with other models (Spearman's rho = 0.67 ± 0.08 and Spearman's rho = 0.57 ± 0.15). In addition, the proposed model was developed with a general framework and could be applied to other rivers as well.



| | | |
|-----------------|------------------------|------------------------------|
| Department: | Mathematics and | Student's Signature |
| | Computer Science | Advisor's Signature |
| Field of Study: | Computer Science and | Co-Advisor's Signature |
| | Information Technology | |
| Academic Year: | 2017 | |

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor and co-advisor, Associate Professor Dr. Rajalida Lipikorn, at Department of Mathematics and Computer Science, and Dr. Supawin Watcharamul, at Department of Environmental Science, Faculty of Science, Chulalongkorn University, for all their great support helping me to improve my researching skill, and giving me the great opportunity in publishing papers.

I would like to thank for data supported by Dr. Chao Nokyoo, at Pollution Control Department, Ministry of Natural Resources and Environment.

I would like to thank all committees for precious comments.

I would like to thank Miss Phasinee Khwanmuang for documentary support, friendly and helpful support in dissertation work.

I would like to thank Prof. Michael E. Houle, Oussama Chelly, Wasin Rojayaroon, Chadaphim Photphanloet, Banphatree Khomkham, Krisorn Chunhapongpipat for technical discussion and useful comments.

Finally I would like to thank my father, mother and friends for everything they suggested and supported me.



CONTENTS

| | Page |
|---|------|
| THAI ABSTRACT | iv |
| ENGLISH ABSTRACT | v |
| ACKNOWLEDGEMENTS | vi |
| CONTENTS | vii |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xiii |
| LIST OF ABBREVIATION..... | xvi |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Rationale..... | 1 |
| 1.2 Problem formulation..... | 1 |
| 1.3 Objectives..... | 2 |
| 1.4 Scope..... | 3 |
| 1.5 Research methodology..... | 3 |
| 1.6 Expected outcomes | 4 |
| CHAPTER 2 LITERATURE REVIEW..... | 5 |
| 2.1 Hydrologic model | 5 |
| 2.2 Empirical model..... | 6 |
| 2.2.1 Imputation..... | 7 |
| 2.2.2 Data transformation..... | 8 |
| 2.2.3 Normalization | 8 |
| 2.2.4 Parameter selection | 9 |
| 2.2.5 Prediction model | 23 |



| | Page |
|--|------|
| 2.2.6 Space and time prediction model | 24 |
| CHAPTER 3 THEORITICAL BACKGROUND | 26 |
| 3.1 Environmental science background | 26 |
| 3.1.1 Water quality parameter description | 26 |
| 3.1.2 Surface water quality standards in Thailand | 34 |
| 3.2 Computer science background | 40 |
| 3.2.1 Descriptive statistics | 40 |
| 3.2.2 Imputation techniques | 42 |
| 3.2.3 Data transformation method | 44 |
| 3.2.4 Normalization methods | 45 |
| 3.2.5 Parameter selection methods | 46 |
| 3.2.6 Prediction models | 50 |
| CHAPTER 4 METHODOLOGY | 52 |
| 4.1 Data preprocessing | 53 |
| 4.1.1 Data collection and descriptive statistics | 53 |
| 4.1.2 Imputation | 58 |
| 4.1.3 Data transformation | 60 |
| 4.1.4 Normalization | 61 |
| 4.2 Modelling | 62 |
| 4.2.1 Parameter selection | 62 |
| 4.2.2 Model comparison | 64 |
| 4.3 Purposed method: Space and time neural network | 66 |
| 4.3.1 Model construction | 67 |



| | Page |
|---|------|
| 4.3.2 Modelling..... | 69 |
| 4.3.3 Performance comparisons | 70 |
| CHAPTER 5 RESULTS | 72 |
| 5.1 Data descriptive statistics | 72 |
| 5.2 Imputation results..... | 78 |
| 5.3 Data transformation results..... | 79 |
| 5.4 Normalization results..... | 83 |
| 5.5 Parameter selection results | 83 |
| 5.6 Prediction models comparisons..... | 84 |
| 5.7 Space and time neural network results | 84 |
| CHAPTER 6 DISCUSSIONS | 92 |
| 6.1 Data descriptive statistics | 92 |
| 6.2 Imputation | 94 |
| 6.3 Data transformation | 95 |
| 6.4 Normalization | 95 |
| 6.5 Parameter selection..... | 97 |
| 6.6 Prediction models comparison | 98 |
| 6.7 Space and time neural network..... | 98 |
| CHAPTER 7 CONCLUSIONS..... | 109 |
| 7.1 Water quality prediction framework | 109 |
| 7.2 Space and time neural network model | 110 |
| 7.3 Future works..... | 111 |
| REFERENCES | 112 |

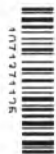


| | Page |
|--|------|
| APPENDIX A DISSERTATION PROPOSAL | 135 |
| APPENDIX B COMPLEMENTARY RESULT | 145 |
| B.1 Historical data chart..... | 145 |
| B.2 Pre-processing result | 153 |
| B.3 Parameter selection result..... | 154 |
| B.4 Selected parameter from proposed model | 156 |
| VITA..... | 158 |



LIST OF TABLES

| | |
|--|----|
| Table 2.1 Detail of reviewed water quality modelling research | 11 |
| Table 3.1 Classification and objectives of classification of water quality standard | 36 |
| Table 3.2 Classification of Chaophraya River | 37 |
| Table 3.3 Surface water quality standards | 37 |
| Table 3.4 Methods for Examination and statistic of each parameter | 39 |
| Table 4.1 Detail of monitoring stations along Chaophraya River | 54 |
| Table 4.2 List of water quality parameters in Chaophraya River, Thailand | 57 |
| Table 4.3 Three imputation methods argument setting | 60 |
| Table 4.4 Argument setting for parameter selection methods | 64 |
| Table 4.5 Argument setting for models | 65 |
| Table 5.1 Basic statistics of the water quality parameters in Chaophraya River, Thailand during 2538-2556 B.E..... | 73 |
| Table 5.2 Spearman correlation of monitoring year and water quality parameters | 74 |
| Table 5.3 Spearman correlation of monitoring month and water quality parameters | 74 |
| Table 5.4 Parameters and monitoring station relationship..... | 75 |
| Table 5.5 Spearman correlation of water temperature and other water quality parameters | 78 |
| Table 5.6 Three imputation methods performance evaluation | 79 |
| Table 5.7 Skewness of parameters at different λ value according to Osborne's transformation..... | 80 |
| Table 5.8 Performance comparison of transformed data and non-transformed data..... | 83 |



| | |
|--|-----|
| Table 5.9 Four normalization methods performance comparison | 83 |
| Table 5.10 Four parameter selection methods performance comparison | 84 |
| Table 5.11 SVM, ANN, MLR performance comparison | 84 |
| Table 5.12 Evaluation of model fits to EC observations | 85 |
| Table 5.13 Evaluation of model fits to TDS observations | 87 |
| Table 5.14 Evaluation of model fits to PO_4^{3-} observations..... | 88 |
| Table 5.15 Top 10 dimensional parameter importance on EC model. | 90 |
| Table 5.16 Top 10 dimensional parameter importance on TDS model..... | 90 |
| Table 5.17 Top 10 dimensional parameter importance on PO_4^{3-} model..... | 91 |
| Table 6.1 Normalization methods performance comparison show by individual model..... | 96 |
| Table 6.2 Top five important parameters on EC model..... | 107 |
| Table 6.3 Top five important parameters on TDS model..... | 107 |
| Table 6.4 Top five important parameters on PO_4^{3-} model..... | 108 |
| Table B.0.1 Three imputation methods performance evaluation show by individual model..... | 153 |
| Table B.0.2 Performance comparison of transformed data and non-transformed data on various models | 154 |
| Table B.0.3 Forward selection performance of various BOD model | 154 |
| Table B.0.4 Backward elimination performance of various BOD model | 155 |
| Table B.0.5 PCA performance of various BOD model..... | 155 |
| Table B.0.6 Genetic algorithm performance of various BOD model..... | 156 |

LIST OF FIGURES

| | |
|--|----|
| Figure 3.1 K-nn imputation algorithm pseudocode..... | 43 |
| Figure 3.2 Artificial neural network structure..... | 43 |
| Figure 4.1 Overview and flowcharts of methods used in this dissertation..... | 52 |
| Figure 4.2 Data collection and descriptive analysis flowchart..... | 53 |
| Figure 4.3 Chaophraya River and location of 19 monitoring stations along the river..... | 56 |
| Figure 4.4 Imputation experiment flowchart..... | 59 |
| Figure 4.5 Transformation experiment flowchart..... | 61 |
| Figure 4.6 Normalization experiment flowchart..... | 62 |
| Figure 4.7 Parameter selection experiment flowchart..... | 63 |
| Figure 4.8 Model comparison experiment flowchart..... | 66 |
| Figure 4.9 Multi-dimensional artificial neural network structure..... | 68 |
| Figure 4.10 Space and time neural network experiment flowchart..... | 71 |
| Figure 5.1 The example of missing value percentage of water quality parameters record of Chaophraya River during 2538-2556 B.E. (%)..... | 72 |
| Figure 5.2 Historical data of salinity from monitoring stations along Chaophraya River during 2538-2556 B.E. | 76 |
| Figure 5.3 Historical data of turbidity from monitoring stations along Chaophraya River during 2538-2556 B.E. | 77 |
| Figure 5.4 Relationship between λ and skewness of each parameter, a) and b) showed different pattern of relationship..... | 82 |
| Figure 5.5 EC model simulator snapshot from Rapid miner studio..... | 86 |
| Figure 5.6 TDS model simulator snapshot from Rapid miner studio..... | 87 |



| | |
|---|-----|
| Figure 5.7 PO ₄ ³⁻ model simulator snapshot from Rapid miner studio | 89 |
| Figure 6.1 Relationship between normalized observed EC and normalized predicted EC from three models | 99 |
| Figure 6.2 Comparisons of the models computed and measured EC in Chaophraya River during 2552 to 2556 B.E. | 99 |
| Figure 6.3 Histogram of difference in EC predicted (observed data – predicted data) from a) time delay model, b) distance model and c) space and time model..... | 100 |
| Figure 6.4 Relationship between observed TDS and predicted TDS from three models..... | 101 |
| Figure 6.5 Comparisons of the models computed and measured TDS in Chaophraya River during 2552 to 2556 B.E. | 102 |
| Figure 6.6 Histogram of difference in TDS predicted (observed data – predicted data) from a) time delay model, b) distance model and c) space and time model..... | 103 |
| Figure 6.7 Relationship between observed PO ₄ ³⁻ and predicted PO ₄ ³⁻ from three models..... | 104 |
| Figure 6.8 Comparisons of the models computed and measured TDS in Chaophraya River during 2552 to 2556 B.E. | 105 |
| Figure 6.9 Histogram of difference in PO ₄ ³⁻ predicted (observed data – predicted data) from a) time delay model, b) distance model and c) space and time model..... | 106 |
| Figure A.0.1 Model overview | 141 |
| Figure B.0.1 Historical data of total dissolved solid from monitoring stations along Chaophraya River during 2538-2556 B.E. | 145 |
| Figure B.0.2 Historical data of total solid from monitoring stations along Chaophraya River during 2538-2556 B.E. | 146 |



| | |
|--|-----|
| Figure B.0.3 Historical data of EC from monitoring stations along Chaophraya River during 2538-2556 B.E..... | 146 |
| Figure B.0.4 Historical data of suspended solid from monitoring stations along Chaophraya River during 2538-2556 B.E. | 147 |
| Figure B.0.5 Historical data of dissolved oxygen from monitoring stations along Chaophraya River during 2538-2556 B.E. | 147 |
| Figure B.0.6 Historical data of NO_2^- from monitoring stations along Chaophraya River during 2538-2556 B.E. | 148 |
| Figure B.0.7 Historical data of NO_3^- from monitoring stations along Chaophraya River during 2538-2556 B.E. | 148 |
| Figure B.0.8 Historical data of PO_4^{3-} from monitoring stations along Chaophraya River during 2538-2556 B.E. | 149 |
| Figure B.0.9 Historical data of fecal coliform from monitoring stations along Chaophraya River during 2538-2556 B.E. | 149 |
| Figure B.0.10 Historical data of total coliform from monitoring stations along Chaophraya River during 2538-2556 B.E. | 150 |
| Figure B.0.11 Historical data of pH from monitoring stations along Chaophraya River during 2538-2556 B.E. | 150 |
| Figure B.0.12 Historical data of water temperature from monitoring stations along Chaophraya River during 2538-2556 B.E. | 151 |
| Figure B.0.13 Historical data of NH_3 from monitoring stations along Chaophraya River during 2538-2556 B.E. | 151 |
| Figure B.0.14 Historical data of biochemical oxygen demand from monitoring stations along Chaophraya River during 2538-2556 B.E..... | 152 |



LIST OF ABBREVIATION

| Abbreviation | Description |
|-------------------------------|--|
| 3-CA | 3-chloroaniline |
| AAD | Average Absolute Deviation |
| AARE | Average Absolute Relative Error |
| AHGA | Aggregation Hybrid Genetic Algorithm |
| ANFIS | Adaptive Neuro Fuzzy Inference System |
| ARIMA | Autoregressive Integrated Moving Average |
| BE | Buddhist Era |
| BOD | Biochemical Oxygen Demand |
| BP-NN | Back-Propagation Neural Networks |
| Ca ²⁺ | Calcium (II) ion |
| CART | Classification And Regression Trees |
| CC | Coefficient of Correlation |
| CE | Coefficient of Efficiency |
| Cl ⁻ | Chloride |
| COD | Chemical Oxygen Demand |
| CODMn | Chemiluminescence Detection of Permanganate Index |
| DO | Dissolved Oxygen |
| Dy | Community Dynamics |
| EC | Electrical Conductivity |
| Fo | Community Evenness |
| GML | Geography Markup Language |
| H ₂ S | Hydrogen Sulfide |
| HCO ₃ ⁻ | Bicarbonate |
| ISSADM | Integrated Seasonal Separate Advection–Diffusion Model |
| K ⁺ | Potassium (I) ion |
| LM | Levenberg–Marquardt algorithm |
| LS-SVM | least Squares Support Vector Machine |

| Abbreviation | Description |
|---------------------------------|--|
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| Mg ²⁺ | Magnesium (II) ion |
| MLP | Multilayer Perceptron |
| MLR | Multiple Linear Regression |
| MNN | Modular Neural Network |
| MRE | Mean Relative Error |
| MSE | Mean Squared Error |
| Na ⁺ | Sodium (I) ion |
| NH ₃ -N | Ammonia-nitrogen |
| NH ₄ ⁺ -N | Ammonium-nitrogen |
| NO ₂ ⁻ -N | Nitrite-nitrogen |
| NO ₃ ⁻ -N | Nitrate-nitrogen |
| NSHGA | Non-Dominated Sorting Hybrid Genetic Algorithm |
| OpenMI | Open Modeling Interface |
| PCA | Principal Component Analysis |
| PMI | Partial Mutual Information |
| PO ₄ ³⁻ | Phosphate |
| PSO | Particle Swarm Optimization |
| R | community richness |
| R ² | Coefficient of determination |
| RBF | Radial Basis Function |
| RGA-SVR | Real-value Genetic Algorithm Support Vector Regression |
| RMSE | Root Mean Square Error |
| RVS | Recursive Variable Selection |
| SDE | Standard Deviation Error |
| SO ₄ ²⁻ | Sulfate |
| SR | Solar Radiation |
| SS | Suspended Solids |
| SVM | Support Vector Machine |

| Abbreviation | Description |
|--------------|---|
| SVR | Support Vector Regression |
| TDS | Total Dissolved Solids |
| TKN | Total Kjehldahl Nitrogen |
| TOC | Total Organic Carbon |
| TP | Total Phosphorous |
| TS | Total Solids |
| WS | Wind Speed |
| WASP | Water Quality Analysis Simulation Program |

