Integrated evaluation model for surface water quality

A case study in Ho Chi Minh City, Vietnam

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Thu Thi Minh Nguyen

Master of Environmental Engineering and Management, Asian Institute of Technology (AIT), Patthumthani, 12120, Thailand

School of Science
College of Science, Engineering and Health
RMIT University

June 2018
Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of this thesis is the result of work which has been carried out since the official commencement date of approved research program; and any editorial work, paid or unpaid, carried out by third party is duly acknowledged.

Thu Thi Minh Thu Nguyen

Date: 30-06-2018
Acknowledgments

I would like to express my very great appreciation of Associate Professor Colin Arrowsmith for his valuable and constructive suggestions during the planning and development of this research work. His professional guidance, constructive recommendations, and willingness to give his time so generously have been very much appreciated. I would also like to express my deep gratitude to Associate Professor David Fraser, my research supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work.

My grateful thanks are also extended to the staff in the Department of Natural Resources and Environment, Department of Science and Technology in Ho Chi Minh City, Vietnam for their assistance and data supporting in helping to fulfil this research. Many thanks are also given to the staff in the Discipline of Geospatial Science, School of Science Royal Melbourne Institute of Technology for their support, and guidance during the time of my study.

I would also like to thank my friends Dr. Brian, Mr. Hai, Mrs. Phuong, Ms. Michelle for their patience and for staying with me during my stressful time, giving support and encouragement in keeping my progress on schedule.

Finally, I wish to thank the children: little Mia and Andy, and my parents for their enormous support and encouragement throughout my study.
Abstract

The assessment of water quality is an important aspect of environmental management. One method commonly applied is the Water Quality Index. The index consists of a group of experimental formulae used for calculating the overall quality of the water environment. In Vietnam, the water quality index is used to calculate water quality based on nine measured parameters, and was officially sanctioned as the standard method for assessing water quality in 2011.

In this study another method, adopting the fuzzy logic concept for calculating water quality, is applied. This method has the advantage of dealing with an unlimited number of quality parameters, and can allocate quality classes by calculating quantitative data obtained through monitoring water quality at a number of monitoring sites. However, the method has some limitations such as data loss during the evaluation, and uncertainties due to the method not considering the overall physical environment of the catchments that feed water into the monitoring sites.

Fuzzy comprehensive analysis comprises of two main parts: the fuzzy membership and the weights. This study applied the simplest and most used membership function which is triangular and semi-trapezoidal shape, and three different methods of assigning weights: Entropy, Overweight standard and F-statistic for testing the quality of water in the study area, Ho Chi Minh City, Vietnam in South-East Asia, using the monitored quality dataset in 2015. It showed big differences among the results when tested with different weighting methods which is the uncertainty of the Fuzzy method.
Therefore, Geographic Information System (GIS) techniques including computer-based sub-watershed delineation modelling using a geographic hydrological model, and geographical regression models, including Ordinary Least Square and Geographically Weighted Regression models, were applied to reduce the uncertainty in the fuzzy process. Results of this study demonstrate the model’s applicability and capability for use in water quality evaluation. From the results, an integrated assessment comprising fuzzy and GIS methods is developed for better assessment of the water quality and therefore contributing to the better management.

**Keywords:** environmental quality management, integrated surface water quality assessment, fuzzy comprehensive analysis, ordinary least square regression, geographically weighted regression, ArcHydro, subwatershed delineation model, GIS.
Table of Contents

 Declaration ............................................................................................................. 1
 Acknowledgments ................................................................................................. 2
 Abstract .................................................................................................................. 3
 List of Tables ......................................................................................................... 10
 List of Figures ........................................................................................................ 12
 List of Abbreviations ............................................................................................ 15
 List of Appendixes ................................................................................................. 18

 1. INTRODUCTION ................................................................................................ 19
 1.1. Objectives and research questions ............................................................... 21
 1.1.1 Objectives .................................................................................................. 21
 1.1.2 Research questions .................................................................................... 22
 1.1.3 Research’s framework ............................................................................... 23
 1.2 Background ..................................................................................................... 24
 1.2.1 Research rationale .................................................................................... 24
 1.2.2 How the research will build on the current body of knowledge .......... 27
 1.2.3 Contribution to the community ................................................................ 28
 1.2.4 Significance and original contribution to knowledge ......................... 30
 1.3 Methodologies ................................................................................................. 31
 1.3.1 Materials .................................................................................................. 31
 1.3.2 Methods .................................................................................................... 32
1.4 Research’s content .................................................................................................................. 34
1.5 Summary .............................................................................................................................. 37

2. LITERATURE REVIEW ........................................................................................................ 38

2.1 Factors influencing water quality ........................................................................................... 39
  2.1.1 Environmental perspective ............................................................................................. 40
  2.1.2 Economic perspective ..................................................................................................... 41
  2.1.3 Social perspective ........................................................................................................... 42
2.2 Water quality indicators ......................................................................................................... 43
2.3 Water quality regulations ...................................................................................................... 47
2.4 Surface water quality assessment ........................................................................................ 50
  2.4.1 Water Quality Index method ........................................................................................ 51
  2.4.2 Fuzzy Comprehensive method ..................................................................................... 52
  2.4.3 Geographical Information System’s techniques ............................................................. 55
2.5 Summary .............................................................................................................................. 61

3. OVERVIEW OF THE STUDY AREA ..................................................................................... 62

3.1 Overview of the study area ................................................................................................... 63
  3.1.1 Climate ............................................................................................................................ 64
  3.1.2 Topography .................................................................................................................... 64
  3.1.3 Administrative districts and population ........................................................................ 66
  3.1.4 Urban development ......................................................................................................... 70
  3.1.5 Land use ......................................................................................................................... 72
3.2 Environmental resources: Surface water ............................................. 74
  3.2.1 The water network ........................................................................ 74
  3.2.2 The surface water quality ............................................................. 80
  3.2.3 Main pollution sources of the surface water quality ....................... 81
  3.2.3 Threats to the surface water quality resource ................................. 82
3.3 Summary ............................................................................................. 86

4. WATER QUALITY AND GEOGRAPHICAL DATASETS .................. 88
  4.1 Surface water quality parameters (Textural dataset) ......................... 89
    4.1.1 Monitoring sites ........................................................................ 89
    4.1.2 Monitored parameters ............................................................... 95
  4.2 Geographical Dataset ...................................................................... 102
    4.2.1 Digital Elevation Model (DEM) ................................................... 102
    4.2.2 Surface water network ............................................................... 105
    4.2.3 Administrative information ....................................................... 105
    4.2.4 Land use information ............................................................... 106
  4.3 Summary ............................................................................................. 107

5. INTEGRATED ASSESSMENT OF SURFACE WATER QUALITY 108
  5.1 Standard values used for quality parameters and the 2015 applied
    dataset .............................................................................................. 109
    5.1.1 Quality standards ....................................................................... 109
    5.1.2 Applied dataset ......................................................................... 113
  5.2 Water quality assessment processes ............................................... 114
5.2.1 Vietnam’s Water Quality Index ............................................. 115
5.2.2. Integrated fuzzy elevation model ........................................ 117
5.3 Summary ................................................................................ 140

6. SURFACE WATER QUALITY RESULTS .............................. 141
6.1 General evaluation of water quality from 2010 to 2015 .......... 142
6.2 Water quality results applying integrating assessment methods ...... 145
6.3 Discussion ............................................................................. 152
6.3.1 Discussion on the water quality of the study area ............... 152
6.3.2 Discussion on the assessment models ............................... 154
6.3.3 Track the uncertainty of the fuzzy model ......................... 156
6.3.4 Proposing an option for reducing the uncertainty of the fuzzy model ................................................................. 158
6.4 Summary ............................................................................. 162

7. GEOGRAPHICAL ANALYSIS OF SURFACE WATER QUALITY AND RESULTS ......................................................... 164
7.1 Applied datasets .................................................................... 165
7.2 Building the regression models ............................................. 166
7.2.1 Regression model build-up .............................................. 167
7.2.2 Watershed delineation model .......................................... 173
7.3 Results .................................................................................. 176
7.3.1 Sub-watershed delineation map ..................................... 176
7.3.2 Regression’s results ......................................................... 179
7.4. Discussion on the watershed delineation and the regression models ........................................ 195
7.4.1 Watershed delineation process .......................................................................................... 195
7.4.2 Results of the regression models ..................................................................................... 197

7.5 Summary .......................................................................................................................... 198

8. PROPOSING INTEGRATED SURFACE WATER QUALITY ASSESSMENT MODEL .......................................................... 200
8.1 Discussion on current surface water quality assessment process of the study area ......................................................................................................................... 201
8.2 Discussion on Fuzzy model and GIS’s techniques ............................................................... 204
8.3 Proposed monitoring program ............................................................................................ 204
8.3.1 Stage 1: Long-term scheduled program ......................................................................... 205
8.3.2 Stage 2: Annual assessment program .............................................................................. 211

8.4 Summary .......................................................................................................................... 213

9. CONCLUSION AND RECOMMENDATION ................................................................. 214
9.1 Conclusion ......................................................................................................................... 215
9.1.1 Conclusion on the research’s objectives ........................................................................ 215
9.1.2 Conclusion on the research’s contribution ................................................................. 224

9.2 Recommendations ........................................................................................................... 226
9.2.1 Recommendations on the assessment model comprising Fuzzy logic and GIS .............. 226
9.2.2 Recommendations on future research .......................................................................... 227

References .................................................................................................................................. 228
## List of Tables

1.1 List of applied methods ........................................... 32
3.1 Twenty four districts of the City ............................. 67
4.1 Details of the monitoring sites of the study area ........ 90
4.2 River monitoring sites ........................................... 94
4.3 Quality parameters in 2015 .................................... 97
4.4 Percentage of Land-use purposes ......................... 106
5.1 Recommended surface water quality standard levels for 18 monitored parameters .................................................... 112
5.2 Monitored values at Ben Cat site (Dong Nai River) .... 114
5.3 Quality assessment based on the WQI’s result .......... 117
5.4 Membership value f(c_i) according to the quality group 126
5.5 Standard thresholds for Biological Oxygen Demand (BOD) ................................................................. 127
5.6 Fuzzy matrix built for Phu Long site ........................ 132
5.7 Weights of the monitored parameters applying overweight standard method for Phu Long monitoring site 138
5.8 Membership value for each quality group at Phu Long site 139
6.1 Results of water quality from the two integrating methods ................................................................. 146
6.2 Weighting values of three different methods applying monitored dataset at low tide ................................. 150
7.1 Transferred scores from land-use type ...................... 169
7.2 Overview of the model’s power of the tested quality parameters applying Ordinary Least Square model ........ 181
7.3 Summary of Ordinary Least Square results of Manganese and Dissolved Oxygen, at high tide 184
7.4 Summary of Ordinary Least Square results of Dissolved oxygen and Coliform, at low tide 189
7.5 The two parameters with highest model’s power ($R^2$) 190
List of Figures

1.1 Research framework 23
2.1 Overview of the water cycle 39
3.1 Overview of Ho Chi Minh City, Vietnam 62
3.2 The City’s shape and location 63
3.3 Overview of the City topography 65
3.4 Population distribution (number of people) 68
3.5 Population density distribution 69
3.6 The City’s developing zones 71
3.7 Land-use in 2015 73
3.8 Overview of the large rivers flowing into the City 75
3.9 The City’s stream lines and monitoring sites 78
3.10 Constructional (in red) development by time 84
3.11 Construction land (in red) in 2011 84
3.12 Heavy urban flooding due to very high rainfall on 22 May 2017 86
4.1 Overview of the datasets applied in the research. 88
4.2 Location of the monitoring sites and the surface water network of the study area 92
4.3 Triangulated Irregular Network of the study area 104
5.1 Overview of the method’s objective 118
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Overview of the fuzzy assessment process</td>
<td>119</td>
</tr>
<tr>
<td>5.3 Semi-trapezoidal and triangular graphs of membership function</td>
<td>123</td>
</tr>
<tr>
<td>5.4 General membership function of the model</td>
<td>124</td>
</tr>
<tr>
<td>5.5 Membership shape for group of 4 limits in the standard level</td>
<td>129</td>
</tr>
<tr>
<td>5.6 Membership shape for group of 3 limits</td>
<td>129</td>
</tr>
<tr>
<td>5.7 Membership shape for group of 2 limits</td>
<td>129</td>
</tr>
<tr>
<td>5.8 Membership function for pH parameter</td>
<td>130</td>
</tr>
<tr>
<td>5.9 Membership function for Manganese parameter</td>
<td>130</td>
</tr>
<tr>
<td>6.1 Overview of the relation of Biological oxygen demand and Chemical</td>
<td>142</td>
</tr>
<tr>
<td>oxygen demand parameters at canal sites from 2010 to 2015</td>
<td></td>
</tr>
<tr>
<td>6.2 Overview of Lead, Cadmium and Cooper values from 2010 to 2015,</td>
<td>143</td>
</tr>
<tr>
<td>at low tide</td>
<td></td>
</tr>
<tr>
<td>6.3 Overview of Coliform and Escherichia Coliform values from 2010 to</td>
<td>144</td>
</tr>
<tr>
<td>2015, at low tide</td>
<td></td>
</tr>
<tr>
<td>6.4 Overview of Ammoniacal nitrogen (NH4-N), and Organic phosphate</td>
<td>145</td>
</tr>
<tr>
<td>(PO4-P) values (mg/l) from 2010 to 2015, at low tide</td>
<td></td>
</tr>
<tr>
<td>6.5 Location of the Nga Bay monitoring sites</td>
<td>161</td>
</tr>
<tr>
<td>7.1 Overview of the regression process</td>
<td>172</td>
</tr>
<tr>
<td>7.2 Overview of the Watershed delineation process</td>
<td>175</td>
</tr>
<tr>
<td>7.3 Computerised Digital Elevation Model</td>
<td>177</td>
</tr>
<tr>
<td>7.4 Computerised Sub-watershed area</td>
<td>178</td>
</tr>
</tbody>
</table>
7.5 Analysis of variance of Manganese at high tide 182
7.6 Analysis of variance of Dissolved Oxygen at high tide 183
7.7 Analysis of variance of Dissolved Oxygen, at low tide 187
7.8 Analysis of variance of Coliform, at low tide 188
7.9 Standardized Residuals distribution for Manganese, at high tide 191
7.10 Standardized Residuals distribution for Dissolved Oxygen at high tide 192
7.11 Standardized Residuals distribution for Dissolved Oxygen at low tide 193
7.12 Standardized Residuals distribution for Coliform at low tide 194
7.13 Detailed distribution of the computerised sub-watershed catchment area 196
8.1 Overview of the water quality assessment program 201
8.2 Overview of the proposed monitoring program 205
8.3 Stage 1: Long term schedule program 207
8.4 Stage 2: Annual assessment program 212
List of Abbreviations

AHP Analytic hierarchy process
ANOVA Analysis of variance
DEM Digital elevation model
DLGs Digital Line Graphs
FCA Fuzzy comprehensive analysis
GDP Gross domestic product
GIS Geographical information system
GWR Geographically weighted regression
LiDAR Light Detection and Ranging
mg/L Milligram per litre
MMT Methylcyclopentadienyl manganese tricarbonyl
MPN/100mL Most probable number per 100 millilitres
NTU Nephelometric Turbidity Unit
OLS Ordinary Least Square
SAW Simple additive weighting
TINs Triangulated Irregular Networks
TOPSIS The Technique for Order Preferences by Similarity to an Ideal Solution
VIKOR The VlseKriterijumska Optimizacija I Kompromisno Resenje
WQI Water quality index
<table>
<thead>
<tr>
<th>No.</th>
<th>Code</th>
<th>Site Name</th>
<th>No.</th>
<th>Code</th>
<th>Site Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LVS</td>
<td>Le Van Sy</td>
<td>22</td>
<td>RT</td>
<td>Rach Tra</td>
</tr>
<tr>
<td>2</td>
<td>DBP</td>
<td>Dien Bien Phu</td>
<td>23</td>
<td>BP</td>
<td>Binh Phuoc</td>
</tr>
<tr>
<td>3</td>
<td>AL</td>
<td>An Loc</td>
<td>24</td>
<td>SG</td>
<td>Sai Gon</td>
</tr>
<tr>
<td>4</td>
<td>TL</td>
<td>Tham Luong</td>
<td>25</td>
<td>PA</td>
<td>Phu An</td>
</tr>
<tr>
<td>5</td>
<td>HB</td>
<td>Hoa Binh</td>
<td>26</td>
<td>PM</td>
<td>Phu My</td>
</tr>
<tr>
<td>6</td>
<td>OB</td>
<td>Ong Buong</td>
<td>27</td>
<td>CL</td>
<td>Cat Lai</td>
</tr>
<tr>
<td>7</td>
<td>RN</td>
<td>Rach Ngua</td>
<td>28</td>
<td>NB</td>
<td>Nha Be</td>
</tr>
<tr>
<td>8</td>
<td>CV</td>
<td>Cha Va</td>
<td>29</td>
<td>VS</td>
<td>Vam Sat</td>
</tr>
<tr>
<td>9</td>
<td>PD</td>
<td>Phu Dinh</td>
<td>30</td>
<td>TTH</td>
<td>Tam Thon Hiep</td>
</tr>
<tr>
<td>10</td>
<td>NTD</td>
<td>Nhi Thien Duong</td>
<td>31</td>
<td>DT</td>
<td>Dong Thanh</td>
</tr>
<tr>
<td>11</td>
<td>C1</td>
<td>Cau so 1</td>
<td>32</td>
<td>N7</td>
<td>Nga Bay</td>
</tr>
<tr>
<td>12</td>
<td>NHC</td>
<td>Nguyen Huu Canh</td>
<td>33</td>
<td>CM</td>
<td>Cai Mep</td>
</tr>
<tr>
<td>13</td>
<td>CY</td>
<td>Cau Chu Y</td>
<td>34</td>
<td>VC</td>
<td>Vam Co</td>
</tr>
<tr>
<td>14</td>
<td>CM</td>
<td>Cau Mong</td>
<td>35</td>
<td>TC</td>
<td>Thay Cai</td>
</tr>
<tr>
<td>15</td>
<td>HD</td>
<td>Hai Duc</td>
<td>36</td>
<td>BD</td>
<td>Binh Dien</td>
</tr>
<tr>
<td>16</td>
<td>BC</td>
<td>Ben Cui</td>
<td>37</td>
<td>AH</td>
<td>An Ha</td>
</tr>
<tr>
<td>17</td>
<td>BS</td>
<td>Ben Suc</td>
<td>38</td>
<td>TA</td>
<td>Trung An</td>
</tr>
<tr>
<td>18</td>
<td>TT</td>
<td>Thi Tinh</td>
<td>39</td>
<td>BL</td>
<td>Binh Loi</td>
</tr>
<tr>
<td>19</td>
<td>PC</td>
<td>Phu Cuong</td>
<td>40</td>
<td>HP</td>
<td>Hoa Phu</td>
</tr>
<tr>
<td>20</td>
<td>HA</td>
<td>Hoa An</td>
<td>41</td>
<td>PL</td>
<td>Phu Long</td>
</tr>
<tr>
<td>21</td>
<td>N46</td>
<td>N46 canal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Water quality parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>Arsenic</td>
</tr>
<tr>
<td>BOD&lt;sub&gt;5&lt;/sub&gt;</td>
<td>Biological oxygen demand (rest in 5 days)</td>
</tr>
<tr>
<td>Cd</td>
<td>Cadmium</td>
</tr>
<tr>
<td>COD</td>
<td>Chemical oxygen demand</td>
</tr>
<tr>
<td>COLI</td>
<td>Total Coliform</td>
</tr>
<tr>
<td>Cr</td>
<td>Chromium</td>
</tr>
<tr>
<td>Cu</td>
<td>Copper</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>ECOLI</td>
<td>Escherichia coli form</td>
</tr>
<tr>
<td>Mn</td>
<td>Manganese</td>
</tr>
<tr>
<td>NH&lt;sub&gt;4&lt;/sub&gt;-N</td>
<td>Ammoniacal nitrogen</td>
</tr>
<tr>
<td>Oil</td>
<td>Oil</td>
</tr>
<tr>
<td>Pb</td>
<td>Lead</td>
</tr>
<tr>
<td>pH</td>
<td>Potential of Hydrogen</td>
</tr>
<tr>
<td>PO&lt;sub&gt;4&lt;/sub&gt;-P</td>
<td>Organic phosphate</td>
</tr>
<tr>
<td>Sal</td>
<td>Salinity</td>
</tr>
<tr>
<td>Temp</td>
<td>Temperature</td>
</tr>
<tr>
<td>TSS</td>
<td>Total suspended solids</td>
</tr>
<tr>
<td>Turb</td>
<td>Turbidity</td>
</tr>
</tbody>
</table>
List of Appendixes

Appendix 1a: Surface water quality dataset measured at low tide in 2015

Appendix 1b: Surface water quality dataset measured at high tide in 2015

Appendix 2a: Regression results of the quality parameters, measured at high tide, applying Ordinary Least Square model.

Appendix 2b: Regression results of the quality parameters, measured at low tide, applying Ordinary Least Square model.

Appendix 3a: Regression results of Dissolved Oxygen (DO) and Manganese, measured at high tide, applying Geographically weighted regression model.

Appendix 3b: Regression results of Coliform (COLI) and Dissolved Oxygen (DO), measured at low tide, applying Geographically weighted regression model.
CHAPTER 1

1. INTRODUCTION

Water quality is a combination of many testing parameters, which can be grouped in chemical, physical, biological, and radiation characteristics. Depending on the purpose of use of the water, different standards regulating threshold values of the quality parameters have been established. For example, thirty two quality parameters were established in Vietnam’s National technical standard for evaluating the surface water quality. 
(National technical standard for surface water quality QCVN 08:2008/BTNMT, 2008). Therefore, a water body is considered qualified for a specific purpose of use such as drinking, domestic, or irrigation when its monitored values meet the requirement of the standards.

However, it is difficult to assess overall quality of the water if more than two monitored values of two different parameters need to be compared to the standard’s values. One of the methods currently used is Water Quality Index, which is a group of experimental functions for evaluating the quality water. In Vietnam, the Water Quality Index was officially established in 2011, comprising of nine fundamental parameters (Decision on Water Quality Index Calculation 2011). However, this method has limitations such as calculating a limited number of parameters, and subjectively setting up the equations.
Another assessment method applied in recent years was developed based on fuzzy theory. Evaluation methods based on a fuzzy algorithm have been applied in many areas such as automatic control, computer science, or environmental management. This method is able to deal with a number of non-linear relations, simplification, transferring ordinal value to verbal form which means it can integrate qualitative and quantitative data, and its capacity in dealing with data loss. However, it has limitations such as data loss during the fuzziness process, or uncertainty when assigning the weight for quality parameters.

Therefore, this research aims to apply the fuzzy concept to build an evaluation model for surface water quality. Furthermore, one of the limitations of the fuzzy model, which is the uncertainty when assigning weights, was managed by using GIS tools, including a computer-generated watershed model and spatial regression models. Monitored water quality dataset in 2015 in Ho Chi Minh City, Vietnam, has been used for testing the model. The results of the model have shown that the model applying fuzzy theory and GIS’ techniques is applicable. From the results, a surface water quality assessment program for an urban developing area was built.
1.1. Objectives and research questions

1.1.1 Objectives

The main objective of this research is to develop an effective method for assessing water quality in an urban water catchment in a developing country. The method to be developed combines two approaches: modified fuzzy comprehensive analysis based on fuzzy logic which provides a range of value for water quality parameters; and geographical information techniques which contribute to manage the uncertainties in the fuzzy model.

Specific goals include:

- To evaluate and set up a process of surface water quality assessment based on fuzzy theory, multi-criteria analysis and GIS techniques.

- To apply and evaluate the surface water quality model built on the studied area and research the main impact factors affecting the water quality.

- To develop a surface water assessment program for an urban water catchment, comprising of the fuzzy model and GIS methods.

Expected deliverables are:
The most important result achieved in this research is establishing a method based on 2 current approaches, fuzzy comprehensive evaluation and GIS tools for assessing the water quality in a developing city. From them, an effective evaluation program for assessing the surface water quality is also built.

Another result expected to be gained from the research is analysing and evaluating the surface water quality in Ho Chi Minh City, Vietnam, including identifying the relationship between impact factors and water quality.

1.1.2 Research questions

In order to achieve the objectives, six research questions have been prepared:

1. What is recent surface water quality in Ho Chi Minh City, Vietnam?
2. What are the most affecting factors for the surface water quality in Ho Chi Minh City, Vietnam?
3. What are the limitations of the environmental quality assessment model applying fuzzy algorithm?
4. How geographical data, land-use information and geographical information techniques can help to lessen the uncertainty of a model built on fuzzy algorithm?
5. How can an effective program be built to evaluate the surface water quality?
6. Is the built program applying fuzzy logic and GIS’s techniques applicable for a developing city, such as Ho Chi Minh city, Vietnam?

1.1.3 Research’s framework

With the research questions and objectives mentioned above, the research framework is organized as followed:

![Research Framework Diagram]

**Figure 1.1: Research framework**
1.2 Background

1.2.1 Research rationale

*Environmental quality assessment using fuzzy logic*

Introduced by L.A.Zadeh, 1965 (Zadeh, 1965), fuzzy logic is a concept for defining the relationship between the value of an object and a relative group of value. (Lotfi A. Zadeh, 2015).

A set is a group of relative objects. A classic set is a set with a number of belonging members. Therefore, there are two options for any object to a set: the object either belongs to the set or it does not (George Bojadziev, 2007; L.A.Zadeh, 1965). Fuzzy concept, in the other hand, is based on the principle that the value of an object can partly belong to a specific set by its membership value, which is the participation level of the object to the set to which it is referred.

Fuzzy logic is widely applied in many fields of study, including economic, social, and technological studies (Chengpo Mu, 2005; Shan Feng, 1997; V.Duraisamy, 2004). Fuzzy theory is also applied and integrated with multi attribute analysis (Matthias Ehrgott, 2010), and many environmental quality management studies (Meng Lihong, 2009), (Runsheng LV, 2011), (Shufei Lin, 2011; Zhenxiang Xing, 2011; Zhou Zhen-min, 2011).
The main advantages of this method are quantifying the qualitative variables such as “fuzzy quantifiers, e.g., most, many, few, not very many, almost all, infrequently, about 0.8, etc” (L. A. Zadeh, 1983), analysing the data loss and integrating spatial and time data, which make this method more practical than other methods (André Lermontov, 2011), (Yazdanbakhsh & Dick, 2018). One of the techniques widely applied in the environmental field is the fuzzy comprehensive evaluation method (Meng Lihong, 2009), (Runsheng LV, 2011), (Shufei Lin, 2011; Zhenxiang Xing, 2011; Zhou Zhen-min, 2011). This method is easy to apply and advantageous when dealing with qualitative variables and loss of data, which are common problems in the field of environmental management.

Nonetheless, this method has some weaknesses which are as follows:

The fuzzy analysis can change the input data which can cause loss of data during the calculation process. The calculation process will thereby change the input data to the membership value, a level of belonging, from 0 to 1 by applying the membership function, and thus the results after applying many steps of calculation may not be precise.

Furthermore, there are different types of shapes to determine the membership function in the fuzzy evaluation. In the application of fuzzy evaluation in environmental assessment, simple continuous graphs such as triangle, trapezoid, or Bayesian (S-shape) were applied. In some research
about environmental evaluation, the simplified semi-trapezoid membership functions were used (Meng Lihong, 2009), (Runsheng LV, 2011), whilst other studies on assessing the impact to the environment set the simplified Bayesian form (triangle) for fuzzy membership degrees such as evaluating the sandy beach macro-fauna patterns (Bozzeda, Zangrilli, & Defeo, 2016), or salt marsh evolution (Taramelli, 2017), and others proposed the super standard weight method for fuzzy membership function (Shen Jihong, 2011). The use of such different membership functions can lead to various results, and therefore makes the method uncertain.

Besides, this method has uncertainties in setting up weight for parameters and during the fuzzy calculation process. Different methods of assigning weights were used in many research, such as F-statistic (L.Ma, 2010), Analytic Hierarchy Process (Jing Cheng, 2010; Zhu Yulin, 2009), Entropy method (Donghua Wang, 2010), or over-standard method (Ji-hong Zhou, 2009).

**GIS’ techniques were used as a means to limit the uncertainty of the assessment method using fuzzy logic**

Because of the advantages of fuzzy logic in the evaluation environmental quality, the research focused on building an integrated model for water quality assessment based on fuzzy theory. Also, this research proposed applying GIS tools, particularly geographical regression model, to manage
the weaknesses of the fuzzy process, particularly the uncertainty of the fuzzy calculation.

GIS techniques have been successfully applied in the field of environmental assessment, especially in the observation and prediction of environmental disasters such as landslide prediction using satellite remote sensing. (Yang, Zonghu, Robert, & Chun, 2012), dust storm observation and monitoring (Tang-Huang, Gin-Rong, Si-Chee, Hsu, & Shih-Jen, 2012), hurricane and drought effects assessment (Ni-Bin & Zhemin, 2012), and forest fire prediction (John & Xianjun, 2012). Assessment tools using geographic information data were also applied. It was shown that they are very practical in evaluation the water quality, showing the sources of pollution and thereby helping to build the evaluation model (Basem Shomar, 2009; Meixler & Bain, 2010; X. Wang, Homer, Dyer, White-Hull, & Du, 2005), or pollution load prediction model (Meixler & Bain, 2010).

1.2.2 How the research will build on the current body of knowledge

The research was built on the current evaluation methods for water quality, Water Quality Index technique and monitoring parameters for water quality. The Water Quality Index has been using as a useful method for evaluating the water quality in many places such as Canada, Chile, England, Taiwan, India, Australia, and Malaysia (Ionus, 2010; Rizwan Reza, 2010; Simone de Rosemond, 2009). However, this method has weaknesses, such as it builds experimental functions for quantifying the surface water quality based on a
limited number of monitored values of quality parameters. Another
technique for analysing the quality of environment is a method that is built
on fuzzy logic. This method is advantageous because it can deal with an
unlimited number of quality parameters, transferring quality fuzzy values to
quantifying values which can be easier to assess, and it can also deal with
data loss or time series dataset. Nevertheless, evaluation method using fuzzy
concept has limitation relating to the uncertainty when using different shapes
of membership function and weighting methods. Therefore, statistical and
geographical regression models for evaluating the relation between human
factors including population rate, GDP, land use and water quality have been
used in this research to lessen this weakness.

### 1.2.3 Contribution to the community

The evaluation model, integrating between fuzzy logic and GIS, has been
tested applying the dataset in 2015 from Ho Chi Minh City, Vietnam as a
case study.

With the area of 2,095 km² and the population of more than 8 million people
in 2015 ([http://www.gso.gov.vn](http://www.gso.gov.vn)), Ho Chi Minh City is one of the most
dynamic places in Vietnam. The City has advantages such as good sources
of water supply, waterway transportation, and water drainage system.
Besides that, the City is undergoing with fast development with more
construction work, such as buildings, bridges, or roads being built and more
people migrating to the City. However, the surface water system of the city
has been biologically polluted by industrial factories and lack of an appropriate domestic wastewater treatment system. Furthermore, with low elevation, a dense canal and river system, and 6 months of a rainy season, the City is facing many problems such as heavy urban flooding, lack of fresh water, and urban sanitary problems.

The Water Quality Index, which was officially established in 2011 (Decision on Water Quality Index Calculation 2011), is currently the only method of assessing the water quality of the City’s authorities. Although this method has been widely studied and applied in Vietnam because of its quick and simple approach to calculation, it has limitations such as its limited number of calculated parameters. Also, the experimental formulae are set subjectively. Moreover, because of fast economic development, many factors, mostly economic impacts, affect the water quality of the City. Therefore, the water quality of the City cannot be assessed or evaluated by simple functions. Consequently, assessing the water quality using the single method Water Quality Index is not appropriate and precise for the study area.

Consequently, this research aims to find the relationship amongst the factors, mostly human influences, especially land-use change onto the quality of the regional water body by applying statistic and geographical regression models. As a result, a suggestion has been made about the use of a better evaluation model by applying multi criteria analysis, fuzzy theory and GIS’ techniques.
1.2.4 Significance and original contribution to knowledge

*Significance of the research*

Main objective of the research is to build an assessment model for surface water quality, applying fuzzy logic and GIS’s techniques. Using the 2015 monitored dataset of the study area: Ho Chi Minh City, Vietnam, the model was tested according to its applicability and possibility. Furthermore, a relation was demonstrated between impact factors such as geographical environment, land-use, population density and business development and water quality.

*Contribution to the knowledge*

Although the assessment model built on fuzzy logic has shown its effectiveness in evaluating the environmental quality in practice, it has limitations including uncertainties. Therefore, the research has proposed using GIS’s techniques to manage this weakness. GIS’s techniques provide an actual environment model to which the water belongs. It also contains the influence factors which affect the quality of water. Therefore, the uncertainty of the fuzzy model when assessing water quality can be reduced.
1.3 Methodologies

1.3.1 Materials

Two types of dataset were used in the research, including on-site water sampling and geospatial datasets. Details about these datasets were presented in Chapter 4.

*Water sampling data:* On-site monitored datasets at the monitoring sites were used in the two methods of assessment: Water Quality Index and Fuzzy Comprehensive approaches. The 2015 dataset of 18 quality parameters from 41 monitoring sites across the study area were used. These values were collected and measured by the Department of Natural Resources and Environment of Ho Chi Minh City, Vietnam.

*Spatial information:* Spatial data was collected about natural characteristics of the City such as rivers and canals systems and topographic information; administrative data including population distribution, districts’ boundary; socio-economic information such as land-use information, number of all size businesses. This data was managed by the Department of Science and Technology of Ho Chi Minh City, Vietnam.
1.3.2 Methods

In order to successfully achieve the research goals and objectives, some of the methods are to be used, as specified in Table 1.1.

<table>
<thead>
<tr>
<th>Main objectives</th>
<th>Methods to be applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set up an evaluation process for surface water quality based on fuzzy theory.</td>
<td>- Fuzzy comprehensive analysis</td>
</tr>
<tr>
<td></td>
<td>- Weighting methods</td>
</tr>
<tr>
<td></td>
<td>- Water quality index formula WQI</td>
</tr>
<tr>
<td>Identify the main influences for the surface water quality of the studied area</td>
<td>- Watershed modelling</td>
</tr>
<tr>
<td></td>
<td>- Regression models including Ordinary least square model and Geographical weighted regression model</td>
</tr>
</tbody>
</table>

**Fuzzy comprehensive analysis method**

Based on the theory about fuzzy algorithm, the fuzzy comprehensive evaluation FCA model comprises of four basic elements (L.Ma, 2010; Liyun Yang, 2010; Xiaohong Wang, 2010):
- A set of evaluation factors U;
- A set of evaluation criteria V;
- Membership function and evaluation matrix of single factor R;
- Weighting factor \( w \).

In this study, the set of evaluation \( U \) consists of the monitored dataset of the quality parameters. The dataset is the monitored values of 18 quality parameters measured in 2015, and divided into two categories: at high and low tide. The criteria set \( V \) is the quality classes from the national standards, including four groups from the highest to the lowest quality: A1, A2, B1, B2. The simplest shapes of the membership function were also used in the model, including triangular and semi-trapezoidal shapes. Also, different types of weighting method were applied to set the weighting factor \( w \).

**Weighting methods**

The weighting method consists of a group of multi attributes analysis methodology. This type of method is set based on the principle of placing a weighting value on a criterion. Normally, total weighting values of all the criteria is equal to 1. Weighting methods are applied for attribute quality parameters as one of the vital stages in the fuzzy evaluation model (Ji-hong Zhou, 2009; Shen Jihong, 2011; Shufei Lin, 2011; Zhou Zhen-min, 2011; L.Ma, 2010; Donghua Wang, 2010; Zhenxiang Xing, 2011).

**Experimental analysis based on Water Quality Index**

The water Quality Index is an algebraic expression of the water quality for decision-makers to set up a management plan. The index is a tool for
synthesizing and simplifying data and information, and the results gained are easy to understand and use whether by local authorities or by people in general (Ionuș, 2010; Rizwan Reza, 2010; Simone de Rosemond, 2009).

**Watershed modelling and regression models**

The watershed modelling is a computerised model built by ArcHydro, an ArcGIS-based system generated to support water resources applications. The watershed model was then used for running the regression. Two models were used in the study, including the Ordinary Least Square and the Geographically Weighted Regression models. These models were applied to test the relation between the impact factors and the water quality. (Nazeer & Bilal, 2018) (Chen, 2015; Tu, 2013), (Yang, Liu, Luo, & Zheng, 2017).

**1.4 Research’s content**

This research is designed to classify the area’s environmental condition according to Vietnam’s national standard for surface water quality, and then use the geospatial tools to identify influences. Therefore, a proposed model for assessment is made. The research consists of nine chapters, including:

Chapter 1: Introduction

*This Chapter introduces main objectives, rationales and contributions to the knowledge of the research.*

Chapter 2: Literature Review
This Chapter reviews recent research about surface water quality methodologies and GIS’ methods. Research’s gaps are also mentioned in this chapter.

Chapter 3: Overview of the study area

General information about Ho Chi Minh City, Vietnam is given in this chapter, including social economic information and environmental threats such as flooding and water contamination.

Chapter 4: Water quality and geographical datasets

Two types of dataset are presented in this chapter. They are textural and geographical datasets. Textural dataset is monitored values of quality parameters obtained in 2015. Geographical information is the data about population, map of watershed areas, land use and number of all size manufacturers which can affect to the surface water quality of the area.

Chapter 5: Integrated assessment of surface water quality

This chapter presents the two methods used for surface water quality assessment as applied to the study. The first one is the Fuzzy Integrated Assessment method, which is considered the main method used for data analysis in the research. The second method is the Water Quality Index, a standard method for assessing the water quality in the study area.

Chapter 6: Surface water quality’s results
This chapter presents the results from the water quality evaluation based on the water quality parameters dataset measured at the monitoring sites in the study area. Strengths and weaknesses of the fuzzy model were identified. Therefore, a suggestion for modification was made.

Chapter 7: Geographical analysis of surface water quality and results

Conducting steps and results for developing a watershed area and applying it to build the regression models are present in this Chapter. The results gained from the regression models show the relationship between the impact factors, such as population and land-use, and the water quality results obtained from the Water Quality Index and Fuzzy comprehensive and assessment methods.

Chapter 8: Proposing integrated surface water quality assessment model;

This chapter proposes an integrating model for assessing the surface water quality of an area, comprising of Fuzzy Comprehensive Method, Water Quality Index, Watershed Delineation Modelling, and Regression Models.

Chapter 9: Conclusion and Recommendation.

Conclusion of the research’s results and recommendation for updating and further studying are discussed in this chapter.
1.5 Summary

Environmental assessment applying fuzzy logic has shown that it is applicable because of many advantages. However, one of its big limitations is the uncertainty. Therefore, GIS’s techniques, which are a group of methods to involve the actual environment and the impact factors on the water quality, were used to manage this weakness. The assessment model applying fuzzy logic and GIS was then tested for its capability with the monitored dataset in 2015 in Ho Chi Minh City, Vietnam. Consequently, an effective program for evaluating the surface water quality in an urban developing area, consisting of the fuzzy comprehensive assessment and GIS’s methods, was built.
CHAPTER 2

2. LITERATURE REVIEW

This chapter discusses the possible influential factors that can affect the water quality, how to evaluate the quality of a water sample according to the quality parameters, or indicators, and the regulated standards depending on the purpose of use of the water body.

Besides, overview of the research’s methods which are the water quality assessment methods including Water Quality Index method, Fuzzy models, multi-criteria analysing, and GIS’s methods are also discussed in this chapter.
2.1 Factors influencing water quality

Natural water quality is primarily influenced by the dissolved minerals of soil and rocks, bio-synthesis and biodegradable organic matters (Sullivan et al., 2005). In the process of the hydrological cycle, when precipitation falls on land, the water will then run off via many paths depending on the terrain, and form the surface water bodies (ponds, springs, rivers). It can also infiltrate into the soil and follow the groundwater pathway to enter to the surface water bodies. Therefore, natural water quality can be affected by features of regional geography, meteorology and climate, rock and soil characteristics. Figure 2.1 shows the overview of the water cycle on Earth.

![Figure 2.1: Overview of the water cycle](https://www.eea.europa.eu)
According to recent research, water usage has tripled since 1950, and it has been mainly used for domestic, industrial and agricultural use (Clark & Hakim, 2014), (Tsuzuki, 2014). Among these uses, irrigation was the largest the consumption of water, comprising more than 70 percent. Human activities significantly contribute to the decline in water quality such as land-use, population growth and socio-economic factors. As regards on the view of sustainability, water quality impact factors can be identified in three dimension including environmental, social and economic perspectives (Ethan.T.Smith & Harry.X.Zhang, 2004).

2.1.1 Environmental perspective

*Climate*: Precipitation, temperature, daylight hours, humidity, wind direction and strength, and hydro-meteorological features such as highest or lowest tide are all parts of the water cycle and therefore they impact significantly on the water quality (Gu et al., 2014; Sullivan, Agardy, & Clark, 2005; Tsuzuki, 2014).

*Geological features*: Soil and rock formation and structure are two factors that affect the infiltration process of the water and these can remain in the water bodies. For example, water quality can contain Arsenic (As) which is a very toxic element to humans, caused by the natural infiltration process from the soil and rocks of an area to the water body. This poisonous substance was detected in the surface water sample in Mekong delta, Vietnam in 2011 (Nga Thi Bui, 2011).
Ecological system: Plants and animals play an important role in the water circle. Without plants, the infiltration process may be less effective and erosion can occur over the area. Local fauna such as fish and algae also make a large contribution to the change in the surface water quality of the area (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

Groundwater level and quality: Groundwater is part of the water cycle and directly affects the surface water bodies since the underground water is a supply source for the surface water. Because of this linkage, consideration of the groundwater level and quality is a good indicator of surface water quality (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

2.1.2 Economic perspective

Consumption of water (for domestic, industrial and agricultural purposes): This is an indirect indicator. Higher usage of water may lead to higher water discharge and a higher pollution rate. Also, over usage of water for a certain period of time can decrease the cleanliness capability of the water body and change the water environment which can in turn harm living organisms (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

Land-use: for example percentage of forest, farmland, urban area, and defragmentation: Changes in land use over years can cause changes in landscape, terrain, ecosystem, and can thus cause changes in water quality.
The high rate of human impact over large areas such as farmland, urban development, etc. may be an indicator of water quality declining (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

**The Gross Domestic Product (GDP) rate:** Gross domestic product is considered as an indicator of national development. Higher GDP means higher living standards, higher consumption, and also a higher waste discharge rate. Therefore, the GDP development rate can be seen as an indirect factor for the environmental cleanliness and impact on the surface water quality (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

2.1.3 Social perspective

**Waste discharged:** As the result of economic development, the general quantity of waste, including solid and waste water from human activities (domestic, industrial and agricultural output), discharged into the environment has a negative impact on the water quality. For example, if the amount of solid waste is high, the pressure on the environmental quality is predictably high and therefore causes negative effects on the surface water (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

**Treatment facilities:** This factor is an indicator for water quality of the surrounding area. For example, if the number of open dumps and the quantity of solid waste landfilled are high, the surface water quality of the
area may be low due to the filtration and pollution transmission processes (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

**Population density:** High population density can cause pressure on the social infrastructure system, including the waste treatment facilities. More people living in the same area can also be a threat to the surrounding environment, including surface water quality, because people tend to discharge more waste directly to the environment if appropriate control is not taken. Thus, the indicator of population density must be taken into account when evaluating the water quality of the area (Gu et al., 2014; Sullivan et al., 2005; Tsuzuki, 2014).

### 2.2 Water quality indicators

In order to determine the water quality suitable for the purpose of use, quality indicators have been applied. Water quality parameters include determination of some of the basic features of the water body and other specific characteristics, such as levels of nutrients, metals and also biological factors (Goncharuk, 2014; Gu et al., 2014; Ngan, 2003; Sullivan et al., 2005; Tien, 2011; WHO, 1996). In general, the water characteristics are classified into five groups:

1. Aesthetic parameters such as odour, colour;
2. Physical parameters such as temperature, turbidity, salinity, suspended solid, conductivity;
3. Chemical parameters such as pH, alkalinity, hardness, dissolved oxygen, biological oxygen demand, nutrient, organic and inorganic substances;
4. Biological parameters such as Coliform, faecal coliform, planktons, benthos;
5. Radiation parameters such as alpha, beta, gamma radiation beam.

All development activities needs water depending on the purpose of use. The quality standard of water is also different for different areas and nations. If the monitored water indicator exceeds the threshold regulated in the national or regional standard, the water body needs some treatment before being used for a specific purpose. Some of the basic elements of the water quality are:

**Temperature:** The temperature directly impacts the dissolved oxygen concentration in the water body. The lower the temperature the higher the concentration of dissolved oxygen. The water temperature is also a factor which affects the process of photosynthesis, metabolism, development rate, and reproduction of organisms. The temperature itself is affected by human activities such as industrial waste dumping, and the construction of dams, or hydro-power facilities (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).
**Potential of Hydrogen** (pH): pH is identified by the negative logarithm concentration of ion H+ in the water, ranging from 0 to 14. Natural water has a pH value from 6.5 to 8.5 (fresh water), and most of the organisms can live in this neutralized range of pH. Water with a pH less than 5 (high acidity) or more than 9 (high alkalinity) can affect living organisms including humans. Human activities associated with industrial or agricultural processes can discharge waste that can change the natural pH of the water, and consequently affect the ecosystem of the area. (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996)

**Turbidity**: Turbidity is identified by the ability of light passing through the water body and measured by the Nephelometric Turbidity Unit (NTU). High turbidity can cause high temperature and prevent the photosynthesis process of the water plants or weeds. High turbidity prevents the photosynthesis process of the water plants or weeds, and since the process produce oxygen as a by-product (at night) it can negatively affect the water ecosystem as the result of this. High turbidity can result from high concentration of suspended solids (SS), or dense algae and plankton in the water body. Waste discharges from human activities or a soil erosion process are other sources of high turbidity in the water (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

**Dissolved oxygen** (DO): DO is determined by the concentration of oxygen dissolved in the water. It is caused by the diffusion of Oxygen during the stirring process, and the photosynthesis activities of the organisms.
is a by-product of the photosynthesis process. It can be predicted that the water environment is less polluted when the DO concentration is high. However, there are some bad effects on the environment if DO value is higher than a specific limit (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

**Biological oxygen demand** (BOD): This parameter is identified by the oxygen concentration needed for bacteria to decompose organic matter. BOD is an indirect indicator that is used to identify decomposable organic matter in the water. This parameter is one of the vital factors that are used to determine the pollution level of a water body, especially when sourced from industrial or urban discharge. BOD is also used to evaluate the treatment efficiency of a waste water treatment system (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

**Nutrients**: Nutrients in a water body are determined by the concentration of three substances Nitrogen, Phosphorous and Potassium. A high concentration of nutrients in a water body results the expansion of algae and water weeds, change the quality of the environment, and harming other species such as fish and water plants (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

**Metal and toxic substances**: The presence of any substances such as heavy metal, cyanic, and phenol can reduce the cleaning ability of a water stream, and can destroy or prevent the development of micro-organisms. Cyanic and
phenol substances are acutely toxic, and heavy metal can accumulate in the body and cause cancer. These substances can also be deposited in the benthic layer of a water body and may accumulate in the tissue of organisms (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

**Faecal coliform**: These are natural intestinal bacteria living in warm-blooded animals. The presence of these bacteria is an indicator that the water body is polluted. The presence of faecal coliform also indicates that other viruses and bacteria may occur in the water body (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996).

### 2.3 Water quality regulations

The quality standards for water are set depending on the purpose of use and nature of the water body. Each standard regulates the specific number of parameters that need to be tested. The quality values of the parameters depend on the discharge sources or water reuse.

Australia is one of the countries that have stringent regulations on the environmental quality protection and management of water. For water quality, in “An Introduction to the Australian and New Zealand Guidelines for Fresh and Marine Water Quality”, (Australian and New Zealand Environment and Conservation Council, 2000), the environmental values of water quality are related to the uses of aquatic ecosystems, primary industry (including irrigation and human consumption), recreation and aesthetics,
drinking, industrial uses, and cultural and spiritual values. In each sector, specific guidelines for biological, physical, chemical, toxic, and sediment factors are outlined as indicators of pollution.

Using the aquatic ecosystem as an example, biological indicators include algae, macrophytes, macro-invertebrates and fish. Physical and chemical parameters contain nutrient, biodegradable organic matter, dissolved oxygen, turbidity, suspended particulate matter, temperature, salinity, pH and flow indicators. Each indicator has a regulated trigger value that is used for quality assessment. The trigger values are referenced and can be different depending on the monitored area. A decision framework is also built to help the user decide which level of protection should be applied.

For drinking water quality, the guidelines provide for a range of measurable compounds that can be tracked in the water. These are categorised into groups of microorganisms, physical characteristics, including radio-nuclides, and chemical parameters. There are two groups of parameter values, one group is related to the consumer’s health over a lifetime, and the other is related to the aesthetic perspective. In “Australian drinking water guidelines 6, 2011, version 3.1”, updated in March 2015, (Australian Drinking Water Guidelines Paper 6 National Water Quality Management Strategy, 2011), the quality of drinking water is regulated by its micro-organisms, physical and chemical characteristics, and guidelines for the monitoring program development for water quality.
In this study, water quality of a specific Asian developing city, Ho Chi Minh City, Vietnam is considered. The Vietnam quality standard contains a threshold for checking whether the water body needs treatment. As regards surface water quality, in 2008 Vietnam promulgated the National technical regulation on surface water quality (*National technical standard for surface water quality QCVN 08:2008/BTNMT*, 2008). This regulation sets the limit thresholds for the water quality in rivers and canals. The quality parameters are categorized as follow:

- Fundamental physical and chemical characteristics including pH, Dissolved Oxygen, Total Suspended Solid, Chemical Oxygen Demand, and Biological Oxygen Demand.
- Nutrient characteristics including Nitrate, Nitric, Ammonia, and Phosphate.
- Heavy metals: Arsenic, Cadmium, Lead, Chromium, Copper, Zinc, Iron, Mercury and surfactants;
- Organic matter including oils and grease, phenols, pesticides (organic chlorine and phosphor) and herbicides;
- Radiation parameters including $\alpha$ and $\beta$ beam activities;
- Bacteria parameters including Coliforms and Escherichia Coliform;
- Others parameters including Chlorine, Fluorine and Cyanide.

Value ranges of these 32 parameters are set according to 4 groups of water quality: A1, A2, B1, and B2. Group A1 is applied for domestic supply and other purposes; A2 is used for domestic supply with little treatment and other
purpose; B1 is applied for agricultural purpose or other equivalent purpose; and B2 is applied for transport purpose or other equivalent purpose.

2.4 Surface water quality assessment

Integrated water quality assessment is a method of comprehensive evaluation of water quality based on a number of quality parameters. Each parameter is considered as a part of the overall water quality. Several methodologies have been developed recently. One of these is based on Artificial Neural Networks (ANN) which is a computation network simulating the animal brains. The method has been well developed and used in the field of data mining, machine learning, and has been recently applied to assess the environmental quality ((Charulatha, Srinivasalu, Uma Maheswari, Venugopal, & Giridharan, 2017; Mengshan Li, 2017). One other method is based on matter-element model. This is a model built on three elements: matter-features-values. Recent studies have integrated fuzzy theory in order to build a modified fuzzy matter-element model, and apply to comprehensively evaluate water quality (Li, Yang, Wan, & Hörmann, 2017; Liu & Zou, 2012). This study, in the other hand, uses Water Quality Index and a combination of multi-criteria analysis method and fuzzy comprehensive assessment to evaluate the water quality.

Combination of methods for more effectively assessing the water quality has been applied, for example integrating linear regression and ANN (Charulatha et al. 2017), or using fuzzy algorithm for modifying the matter-element
model (Li et al. 2017). One of the weaknesses of assessment methods such as ANN, linear regression, or multi-criteria methods, is they consider each of water quality’s parameters as a single independent variable. However, in fact there are many quality parameters having correlation to each other, for example Biological oxygen demand (BOD indicator) is highly related to Chemical oxygen demand (COD indicator). Therefore, some researches have been using fuzzy theory as a supporting tool to modify these methods (Li et al. 2017; Liu & Zou 2012). This study also uses fuzzy and multi-decision methods as a combination technique for assessing the water quality. One of the advantages of fuzzy concept is it helps to define the water quality based on the level of belonging, and it does not require the parameters to be independent

2.4.1 Water Quality Index method

The Water Quality Index is a popular method for assessing the overall quality of water. It is a set of experimental formulae designed for a specific region. This approach has been developed from different kinds of formula and applied in various areas such as Canada, Chile, England, Taiwan, India, Australia, and Malaysia (Ionus, 2010; Rizwan Reza, 2010; Simone de Rosemond, 2009).

In Vietnam, this method was applied in Ho Chi Minh City in 2007 (Study on surface water partition in Ho Chi Minh City using water quality index WQI and suggestion possible use, 2007) and other areas (Pham, 2009), (That,
2007). In 2011, the national Water Quality Index was officially established. It is consisting of nine fundamental parameters (Decision on Water Quality Index Calculation 2011). It is currently a recognized method for assessing the water quality of the study area.

2.4.2 Fuzzy Comprehensive method

Fuzzy logic is a concept defining the relationship between the value of an object and a relative group of value (Zadeh, 1965). Fuzzy Comprehensive method is an approach for assessing the quality of the water based on fuzzy logic. It consists of two parts: the membership and the weight.

In some recent research, the membership function has been applied for environmental evaluation is a mixed type of linear lower semi-trapezoidal and triangular distribution (L.Ma, 2010; Liyun Yang, 2010; Runsheng LV, 2011; Shen Jihong, 2011; Shufei Lin, 2011; Zhenxiang Xing, 2011). For each parameter, the membership function is set so as to be based on the standard limit.

Weight is the level of importance given to each quality parameter, and it has a value ranging from 0 to 1, equivalent to 0 – 100%. Therefore, if a parameter is a criteria for identifying water quality, methods for assigning weights are multi-criteria analysis (Gwo-Hshiung Tzeng, 2011).
Some popular methods in this family of multi-criteria methodologies include Analytic Hierarchy Process (AHP) (Jing Cheng, 2010; Xiang Zhao, 2011; Yingjie, 2011), Simple addictive weighting method (SAW) (Azizollah Memariania, 2009), The Technique for Order Preferences by Similarity to an Ideal Solution (TOPSIS) (Liu, 2009; Nüfer Yasin Ates, 2006), The VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) (Jin Han Park, 2011; Serafim Opricovic, 2007).

In this research, because the water is a natural component, the weights of its quality parameters cannot be subjectively assigned. Therefore, depending on the monitored dataset and the standard’s threshold, three different weighting methods were used for testing the applicability of the fuzzy model. They include:

**Over-weight weighting method:** This method emphasizes the difference between the main factor and the standard value. In general, this approach is bases on the principle of "the more or less the better". For example, with the parameter such as Biological Oxygen Demand the principle "the less the better" is applied, whereas "the more the better" is applied in the case of Dissolve Oxygen parameter (Ji-hong Zhou, 2009; Shen Jihong, 2011; Shufei Lin, 2011; Zhou Zhen-min, 2011).

**F-statistic:** F-statistic is defined as a ration between square standard deviation amongst groups and square standard deviation of each group. Higher F means a higher deviation amongst groups, and the higher
concentration within a group, which can make the better discrimination (L.Ma, 2010).

Entropy: Entropy is a physical concept in a thermal dynamic system for defining the turbulence of a system. In environmental fields, entropy is used for quantifying the information. The more information in a specific index the more importance is the effect of this index on the system (Donghua Wang, 2010; Zhenxiang Xing, 2011).

Assessment method integrated with Fuzzy concept has limitations in its membership functions and setting weights for the parameters. Different types of equations were used to build the membership function, such as trapezoidal shape (Shufei Lin, 2011; Zhenxiang Xing, 2011) or bell shape (Li et al. 2017). Also, different methods for setting weights were used, including using different multi-decision methods such as: F-stat (L.Ma 2010), Analytical Hierarchy Process (AHP) (Jing Cheng 2010), Entropy (Donghua Wang 2010). All of the above researches have proven its capability applicability into the research studies.

However, using fuzzy model with the most simple membership function (triangle and trapezoidal) and different weighting methods (F-stat, Entropy, and Overweight standard) into this research study (Ho Chi Minh City, Vietnam), the results were much different and seemed not demonstrate the real water quality of the area. Therefore, watershed based model had built to lessen this vagueness.
2.4.3 Geographical Information System’s techniques

Applying GIS techniques for assessing overall quality of water has been developing in recent years. Some of the models that have been recently used are: Inverse Distance Weighted (IDW) interpolation model to map spatial distribution of contaminants (Nath et al., 2018; Yazidi, Saidi, Ben Mbarek, & Darragi, 2017), Soil and Water Assessment Tool (SWAT) model for evaluating the effect of climate change on the surface water quality (Shrestha, Bhatta, Shrestha, & Shrestha, 2018; Wuxia Bi, 2018). Also, there has been use of integrating Time series modelling, multivariate statistical/Geostatistical analyses, and artificial intelligence techniques for evaluating the ground water quality (Deepesh Machiwal, 2018).

This research used Fuzzy theory to test the water quality of the study area. However, as mentioned above one of the weaknesses of the Fuzzy Comprehensive method is the uncertainty of the results. Therefore, Geographical Information System has been applied to manage this limitation. These include Digital elevation model, Watershed delineation model and the Regression models.

*Digital Elevation Model (DEM)*
As an important part of the watershed modelling, Digital Elevation Model (DEM) is a surface terrain representation. According to Wilson 2018, this model includes different types of modelling such as Digital Line Graphs (DLGs), Triangulated Irregular Networks (TINs), Grids, and Light Detection and Ranging (LiDAR) point clouds (J. P. Wilson, 2018). It can also be derived from satellite modelling (Jarihani, Callow, McVicar, Van Niel, & Larsen, 2015).

In this research, DEM was built by integrating a number of elevation points measured in 2000 by the Department of Sciences and Technology. Data showing impact factors, such as population, land-use, number of enterprises are also incorporate into the DEM model. The model is used as a fundamental layer of the watershed delineation map in which the spatial regression, that is Geographically Weighted Regression, is running. The watershed delineation model is a computerised model simulating the surface water catchment area according to elevation. Therefore, it can represent the actual water environment of the study area. In this study, in order to lessen the vagueness of the fuzzy integrated assessment, it is important to consider the human impacts on the water quality by using the spatial regression model run onto the watershed delineation map. Consequently, the possible impacts and potential quality parameters of the water can be identified and chosen for a better fuzzy integrated assessment.

Although the model has limitations such as errors of artificial sinks (William, Mats, Tomas, & M., 2017); or inaccuracies in mapping flat areas
(Hou, Yang, Sun, & Sun, 2011; H. Zhang & Huang, 2009), it has been used in building the streamline and watershed network in many research and its capabilities have been proved (Ames, Rafn, Van Kirk, & Crosby, 2009; Konadu & Fosu, 2009; Lin, Chou, Lin, Huang, & Tsai, 2008; Murphy, Ogilvie, Meng, & Arp, 2007; Nardi, Grimaldi, Santini, Petroselli, & Ubertini, 2008; Wu, Li, & Huang, 2008).

**Watershed delineation model**

This research uses the method of computerising the watershed area built by ArcHydro, a sub-program of ArcGIS-based system (*Arc hydro tools overview*, 2004; J. Zhang et al., 2010). In the model, other data of water quality’s influence such as land-use, population density, number of all-size manufacturers was also included for further testing.

**Regression models**

Two types of regression models were used: Ordinary Least Square and Geographically Weighted Regression model. These two regression models are the ones of the simple linear models. They have been successfully applied in determining environmental quality evaluations, such as salinity monitoring (Nazeer & Bilal, 2018), modelling the relationship between land-use and water quality (Chen, 2015; Tu, 2013), or in estimating the pollution level in point and non-point sources (Yang et al., 2017).
Ordinary Least Square model (OLS)

In order to test the inter-dependence between water quality and impact factors, and show the relationship between these factors, the research aimed to apply the OLS model. This statistical technique has been well developed and applied to many field of study such as business, risk management, and also environmental management (J. H. Wilson, 2012).

In the research, the power of the test $R^2$ (coefficient determination), serial correlation (time series problem), and status of multicollinearity have been checked together with testing of the relationship between the dependent variable (water quality) and independent variables (such as land use, geospatial data, etc.).

According to Pratt and Chang, 2012, general equation for OLS model is shown in function 1 (Pratt & Chang, 2012):

$$y = \beta_o + \sum_{i=1}^{n} \beta_i x_i + \epsilon$$  
(Equation 1)

Where:

$y$: dependent variable

$x$: independent variables, ranging from $i = 1$ to $n$

$n$: number of independent variables.

$\beta$: coefficient
ε : Error term (or Residuals)

The OLS model was tested by ordinary software such as Microsoft Excel (version 2007 or later) or SPSS.

Geographically Weighted Regression model (GWR)

The OLS model is based on the assumption that the relationship between a dependent variable and independent variables are static and consistent across the study area. The model creates a single function for all the features in the dataset. Thus it is called “global model” (Fischer & Getis, 2010). However, with the geospatial data such as distance, or population density, the relationship is different within the space.

Therefore, given the above, the Geographically Weighted Regression model is more applicable. According to Murayama, 2012, GWR “is a technique for spatial statistical modeling used to analyze spatially varying relationships between geographical variables. Unlike the traditional regression framework, GWR allows local rather than global parameters to be estimated”. Introduced as one of the Modeling Spatial Relationships toolsets, including OLS and GWR (Fischer & Getis, 2010), GWR is based on the assumption that a nearer object will get a higher weight in the modelling (Murayama, 2012). GWR builds an equation that is based on the concept that nearer objects will have higher weights than those that are further away. A general calculation for GWR model is as follow (Pratt & Chang, 2012):
\[ y_j = \beta_o (u_j + v_j) + \sum_{i=1}^{r} \beta_i (u_j + v_j) x_{ij} + \mathcal{E}_j \]

(Equation 2)

Where

\( j \): location of a point

\( y_j \): dependent variable at a location \( j \)

\((u_j, v_j)\): coordinates of location \( j \)

\( x_{ij} \): independent variables

This model is called a “local model” because it creates a regression function for every feature in the dataset. It calibrates each one by using a target feature and its neighbour by using exponential distance decay function (Pratt & Chang, 2012):

\[ w_{ij} = \exp^{-\frac{d_{ij}^2}{b^4}} \]

(Equation 3)

Where

\( d \): distance between point \( i \) and \( j \)

\( b \): kernel bandwidth

Therefore the weight decreases when distance \( d \) is higher than bandwidth kernel \( b \).

This model has been developed on the traditional regression framework and applied in many areas of research including environmental quality evaluation.
(Pratt & Chang, 2012; Tu, 2011). Although some critical ideas to this model such as lack of a unified statistical framework, different kernel function and bandwidth, or the collinearity have been pointed out (Fischer & Getis, 2010), it can be used in this research to emphasis the relationship between the variables. GWR can be tested by AcrGIS tool.

2.5 Summary

There are many factors that can possibly affect the quality of a water body, including natural and man-made impacts. Water quality parameters therefore were established to identify the natural characteristics of the water body, and the concentration of the unnatural substances, including toxic and harmful substances. Water quality standards have been official published according to the quality parameters in order to determine what purpose of use is appropriate for the water body or what further treatment is needed. Besides, Water Quality Index has been widely applied for testing the overall water quality. For the purpose of building a better quality assessing method, this study has aimed at using another method: Fuzzy Comprehensive analysis and GIS’s techniques for testing the water quality.
CHAPTER 3

3. OVERVIEW OF THE STUDY AREA

This chapter introduces the study area, Ho Chi Minh City in Vietnam, shown in Figure 3.1. Located in the Southern East of Asia, the City is one of the fastest developing cities of the country and it contributes the highest GDP to the country’s economy. However, with its fast economic development, the City has been facing many problems relating to environmental decline particularly noise, air and water quality pollution.

Source: Esri’s world map

Figure 3.1: Overview of Ho Chi Minh City, Vietnam
3.1 Overview of the study area

Ho Chi Minh City is located at 10°10’ – 10°38’ North Latitude and 106°22’ – 106°54’ East Longitude. The City has an area of 2,095 km² and a population of about 8 million in 2015, and is one of the most dynamic cities in Vietnam. GDP development rate is 12% on average, and contributes more than 22.5% of the national budget (http://www.gso.gov.vn). Figure 3.2 shows the location and shape of the City’s area.

Source: Esri’s world map

Figure 3.2: The City’s shape and location
3.1.1 Climate

The area is in a tropical region which is mainly affected by monsoon. There are two seasons in area: the rainy season is from November to May, and the dry season is from June to October. The average temperature ranges from 27 – 28°C, with the highest in April and the lowest in December. Rainfall is on average 1,950 – 2,100 mm, with most rain falling in the rainy season (Report of environmental quality monitoring results in 2011, 2012).

3.1.2 Topography

The City is mainly flat, with some high areas in the North-West and North-East. The elevation decreases from the North-West to the South and South East (Figure 3.3). More than 70% of the area is lower than 2 metres above mean sea level (Nigel K.Downes, 2014), and therefore regular flooding occurs, especially in the rainy season from November to May, or during high tides. Figure 3.3 shows the overview of the City’s topography.
Source: Esri’s world map

Figure 3.3: Overview of the City topography
3.1.3 Administrative districts and population

With a total area of about 2,109 km$^2$, the City has 24 administrative districts and it had more than 8.2 million people in 2015. Table 3.1 presents details about the 24 districts of the City, from the highest to the lowest populated (Ho Chi Minh City’s General Statistical Office http://www.pso.hochiminhcity.gov.vn).

The highest population area is on the northern and eastern edge of the city. These areas have been the most recently developed zones, and they have had very high number of immigrants from the surrounding provinces. Figure 3.4 and 3.5 show the population and density distribution among the City in 2015.

The average population density is approximately 20,000 people per km$^2$. The central districts are smaller in size but have a higher population density, whilst the suburban areas are much larger in size but less dense. The highest density is in the centre of the city, shown in Figure 3.5.
Table 3.1: Twenty four districts of the City  
(from the highest to the lowest populated in 2015)

<table>
<thead>
<tr>
<th>No.</th>
<th>District’s name</th>
<th>Density (people per square kilometre)</th>
<th>Population (number of people)</th>
<th>Area (square metres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>44,937</td>
<td>230,596</td>
<td>5,131,515</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>44,686</td>
<td>186,727</td>
<td>4,178,613</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>41,793</td>
<td>238,558</td>
<td>5,708,069</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>41,770</td>
<td>178,615</td>
<td>4,276,157</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>39,862</td>
<td>196,333</td>
<td>4,925,262</td>
</tr>
<tr>
<td>6</td>
<td>Phu Nhuan</td>
<td>37,317</td>
<td>182,477</td>
<td>4,889,932</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>36,257</td>
<td>258,945</td>
<td>7,141,956</td>
</tr>
<tr>
<td>8</td>
<td>Go Vap</td>
<td>32,164</td>
<td>634,146</td>
<td>19,716,167</td>
</tr>
<tr>
<td>9</td>
<td>Tan Phu</td>
<td>29,021</td>
<td>464,493</td>
<td>16,005,570</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>25,098</td>
<td>193,632</td>
<td>7,714,992</td>
</tr>
<tr>
<td>11</td>
<td>Binh Thanh</td>
<td>23,479</td>
<td>487,985</td>
<td>20,783,947</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>22,554</td>
<td>431,969</td>
<td>19,152,843</td>
</tr>
<tr>
<td>13</td>
<td>Tan Binh</td>
<td>20,460</td>
<td>459,029</td>
<td>22,435,806</td>
</tr>
<tr>
<td>14</td>
<td>Binh Tan</td>
<td>13,223</td>
<td>686,474</td>
<td>51,913,785</td>
</tr>
<tr>
<td>15</td>
<td>Thu Duc</td>
<td>11,065</td>
<td>528,413</td>
<td>47,755,362</td>
</tr>
<tr>
<td>16</td>
<td>12</td>
<td>9,670</td>
<td>510,326</td>
<td>52,774,957</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>8,799</td>
<td>310,178</td>
<td>35,252,332</td>
</tr>
<tr>
<td>18</td>
<td>Hoc Mon</td>
<td>3,869</td>
<td>422,471</td>
<td>109,181,189</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2,947</td>
<td>147,168</td>
<td>49,937,943</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>2,544</td>
<td>290,620</td>
<td>114,259,517</td>
</tr>
<tr>
<td>21</td>
<td>Binh Chanh</td>
<td>2,338</td>
<td>591,451</td>
<td>252,932,046</td>
</tr>
<tr>
<td>22</td>
<td>Nha Be</td>
<td>1,385</td>
<td>139,225</td>
<td>100,489,795</td>
</tr>
<tr>
<td>23</td>
<td>Cu Chi</td>
<td>927</td>
<td>403,038</td>
<td>434,725,682</td>
</tr>
<tr>
<td>24</td>
<td>Can Gio</td>
<td>104</td>
<td>74,960</td>
<td>717,732,937</td>
</tr>
</tbody>
</table>
Figure 3.4: Population distribution (number of people)
Figure 3.5: Population density distribution

(number of people per square kilometre)
3.1.4 Urban development

In general, the City can be divided into the following three development zones, as shown in Figure 3.6 (Report of environmental quality monitoring results in 2011, 2012).

The City central’s area (Development zone): The City’s centre includes 13 central districts. The total area is about 14,200 ha (6.8% of the whole City), and the population is about 3.9 million people (52% of the population). This zone has the highest population density (more than 27,000 people/km²) and is stable.

Recent development areas: including six districts, that is district 2, 7, 9, 12, Thu Duc and Binh Tan. The total area of these areas is about 35,183ha (16.8% of the whole City) and population is about 2.2 million (29.4% of the population). These areas were planned and constructed a few years ago. The population growing rate has been 6.2% with high density (6,279 people/km²) in the period 2006-2010.

The developing area: including five districts, that is Binh Chanh, Cu Chi, Hoc Mon, Nha Be and Can Gio. Total area is about 160,172 ha (76.4% of the whole City) and the population was 1.4 million (18.2% of the population) in 2015. These areas had the lowest population density (586 people/km²) and slow growth rate of 5.2% in the period 2006-2010. However it is anticipated that this area will grow rapidly in the near future.
Figure 3.6: The City’s developing zones
3.1.5 Land use

In 2015, agriculture was the largest land use within the City’s boundary. This was located mostly in the northern and western regions, predominantly short term vegetables, flowers, and rice. There was aqua agriculture in the north, and farms for salt harvesting in the south. In recent years, however, there has been a shift from agricultural to residential and commercial land use. The percentage of land used for construction has increased rapidly since 2006. Many new urban areas have been formed, and new bridges and highways have been constructed such as the Vo Van Kiet highway, Thu Thiem tunnel, and metro construction.

In 2015 more than 20% of the City’s area was a natural zone which was officially sanctioned as a natural biosphere: from the Can Gio mangrove forest in the south east of the City to the edge of the East Sea.

Residential and manufacturing zones covered more than 26% of the City’s area. At that time, although the City’s residential area was still relatively small, it had the highest population density. This high density in population had caused many environmental problems such as pollution, sanitation issues, and a lack of environmental resources.

Detailed on purposes of land using is presented in Figure 3.7.
Figure 3.7: Land-use in 2015
3.2 Environmental resources: Surface water

Because the City is mainly flat with a low elevation, and its location is at the frontage of the East Sea (see Figure 3.1), the City has a dense water network and a rich volume of surface water.

3.2.1 The water network

In general, the surface water network of the City can be classified into two categories, that is: rivers and canals.

The river system

The City is located downstream from many large rivers. Main Rivers flowing into the City are Dong Nai, Sai Gon, and Vam Co (Figure 3.8).
Figure 3.8: Overview of the large rivers flowing into the City

Dong Nai River network (red-covered zone in Figure 3.8)

Dong Nai River network starts in the Lang Biang highland, in the Lam Dong Province, and it extends to the Soai Rap outfall in the South East of the City. The total length of this network is about 630 km, with 40 km of this flowing toward the City. The catchment area is about 38,000 – 39,000 km$^2$ with the water volume of 36 – 37 billion cubic metres per year.
The sub-watersheds of this river network include the Da Nhìm, Da Dang, La Nga, Be, Sai Gon, and Vam Co Rivers. Among them, the Sai Gon and Vam Co rivers are the main influence on the City’s hydrology.

**Sai Gon River network** *(black-covered zone in figure 3.8)*

The network, to the North of the City, starts in the Hon Quang highland to the West of the Dong Nai catchment area. The river flows parallel to the Be River, and it merges with the Dong Nai River in the town of Nha Be, Ho Chi Minh City.

The catchment area is about 4,500 km², with the river length of 250 km. There is also a big lake, named Dau Tieng, in the upstream of the river, and it plays an important role for agriculture and water supply for the abutting provinces, including Ho Chi Minh City and the surroundings, including Tay Ninh and Binh Duong provinces.

**Vam Co River network** *(violet-covered zone in figure 3.8)*

Vam Co River network to the West of the City includes two branches the East and West, and it has a total catchment area of about 6,300 km². The two Rivers, East and West Vam Co, join together in Long An Province. This junction is 36 km from the Soai Rap River outfall which is located in the South East of the City.
The East Vam Co River has a length of 283 km. It starts in Cambodia, flowing through Tay Ninh and Long An Provinces and it merges with the Dong Nai River. The West Vam Co River has the length of 186 km and has the main inflow from Tien River through the Hong Ngu Canal in Dong Thap, a province in Cuu Long Delta in Southern of Vietnam.

*The canal system*

Because the City is downstream from three main large river networks, and has very low elevation, it has a dense and narrow canals system. In the past, transporting goods by boat on these canals was one of the main characteristics of the City. However, with fast economic development and urbanization in recent years, a proportion of the natural canals in the City’s centre have been filled for residential construction, while more irrigation waterways have been built to the north and west of the City for agricultural purposes. Figure 3.9 shows detailed surface water network and monitoring sites.
Figure 3.9: The City’s stream lines and monitoring sites
In general, the inner City has five main canal systems which mainly contribute to the drainage of the whole City. With a total length of 55km, they are the Nhieu Loc – Thi Nghe, Tan Hoa – Lo Gom, Tau Hu – Kenh Doi – Kenh Te, Ben Nghe, and Tham Luong – Ben Cat – Vam Thuat. These canal systems have a low gradient, and therefore it is easy for the sediments from domestic and industrial waste to be deposited in the canals.

Because the City has very dense river and canal network, the surface water resource supplies large number of water volume to the City’s activities such as domestic consumption, industrial and agricultural use. Among the three biggest river systems, Dong Nai River contributes the largest water supply source for the whole area. With a total volume of about 36-37 billion m$^3$, the river provides around 15 billion m$^3$ per day for the City.

Along with using the water for agricultural and aqua-cultural purposes, the City is currently using about 1.15 million m$^3$ per day of water for domestic purposes. The water is processed at three water treatment plants: Thu Duc water supply plant in Dong Nai River (750,000 m$^3$/day); Binh An water supply plant in Dong Nai River (100,000 m$^3$/day); and Sai Gon water supply plant phase 1 in Sai Gon River (300,000 m$^3$/day). Together with the underground water supply of about 114,000 m$^3$/day, the total domestic water use of the city is around 1,264 km$^3$ per day.
Although the volume of surface water supply is plentiful, the quality of this resource has been declining over the years because of inadequate use associated with human activities. Details about the quality of surface water quality of the City are presented as followed.

### 3.2.2 The surface water quality

**Water quality at river**

According to the Report of Environmental status 2005 – 2009, the upper-stream area of the Dong Nai – Sai Gon river system, used for water supply purpose, had a good quality of water. Most of the monitoring values met the required standard; however there were some parameters below the standard including coliform, dissolved oxygen, oil and grease. In the report of the City’s water quality in 2011, the Oil and Coliform value were also higher than the standard’s value (*Report of environmental quality monitoring results in 2011, 2012*). Monitored values in 2014 and 2015 also detected very high concentrations of Coliform and Escherichia Coliform in the area. This is a sign of organic and micro-biological pollution in the water.
Water quality at canal

The City has a very dense and narrow canal network which can easily deposit pollutants in the water. According to the Environmental Report in 2011, the water quality from the period 2005 – 2009 had very high organic pollution concentrations. The concentration of organic and microbiological pollution substances was also very high in 2011 (Report of environmental quality monitoring results in 2011, 2012). Besides, the monitored values from 2014 and 2015 also showed high values of Ammoniacal Nitrogen (NH$_4$-N) in the water. This was a sign of free discharge of domestic waste water, including urine, into the canal system without pre-treatment.

3.2.3 Main pollution sources of the surface water quality

With higher pollution levels in the water quality over the years, particularly in term of human-made pollutants such as oils and grease and organic matters, possible factors contributing to this declining were addressed as follow (Report of environmental quality monitoring results in 2011, 2012)

**Domestic waste water:** It appears waste water from households, hospitals, hotels, and other institutions, is being discharged directly into the canals without treatment. Also, given the canal and river system of the City is downstream from the watershed, it therefore receives a large amount of polluted substances from the upstream areas. It has been shown that from
2005 to 2010, the volume of wastewater discharged into the water system of the City has been increased from 1.2 to 1.5 million cubic metres per day. It is anticipated that the wastewater will increase to 2.2 million cubic metres per day in 2020.

**Industrial waste water:** The total volume of industrial waste water from the industry, including the livestock industry, small scale businesses and transport which was discharged to the Dong Nai River was 342 km$^3$ in 2007. It is predicted to increase by more than 8 times in 2020, and will become the largest source of pollution into surface water network of the City.

**Agricultural waste water and run-off water:** In 2015, most of the City’s area was used for agriculture, as shown in Figure 3.7. Therefore, agricultural and aqua-cultural wastewater was the source for polluting the surface water resource. These types of waste water have high concentrations of oil and grease, chemical compounds, organic matters and micro-organisms.

**Other natural factors:** Most of the canals are narrow and of a low gradient, and therefore they can be easily affected by tides, salt water or alum intrusion.

**3.2.3 Threats to the surface water quality resource**
Due to the awareness that the water quality of rivers and canals in the City has declined recently, threats that negatively affect the water source have been identified as follows (Report of environmental quality monitoring results in 2011, 2012):

**Population growth rate**

In 2011, the total population of the City was about 7.5 million. In fact the natural population growth rate decreased from 1.52% in 1989-1999 to 0.98% in 2011. However, the population itself has increased because the immigration rate has been high. A large number of immigrants from the nearby provinces have gathered in the developing districts such as 2, 7, 9, 12, Thu Duc, Binh Tan, shown in Figure 3.6 (light yellow area). In this area, the population rate has been increasing rapidly, from 15.9% (2001-2005) to 16.2% (2006-2010). In average, this developing area has an increase of around 190,000 people per year, which results in a high pressure on the City’s infrastructure including public transport, health services and urban sanitation.

Besides, according to the Department of Natural Resources and Environment and the Agency of Flooding control of the City, there has been an increase in flooding sites in the City. This has been because many more residential areas have been constructed by filling in the canals and rivers, as shown in Figure 3.10 and 3.11. The canals are also being polluted by domestic and industrial waste (Report of environmental quality monitoring results in 2011, 2012).
Figure 3.10: Constructional (in red) development by time

Figure 3.11: Construction land (in red) in 2011
Domestic waste

In 2012 domestic solid waste discharge was about 6,500 – 7,000 tons/day, equivalent to 7,500 – 8,400 m$^3$, with the development rate of 6%/year on average. The solid waste has been mainly landfilled which has a high potential to pollute the environment. Also, the high volume of untreated industrial waste water yet another risk to the water resources of the City (Report of environmental quality monitoring results in 2011, 2012).

Urban flooding

Flooding has been a serious problem of the City because of its dense water network, heavy rain frequency, tide, and poor designed drainage system (Figure 3.12). Furthermore, it is predicted that when the sea level rises as a result of climate change, new urbanized areas will be severely impacted by flood because of the very low terrain and dense housing construction (Nigel K. Downes, 2014). Flooding also causes secondary issues such as public hygienic problems, lack of fresh water, and transport difficulties.

Poor canal restoration

Although the City has made a big effort to renovate the canal systems, including dredging, restoration of water outlets and building embankments, the Nhieu Loc – Thi Nghe canal renovation project has been considered the
only successful one. The other canal restoration projects have faced many obstacles such as being poorly managed, and market reform. They are therefore of poor quality. (Givental, 2014).

![Image of heavy urban flooding](https://www.dantri.com)

**Figure 3.12: Heavy urban flooding due to very high rainfall on 22 May 2017 (www.dantri.com)**

3.3 Summary

Ho Chi Minh City is one of the highest populated and fastest developing areas of the country. Most of the City’s area is less than two metres above sea level, has a dense canal and river system, and has a rainy season of six
months. With the dense canals and river systems throughout the area, the City has advantages such as having an excellent water supply, a good waterway transportation system, and an effective water drainage system. However, the surface water system of the City has been biologically polluted by industry and a lack of an appropriate domestic waste water treatment system. In the near future, there will be more construction work (buildings, bridges and roads) and more people will migrate to the City. The City will also be highly impacted when the sea level rises as a result of the greenhouse effect. As a result the City will suffer from heavy flooding and their effects such as poor urban sanitation, lack of fresh water and shortage of transportation.
CHAPTER 4

4. WATER QUALITY AND GEOGRAPHICAL DATASETS

This chapter presents details about the dataset applied in the research. In this research, the dataset is classified into two groups (Figure 4.1).

![Diagram of datasets]

**Figure 4.1: Overview of the datasets applied in the research.**

Textural dataset: They include values of quality parameters. In general, they are a set of average annual data of quality parameters which were tested in the laboratory from the monitoring sites in 2015. This information was used in the research to test the actual quality of water in comparison to the quality standard. By applying the two methods which are the Water Quality Index and the Fuzzy Comprehensive Analysis, the water quality of every monitoring site was classified according to the quality groups from the
highest to the lowest quality. Details about the measured dataset of the quality parameters in 2015 is presented in Appendix 1.

**Geographical dataset:** They are a set of geographical information that has been prepared for setting up the watershed area and building the regression models for identifying the relationship between water quality and impact factors such as population, land-use or the number of manufacturers in the study area. They include elevation points, river and canal systems, administrative, land use, and number of all size business information of the study area. This information was gathered in 2015.

**4.1 Surface water quality parameters (Textural dataset)**

**4.1.1 Monitoring sites**

In the first quarter of each year, the City’s Environmental Protection Agency reports the water quality in the previous year to the Department of Natural Resources and Environment. The Department will then report to the City Council and decide on the annual budget for the monitoring program. The Department also recommends the number of parameters and monitoring frequency for the coming year according to the pollution trends that have been shown in the report. After being approved the budget limit by the City’s Council, the Department will then decide on the number and the frequency of the monitored parameters in the program.
Water quality of the area is evaluated under two different conditions: at the highest and lowest tides, according to the time of taking the sample. The water quality of the area is also assessed in groups of physical, chemical, biological, heavy metals, bacteria, oils and nutrients. In 2015, there were 15 monitoring sites for canal and 26 sites for river monitoring. All of the canal sites were concentrated at the City’s centre. The river sites were placed mostly in the main rivers and were dispersed across the whole City. Table 4.1 presents the names and coordination of each site. Also, Figure 4.2 shows the location of the sites with the surface water network of the study area.

### Table 4.1 Details of the monitoring sites of the study area

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of site</th>
<th>Name</th>
<th>Label</th>
<th>X coordination</th>
<th>Y coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Canal</td>
<td>Le Van Sy</td>
<td>LVS</td>
<td>10.78599689</td>
<td>106.681</td>
</tr>
<tr>
<td>2</td>
<td>Canal</td>
<td>Dien Bien Phu</td>
<td>DBP</td>
<td>10.79336056</td>
<td>106.7006028</td>
</tr>
<tr>
<td>3</td>
<td>Canal</td>
<td>An Loc</td>
<td>AL</td>
<td>10.85108778</td>
<td>106.6789253</td>
</tr>
<tr>
<td>4</td>
<td>Canal</td>
<td>Tham Luong</td>
<td>TL</td>
<td>10.82480722</td>
<td>106.6280383</td>
</tr>
<tr>
<td>5</td>
<td>Canal</td>
<td>Hoa Binh</td>
<td>HB</td>
<td>10.77003667</td>
<td>106.6356342</td>
</tr>
<tr>
<td>6</td>
<td>Canal</td>
<td>Ong Buong</td>
<td>OB</td>
<td>10.75447944</td>
<td>106.6367717</td>
</tr>
<tr>
<td>7</td>
<td>Canal</td>
<td>Rach Ngua</td>
<td>RN</td>
<td>10.73252306</td>
<td>106.6192675</td>
</tr>
<tr>
<td>8</td>
<td>Canal</td>
<td>Cha Va</td>
<td>CV</td>
<td>10.74874028</td>
<td>106.6605467</td>
</tr>
<tr>
<td>9</td>
<td>Canal</td>
<td>Phu Dinh</td>
<td>PD</td>
<td>10.73321861</td>
<td>106.6337675</td>
</tr>
<tr>
<td>10</td>
<td>Canal</td>
<td>Nhi Thien Duong</td>
<td>NTD</td>
<td>10.74194667</td>
<td>106.6559922</td>
</tr>
<tr>
<td>11</td>
<td>Canal</td>
<td>Cau so 1</td>
<td>C1</td>
<td>10.79319167</td>
<td>106.6595008</td>
</tr>
<tr>
<td>12</td>
<td>Canal</td>
<td>Nguyen Huu Canh</td>
<td>NHC</td>
<td>10.78749639</td>
<td>106.7093903</td>
</tr>
<tr>
<td>13</td>
<td>Canal</td>
<td>Cau Chu Y</td>
<td>CY</td>
<td>10.75046917</td>
<td>106.6838897</td>
</tr>
<tr>
<td>14</td>
<td>Canal</td>
<td>Cau Mong</td>
<td>CM</td>
<td>10.76802389</td>
<td>106.7037233</td>
</tr>
<tr>
<td>15</td>
<td>Canal</td>
<td>Hai Duc</td>
<td>HD</td>
<td>10.797025</td>
<td>106.6860569</td>
</tr>
<tr>
<td>No.</td>
<td>Type of site</td>
<td>Name</td>
<td>Label</td>
<td>X coordination</td>
<td>Y coordination</td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>-----------</td>
<td>-------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>16</td>
<td>River</td>
<td>Ben Cui</td>
<td>BC</td>
<td>11.27947222</td>
<td>106.3552222</td>
</tr>
<tr>
<td>17</td>
<td>River</td>
<td>Ben Suc</td>
<td>BS</td>
<td>11.15611111</td>
<td>106.45175</td>
</tr>
<tr>
<td>18</td>
<td>River</td>
<td>Thi Tinh</td>
<td>TT</td>
<td>11.04011111</td>
<td>106.6035556</td>
</tr>
<tr>
<td>19</td>
<td>River</td>
<td>Phu Cuong</td>
<td>PC</td>
<td>10.981</td>
<td>106.6433333</td>
</tr>
<tr>
<td>20</td>
<td>River</td>
<td>Hoa An</td>
<td>HA</td>
<td>10.94752778</td>
<td>106.8059444</td>
</tr>
<tr>
<td>21</td>
<td>River</td>
<td>N46 canal</td>
<td>N46</td>
<td>10.94594444</td>
<td>106.5037222</td>
</tr>
<tr>
<td>22</td>
<td>River</td>
<td>Rach Tra</td>
<td>RT</td>
<td>10.91961111</td>
<td>106.6489167</td>
</tr>
<tr>
<td>23</td>
<td>River</td>
<td>Binh Phuoc</td>
<td>BP</td>
<td>10.86180556</td>
<td>106.7168611</td>
</tr>
<tr>
<td>24</td>
<td>River</td>
<td>Sai Gon</td>
<td>SG</td>
<td>10.79883333</td>
<td>106.7271667</td>
</tr>
<tr>
<td>25</td>
<td>River</td>
<td>Phu An</td>
<td>PA</td>
<td>10.77116667</td>
<td>106.7089722</td>
</tr>
<tr>
<td>26</td>
<td>River</td>
<td>Phu My</td>
<td>PM</td>
<td>10.74511111</td>
<td>106.7447222</td>
</tr>
<tr>
<td>27</td>
<td>River</td>
<td>Cat Lai</td>
<td>CL</td>
<td>10.75636111</td>
<td>106.7943056</td>
</tr>
<tr>
<td>28</td>
<td>River</td>
<td>Nha Be</td>
<td>NB</td>
<td>10.67358333</td>
<td>106.7754167</td>
</tr>
<tr>
<td>29</td>
<td>River</td>
<td>Vam Sat</td>
<td>VS</td>
<td>10.48988889</td>
<td>106.7981944</td>
</tr>
<tr>
<td>30</td>
<td>River</td>
<td>Tam Thon Hiep</td>
<td>TTH</td>
<td>10.606</td>
<td>106.8671667</td>
</tr>
<tr>
<td>31</td>
<td>River</td>
<td>Dong Tranh</td>
<td>DT</td>
<td>10.44905556</td>
<td>106.8666944</td>
</tr>
<tr>
<td>32</td>
<td>River</td>
<td>Nga Bay</td>
<td>N7</td>
<td>10.47572222</td>
<td>106.9405</td>
</tr>
<tr>
<td>33</td>
<td>River</td>
<td>Cai Mep</td>
<td>CM</td>
<td>10.50788889</td>
<td>107.0025556</td>
</tr>
<tr>
<td>34</td>
<td>River</td>
<td>Vam Co</td>
<td>VC</td>
<td>10.50788889</td>
<td>107.0025556</td>
</tr>
<tr>
<td>35</td>
<td>River</td>
<td>Thay Cai</td>
<td>TC</td>
<td>10.94186111</td>
<td>106.4535833</td>
</tr>
<tr>
<td>36</td>
<td>River</td>
<td>Binh Dien</td>
<td>BD</td>
<td>10.70161111</td>
<td>106.5973056</td>
</tr>
<tr>
<td>37</td>
<td>River</td>
<td>An Ha</td>
<td>AH</td>
<td>10.79008333</td>
<td>106.437</td>
</tr>
<tr>
<td>38</td>
<td>River</td>
<td>Trung An</td>
<td>TA</td>
<td>11.0485</td>
<td>106.6110833</td>
</tr>
<tr>
<td>39</td>
<td>River</td>
<td>Binh Loi</td>
<td>BL</td>
<td>10.82566667</td>
<td>106.7101667</td>
</tr>
<tr>
<td>40</td>
<td>River</td>
<td>Hoa Phu</td>
<td>HP</td>
<td>10.98647222</td>
<td>106.6193333</td>
</tr>
<tr>
<td>41</td>
<td>River</td>
<td>Phu Long</td>
<td>PL</td>
<td>10.89858333</td>
<td>106.6941111</td>
</tr>
</tbody>
</table>

*(Source: Department of Natural Resources and Environment)*
Figure 4.2: Location of the monitoring sites and the surface water network of the study area
As shown in Table 4.1 and Figure 4.2, there are two types of monitoring sites: rivers and canals. These sites are under the management of, and operated by, the City Environmental Protection Agency. This is a division of the Department of Natural Resources and Environment (Ho Chi Minh City Environmental Protection Agency, 2012).

**Monitoring sites for river water quality**

The river system of the city includes three main rivers Dong Nai, Sai Gon and Vam Co. The monthly recorded value for a quality indicator is the average value of the measured data, at the time of highest and lowest tide. Also, the yearly value of a parameter is the average value of the 12 months of a year; divided into two different sets according to high and low tides.

Since 1993, ten monitoring sites have been constructed and operated to periodically monitor the water quality and hydrological characteristics of the rivers. This number of monitoring sites has been increased to twenty two from 2011 and twenty six in 2015 (red dots in figure 4.1). Details about the river monitoring sites are presented in Table 4.2.

Some sites located in the big rivers, such as the Dong Nai, Sai Gon, or Vam Co River, monitor the river tide levels, and these are also referred to as hydrological sites. A water sample is taken from each site on the 1st, 8th, 15th and 22nd of the month, at both the times of the highest and lowest tides.
during the day. The water sample is then tested at the laboratory for the monitored values.

Table 4.2: River monitoring sites

<table>
<thead>
<tr>
<th>No</th>
<th>Name of site</th>
<th>Label</th>
<th>Year of operation</th>
<th>Hydrological monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phu Cuong</td>
<td>PC</td>
<td>1993</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Binh Phuoc (Sai Gon River)</td>
<td>BP</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>Phu An</td>
<td>PA</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>Hoa An</td>
<td>HA</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>Cat Lai (Dong Nai River)</td>
<td>CL</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>Binh Dien (Cho Dem River)</td>
<td>BD</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>Nha Be</td>
<td>NB</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>Ly Nhon (Nha Be River)-Vam Sat</td>
<td>VS</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>9</td>
<td>Tam Thon Hiep (Dong Tranh River)</td>
<td>TTH</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>10</td>
<td>Vam Co (Vam Co River estuary)</td>
<td>VC</td>
<td>1993</td>
<td>yes</td>
</tr>
<tr>
<td>12</td>
<td>Ben Suc</td>
<td>BS</td>
<td>3/2007</td>
<td>yes</td>
</tr>
<tr>
<td>13</td>
<td>Thi Tinh</td>
<td>TT</td>
<td>3/2007</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>Rach Tra (Sai Gon River)</td>
<td>RT</td>
<td>3/2007</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Thay Cai (Tan Thai)</td>
<td>TC</td>
<td>3/2007</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>An Ha</td>
<td>AH</td>
<td>3/2007</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Kenh Dong (Dong Canal) water facility – N46 canal (a part of Dong Canal)</td>
<td>N46</td>
<td>3/2007</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Dong Tranh Estuary</td>
<td>DT</td>
<td>3/2007</td>
<td>yes</td>
</tr>
<tr>
<td>19</td>
<td>Nga Bay Estuary</td>
<td>N7</td>
<td>3/2007</td>
<td>yes</td>
</tr>
<tr>
<td>20</td>
<td>Cai Mep Estuary</td>
<td>CM</td>
<td>3/2007</td>
<td>yes</td>
</tr>
<tr>
<td>21</td>
<td>Sai Gon (Sai Gon River)</td>
<td>SG</td>
<td>06/2011</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Phu My (Sai Gon River)</td>
<td>PM</td>
<td>06/2011</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Trung An</td>
<td>TA</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Binh Loi</td>
<td>BL</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Hoa Phu</td>
<td>HP</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Phu Long</td>
<td>PL</td>
<td>2015</td>
<td></td>
</tr>
</tbody>
</table>

(Source: Department of Natural Resources and Environment)
**Monitoring sites for canal water quality**

Ten monitoring sites were set up in 2011 for monitoring water quality in the canals throughout 5 main networks in the city. Prior to January 2015 water samples were taken twice in a year: April (dry season) and September (rainy season).

Since January 2015, the number of monitoring sites has been increased to 15 (yellow dots in Figure 4.2). The water samples have been taken four times per year in February, April, September, and November. Monthly and yearly values of a parameter are calculated in a manner similar to those from the river monitoring sites. In this research, laboratory canal water quality data from 2015 has been applied in the model, particularly the average annual value.

**4.1.2 Monitored parameters**

The quality parameters applied for testing the water sample have been referenced from the National Standard for surface water quality (*National technical standard for surface water quality QCVN 08:2008/BTNMT*, 2008). The research used the monitored dataset in 2015, categorized into two sets: at the highest and lowest tide, in order to test the model.
Also, different types of monitoring sites had different tested quality parameters. Chosen parameters were decided upon by the Department of Natural Resources and Environment. Depending on the budget and the quality of water of the previous year, the Department decides on which parameters are tested and which monitoring sites are used for taking samples.

In 2015, 18 parameters were used for testing the water samples taken from the monitoring sites, presented in the Table 4.3. Among these parameters, canal monitoring sites measured 14 parameters and river sites monitored 17 parameters.

Sites located at canals monitored **14 parameters** including: pH, Total suspended solids, Turbidity, Dissolved oxygen, Biological oxygen demand, Chemical oxygen demand, Lead, Cadmium, Copper, Chromium, Total Coliform, Escherichia Coliform, Ammoniacal nitrogen, and Organic phosphate.

River monitoring sites measured **17 parameters** including Temperature, Salinity, pH, Total suspended solids, Turbidity, Dissolved oxygen, Biological oxygen demand, Chemical oxygen demand, Lead, Cadmium, Copper, Manganese, Total Coliform, Escherichia coli form, Ammoniacal nitrogen, Organic phosphate, and Oil.
Table 4.3: Quality parameters in 2015

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Abbreviation</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Temperature</td>
<td>Temp</td>
<td>°C</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Salinity</td>
<td>Sal</td>
<td>g/L</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>pH</td>
<td>pH</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Total suspended solids</td>
<td>TSS</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Turbidity</td>
<td>Turb</td>
<td>NTU*</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Physical characteristics

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Abbreviation</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Dissolved oxygen</td>
<td>DO</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>Biological oxygen demand</td>
<td>BOD$_5$</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Chemical oxygen demand</td>
<td>COD</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Chemical and biological characteristics

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Abbreviation</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Lead</td>
<td>Pb</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Cadmium</td>
<td>Cd</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>Copper</td>
<td>Cu</td>
<td>mg/L</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>Chromium</td>
<td>Cr</td>
<td>mg/L</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>Manganese</td>
<td>Mn</td>
<td>mg/L</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Heavy metal contamination

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Total Coliform</td>
<td>COLI</td>
<td>MPN*/100 mL</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>Escherichia coliform</td>
<td>ECOLI</td>
<td>MPN*/100 mL</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Bacteria indicators

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Ammoniacal nitrogen</td>
<td>NH$_4$-N</td>
<td>mg/L</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>Organic phosphate</td>
<td>PO$_4$-P</td>
<td>mg/L</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Nutrient pollutants

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Unit</th>
<th>Measured in Canal</th>
<th>Measured in River</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Oil</td>
<td>Oil</td>
<td>mg/L</td>
<td>No</td>
</tr>
</tbody>
</table>

Notes:

*: Nephelometric Turbidity Units

**: Most Probable Number
Details about the characteristics of the tested quality parameters are present as follow (Sullivan, Agardy, & Clark, 2005; Tien, 2011; WHO, 1996)

**Physical characteristic**

*Temperature (Temp, °C) and Salinity (Sal, g/L):* These two indicators are basic characteristics of the water environment for living organisms. Also, these parameters relate to other water quality indicators. For example high water temperature can increase the concentration of dissolved oxygen, or another example, salinity is related to the conductivity of the water environment.

*pH:* a measure of the Hydrogen ion concentration in the water, ranging from 0 to +14. This measurement is an indicator of the acceptable water condition for living organisms to survive. A pH level in a range from 6 to 8 is suitable for most aquatic life. A pH of less than 5 means the water is more alkaline and a pH of more than 8 indicates that the water is more acidic. Both ranges are not suitable for aquatic life, such as fish or water plants.

*Total suspended solids (TSS, mg/L):* A measure of the concentration of the undissolved solid in the water, such as particulates. This parameter is closely related to the turbidity. A high TSS reading, high level of undissolved solids, can affect the photosynthesis activity of plants.
Turbidity (Turb, Nephelometric Turbidity Units (NTU)): This indicates the cloudiness of the water sample which is caused by suspended particles or dissolved substances.

**Chemical and biological characteristics**

*Dissolved oxygen (DO, mg/L):* DO is a measure of the saturated concentration of oxygen in the water. This can fluctuate depending on the water’s temperature and movement. The amount of the dissolved oxygen increases with increasing temperature and diffusion. The higher the level of oxygen dissolved in the water, the more suitable the water is for supporting aquatic life.

*Biological oxygen demand (BOD₅, mg/L):* BOD shows the amount of oxygen required for oxidizing biological substances in the water sample. Chemical substances are added to the water sample at the beginning of the testing, the mixture is then rested in a standardized condition of darkness and at room temperature for 3 to 5 days, or more depending on the test. The sample is then tested to determine the result. The higher the amount of consumed oxygen, the higher the biological substances that have been dissolved in the water sample. Therefore, this parameter is an indirect indicator of biological pollution in the water.

*Chemical oxygen demand (COD, mg/L):* COD shows the amount of oxygen required for oxidizing all the chemical substances in the water sample.
Because some of the biological substances can also be oxidized in the test, this indicator is closely related to the BOD parameter. A COD value is usually higher than a BOD value. The COD value is also another indirect indicator for chemical and biological pollution.

**Heavy metal contamination**

*Lead (Pb), Cadmium (Cd), Copper (Cu), Chromium (Cr) and Manganese (Mn) (mg/L):* These heavy metals are toxic for humans. High levels of these heavy metals can cause immediate reactions such as severe nerve trauma or death. Low concentrations of these compounds can remain in the water environment for a long time and they can be accumulated in fish or water plants. If consumed by humans they can cause long term effects such as cancer, mutation or foetal malformation.

**Bacteria indicators**

*Total Coliform (COLI, Most Probable Number MPN/100mL):* This is a group of many kinds of bacteria, and these can easily be found in the water or soil environment. Although they are usually harmless to human health, if the amount of coliform reaches a higher level than the defined upper threshold then this is a sign of the presence of pathogenic organisms which can be a threat to human well-being.
**Escherichia Coliform (ECOLI, Most Probable Number MPN/100mL):** This is one of a group of faecal coliform which are present only in the gut and faeces of warm-blooded animals. Therefore, this parameter is a direct indicator of faecal pollution and pathogenic presence.

**Nutrient pollutants**

**NH$_4$-N and PO$_4$-P (mg/L):** Agricultural activities require nutrients to enrich the soil. The most commonly used nutrients include Nitrogen, Phosphorous, and Potassium (N-P-K) in combination. During the run-off process, these nutrients can dissolve into the water and form the NH$_4^+$ and PO$_4^{3-}$ compounds. These substances are very mobile, they promote eutrophication and algae development which can lead to oxygen deficiency, an increased concentration of organic matter in the water column and encourage microbial growth. The NH$_4^+$ compound can be as volatile as NH$_3$, which is particularly toxic for living organisms.

**Oils and greases**

**Oils (Oil, mg/L):** Oils and greases are the by-products of industry and can be discharged without treatment to the water environment. In the residential areas, especially in the developing cities, cooking oils can also be found in the drainage system and can cause serious problems. They block pipes and prevent the water purifying process at the water treatment plant. Also, to remove them special chemicals have to be added to the water to allow
precipitation which can increase the cost of treatment. In the natural environment, oils prevent sunlight penetration, block air diffusion to the water, and thus can have a detrimental impact on living organisms.

### 4.2 Geographical Dataset

This part discusses details about the topographical dataset applied in the research. They include elevation points, detailed information on administrative data including population and density, and data about economic development such as land-use and number of all size businesses.

#### 4.2.1 Digital Elevation Model (DEM)

More than six hundred thousand points (608,797) around the City were collected in 2000 by the City’s Department of Science and Technology. These points have been applied for elevation measurement in this research. The altitude is measured in metres, above sea level. The average height of the whole area is around 3 metres above sea level, shown in Figure 3.3 and Figure 4.3.

Based on these elevation points, a Triangulated Irregular Network (TIN) has been created, from which a contour map can be generated in Geographic Information System software named ArcGIS. The data can then be
transformed into a DEM. The topography ranges from the highest altitude in the North West to the lowest in the South East.
Figure 4.3: Triangulated Irregular Network of the study area
4.2.2 Surface water network

Digitised stream lines for the City created in 2015 by the City’s Department of Science and Technology were used in the model, as shown in Figure 3.9. The stream network is dense, with most of the area being covered by canal lines, except in some of the urbanized areas where they have been removed. This removal of canals has occurred mostly in the centre of the City. The DEM was then adjusted to take into account the streamline network of river and canal systems. The DEM was then used to build the watershed delineation map.

4.2.3 Administrative information

The City’s area is approximately 2,109 km², consisting of 24 districts, in which the central districts are small in size, whilst the sub-urban areas such as Thu Duc, Can Gio, Hoc Mon, are much larger. The population is concentrated in the smaller districts in the central areas. There were about 8 million people in the city in 2015 with an average population density of approximate 20,000 people per km², shown in Table 3.1 and Figure 3.4, 3.5 and 3.6. The highest population concentration is on the northern and eastern edges of the City. The highest density is in the centre of the City where there is also the highest number of the manufacturing industries.
4.2.4 Land use information

In 2015, most of the central areas of the City were residential. The northern and eastern areas were used for short term annual agriculture such as rice, flowers, or vegetable fields, as shown in Figure 3.7. Table 4.3 shows the percentage of land-use in the study area in 2015, organized from the highest to the lowest percentage of land-use.

Agriculture occupied the largest part of the City’s area, mostly in the north and west of the city. Short term agriculture and horticulture such as vegetables, flowers, rice, and aqua agriculture occurred in the North, and farms for salt harvesting were found in the South of the City.

Table 4.4: Percentage of Land-use purposes

<table>
<thead>
<tr>
<th>No</th>
<th>Land-use Types</th>
<th>Area (m²)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agriculture</td>
<td>770,639,067</td>
<td>45.7</td>
</tr>
<tr>
<td>2</td>
<td>Natural Zone</td>
<td>345,662,995</td>
<td>20.5</td>
</tr>
<tr>
<td>3</td>
<td>Urban Area</td>
<td>236,314,333</td>
<td>14.0</td>
</tr>
<tr>
<td>4</td>
<td>All size manufacturers</td>
<td>210,761,692</td>
<td>12.5</td>
</tr>
<tr>
<td>5</td>
<td>Governmental Area</td>
<td>75,706,475</td>
<td>4.5</td>
</tr>
<tr>
<td>6</td>
<td>Commerce and other social facilities</td>
<td>22,840,911</td>
<td>1.4</td>
</tr>
<tr>
<td>7</td>
<td>Waste treatment &amp; Graveyard</td>
<td>13,388,764</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>Reserved Area</td>
<td>7,819,406</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>Relic and Tourism Area</td>
<td>5,576,334</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>Other purposes</td>
<td>2,389,915</td>
<td>0.1</td>
</tr>
</tbody>
</table>
4.3 Summary

Two types of datasets were used in the research: monitored values of the water quality and the GIS’s dataset.

The monitored dataset of the quality parameters was used for testing the overall quality of the water, applying Fuzzy Comprehensive analysis and Water Quality Index methods.

The geographical dataset was used for testing the applicability of the water quality results. Because the study focuses on the water quality of the urban developing area, the information of human influences was collected and used for testing the relation to the water quality. These influences include population density, land-use, number of all size businesses and the sub-watershed area created by ArcHydro.
CHAPTER 5

5. INTEGRATED ASSESSMENT OF SURFACE WATER QUALITY

In general, the quality parameters for assessment are applied to identify the physical, chemical and biological characteristics of the water sample. Details about the quality parameters were outlined in Chapter 2, part 2.3: Water quality indicator and the monitored parameters sampled in 2015 were discussed in Chapter 4, part 4.2: Surface water quality parameters.

This chapter presents the methods used for surface water quality assessment as applied to the study. Two methods are outlined in this chapter. The first one is the Fuzzy Integrated Assessment method, which is considered the main method used for data analysis in the research. The second method is the Water Quality Index, a standard method for assessing the water quality in the study area. These methods are applied to determine the water quality based on the monitored values of different quality parameters.
5.1 Standard values used for quality parameters and the 2015 applied dataset

Water quality regulations were presented in *Chapter 2, part 2.4 Water quality regulations*. Information about the water quality dataset (2015) was also described in *Chapter 4, part 4.2 Surface water quality parameters*. This section discusses the applied standard values for assessment of the two aforementioned methods.

5.1.1 Quality standards

Established in 2008, Vietnam’s National Standard for Surface Water Quality, *(National technical standard for surface water quality QCVN 08:2008 BTNMT, 2008)*, consists of 32 quality parameters, and it is the official standard for determining the quality of the surface water in Vietnam. This standard provides the benchmark for the assessment of surface water quality in this study. According to this standard, water quality is classified into 4 groups of quality from the highest (A1, A2) to the lowest (B1 and B2). Furthermore, depending on the use for the water, Vietnam’s other standards are also applied for setting the thresholds of the quality parameters, such as:

Industrial wastewater quality (National Technical Regulation on Industrial Wastewater QCVN 40:2011/ BTNMT, 2011): Established in 2011, this standard regulates the highest acceptable pollution level of the wastewater of industries in general when the wastewater is discharged.

Water quality standard for drinking (National technical regulation on drinking water quality QCVN 01:2009/BYT, 2009): Established in 2009, this standard regulates the water quality for the output of water supply facilities, food processing facilities use and household use for drinking and cooking.

Water quality standard for domestic use (National technical regulation on domestic water quality QCVN 02:2009/BYT, 2009): Established in 2009, this regulation is applied to water that is used for domestic purposes, but not including direct drinking and food processing.


The limit in the annual budget for water quality assessment in Vietnam means that not all the standard parameters are measured. In general, there were 18 parameters applied for on-site measurement in 2015. These
included: Biological oxygen demand, Chemical oxygen demand, Coliform, pH, Total suspended solid, Dissolved oxygen, Oil and grease concentration, E-Coliform, Lead, Cadmium, Chromium, Copper, Ammonia Nitrogen concentration, Phosphate Phosphorous concentration, Turbidity, Salinity, Manganese and Temperature.

Using these parameters, 17 parameters were chosen for tested their values at the river sites. Samples taken from the canals were used for testing 14 parameters.

Among the 18 monitored parameters except for Temperature, 15 parameters are regulated in the National standard for surface water quality: QCVN 08:2008 BTNMT. The other parameters, including Turbidity, Salinity and Manganese relate to other water quality standards as listed below:

- Water quality standard for drinking (QCVN 01:2009/BYT),
- Water quality standard for domestic use (QCVN 02:2009/BYT),
- Water supply regulation: Distribution and Facilities design (TCVN 33-2006).

Parameters must be referred to more than two standards before being set up as thresholds for each quality group. This is because there is no specific guideline appropriate to the quality level A1, A2, B1, and B2 for these
parameters. Table 5.1 presents details about standard levels for all the monitored parameters.

**Table 5.1 Recommended surface water quality standard levels for 17 monitored parameters (not including Temperature)**

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameters</th>
<th>Label</th>
<th>Unit</th>
<th>Standard quality group</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Biological oxygen demand</td>
<td>BOD</td>
<td>mg/L</td>
<td>4</td>
<td>6</td>
<td>15</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Chemical oxygen demand</td>
<td>COD</td>
<td>mg/L</td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Coliform</td>
<td>COLI</td>
<td>MPN/100m</td>
<td>20</td>
<td>50</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>pH</td>
<td>pH</td>
<td>-</td>
<td>6-8.5</td>
<td>6-8.5 or 8.5-9</td>
<td>&lt;5.5 or &gt;9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Total suspended solid</td>
<td>TSS</td>
<td>mg/L</td>
<td>20</td>
<td>30</td>
<td>50</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Dissolved oxygen</td>
<td>DO</td>
<td>mg/L</td>
<td>&gt;=6</td>
<td>&gt;=5</td>
<td>&gt;=4</td>
<td>&gt;=2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Oil and grease concentration</td>
<td>OIL</td>
<td>mg/L</td>
<td>0.01</td>
<td>0.02</td>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>E-coliform</td>
<td>ECOLI</td>
<td>MPN/100m</td>
<td>20</td>
<td>50</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Lead</td>
<td>Pb</td>
<td>mg/L</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Cadmium</td>
<td>Cd</td>
<td>ug/L</td>
<td>0.005</td>
<td>0.005</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chromium</td>
<td>Cr&lt;sup&gt;6+&lt;/sup&gt;</td>
<td>mg/L</td>
<td>0.010</td>
<td>0.020</td>
<td>0.040</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Cooper</td>
<td>Cu</td>
<td>mg/L</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Ammonia Nitrogen concentration</td>
<td>NH&lt;sub&gt;4&lt;/sub&gt;_N</td>
<td>mg/L</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Phosphate</td>
<td>PO&lt;sub&gt;4&lt;/sub&gt;_P</td>
<td>mg/L</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

**Applied Domestic water quality (QCVN 02:2009/BYT) and Water supply regulation (TCVN 33-2006)**

<table>
<thead>
<tr>
<th>No.</th>
<th>Turbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Turb</td>
</tr>
</tbody>
</table>

**Applied Water supply regulation (TCVN 33-2006), Drinking quality standard (QCVN 01:2009/BYT) and domestic use standard (QCVN 02:2009/BYT)**
### Table 5.2

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameters</th>
<th>Label</th>
<th>Unit</th>
<th>Standard quality group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>16</td>
<td>Salinity</td>
<td>Sal</td>
<td>mg/L</td>
<td>&lt;250</td>
</tr>
</tbody>
</table>

**Applied Water supply regulation (TCVN 33-2006) and Drinking quality standard (QCVN 01:2009/BYT)**

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameters</th>
<th>Label</th>
<th>Unit</th>
<th>Standard quality group</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Manganese</td>
<td>Mn</td>
<td>mg/L</td>
<td>&lt;=0.2</td>
</tr>
</tbody>
</table>

#### 5.1.2 Applied dataset

The data as applied is the average annual data from canal and river monitoring sites in 2015. The data is classified into two sets: at the highest and lowest tide of the day the sample was taken. The dataset includes the quality values for 18 parameters tested in the laboratory from 15 canal and 26 river sites. Details about these monitoring sites were presented in *Chapter 4, part 4.2.1 Monitoring sites*. Table 5.2 shows the measured values from 17 parameters for the Ben Cat river site, as an example of the monitoring dataset.
### Table 5.2 Monitored values at Ben Cat site (Dong Nai River)

<table>
<thead>
<tr>
<th>No</th>
<th>Parameters</th>
<th>Label</th>
<th>Unit</th>
<th>At high tide</th>
<th>At low tide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Biological oxygen demand</td>
<td>BOD₅</td>
<td>(mg/l)</td>
<td>2.815</td>
<td>2.68</td>
</tr>
<tr>
<td>2</td>
<td>Chemical oxygen demand</td>
<td>COD</td>
<td>(mg/l)</td>
<td>4.575</td>
<td>4.589</td>
</tr>
<tr>
<td>3</td>
<td>Coliform</td>
<td>COLI</td>
<td>MPN/100ml</td>
<td>3,339.91</td>
<td>2,564.25</td>
</tr>
<tr>
<td>4</td>
<td>pH</td>
<td>pH</td>
<td></td>
<td>6.27</td>
<td>5.98</td>
</tr>
<tr>
<td>5</td>
<td>Total suspended solid</td>
<td>TSS</td>
<td>(mg/l)</td>
<td>54.026</td>
<td>48.071</td>
</tr>
<tr>
<td>6</td>
<td>Turbidity</td>
<td>Turb</td>
<td>(N.T.U)</td>
<td>12.919</td>
<td>8.48</td>
</tr>
<tr>
<td>7</td>
<td>Dissolved oxygen</td>
<td>DO</td>
<td>(mg/l)</td>
<td>4.126</td>
<td>3.743</td>
</tr>
<tr>
<td>8</td>
<td>Lead</td>
<td>Pb</td>
<td>(mg/l)</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>9</td>
<td>Cadmium</td>
<td>Cd</td>
<td>(ug/l)</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>Cooper</td>
<td>Cu</td>
<td>(mg/l)</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>11</td>
<td>Oil</td>
<td>OIL</td>
<td>(mg/l)</td>
<td>0.018</td>
<td>0.017</td>
</tr>
<tr>
<td>12</td>
<td>E.Coli</td>
<td>ECOLI</td>
<td>MPN/100ml</td>
<td>6,233</td>
<td>10,624.4</td>
</tr>
<tr>
<td>13</td>
<td>Salinity</td>
<td>Sal</td>
<td>(g/l)</td>
<td>0.046</td>
<td>0.052</td>
</tr>
<tr>
<td>14</td>
<td>Manganese</td>
<td>Mn</td>
<td>(mg/l)</td>
<td>0.035</td>
<td>0.037</td>
</tr>
<tr>
<td>15</td>
<td>Chlorine</td>
<td>Chlorine</td>
<td>(mg/l)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>Ammonia Nitrogen concentration</td>
<td>NH4-N</td>
<td>(mg/l)</td>
<td>0.651</td>
<td>0.709</td>
</tr>
<tr>
<td>17</td>
<td>Phosphate Phosphorous concentration</td>
<td>PO4-P</td>
<td>(mg/l)</td>
<td>0.204</td>
<td>0.1</td>
</tr>
<tr>
<td>18</td>
<td>Temperature</td>
<td>Temp</td>
<td>°C</td>
<td>28.74</td>
<td>29.146</td>
</tr>
</tbody>
</table>

### 5.2 Water quality assessment processes

This section discusses two main methods of assessing water quality using the monitored dataset. Taking the monitored values at the Ben Cat site (Table 5.2) as an example, all the numeric values must be processed to determine whether the water in the sample site, Ben Cat at Dong Nai River, is in the
group of the highest quality (A1 or A2), or needs to have further treatment which places it in the lowest quality levels B1 or B2.

5.2.1 Vietnam’s Water Quality Index

This section presents the Water Quality Index (WQI) method of assessment. WQI is an experimental function developed individually for a specific region. WQI’s method applied in this research is based on the instruction of the Handbook of WQI’s application in Vietnam, promulgated by Vietnam Environmental Agency in 2011. This method is widely applied in the study area and the results of using this method are presented in the official environmental report, and used for reference for the City’s Council.

This method is comprised of a group of formulas, calculating nine common parameters: Biological oxygen demand (BOD), Chemical oxygen demand (COD), Coliform, Total suspended solid (TSS), Conductivity, N-NH₄, P-PO₄, Dissolved oxygen (DO), and pH. The general calculation is:
\[
WQI = \frac{WQI_{pH}}{100} \left[ \frac{1}{5} \sum_{a=1}^{5} WQI_a \times \frac{1}{2} \sum_{b=1}^{2} WQI_b \times WQI_c \right]^{1/3}
\]

(Equation 4)

Source: Handbook of WQI’s application in Vietnam, 2011

where:

WQI\textsubscript{a}: WQI results of DO, BOD\textsubscript{5}, COD, N-NH\textsubscript{4}, P-PO\textsubscript{4}

WQI\textsubscript{b}: WQI results of TSS and Conductivity

WQI\textsubscript{c}: WQI results of Coliform

WQI\textsubscript{pH}: WQI result of pH.

(Each of WQI\textsubscript{a}, WQI\textsubscript{b}, WQI\textsubscript{c}, WQI\textsubscript{pH} has its different calculation shown in the Handbook)

The overall WQI’s result is a number ranging from 0 to 100, which means the higher the value the better the environment. According to the Handbook, the corresponding quality of water is described in Table 5.3. Also, because the results of the WQI’s method is applied as a mean for comparison to the other method, the Fuzzy Comprehensive Evaluation method, equivalent quality groups are set for each range of WQI’s values. The lowest quality group B2 is divided into two sub-set: upper and lower B2 for the purpose of defining the level of pollution of the water sample.
Table 5.3: Quality assessment based on the WQI’s result

<table>
<thead>
<tr>
<th>WQI’s value</th>
<th>Level of quality</th>
<th>Classified into quality group</th>
</tr>
</thead>
<tbody>
<tr>
<td>91 - 100</td>
<td>Good for domestic supply</td>
<td>A1</td>
</tr>
<tr>
<td>76 - 90</td>
<td>Applicable for domestic supply with suitable treatment.</td>
<td>A2</td>
</tr>
<tr>
<td>51 - 75</td>
<td>Applicable for irrigation and other relevant purposes.</td>
<td>B1</td>
</tr>
<tr>
<td>26 - 50</td>
<td>Applicable for water transport and other relevant purposes.</td>
<td>B2+</td>
</tr>
<tr>
<td>0 - 25</td>
<td>Heavily polluted and need further treatment.</td>
<td>B2-</td>
</tr>
</tbody>
</table>

5.2.2. Integrated fuzzy elevation model

Overview of the method

In general, the model tests all the quality parameters and groups the monitoring sites into four quality sets according to the Vietnam standard for surface water quality: A1, A2, B1, B2. This is equivalent to the highest to the lowest quality. Figure 5.1 shows the outline of the method’s purpose.
In Figure 5.1, \( V = (A_1, A_2, B_1, B_2) \) is considered the evaluation set, which can be understood as being \( V = \) (Very good, Good, Moderate, Bad) quality group. Each quality group \( A_1, A_2, B_1 \) and \( B_2 \) has its own thresholds according to the reference standards. For example Table 5.1 presents details about the different threshold value of 18 monitored parameters.

The fuzzy model is built on 2 main parts: the matrix and the weight. Figure 5.2 shows the overview of the assessment method.
Figure 5.2: Overview of the fuzzy assessment process

Fuzzy matrix

One important part of the model is the Fuzzy matrix or R-matrix. In this research, this matrix is initially built on the simplest functions triangular and trapezoidal shapes. These are the simplest functions, and they have been proven to be appropriate for environmental quality assessment.
Weights

Weight is the level of importance, ranging from 0 to 1, or 0% to 100%. Testing the quality of a water body, particularly at a monitoring site, needs to monitor a number of quality parameters, for example, there were 18 parameters tested at each monitoring site in the study area. Therefore, as an important part of fuzzy comprehensive analysis, these quality parameters must be assigned the weights for calculation. In this research, three different weighting methods have been used for testing, they include: Entropy, F-statistic, and Overweight. Each method of assigning weight has its strength and weakness. Therefore, depending on the purpose of analysis, characteristic of the datasets, and the study area, possible weighting methods can be used for more precise fuzzy evaluation.

**Overweight weighting method:** The higher the sample's mean as compared to the standard value’s mean, the higher the weight. The weight of a parameter is the ratio between the mean of the samples and the mean of the standard values of that parameter. (Shen Jihong, 2011, Zhou Zhen-min, 2011, Shufei Lin, 2011, Ji-hong Zhou, 2009)

**Entropy weighting method:** Entropy is a physical concept in a thermal dynamic system for defining the system’s turbulence. Therefore, a group of values of one parameter which has the highest diversion within the group will gain the highest weight.
If all the members of a set are of equal value, for example Mercury concentration measured each month being 0.02 mg/l for a whole year, the weight is 0. Therefore, the weight of a parameter is identified by the entropy value of each parameter (sum of natural logarithm (Ln) of all the monitored values of that parameter) (Zhenxiang Xing, 2011, Donghua Wang, 2010, Liu Jun, 2006)

_F-statistic weighting method:_ Higher weights will be assigned to the parameters that have a smaller standard deviation compared to the whole group of all parameters. This means that the sample’s values within that group are centralized to the mean value. The weight of a parameter is therefore identified by the F-ratio (F-statistic). F-ratio is the ratio between the sum of the square standard deviation between groups (groups of parameters) and the sum of the square standard deviation within a group (group of values for one parameter)(L. Ma, 2010)

In general, the calculation of the Fuzzy matrix involves three steps:

_Step 1: Setting up the membership function for each parameter:_ The function applied in this trial model is the simplest one, the semi-trapezoidal graph, then comprising the matrix.
Step 2: Setting up the weighting vector for each of the parameters: Several methods are applied to test the results, including Overweight standard, Entropy, and F-statistic, all of these are based on the statistical characteristics of the dataset. Among these approaches, the over-weight standard method is considered to be the most appropriate.

Step 3: Multiply the fuzzy matrix and the weighting vector to get the final result and then compare the result to the results of the Water Quality Index.

Details about the conducting steps are discussed below:

Step 1: Membership and matrix build-up

Membership build-up

For each parameter, the membership function is based on the standard limit. In recent research, the membership function applied for environmental evaluation is a mixed type of linear lower semi-trapezoidal and triangular distribution (Runsheng LV, 2011, Shufei Lin, 2011, Zhenxiang Xing, 2011, Shen Jihong, 2011, Liyun Yang, 2010, L.Ma, 2010).

General formulas for these memberships are presented in Figure 5.3:
For Equation 5 and 6, where:

- $f_a(x)$: the membership value of a specific parameter at a monitored value $x$
- $x$: monitored value
- $a_1, b_1, a_2, b_2, a, b$: threshold values of a quality parameter
- $m$: exact value of the quality parameter that receives the highest membership value, 1.

**Figure 5.3: Semi-trapezoidal and triangular graphs of membership function**
The shape of the membership function of the fuzzy comprehensive analysis is then set as followed:

![General membership function of the model](image)

**Figure 5.4: General membership function of the model**

Where:

- $c_i$: monitored value at one site
- $f(c_i)$: membership function value at a specific $c_i$
- $a_1, a_2, a_3, a_4$: standard thresholds from the lowest to the highest.
- $V1$, $V2$, $V3$, $V4$: Equivalent evaluation standard level Excellent, Good, Moderate, Bad, equivalent to quality group A1, A2, B1, B2.

Membership function for each of the standard levels as follows:
\[ f_A(c_i) = \begin{cases} 
1, & c_i < a_1 \\
\frac{a_2 - c_i}{a_2 - a_1}, & c_i \in [a_1, a_2] \\
0, & c_i > a_2 
\end{cases} \] (Equation 7)

Membership function for standard \textit{V1} (Lower semi-trapezoidal shape)

\[ f_A(c_i) = \begin{cases} 
0, & c_i \leq a_3 \\
\frac{c_i - a_3}{a_4 - a_3}, & c_i \in [a_3, a_4] \\
1, & c_i > a_4 
\end{cases} \] (Equation 8)

Membership function for standard \textit{V4} (Lower semi-trapezoidal shape)

\[ f_A(c_i) = \begin{cases} 
0, & c_i < a_1 \\
\frac{c_i - a_1}{a_2 - a_1}, & c_i \in [a_1, a_2] \\
\frac{a_3 - c_i}{a_3 - a_2}, & c_i \in [a_2, a_3] \\
0, & c_i > a_3 
\end{cases} \] (Equation 9)

Membership function for standard \textit{V2} (Triangular shape)
\[
 f_A(c_i) = \begin{cases} 
 0, & c_i \leq a_2 \\
 \frac{c_i - a_2}{a_3 - a_2}, & c_i \in [a_2, a_3] \\
 \frac{a_4 - c_i}{a_4 - a_3}, & c_i \in [a_3, a_4] \\
 0, & c_i > a_4 
\end{cases}
\]

(Equation 10)

Membership function for standard V2 (Triangular shape)

For Equation 7 to 10, where:
\( f_a(c_i) \): the membership value of a specific parameter at a monitored value \( x \)
\( c_i \): monitored value
\( a_1, a_2, a_3, a_4 \): threshold values of a quality parameter

In short, the calculation of one monitored parameter at one site based on four levels of standard V is presented in Table 5.4:

**Table 5.4: Membership value \( f(c_i) \) according to the quality group**

<table>
<thead>
<tr>
<th>( c_i )</th>
<th>( f(c_i) ) value of group A1</th>
<th>( f(c_i) ) value of group A2</th>
<th>( f(c_i) ) value of group B1</th>
<th>( f(c_i) ) value of group B2</th>
<th>( f(c_i) + f(c_i) ) value of group B3 and B4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_i \leq a_1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( a_1 \leq c_i \leq a_2 )</td>
<td>( \frac{a_2 - c_i}{a_2 - a_1} )</td>
<td>( \frac{c_i - a_1}{a_2 - a_1} )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( a_2 \leq c_i \leq a_3 )</td>
<td>0</td>
<td>( \frac{a_3 - c_i}{a_3 - a_2} )</td>
<td>( \frac{c_i - a_2}{a_3 - a_2} )</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( a_3 \leq c_i \leq a_4 )</td>
<td>0</td>
<td>0</td>
<td>( \frac{a_4 - c_i}{a_4 - a_3} )</td>
<td>( \frac{c_i - a_3}{a_4 - a_3} )</td>
<td>1</td>
</tr>
<tr>
<td>( c_i &gt; a_4 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Where:

$c_i$: monitored value at the monitoring site

$a_1, a_2, a_3, a_4$: standard’s threshold, from the lowest to the highest value

$V_1, V_2, V3, V4$: quality group, from the highest to the lowest quality.

$f(c_i)_1, f(c_i)_2, f(c_i)_3, f(c_i)_4$: fuzzy membership value of each quality group.

For one monitored value, there could be a possibility of two membership values relating to the level of standard. For example, in Table 5.4 with $(a_1 \leq c_i \leq a_2)$ there are two membership values, included: in the group of $V_1$; and in the group of $V_2$. However, the summation of membership values of every $c_i$ is always equal to 1.

Taking the BOD parameter as an example, the thresholds for each of the standard levels is shown in Table 5.5:

**Table 5.5: Standard thresholds for Biological Oxygen Demand (BOD)**

<table>
<thead>
<tr>
<th>Standard group</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_1 = 4$</td>
<td>$a_2 = 6$</td>
<td>$a_3 = 15$</td>
<td>$a_4 = 25$</td>
</tr>
</tbody>
</table>

Applying the calculation in Table 5.4, if the monitored value of BOD is 18 mg/L, $c_i$ (BOD)=18, there are two membership values $f_{B2}(c_{BOD}) = 0.3$ and $f_{B1}(c_{BOD}) = 0.7$. This means $c_i$ (BOD)=18 that is 30% belonging to group B2.
(30%) and 70% belonging to group B1 (70%). Summation of the membership values for every $c_i$ is 1 (100%).

In this study, according to the standard thresholds from $a_1$ to $a_4$, which $a_1 < a_2 < a_3 < a_4$, graphs of membership can be classified into four groups:

- Group having 4 limits (thresholds) $a_1$, $a_2$, $a_3$, $a_4$: including 10 parameters: Biological oxygen demand (BOD), Chemical oxygen demand (COD), Total suspended solid (TSS), Oil, Coliform, E.Coli, Cooper (Cu), Dissolved oxygen (DO), Ammoniacal nitrogen (NH4-N), and Organic phosphate (PO4-P);
- Group having 3 limits (thresholds) $a_1$, $a_2$, $a_3$: three parameters Turbidity, Salinity, and Chlorine;
- Group having 2 limits (thresholds) $a_1$, $a_2$: three parameters Lead (Pb), Cadmium (Cd), Chromium ($\text{Cr}^{6+}$);
- Group having a range of thresholds: Two parameters pH and Manganese.

Figure 5.5 to 5.9 show the membership shapes for each group.
Figure 5.5: Membership shape for group of 4 limits in the standard level

Figure 5.6: Membership shape for group of 3 limits

Figure 5.7: Membership shape for group of 2 limits
Where:
c\_i: monitored value at 1 site
f(c\_i): membership function value at a specific c\_i
a\_1, a\_2, a\_3, a\_4: standard thresholds from the lowest to the highest value.
A1, A2, B1, B2: Quality group, equivalent to Very good, Good, Moderate and Bad quality.
Fuzzy matrix set up $R_{nxm}$

After setting up the membership functions and calculating the membership value for each parameter at every monitoring site, the fuzzy matrix is then built for every monitoring site applying the membership value of all the monitored parameters.

$$R_{nxm} = \begin{pmatrix}
    r_{11} & \cdots & r_{1m} \\
    \vdots & \ddots & \vdots \\
    r_{n1} & \cdots & r_{nm}
\end{pmatrix}
$$

(Equation 11)

Where:

- $n$ (row): number of the evaluation parameters ($u_i$) (18 parameters)
- $m$ (column): number of standard levels ($v_j$)
- $r_{ij}$: $f(c_{ij})$ value

Summation of the matrix row is equal to 1:

$$\sum_{j=1}^{m} r_{i,j} = 1$$

(Equation 12)
Taking the monitored value of 17 parameters at Phu Long, which is a river monitoring site in Sai Gon River, as an example for building the matrix, the fuzzy matrix is built as in Table 5.6. It is also noted that the data was the average values of monthly monitored data in 2015 and taken at the lowest tide.

**Table 5.6: Fuzzy matrix built for Phu Long site**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Monitored value</th>
<th>Membership value of group A1</th>
<th>Membership value of group A2</th>
<th>Membership value of group B1</th>
<th>Membership value of group B2</th>
<th>Sum of membership values of all group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BOD₅</td>
<td>3.857</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2. COD</td>
<td>6.641</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3. COLI</td>
<td>124,152.308</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4. pH</td>
<td>6.311</td>
<td>0.31</td>
<td>0.69</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5. TSS</td>
<td>80.077</td>
<td>0.00</td>
<td>0.00</td>
<td>0.40</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>6. Turb</td>
<td>24.597</td>
<td>0.00</td>
<td>0.35</td>
<td>0.65</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7. DO</td>
<td>3.042</td>
<td>0.48</td>
<td>0.52</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>8. Pb</td>
<td>0.005</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>9. Cd</td>
<td>0.001</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10. Cu</td>
<td>0.005</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>11. OIL</td>
<td>0.023</td>
<td>0.00</td>
<td>0.96</td>
<td>0.04</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>12. ECOLI</td>
<td>63,993.077</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>13. Salinity</td>
<td>0.403</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>14. Mn</td>
<td>0.034</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>15. Chlorine</td>
<td>0</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>16. NH4-N</td>
<td>0.671</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
<td>0.34</td>
<td>1.00</td>
</tr>
<tr>
<td>17. PO4-P</td>
<td>0.068</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Step 2: Weighting vector set up

The purpose of this step is to identify the weight of each calculated parameter. Three different approaches are applied in the model, including Over-weight standard, F-statistic, and Entropy methods.


This method is used to calculate the weight of each parameter based on the actual values of the monitored samples. This is based on the principle “the more/less, the better”.

For the parameters such as Biological or Chemical oxygen demand (BOD or COD), the higher the value of BOD or COD the higher the pollution level. Therefore these parameters are applied the principle of “the less the better”. In the contrary, some parameters such as DO, the higher the value the better the environment, thus the principle of “the more the better” can be applied in such case.

The calculation is as followed:
With the principle “the less the better”:
\[
W_i = \frac{c_i}{S_i} \quad \quad S_i = \frac{1}{m} \sum_{j=1}^{m} s_{i,j}
\]

(Equation 13)

With the principle “the more the better”:

\[
W_i = \frac{S_i}{c_i}
\]

(Equation 14)

Where:

- \( w_i \): weight of the \( i \)th parameter (un-standardized)
- \( c_i \): mean of the monitored values of the \( i \)th parameter
- \( S_i \): mean of all the standard limits
- \( m \): number of monitored values of one parameter.

Standardize the \( w_i \):

\[
\overline{W}_i = \frac{1}{n} \sum_{i=1}^{n} w_i
\]

(Equation 15)

Where:

- \( n \): number of \( w_i \) of the \( i \)th parameter
**Entropy method** (Donghua Wang, 2010; Liu Jun, 2006; Zhenxiang Xing, 2011)

According to the method, primary matrix $X = (x_{ij})_{m \times n}$ $(i = 1, \ldots, m; j = 1, \ldots, n)$ is built, where:

$m$: number of monitoring sites, starting from $i = 1$ to $m$.

$n$: number of criteria (parameters), starting from $j = 1$ to $n$.

$x_{ij}$ the monitored value of the evaluation parameters

After the matrix $X$ is built, standardize the column of the primary matrix $X$ to the establish the matrix $Y$: $Y = (y_{ij})_{m \times n}$, with:

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}$$

(Equation 16)

<table>
<thead>
<tr>
<th>$u_1$ (pH)</th>
<th>$u_2$ (COD)</th>
<th>...</th>
<th>$u_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_1$</td>
<td>$x_{11}$</td>
<td>$x_{12}$</td>
<td>...</td>
</tr>
<tr>
<td>$Q_2$</td>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$Q_m$</td>
<td>$x_{m1}$</td>
<td>$x_{m2}$</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$y_{11}$</th>
<th>$y_{12}$</th>
<th>...</th>
<th>$y_{1n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_{21}$</td>
<td>$y_{22}$</td>
<td>...</td>
<td>$y_{2n}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>$y_{m1}$</td>
<td>$y_{m2}$</td>
<td>...</td>
<td>$y_{mn}$</td>
</tr>
</tbody>
</table>

X matrix

Y matrix
Steps include:

Identify the entropy value of each parameter:  
\[
e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} y_{ij} \ln y_{ij}
\]
(Equation 17)

Identify the weight of the j\textsuperscript{th} parameter:  
\[
a_j = \frac{1-e_j}{\sum_{j=1}^{n} (1-e_j)}
\]
(Equation 18)

Identify the comprehensive weight of the j\textsuperscript{th} parameter:  
\[
w_j = \sum_{j=1}^{n} a_j y_{ij}
\]
(Equation 19)

From Equation 17 to 19, where:

- m: number of monitoring sites, starting from i = 1 to m.
- n: number of criteria (parameters), starting from j =1 to n.
- y\textsuperscript{ij}: standardized x\textsuperscript{ij} which is the monitored value of the evaluation parameters

**Statistical F method (L.Ma, 2010)**

F – ratio is the ratio between the square standard deviation of each group and the square standard deviation within the group.

\[
F = \frac{S_A^2}{S_E^2} = \frac{Q_A / (r - 1)}{Q_E / (n - r)}
\]
(Equation 20)
Where

$S_A^2$: Square standard deviation of each of the group

$S_E^2$: Square standard deviation within the groups

$Q_A$: Deviation square sum in group

$Q_E$: Deviation square sum within the groups

$r$: the number of groups

$n$: the number of member in each group

Higher F shows the higher differences between groups. This means higher concentration within a group, and shows better differentiation among groups. Lower F shows the higher differences within a group.

Taking the monitored value at Phu Long, a river monitoring site in Sai Gon River, which is mentioned in Table 5.6, the over-weight weighting values for the monitored parameters are set in Table 5.7.
Table 5.7: Weights of the monitored parameters applying overweight standard method for Phu Long monitoring site

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Weight</th>
<th>Parameters</th>
<th>Weight</th>
<th>Parameters</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BOD5</td>
<td>0.0008</td>
<td>7. DO</td>
<td>0.0027</td>
<td>13. Salinity</td>
<td>0.0000</td>
</tr>
<tr>
<td>2. COD</td>
<td>0.0007</td>
<td>8. Pb</td>
<td>0.0004</td>
<td>14. Mn</td>
<td>0.0004</td>
</tr>
<tr>
<td>3. COLI</td>
<td>0.0881</td>
<td>9. Cd</td>
<td>0.0004</td>
<td>15. Chlorine</td>
<td>0.0000</td>
</tr>
<tr>
<td>4. pH</td>
<td>0.0029</td>
<td>10. Cu</td>
<td>0.0000</td>
<td>16. NH4-N</td>
<td>0.0034</td>
</tr>
<tr>
<td>5. TSS</td>
<td>0.0074</td>
<td>11. OIL</td>
<td>0.0007</td>
<td>17. PO4-P</td>
<td>0.0006</td>
</tr>
<tr>
<td>6. Turb</td>
<td>0.0092</td>
<td>12. ECOLI</td>
<td>0.8821</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 3: Integrated evaluation

To calculate the quality group that the monitoring site belong to, multiply the weighting vector to the fuzzy matrix to get an evaluation matrix $B$:

$$B = w_{1x_m} \cdot R_{n x m} = \begin{bmatrix} w_1, w_2, \ldots, w_n \end{bmatrix} \cdot \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix}$$

$$B = [b_1, b_2, \ldots, b_m]$$

(Equation 21)

With $B = B_{1x_m}$ (m is the number of standard levels $V (v_1, v_2, \ldots, v_m)$).
Where:

\( w_{1xn} \) : weighting vector of one quality parameter.

\( R_{nxm} \) : built fuzzy membership matrix gained in step 1, Equation 11.

\( m \): the number of standard levels \( V (v_1, v_2, ..v_m) \).

\( n \): number of monitored quality parameters.

Taking an example at Phu Long site, multiplying the fuzzy matrix in Table 5.6 with the weight vector in Table 5.7, the final membership value for each quality group is presented in Table 5.8

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0056</td>
<td>0.0072</td>
<td>0.0113</td>
<td>0.9759</td>
</tr>
</tbody>
</table>

According to Table 5.8, the monitored dataset of 18 parameters which were measured in Phu Long site shows 56% belonging to group A1, 72% belonging to group A2, 1.13% belonging to group B1 and 97.59% belonging to group B2. Applying the principle “maximize the membership value”, the quality of the water in Phu Long site is in group B2, that is of Bad quality.

Also, to determine the level of pollution at the site, if the membership value of B2 is higher than 0.6, the quality of water is classified into group lower B2 (B2-). This means the water quality is highly polluted and needs special treatment.
5.3 Summary

The Water quality index is a comprehensive approach for quality evaluation based on a group of experimental functions. This method is widely applied and results are officially approved in many countries. In 2011, Vietnam has promulgated a Handbook of WQI’s instruction and application (Decision no.879/QD-TCMT, 2011). The advantage of the WQI method is that there is quick assessment, and the result is approved by the authority. However, with the limited number of assessed parameters, for example nine parameters as regulated in the Vietnam handbook, the environmental quality cannot be totally covered.

Besides, Fuzzy assessment is a model which enables wide quality assessing. It is based on two important elements: fuzzy matrix and weights. Although the concept of fuzzy set was long developed, its application on the environmental management and assessment is still being researched. Fuzzy model can apply unlimited number of indicators, thus it can comprehensively assess all the relative parameters. Nevertheless this method should be further researched for more comprehensive membership functions and limiting the uncertainty of the weights and the final results.

Consequently, this Chapter simply demonstrates the conducting steps of the two methods. The next chapter discusses the results gained from Fuzzy assessment and Water Quality Index. From these results, strengths and weaknesses of the two methods are identified.
CHAPTER 6

6. SURFACE WATER QUALITY RESULTS

This chapter presents the results from the water quality evaluation based on the water quality parameters dataset which were measured at the monitoring sites in the study area. When the dataset of quality parameters were tested by the fuzzy model and WQI, two outputs were gained:

- The overall water quality of the study area was found using the results of the testing with the two methods: WQI and fuzzy model.

- The strengths and weaknesses of the fuzzy model were identified. Therefore, a suggestion for modification was made.

This chapter contents three main parts. The first part involves discussing the pollution trend of the water quality in the period from 2010 to 2015. By scanning the monitored datasets from 2010 to 2015, types and developing trends of pollutants were identified. Therefore, they support the quality results gained from the integrated models using the 2015 dataset. The second part mainly discusses the results of Fuzzy model and WQI, using the monitored dataset in 2015. Therefore, discussions on the strengths and weaknesses of Fuzzy model, and recommendations on modification of the model are presented in the third part of this chapter.
6.1 General evaluation of water quality from 2010 to 2015

In general, when skimming the overall water quality of the study area in a period of time from 2010 to 2015, it could be seen that the pollution concentration for most of the substances was much higher at canal sites that are located in the central area of the City rather than the river sites; and higher when monitored at high tide rather than low tide.

For biological pollution as indicated by Biological and Chemical oxygen demand parameters, canals sites have shown to have very high values which are an indication of biological pollution (Figure 6.1). Measured values of these two parameters met the standard for the river sites.

![Figure 6.1: Overview of the relationship between Biological oxygen demand and Chemical oxygen demand parameters at canal sites from 2010 to 2015](image-url)
For metal contamination, a very high measurement of Lead (Pb) was detected in 2015 at high tide for canal sites. Figure 6.2 shows the concentration of heavy metals, including Lead, Cadmium and Cooper from 2010 to 2015, which Cooper shows its highest values among the substances.

![Figure 6.2 Overview of Lead, Cadmium and Cooper values from 2010 to 2015, at low tide](image)

For bacteria pollution, extreme values of Coliform and Escherichia Coliform were measured at all sites in all years, as shown in Figure 6.3. Although the pollution trend was going down in 2014 and 2015, the monitored values were still very high. This is a sign of pollution that needs immediate treatment.
Figure 6.3 Overview of Coliform and Escherichia Coliform from 2010 to 2015, at low tide

For nutrient pollution, high values of Ammoniacal Nitrogen (NH₄-N) were found at all canal sites for all years, as shown in Figure 6.4. This pollution can be a result of a free discharge of domestic waste water, particularly agricultural fertilizers, into the canal system without pre-treatment, however further studies should be conducted to identify the pollution sources.
6.2 Water quality results applying integrating assessment methods

According to the Vietnam standard for Surface water quality in 2008 (Environment, 2008), water is classified into four groups of quality, including A1, A2, B1, B2, from the highest to the lowest quality.

Because of the discontinuation of the monitored parameters and changes in the number of monitoring sites each year, datasets from 2010 to 2014 could not be used for integrated assessment. Therefore, the dataset for 2015, which is considered to have the highest number of monitoring sites and monitored parameters, is applied in the integrated assessment.

Two methods of evaluation are applied, including the Water Quality Index and the Fuzzy Comprehensive Analysis. The Water Quality Index is an
officially approved method for assessing overall water quality of the study area. Its results are compared to the results of the Fuzzy Comprehensive Analysis. For the Fuzzy Comprehensive Analysis, three different weight methods are applied for testing the sensitivities, that is the uncertainties of the results. They are Entropy, F-statistic and Overweigh methods. Details about these calculations were presented in Chapter 5.

**Overall surface water quality results**

In general, four methods were applied. These were (1) the Water Quality Index, (2) the Fuzzy Comprehensive method applying Overweight standard method, (3) the Fuzzy Comprehensive method, applying the Entropy weighting method, and (4) the Fuzzy Comprehensive method, F-statistic weighting method. Table 6.1 shows the overall results from these methods applied to the monitored dataset in 2015.

**Table 6.1: Results of water quality from the two integrating methods**

<table>
<thead>
<tr>
<th>Type of site</th>
<th>Name of site</th>
<th>Label</th>
<th>At high tide</th>
<th>At low tide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Type of site</td>
<td>Name of site</td>
<td>Label</td>
<td>At high tide</td>
<td>At low tide</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1) (2) (3) (4)</td>
<td>(1) (2) (3) (4)</td>
</tr>
<tr>
<td>River</td>
<td>Dong Tranh</td>
<td>DT</td>
<td>B2+ B2- A1 A1</td>
<td>B1 B1 A1 A1</td>
</tr>
<tr>
<td>River</td>
<td>Ngay Bay</td>
<td>N7</td>
<td>B1 B2- A2 A1</td>
<td>B1 B2+ A1 A1</td>
</tr>
<tr>
<td>Type of site</td>
<td>Name of site</td>
<td>Label</td>
<td>At high tide</td>
<td>At low tide</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>36</td>
<td>Canal Nhi Thien Duong</td>
<td>C10</td>
<td>B2-</td>
<td>B2-</td>
</tr>
<tr>
<td>37</td>
<td>Canal Cau so 1</td>
<td>C11</td>
<td>B2-</td>
<td>B2-</td>
</tr>
<tr>
<td>38</td>
<td>Canal Nguyen Huu Canh</td>
<td>C12</td>
<td>B2-</td>
<td>B2-</td>
</tr>
<tr>
<td>40</td>
<td>Canal Cau Mong</td>
<td>C14</td>
<td>B2-</td>
<td>B2-</td>
</tr>
</tbody>
</table>

**Notes for the columns:**

(1): **Applied Water Quality Index,**

(2): **Applied Fuzzy Comprehensive method, Overweight standard method,**

(3): **Applied Fuzzy Comprehensive method, Entropy weighting method,**

(4): **Applied Fuzzy Comprehensive method, F-statistic weighting method.**

When applying the Water Quality Index method, most of the sites meet the lowest quality group. Similar results are found when applying the Water Quality Index and the Fuzzy comprehensive with the Over-weight standard method. However, results are different among the three combinations of methods which apply the integrating fuzzy method. This variation comes from the approach used in assigning weights, including Over-weight, Entropy and F-statistic.
Model’s uncertainties

Taking a river monitoring site, Nga Bay (N7) as an example, as shown in the Table 6.1, row number 15, the quality of water is allocated four different quality groups: B1, B2, A2, A1 when the four different assessment methods (1) to (4) are applied. Scanning through the columns 1 to column 4 of Table 6.1, it can be seen that there is a big variation amongst the results of the groups. This difference comes from applying the various weighting methods in the fuzzy assessment.

Weights are allocated according to the perceived level of importance of the parameters, and these have values ranging from 0 to 1, being equivalent to 0% to 100%. Therefore, if the parameter has a higher weighting value, its effect on the final result is higher. Table 6.2 shows the weighting values of the parameters when applying the three different methods at low tide. It can be seen from this that the weights are not evenly distributed. There are parameters, such as Manganese or Oil, which receive very low values whereas Coliform or Escherichia Coliform has much higher values. Discussion on the results of the three weighting methods is as follows.
Table 6.2 Weighting values of three different methods applying monitored dataset at low tide

<table>
<thead>
<tr>
<th>Quality parameters</th>
<th>Overweight</th>
<th>Entropy</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>River sites</td>
<td>Canal sites</td>
<td>River sites</td>
</tr>
<tr>
<td><strong>BOD₅</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.015</td>
</tr>
<tr>
<td><strong>COD</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>COLI</strong></td>
<td><strong>0.088</strong></td>
<td><strong>0.042</strong></td>
<td><strong>0.220</strong></td>
</tr>
<tr>
<td><strong>pH</strong></td>
<td>0.003</td>
<td>0.000</td>
<td>0.052</td>
</tr>
<tr>
<td><strong>TSS</strong></td>
<td>0.007</td>
<td>0.000</td>
<td>0.042</td>
</tr>
<tr>
<td><strong>Turb</strong></td>
<td><strong>0.009</strong></td>
<td>0.000</td>
<td>0.049</td>
</tr>
<tr>
<td><strong>DO</strong></td>
<td>0.003</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td><strong>Pb</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Cd</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td><strong>Cu</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td><strong>OIL</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td><strong>ECOLI</strong></td>
<td><strong>0.882</strong></td>
<td><strong>0.954</strong></td>
<td><strong>0.198</strong></td>
</tr>
<tr>
<td><strong>Salinity</strong></td>
<td>0.000</td>
<td>0.000</td>
<td><strong>0.211</strong></td>
</tr>
<tr>
<td><strong>Mn</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
</tr>
<tr>
<td><strong>NH₄-N</strong></td>
<td>0.003</td>
<td><strong>0.003</strong></td>
<td>0.141</td>
</tr>
<tr>
<td><strong>PO₄-P</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.015</td>
</tr>
<tr>
<td><strong>Cr⁶⁺</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.092</td>
</tr>
</tbody>
</table>

*Over-weight method*: This method calculates the weight of each parameter based on the actual values of the monitored samples in comparison to the standard, and it is based on the principle “the more or less, the better”. Therefore, the quality parameter will receive a higher weight if its monitored dataset is increasingly different from the regulated standard.
In Table 6.2, the three highest weights are associated with Escherichia Coliform, Coliform, and Turbidity for the river sites, and Escherichia Coliform, Coliform and Ammoniacal Nitrogen for canal sites. These highest weighting values show that the monitored datasets of these parameters have the highest variation from the standard. On the other hand, the pollution levels associated with these parameters are the highest as compared with other pollution levels.

**Entropy and F-statistic**: These are the two methods focusing on the dataset of the parameter. There are 18 datasets, representing 18 parameters, being tested in the model. Each dataset shows a group of monitored values from 41 monitoring sites.

Entropy is a physical concept in a thermal dynamic system that is used for defining the system’s turbulence. The entropy method defines the weight according to the variation of the values within the dataset. A dataset will receive a higher weight if its member values have a higher variation.

For example, in the Table 6.2, Escherichia Coliform, Coliform and Salinity have the three highest weights for river sites after applying the Entropy method. The highest weights mean that these three parameters have the biggest spread amongst the monitored values and therefore it can be predicted that they have extreme values.
For the F-statistic method, as F is the ratio between the square of the standard deviation of each group and the square of the standard deviation within the group, a higher F value indicates higher differences between groups. Therefore, this method will give high scores for the datasets having high variation than other datasets.

For example, applying the F-statistic for the river site in Table 6.2, three highest weights are assigned to the metal pollution parameters including Lead, Cadmium and Cooper. These highest weights mean that these parameters are more consistent amongst the monitored values than other parameters.

6.3 Discussion

6.3.1 Discussion on the water quality of the study area

In the period from 2010 to 2015, the water was found to have high turbidity, and was contaminated by biological pollutants, nutrients, Lead, and extremely high values of faecal bacteria. The pollution level was much higher when measured at the canal sites than at river sites, and there were a higher level of pollution at low tide as compared with high tide. Also, the trend was similar amongst the pollutants: very high in 2010, and going down rapidly from 2010 to 2013, and reaching the lowest values in 2013. Then they trended upwards from 2014 to 2015.
Furthermore, applying the Water Quality Index method to the dataset in 2015 the results shows that all the sites are in the group of lowest quality.

For the Fuzzy model, when applying three different weighting methods, the overweight standard weighting method shows similar results to the Water Quality Index’s outputs. The other two combinations show no relationship to each other and the Water Quality Index’s results. Also, at some sites, there is a big difference in the quality results between the methods, for example four quality groups are set at one monitoring site, Nga Bay (N7) site as shown in Table 6.1. These differences demonstrate the uncertainty of the fuzzy model.

On the other hand, checking the weighting results shown in Table 6.2, similarities can be seen between the weights and the pollution trend from 2010 to 2015.

For example, Escherichia Coliform, Coliform and Turbidity have the highest weight values when applying the overweight standard method. This means that these variations between the monitored values of these parameters and the standard values are the highest, and therefore indicate the highest pollution level. Also, testing using the Entropy weighting method, Escherichia Coliform, Coliform and Salinity are found to have the highest weight. This means these parameters have high variation within the monitored values, and thus have extreme values in their datasets.
6.3.2 Discussion on the assessment models

Water Quality Index

The WQI is a group of tried and tested equations built especially for a specific area and based on a number of quality parameters. Therefore it is stable, and the results are reliable in the area where the equations are built.

However, the method has big limitations. One of the limitations is that the equation cannot be applied to other areas unless a new group of equations are set up. However, the environment, particularly water, is changeable because of many different and unpredictable impact factors such as weather, human activities, or natural disaster. Therefore, it is difficult to set a function which can precisely assess the quality of the environment, especially water.

One other limitation of the Water Quality Index method is the fact that it focuses on a limited number of parameters, for example nine parameters are regulated in the Guidelines for the Water Quality Index of Vietnam. However, the number of quality parameters are many and these are increasing because more new quality parameters are added for assessment. Furthermore, choosing which parameters should be assessed is determined by the water use. For example the Environmental Protection Agency of Ireland regulates 48 parameters in the Microbiological, Chemical and Indicator groups for drinking water in 2014. Therefore, the choice of
parameters imposes big limitations on the Water Quality Index because it is fixed by the number and type of parameters.

**Fuzzy Comprehensive Analysis**

This is a method of assessing the overall water quality based on given parameters. Its advantage is that it can deal with an unlimited number of quality parameters, and that the membership level is built based on the regulated standard. One other advantage of this method is that unlike the Water Quality Index that relies on expert knowledge in building the formulas, this method relies on the monitored values of the quality parameters to test the quality of the water. Taking the water quality dataset in 2015 in Ho Chi Minh City for testing, the results have shown similarities between the Water Quality Index and the Fuzzy model applying the overweight method.

However, one of the big limitations of the Fuzzy method is the uncertainty. Although there are similarities in the results gained from the combination of the Fuzzy and Overweight standard method and the Water Quality Index method, testing the fuzzy model with three different weighting methods shows great variation. Furthermore, when working on the quality dataset, no evidence can be collected to identify which results are more precise.

Taking the Nga Bay (N7) river site as an example, this site is allocated four quality groups B2, B1, A2 and A1 when being assessed by the Water Quality
Index; the combination of Fuzzy Comprehensive and Overweight standard method; the combination of Fuzzy Comprehensive and Entropy weighting method; and the combination of Fuzzy Comprehensive and F-statistic weighting method. However, no evidence can be collected to check which of these assessment methods is giving the most precise output.

Therefore, in order to identify the limit of the uncertainty of the Fuzzy model, scanning elements of the model and tracking the uncertainties of these elements are discussed.

### 6.3.3 Track the uncertainty of the fuzzy model

Details about the steps involved in conducting the Fuzzy model was described in Chapter 5. In general, the model comprises two parts: Fuzzy membership and Weight.

**Fuzzy membership**

Fuzzy membership is built based on the basic shapes: triangle and trapezoid. Although they are the simplest shapes of membership, past experience has shown that they are suitable for environmental assessment with quality parameters in past research such as that conducted by Meng Lihong in 2009 and Shen Jihing in 2011 (Meng Lihong, 2009; Shen Jihong, 2011). This part of the model can be considered stable for the following reasons:
Although other shapes such as Gaussian or Bayesian can be tested for their applicability to the environmental assessment using quality parameters, basic membership shapes such as the triangle and trapezoidal can be used because of their simplicity and applicability to the environmental assessment.

In the model, the standard values are used for setting the limit values of the shapes. Standard values are the official regulated values, therefore they are stable and reliable.

Furthermore, the shapes are adjusted to suit the characteristics of the parameters. For example, a pH value ranging from 6 to 8.5 is considered to be of good quality, whereas pH values less than 6 or higher than 8.5 can be classified as being of bad quality. Therefore, the membership shape of pH is modified to be appropriate for this characteristic (Figure 5.8, chapter 5). Because the quality parameters are set on the characteristics of the environment, particularly water, they are considered to be stable.

**Weights**

Weight is the level of importance of a parameter. There are different methods used to define weights including:
Subjective assigning: Weight can be assigned by an expert knowing the importance of the parameter. Methods used include the Analytic Hierarchy process (AHP), Similarity to an Ideal Solution (TOPSIS) or The VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR).

Based on the dataset’s characteristics: for example comparison within the dataset (Entropy method), or with the datasets (F-statistic method), or with the standard (Overweight method).

Different weighting values lead to different result despite the membership. Weight, for that reason, becomes uncertain. As a result, in order to limit the uncertainty of the model, finding an appropriate method for defining weight is essential.

6.3.4 Proposing an option for reducing the uncertainty of the Fuzzy model

Water is a natural component, and it is characterized by quality parameters. In the Fuzzy model, because weights represent the importance of the quality parameters, which are parts of the water quality features, the weight for that reason cannot be assigned subjectively.

Weight, therefore, relies on the characteristic of the parameter’s dataset being defined. This can be determined by the purpose of the assessment. For
example the overweight method is used to check the pollution level of the parameter against the standard; or the Entropy method is applied to check what parameter has the highest possibility of having extreme values which means that there may be an environmental incident happening.

Also, weight depends on the types of parameters being assessed. For example, in Table 6.2, in applying the Overweight method to the dataset at a canal site, 15 parameters receive no weight (having zero value), and the other three parameters Coliform, Escherichia Coliform, and Ammoniacal Nitrogen have 100% of weight. When checking the results from other weighting methods in Table 6.2, it can be seen that the majority of the weight values focus on two or three parameters. The final results of the fuzzy assessment, therefore, depend largely on these parameters, and thus the method can be imprecise.

As a result, choosing a suitable parameter is necessary to determine better weighting and thus this would lessen the vagueness of the model. Consequently, to limit the uncertainty of the fuzzy model, one approach would be to apply a Geographical Information System. By building a watershed network and running a geographically regression model, the suitable impact factors for the water quality assessment, the most probable pollution types and the sources of the pollution can be identified. This would then allow the determination of a group of possible quality parameters that should be tested in the fuzzy model.
Furthermore, the geographic model can also be used to check whether or not the results from the fuzzy model are reliable. Taking the example of the Nga Bay site (N7), the monitoring site is allocated 4 different water quality groups when the Water Quality Index method is applied along with the three combinations of fuzzy method and weight *(Table 6.1, row 15).* By applying the map of the region and the land-use information, it is apparent the site is located in the Can Gio mangrove area which is a national nature reserve. Therefore its quality value can be predicted in one of the groups of A2 or B1, (see Figure 6.5).
Figure 6.5 Location of the Nga Bay monitoring sites
6.4 Summary

In the period of 2010 – 2015, water quality of the study area was affected by biological, bacteria, and Lead pollutants. Testing the monitored dataset of 2015 using two integrating methods: the Water Quality Index and the Fuzzy Comprehensive Analysis, including three combinations: Fuzzy and overweight method, Fuzzy and Entropy method, Fuzzy and F-statistic method, the results show the similarity between the Water Quality Index and the combination of Fuzzy assessment and overweight methods. However, for the three combined Fuzzy assessment methods, the results are much different. With the membership shape, one of the components of the fuzzy model, is steady, weight is the uncertainty of the Fuzzy method.

Therefore, finding an appropriate method for assigning weights is one way to limit the uncertainty. Because there are many methods for defining the weights, two factors should be considered in choosing the possible method.

The first is to determine the purpose of the assessment, for example: assessment for checking the pollution levels, or finding an excess value which can represent an environmental incident. Identifying this purpose will help in selecting the right weighting method. For example the overweight method focuses on the difference between the monitored values and the standard, or the Entropy method relies on the differences within the dataset which can help to track the excess values.
The second factor is choosing suitable parameters for the assessment. In the test of the two models with the dataset in 2015, some parameters such as Oils or Manganese show their weighting values are 0 or close to 0 in all the weighting methods (Table 6.2), and therefore they do not contribute to the assessment models. Furthermore, some others have weighting values that are too high, such as 88\% or 95\% for Escherichia Coliform. As a result, the results mostly depend on these very high weighting parameters, and thus it can lead to bias.

Consequently, choosing the type of quality parameter for assessment becomes very important for the effectiveness of the model, especially in regards to the weight. In order to select the appropriate quality parameters, an understanding of the study area including the factors affecting to the water quality is essential. These factors can be tested by a GIS technique, including building the watershed network and running the regression models. Then, these models can be used to check the reliability of the results of the Fuzzy method by checking the suitability between the practical conditions and quality groups’ results.
CHAPTER 7

7. GEOGRAPHICAL ANALYSIS OF SURFACE WATER QUALITY AND RESULTS

This chapter discusses the steps for developing a watershed area and applying its characteristics to build the regression models. The results gained from the regression models show the relationship between the impact factors, such as population and land-use, and the water quality results obtained from the Water Quality Index and Fuzzy comprehensive and assessment methods.

Two stages are required, the first being the building of the sub-watershed delineation model and the second involves running the regression model. The watershed delineation model is a calculated simulation of the actual watershed areas based on elevation and inlet/outlet points. In this stage, the impact factors such as land-use, population and number of manufactures are also added to the map as support attribute data for running the regression model in the next stage. Running the regression model is the next stage of the process. Two regression models are applied in the test; they are the Ordinary Least Square and the Geographical Weighted Regression models. The Ordinary Least Square model is a linear regression model that tests the relationship between the dependent variable, water quality, and a number of impact factors such as land-use and population density. The Geographically
Weighted Regression is also a linear regression model which takes into account the distance between the tested objects.

7.1 Applied datasets

In this research, two types of datasets were applied to building the watershed model and applying the regression analysis. They are the coordinated geographical dataset and the textural quality dataset. Details about these datasets were presented in Chapter 4: Water quality and geographical datasets.

*Geographical dataset:* This type of data was applied for building the watershed area’s map. They include the actual elevation, rivers and canals system, monitoring site locations, administrative boundaries and population density.

*Elevation map:* The average altitude of the study area is about 3 metres above sea level. The topography is highest in the North West and decreases to the South East.

*Canals and rivers network:* The digitized stream lines of the City in 2015, built by the City’s Department of Science and Technology, were applied in the model. The stream network is dense, and most of the area is covered by
the canal lines, except in some of the urbanized areas where the canals have been filled in.

*Monitoring sites:* 41 monitoring sites, 26 located in rivers and 15 in canals, have been used and placed onto the watershed map as the inlet/outlet sites in order to determine the sub-watershed areas.

*Administrative information:* relating to population, density, and the development of 24 administrative districts.

*Land-use information:* Data about the 2015 land use distribution and percentage of land cover types were also applied in the regression models.

*Textural dataset:* Quantitative values for the 18 water quality parameters, except for Temperature and Chromium, from the 41 monitoring sites are applied for testing the relationship with the impact factors. Temperature and Chromium were not put into the regression test because of their discontinuation in the dataset. Attribute information about the population distribution, population density, and the number and size of manufacturing establishments were also applied as supportive data.

### 7.2 Building the regression models
7.2.1 Regression model build-up

Regression models are the traditional method for testing the relationship between many factors. The models can be simply explained as:

\[ y = f(x) \]

(Equation 22)

Where:

- \( y \) : dependent variables, which are influenced by other factors;
- \( x \) : independent variables, which influence the dependent variables;
- \( f(x) \) : function showing the relation between \( y \) and \( x \).

**Dependent and Independent variables**

Dependent variables \((y)\) are the data applied for testing the influence they have on the independent variable. In this study, they are water quality parameters from the 41 monitoring sites including \( \text{pH} \), Salinity, Temperature, Total suspended solids, Turbidity, Dissolved oxygen, Biological oxygen demand, Chemical oxygen demand, Lead, Cadmium, Cooper, Manganese, Total Coliform, Escherichia Coliform, \( \text{NH}_4^-\text{N} \), \( \text{PO}_4^-\text{P} \), and Oil. These are categorized into two sets of values at the highest and lowest tide. The dataset was collected in 2015.
Independent variables \((x)\) are also applied in order to test the influence of the socio-economic factors on the water quality. They include population density, number of all size manufacturers, sub-watershed area, and land-use. Among them, the land-use dataset is a set of textual information such as “Urban area”, “Agriculture area” or “Natural reserve area”. Therefore, the dataset must be transformed into numerical values before running the regression.

In the study, the scoring is applied subjectively based on the principle “the higher the pollution possibility, the higher the score”. The scoring principle is applying the higher score for the land-use purpose which has the higher possibility of causing pollution. For example, the natural reserved area receives a lower score than the cemetery or waste treatment area. This is because the possibility of causing pollution of the reserved area is much less than it is in the cemetery or waste treatment area. The transferred scores of the land-use types are therefore shown in Table 7.1.
### Table 7.1: Transferred scores from land-use type

<table>
<thead>
<tr>
<th>Land-use type</th>
<th>Transferred numeric value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Governmental area, Natural zone, Reserved area</td>
<td>0</td>
</tr>
<tr>
<td>2. Military area, Amusement park, Religious area, Salt farm, Administrative offices</td>
<td>1</td>
</tr>
<tr>
<td>3. Production forest, Commercial service, Infrastructure area, Hospitals, Universities</td>
<td>2</td>
</tr>
<tr>
<td>5. Waste treatment facilities, Cemetery area</td>
<td>4</td>
</tr>
</tbody>
</table>

Dependent and independent variables being tested in the regression model must be in the same unit. However, the data applied in the study uses different scales of measurement because of the different characteristics. For example, monitored values of the quality parameters are mostly measured in mg per litre, whereas some others use different units of scale, such as Turbidity applying Nephelometric Turbidity Units (NTU) or Coliform applying Most Probable Number per 100 millilitres (MPN/100mL).

Furthermore, independent variables do not use the same unit. For example:

- Population density (people/m²)
- Number of all size manufacturers
Sub-watershed area (square m²)

- Land-use value (transferred score)

Therefore, an additional step, standardizing the variables, must be conducted before running the regression.

**Data standardization**

Statistically, a normally distributed dataset is displayed by its features including mean value ($\mu$) and standard deviation ($\sigma$). The standardization process transfers the original dataset with a certain mean and standard deviation to another dataset which has a mean value is 1 and a standard deviation is 0. The distribution of the modified dataset is not changed. The calculation is simple:

\[
 x_{\text{standardized}} = \frac{x_{\text{mean}}}{\sigma}
\]

(Equation 23)

Where:

- $x_{\text{standardized}}$: the standardized value of $x$
- $x_{\text{mean}}$: mean value of the dataset containing $x$
- $\sigma$: standard deviation of the dataset containing $x$.

**Linear regression models**
There are different types of regression equations, such as linear, Gaussian, or exponential. In the study, simple linear regression models being the Ordinary Least Square and the Geographically Weighted Regression models are applied. The general expression of the two models is:

\[ y = a + bx_1 + cx_2 + \ldots + nx_n \]  

(Equation 24)

where:

a, b, c, ..., n: coefficient values of the equation

Overview of the calculation process is presented in Figure 7.1.
Figure 7.1: Overview of the regression process

- **Dependent variables**
  - Quality parameters’ values

- **Independent variables**
  - Four influence factors

- **Sub-watershed delineation map**
  - (detailed in 7.2.2)

- **Ordinary Least square Regression model**

- **Geographically Weighted Regression model**

- **List of highly related dependent variables**
  - (highest $R^2$)

: Data input/output

: Processing tools
The Ordinary Least Square model is a simple linear regression model testing the correlation between independent variables and dependent variable (Wilson, 2012). However, the model is based on the assumption that the relation between the dependent variable and the independent variables is static and consistent across the study area. The model creates a single function for all the features in the dataset, thus it is called a “global model” (Fischer and Getis, 2010). Nevertheless, with the geospatial data such as distance or population growth, the relation is changeable within the space.

In order to lessen the weakness of the Ordinary Least Square model, the Geographically Weighted Regression model is applied to increase the power of the model ($R^2$). This model is based on the assumption that a nearer object will get a higher weight (Murayama, 2012). This model is called a “local model” because it creates a regression function for every feature in the dataset. The model calibrates each feature by using a target feature and its neighbour.

### 7.2.2 Watershed delineation model

As an important part of the Geographically Weighted Regression model, a sub-watershed area map was built based on the elevation and defined sub-catchments. An overview of the procedure followed is shown in the Figure 7.2. The approach starts with the Digital Elevation Model (DEM) which is a raster altitude map built from the elevation points. The model is then overlain with the actual rivers and canal networks to determine break lines in
the model (see *Chapter 3, part 3.2 Environmental resources: Surface water* for more details about the streamlines network). Consequently, pouring points (inlets or outlets) are then manually added to the map in order to generate the final watershed model. Because the study’s objective is to identify influences on the water quality, monitoring points are assigned to be at the pouring points of the catchment areas, for the whole catchment area.
Figure 7.2: Overview of the Watershed delineation process
7.3 Results

7.3.1 Sub-watershed delineation map

In order to test the impact factors to the water quality, a model of the watershed area must be built. A watershed model is a model presenting the actual water network according to the elevation model, which is the Digital Elevation Model (DEM), as shown in Figure 7.3. This is a raster elevation map built on a number of elevation points measured in the year 2000 around the City. Information about influences such as population, land-use, or industrial development is also integrated into the model.

By integrating the actual streamline network, a computerised sub-watershed network was built (see Figure 7.4). This is a model of water flow that is based on the elevation. In the sub-watershed model, monitoring points are set to be batch points (inlet/outlet of a watershed). Among the 41 monitoring sites, 36 sites are automatically included in the model. The others are removed because they are placed outside the City’s boundary, or too close to each other, or there is not enough information about the elevation to set up an inlet/outlet point.
Figure 7.3: Computerised Digital Elevation Model
Figure 7.4: Computerised Sub-watershed area
7.3.2 Regression’s results

The purpose of the regression tests is to check the relationship between the monitored values and the influence factors, and also explore the possibilities of reducing the uncertainty of the fuzzy model when testing the overall water quality. This regression was used to test the relation between the dependent variables, quality parameters, and the influence of socioeconomic factors including: Population density; Sub-watershed area; Number of all size manufacturers; and Land-use (transferred to pollution points). Details about the results of the two regression models are presented in the Appendixes 2 and 3.

Two of the most important factors to be considered in the regression model are the power of the model and the significant F value. The power of the model ($R^2$) is the percentage of values that can be explained by the regression model, ranging from 0 to 1 (equivalent from 0% to 100%) with the higher the percentage $R^2$ the more powerful the model. The significant F, also known as the p-value, is a statistical value which tests the significance of the model. Usually, a p-value less than 0.05 is considered statistically significance whilst the linear regression should be removed if the p-value is higher than 0.05.

These four impact parameters are the fundamental socio-economic factors that can most possibly influence the surface water quality. Although the powers $R^2$ of the tested models are relatively low, all regression models are statistically significant. Therefore, some other specific factors could be
added to the models in the future research, such as types of agricultural crops, traffic congestion rate, or frequency of urban flooding relating to rainfalls, for increasing the models’ powers.

**Results from the Ordinary least square regression model**

Testing with the quality parameters, Temperature and Chromium have been excluded because of interruption of the dataset. Also, dataset of Oil and Manganese parameters at low tide have not been put into the regression test because of the discontinuation of the dataset. For all the remaining tested parameters, the regression p-values are less than 0.05. Overview of the model’s powers of the parameters is shown in Table 7.2.

**At high tide**

At high tide, when testing with 18 parameters (see Appendix 2a), the two parameters with the highest power ($R^2$) are Manganese (Mn) and Dissolved Oxygen (DO), with the highest power of 21% for Mn, and 18% for DO.
Table 7.2: Overview of the model’s power of the tested quality parameters applying Ordinary Least Square model

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameters</th>
<th>Label</th>
<th>Unit</th>
<th>$R^2$ at high tide</th>
<th>$R^2$ at low tide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Biological oxygen demand</td>
<td>BOD$_5$</td>
<td>(mg/l)</td>
<td>0.12</td>
<td>0.134</td>
</tr>
<tr>
<td>2</td>
<td>Chemical oxygen demand</td>
<td>COD</td>
<td>(mg/l)</td>
<td>0.087</td>
<td>0.109</td>
</tr>
<tr>
<td>3</td>
<td>Coliform</td>
<td>COLI</td>
<td>MPN/100ml</td>
<td>0.065</td>
<td>0.302</td>
</tr>
<tr>
<td>4</td>
<td>pH</td>
<td>pH</td>
<td></td>
<td>0.017</td>
<td>0.047</td>
</tr>
<tr>
<td>5</td>
<td>Total suspended solid</td>
<td>TSS</td>
<td>(mg/l)</td>
<td>0.049</td>
<td>0.136</td>
</tr>
<tr>
<td>6</td>
<td>Turbidity</td>
<td>Turb</td>
<td>(N.T.U)</td>
<td>0.059</td>
<td>0.072</td>
</tr>
<tr>
<td>7</td>
<td>Dissolved oxygen</td>
<td>DO</td>
<td>(mg/l)</td>
<td><strong>0.18</strong></td>
<td><strong>0.335</strong></td>
</tr>
<tr>
<td>8</td>
<td>Lead</td>
<td>Pb</td>
<td>(mg/l)</td>
<td>0.10</td>
<td>0.201</td>
</tr>
<tr>
<td>9</td>
<td>Cadmium</td>
<td>Cd</td>
<td>(ug/l)</td>
<td>0.094</td>
<td>0.259</td>
</tr>
<tr>
<td>10</td>
<td>Cooper</td>
<td>Cu</td>
<td>(mg/l)</td>
<td>0.095</td>
<td>0.055</td>
</tr>
<tr>
<td>11</td>
<td>Oil</td>
<td>OIL</td>
<td>(mg/l)</td>
<td>0.053</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>E.Coli</td>
<td>ECOLI</td>
<td>MPN/100ml</td>
<td>0.091</td>
<td>0.160</td>
</tr>
<tr>
<td>13</td>
<td>Salinity</td>
<td>Salinity</td>
<td>(g/l)</td>
<td>0.170</td>
<td>0.148</td>
</tr>
<tr>
<td>14</td>
<td>Manganese</td>
<td>Mn</td>
<td>(mg/l)</td>
<td><strong>0.21</strong></td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Ammonia Nitrogen concentration</td>
<td>NH4-N</td>
<td>(mg/l)</td>
<td>0.14</td>
<td>0.145</td>
</tr>
<tr>
<td>16</td>
<td>Phosphate Phosphorous</td>
<td>PO4-P</td>
<td>(mg/l)</td>
<td>0.094</td>
<td>0.133</td>
</tr>
</tbody>
</table>

(Note: (-) not enough data for running the regression model)

Although the two highest powers are quite low, it is shown that the regression models of these two parameters are significant because the significant F value is less than 0.01, shown in Figure 7.5 and 7.6.
### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of square</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig_F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>177.635</td>
<td>4</td>
<td>44.409</td>
<td>63.260</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>650.049</td>
<td>926</td>
<td>.702</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>827.684</td>
<td>930</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: StD_Mn

b. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std.Error of the Estimate</th>
<th>Change Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R Square change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F Change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>df1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>df2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sig. F change</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std.Error of the Estimate</th>
<th>R Square change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.463&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.215</td>
<td>.211</td>
<td>.8378527737</td>
<td>.215</td>
<td>63.260</td>
<td>4</td>
<td>926</td>
<td>.000</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

b. Dependent Variable: StD_Mn

**Figure 7.5: Analysis of variance of Manganese at high tide**
### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of square</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig_F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>122.109</td>
<td>4</td>
<td>30.527</td>
<td>51.724</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>546.522</td>
<td>926</td>
<td>.590</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>668.631</td>
<td>930</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: StD_DO

b. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std.Error of the Estimate</th>
<th>Change Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.427&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.183</td>
<td>.179</td>
<td>.7682426653</td>
<td>.183</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

b. Dependent Variable: StD_DO

**Figure 7.6: Analysis of variance of Dissolved Oxygen at high tide**

Besides, checking the coefficients of these two parameters, it is shown that the all the influences, including population density, number of all size manufacturers, sub-watershed area and Land-use value (transferred score), have negative impact on the quality value, as shown in Table 7.3.
Table 7.3: Summary of Ordinary Least Square results of Manganese and Dissolved Oxygen, at high tide

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Probability (p-value) (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manganese</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.419814</td>
<td>0.081375</td>
<td>5.159032</td>
<td>0.000001</td>
</tr>
<tr>
<td>Number of Enterprise</td>
<td>-0.000038</td>
<td>0.000008</td>
<td>-4.803972</td>
<td>0.000003</td>
</tr>
<tr>
<td>Pollution Point</td>
<td>-0.019017</td>
<td>0.026929</td>
<td>-0.706214</td>
<td>0.480227</td>
</tr>
<tr>
<td>Standardized Area</td>
<td>-0.154128</td>
<td>0.028207</td>
<td>-5.464170</td>
<td>0.000000</td>
</tr>
<tr>
<td>Standardized Density</td>
<td>-0.367231</td>
<td>0.030598</td>
<td>-12.001841</td>
<td>0.000000</td>
</tr>
<tr>
<td><strong>Dissolved Oxygen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.072040</td>
<td>0.074614</td>
<td>-0.965502</td>
<td>0.334533</td>
</tr>
<tr>
<td>Number of Enterprise</td>
<td>-0.000031</td>
<td>0.000007</td>
<td>-4.288871</td>
<td>0.000024</td>
</tr>
<tr>
<td>Pollution Point</td>
<td>-0.031278</td>
<td>0.024691</td>
<td>-1.266779</td>
<td>0.205558</td>
</tr>
<tr>
<td>Standardized Area</td>
<td>-0.357073</td>
<td>0.025863</td>
<td>-13.806065</td>
<td>0.000000</td>
</tr>
<tr>
<td>Standardized Density</td>
<td>-0.097432</td>
<td>0.028056</td>
<td>-3.472781</td>
<td>0.000554</td>
</tr>
</tbody>
</table>

**Note:**

(*): Probability (p-value) indicates a coefficient is statistically significant when its value less than 0.01.

In Table 7.3, it can be seen that the probability values of the coefficient of the “Pollution point” variable are higher than 0.01 for both parameters: Manganese and Dissolved oxygen. It can be explained that the “Pollution
point” variable is the transferred pollution scores from the land-use information and they are subjectively assigned. Although the coefficient of this variable is not significant, the regression models of the two parameters are statistically significant, shown in Figure 7.5 and 7.6. However, more research on assigning the pollution score to the land-use purposes should be conducted in order to have a more effective regression model.

For Manganese, according to World Health Organisation (WHO) in 2011 “Manganese compounds may be present in the atmosphere as suspended particulates resulting from industrial emissions, soil erosion, volcanic emissions and the burning of MMT-containing petrol. In surface waters, manganese occurs in both dissolved and suspended forms, depending on such factors as pH, anions present and oxidation–reduction potential. Anaerobic groundwater often contains elevated levels of dissolved manganese” (WHO, 2011). Also, according to Vietnam’s national regulation (see Table 5.1 Chapter 5), a concentration of Manganese that is higher than 0.3 mg per litres is a sign of pollution.

Therefore, the regression model is supposed to show positive impact between the Manganese values and urban growth such as population density, land-use, or the increasing in the number of businesses in the area. However, the negative values of the coefficient (Table 7.3) of Manganese show the reverse impact of urban development on the level of this substance. It can be explained that the residents of the area have been using more tap water than ground water, using different types of gasoline, or more industries that
discharge heavy metal have been moving out of the City’s centre. Therefore, as the model power is low, 21%, more influences such as percentage of using groundwater, types of industries, number and types of transportation vehicle, or types of rock and soil in the area should be added to the model for identifying the relation of this substance’s concentration to these factors.

For Dissolved oxygen (DO), DO is a parameter showing the concentration of oxygen in the water. This parameter is very important to the aquatic life; therefore the higher the level of oxygen the better the environment. According to Vietnam’s standard for surface water (Table 5.1, Chapter 5), a level of higher than 6 mg per litre is desirable, whereas a concentration of lower than 4 mg per litre is considered polluted. Therefore, a negative impact of the influences to the level of DO is a sign of biological pollutants in the water which can trigger the growth of microorganism that consume the oxygen and therefore lower the concentration of this substance in the water.

*At low tide*

Coliform (COLI) and Dissolved oxygen (DO) are the two indicators having the highest power ($R^2$) for the model, with the highest power being 0.34 for Dissolved oxygen and 0.30 for Coliform, as shown in Figure 7.7 and 7.8.
### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of square</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig_F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>221.027</td>
<td>4</td>
<td>55.257</td>
<td>117.938</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>433.852</td>
<td>926</td>
<td>.469</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>654.879</td>
<td>930</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: StD_DO  
b. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std.Error of the Estimate</th>
<th>Change Statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R Square change</td>
<td>F Change</td>
<td>df1</td>
<td>df2</td>
</tr>
<tr>
<td>1</td>
<td>.581</td>
<td>.338</td>
<td>.335</td>
<td>.6844872268</td>
<td>.338</td>
<td>117.938</td>
<td>4</td>
<td>926</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density  
b. Dependent Variable: StD_DO

**Figure 7.7: Analysis of variance of Dissolved Oxygen, at low tide**
### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of square</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig_F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>4</td>
<td>114.848</td>
<td>101.590</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>926</td>
<td>1.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>930</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- c. Dependent Variable: StD_COLI
- d. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std.Error of the Estimate</th>
<th>Change Statistics</th>
<th>R Square change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.552a</td>
<td>.305</td>
<td>.302</td>
<td>1.063251994</td>
<td></td>
<td>.305</td>
<td>101.590</td>
<td>4</td>
<td>926</td>
<td>.000</td>
</tr>
</tbody>
</table>

- c. Predictors: (Constant), StD_Area, Pollution_Point, No_Enterpr, Density
- d. Dependent Variable: StD_COLI

**Figure 7.8: Analysis of variance of Coliform, at low tide**

Table 7.4 shows the coefficients of the four independent variables to the two parameters having the highest powers $R^2$. Similar to the results of Dissolved oxygen at high tide (Table 7.3), the model shows negative impact of urban
growth to the concentration of oxygen in the water, which is a sign of biological pollution.

**Table 7.4: Summary of Ordinary Least Square results of Dissolved oxygen and Coliform, at low tide**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Probability (p-value) (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dissolved Oxygen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.227068</td>
<td>0.064364</td>
<td>-3.527848</td>
<td>0.000454</td>
</tr>
<tr>
<td>Number of Enterprise</td>
<td>-0.000041</td>
<td>0.000007</td>
<td>-6.215226</td>
<td>0.000000</td>
</tr>
<tr>
<td>Pollution Point</td>
<td>-0.033346</td>
<td>0.021999</td>
<td>-1.515758</td>
<td>0.129937</td>
</tr>
<tr>
<td>Standardized Area</td>
<td>-0.490790</td>
<td>0.023056</td>
<td>21.286716</td>
<td>0.000000</td>
</tr>
<tr>
<td>Standardized Density</td>
<td>-0.081675</td>
<td>0.025011</td>
<td>-3.265632</td>
<td>0.001147</td>
</tr>
<tr>
<td><strong>Coliform</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.183607</td>
<td>0.104027</td>
<td>-1.764993</td>
<td>0.077901</td>
</tr>
<tr>
<td>Number of Enterprise</td>
<td>0.000059</td>
<td>0.000010</td>
<td>5.795582</td>
<td>0.000000</td>
</tr>
<tr>
<td>Pollution Point</td>
<td>0.061693</td>
<td>0.034173</td>
<td>1.805325</td>
<td>0.071352</td>
</tr>
<tr>
<td>Standardized Area</td>
<td>0.705896</td>
<td>0.035814</td>
<td>19.709815</td>
<td>0.000000</td>
</tr>
<tr>
<td>Standardized Density</td>
<td>0.000009</td>
<td>0.000003</td>
<td>3.157420</td>
<td>0.001657</td>
</tr>
</tbody>
</table>

*Note:*

(*): Probability (p-value) indicates a coefficient is statistically significant when its value less than 0.01)

Also, the regression model shows the positive impact of human influences on Coliform levels. Because Coliform is an indicator of biological pollution,
it is a clear evidence that urban developing including population growth, the number of businesses, or land-use purposes negatively impacts to the surface water quality of the area.

**Results from the Geographically Weighted Regression model**

Geographically weighted regression was used to check the spatial relationship between the variables by improving the power of the model ($R^2$). By applying the spatial regression to the two parameters which have the highest power from the Ordinary Least Square’s model, the considerable increase of $R^2$ shows that there is a spatial correlation between the tested dependent and independent variables. In Table 7.5, the power level from Geographically weighted regression model is significantly higher than that of the Ordinary least square. This also means that the quality values and the influence factors are spatially correlated.

<table>
<thead>
<tr>
<th>Table 7.5: The two parameters with highest model’s power (R$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters</strong></td>
</tr>
<tr>
<td>Manganese (Mn)</td>
</tr>
<tr>
<td><strong>Ordinary Least Square Model</strong></td>
</tr>
<tr>
<td>Model’s power (R2)</td>
</tr>
<tr>
<td><strong>Geospatial Weighted Regression model</strong></td>
</tr>
<tr>
<td>Model’s power (R2)</td>
</tr>
</tbody>
</table>
Figures from 7.9 to 7.12 present the standardized Residuals distribution when applying the two regression models. Residuals are the gaps between the tested values in the dataset and the linear generated values. With a refined model, Residuals are preferably close to 0. In the figures from 7.6 to 7.9, Residuals ranging from -0.5 to 0.5 (yellow area) are desirable while the residual values far from 0 (less than -0.5 or more than 0.5) are in the areas where the models have large biases.
Applying the regression models to the dataset of Manganese at high tide results in appropriate outcomes in the South and some small and dispersed areas in the centre of the City. This means that there can be one or more influencing factors that are missing in the model for Manganese.

Dissolved oxygen shows a high power $R^2$ and this matches the central zone of the study area at low and high tide (Figure 7.10 and 7.11). This parallel
can be explained by the discharge of untreated waste water directly into the water effluent, thereby causing high turbidity and an increased growth of organisms that consume oxygen. One of the other reasons is that the canal system in the City’s centre is dense and narrow. Also, it flows slowly which leads to very low turbulence. This in turn causes low concentrations of dissolved oxygen. At low tide (Figure 7.11), the Geographically weighted regression model of Dissolved oxygen works well in the Northern area of the City, which is mostly the agricultural area.

**Ordinary least square model**

**Geographical weighted model**

\[
\text{Dissolved oxygen, } R^2 = 0.34
\]

\[
\text{Dissolved oxygen, } R^2 = 0.56
\]

**Figure 7.11: Standardized Residuals distribution for Dissolved Oxygen at low tide**
This can be a result of discharge of fertilizer and nutrients to the water, leading to a high number of organisms and therefore lessening the level of oxygen dissolved in the water.

For Coliform, the model shows the highest power $R^2$. This is a sign of a strong connection to human activities. At low tide, the model results in appropriate outcomes at the eastern and western side of the City, which is developing quickly. This is also an indication of a fast and unmanaged
growth in population which in turn poses a potential threat to the environment and human health.

**7.4. Discussion on the watershed delineation and the regression models**

**7.4.1 Watershed delineation process**

The model results show it works well within the study area and it can be applied for identifying the influences between water quality and impact factors. However because of the special characteristics of the area such as the very dense surface water network, low elevation, and flat terrain, there are some problems which occur when setting the parameters of the watershed model.

*Issue about the flat terrain*

The average height of the study area is about 3 meters above sea level, and the whole area is almost flat. This leads to ill-defined sub-watershed delineation which is visually manifested by the zig-zac and parallel shaped boundaries, as shown in Figure 7.13
Figure 7.13: Detailed distribution of the computerised sub-watershed catchment area
Lack of information

At first, all 41 monitoring points were chosen for the model. However, the software generated only 35 batch points and 30 sub-watershed areas, as shown in Figure 7.4.

Also, the total area of coverage of the watershed created by the software is 1,500 km², equivalent to 71% of the total area of the city of 2,100 km². About 30% of the city’s area is left without assessment. This can be explained by the lack of monitoring data in the left-over area.

With the area of more than 2,100 km² with very dense river and canal system, 41 monitoring points for surface water evaluation were set in the computerised sub-watershed model. The monitoring points are also inappropriately distributed being largely concentrated at the city’s centre.

7.4.2 Results of the regression models

With the significant increase in the model’s power $R^2$ when tested with the Geographically Weighted Regression (Table 7.5), there is possibly a spatial relationship between the impact factors, which represents the socio-economic influence, and the pollution levels, which are characterised by the quality parameters.
Dissolved oxygen (DO) is present at low and high tide in the groups having the highest model’s power. This is an indicator which possibly displays chemical or biological pollutants; however more research should be conducted to identify the main sources of pollutants. Coliform (COLI) shows the highest power (60% at low tide) which also indicates that this bacteria is closely related to human activities such as residency (high population density), economic growth (high number of businesses), and land-use. This is similar to the results gained from the Fuzzy Comprehensive Assessment and Water Quality Index.

From the tested parameters, the highest model’s power is less than 60% for Coliform at low tide, tested with the Geographically weighted regression model. Although most of the tested parameters are statistically significant with zero significant-f, the relative powers of model are quite low, ranging from 0.1 to 0.2. Therefore, more impact factors should be added to the models to improve the power $R^2$.

### 7.5 Summary

The watershed delineation model shows its applicability to the study area. However, because the area is flat with a low elevation and a dense surface water network this prevents the model from running more efficiently. It needs improvement to increase the cover area, and to improve its effectiveness by dividing the small sub-watershed areas.
The two regression models show that they are suitable for finding the correlation between the quality parameters and the impact factors. The power levels of the model are low, where the highest power is 0.6 for the Coliform parameter. Therefore there needs to be improvement such as by adding more impact factors or by testing with more quality parameters. In general, the models can be used for identifying the most influential parameters that can be applied in order to reduce the uncertainty of the fuzzy model.
CHAPTER 8

8. PROPOSING INTEGRATED SURFACE WATER QUALITY ASSESSMENT MODEL

This chapter proposes an integrating model for assessing the surface water quality of an area, consisting of Fuzzy Comprehensive Method, Water Quality Index, Watershed Delineation Modelling, and Regression Models.
8.1 Discussion on the current surface water quality assessment process of the study area

Water is an environmental component and has specific features. Therefore, quality parameters have been designed to quantify the water quality’s characteristics. A parameter is a measurement of one portion of the water quality. For example, Temperature is a parameter showing one feature of the water. Therefore, in order to answer the question “What is the quality of the surface water of the area?”, a common method is to set up a monitoring program, integrate assessing the monitored values and get final results of the water quality. Overview of the process is presented in Figure 8.1.

Figure 8.1: Overview of the water quality assessment program

In the study area, Ho Chi Minh City, Vietnam, the monitoring program is set annually. Starting from the budget approval of the City’s Council, the Department of Natural Resources and Environment designs the program and assigning the City’s Environmental Protection Agency to conduct. The program consists of a set of number of sites, location of sites, monitored parameters, time and frequency of taking sample. The monitored values of all the monitoring sites are then evaluated by an integrated method such as
Water Quality Index to get an overall water quality. Then the result is reported to the City’s Council for setting the management plan.

However, the process has some problems, as discussed below:

**Inconsistent and improper quality parameters**

The monitoring program for surface water quality in the study area is managed by the Department of Natural Resources and Environment. Each year the Department is given a limited budget approved by the City’s Council. The funds are used for setting up the monitoring program, including the number of sites, location of sites, and the quality parameters which need to be monitored.

Budget limitations lead to difficulties in processing the monitored values from year to year and they disrupt the possibility of continuously assessing the water quality in a time series, including introducing discontinuity in the measured parameters and number and location of monitoring sites. Taking the canal monitoring sites as an example, Turbidity, one of the quality parameters, was measured in 2010 and 2011. However it was omitted in 2012 before being placed back on the monitoring list from 2013 to 2015. Also, the monitored parameters are not always the same from site to site. For example in 2012, Chemical Oxygen Demand (COD), one of the quality parameters, was measured in all of the sites except Cai Mep, a river monitoring site in the Sai Gon River network. Furthermore, the number of
monitoring sites has increased every year. For example the number of river monitoring sites has increased from 22 in 2014 to 26 sites in 2015

*Lack of integrated assessment method*

The integrated water quality assessment method currently used is the Vietnam’s Water Quality Index, officially established in 2011 (*Decision on Water Quality Index Calculation 2011*). Although this method has advantages such as quick assessment and reliability based on the experimental functions designed especially for surface water environment in Vietnam, it has weaknesses as follows:

The number of tested parameters included in the Index comprises of nine parameters. Although these nine parameters are fundamental for the water quality, they do not represent the overall water quality of the area. Furthermore, the number of actual parameters being tested is not limited and increasing every year. For example, with 18 parameters monitored in 2015 the Water Quality Index could not integrated assesses all the monitored values.

Besides, because the assessment completely depends on the calculation of the monitored datasets, the result could be biased because the assessment does not cover unpredictably changeable environment, for example weather, or human activities such as land-use or population growth, which can be predicted by GIS’s techniques.
Therefore, in this chapter a better program for assessing the water quality is recommended applying fuzzy model and GIS’s techniques as a supportive method for water quality assessment.

8.2 Discussion on Fuzzy model and GIS’s techniques

Assessment of the surface water quality using Fuzzy theory has the advantages such as dealing with unlimited number of parameters, and converting the quantified monitored values into quality groups. However, one of its limitations is its uncertainty because of different methods of assigning weights. When applying fuzzy assessment with three different weighting methods to the dataset in 2015 of the study area, big variations were found in the results as discussed in Chapter 7. Therefore, GIS’s techniques including computerized watershed and spatial regression models are suggested to reduce this vagueness of fuzzy method.

8.3 Proposed monitoring program

Consequently, in order to reduce the inconsistency in the evaluation process and have the most exact results of the water quality, the proposed monitoring program is recommended to have 2 main parts: long-term schedule for 5 years and annual assessment program, shown in Figure 8.2.
8.3.1 Stage 1: Long-term scheduled program

To understand the quality of a water body, especially surface water, is a complicated problem because quality of water is affected by many impact factors including human and non-human influences. Besides, because a parameter is a measurement of a small part of water features, it is not certain that the water quality is accurately identified even if all the regulated
parameters are monitored. Consequently, in order to build a model for quality assessment that can make a closest prediction of the water quality, two following questions should be answered as a priority:

- What are the most likely impact factors to the water quality in the study area? For example: industrial activities, population growth, or heavy rain causing flooding?

- What is the purpose of the assessment? For example: is it to find the pollution level against the standard, or is it to track the extreme values to investigate the instability of the water environment?

Therefore, this schedule program should be designed and conducted before the actual monitoring program to control the vagueness of the assessment process. This stage is set and revised every 5 years. Consequently, by stabilizing the number and types of quality parameters, this stage can also helpful for assessing the water quality trend in a period of 5 years. The process of this stage is presented in Figure 8.3. Three outputs are set in this stage. They include the integrated watershed model, the list of chosen quality parameters being tested and the method for assigning weight which is the level of importance of the selected parameters.
Figure 8.3: Stage 1: Long term schedule program
The integrated watershed model

This is a watershed delineation map that is integrated between the computerized watershed map and the possible impact factors that can affect the water quality. This map is necessary in defining the list of quality parameters by applying the regression models onto it. It is also a key component in managing the uncertainty of the results that would be gained from the annual assessment applying the fuzzy model.

Therefore, one important condition in building this integrated watershed model is to identify the possible impact factors. As this map is periodically revised every 5 years, information of activities such as urban growth, residential development for urban areas, or natural incidents such as rain, flooding, drought or landslides at the riverbank for natural areas should be gathered and incorporated into the model.

In the test with quality dataset in 2015 in Ho Chi Minh City, Vietnam, as the study area is an urban fast developing region information about population growth, business development and land-use have been incorporated in the watershed map. The model showed its capability in testing the relationship between urban growth and water pollution. However, because the model’s powers, \( R^2 \), were not high, more data about human activities needed to added to strengthen the power of the model.
List of quality parameters for monitoring

The list of the quality parameters need to be identified and used for monitoring program over five years. Compilation of this list will help to assess the water quality continuously over a period of time, and therefore evaluate the trend in pollution.

Furthermore, the quality parameters would not be chosen subjectively. Past monitored values would be applied in the regression models, including Ordinary Least Square and Geographically Weighted Regression models, to test the correlation between the influences and the monitored dataset in the past. From then, a number of parameters are chosen according to their highest power’s model (R^2) and statistical significance.

This purposely selection is to amplify the impact of influences to the water quality through the quality parameters; therefore the results will be more precise. For example, if it is found that there is a high relation between the past monitored values of heavy metals and the population growth, the list of parameters should include the heavy metal indicators in the annual monitoring program for 5 years to check whether or not the quality of water is still being polluted by heavy metal caused by human activities.

Furthermore, with the limit in budget and the increasing in the number of parameters need to be monitored regulated in the standard, it is not practical for all the parameter in the standard to be monitored. Therefore, this chosen
list is appropriate and can also represent the influences of the impact factors to the water quality.

**The weighting method**

A fuzzy model is comprised of two parts: the fuzzy matrix and the weight. Weight is the level of importance of a parameter, and is an essential component in the assessment process applying fuzzy theory. As discussed in Chapter 7, when applying 2015 dataset of Ho Chi Minh City, Vietnam, to the fuzzy model with three different weighting methods, the results were much different. Therefore, when the fuzzy matrix is stable, weight becomes uncertain. As a result, two conditions were identified to manage the uncertainty of the weight. They are the types of chosen parameters and the purpose of the assessment.

*Types of chosen parameter:* As discussed above, the list of chosen parameters is identified by applying the regression models onto the integrated watershed model, which includes the information about the influent factors.

*The purpose of the assessment:* Depending on the applying method for defining the weight, the result is different. Therefore, purpose of the assessment becomes one way for choosing a suitable method for defining weights. For example, because the Entropy method is based on the principle of identifying the turbulence within the dataset of a parameter, it is
applicable for defining weights if the purpose of the assessment is to track the extreme values as being a sign of environmental incidents. For most of cases, the purpose of water quality assessment is to identify the pollution by comparing the monitored values to the standard threshold. The appropriate method for defining weight, therefore, is overweight standard method, which its principle is the higher the ratio between the monitored and the standard values, the higher the weight.

8.3.2 Stage 2: Annual assessment program

After defining the integrated watershed model which includes information about influences, types of parameters need to be monitored, and the weights of the chosen parameters in Stage 1, the monitored values are then collected. The collection is carried out by taking samples at the monitoring sites and testing its values at the laboratory.

The monitored datasets of the quality parameters are then assessed by two methods: Water Quality Index and Fuzzy Comprehensive methods, as shown in Figure 8.4. Details about the calculation of the two methods are presented in Chapter 5: Integrated assessment for surface water quality.

Water Quality Index is an official method for assessing the overall water quality. Therefore, its results can be used for reference and check on whether
or not the nine fundamental quality parameters calculated in the Index are in the normal range of values.

Fuzzy model, on the other hand, will calculate all the monitored values of the chosen parameters, applying the weights assigned in Part 1. Furthermore, the uncertainty of fuzzy method is managed by carefully choosing the monitored parameters, assigning weights depending on the purpose of assessment, and testing the reliability of the results with the integrated

**Figure 8.4: Stage 2: Annual assessment program**
watershed model. Consequently, the results can be used for evaluating the overall water quality of the area.

8.4 Summary

Building an assessment model for surface water quality is not simple because water quality is affected by many influent factors, such as land-use, urban growth or natural incidents such as rain or flood. Therefore, the quality results can be biased if the assessment focuses only on the monitored datasets of the quality parameters and does not cover the impact factors.

Consequently, a two stages assessment program is proposed to more precisely evaluating the surface water quality. Stage 1 is set for a period of five years. Main target of this stage is to cover the possible impact factors, manage the uncertainty, and refine the results of the integrated assessment process in stage 2. Stage 2 is the main assessment of the overall water quality based on the monitored datasets. Two methods are applied in this stage, Water Quality Index and Fuzzy Comprehensive Analysis. As the results from Water Quality Index are used for reference, the quality class, gained from Fuzzy Comprehensive Analysis, after being refined by the integrated watershed model can be reliable. Therefore, this two stages assessment program can be applied for a better evaluation of the overall water quality at the study area.
CHAPTER 9

9. CONCLUSION AND RECOMMENDATION

This chapter discusses the conclusion of the research’s results and recommendation for updating and further studying.
9.1 Conclusion

9.1.1 Conclusion on the research’s objectives

As the main objective of this research was to build a method for assessing the water quality in an urban water catchment in a developing country, the research has applied the dataset of 2015 in Ho Chi Minh City, Vietnam for testing the applicability of the model comprising fuzzy evaluation and GIS. The results have shown that the integrated model was appropriate for building the assessment model for surface water quality and it can therefore be appropriate for establishing a surface water quality assessment program for an urban developing area.

Therefore, the answers for the research’s questions are as follow:

1. What is recent surface water quality in Ho Chi Minh City, Vietnam?

In the period from 2010 to 2015, the quality of surface water in Ho Chi Minh, Vietnam was affected by biological, bacteria, and Lead (Pb) pollutants. The pollution level was much higher when measured at the canal sites than at river sites, and it had a higher level at low tide than high tide. Also, the trend was similar among the pollutants: it was very high in 2010, going down rapidly from 2010 to 2013. Reaching the lowest values in 2013, they then trended upwards from 2014 to 2015. This was discussed in Chapter 6.
Applying the integrated assessment models, including Water Quality Index and Fuzzy Comprehensive analysis, with the quality dataset in 2015, the results have shown that all the sites were in the group of lowest quality. The presence of Escherichia Coliform and Coliform parameters were found to have the highest weights when applying the Overweight standard method which is based on the principle that the higher the difference to the standard, the higher the weight. This also means that the water quality has been biologically polluted by very high concentration of the bacteria, particularly faecal bacteria, in the water.

2. **What are the most affecting factors for the surface water quality in Ho Chi Minh City, Vietnam?**

Testing the relation between the influence factors, including land-use, population density, watershed area, and number of all size businesses, and the quality of water, the results has shown that these impact factors have the highest connection to Dissolved oxygen at both the highest and lowest tide, as discussed in Chapter 7. The highest power of regression model was 60%, when testing for Coliform at low tide applying Geographically weighted regression method. This bacteria is closely related to human activities such as residency (high population density), economic growth (high number of businesses), and land-use. This result is similar to the results gained from using the Fuzzy Comprehensive Assessment and Water Quality Index’s assessment, as discussed in Chapter 6.
Furthermore, the significant improvement in the power of the model when using the Geographically weighted regression approach in comparison to the Ordinary Least Square has shown that there is a spatial relationship between the influences and the water quality.

When testing for Dissolved oxygen, the geographically weighted regression has shown the second highest power $R^2$, and fitted the centre of the study area. Particularly, at low tide the Geographically weighted regression model of Dissolved oxygen has worked well in the Northern area of the City, which is mostly the agricultural area. This can be a result of free discharge of fertilizer and nutrients to the water, leading to a high number of organisms and therefore lessening the level of oxygen dissolved in the water.

For Coliform, the model has shown the highest power $R^2$ which is a sign of a very high relationship with human activities. At low tide, the regression model of Coliform has fitted the eastern and western side of the City, which are developing quickly. This is also an indication of a fast and unmanaged increasing population which in turn poses a potential threat to the environment and human health.

Besides, from the tested parameters, the highest model’s power is less than 60% for Coliform at low tide when it is tested with the Geographically weighted regression model. Although most of the tested parameters are statistically significant with zero significant-f, the relative powers of the
models are quite low, ranging from 0.1 to 0.2. Therefore, more impact factors should be added to the models to improve the power $R^2$.

3. **What are the limitations of the environmental quality assessment model applying fuzzy algorithm?**

The integrated assessment using fuzzy algorithm has shown its capability in the field of environmental quality evaluation in some recent research (Olu-Owolabi, Agunbiade, Oseghe, & Adebowale, 2012; Ostovari, Beigi-Harchegani, & Asgari, 2014; Runsheng LV, 2011; Shen Jihong, 2011; Shufei Lin, 2011; Y. Wang et al., 2014; Xiang Zhao, 2011). However, it has limitation such as data loss due to the fuzziness and uncertainty. This research focuses on finding a method which can reduce the uncertainty of the results gained from the fuzzy process.

*Table 6.1: Results of water quality from the two integrating methods* shows the results from the Water Quality Index and three combinations of fuzzy model. It can be seen from the Table 6.1 that the results are different. A clear example of the uncertainty of the fuzzy model is shown in row 15 of Table 6.1, which is the surface water quality of a river site: Nga Bay (N7). When checking the water quality according to the standard which is classified into four quality groups from the lowest to the highest quality: B2, B1, A2 and A1, this site falls into all four quality groups. Therefore, a method to lessen the uncertainty and thus support the integrated fuzzy assessment of water quality is needed.
4. How geographical data, land-use information and geographical information techniques can help to lessen the uncertainty of a model built on fuzzy algorithm?

Fuzzy comprehensive analysis consist of two main parts: fuzzy matrix and weights.

*Fuzzy matrix*

Fuzzy matrix, discussed in Chapter 5, is based on the thresholds regulated in the national standards which are officially approved. Furthermore, another part of the matrix is the membership function. Fuzzy matrix is based on different membership shapes including simple graphs such as triangular or trapezoidal, or curved shapes such as Bayesian (S-shape) or Gaussian (bell-shape). Among these shapes, simple graphs such as triangle and trapezoidal have been researched and showed its capability in environmental quality assessment (L.Ma, 2010; Runsheng LV, 2011; Shen Jihong, 2011). This research applied the combination of simple shapes including triangle and low-semi trapezoidal graphs, and the graphs were built based on the characteristics of the quality parameters (see Figure from 5.4 to 5.9). Therefore, because of their simplicity and solid based, these membership functions can be considered to be steady.
**Weights**

In Chapter 6, weight of each quality parameter was demonstrated uncertain. Weight is the level of importance of a quality parameter. It can be defined by subjectively assigned or by using one of the methods in the multi-criteria group. Applying three different combination of weighting methods and fuzzy model, the results are various, shown in Table 6.1. Therefore, one approach to manage the uncertainty of the fuzzy model is finding an appropriate method for defining weight.

Consequently, two conditions for defining weight have been discussed: purpose of the assessment and types of parameters being tested (*Chapter 6, part 6.3.3: Track the uncertainty of the fuzzy model*).

The purpose of the assessment plays important part when choosing which method of defining weights is used. For example, if the assessment is for checking how much the monitored dataset is different from the standard value, or in another word is testing the pollution level of the tested parameter against the standard, then the over-weight standard method is appropriate. In other cases, Entropy method is suitable if the assessment is to check what parameter has the highest possibility of having extreme values which means that there may be an environmental incident such as flooding or landslide.

Assigning weight also depends on the types of parameters being assessed. For example, when checking the results from other weighting methods in
Table 6.2, it can be seen that the majority of the weight values focus on two or three parameters. The final results of the fuzzy assessment, therefore, depend largely on these parameters, and thus are ambiguous.

As a result, choosing a suitable parameter is necessary to determine better weighting and thus lessen the vagueness of the model. Consequently, one approach is to apply a Geographical Information System. By building a watershed network and running a geographically regression model, the suitable impact factors for the water quality assessment, the most probable pollution types and the sources of the pollution can be identified. This then allows the determination of a group of possible quality parameters that should be tested in the fuzzy model. Furthermore, the geographic model can also be used to check whether the results from the fuzzy model are reliable. For example, quality value of Nga Bay (N7) site in Table 6.1, row 15 can be predicted in one of the groups of A2 or B1, when applying the regional map and land-use information (see Figure 6.5).

5. How can an effective program be built to evaluate the surface water quality?

Assessment of the surface water quality using Fuzzy theory has the advantages such as dealing with an unlimited number of parameters, and converting the quantified monitored values into quality groups. However, one of its limitations is the uncertainty because of different methods of assigning weights. Therefore, GIS’s techniques including computerized
watershed and spatial regression models are suggested to reduce this vagueness of fuzzy method. Consequently, in order to reduce the inconsistency in the evaluation process and have the most exact results of the water quality, the proposed monitoring program is recommended to have 2 main parts: long-term schedule for 5 years and annual assessment program, shown in Figure 8.2.

*Long-term scheduled program*

In order to build a model for quality assessment that can make a closest prediction of the water quality, two conditions must be met: the most probable impact factors to the water quality, such as population growth or land-use impact, and the purposes of the assessment, for example: to find the pollution level against the standard, or to track the extreme values to investigate the instability of the water environment.

Therefore, this schedule program is needed to be designed and conducted before the actual monitoring program to control the vagueness of the assessment process. This stage is set and revised every 5 years. Furthermore, by stabilizing the number and types of quality parameters, a pollution trend for 5 years can be conducted. Three outputs are integrated watershed model, list of chosen quality parameters being tested and chosen weighting method (Figure 8.3).
Annual assessment model

After defining the integrated watershed model which includes the information of the influences, types of parameters need to be monitored, and the weights of the chosen parameters, the monitored values are then collected by taking samples at the monitoring sites and testing its values at the laboratory. The monitored datasets of the quality parameters are then assessed by two methods: Water Quality Index and Fuzzy Comprehensive methods, shown in Figure 8.4. Details about the calculation of the two methods are presented in Chapter 5: Integrated assessment for surface water quality.

6. Is the built program applying fuzzy logic and GIS’s techniques applicable for a developing city, such as Ho Chi Minh city, Vietnam?

The monitoring program of the study area is managed by the Department of Natural Resources and Environment. The monitored values of all the monitoring sites are then evaluated by an integrated method such as Water Quality Index to get an overall water quality, and the result is reported to the City’s Council for setting the management plan. However, the program has weaknesses including inconsistent and improper quality parameters, and Lack of integrated assessment method. Details are discussed in Chapter 8, part 8.1.
Therefore, the built program for assessing the water quality based on fuzzy logic and GIS’ techniques, discussed in Chapter 8, part 8.3, is appropriate for the study area because it limits the weaknesses of the current program including inconsistency and improperartion of the chosen parameters and limitation in the assessment method.

Besides, because the proposed program is built for long term assessment, and it is also taking to account the impact factors (in the watershed area model), it is appropriate for any study area which has fast economic development and thus has high influence to the water quality. Furthermore, the method has been developed and well applied in a developing city: Ho Chi Minh City, Vietnam, where sampling data is still limited; therefore it could be applied to any area which has problem about data shortage.

9.1.2 Conclusion on the research’s contribution

Contribution to the community of the study area

The evaluation model, integrating fuzzy logic and GIS, has been tested applying the dataset in 2015 from Ho Chi Minh City, Vietnam as a case study. Water quality of the study area in 2015 was found to be highly polluted. Also by applying the regression models, population density, land-use, business development and watershed area were identified to influence the surface water quality. Furthermore, a proposed program for surface water quality assessment is then built comprising of long term program and annual
assessment model. This built program is supposed to be more precise and effective than the current monitoring program.

**Contribution to the knowledge on surface water quality assessment**

Environmental quality assessment based on fuzzy logic has been researched and shown its appropriateness. However, it has weaknesses such as data loss, and uncertainty. Therefore, this study focuses on finding a method to reduce the uncertainty of the fuzzy model.

By researching the conducting steps and using the monitored water quality dataset in 2015 of the study area for testing the model, the uncertainty was found to be the weight. Therefore, defining an appropriate weighting method is a way to reduce the uncertainty.

As mentioned above, two conditions need to be fulfilled in order to identify a suitable weighting method: the purpose of the assessment and the types of parameters being tested. As a result, beside the assessment’s purpose, GIS’ techniques are proposed to be used to predict the chosen parameters, thus help to choose the possible weighting method, as a results support to lessen the uncertainty of the fuzzy model.

By integrating the actual environment in the watershed model, possible impact factors and pollution trend can be recognized, and therefore help to
build a precise list of tested parameters. For example, if the land-use information is residential and population growth rate is highly related to the water quality, then anticipated pollution trend can be biological and the chosen parameters can be Coliform and Biological oxygen demand, whereas heavy metal indicators such as Cadmium or Mercury can be omitted.

Furthermore, the integrated model of the actual environment can also help to clarify the quality results when the monitoring site receive more than two quality group. For example, quality value of Nga Bay (N7) site in Table 6.1, row 15 can be predicted in one of the groups of A2 or B1, when applying the regional map and land-use information (Figure 6.5).

9.2 Recommendations

9.2.1 Recommendations on the assessment model comprising Fuzzy logic and GIS

The fuzzy process comprises of two main parts: the matrix and the weight. For the matrix, this study has applied the simplest graphs of membership to build the matrix. These are triangle and semi-trapezoidal shapes. Although these shapes have proved their capability in some recent studies (L.Ma, 2010; Runsheng LV, 2011; Shen Jihong, 2011), more research on different types of graphs should be conducted to test which types of shape are more suitable for assessing the environmental quality, especially surface water. For the weights, results from the regression models by GIS showed that they
can probably lessen the vagueness of the weights of the fuzzy model. Therefore future research focusing on multivariate regression should be conducted to test the relation among the quality parameters and the group of human impact factors. Also, as not many studies on fuzzy analysis have been conducted in the study area of Ho Chi Minh City, more research about using this methodology on environmental assessment in this area should be made to more precise analysing the environmental quality of the area.

Besides, when testing the assessment model combining fuzzy logic and GIS with the water dataset in the study area: Ho Chi Minh City, Vietnam, the watershed area has not been perfectly built because of the naturally geographical feature of the study area which is low elevation and flat. Therefore, more research should be conducted to refine the DEM, a fundamental map to build the watershed model, in a very flat and low elevated area.

9.2.2 Recommendations on future research

In adding the monitoring sites to the watershed model to build the sub-watershed map for testing the relation between the water quality and the impact factors, 35 out of 41 sites have been put into the model as the inlets/outlets of the sub-watershed area. Among the six omitted sites, two have been placed out of the watershed area, and the other four have been deleted because they have been placed too close to each other. Furthermore, when adding the monitoring sites to the watershed area, the built sub-
watershed model had total covered area of about 70% of the watershed area (see Figure 7.3). These problems resulted in the model not working perfectly.

Therefore, serious research should be undertaken to calculate the optimal number of the monitoring sites and their location, particularly in a very flat area such as Ho Chi Minh City, Vietnam. From then, a 100% covered sub-watershed area could be built and therefore enable the model to work more perfectly.
References


Pham, H. G. (2009). Study on water quality index calculation for planning and usage *Journal of Science and Technology for hydrological planning, 3*.


Study on surface water partition in Ho Chi Minh City using water quality index WQI and suggestion possible use. (2007). Retrieved from Ho Chi Minh City Department of Sciences and Technologies


Wuxia Bi, B. W., Zhe Yuan, Mao Ye, Cheng Zhang, Yu Zhao, Dengming Yan, Ting Xu. (2018). Evolution Characteristics of Surface Water Quality Due to Climate Change and LUCC under Scenario Simulations: A Case Study in the Luanhe River Basin.


Yingjie, C. (2011). The research of the evaluation indicator system for enterprise scientific information services. 5.


