2013-09-09

Development and Application of Water Quality Classification Models

Akbar, Tahir Ali

UNIVERSITY OF CALGARY

Development and Application of Water Quality Classification Models

by

Tahir Ali Akbar

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CIVIL ENGINEERING, SCHULICH SCHOOL OF

ENGINEERING

CALGARY, ALBERTA

SEPTEMBER, 2013

Abstract

Though surface water quality is a dynamic quantity; factors, such as increase in population, changes in climate, and anthropogenic activities impose more variability in recent times. The main objectives of this thesis were to: (i) develop models for classification of raw surface water quality, (ii) analyze the spatial patterns and temporal trends of surface water quality, (iii) obtain exceedances of parameters in each class; and (iv) develop remote sensing based models for Canadian Water Quality Index (CWQI) and turbidity. A methodology was developed using principal component analysis (PCA) and clustering techniques on the basis of 19 water quality parameters for 18 lakes of Alberta. Three principal components (PCs) were indicators of hardness, salinity and biological activities for lakes. The surface water quality showed deterioration as the cluster number increased from 1 to 5. The most deteriorated quality of water was found in Cardinal Lake, Moonshine Lake, Winagami Lake, Miquelon Lake and Saskatoon Lake. A total exceedance model was developed for clusterization of surface water quality for 12 major rivers of Alberta. The PCs were the indicators of watershed geology, mineralization and anthropogenic activities related to land use/cover for rivers. The clusters showed a strong relationship with CWQI classes. Snow melting deteriorated the surface water quality of rivers due to anthropogenic activities from different land uses/covers. There was increasing trend for the mean exceedance of the parameters as the cluster number increased from low to high. Empirical models were developed for Canadian Water Quality Index and turbidity using 31 scenes of Landsat-5 TM satellite data for the Bow River. The significant models were 14 for CWQI and 12 for turbidity. 100% matching was found for 72% and 83% of data in best-fit models for CWQI and turbidity respectively. The variation in the Bow River water quality was due to climatic changes and irrigation.

Preface

The research presented in this thesis have been published in journals and conferences as listed below:

Journal publications

1. Akbar T. A., Hassan Q. K., Achari G., Development of Remote Sensing Based Models for Surface Water Quality, *CLEAN-Soil, Air, Water* **2013**. DOI: 10.1002/clen.201300001 (In Press).

2. Akbar T. A., Hassan Q. K., Achari G., Clusterization of Surface Water Quality and Its Relation to Climate and Land Use/Cover. Journal of Environmental Protection **2013** Volume 4, Issue 4, pp. 333–343.

3. Akbar T. A., Hassan Q. K., Achari G., A Methodology for Clustering Lakes in Alberta on the Basis of Water Quality Parameters. *CLEAN-Soil, Air, Water* **2011** Volume 39, Issue 10, pp. 916–924.

Conference proceedings

4. Akbar T. A., Hassan Q. K., Achari G. (2010). A Remote Sensing Based Framework for Predicting Water Quality of Different Source Waters. Oral presentation. Canadian Geomatics Conference and Symposium of Commission I, Convergence in Geomatics – Shaping Canada's Competitive Landscape, International Society for Photogrammetry and Remote Sensing (ISPRS), Calgary, Alberta, Canada, May 15-18, 2010.

Abstract presentations

5. Akbar T. A., Hassan Q. K., Achari G. (2013). Developing Remote Sensing-Based Models for Quantifying Water Quality. Oral presentation. Remote Sensing & Monitoring Forum, Petroleum Technology Alliance Canada (PTAC) and Leading Operational Observations & Knowledge for the North (LOOKNorth), Calgary

Petroleum Club, Devonian Room, Calgary, May 27 2013.

6. Akbar T. A., Achari G., Hassan Q. K. (2012). Surface Water Quality Classification and Analysis Using Machine Learning and MODIS. Oral presentation. Annual scientific meeting of Canadian Geophysical Union (CGU) and The Canadian Society of Agricultural and Forest Meteorology (CSAFM), Banff Park Lodge, Banff, Alberta, Canada, June 5-8, 2012.

7. Akbar T. A., Achari G., Hassan Q. K. (2011). Multivariate Statistical and GIS based Methodology for Source Water Quality Classification. Oral presentation. National conference, Canadian Geophysical Union, Banff Park Lodge, Alberta, Canada, May 15-18, 2011.

Poster presentation

8. Akbar T. A., Achari G., Hassan Q. K. (2009). GIS-RS Based Decision Support System for Water Quality Management in Alberta. Poster presentation. 3rd International poster conference on advances in civil engineering, Schulich School of Engineering, University of Calgary, Calgary, Canada, Nov. 2009.

Acknowledgements

I am sincerely grateful to Dr. Gopal Achari and Dr. Quazi K. Hassan for their outstanding supervision, sincere support and encouragement during my PhD program. The technical inputs in research work and support in journal publications from Dr. Quazi K. Hassan is highly appreciated. I would also like to thank to my supervisory committee members for their support in my research work.

I would like to acknowledge funding from: (i) Natural Sciences and Engineering Research Council of Canada (NSERC) for Postgraduate Scholarship at Doctoral level, and (ii) University of Calgary. I also acknowledge: (i) Alberta Environment and Environment Canada for giving the water quality data, (ii) United States Geological Survey (USGS) for supplying the satellite data for Landsat TM-5, (iii) NASA for providing MODIS-based annual composite land use/cover map, (iv) Atlas of Canada and Geobase for giving the geospatial data, and (v) N. Sekhon for providing MODIS-derived snow melting map respectively. My appreciation goes to my father, mother, wife, children, brother, and sisters for their love, support, encouragement, and patience.

Dedication

To my parents and family

For

Their support, encouragement & inspiration.

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AHP	Analytic Hierarchy Process
AISA	Airborne Imaging Spectrometer for Applications
ALK	Alkalinity
ANOVA	Analysis of variance
AR-1 AR-2 AR-3	Three sampling sites of Athabasca River
AVHRR	Advanced Very High Resolution Radiometer
B	Blue
BCWQI	British Columbia Water Quality Index
BOR-1 BOR-2 BOR-3 BOR-4	Four sampling sites of Bow River
BR-1 BR-2	Two sampling sites of Battle River
CASI	Compact Airborne Spectrographic Imager
Ca	Calcium
CCME	Canadian Council of Ministers for the Environment
CHL-a	Chlorophyll-a
Cl	Chloride
CO_3	Carbonate
CWQI	Canadian Water Quality Index
CZCS	Coastal Zone Color Scanner
DA	Discriminant Analysis
DEM	Digital Elevation Model
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
EM	Expectation Maximization
ETM+	Enhanced Thematic Mapper Plus
ER-1	One sampling site of Elbow River
F	Fluoride
F_1	Factor 1(scope)
F_2	Factor 2 (frequency)

F_3	Factor 3 (amplitude)
FA	Factor Analysis
FC	Fecal Coliforms
Fe	Iron
G	Green
GAC	Granular Activated Carbon
GIS	Geographic Information System
GloVis	Global Visualization Viewer
HACA	Hierarchical Agglomerative Cluster Analysis
HARD	Hardness
HCO_3	Bicarbonate
HGA	Hierarchical Group Analysis
HRV	High Resolution Visible
K	Potassium
$KMnO_4$	Potassium Permanganate
LIDAR	Laser Imaging Detection and Ranging
LTRN	Long-Term River Network
MANOVA	Multivariate Analysis of Variance
MERIS	Medium Resolution Imaging Spectrometer
Mg	Magnesium
Mn	Manganese
MODIS	Moderate Resolution Imaging Spectroradiometer
MOS	Marine Observation Satellite
MR-1	One sampling site of Milk River
MSS	Multispectral Scanner
Na	Sodium
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
NR-1, NR-2	Two sampling sites of North Saskatchewan River

NOM	Natural Organic Matter
NTU	Nephelometric Turbidity Unit
OIP	Overall Index of Pollution
OR-1 OR-2	Two sampling sites of Oldman River
OWQI	Oregon Water Quality Index
PC	Principal Component
PCs	Principal Components
PCA	Principal Component Analysis
PR-1	One Sampling Site of Peace River
R	Red
r^2	Co-efficient of Determination
RADAR	Radio Detection and Ranging
RDR-1, RDR-2	Two sampling sites of Red Deer River
RS	Remote Sensing
SC	Specific Conductivity
SD	Secchi Depth
SiO_2	Silica
SO_4	Sulfate
SR-1	One Sampling Site of Smoky River
SSR-1	One Sampling Site of South Saskatchewan River
SSE	Sum of Squared Error
TC	True Color
TDS	Total Dissolved Solids
TH	Total Hardness
TKN	Total Kjeldahl Nitrogen
TM	Thematic Mapper
TN	Total Nitrogen
TP	Total Phosphorus
TUR	Turbidity
USGS	United States Geological Survey
UV	Ultraviolet

WQI	Water Quality Index
WR-1, WR-2	Two Sampling Sites of Wapiti River
WT	Water Temperature

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Background

In Alberta, Canada like in many jurisdictions of the world, the demand for safe drinking water is growing with a growth in population and in the economy (Alberta Environment 2010). The main source of water for Albertans is surficial such as lakes and rivers. The major rivers of this province originate from glaciers and snowpacks located in high elevation of the Rockies. Snowmelt and rainfall are the largest contributors to the annual flows of the rivers (Alberta Environment 2010). The quality of water in the rivers changes as water flows through different land covers and watersheds (Bolstad & Swank 1997). Land-use defines the type and amount of contaminants that flow into the water bodies (Moss 1998). The deterioration in surface water quality occurs due to various land uses like agriculture, forests, residential, commercial and industrial and the runoff from these into water bodies. There are six different natural regions in Alberta, which impact the quality of surface water. These include: (i) Boreal Forest, (ii) Canadian Shield, (iii) Parkland, (iv) Foothills, (v) Rocky Mountain, and (vi) Grasslands. The precipitation in boreal forests is higher as compared to that in grasslands and parklands. The dark brown color of surface waters in lakes and rivers of boreal forests are caused by presence of natural organic carbon. The presence of fine rock particles and its dissolved components impact the quality of surface waters in the Canadian Shield. The surface water in the Shield is neutral to slightly alkaline and low in nutrients. The water in the Rocky Mountains is low in nutrients due to absence of rich soil. Due to runoff from agricultural land, the surface waters have higher concentrations of nutrients. Similarly, the impact of parkland and grasslands lead to changes in surface water quality (Alberta Environment 2010).

Climate changes during to global warming can lead to floods, drought, biodiversity loss, and even increase in infectious diseases which can degrade the water quality (Murdoch et al. 2000, Watson et al. 1996, McKnight et al. 1996, Cushing 1997). Thus, land use changes when combined with climatic changes can lead to significant deterioration in the water quality of surface water bodies (Delpla et al. 2009). In addition, the temperature changes due to seasonal impacts will impact the quality of surface waters in Alberta. These variations are substantial and can cause increased dissolution/precipitation of minerals significantly impacting the water quality. Because of these variations, it is important to understand the surface water quality and to classify it based on its characteristics such that the impact of land use/cover can be understood and targeted treatment can be introduced. While changes to water quality and its classification can be studied anywhere, Alberta has been identified as a target jurisdiction for detailed study and classification.

In Alberta, the surface water quality is monitored for lakes and rivers on a regular basis. The water samples are collected monthly for the rivers while for lakes the water samples are collected only in the summer months. The quality of water is judged by comparing the concentrations of various parameters against the water quality guidelines set by Health Canada and Alberta Environment. The surface water quality data can be utilized to obtain and study patterns in quality changes (Eneji et al. 2012; Singh et al. 2010). Classification methods can be utilized to identify dominant factors that impact water quality. With geographic information system (GIS), the water quality data can also be used to obtain spatial patterns and temporal trends to analyze the impact of land use/cover and climate on surface water quality. Any water, prior to consumption has to be treated. It is possible to target the treatment based on the quality of source waters to be most effective. This has significant and almost immediate cost benefit to the community.

It is understood that physical collection and analysis of water samples is labour intensive and time consuming. . It is impractical to physically collect and analyse the samples over a large geographic area. Further, to analyze the surface water quality for the whole water body spatio-temporal aspects have to be studied. To address both the spatial and temporal variability, remote sensing based models are useful (Mancino et al. 2009; Vignolo et al. 2006; Wang et al. 2006). In our study, application of remote sensing will have significant benefits over traditional water quality monitoring methods. These benefits include: (i) obtaining spatial coverage for large water bodies, (ii) developing maps for water quality parameters, (iii) classifying the surface water quality, and (iv) conducting the spatial and temporal analysis.

1.2 Thesis objectives

While it is understood that this is a large problem and everything can not be answered in this thesis, the main objectives of this research are as following:

1. Develop methodologies and models to cluster Alberta waters based on water quality.
2. Analyze the spatial patterns and temporal trends of surface water quality.
3. Obtain exceedances of parameters in each cluster.
4. Develop remote sensing based models for Canadian Water Quality Index (CWQI) and turbidity.

To meet these main objectives there are six specific objectives:

1. Development of a methodology using principal component analysis (PCA) and clustering techniques on the basis of water quality parameters for 18 lakes of Alberta.
2. Development of a model for clusterization of surface water quality of 12 major rivers of Alberta. Validation of clusters using CWQI. Application of clusters for spatio-temporal analysis and impact of climate and land use/cover.
3. Development of a model for obtaining parameter exceedance in surface water quality for 12 major rivers of Alberta. Review of literature on treatment technologies for the exceeded parameters.
4. Development of remote sensing based models for obtaining CWQI classes using the planetary reflectance of Landsat-5 TM and ground-measured data for Bow River of Alberta.
5. Development of remote sensing based models for obtaining turbidity using the planetary reflectance of Landsat-5 TM and in-situ data for Bow River of Alberta.
6. Apply the selected remote sensing models to classify the surface waters of the Bow River into CWQI and turbidity classes for spatial and temporal analysis.

1.3 Thesis structure

This thesis consists of seven chapters. The brief description for each chapter is given below:

Chapter 1: In this chapter the background information on the Alberta surface water quality is given. It also provides objectives and structure of this thesis.

Chapter 2: This chapter presents literature review on water quality index, multivariate statistical analysis techniques, remote sensing, and geographic information system related to the surface water quality.

Chapter 3: This chapter provides the research work accomplished on the development and application of a methodology for clustering 18 lakes in Alberta using the mean annual data of 19 water quality parameters during the period of 11 years (1988–2002).

Chapter 4: This chapter presents the research work accomplished on the development and application of a clusterization model for analyzing the surface water quality of 12 major rivers of Alberta on the basis of 17 parameters during the period of five years (i.e., 2004-2008).

Chapter 5: This chapter provides the research work accomplished on the (i) development and application of exceedance model for obtaining the exceeded parameters for the surface water quality of 12 major rivers of Alberta, and (ii) review on treatment technologies for the exceeded parameters.

Chapter 6: This chapter presents the research work accomplished on the development and application of remote sensing based models for Canadian Water Quality Index and turbidity for the Bow River of Alberta using the satellite and ground measured data during the period of five years (i.e., 2006-2010).

Chapter 7: This chapter provides the concluding remarks, contribution and recommendations on the research work accomplished in this thesis.

It should be noted that this thesis has led to a number of publications. This thesis is written in "paper format" which means all chapters (except 1, 2 and 7) have been presented as stand-alone papers. Figure 1.1 shows a schematic diagram of the objectives and chapters of the thesis.

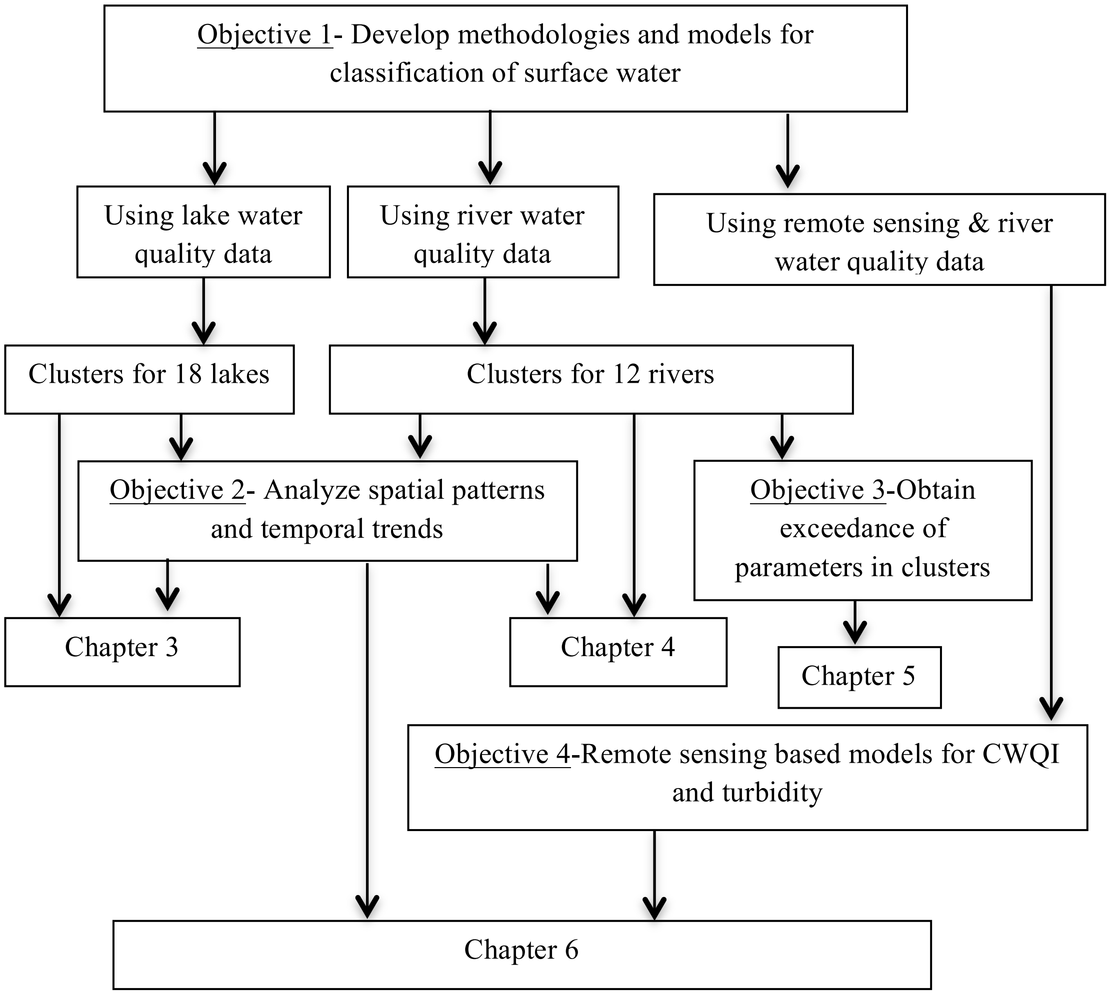

Figure 1.1: Schematic diagram of the objectives and chapters of thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Canada is a water-rich country and it has 20% of world's fresh water. Alberta has 2.2 % of Canada's freshwater (Alberta Environment 2010). In Alberta, 97.5% of consumptive use of water is from surface water. The major uses of surface water in Alberta are drinking, agriculture, industry and recreation. The source for various major rivers in Alberta is glaciers in Banff and Jasper National Parks.

Many natural and anthropogenic land use activities influence the water quality of the rivers. Snowmelt and precipitation runoff from various land use activity areas like wood logging, agricultural, mining and urban development can impact the water quality (Alberta Environment 2010). Non-point point source pollutants from land-use activities might include sediments, nutrients and other contaminants. Examples include: Intense agricultural activities found in South Saskatchewan River Basin, which consists of six major rivers (i.e., Bow River, Elbow River, Oldman River, Red Deer River and South Saskatchewan River (Bruneau et al. 2009; Toth et al. 2009); deteriorating water quality of Elbow River due to the runoff from the agriculture and residential developments (Sosiak and Dixon 2006); the river water quality for the Oldman River Basin deteriorating due to the anthropogenic activities like forestry, recreation, oil and gas development, and agriculture (Koning et al. 2006); the naturally occurring process of sulfide oxidation in Oldman River Basin observed due to the presence of extensive network of drainage and irrigation canals (Rock and Mayer, 2008); Runoff from forested and agricultural lands as a major potential source of contamination for North Saskatchewan River (Zhang and Stanley 1997); Deterioration of the surface water quality for Athabasca River and Wapiti River due to the discharge of sewage effluent from pulp mill and municipalities (Chambers et al. 2001); The deterioration in the water quality of Peace River and Smoky River due to oil sands refinery discharges and runoff from the forest and agricultural activities (Wrona et al. 2000) and the discharge of wastewater from various towns and cities deteriorating the water quality of Battle River (Anderson 1999).

Due to different natural and climatic conditions and associated anthropogenic activities nearby, the water quality varies significantly in the different rivers of Alberta.

This means the water may require varying levels of treatments prior to human consumption. It is quite possible that in some places the water quality is intrinsically good and advanced levels of treatment may not be necessary. For cost effectiveness, the water treatment should be targeted towards the pollutants of concerns that exist in a particular water body. This will lead to effective savings and proper utilization of resources. To this end, it is important to classify the surface water quality in Alberta. The classification can be used to analyze the surface water quality spatially and temporally.

2.2 Water Quality Index

Water quality indices are used to monitor the water quality. It is a mechanism based on numerical expression for defining the level of water quality (Bordalo et al. 2006). The large amount of complex data is summarized into simplified mathematical numbers, which can be interpreted into text classes (e.g., excellent, very good, good, moderate, poor etc.). In a study the water quality index was developed by considering 10 most commonly measured parameters, which are, dissolved oxygen, pH, coliforms, specific conductance, alkalinity, and chloride. The index score was obtained with a linear sum aggregation and index score range was from 1 to 4 (Horton 1965). A multiplicative water quality index was developed on the basis of weights assigned to every parameter. The weights were given to the parameters subjectively. The weight-based index was found useful and had significant impact on the indices (Brown et al. 1972). Other studies also incorporated weight-based schemes in their indices (e.g. Bolton 1978; Inhaber 1975). An index was developed on the basis of empirical data for recreational waters. In this index, sensitivity functions were used to assign a numerical value between 0 and 1. Negative exponential curves were used to represent the sensitivity functions. Sub-indices were defined which were combined to obtain the geometric mean (Walski and Parker 1974).

The surface water quality for the major rivers of Alberta is monitored using Long-Term River Network (LTRN) program and the Canadian Water Quality Index (CWQI). In LTRN program, the representative water is collected for each of the major rivers at fixed sampling sites. The samples are tested for a large number of parameters every month. For the suitability of water for specific uses, the water quality is evaluated on the basis of Canadian water quality guidelines and the quality is considered to be acceptable when the measured values are within the limits of guidelines (Alberta Environment 2010). The Canadian Water Quality Index (CWQI) is a tool implemented by Canadian Council of Ministers of the Environment (CCME) to provide reports on water quality in Canada. There are three important factors in CWQI. All factors are calculated on the basis of objectives, which provide guideline values developed by Federal-Provincial-Territorial Committee on Drinking Water (CCME 2001). The different equations used for calculation of CWQI are given in Table 2.1.

Table 2.1: Equations used for calculation of CWQI

$$F_1\,(\text{Scope}) = \left(\frac{\text{Number of failed parameters}}{\text{Total number of parameters}}\right) \text{ x } 100$$

$$F_2\,(\text{Frequency}) = \left(\frac{\text{Number of failed tests}}{\text{Total number of tests}}\right) \text{ x } 100$$

$$\text{excursion}i = \left(\frac{\text{Objective}j}{\text{Failed test value}i}\right) - 1$$

$$\text{excursion}_i = \left(\frac{\text{Failed test value }_i}{\text{Objective}_j}\right) - 1$$

$$\text{nse} = \left(\frac{\sum_{i=1}^{n} \text{excursion}i}{\text{number of tests}}\right)$$

$$F_3\,(\text{amplitude}) = \left(\frac{\text{nse}}{0.01\text{nse} + 0.01}\right)$$

$$\text{CWQI} = 100 - \left(\frac{\sqrt{F1^2 + F2^2 + F3^2}}{1.732}\right)$$

Where:

F_1 or Factor 1(scope) is the numbers of parameters for which objectives were not met;

F_2 or Factor 2 (frequency), is the percentage of tests that do not meet the objectives;

F_3 or Factor 3 (amplitude), which shows the amount by which failed tests do not meet the objectives; and nse is, normalized sum of excursion.

The CWQI produces value between 0 and 100 where 0 indicates poor water quality and 100 represents excellent water quality. The water quality is ranked into five categories, which are 1-Excellent (95-100), 2-Good (80-94), 3-Fair (60-79), 4-Marginal (45-59), 5-Poor (0-44). The advantages of CCME WQI include: (i) user-friendly format for understanding the overall general water quality (Rosemond et al. 2009), and (ii) representation for the measurements of large number of parameters into a single index value (CCME 2001). The limitations in using CCME WQI are (i) loss of information by combining the different parameters to obtain a single index value (Rosemond et al. 2009), (ii) loss of interaction between the parameters (Zanderbergen and Hall 1998), (iii) sensitivity to input parameters (Khan et al. 2004), and (iv) Limitation for evaluating spatial changes in water quality (Rosemond et al. 2009). In addition to these limitations, CCME WQI needs large number of parameters obtained by physical monitoring of water quality. Such type of physical monitoring is labour intensive, time consuming and costly.

In a study conducted for Mackenzie River basin of Canada, CCME WQI and a statistical approach was used to monitor water quality and it was found that the river is influenced by high turbidity and total trace elements due to high suspended sediment loads (Lumb et al. 2006). Another study used CCME WQI for comparative analysis of regional water quality in Canada and found it to be a good tool for assessment of water quality (Rosemond et al., 2009). CCME WQI values are calculated annually for each sampling site of a major river on the basis of data collected monthly or quarterly (Alberta Environment 2010).

In addition to CWQI, we have reviewed some other indices here. The Canadian Ministry of Environment developed the British Columbia Water Quality Index (BCWQI) (Rocchini and Swain, 1995). It is given in Eq. (2.1)

$$\mathrm{BCWQI} = \left(\frac{\sqrt{F1^2+F2^2+(\frac{F3}{3})^2}}{1.453}\right) \qquad (2.1)$$

Where:

F_1 is the numbers of parameters for which objectives were not met;

F_2 is the percentage of tests that do not meet the objectives; and

F_3 is the amount by which failed tests do not meet the objectives.

Like CWQI, BCWQI works on the basis of Water quality guidelines. The accuracy of BCWQI improves by increasing the frequency of sampling. The disadvantage of this index is does not indicate the water quality trend until it deviates from the guideline values (Salim et al. 2009).

The Oregon Water Quality Index (OWQI) determines the water quality on the basis of integrated measurements of eight parameters which are temperature, dissolved oxygen, biochemical oxygen demand, pH, ammonia+nitrate nitrogen, total phosphorus, total solids, and fecal coliform (Cude 2001). OWQI is applied for water quality for recreation, swimming and fishing. Mathematically the model is represented as in Eq. (2.2) (Cude 2001):

$$\mathrm{OWQI} = \sqrt{\frac{n}{\sum_{i=1}^{n} \frac{1}{SI_i^2}}} \qquad (2.2)$$

Where n is the number of sub-indices and SI_i is sub-index i.

The benefits of this model include: (i) simplicity, (ii) easy interpretation, (iii) spatial analysis, and (iv) temporal patterns. The disadvantages are: (i) loss of information, (ii) parameter specific, (iii) site specific, and (iv) usage specific.

Overall Index of Pollution (OIP) was developed by Sargaonkar and Deshpande (2003) for Indian Rivers on the basis of measurements and classification of pH, turbidity, dissolved oxygen, BOD, hardness, total dissolved solids, total coliforms, arsenic, and fluoride. The water quality observation is scored as excellent, acceptable, slightly polluted, polluted, and heavily polluted on the basis of water quality guidelines of India, World Health Organization and European Community. After categorization, each sampling record for a parameter is given a pollution index. OIP is calculated using Eq. (2.3) (Sargaonkar and Deshpande 2003):

$$\mathrm{OIP} = \sum_{i=1}^{n} \frac{P_i}{n} \qquad (2.3)$$

Where P_i is pollution index for i^{th} parameter and n is number of parameters.

2.3 Multivariate statistical analysis techniques

There are various water quality parameters, which are determined to monitor overall water quality. Water samples are obtained from water bodies to test physical, chemical and biological properties in the laboratories. Multivariate statistics is a useful way to analyze source water quality on the basis of historical water sampling data. There are different multivariate statistical methods, e.g., Multivariate analysis of variance (MANOVA), Canonical correlation analysis, Discriminant analysis, Principal components analysis (PCA), and Cluster analysis. Multivariate analysis of variance (MANOVA) is used when there are two or more dependent variables. Canonical correlation analysis is used to find linear relationship in two sets of variables. Discriminant analysis is used to differentiate between two or more groups of cases. The advantages of multivariate statistical methods include: (i) reduction in complexity of data (Bengraïne and Marhaba 2003), (ii) unbiasedness in methods (Wenning and Erickson 1994), (iv) usefulness in water quality studies (Areerachakul and Sanguansintukul 2010). The disadvantages of multivariate statistical methods include (i) subjectivity (Liu et al. 2003), (ii) unreliability, and (iii) repetition of data. PCA and cluster analyses are discussed in the following sub-sections.

2.3.1 Principal Component Analysis (PCA)

If a specified number of parameters are recorded but there exists inter-correlations between the parameters such that they move in tandem, then this can lead to erroneous conclusions. Principal Component Analysis (PCA) can be used to create a new set of orthogonal variables, which contain the same information as the original set. In PCA, all correlated parameters are combined into different principal components (PCs) with positive and negative loading values, which can be used to interpret major processes, involved in analyzing and characterizing the water quality. The use of PCA before clustering was suggested by Ben-Hur and Guyon (2003). It is useful to make use of data with higher variance and remove the data with low variance. The application of PCA in several studies has been discussed in the subsequent paragraphs:

(i) In a study the spatial and temporal variations of water quality was analyzed in Sanya Bay of China using three-way principal component analysis. The water quality of one sampling station was influenced by Sanya River and the water qualities of other nine stations were impacted by South China Sea. It was also found that Sanya River as a source of pollution. The influence of dry and rainfall season was observed on the water quality of Sanya River. (Dong et al. 2010).

(ii) The principal component analysis was applied for analyzing the water quality of Neckar River, Germany on the basis of ten parameters during the period of five years (1993-1998). Four principal components were identified accounting for 72% of total variance. The principal component analysis was interpreted for: (i) biological activity, (ii) dilution by high discharge, (iii) seasonal effects, and (iv) wastewater impact. Eutrophication was the reason for the deteriorated water quality (Haag et al. 2002).

(iii) Groundwater samples from 10 different sources were collected in three different years for 10 parameters. Q-mode principal component analysis was used to classify water samples into four principal components. This classification could help planners and field engineers for improvement of field data collection and preventing groundwater contamination (Mahapatra 2012).

(iv) Principal component analysis was used to identify the factors, which caused variation in water quality of Porsuk Tributary in the Sakarya river basin. Six principal components explained 70% of the total variance of the data. It was found that small domestic waste discharge, industrial waste discharge, nitrification and seasonal effects were responsible for the variation of the water quality (Mazlum et al. 1999).

(v) Principal Component Analysis was used for water characterization and seasonal heavy metal distribution in the Odiel River of Spain. PCA showed that the first component accounted for 40.88% of total variance. The second PC showed the minority metals, such as nickel, cobalt, and cadmium. Heavy metals showed three different seasonal patterns. (Montes-Botella and Tenorio 2003).

(vi) The evaluation of river water quality monitoring stations was done by PCA for assessment on annual variations for the water quality of St. Johns River in Florida, USA. The principal factor analysis (PFA) was used to identify the important water quality parameters. It was found that the important parameters were total organic carbon, dissolved organic carbon, total nitrogen, dissolved nitrate and nitrite, orthophosphate, alkalinity, salinity, calcium, and magnesium (Ouyang 2005).

(vii) PCA was applied to identify six major factors, which explained 71% of total variance using the water quality data of 24 parameters for a period of 5 years (1994-1998) at three sampling sites of Gomti River in India. Three groups of similarity for sampling sites were identified using cluster analysis. The large variations in temporal and spatial analysis were obtained using discriminant analysis (Singh et al. 2004).

(viii) Statistical techniques were used to find spatial variation and source water pollution for Qiantang River. Three pollution zones were developed, low, moderate and high. With factor analysis, it was found that there are two pollution sources in each of low and moderate pollution sources that explained 67% and 73% of total variance respectively. It was found that there are three pollution sources in high pollution zone with a total variance of 80%. The potential sources of pollution were industrial wastewater, agricultural activities and urban runoff (Huang et al. 2010).

2.3.2 Cluster analysis

In cluster analysis, the objects are grouped on the basis of similarities within a class and dissimilarities among different classes (Panda et al. 2006). The similarities and dissimilarities are obtained on the basis of distance measures which are Euclidean and Manhattan (Kaufman and Rousseeuw 1990). There are two types of cluster analysis, which are partitioning and hierarchical methods.

2.3.2.1 Partitioning methods

The partitioning methods belong to a class of cluster-based methods, which assign weight vectors to the clusters. The common partitioning methods are discussed in the subsequent paragraphs. K-means is a simple and efficient algorithm. It divides *n* observations into K clusters and each observation belongs to cluster with nearest mean. It uses the sum of square error criteria. The cluster pattern is assigned when sum of square error is minimum. The sum of square error equation (SSE) for K-means is given in Eq. (2.4) (Kanungo et al. 2002):

$$SSE = \sum_{C_i} \sum_{x \in C_i} \left\| x - m_i \right\|^2 \qquad (2.4)$$

where m_i is the mean of the i^{th} cluster and $x \varepsilon C_i$ is a pattern assigned to that cluster. The K-means clustering has advantage over other methods as it can be used to assign new cases to the existing clusters.

K-mediod selects data point as centers (medians). After finding medians of clusters, the clusters are developed by assigning each object of dataset to the nearest medians of the clusters. The dissimilarities from each of the objects in the dataset from these medoids of the clusters are determined using Euclidean distance or Manhattan distance. Medoids are selected on the basis of the minimum distance. Silhouette is used for interpretation and validation of clusters. Silhouette is a graphical representation for defining the position of an object within its cluster (Kaufman and Rousseeuw 1990). The cluster centres are not affected by outliers in K-mediods. K-mediod can be applied to obtain the cost between any two points using Eq. (2.5) (Theodoridis and Koutroumbas 2006).

$$\text{cost}\,(\text{x}, \text{c}) = \sum_{i=1}^{d} |X_i - C_i| \qquad (2.5)$$

where x is any data object, c is the medoid, and d is the dimension of the object.

Fuzzy partitioning clustering has the capacity to deal with the ambiguity of data. In this clustering, each object can be placed into different clusters. The placement of objects is quantified by membership coefficient, which lies between 0 to 1. This is called fuzzification of the cluster and it is termed as Fanny cluster analysis. The benefit of using this technique is that the objects are not forced to be the part of a specific cluster. The disadvantage of this method is that there is a lot of information for interpretation. Fanny aims at the minimization of the objective function as given in Eq. (2.6) (Theodoridis and Koutroumbas 2006):

$$\text{Objective function} = \sum_{v=1}^{k} \frac{\sum_{ij=1}^{\mathrm{n}} u_{iv}^2 u_{jv}^2}{2\sum_{j=1}^{n} u_{jv}^2} d(i,j) \qquad (2.6)$$

where

d (i , j) is the dissimilarities between objects i and j,

u_{iv} is the unknown membership of object i in cluster v .

The membership functions are subject to the following two constraints:

$$(i)\ u_{iv} \geq 0\ for\ all\ i = 1, -, n\ and\ all\ \mathrm{v} = 1, -, \mathrm{k}.$$

$$(ii)\ \sum_{v=1}^{k} u_{iv} = 1 = 100\%\ for\ all\ i = 1, -, n$$

Both these constraints indicate that the membership cannot be negative and that the total membership of each object was distributed over various clusters. On the basis of this convention, the total membership is normalized to 1.

2.3.2.2 Hierarchical methods

In this method, hierarchy of clusters are made. These are of two types: (i) agglomerative and (ii) divisive (Fielding, 2007). The agglomerative is bottom up approach. In this each observation has pairs of clusters, which are merged as one moves up in hierarchy.

Divisive is top down approach. In this, the clusters are separated as one moves down the hierarchy. In this approach we need to select suitable distance measure and linkage algorithm. Some of the hierarchical methods are discussed here.

In single linkage (nearest neighbour) method, the distance between the two clusters is obtained by the distance of the two closest objects in the different clusters as given in Eq. (2.7) (SAS/STAT 9.2 Users Guide 2009; Székely and Rizzo 2005).

$$\min\{\, d(a,b): a \in A, b \in B\} \tag{2.7}$$

For complete linkage (furthest neighbour) method, the distances between clusters are identified by the largest distance between any two objects in the different clusters as provided in Eq. (2.8) (SAS/STAT 9.2 Users Guide 2009; Székely and Rizzo 2005).

$$\max\{\, d(a,b): a \in A, b \in B\} \tag{2.8}$$

The average distance is calculated between all pairs of objects in the two different clusters in unweighted pair-group average method (See Eq. 2.9) (SAS/STAT 9.2 Users Guide 2009; Székely and Rizzo 2005).

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a,b) \tag{2.9}$$

Cluster based pattern recognition techniques were widely used in the water quality studies and the related examples are:

(i) Cluster analysis was used for interpretation of atmospheric and surface water pollution using major inorganic ions, electrolytic conductivity and pH. They found that the potential sources of pollution were fertilizers usage, road salting and erosion of construction materials (Dubiella-Jackowska et al. 2010).

(ii) The water quality variance was analyzed in fresh and brackish water for 36 sampling stations in Richibucto River drainage basin of New Brunswick using water quality data of 6 parameters from 1996 to 2001. With PCA, it was found that variance in fresh water was due to pH, total organic carbon and salinity with high nutrient concentrations being causative parameters for variance in brackish water. Cluster analysis explained the importance of high concentrations of phosphorous and nitrate in water bodies from treated municipal effluent (St-Hilaire et al. 2004).

(iii) In a study conducted for Jajrood River of Iran, the PCA and cluster analysis were used to evaluate spatial and temporal variation in using monthly water quality monitoring data for 18 water quality variables. PCA identified five factors, which showed 85% of variability. The water sampling monitoring stations were classified using cluster analysis technique and it was found that the most polluted monitoring station was Out-Meygoon. The organic pollution was the source of pollution from Ahar, Baghgol, Rooteh, before Zaygan, Fasham, Roodak and Lashgarak (Razmkhah et al. 2010).

(iv) For partitioning process of Ulansuhai Lake, a multiplex model of fuzzy clustering was developed and applied. The model was developed by integrating transitive closure method, ISODATA algorithm in fuzzy clustering and fuzzy pattern recognition. The model was useful for the determination of functional zones of the lake (Chuntao et al. 2008).

(v) Chemical classification of water was done for Salado River of Argentina. PCA and K-means clustering were applied to the percentages of the major ions. The authors found seven types of waters related to discharges from different sub-catchments. The major reasons for variation in the river water quality were: (i) salts from groundwater, (ii) weathering, and (iii) anthropogenic effects due to diversion (Gabellone et al. 2008).

(vi) The spatial water quality assessment for seven stations of Langat River was investigated using hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA), and the factor analysis (FA). HACA was used to develop three spatial clusters. DA was applied to discriminate six and seven water quality variables. PCA and FA were used to obtain the impact of land use activities on clustered region (Juahir et al. 2011).

(vii) The water quality sampling frequency was calculated by analytic hierarchy process (AHP) for Jingmei and Xindian Rivers of Taiwan. The weighing factors of variables were combined with the relative weights of stations to select sampling frequency for each station. The results revealed that the frequency of sampling should be increased for high weighted stations and it should be decreased for low weighted stations (Do et al. 2012).

(viii) The spatial and temporal variations of main pollutants for water quality in Wen-Rui Tang River watershed was obtained using the geographic information system, cluster analysis and principal component analysis. The results showed that the concentrations of certain parameters were obviously high in tertiary rivers as compared to primary and secondary rivers. The correlation analysis showed that there is negative correlation with 5-day cumulative rainfall and monthly rainfall. The results of cluster analysis indicated that the northern part of the river was highly polluted. PCA indicated that water quality is deteriorated due to anthropogenic activities and poor wastewater management (Lu et al. 2011).

(ix) Cluster analysis, discriminant analysis, and support vector machines were used for analyzing water bodies in the Polish Tatra Mountains. The results from cluster analysis indicated reconsideration for the geographical distinction. With discriminant analysis it was confirmed the geographical separation of water bodies and it also stated that the sampling time is a crucial factor in environmental analysis (Prikler et al. 2003).

2.4 Remote Sensing

Remote sensing is the science of getting information about the earth's surface remotely. It is obtained by recording the reflected energy using electromagnetic radiation by sensors which is sent to ground station where it is converted into satellite data. The satellite data is analyzed, processed and interpreted for different types of applications. The two main types of remote sensing are passive and active. Passive sensors detect the natural radiation, which is reflected by the object e.g. reflected sunlight. Active sensors emit energy to scan the objects of interest e.g. RADAR and LIDAR. The remote sensing data consists of spatial, spectral, radiometric and temporal resolutions. The spatial resolution represents the pixel size and it can be coarse or fine. The large features are visible in coarse resolution whereas the small objects are visible in fine resolution. The spectral resolution describes the wavelength intervals of the different frequency bands recorded. The radiometric resolution gives the number of different intensities of radiation of the sensor that ranges between 8 to 14 bits. The absolute temporal resolution is defined as the time period required to image the same area at same viewing angle for the second time. In Alberta, the water quality is monitored by collecting the water samples physically. The manual water sample collection is laborious and time consuming. The analysis is also site specific and it cannot provide information for large geographic areas. Remote sensing is an alternate and effective way to analyze the spatial and temporal aspects of the surface water quality. The benefits of remote sensing include: (i) spatial analysis for large geographic areas, (ii) temporal analysis for specific period of time or season, (iii) accessibility to remote areas, (iv) economical, and (v) efficient. The disadvantages include: (i) coarse resolution, (ii) difficult data interpretation, (iii) finding the suitable images matching with the date of sampling, and (iv) measurement uncertainty.

The Landsat program is a series of earth-observing satellites under the control of NASA and the US Geological Survey since 1972. Landsat has launched eight satellites for earth observation, which are Landsat 1, Landsat 2, Landsat 3, Landsat 4, Landsat 5, Landsat 6, Landsat 7, and Landsat 8. These satellites were launched in 1972, 1975, 1978, 1982, 1984, 1993, 1999 and 2013 respectively. In this research, we have used the satellite data of Landsat 5-TM. This satellite is still functioning and it has seven spectral bands. The spectral and spatial resolutions of these bands are given in **Table 2.2**. The temporal resolution is 16 days and the image size is 185km x 172km (NASA 2013).

Table 2.2: Spectral and spatial resolution for Landsat 5-TM (NASA 2013).

Band Number	Wavelength (μm)	Resolution (m)
1	0.45-0.52	30
2	0.52-0.60	30
3	0.63-0.69	30
4	0.76-0.90	30
5	1.55-1.75	30
6	10.4-12.5	120
7	2.08-2.35	30

The Moderate-resolution Imaging Spectroradiometer (MODIS) was launched by NASA in 1999 and it has two satellites, which are Terra and Aqua. MODIS capture data in thirty six spectral bands with wavelengths range from 0.4 μm to 14.4 μm and spatial resolution varies between 250 m to 1 km. The spatial resolutions of two bands are at 250 m, five bands at 500 m and twenty-nine bands at 1km for MODIS. The application of Landsat satellite and MODIS data in various water quality studies has been discussed in the subsequent paragraphs:

(i) In a study, the ground measured data of Secchi depth, turbidity and chlorophyll were used to develop models using the surface reflectance of Landsat-5 TM and MODIS for water quality studies in New York Harbor. The red reflectance correlated positively with turbidity for areas affected by river runoff. r^2 was 0.85 for Secchi depth using red band. r^2 was 0.78 for chlorophyll using green/red band (Hellweger et al. 2004).

(ii) In another study conducted for Monticchio lakes of southern Italy, the Secchi disk depth and chlorophyll were investigated using Landsat TM data. The statistical techniques were used to determine the relationship between TM data and water quality parameters. The visible bands (particularly blue and green band) and especially some of their ratios (blue/red, red/blue, red/green) were significantly correlated with transparency and chlorophyll concentration. The blue band showed the best relationship with Secchi Disk depth as suggested by the correlation analysis; r^2 was 0.82 for Secchi Disk and it was 0.72 for chlorophyll (Mancino et al. 2009).

(iii) The water quality of Ömerli Dam was assessed using blue, green, red and infrared bands of Landsat 7-ETM satellite data. The water quality parameters, which were analyzed, include chlorophyll, suspended solid matter, Secchi disk and total phosphate. The regression analysis has been used to develop empirical equations using the ETM satellite data and ground measured water quality data. r^2 values for suspended solid matter, Secchi disk and total phosphate were 0.9999, 0.9996 and 0.9906 (Alparslan et al. 2007).

(iv) The Landsat-5 TM images and in-situ observations were used from 15 stations on Lake Simcoe, Ontario, Canada for estimation of Secchi Disk Transparency. TM based Linear regression model has been developed using ratios of blue to red band and red band. The results are validated using in-situ data by linear regression and the accuracies are measured by the coefficient of determination (r^2) (Guan et al. 2011).

(v) The water quality in Reelfoot Lake, Tennessee, USA was investigated for Secchi disk depth, turbidity, chlorophyll, and total suspended solids. The relationship between green band and turbidity, chlorophyll, and total suspended solid was positive. The relationship was negative between red band and turbidity, chlorophyll, and Total suspended solid was negative. The values of r^2 were 0.705, 0.588, 0.537, and 0.522 for chlorophyll, Secchi disk depth, turbidity and total suspended solid respectively (Wang et al. 2006).

(vi) Two Landsat-7 ETM+ bands (blue and green) were used to study the contaminated waters of Medrano Creek, Argentina. Vignolo et al. (2006) developed a model that predicts the water quality index (WQI) of surface waters in the study area and uses linear regression analysis. The model has been validated using a data set of 12 physicochemical parameters obtained during the last 3 years. The physicochemical parameters used for this study were temperature, hardness, dissolved oxygen, pH, conductivity, alkalinity, turbidity, and the concentration of nitrate, nitrite and ammonia, chloride and sulfate. With r^2 values as 0.82, the best correlation of water quality index was found with blue and green bands (Vignolo et al. 2006).

(vii) The water quality mapping was accomplished for Secchi disk depth, turbidity, chlorophyll, and temperature using the ground measured data and Landsat-5 TM data. Blue and green bands were used for Secchi disk depth and their r^2 was 0.83. For turbidity green band was used and its r^2 was 0.52. Blue and green bands were used for chlorophyll and its r^2 was 0.84. The temperature has been obtained from green band and its r^2 was 0.55 (Khorram et al. 1991).

(viii) In a study, the chlorophyll-a concentration was obtained for Pearl River using MODIS land bands (band 1 and band 2). A model was established between the ratio of band 2 to band 1 and water sampling data for chlorophyll-a. The coefficient of determination (r^2) was 0.85 and the range of chlorophyll-a concentration was between 5 and 60mg-m^{-3} (Liu et al. 2010).

(ix) A study was conducted for Lake Erie, MODIS red/near-infra-red for estimation of suspended particulate matter. Remote sensing maps were generated for monthly mean distribution of surface concentrations of suspended particulate matter using MODIS water-leaving radiance at 748 nm for the period of five years (2003 to 2007) (Binding et al. 2010).

(x) The MODIS data was used for quantitative measurement and characterization for inland freshwater Lake Taihu, China. Seasonal, inter-annual variability and spatial distributions of lake water properties were analyzed from 2002 to 2008. Climatological water property maps, including normalized water-leaving radiance spectra, chlorophyll-a concentration, and water diffuse attenuation coefficient at the wavelength of 490 nm were derived from MODIS-aqua data. MODIS-Aqua–based water-leaving radiance at the blue band was constantly low in various regions, which indicate high algae concentration used (Wang et al. 2011).

(xi) The turbidity in Tampa Bay of Florida was estimated using MODIS Band 1 (620–670 nm) for estimation of turbidity in Tampa Bay, Florida. The coefficient of determination (r^2) was 0.76 between MODIS based surface reflectance and in situ turbidity after rainfall events (Moreno-Madrinan et al. 2010).

A literature survey has been done to compile different types of sensors in **Table 2.3**, which have been used for analyzing water quality parameters in inland waters. Most researchers used Landsat TM/ETM data for analyzing water quality parameters in inland waters, which clearly indicates the usefulness of this sensor for such types of studies. The use of environmental satellite sensors such as Landsat TM to assess water quality is one such technology as it offers wealth of remotely sensed data from the earth's surface at a resolution practical for the sensing of inland water bodies. Positive correlation of TM imagery to in situ measures of optically and thermally sensible water quality parameters (i.e. suspended sediments, chlorophyll and surface temperature) indicate satellite imagery as a potential data source for inland water quality monitoring (Cox et al. 1998). There are five most dominant water quality parameters, which were studied in various research. These are chlorophyll, turbidity, secchi disk depth, total suspended solids and temperature.

Table 2.3: Satellite measurements of water quality in inland waters

References	Locations	Sensors	Water Quality Parameters
(Wang et al. 2006)	Reelfoot Lake, Tennessee, USA	TM	chlorophyll-a, turbidity, Secchi disk depth, and total suspended solids
(Cox et al. 1998)	Catawba River, North Carolina, USA	TM	turbidity, secchi desk depth, chlorophyll, temperature
(Koponen et al. 2002)	Finland Lakes	AISA, MERIS	Secchi depth, turbidity, and chlorophyll a.
(Mancino et al. 2009)	Monticchio Lakes, Italy	TM	Secchi Disk depth and chlorophyll concentration
(Ostlund et al. 2001)	Lake Erken, Sweden	CASI, TM	chlorophyll and turbidity
(Alparslan et al. 2007)	Ömerli Dam, Istanbul City, Turkey	ETM	chlorophyll-a, suspended solid matter, secchi disk and total phosphate
(Vignolo et al. 2006)	Medrano Creek, Argentina	ETM	water quality index
(Lillesand et al. 1983, Kloiber et al. 2002)	Lakes in Minnesota, USA	TM, MSS	Secchi depth, chlorophyll, turbidity
(Verdin 1985)	Flaming Gorge Reservoir, Wyoming , Utah, USA	MSS	Secchi depth, chlorophyll
(Baruah et al. 2002)	Lake Kasumigaura, Japan	TM	total suspended solids, chlorophyll
(Schiebe et al. 1992)	Lake Chicot, Arkansas, USA	MSS	total suspended solids

(Giardino et al. 2001)	Lake Iseo, Lombardy, Italy	TM	total suspended solids, chlorophyll
(Fraser 1998)	Lakes in Nebraska, USA	TM	turbidity
(Dekker et al. 2001)	Frisian Lakes, Netherlands	TM, HRV	total suspended solids
(Brivio et al. 2001)	Lake Garda, Italy	TM	chlorophyll
(Mayo et al. 1995)	Lake Kinneret, Israel	TM	chlorophyll
(Barale et al. 2002)	Black Sea	CZCS, MOS	chlorophyll
(Lathrop 1992)	Lakes, Green,Bay, Lake Michigan, USA	TM	Secchi depth, total suspended solids, chlorophyll, turbidity

2.5 Geographic Information System (GIS)

Geographic information system (GIS) is used for collection, storage, and analysis of processes where geographic location is involved (Aronoff 1993). GIS was used to assess the relationship between land use and water quality (Osborne and Wiley 1988; Rhodes et al. 2001; Tufford et al. 2003). The spatial patterns of different places impact the relationship of land use with the surface water quality (Zampella et al. 2007). The water is acidic and nutrients are low for the water bodies surrounded by forests whereas pH and dissolved solids are higher for water around agricultural lands in the surroundings (Johnson and Watt 1996; Dow and Zampella 2000; Hunchak-Kariouk and Nicholson 2001). There was significant impact from agricultural lands on pH, specific conductance, and chloride (Zampella et al. 2007). Another study found that the increase in phosphorus and ammonia was due to sewage (Zampella 1994). The higher concentration of nitrates was due to urban land and agriculture (Smith et al. 1987). Calcium and magnesium were found as indicators of geological formation (Patrick 1996; Rhodes et al. 2001). A study showed that after obtaining temporal trends and sources of water pollution using multivariate statistical techniques, GIS was applied to develop yearly pollution index maps (Su et al. 2011). In another study, clusters were developed for different water pollution levels for the Nile Delta in Egypt using multivariate clustering technique and GIS (Shaban et al. 2010).The concentration of nonpoint source pollutants in surface water depends upon the spatio-temporal variations due to impact of land cover/use and precipitation (Wilson and Weng 2010). The climatic changes impact the quality of surface water adversely (Alvarez-Cobelas et al. 2005). It was found from the data for long-term ecosystem monitoring and research stations in North America that changes in climate (i.e. precipitation and temperature) affect the quality of surface waters significantly (Murdoch et al. 2000). GIS and remote sensing tools were used to develop simple predictive models that define relationships between watershed variables known to influence lake DOC concentrations and lake water color in the Absaroka-Beartooth Wilderness in Montana and Wyoming, USA. The resulting GIS model predicts DOC concentrations at the lake watershed scale with a high degree of accuracy ($r^2 = 0.92$; $P \leq 0.001$) by including two variables: vegetation coverage (representing sites of organic carbon fixation) and areas of low slope (0–5%) within the watershed (wetland sites of DOC production). Modeling with Advanced Land Imager satellite remote sensing data provided a weaker relationship with water color and DOC concentrations ($r^2 = 0.725$; $P \leq 0.001$). Model extrapolation is limited by small sample sizes but these models show promise in predicting lake DOC in subalpine and alpine regions (Winn et al. 2009).

CHAPTER 3

A METHODOLOGY FOR CLUSTERING LAKES IN ALBERTA ON THE BASIS OF WATER QUALITY PARAMETERS

Akbar T. A., Hassan Q. K., Achari G., A Methodology for Clustering Lakes in Alberta on the Basis of Water Quality Parameters. *CLEAN-Soil, Air, Water* **2011** Volume 39, Issue 10, pp. 916–924.

Abstract

In this study, a methodology for clustering 18 lakes in Alberta, Canada using the data of 19 water quality parameters for a period of 11 years (1988–2002) is presented. The methods consist of (i) principal component analysis (PCA) to determine the dominant water quality parameters, (ii) cluster analysis techniques to develop the characteristics of the clusters, and (iii) pattern-match lakes to determine the appropriate cluster for each of the lakes. The PCA revealed that three principal components (PCs) were able to explain ~88% of the variability and the dominant water quality parameters were total dissolved solids, total phosphorus, and chlorophyll-a. We obtained five clusters for the period 1994–1997 by using the dominant parameters with water quality deteriorating as the cluster number increased from 1 to 5. Upon matching cluster patterns with the entire dataset, it was observed that some of the lakes belonged to the same cluster all the time (e.g., cluster 1 for lakes Elkwater, Gregg, and Jarvis, cluster 3 for Sturgeon, cluster 4 for Moonshine, and cluster 5 for Saskatoon), while others changed with time. This methodology could be applied in other regions of the world to identify the most suitable source waters and prioritize their management. It could be helpful to analyze the natural controlling processes, pollution types, impact of seasonal changes and overall quality of source waters. This methodology could be used for monitoring water bodies in a cost effective and efficient way by sampling only less number of dominant parameters instead of using a large set of parameters.

Keywords: Chlorophyll-a; K-means clustering; Principal component analysis; Total dissolved solids; Total phosphorus

3.1 Introduction

In Alberta, Canada, approximately 97.5% of consumptive use of water comes from surface water, which serves numerous communities big and small (www.water.ca/sat-82308.asp) (The Water Chronicles, 2008). With an increase in population, there is an increase in pressure on water resources. A good understanding of the quality of different source waters is essential for effective treatment prior to distribution. The water quality may potentially be affected by several natural processes such as precipitation, weathering, soil types, and watershed geology as well as anthropogenic activities such as agricultural runoff, municipal sewage, and industrial waste (Reghunath et al. 2002; Akbar and Lin 2010; Koç 2010; Tokalıoğlu et al. 2010). Since these processes impact the quality of water, each water body can be considered unique. If a particular water body is to serve as source for a drinking water system then it has to be treated to meet the Canadian drinking water guidelines. While large municipalities have the financial resources to remove all kinds of pollutants from source waters, most small water systems do not. These financially constrained water systems have to use technologies that are economical, effective and targeted to pollutants of concern in its source waters. In this paper, a methodology is presented to cluster water bodies so that these clusters could be used to identify technologies for effective treatment. At this stage, we intend to use water quality data from 18 lakes in Alberta to demonstrate the methodology and develop the clusters.

In determining the water quality, water samples are collected and analyzed in a laboratory to determine the physical, chemical, and biological properties. These water quality-related parameters are analyzed further using statistical methods in order to develop clusters having similar characteristics (Barreto et al. 2008; Shrestha and Kazama 2007). The most commonly employed statistical methods are the use of principal component analysis (PCA) and cluster analysis techniques (e.g., K-means, K-medoid, fuzzy, etc.) (Cunjie et al. 2010; Panda et al. 2006; Ragno et al. 2007; Westra et al. 2007) These techniques have also been proven to be effective techniques to analyze large and complicated data with numerous parameters and different units (Ragno et al. 2007). As a variation to statistical methods, fuzzy logic may also be used to investigate the uncertainties (Garg et al. 2007; Senevirathna et al. 2011).

PCA is an effective way for separating variables in subgroups (Tokalıoğlu et al. 2010). In PCA correlations among various parameters are investigated and all correlated parameters are combined into different principal components (PCs). Each PC comprises of groups of correlated parameters with positive and negative loading values, which are used for the interpretation of major processes involved in analyzing and characterizing the surface water quality.

In cluster analysis, the objects (e.g., a set of water quality parameters) are grouped on the basis of similarities within a class and dissimilarities among different classes (Panda et al. 2006). This technique has been used to evaluate spatio-temporal variations of surface water quality-related studies (Ragno et al. 2007; Kambe et al. 2007; Singh et al. 2004; Sundaray et al. 2004). A distance measure is used to determine the similarity and/or dissimilarity among objects of interest. The two most commonly used distance measures are Euclidean and Manhattan (Kaufman and Rousseeuw 1990).

The advantages of multivariate statistical methods include: (i) reduction in complexity of large-scale dataset (Bengraïne and Marhaba 2003), (ii) unbiasedness in methods which help in natural association between samples and parameters and it reveals the information which cannot be observed from the dataset at first glance (Wenning and Erickson 1994), (iii) related parameters can be identified by reducing and organizing large dataset into groups with similar characteristics (Jayakumar and Siraz 1997**)**, and (iv) usefulness and efficiency in water quality studies (Areerachakul and Sanguansintukul 2010).

The disadvantages of multivariate statistical methods include (i) subjectivity in terms of interpretation for controlling the sources and processes (Liu et al. 2003), (ii) unreliability in water quality data which might not give appropriate results, (iii) existence of same parameters in different PCs which might change the interpretations, and (iv) difficulty in determination of suitable number of clusters.

Previous research work accomplished for analyzing water quality using multivariate statistical methods is discussed in the subsequent paragraphs. Prikler et al. (2003) applied cluster analysis, discriminant analysis, and support vector machines for analyzing water bodies in the Polish Tatra Mountains. The results from cluster analysis suggested reconsideration for the geographical distinction. Discriminant analysis also confirmed the geographical separation of water bodies and it also revealed that the sampling time is a crucial factor in environmental analysis. Dubiella-Jackowska et al. (2010) used multivariate statistical techniques for interpretation of atmospheric and surface water pollution. The hierarchical cluster analysis was used for major inorganic ions, electrolytic conductivity, and pH. The potential sources of pollution were anthropogenic activities like fertilizers usage and transport, road salting in winter and semi-natural like sea salt aerosols, and erosion of construction materials. Zhao et al. (2010) used statistical methods to analyze spatial and temporal patterns of phytoplankton in Namuka Co saline Lake of Tibet, China. The investigation was performed on a monthly water quality data collected between June 2001 and July 2002. It was found that total phytoplankton was lower in winter and higher in spring and summer. A negative correlation was found for phytoplankton with salinity. Jose Barreto et al. (2008) analyzed 17 parameters for a hydroelectric reservoir in South Brazil. Multivariate PCA and hierarchical group analysis (HGA) were used to identify the major parameters for differentiating between origin of source water for Tibagi and the Primeiro de Maio River. The major discriminating parameters were the absorbance relation, Fe(III), Mn(III), and Ni(II). The potential anthropogenic source of pollution due to Ni(II) and orthophosphate were town sewage discharge.

Su et al. (2011) obtained temporal trends and sources of water pollution in functional zones of Qiantang River, China. The water quality monitoring data for 13 parameters from 41 monitoring sites for a period of 9 years (1996–2004) were used. The discriminant analysis was used to identify the four significant parameters. A geographic information system (GIS) based yearly pollution index was used to develop maps for all monitoring sites. PCA was used to obtain information on potential pollution sources. Huang et al. (2010) applied statistical techniques to obtain spatial variation and source water pollution for Qiantang River of China. In this study 13 water quality variables were analyzed for 46 monitoring sites. Three pollution zones (low, moderate, and high) based on national quality standards were developed using fuzzy comprehensive analysis. Factor analysis was used to identify two pollution sources that explained 67% of total variance in low pollution zone, two pollution sources that explained 73% of the total variance in moderate pollution source, and three pollution sources that explained 80% of total variance in high pollution source. The potential sources of pollution for the most water quality variables were industrial waste-water, agricultural activities, and urban runoff. The statistical-based approach supported the idea for developing better pollution control strategies. Astel et al. (2006) used chemometrics (multivariate techniques) in monitoring spatial and temporal variations in drinking water quality by interpretation of monitoring data for chloro/bromo disinfection by-products in drinking water at 12 locations in Gdańsk area of Poland for a period of 8 years (1993–2000). Cluster analysis showed two different groups of sources of drinking water for sampling locations. Analysis of variance (ANOVA) was applied to classify and confirm the cluster groups and it was also used to prove the existence of difference between the concentrations of $CHCl_3$, $CHBrCl_2$+C_2HCl_3, $CHBr_2Cl$, and CH_2Cl_2. The temporal changes showed improvement in drinking water quality. The statistical methods proved powerful tools to understand spatial and temporal variations in water quality. Singh et al. (2004) used water quality data of 24 parameters for a period of 5 years (1994–1998) at three sampling sites for Gomti River of India. PCA identified six major factors, which explained 71% of total variance. Cluster analysis defined three groups of similarity among sampling sites. Discriminant analysis indicated major parameters for large variations in temporal and spatial analysis.

Shaban et al. (2010) detected and mapped variation of water pollution in the Nile Delta of Egypt using multivariate clustering and GIS techniques. The pollution levels of different drainage system were categorized using clustering method and these clusters were visualized in GIS. Finally, a GIS-based decision support system was developed by producing thematic maps for entire Nile Delta. Razmkhah et al. (2010) evaluated spatial and temporal variation in Jajrood River of Iran using pattern recognition techniques. The monthly water quality monitoring data for 18 water quality variables were analyzed for a period of 3 years. PCA identified five varifactors, which showed 85% of both temporal and spatial changes. Cluster analysis was used to classify water quality of monitoring stations and results indicated that Out-Meygoon was the most polluted one. Ahar, Baghgol, Rooteh, before Zaygan, Fasham, Roodak, and Lashgarak were found affected by organic pollution. Yidana et al. (2008) applied hierarchical cluster and PCA to surface water hydrochemical data for three locations in Ghana. PCA reduced 30, 33, and 33 data points, respectively, for Ankwaso, Dominase, and Prestea to four, three, and four PCs. It was found that hydrochemistry of basin was controlled by weathering of minerals like silicates, carbonates, gypsum and apatite, and the decay of organic matter due to forest region.

Lumb et al. (2006) applied a water quality index (WQI) developed by Canadian Council of Ministers for the Environment (CCME) and statistical approach to monitor water quality for Mackenzie River basin of Canada. CCME WQI consists of three elements which are scope, frequency and amplitude. The scope indicates the number of water quality parameters not meeting water quality guidelines. The frequency shows the number of times the water guidelines are not met and amplitude identifies the extent to which the guidelines are not met. CCME WQI shows water quality range between 0 (worst) to 100 (best). The result indicated that river is impacted by high turbidity and total trace element due to high suspended sediment loads. Rosemond et al. (2009) found CCME WQI as a good tool to assess absolute water quality guidelines for the protection of aquatic life but it had limitation for evaluating the spatial changes in water quality downstream of point source discharges. Usunoff and Guzman (1989) used statistical techniques called factor analysis and correspondence analysis and it was found that three factors accounted for the ion variations in the Milk River, Alberta Canada. St-Hilaire et al. (2004) used data of 6 parameters for 36 stations in Richibucto River drainage basin of New Brunswick province of Canada for a period of 6 years (1996–2001). PCA was applied on the data for a period of 6 years to analyze water quality variance in fresh and brackish water. The results showed that the most of variance in fresh water was due to pH, total organic carbon, and high nutrient concentrations whereas salinity with high nutrient concentrations explained the variance in brackish water. Cluster analysis showed the importance of high phosphorous and nitrate concentrations in water bodies receiving treated municipal effluent.

In this paper, our objectives are to (i) determine the dominant water quality parameters impacting Alberta lakes, (ii) identify the optimum number of clusters and their characteristics, and (iii) pattern-match the lakes and investigate their temporal dynamics.

3.2 Materials and methods

3.2.1 Study area and data used

In this study, data from 18 lakes in Alberta were considered. The relative position of Alberta, in the Canadian context and geographical location of the lakes are shown in **Figure 3.1**. The lakes of interest are: Beauvais, Cardinal, Crimson, Dillberry, Elkwater, Gregg, Gregoire, Jarvis, Long, Mcleod, Miquelon, Moonshine, Reesor, Saskatoon, Spruce Coulee, Steele, Sturgeon, and Winagami. The total surface area of these lakes is 216.08 km^2 and their catchment area is 2174.45 km^2. The surface area of the lakes ranges from 0.21 to 52 km^2 with an average of 3 km^2. The catchment areas of the lakes vary between 1.75 and 571 km^2 with an average of 25 km^2. The mean depth of these lakes varies between 1.3 and 25 m.

The water quality data for the lakes were obtained from Alberta Environment, the regulatory body responsible for collecting and storing the data. The data included mean annual values for 19 water quality parameters including alkalinity (ALK), bicarbonate (HCO_3), calcium (Ca), carbonate (CO_3), chloride (Cl), chlorophyll-a (CHL-a), fluoride (F), hardness (HARD), iron (Fe), magnesium (Mg), pH, total phosphorus (TP), potassium (K), Secchi depth (SD), silica (SiO_2), sodium (Na), specific conductivity (SC), sulfate (SO_4), and total dissolved solids (TDS). In this study, data of water quality parameters for 11 years (i.e., 1988–1989, 1992–1999, and 2002) are used. These years had the maximum amount of required data.

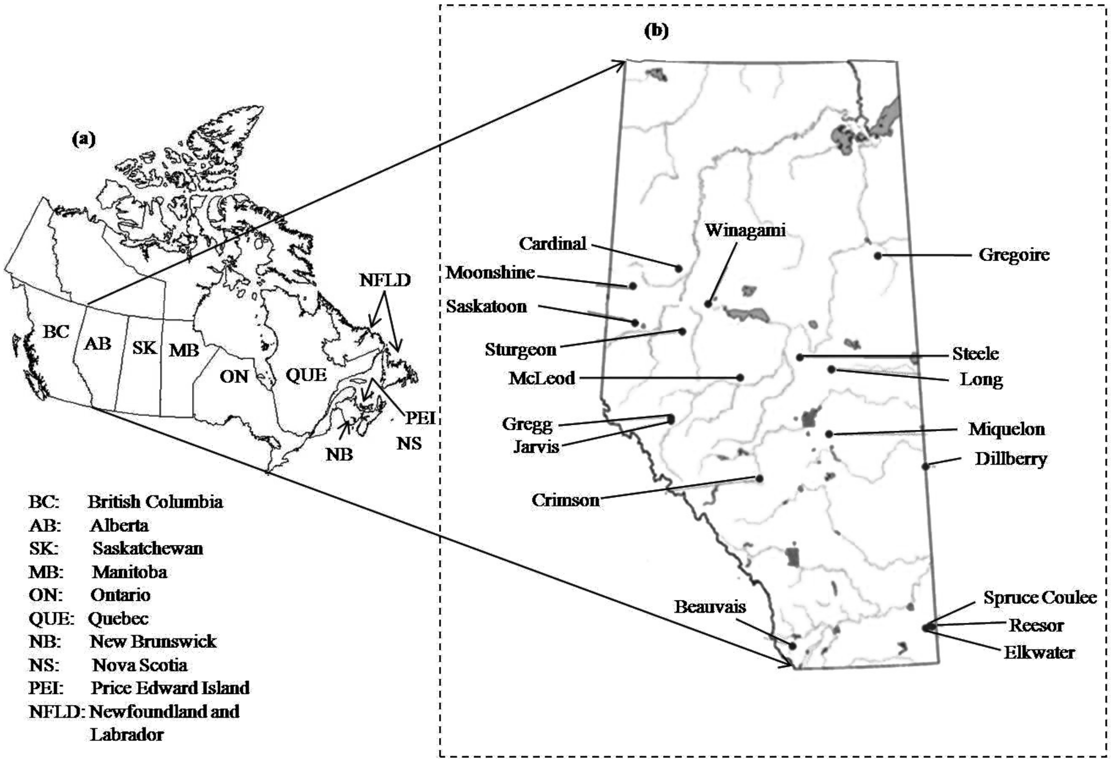

Figure 3.1: a) Location of Alberta in Canada and (b) Location of the 18 lakes in Alberta

3.2.2 Methods

Figure 3.2 shows a schematic diagram of the methodology used in this study. The major components of the methodology are (i) the use of PCA on all eighteen lakes with nineteen water quality-related parameters for the period 1994–1997 to determine the dominant parameters, (ii) the use of cluster analysis to determine the characteristics of the clusters using data from 1994 to 1997, and (iii) pattern-match lakes to allocate clusters and investigate their temporal dynamics. A brief description of the three major components is provided in the subsequent paragraphs. PCA is used to obtain major PCs using an eigenvalue of 1 as a cutoff (Cunjie et al. 2010; Panda et al. 2006). The loading values for all the parameters under major PCs were obtained using varimax normalized rotation (Shrestha and Kazama 2007; Panda et al. 2006).

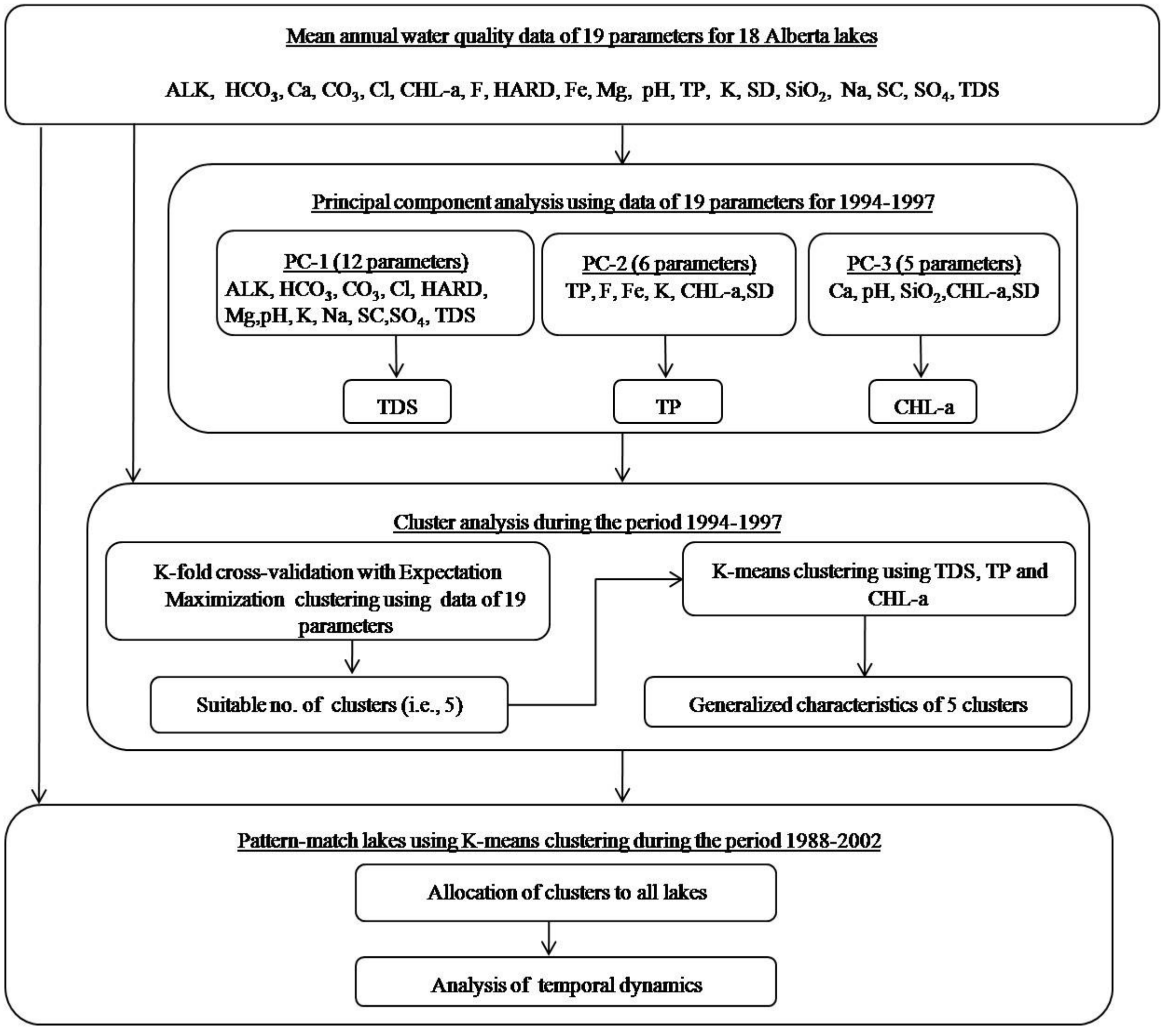

Figure 3.2: Schematic diagram for PCA and cluster analysis.

The cluster analysis is conducted to determine (i) the suitable number of clusters, (ii) identify the characteristics of clusters using the dominant parameters, and (iii) pattern-match the lakes to allocate clusters and analyze the temporal dynamics. K-fold cross-validation with expectation maximization (EM) clustering algorithm (Boyce et al. 2002; Dempster et al. 1977; Fielding et al. 1997; McLachlan et al. 2004) on raw water quality data of 4 years (1994-1997) by considering all of the 19 parameters for the period 1994–1997 to decide the total number of suitable clusters, is conducted. In K-fold cross-validation, the original data are randomly divided into K sub-samples. In this process, one sub-sample is retained for validation and the remaining (K-1) sub-samples are used as training data set. The cross-validation is conducted K times (folds) with each of K sub-samples being used once for validation. Finally a single value obtained by averaging or combining the K results is determined. In this study, 10-fold cross-validation is used as it is the most preferred method in literature (McLachlan et al. 2004). The EM algorithm is a statistical method, which uses the means and standard deviations of each cluster for obtaining the maximum likelihood estimates of parameters (Dempster et al. 1977). We considered all 19 parameters because it allowed for optimizing the number of clusters between 3 and 11. On the contrary, considering the dominant three parameters we are only able to select clusters between 2 and 3, which may not be suitable representation of clusters for the whole data. This step revealed that the optimum number of clusters should be 5 (see **Figure 3.4**, discussed later).

Prior to employing the dominant parameters in determining the cluster characteristics, those are normalized over the entire dataset for the period 1988–2002. Normalization is done by dividing the raw values of individual parameter by its respective maximum value (Panda et al. 2006; Liu et al. 2003). The maximum values of TDS, TP and CHL-a are 6371mg/L, 1.07 mg/L, and 173.55 mg/L, respectively. Note that the highest value of TDS (i.e., 6371 mg/L) is found in Miquelon Lake, which indicated that it is saline. As salinity is uncommon in inland lakes, the second highest value of TDS in the dataset (757mg/L) is used and the normalized value of TDS for Miquelon Lake is set to 1. On the normalized dataset for the period 1994–1997, K-means clustering technique is applied to generate five clusters for each of the year individually. In K-means algorithm, a cluster centroid is obtained randomly according to initial value. The clusters of objects are assigned based on the distance between the mean value of the object and centroid of the cluster. The distance is a measure of either Euclidean or L1 distance (Kanungo et al. 2002). K-means clustering algorithm showed five different patterns of clusters using the mean values of TDS, TP, and CHL-a for each of the years individually. As, the K-means algorithm was applied on the normalized data from individual years, thus it was possible that a particular cluster might have different patterns over the period of interest. In order to address this issue, we aggregated the clusters of similar patterns for the period 1994–1997 by averaging the mean values of parameters; and generated the generalized patterns for the characteristics of all five clusters (see **Figure 3.5**, discussed later).

The generalized cluster characteristics are applied over the entire normalized dataset for the period 1988–2002. For a particular lake during a year of interest, the sum of squared error (SSE) with respective to the each of the five generalized patterns, is computed using Eq. (3.1):

$$SSE = \sum_{C_i} \sum_{x \in C_i} \left\| x - m_i \right\|^2 \tag{3.1}$$

where m_i is the mean of the i^{th} cluster and $x \varepsilon Ci$ is a pattern assigned to that cluster. A particular lake is assigned to a cluster, where the SSE is observed to be the minimum.

3.3 Results and discussion

3.3.1 Principal component analysis

Using the data from the period 1994–1997, the PCA produced a set of PCs and their respective eigenvalues (see **Figure 3.3**). The first three PCs (i.e., PC-1, PC-2, and PC-3) have eigenvalues >1 and are considered as major PCs like other studies (e.g., Tokalıoğlu et al. 2010; Cunjie et al. 2010). These three PCs captured approximately 88% of the variability in the dataset (see Tab. 1). PC-1, PC-2, and PC-3 accounted from 61.97 to 62.73%, 18.17 to 18.40%, and 6.06 to 9.07% of the total variance, respectively. **Table 3.1** also reveals the corresponding loading values for each of the major three PCs. The loading values are categorized into three classes (i.e., strong > 0.75, 0.75 > moderate > 0.5, and 0.5 > weak > 0.4) (Panda et al. 2006; Liu et al. 2003) with parameter loading values less than 0.40 not being considered because of their less significance.

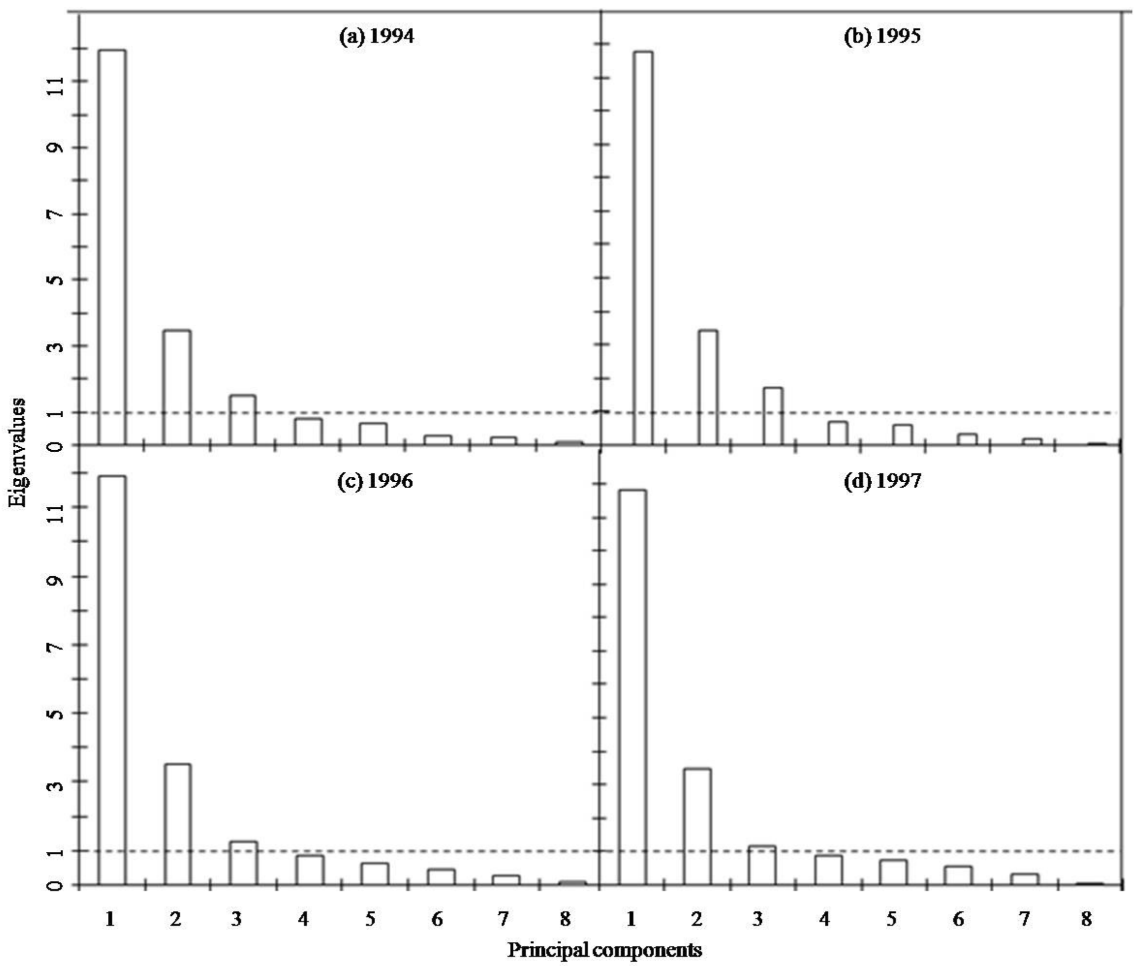

Figure 3.3: Eigenvalues to obtain major PCs in water quality data for the period of 1994–1997.

Table 3.1: PCs with loading values for 19 water quality parameters in 4 years (1994–1997).

Year	1994			1995			1996			1997		
Parameter	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3	PC-1	PC-2	PC-3
ALK	0.925			0.949			0.961			0.989		
HCO_3	0.912			0.908			0.916			0.952		
Ca			-0.614			-0.694	-0.419		-0.661	-0.498		0.583
CO_3	0.957			0.975			0.986			0.995		
Cl	0.968			0.975			0.980			0.897		
CHL-a		0.401	-0.763			-0.874			-0.883		0.646	
F		0.834			0.845			0.863		0.439	0.725	
HARD	0.958			0.974			0.962			0.953		
Fe		0.940			0.948			0.917			0.926	
Mg	0.966			0.983			0.976			0.988		
pH	0.581		0.600	0.675		0.398	0.692			0.700		0.432
TP		0.980			0.972			0.969			0.954	
K	0.700	0.655		0.780	0.586		0.779	0.606		0.874	0.461	
SD		-0.660	0.487		-0.660	0.484		-0.597	0.472		-0.693	0.453
SiO_2			-0.663			-0.660	-0.431		0.430	-0.441		
Na	0.978			0.982			0.989			0.993		
SC	0.980			0.988			0.992			0.994		
SO_4	0.984			0.986			0.992			0.980		
TDS	0.981			0.987			0.993			0.993		
Variance (%)	62.68	18.30	7.76	61.97	18.19	9.07	62.73	18.40	6.66	62.14	18.17	6.06
Cumu.(%)	62.68	80.98	88.74	61.97	80.16	89.23	62.73	81.13	87.79	62.14	80.31	86.37

PC-1 revealed that 12 parameters (i.e., ALK, HCO_3, CO_3, Cl, HARD, Mg, pH, K, Na, SC, SO_4, and TDS) are correlated with each other. Apart from pH and K in 1994, all other parameters are found to be loaded as both strong (i.e., >0.75) with positive values. During the years of 1996–1997, it is also observed that the parameters of Ca and SiO_2 are included in PC-1; however, both are loaded as weak (i.e., <0.50) with negative values. The PC-1 may be interpreted as alkaline (due to presence of ALK, HCO_3, and CO_3), hard (due to the presence of HARD, and Mg), and saline (due to the presence of Cl, Na, SC, SO_4, and TDS) water. Note that the factors that might affect HARD, also affect ALK. In Alberta, most of the watersheds are enriched with carbonates, thus the lakes are basically alkaline (Alberta Environment 2006). In addition, the bedrock of Alberta is sedimentary (such as limestone and dolomite), thus the underlying basin of the lakes consists of high concentration of TDS (Alberta Environment 2006). In particular, gradual increment of pH (i.e., 0.581 in 1994 to 0.70 in 1997) may be associated with the changes in water levels and the effect of anthropogenic activities such as agricultural, mining, and industrial. Another multivariate-based water quality study showed that the main anthropogenic activities that affected the lake water were domestic and industrial effluents (Barreto et al. 2008).

The second major component of PC-2 revealed that six parameters (i.e., F, Fe, TP, K, SD, and CHL-a) are correlated with each other (see Tab. 1 with the columns for PC-2). The parameters of Fe, TP, and F are loaded strong (i.e., >0.75) with positive values with exception of F in 1997 (i.e., 0.725). The K is found to be both in PC-1 and PC-2, however, it is more significant in PC-1 (i.e., loaded strong with positive values for most years except 1994).

The SD is found to be loaded moderate (i.e., in the range of 0.50 – 0.75) with negative values. It is worthwhile mentioning that SD is also found to be loaded weak (i.e., <0.50) with positive values in PC-3. In addition, it is also found that CHL-a is loaded weak (i.e., 0.401 in 1994), and moderate (i.e., 0.646 in 1997) with positive values. F can be present due to geochemical weathering of rocks and soils and also from municipal wastewater because of fluoridation of drinking water (Alberta Environment 2006). The Fe can be due to weathering of sulfide-rich ores, rocks, and leaching of sandstones and the other potential sources including acid mine drainage, coal burning, processing of minerals, sewage, and landfill leachates (Alberta Environment 2006). The TP can be leaching from nutrient-enriched soils and anthropogenic activities. Jose Barreto et al. (2008) showed phosphorus as indicator of anthropogenic activities for Capivara Hydroelectric Dam Lake in Brazil using PCA technique. The SD is an indicator for water clarity and algal biomass production (Alberta Environment 2006). In general, the PC-2 can be considered as an indicator of biological activities in the lakes.

The third major component, PC-3, comprises of five parameters (i.e., Ca, CHL-a, pH, SD, and SiO_2) that are correlated with each other (see Tab. 1 with the columns for PC-3). Among the parameters, only CHL-a is found to be loaded strong (i.e., >0.75) with negative values. In most cases, both Ca and SiO_2 are loaded moderate (i.e., in the range 0.50–0.75) with negative values while both pH and SD are loaded weak (i.e., <0.50) with positive values. The presence of CHL-a can be enhanced by higher amounts of TP, thus PC-3 can also be considered as an indicator of biological activities like PC-2. Baborowski et al. (2011) observed that concentration of CHL-a is directly related to the development of phytoplankton using factor analysis technique. The potential source of Ca is the sedimentary bedrock. The SiO_2 may be from industrial effluents.

3.3.2 Dominant water quality parameters

The dominant parameters identified by the PCA are: TDS, TP and CHL-a (see Tab. 1). The previous discussion indicated that different major ions (i.e., Mg, Na, K, HCO_3, CO_3, SO4, and Cl) were loaded with positive values, and they have strong effects on PC-1. These ions account for most of the TDS present in surface water quality (Alberta Environment 2006). TDS is also loaded strong in PC-1 and has the highest positive values. Thus, TDS is considered as a dominant parameter. The TP is considered as the next dominant water quality parameter as it is loaded strong in PC-2 with the highest positive values. The CHL-a is considered as the third dominant water quality parameter as it is loaded strong in PC-3 with highest negative values. These two parameters (TP and CHL-a) may have a strong relationship with each other indicating the effect of anthropogenic activities on water quality (Alberta Environment, 2006). TP and CHL-a concentrations along with SD are also used as indicators of trophic states in the lakes by the Alberta Environment (Alberta Environment, 2006). Liu et al. (2010) found that TP had the greatest positive influence on

CHL-a, whereas SD had the negative influence.

3.3.3 Cluster analysis

Figure 3.4 shows the relation between number of clusters and the cost function (-2*log-likelihood) of the K-fold cross-validation with EM clustering algorithm for the period 1994–1997. It shows that the magnitude of the cost function reached a minimum-value when the number of clusters was 5 in most of the instances, except for 1997 (i.e., number of clusters was 4 if the minimum value cost function is considered). On the basis this we considered five as the suitable number of clusters in this study. **Figure 3.5** shows the generalized characteristics of the five clusters obtained for the period 1994–1997. It is obvious that TDS, TP, and CHL-a increased from cluster 1 to cluster 5. From this, we interpret that the water quality for all lakes in Alberta decreases from cluster 1 to cluster 5. Upon implementing these generalized characteristics as shown in **Figure 3.5** over the entire dataset, the temporal change in cluster number for each lake is investigated (see **Table 3.2**). The temporal variation was analyzed in surface water quality studies using multivariate statistical techniques (Barreto et al. 2008; Dubiella-Jackowska et al. 2010; Liu et al. 2010).

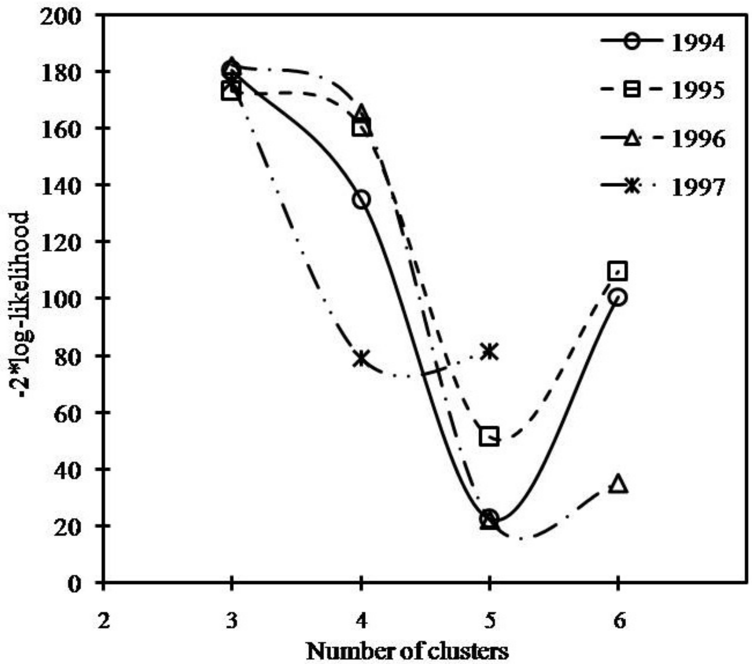

Figure 3.4: Determining the suitable number of clusters using K-fold cross-validation with algorithm of EM clustering in water quality data for the period of 1994–1997.

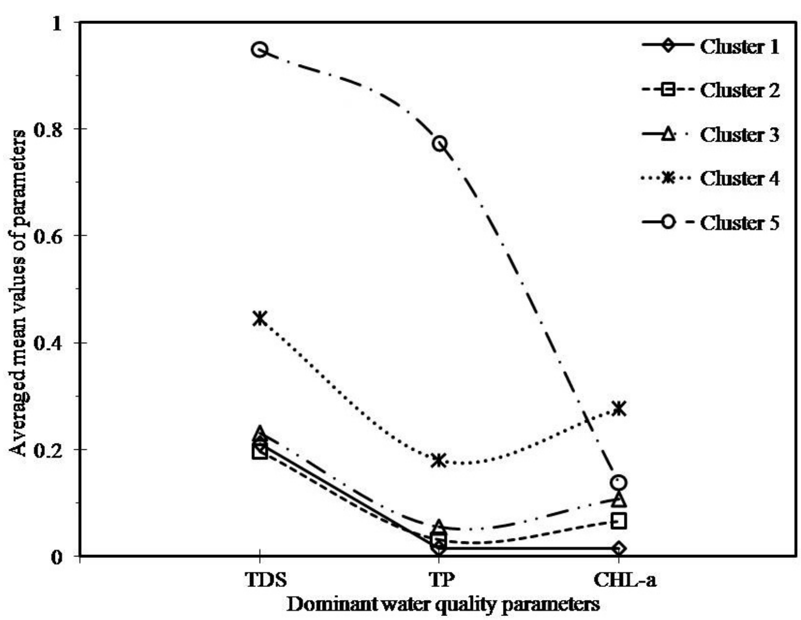

Figure 3.5: Generalized characteristics for the five clusters produced from the data for the period 1994–1997.

Table 3.2: Temporal changes of lakes (1988–2002).

Lakes	1988	1989	1992	1993	1994	1995	1996	1997	1998	1999	2002
Beauvais	3	1	1	1	1	1	1	1	ND*	1	ND
Crimson	1	1	3	1	1	1	1	1	2	ND	1
Dillberry	1	ND	1	1	1	1	1	1	1	ND	1
Elkwater	1	1	1	1	1	1	1	1	1	1	1
Gregg	1	1	1	1	1	1	1	1	1	1	1
Jarvis	1	1	1	1	1	1	1	1	1	1	1
Gregoire	ND	2	2	2	2	2	2	2	2	2	2
Mcleod	1	1	2	2	2	2	2	2	2	2	1
Reesor	2	1	1	2	2	2	3	2	2	1	1
Spruce Coulee	2	2	1	1	2	1	2	2	2	1	2
Long	3	3	3	3	3	3	3	3	3	1	3
Steele	3	3	3	3	3	3	3	3	3	4	4
Sturgeon	3	3	3	3	3	3	3	3	3	3	3
Cardinal	ND	ND	4	4	4	4	4	4	ND	ND	ND
Moonshine	4	4	4	4	4	4	4	4	4	4	4
Winagami	4	4	4	3	4	4	3	1	4	4	1
Miquelon	ND	ND	5	ND	5	5	4	4	4	ND	ND
Saskatoon	5	5	5	5	5	5	5	5	5	5	5

It is found that Beauvais, Crimson, Dillberry, Elkwater, Gregg, and Jarvis Lakes are in cluster 1 (i.e., relatively good water quality). There are no significant changes in clusters for most of these lakes, which indicate the stability in lake water quality. Beauvais Lake was in cluster 3 in 1988 whereas Crimson Lake was in cluster 3 and cluster 2 in 1992 and 1998, respectively. The slight changes in clusters of these two lakes are due to variations in water levels (Alberta Environment 2006).

It is found that cluster 2 is the most dominant cluster of Gregoire, Mcleod, Reesor, and Spruce Coulee Lakes. There are some changes of clusters like Mcleod Lake was in cluster 1 in 1988–1989 and 2002, Reesor Lake was in cluster 1 in 1989, 1992, 1999, and 2002, Reesor Lake was in cluster 3 in 1996, Spruce Coulee Lake was in cluster 1 in 1992–1993, 1995, and 1999. The changes to cluster for these five lakes in all years except 1996 were from cluster 2 to cluster 1, which indicated improved water quality. This improvement is probably due to increased water levels in lakes due to snowmelt and precipitation (Alberta

Environment 2006). Another study also showed the effect of seasonal fluctuations on the river water quality (Baborowski et al. 2011).

Long, Steele, and Sturgeon Lakes belonged to cluster 3. The recreational activities around Long Lake impact its water quality. Activities such as camping, residential developments and cropping have negative impacts on water quality of Sturgeon Lake. There is a decrease in water quality of Steele Lake in 1999 and 2002 as it moved to cluster 4. Coincidently, in both these years there is an abrupt decrease in water levels because of drought (Alberta Environment 2006). Koç (2010) observed an increase in polluted concentration in surface water quality due to decrease in water flow rate because of drought.

Cardinal, Moonshine, and Winagami Lakes are found in cluster 4 predominantly. Cardinal Lake is a shallow lake located in low topographical relief surrounded by mixed forest (Alberta Environment 2006), due to which there is a possibility of higher negative impact of watershed soils and geology on Cardinal Lake water quality as compared to other lakes. Moonshine Lake is a shallow reservoir with high recreational activities and its drainage basin is a mixture of forests and wetlands (http://environment.gov.ab.ca/info/home.asp) (Alberta Environment 2006). Wood logging and agricultural activities in the surrounding areas of Winagami Lake may have provided negative impact on its water quality. The water quality of Winagami Lake is improved as it moved from cluster 3 in 1993 and 1996 to cluster 1 in 1997 and 2002. It also supplies water for residential use to small communities due to which several canals and control structures are built to increase water level and flushing rate. This may have led to an enhancement in its water quality. The clusters for these lakes could be interpreted as impacts of anthropogenic and natural processes as also observed in another study (Dubiella-Jackowska et al. 2010).

Miquelon Lake is in cluster 5 and cluster 4 consistently. It is a shallow, highly saline lake surrounded by forest parkland and agriculture land, which is used for cereal crops and livestock operation (Alberta Environment 2006). The Saskatoon Lake drainage basin has agricultural land well known for top quality berry crop (Alberta Environment 2006), which probably has negatively impacted its quality. Koç (2010) found the negative impact of agricultural activities on surface water quality.

3.4 Concluding remarks

In this study, a methodology for clustering eighteen lakes in Alberta, Canada using PCA and clustering techniques is presented. With PCA, three PCs were identified. The three most dominant parameters, which were obtained from the PCs, are TDS, TP, and CHL-a. K-fold cross-validation with EM clustering indicated five as the most suitable number of clusters. K-means clustering technique is used on the normalized data of the dominant parameters to obtain the generalized characteristics of five clusters. The water quality deteriorated as the cluster number increased from 1 to 5. Pattern-match using K-means clustering technique was done to allocate clusters to all lakes for the period 1988–2002. From the results, it is found that clusters remained same for nine lakes (i.e., Dillbery, Elkwater, Gregg, Jarvis, Gregoire, Sturgeon, Cardinal, Moonshine, and Saskatoon), which indicate stability in water quality whereas the remaining nine lakes are found to change the clusters over time.

This methodology is useful for (i) monitoring and sampling the source waters using the less number of dominant parameters, (ii) analyzing the impact of natural processes and anthropogenic activities on water bodies, (iii) analyzing the temporal dynamics to observe the changes in water quality over a specified period of time, (iv) identifying the specific pollutants in source waters for designing economical, targeted and effective drinking water treatment facilities for smaller communities, (v) providing good understanding of water bodies which can help in the management of source waters.

This methodology might be applied to understand quality of source waters (i.e., lakes, rivers, and other water bodies) that supply drinking water to big and small communities in any region of the world. PCA could be applied on the monthly, seasonal or yearly water quality sampling data to identify major PCs and extract dominant parameters. The natural controlling processes and pollution types which impact the quality of source waters could be defined and explained by interpretation of correlated parameters combined under PCs of PCA. K-means algorithm could be applied to produce the generalized characteristics using the dominant parameters for developing monthly, seasonal, or yearly clusters for source water quality. The allocation of clusters to source waters might be helpful to understand the effect of natural processes, pollution types, and seasonal changes on the water quality of source waters. The results could indicate the treatment required for the specific parameters at various sampling locations before supplying water to communities. On the basis of cluster results, sampling strategies could be revised to focus on the monitoring of dominant parameters, which could make water quality sampling economical and targeted. This methodology might also help to determine the required frequency of monitoring sampling sites in different periods of a year. This methodology might identify most suitable source waters that would require minimum level of water treatment. This strategy could make water treatment economical and targeted for the communities, which are constrained with limited funds. On the basis of clustering results, the management of water bodies could be prioritized. Koç (2010) suggested the implementation of appropriate management principles for all point and non-point pollution sources to enhance the surface water quality. The existing methodology might be enhanced by identifying the point and non-point source pollution for the source waters, which are in cluster 3, 4, or 5. The non-point source pollution for clustered water bodies could be identified using GIS by considering the watershed characteristics like land cover, land use, soil, surface geology, drainage area, and topography. The information on point source pollution for these clustered water bodies could be identified by considering GIS layers of industrial discharge and sewage discharge. GIS was applied with PCA and clusters to identify source water pollution and visualize results in the form of maps (e.g., Su et al. 2011; Razmkhah et al. 2010). Remote sensing (RS) techniques could be used for mapping the dominant parameters for the whole water bodies. Finally, a decision support system could be developed using multivariate, GIS, and RS techniques that help decision makers to develop economical, feasible, and targeted water treatment systems for the treatment of source waters.

CHAPTER 4

CLUSTERIZATION OF SURFACE WATER QUALITY AND ITS RELATION TO CLIMATE AND LAND USE/COVER

Akbar T. A., Hassan Q. K., Achari G., Clusterization of Surface Water Quality and Its Relation to Climate and Land Use/Cover. *Journal of Environmental Protection* **2013** Volume 4, Issue 4, pp. 333–343.

Abstract

The quality of surface water is rapidly changing due to climatic variations, natural processes, and anthropogenic activities. The objectives of this study were to classify and analyze the surface water quality of 12 major rivers of Alberta on the basis of 17 parameters during the period of five years (i.e., 2004-2008) using principal component analysis (PCA), total exceedance model and clustering technique. Seven major principal components (PCs) with variability of about 89% were identified. These PCs were the indicators of watershed geology, mineralization and anthropogenic activities related to land use/cover. The seven dominant parameters revealed from the seven PCs were total dissolved solids (TDS), true color (TC), pH, iron (Fe), fecal coliform (FC), dissolved oxygen (DO), and turbidity (TUR). The normalized data of dominant parameters were used to develop a model for obtaining total exceedance. The exceedance values acquired from the total exceedance model were used to determine the patterns for the development of five clusters. The performance of the clusters was compared with the classes obtained in Canadian Water Quality Index (CWQI). Cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5 showed agreements of 85.71%, 83.54%, 90.22%, 80.74%, and 83.40% with their respective CWQI classes on the basis of the data for all rivers during 2004-2008.The water quality was deteriorated in growing season due to snow melting. This methodology could be applied to classify the raw surface water quality, analyze the spatio-temporal trends and study the impacts of the factors affecting the water quality anywhere in the world.

Keywords: Alberta rivers; Canadian Water Quality Index; Clustering; Geographic Information System; Pattern recognition; Principal component analysis; River water quality.

4.1 Introduction

In general, the quality of waters in rivers and lakes depend on climate, land use, land cover, geographical and anthropogenic factors (Mahapatra and Mitra 2012; García-Reiriz et al. 2011; Toth et al. 2009; Zhu et al. 2012). Climatic factors, such as melting snow over high latitudes and precipitation wash material from the land surface into the water bodies. Various land use activities (e.g., wood logging, agricultural, mining and urban development) can be potential sources of pollutants, which impact the water quality. Thus, it is important to classify the raw surface water quality and study the spatio-temporal impacts due to anthropogenic activities and climatic factors.

In Alberta, 17 water quality-related parameters are periodically measured for 12 major rivers at 23 fixed sampling sites. These data are then analyzed using the Canadian Water Quality Index (CWQI) system developed by the Canadian Council of Ministers of the Environment (CCME); and represented as an index-value (CCME 2001). Despite the robustness and acceptance of CWQI, the data acquisition is labour intensive, time consuming and costly. Thus, it is worthwhile to investigate whether a lesser number of water quality-related parameters would produce similar CWQI-values.

In order to determine data redundancy in any dataset, one of the most commonly used methods is the employment of pattern recognition algorithms (Eneji et al. 2012; Singh et al. 2010). Examples of such algorithms are principal component analysis (PCA) and clustering techniques. In PCA, the original set of parameters is transformed into uncorrelated principal components (PCs), which decrease the total variance. Each parameter contributes towards its respective PC and its contribution is determined by the loading values. PCA has been used in many water quality studies, such as (i) determining spatio-temporal changes in the water quality of Jajrood River (Razmkhah et al. 2010), (ii) comparing water quality of regional sites of Canada for spatial and temporal changes (Rosemond et al. 2009), (iii) seasonal and spatial variations for surface water quality of Mid-Black Sea Coast in Turkey (Akbal et al. 2011), and (iv) impact of agricultural activities for Nathan Creek Watershed, British Columbia, Canada (Furtula et al. 2012).

The clustering techniques are used to find structure in data by identifying the groups (clusters) in the data and the objects are grouped on the basis of similarities within a class and dissimilarities among different classes. The similarities and dissimilarities are obtained on the basis of distance measures (e.g., Euclidean, Manhattan, etc.) using various clustering methods (Kaufman and Rousseeuw 1990). The clustering methods have been widely used in the water quality studies. For example: (i) clustering for chemical classification of water in Salado River (Gabellone et al. 2008), (ii) Hierarchical agglomerative cluster analysis for delineating and grouping pollution causing areas (Srivastava et al. 2011), and (iii) Fuzzy clustering of water quality parameters for Ulansuhai Lake (Ren et al. 2008). In addition to classification of water quality, it is also important to understand the impact of causative factors on the surface water quality of rivers in Alberta. For this purpose geographic information system (GIS) was used as its application was found useful in studying the water quality (Akbar and Akbar 2013; Akbar and Lin 2010). The objectives of this paper are to: (i) develop clusters for major rivers in Alberta on the basis of monthly water quality data, (ii) evaluate the clusters using Canadian Water Quality Index (CWQI) system, (iii) apply clusters for spatio-temporal analysis, and (iv) study the impact of climatic factor (i.e., snow-melting) and land use activities on the water quality of the rivers.

4.2 Materials and methods

4.2.1 Study area and data requirements

The study area consists of 12 major rivers in Alberta as shown in **Figure 4.1**. Alberta is a western province in Canada, which borders the province of British Columbia in west, and Saskatchewan in east. The mean annual temperature in winter varies from −25.1 ^{o}C to −9.6 ^{o}C and in summer it ranges from 8.7 ^{o}C to 18.5 ^{o}C. The mean average annual precipitation ranges from 333 mm to 989 mm (Downing and Pettapiece 2006). The major land use/cover types are needle leaf forests (57.57%), grasses/cereal crops (30.11%) and broad leaf forests (5.25%). The province is dominated by boreal forest in the north and agriculture in the south. At each of the sites, we obtained the monthly values of the 17 water quality-related parameters for the period 2004-2008 from Alberta Environment. These parameters included: chloride (Cl), dissolved organic carbon (DOC), dissolved oxygen (DO), fecal coliforms (FC), fluoride (F), iron (Fe), manganese (Mn), pH, sodium (Na), sulfate (SO_4), total dissolved solids (TDS), total hardness (TH), total nitrogen (TN), total phosphorus (TP), true color (TC), turbidity (TUR) and water temperature (WT). There are guideline values for each of these parameters in the context of determining the water quality (Health Canada 2010; Ministry of the Environment 2006; Alberta Environment and Sustainable Resource Development 2011). Those guidelines are summarized in **Table 4.1**. Different types of land uses/covers could change the water quality due to flow of various types of contaminants in the rivers (Bolstad & Swank 1997; Moss 1988). In Alberta, the mean annual temperature varies between −25.1℃ to −9.6℃ in winter and it ranges between 8.7℃ to 18.5℃ in summer. The temperature changes in both the seasons initiate the snow melting which could impact the quality of surface water. Thus we used the maps for land use/cover and snow-melting time period to understand the impact on the surface water quality. Those included: (i) Moderate Resolution Imaging Spectroradiometer (MODIS)-based annual composite land use/cover map at 1 km spatial resolution (MOD12Q1 ver. 004) during 2004 available from National Aeronautics and Space Administration (NASA 2004), and (ii) MODIS-derived snow melting time period map at 500 m spatial resolution during 2008 (Sekhon et al. 2010).

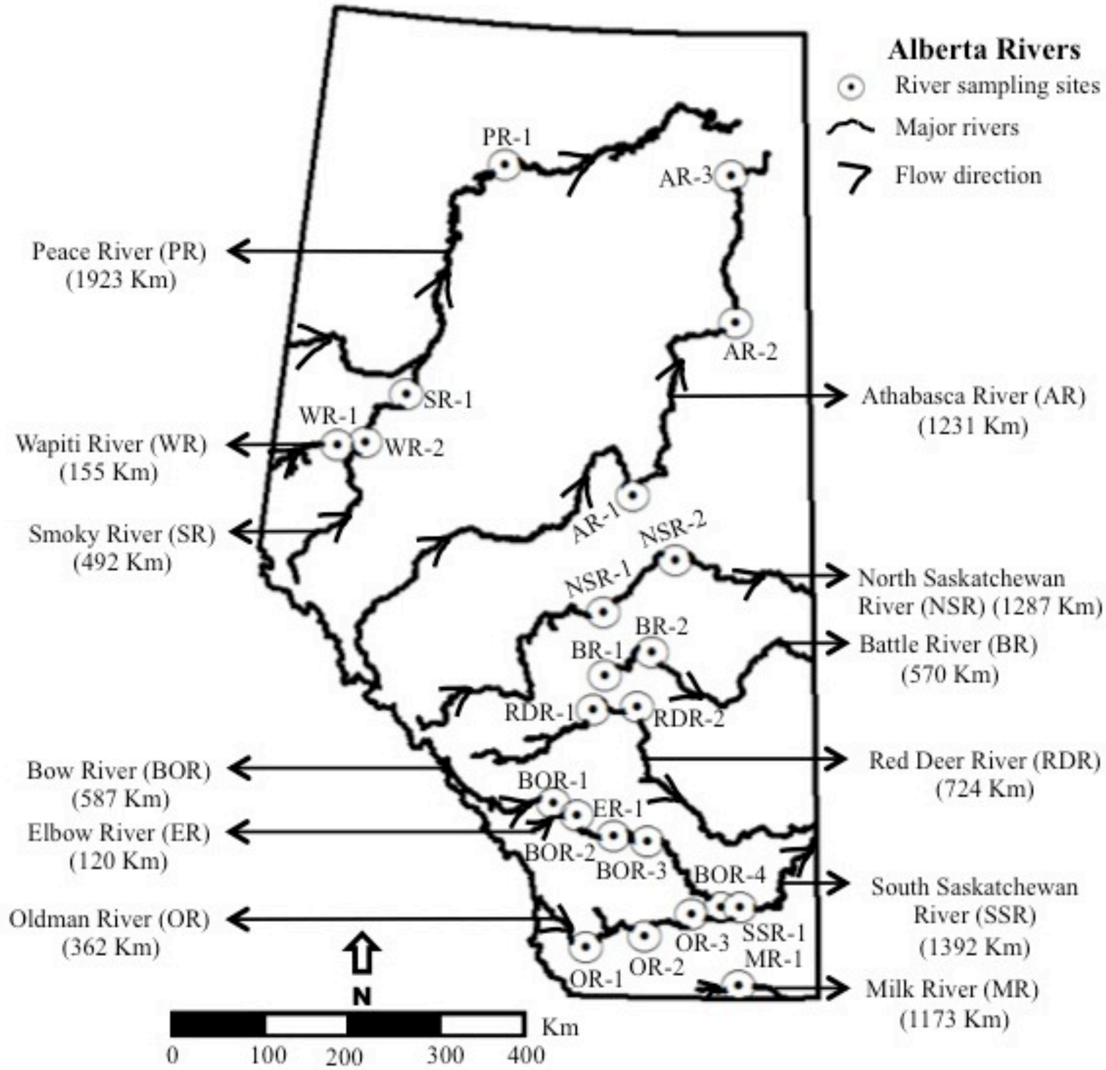

Figure 4.1: Location of 23 sampling sites across the twelve major rivers in Alberta. The lengths of rivers are provided in the parenthesis and the arrows show the directions of rivers' flow.

Table 4.1: Guidelines for Canadian drinking water quality (Health Canada 2010; Ministry of the Environment 2006; Alberta Environment and Sustainable Resource Development 2011).

Parameter	Non-compliance if guideline value:
WT	>15°C
DO	<6.5 mg/l
TUR	>1 NTU
TC	>15 Pt Co units
DOC	>5 mg/l
TDS	>500 mg/l
TP	>0.05 mg/l
TN	>1 mg/l
pH	<6.5 or >8.5
TH	>500 mg/l
Cl	>250 mg/l
SO_4	>500 mg/l
Na	>200 mg/l
F	>1.5 mg/l
FC	>0
Mn	>0.05 mg/l
Fe	>0.3 mg/l

4.2.2 Methods

The methods consisted of three major components, such as: (i) development of clusters, (ii) evaluation of clusters, and (iii) application of clusters. Brief descriptions of these components are as follows:

4.2.2.1 Development of clusters

For the development of clusters, we followed four steps, i.e., (i) normalizing water quality data, (ii) obtaining dominant parameters, (iii) developing total exceedance model, and (iv) identifying the cluster patterns. In the first step, the data was normalized for: (i) WT, TUR, TC, DOC, TP, TN, TDS, TH, Cl, SO_4, pH > 8.5, Na, F, Mn and Fe using Eq. (4.1), and (ii) DO and pH < 6.5 using Eq. (4.2).

$$(Parameter)_{normalization} = \left(\frac{(Parameter)_{measured}}{(Parameter)_{guideline}} \right)^{0.25} \qquad (4.1)$$

$$(Parameter)_{normalization} = \left(\frac{(Parameter)_{guideline}}{(Parameter)_{measured}} \right)^{0.25} \qquad (4.2)$$

In both the above equations (i.e., Eq. (4.1) and Eq. (4.2)) we used the power of a constant number (i.e., 0.25) to reduce the spread between the parameters due to large variations in their measured values. As the guideline was 0 for FC therefore we normalized it by exponention with exponent equal to 0.25.

In the second step, we used PCA to identify the major PCs and obtain the dominant parameters using the normalized data (Akbar et al. 2011). The numbers of PCs were decided by setting eigenvalue to 0.5 and the loading values of parameters were obtained using varimax normalized rotation (Razmkhah et al. 2010). The loading values were divided into three classes (i.e., strong > 0.75, 0.75 > moderate > 0.5 and 0.5 > weak > 0.4). Parameter loading values less than 0.40 were not considered because of their minor significance in the data (Panda et al. 2006). From each of the PCs, one of the parameters was selected as the dominant one on the basis of the highest loading values. In the third step, the normalized values of dominant parameters were used to develop a model for obtaining the total exceedance for each monitoring day during the period 2004-2008. In the fourth step, the exceedance values (obtained from the third step) were used to identify the patterns to develop clusters for the classification of surface water quality of the rivers. Seventy percent of the results obtained from the total exceedance model were used to develop the clusters and the remaining thirty percent of the results were used to evaluate them.

4.2.2.2 Evaluation of clusters

For quantitative evaluation, the percent cumulative agreements (in the form of deviation) between the clusters and CWQI classes were calculated for: (i) all rivers during the period 2004-2008, and (ii) each river during the whole period of 2004-2008. Several equations were used for calculating CWQI (see Table 4.2 for more details). On the basis of quantitative values (i.e. 0 to 100) calculated using CWQI, the water quality at the sampling sites of rivers was categorized into five classes which are (i) 1: excellent (95 - 100), (ii) 2: good (80 - 94), (iii) 3: fair (60 - 79), (iv) 4: marginal (45 - 59), and (v) 5: poor (0 - 44) (CCME 2001).

Table 4.2: Equations used for calculation of CWQI and identifying classes using the data of 23 sampling sites for 12 rivers during the period 2004-2008 (CCME 2001).

$$F_1\,(\text{Scope}) = \left(\frac{\text{Number of failed parameters}}{\text{Total number of parameters}}\right) \text{ x } 100$$

$$F_2\,(\text{Frequency}) = \left(\frac{\text{Number of failed tests}}{\text{Total number of tests}}\right) \text{ x } 100$$

$$\text{excursion}_i = \left(\frac{\text{Objective}_j}{\text{Failed test value}_i}\right) - 1$$

$$\text{excursion}_i = \left(\frac{\text{Failed test value}_i}{\text{Objective}_j}\right) - 1$$

$$\text{nse} = \left(\frac{\sum_{i=1}^{n} \text{excursion}_i}{\text{number of tests}}\right)$$

$$F_3\,(\text{amplitude}) = \left(\frac{\text{nse}}{0.01\text{nse} + 0.01}\right)$$

$$\text{CWQI} = 100 - \left(\frac{\sqrt{F1^2 + F2^2 + F3^2}}{1.732}\right)$$

4.2.2.3 Application of clusters

The dominant clusters were identified for the growing season (April 1-September 30) and the winter months (Oct 1-March 31) for all the sampling sites during 2004-2008. These dominant clusters were used to understand the: (i) spatio-temporal patterns of the surface water quality of rivers, and (ii) impact of land use/cover and snowmelt. To understand the influence of both the factors, all the rivers with their respective sampling sites were overlain in GIS on: MODIS based (i) land use/cover map, and (ii) snowmelt time period map.

4.3 Results and discussion

4.3.1 Major principal components and the dominant water quality parameters

PCA led to a set of seven principal components (PCs) using the normalized data during the period 2004-2008. These PCs had eigenvalues greater than 0.5. Individually they captured 31.5%, 20.8%, 12.6%, 9.1%, 6.1%, 5.6%, and 3.4% of the total variance (See **Table 4.3**). PC-1 revealed that four ions (i.e., Cl^-, SO^{-2}, Na^+ and F^-) accounted for most of the TDS, which was also related to the variation in TH. Thus it is interpreted as indicator of the watershed geology (Anderson 1999). TDS was considered as the first dominant parameter due to having highest loading value (i.e., 0.94). PC-2 indicated three correlated parameters (i.e., TC, DOC, and TP). This could be an indicator of natural and anthropogenic mineralization of water quality (Anderson 1999; Wolfe et al. 2007). In this category, two parameters (i.e. TC and DOC) are strongly positively loaded with TC having the highest loading (i.e., 0.95). TC was considered as the second dominant parameter. PC-3 indicated that pH > 8.5 and pH < 6.5 are strongly loaded with similar magnitudes (i.e., 0.98). WT was weakly negatively loaded in PC-3. In general, temperature increase during the spring season would initiate the process of snow melting, which contributes to the variation of pH in the water. Thus it could be the indicator of anthropogenic activities related to different types of land use/cover (Bruneau et al. 2009). In this component, pH was considered as the third dominant parameter. PC-4 indicated that two parameters (i.e., Mn and Fe) were strongly positively correlated. PC-4 was considered as an indicator of natural mineralization (Anderson 1999). Fe was considered as the fourth dominant parameter due to its highest loading (i.e., 0.96). PC-5 indicated solely FC as a strongly positively loaded parameter (i.e., 0.98). As FC is related to land cover activities therefore PC-5 could also be the indicator of anthropogenic activities like PC-3. FC was identified as the fifth dominant parameter. PC-6 showed DO as exclusive strongly positively loaded parameter having loading value of 0.93. DO was identified as the sixth dominant parameter. PC-6 was considered as an indicator of natural mineralization like PC-4 (Anderson 1999). PC-7 indicated TUR as strongly positively loaded and TP as moderately positively loaded parameter. TUR was considered as the seventh dominant parameter due to its highest loading value (i.e., 0.86). The snow melting and precipitation from the different types of land use/cover increase the sediment levels in the surface waters, which increase TUR. In PC-7, both the parameters (i.e., TUR and TP) are related to land cover activities. Like PC-3 and PC-5, it could also be considered as an indicator of anthropogenic activities related to different land cover types (Bruneau et al. 2009; Sosiak and Dixon 2006). Thus, the seven dominant parameters obtained from PCA were: TDS, TC, pH, Fe, FC, DO, and TUR.

Table 4.3: PCs with loading values for 17 water quality parameters in the five years (2004-2008).

Parameter	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7
WT			-0.44				
DO						0.93	
TUR							0.86
TC		0.95					
DOC		0.92					
TP		0.57					0.52
TN	0.43						
TDS	0.94						
pH>8.5			-0.98				
pH<6.5			0.98				
TH	0.92						
Cl	0.66						
SO_4	0.81						
Na	0.74						
F	0.81						
FC					0.98		
Mn				0.91			
Fe				0.96			
Var. (%)	31.5	20.8	12.6	9.1	6.1	5.6	3.4
Cum. (%)	31.5	52.3	64.9	74.0	80.1	85.7	89.1

Note: Var: Variance; Cum: Cumulative.

4.3.2 Databases of clusters and CWQI classes for classification of water quality

The normalized values of dominant parameters, obtained using Eq. (4.1) and Eq. (4.2), were used to develop a model for obtaining total exceedance as given in Eq. (4.3):

$$(\text{Exceedance})_{\text{total}} = \sum[(\text{Dominant parameter})_{normalized} - 1] \quad (4.3)$$

Using Eq. (4.3), we calculated the total exceedance values for the normalized data of the dominant parameters during 2004-2008. All of these exceedance values were then used to identify the patterns for the development and evaluation of five clusters. For presentation of cluster patterns in this paper, we used the minimum, maximum and mean exceedance values of dominant parameters as shown in **Figure 4.2.** It is obvious that minimum, maximum, and mean increase from cluster 1 to cluster 5.

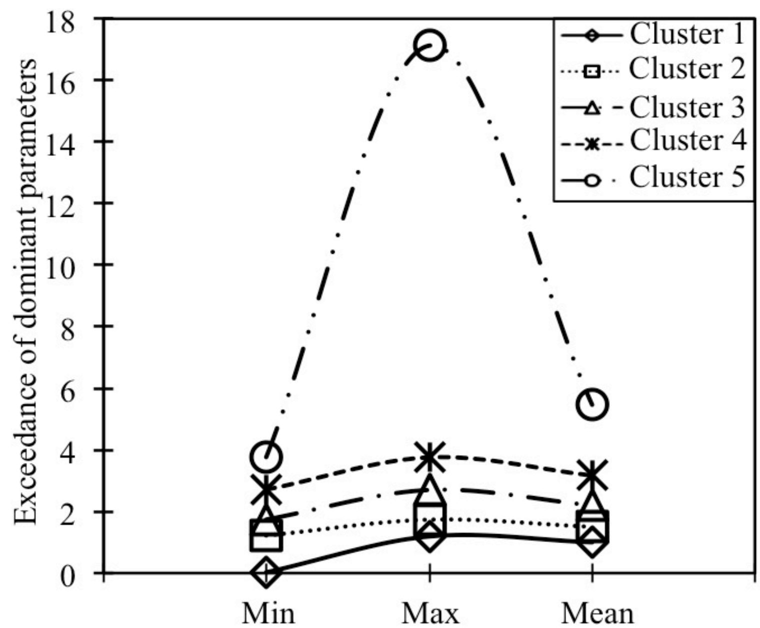

Figure 4.2: Patterns of five clusters produced from minimum, maximum and mean of the exceedance values of dominant parameters during the period 2004-2008. The exceedance values were calculated using the total exceedance model given in Eq. (4.3).

We used these clusters to define the water quality of rivers, which could change from cluster 1 towards cluster 5. The water quality deteriorates from cluster 1 to cluster 5. A database of clusters was developed by obtaining the clusters for all the sampling sites of rivers to classify the water quality in each month during 2004-2008. Another database of CWQI classes was also developed for the classification of water quality of rivers during the same time period.

4.3.3 Comparison of clusters with CWQI classes

Figure 4.3 (a) and **(b)** shows a comparison between % cumulative agreement and deviation for clusters with CWQI classes using the data of all rivers during the period 2004-2008. In the cluster development, the agreements for 0 deviation were 85.71%, 83.54%, 90.22%, 80.74%, and 83.40% for cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5 respectively as shown in **Figure 4.3(a)**. For the respective five clusters, the agreements for ±1 deviation were 14.29%, 16.46%, 8.83%, 19.26%, and 16.60%. An agreement of 0.95% was observed for ±2 deviation in cluster 3. In the cluster evaluation, the agreements for 0 deviation were 87.50%, 81.82%, 89.51%, 80.64% and 81.63% for cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5 respectively as shown in **Figure 4.3(b).** In these five clusters, the agreements for ±1 deviation were 12.50%, 18.18%, 9.09%, 19.36%, and 18.37% respectively. The agreement of 1.40% was found for ±2 deviation in cluster 3. These percentages of agreements showed very close match of clusters with CWQI classes. **Table 4.4** shows the % agreement for the deviations calculated for each river during the period 2004-2008. From **Table 4.4**, we found 0 deviation (i.e., 100% agreement) for majority of the rivers for most of the times (*i.e.,* in between 80-100% of the cases). In limited number of cases, we observed that the agreement for 0 deviation was between 20% - 73% of the cases for Battle River, Elbow River, Milk River, South Saskatchewan River and Peace River. This difference in agreements from majority of the rivers could be related to the impact of exceedance for parameters other than the dominant once. The quantitative evaluation showed a reasonably strong match between clusters and CWQI classes, which indicates the suitability and usefulness of cluster based classification system for the surface water quality of major rivers of Alberta. The clusters were plotted against CWQI classes for a sampling site of Bow River (i.e., BOR-1) over a period of five years (i.e. 2004-2008) as shown in **Figure 4.4**. In this figure, about 90% of observed data showed complete match between clusters and classes whereas only 10% of observed data showed the deviation of ±1. Overall, the patterns of clusters matched quite well with the patterns of CWQI classes as shown in the **Figure 4.4**.

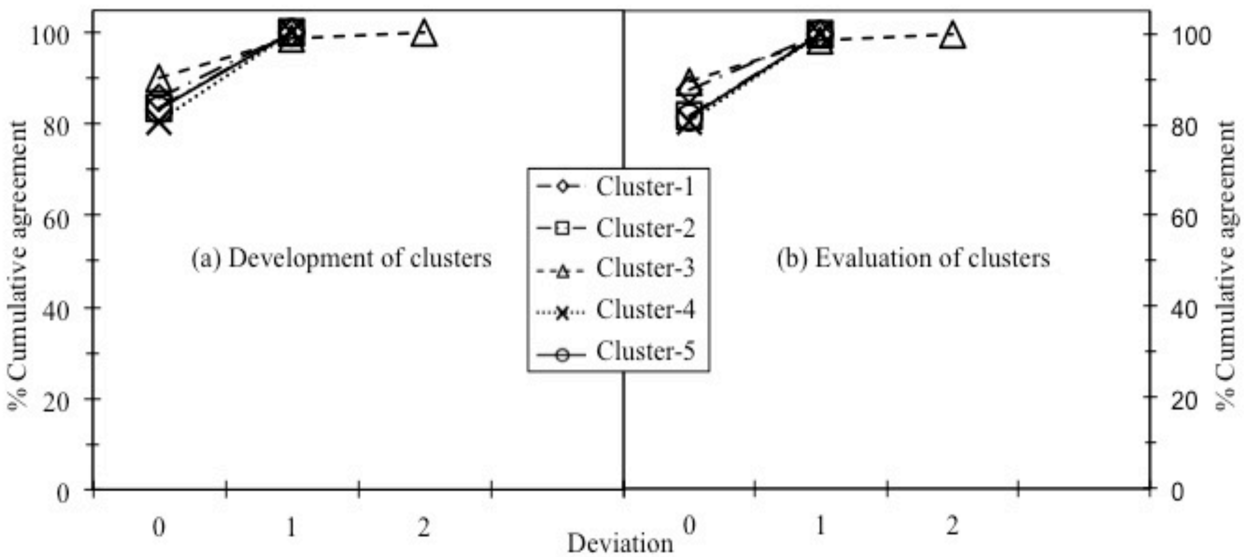

Figure 4.3: Percentage cumulative agreement between clusters and CWQI classes on the basis of deviations for: (a) Development of clusters, and (b) Evaluation of clusters.

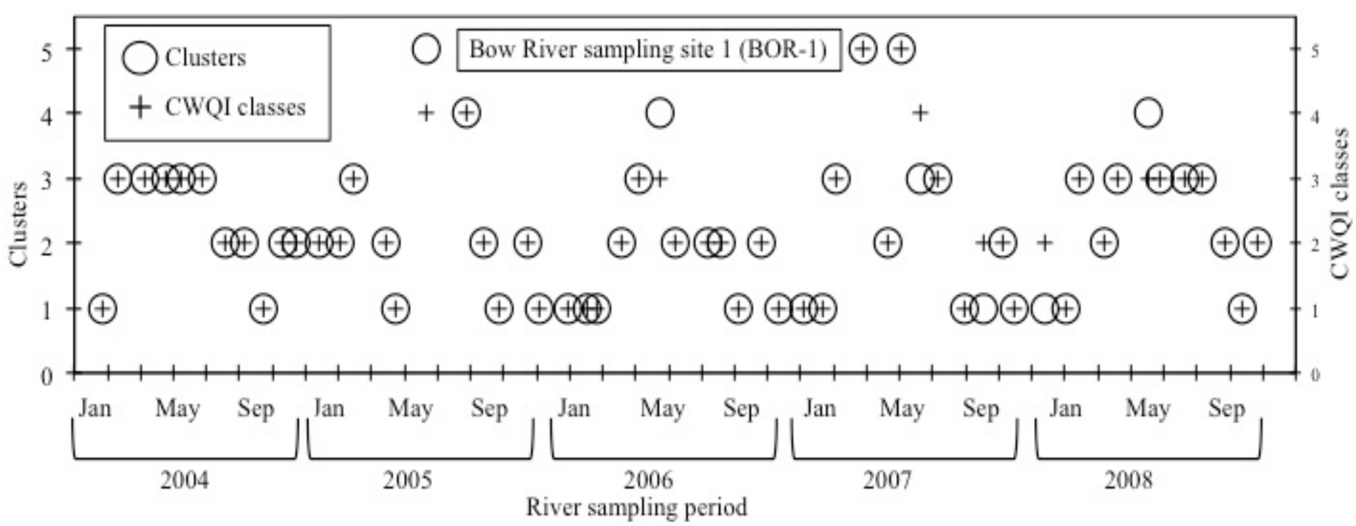

Figure 4.4: Comparison between clusters and CWQI classes for a sampling site (BOR-1) of the Bow River during the period of 2004-2008.

Table 4.4: Percentage agreement for deviation of clusters on the basis of quantitative evaluation for each river during the period 2004-2008.

River	Cluster	% Agreement for deviation			River	Cluster	% Agreement for deviation	
		0	±1	±2			0	±1
	3	91.18	7.35	1.47		1	100.00	
AR	4	93.33	6.67			2	100.00	
	5	89.13	10.87		OR	3	92.59	7.41
						4	88.24	11.76
	3	29.41	52.94			5	72.00	28.00
BR	4	39.13	60.87					
	5	100				3	96.15	3.85
					PR	4	50.00	50.00
	1	84.21	15.79			5	100.00	
	2	82.61	17.39					
BOR	3	94.29	5.71			1	100.00	
	4	87.80	12.20			2	87.50	12.50
	5	61.11	38.89		RDR	3	87.88	12.12
						4	96.43	3.57
	3	60.00	40.00			5	83.33	16.67
ER	4	96.43	3.57					
	5	73.33	26.67			3	92.59	7.41
					SR	4	100.00	
	4	20.00	80.00			5	100.00	
MR	5	100.00						
						2	60.00	40.00
	3	95.52	2.99	1.49	SSR	3	85.00	15.00
NSR	4	90.91	9.09			4	84.62	15.38
	5	83.33	16.67			5	66.67	33.33
						3	91.30	8.70
					WR	4	86.21	13.79
						5	80.00	20.00

4.3.4 Application of clusters for spatio-temporal trends

We discussed below the classified water quality for five of the twelve major rivers in Alberta on the basis of clusters. The discussion on the application of clusters for the remaining seven rivers is given in appendix-I. The monthly clusters obtained for these five rivers during the period 2004-2008 are shown in **Tables 4.5-4.7.** An example for studying the spatio-temporal trends from the clusters is presented in **Figure 4.5** for all the sampling sites on the Bow River (see Section 3.4.2). The impacts of land cover **(Figure 4.6(a))** and snow melting **(Figure 4.6(b))** on the water quality of rivers was also discussed in the same sub-sections.

4.3.4.1 Athabasca River

The dominant cluster for all three sampling sites (AR-1, AR-2 and AR-3) of Athabasca River was cluster 5 during the growing season and it was cluster 3 during the winter season from 2004 to 2008 as shown in **Table 4.5**. In 2008, the snowmelt period ranged from 16-May-08 to 24-Jun-08 and 16-May-08 to after 25-Jun-08. This was dominant on the downstream and upstream sides of Athabasca River respectively as shown in **Figure 4.6(b).** It indicates that the melting snow is contributing more towards the deterioration in water quality during the growing season as compared to the winter months. The potential sources of deterioration in this river are also the surface runoff from the different land cover types that include needle leaf forests, broad leaf forests and cereal crops/grasses as shown in **Figure 4.6(a).** A study done for Athabasca River found that the contamination was associated to land-use related run-off from the forestry and agricultural activities (Wrona et al. 2000).

4.3.4.2 Bow River

We obtained the dominant clusters from **Table 4.6** for the four sampling sites (BOR-1, BOR-2, BOR-3 and BOR-4) of Bow River during the growing season in the period 2004-2008: (i) BOR-1 belonged to cluster 2 during the period 2005-2006 and cluster 3 in 2004 and 2008, (ii) BOR-2 belonged to cluster 4 in 2004 and 2007 and cluster 5 during the periods of 2005-2006 and 2008, (iii) BOR-3 fitted in cluster 4 during 2004-2005 and 2008, cluster 3 in 2006 and cluster 5 in 2007, (iv) BOR-4 be- longed to cluster 4 in 2004, cluster 3 in 2005 and cluster 5 in 2006-2008. From **Table 4.6**, it was obvious that during the winter season, the dominant cluster was (i) cluster 1 for BOR-1 in 2004 and 2006-2008, (ii) cluster 3 for BOR-2 in 2004-2008, (iii) cluster 2 for BOR-3 in the periods of 2006-2007 and cluster 3 in 2008, and (iv) cluster 3 for BOR-4 in 2004-2007. In 2008, the change in clusters from winter to growing season for all sampling sites was related to snow melting period. **Figure 4.6(b)** indicates that the snow melting period in year 2008 started earlier (i.e., before 5-Apr-08) for BOR-2, BOR-3 and BOR-4 as compared to snow melting period of BOR-1 (i.e., 6-Apr-08 to 15-May-08). The snow melting period could also contribute towards the deterioration of surface quality of Bow River in 2004-2007. The cluster results also revealed that the surface water quality of Bow River in BOR-2, BOR-3 and BOR-4 deteriorated as compared to BOR-1 during the growing season. This was related to the agricultural activities of cereal and broad leaf crops as these three sites are located in adjacent agricultural areas as shown in **Figure 4.6(a)**. In comparison, BOR-1 is located near a needle leaf forest. Agriculture consumes 90% of the total water usage in South Saskatchewan River Basin and the Bow River is one the major rivers of this basin (Bruneau et al. 2009).

4.3.4.3 Milk River

For the sampling site (MR-1) of Milk River, the dominant cluster was cluster 5 in growing season as well as in winter during the period 2004-2008 as given in **Table 4.7**. The dominant land cover type around Milk River is cereal crops/grasses and the snow melting period around this river was before 5-April-08 as shown in **Figures 4.6(a)** and **(b)** respectively. The deteriorated water quality of Milk River in growing season was because of agricultural activities and surface runoff due to snow melting. The natural mineralization in Milk River due to manganese and iron could be a significant factor for unsatisfactory water quality throughout the year (Anderson 1999).

4.3.4.4 North Saskatchewan River

Table 4.7 shows that the dominant cluster for both sampling sites (NSR-1 and NSR-2) of North Saskatchewan River was cluster 3 each year in winter during the period 2004-2008 except 2004 for NSR-2 in which it was cluster 4. During the growing season, the dominant cluster was: (i) cluster 4 in 2004, cluster 5 in 2005 and 2007, and cluster 3 in 2006 and 2008 for NSR-1, and (ii) cluster 4 in 2004, cluster 5 in 2005-2006 and 2008 and cluster 3 in 2007 for NSR-2. A major portion of North Saskatchewan River along with their sampling sites is dominated by cereal crops/grasses on downstream side of the river and on the upstream side it is covered mostly by needle leaf and broad leaf forests according to the land cover classes shown in **Figure 4.6(a).** Cluster 4 and cluster 5 for NSR-1 and NSR-2 in the growing seasons during the period 2004-2008 were due to the agricultural activities. In 2008, the snow melting period was between 6-Apr-08 to 15-May-08, which changed the cluster from (i) cluster 3 in April to cluster 4 in May for NSR-1, and (ii) cluster 3 in April to cluster 5 in May for NSR-2 as shown in Figure 6(b). The variation of clusters in different months during the period 2004-2008 was related to snow melting. The potential sources of contamination for the North Saskatchewan River could be the pollutants carried by snowmelt from the activities related to agriculture and forestry (Mitchell 1994).

4.3.4.5 Peace River

The dominant cluster was cluster 3 for PR-1 in winter season during the period 2004-2008 as obvious from **Table 4.7**. From this table, we also observed that during the growing seasons, the dominant cluster for PR-1 was: (i) cluster 5 in 2004-2005, (ii) cluster 3 in 2006, and (iii) cluster 4 in 2007-2008. Most of Peace River is in the snow melting period of 6-Apr-08 to 15-May-08 as shown in **Figure 4.6(b)**, due to which it was observed that PR-1 had cluster 3 from January to March and cluster 4 in April and cluster 5 from May to June during the year 2008. The reason for the variation in the cluster during the winter and growing seasons for the period 2004-2007 is related to snowmelt period as it was observed for the year 2008. The land cover map **(Figure 4.6(a))** shows that the upstream of Peace River and the area surrounded by the sampling site (PR-1) have cereal crops/grasses whereas the downstream of Peace River is dominated by needle leaf forests. The potential sources of contamination were runoff due to the forests and the agricultural activities (Wrona et al. 2000).

Table 4.5: Clusters for three sampling sites (AR-1, AR-2, AR-3) of Athabasca River during the period 2004-2008.

	2004			2005			2006			2007			2008		
Month	AR-1	AR-2	AR-3	AR-1	AR-2	AR-3	AR-1	AR-2	AR-3	AR-1	AR-2	AR-3	AR-1	AR-2	AR-3
Jan	3	3	3	3	3	4	3	5	N	3	N	3	3	3	3
Feb	3	3	3	3	3	4	3	3	3	3	3	4	4	3	3
Mar	3	5	4	4	3	4	3	3	3	3	3	N	3	4	3
Apr	5	5	N	5	3	N	3	3	N	5	N	N	3	3	4
May	4	5	4	4	5	5	4	5	4	5	5	5	5	5	N
June	5	5	5	4	5	5	4	5	5	5	5	5	5	5	N
July	5	4	5	5	4	5	5	4	5	5	4	5	5	5	4
Aug	4	5	4	4	4	3	4	5	4	5	5	5	3	5	N
Sep	5	4	5	4	3	4	3	3	3	3	3	4	3	4	5
Oct	4	3	2	3	3	4	4	4	4	4	3	4	3	3	4
Nov	4	3	N	3	N	4	3	3	3	3	N	3	3	3	N
Dec	3	3	4	3	N	3	3	5	3	3	4	3	3	N	4

Note: N: No data; B1: BOR-1; B2: BOR-2; B3: BOR-3; B4: BOR-4.

Table 4.6: Clusters for four sampling sites of Bow River during the period 2004-2008.

	2004				2005				2006				2007				2008			
Mon	B1	B2	B3	B4	B1	B2	B3	B4	B1	B2	B3	B4	B1	B2	B3	B4	B1	B2	B3	B4
Jan	N	3	N	3	2	3	N	3	1	3	2	2	1	4	2	2	1	4	2	2
Feb	1	3	N	3	2	3	N	2	1	4	3	3	1	N	N	3	1	3	3	3
Mar	3	4	N	4	3	4	N	3	1	3	2	3	3	5	2	N	3	4	3	2
Apr	3	4	N	4	2	4	N	3	2	2	3	4	5	3	5	5	2	3	2	2
May	3	4	4	4	1	5	N	3	3	N	3	5	2	4	3	5	3	4	4	5
June	3	4	N	4	5	5	N	5	4	5	5	5	5	5	5	5	4	5	5	5
July	3	4	N	3	N	5	N	4	2	4	4	4	3	4	5	3	3	4	4	5
Aug	2	4	4	5	4	5	4	5	2	4	3	3	3	N	3	5	3	5	3	4
Sep	2	4	3	3	2	4	3	4	2	5	4	5	1	5	5	2	3	5	3	5
Oct	1	4	2	5	1	4	2	3	1	3	2	N	1	3	3	N	2	3	3	4
Nov	2	2	3	2	2	2	3	3	2	3	3	3	2	3	2	3	1	2	2	2
Dec	2	3	N	3	1	3	N	1	1	3	2	2	1	3	2	N	2	3	3	3

Note: N: No data; B1: BOR-1; B2: BOR-2; B3: BOR-3; B4: BOR-4.

Table 4.7: Clusters for (i) one sampling site of Milk River, (ii) two sampling sites of North Saskatchewan River, and (iii) one sampling site of Peace River during the period 2004-2008.

	Milk River					North Saskatchewan River										Peace River				
	04	05	06	07	08	04		05		06		07		08		04	05	06	07	08
Month	M1	M1	M1	M1	M1	N1	N2	N1	N2	N1	N2	N1	N2	N1	N2	P1	P1	P1	P1	P1
Jan	N	4	N	N	5	3	4	3	4	3	3	3	3	3	3	3	N	3	3	3
Feb	N	N	N	N	N	3	3	3	4	3	3	3	4	3	3	3	3	3	3	3
Mar	5	4	5	5	5	3	5	3	5	3	3	3	4	3	5	3	5	3	3	3
Apr	5	5	5	5	N	4	4	4	5	4	4	5	3	3	3	5	5	3	3	4
May	5	5	5	N	N	3	3	5	5	3	3	5	5	4	5	5	5	N	5	5
June	5	5	5	N	N	4	4	5	5	3	5	5	5	5	5	5	5	5	4	5
July	5	5	5	N	N	4	5	5	5	3	5	4	4	5	5	5	4	N	4	4
Aug	5	5	5	N	5	3	3	3	3	3	3	3	4	3	5	4	4	N	N	3
Sep	5	5	5	5	5	4	5	5	5	3	5	3	3	3	3	5	4	3	4	4
Oct	4	4	N	N	N	3	4	3	3	3	5	3	3	3	3	5	3	4	4	3
Nov	N	4	5	5	N	3	4	3	3	3	3	3	4	3	3	4	4	3	3	3
Dec	N	N	N	5	N	3	4	3	3	3	5	3	3	3	3	N	3	3	3	3

Note: N: No data; M1: MR-1; N1: NSR-1; N2: NSR-2; P1: PR-1

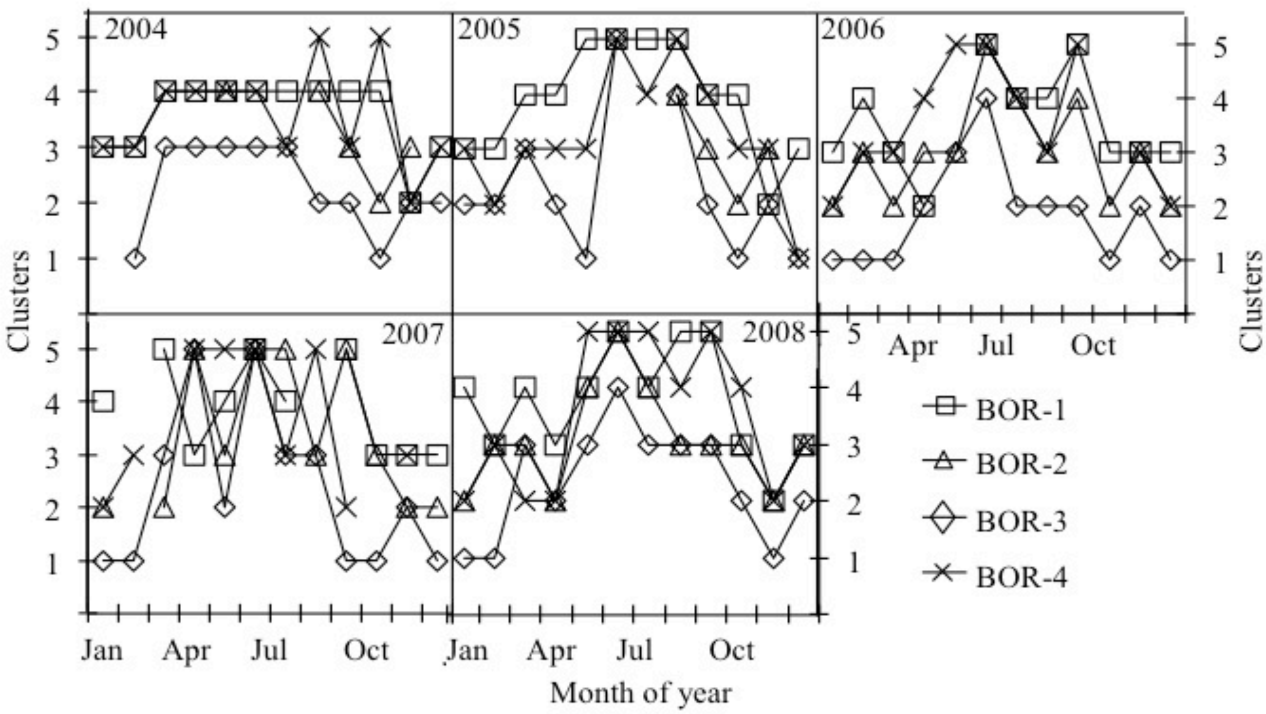

Figure 4.5: The spatial and temporal trends for the four sampling sites (i.e., BOR-1, BOR-2, BOR-3, and BOR-4) of the Bow River using the clusters during the period of 2004-2008.

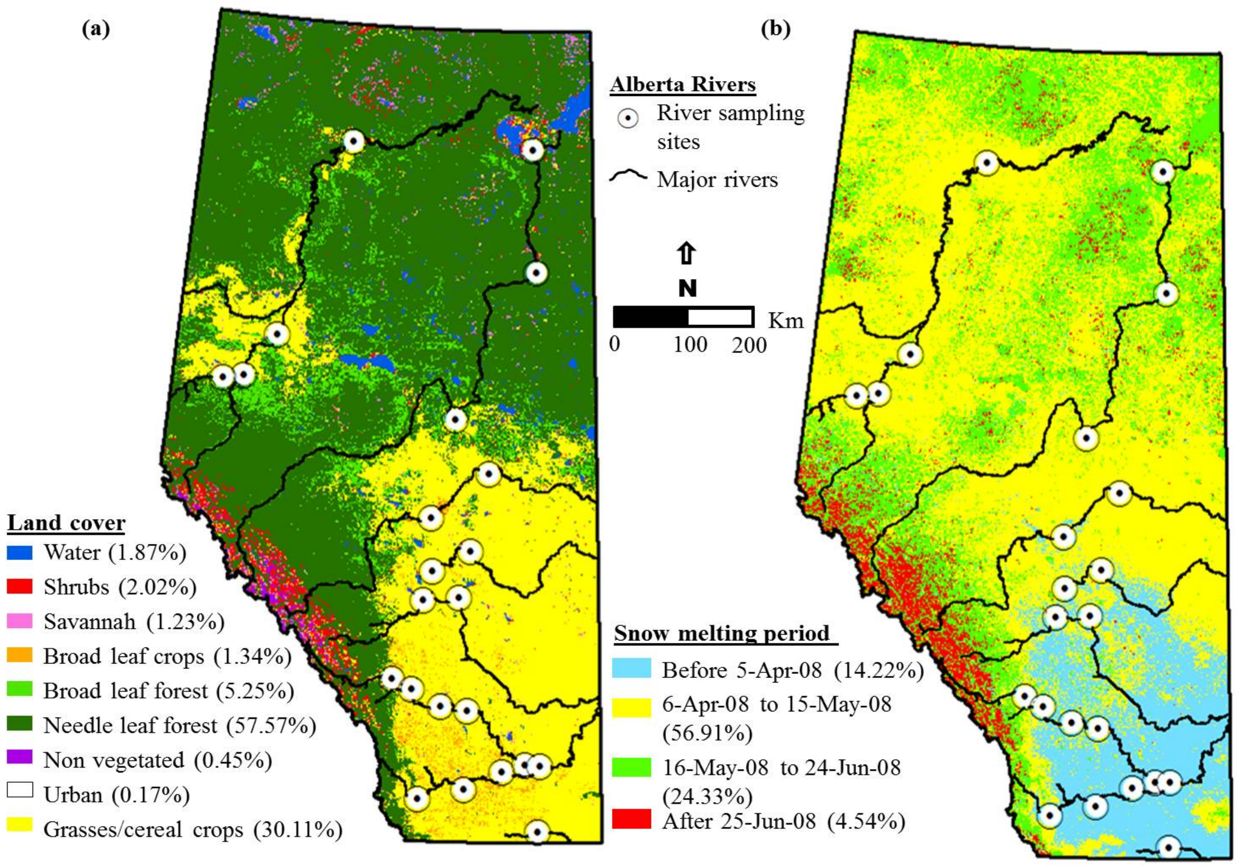

Figure 4.6: Overlay of the major rivers with their sampling sites on: (a) Land use/cover classes, and (b) Snow melting periods.

4.4 Conclusions

In this paper, we classified and analyzed the surface water quality for 12 major rivers in Alberta using the data of 17 parameters for 23 sampling sites during 2004-2008. For classifying the water quality, the clusters were developed and evaluated using CWQI. We developed the normalization models on the basis of Canadian water quality guidelines. The normalized data was then used for PCA to obtain the PCs and identify the dominant parameters. The dominant parameters were used to develop the total exceedance model. The exceedance values of dominant parameters were used to generate the clusters on the basis of identified patterns. The clusters were applied for spatio-temporal analysis. From PCA, we found that PC-1 was indicator of watershed geology. PC-2, PC-4, and PC-6 were indicators of natural and anthropogenic mineralization. PC-3, PC-5 and PC-7 were indicators of activities related to land use/cover. The clusters for all the rivers showed a very strong relationship with CWQI classes. From the cluster analysis, mostly higher (worse condition) cluster number (i.e. 4, 5) were observed for majority of the rivers in the growing seasons as compared to the lower cluster numbers (i.e. 1, 2, 3) in the winters. These would be related to the fact that the snow melting would potentially deteriorate the water quality due to anthropogenic activities from different land use/cover as interpreted in PC-3, PC-5 and PC-7. The agricultural activities were also responsible for deteriorating the water quality of rivers during the growing seasons. We observed the most deteriorated water quality for Battle River and Milk River. The methodology of this study was useful in: (i) grouping a large set of parameters into smaller set of meaningful PCs, (ii) interpreting each PC for some natural or anthropogenic activity, (iii) identifying the dominant parameters, (iv) classifying the large water bodies into clusters, (v) identifying the patterns of clusters, (vi) performing the spatial analysis, (vii) obtaining the temporal trends, and (viii) identifying the potential contamination sources. We suggest applying this method for monitoring, classifying and analyzing the surface water quality in an economical, efficient and user-friendly manner.

CHAPTER 5

EXCEEDANCE MODELING FOR SURFACE WATER QUALITY PARAMETERS

Abstract

The drinking water treatment technology can be expensive and ineffective if implemented without identifying the patterns of parameter exceedances. The objectives of this study were to develop the exceedance model for: (i) identifying the parameters, which exceeded the Canadian drinking water quality guidelines, and (ii) obtaining the exceedance patterns of parameters for clusters of 12 major rivers of Alberta during the five year period (2004-2008). The clusters were obtained using the total exceedance model developed and presented in Chapter 4. The monthly water quality data for seventeen parameters was normalized using the normalization models. In this study, a mean exceedance model was developed for obtaining the exceedance of parameters for the clusters of rivers. The mean exceedance for the parameters increased as the cluster number increased from low to high for all the rivers. Overall, the mean exceedance was higher for FC, TUR, TP, TN, TC, DO, Fe and Mn. The exceedance in FC, TUR, TP, TN, TC, and DO was related to anthropogenic activities of land cover/uses. The exceedance in Fe and Mn was due to natural mineralization. The mean exceedance model was found useful for obtaining the specific parameters with their exceedance levels. The parameter exceedance patterns could be utilized for the development of economical, efficient and targeted treatment technology for the source waters.

Keywords: Alberta Rivers; Canadian water quality guidelines; Drinking water treatment; Parameter exceedance

5.1 Introduction

In Alberta, the main source of drinking water is surface water (The Water Chronicles 2008). There is an ever-increasing pressure on surface water resources due to rapidly increasing population and urbanization. It is well understood that the water quality varies significantly at different sources, which requires different level of treatment. It is quite possible that in some places advanced level of treatment might not be necessary as the water quality is intrinsically good. This will lead to more effective savings and proper utilization of the resources.

In Canada, the selection and use of water treatment technologies is mainly driven by the Canadian drinking water guidelines and regulatory requirements. While the large municipalities those that have significant financial resources can opt for state-of-the-art technologies, which might be very expensive, most small water systems cannot afford that because of financial and other resource constraints. These small and financially constrained water systems need technologies that are effective and yet economical. Thus, the treatment technology should be targeted to pollutants of concerns, which will depend on the source water quality.

Different treatment technologies target different pollutants. For example in-line filtration is useful for water having low turbidity, direct filtration is beneficial for low to moderate turbidity and conventional treatment is good for high turbidity (Crittenden et al. 2005). Different membrane filtration technologies can be used for treating different pollutants e.g. (i) nanofiltration can be used for removal of calcium and magnesium ions, (ii) ultrafiltration can be used for removal of calcium and magnesium ions, and (iii) microfiltration can be used for removing pathogens (Jacangelo 1991; Taylor 1990). Different types of disinfectants (e.g., chlorine, chloramine, UV light, ozone, and chlorine dioxide) are used for controlling bacteria, virus and other organisms (LeChevallier et al. 1990; Wilczak et al. 1996).

It is important to target the pollutants that are endemic to a particular area and accordingly devise a suitable treatment technology. The objectives of this chapter are to: (i) develop the mean exceedance model, (ii) identify the parameters which exceed the Canadian drinking water quality guidelines, (iii) obtain the exceedance patterns of parameters in clusters of 12 major rivers of Alberta during the period of five years (i.e., 2004-2008), and (iv) conduct a review of treatment technologies for the exceeded parameters.

5.2 Materials and methods

5.2.1 Study area and data requirements

The study area has 12 major rivers of Alberta (See **Figure 4.1** in chapter 4**)**. Alberta is a western province of Canada where the average annual temperature in winter ranges from −25.1℃ to −9.6℃ and in summer it varies from 8.7℃ to 18.5℃ (Downing and Pettapiece 2006). The major land use/cover types are needle leaf forests (57.57%), grasses/cereal crops (30.11%) and broad leaf forests (5.25%) (NASA 2004). There are 23 sampling sites of the rivers. For each sampling site, we obtained the monthly values of the 17 water quality parameters for the period 2004-2008 from Alberta Environment. These parameters included: chloride (Cl), dissolved organic carbon (DOC), dissolved oxygen (DO), fecal coliforms (FC), fluoride (F), iron (Fe), manganese (Mn), pH, sodium (Na), sulfate (SO_4), total dissolved solids (TDS), total hardness (TH), total nitrogen (TN), total phosphorus (TP), true color (TC), turbidity (TUR) and water temperature (WT). The guideline values for each of these parameters are given in **Table 4.1** (Health Canada 2010; Ministry of the Environment 2006; Alberta Environment and Sustainable Resource Development 2011).

5.2.2 Methods

We used the data of monthly clusters developed in chapter 4 for all the twelve rivers of Alberta during 2004-2008. For each monthly cluster of a river, we obtained the mean exceedance for seventeen parameters during the period 2004-2008. For this purpose the measured water quality data was normalized for: (i) WT, TUR, TC, DOC, TP, TN, TDS, TH, Cl, SO_4, pH > 8.5, Na, F, Mn and Fe using Eq. (5.1), and (ii) DO and pH < 6.5 using Eq. (5.2). Both of these equations were developed in chapter 4:

$$(\text{Parameter})_{\text{normalization}} = \left(\frac{(\text{Parameter})_{\text{measured}}}{(\text{Parameter})_{\text{guideline}}}\right)^{0.25} \quad (5.1)$$

$$(\text{Parameter})_{\text{normalization}} = \left(\frac{(\text{Parameter})_{\text{guideline}}}{(\text{Parameter})_{\text{measured}}}\right)^{0.25} \quad (5.2)$$

We developed the mean exceedance model given in Eq. (5.3) to determine the magnitude of exceedance for the parameters of the clusters in all the rivers of study area.

$$(\text{Mean Exceedance})_{\text{parameter}} = \text{Mean}[(\text{parameter})_{\text{normalized}} - 1] \quad (5.3)$$

5.3 Results and discussion

In this section, we discussed about the: (i) mean exceedance of parameters and (ii) treatment technologies for the exceeded parameters on the basis of literature review.

5.3.1 Mean exceedance of parameters for rivers

We discussed below the mean exceedance of the parameters for the respective clusters of the twelve major rivers in Alberta (See sub-section 5.3.1 to 5.3.12). On the basis of mean exceedance, we identified the parameters exceeded for the clusters.

5.3.1.1 Athabasca River

For Athabasca River, the mean exceedance for the parameters is given in **Fig. 5.1**. The mean exceedance for FC was above: (i) 1 for cluster 3 and cluster 4, (ii) 2 for cluster 5. The mean exceedance for TUR was above: (i) 0.40 for cluster 3, (ii) 1 for cluster 4, (iii) 2 for cluster 5. For cluster 5, the mean exceedance for (i) TP was above 0.30, (ii) TC and DOC was above 0.20, (iii) Fe was above 0.10. For cluster 4, the mean exceedance for: (i) Fe was above 0.20, (ii) TP, TC and DOC was above 0.10. For cluster 3, the mean exceedance for: TP, TC, and DOC was above 0.10. Overall the parameters with the higher mean exceedance values were FC and TUR. The lowest to highest exceedance were observed from cluster 3 towards cluster 5.

5.3.3.2 Battle River

The mean exceedance for the parameters of Battle River is given in **Fig. 5.2**. The mean exceedance for FC was above: (i) 1 for cluster 3, (ii) 2 for cluster 4 and cluster 5. The mean exceedance for TUR was above: (i) 0.50 for cluster 3, (ii) 0.80 for cluster 4 and (iii) 1 for cluster 5. The mean exceedance for Mn was above: (i) 0.10 for cluster 3, (ii) 0.40 for cluster 4, (iii) 1 for cluster 5. The mean exceedance for DO was above: (i) 0.20 for cluster 4, and (ii) 0.60 for cluster 5. The mean exceedance for TP was above: (i) 0.20 for cluster 3, (ii) 0.30 for cluster 4, and (iii) 0.50 for cluster 5. The mean exceedance for DOC was above: (i) 0.20 for cluster 3, (ii) 0.30 for cluster 4, and (iii) 0.40 for cluster 5. The mean exceedance for TC was above: (i) 0.20 for cluster 3, (ii) 0.10 for cluster 4, and (iii) 0.20 for cluster 5. The mean exceedance for TN was above: (i) 0.20 for cluster 4 and cluster 5. We also observed mean exceedance over 0.10 for TDS and Fe in cluster 5. The parameters, which showed higher mean exceedance as compared to others parameters, were FC, TUR, Mn, DO, and TP. The range of mean exceedance for these parameters was from: (i) 0.28 to 1.77 in cluster 3, (ii) 0.38 to 2 in cluster 4, and (iii) 0.58 to 2.97 in cluster 5.

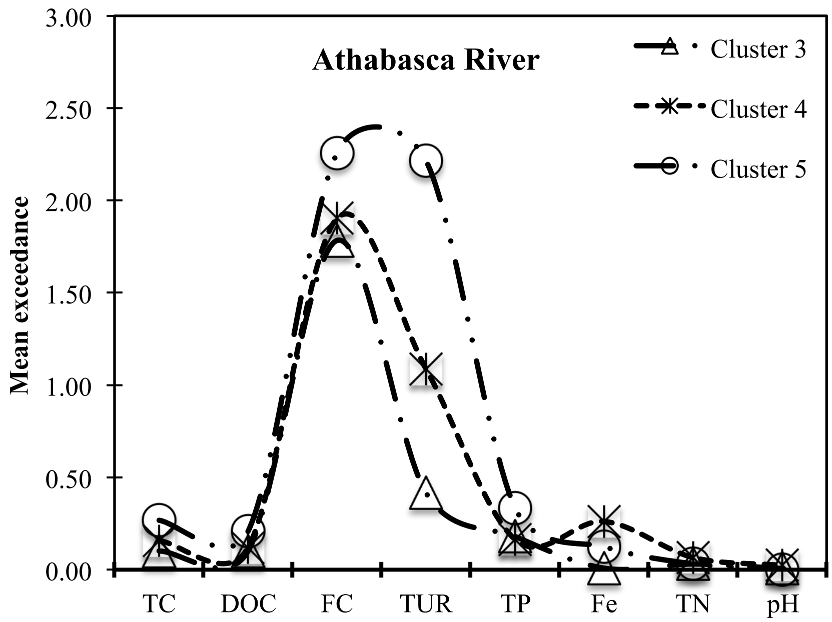

Figure 5.1: Mean exceedance of parameters in the clusters of Athabasca River.

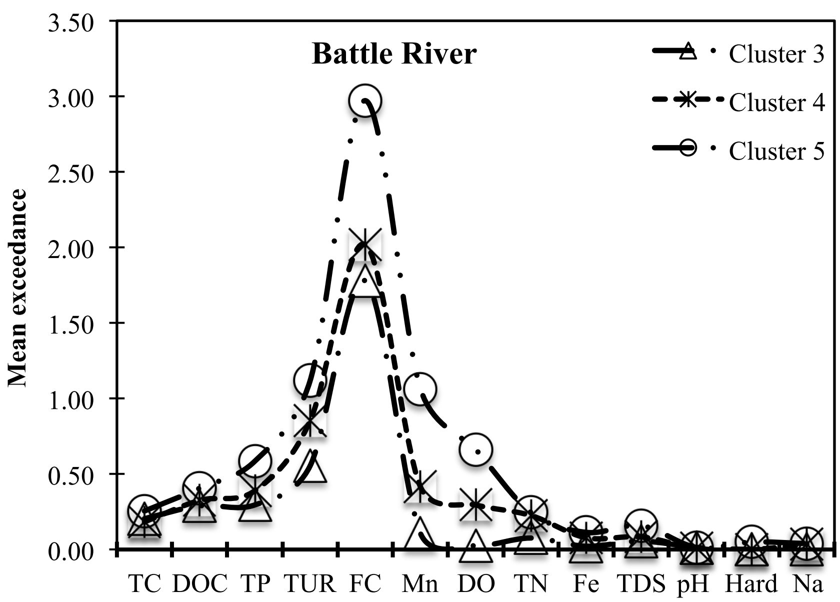

Figure 5.2: Mean exceedance of parameters in the clusters of Battle River.

5.3.3.3 Bow River

For Bow River, the mean exceedance for the parameters is given is **Fig. 5.3**. The lowest to highest exceedance for the parameters were observed from cluster 1 towards cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 1, cluster 2 and cluster 3, (ii) 2 for cluster 4, and (iii) 3 for cluster 5. The mean exceedance for TUR was above: (i) 0.2 for cluster 2, (ii) 0.4 for cluster 3, (iii) 0.5 for cluster 4, and (vi) 1 for cluster 5. The mean exceedance for TP was above: (i) 0.1 for cluster 3 and cluster 4, and (ii) 0.3 for cluster 5. Fe also showed exceedance of about 0.30 in cluster 4. FC and TN showed exceedance in all five clusters. FC, TUR and TP showed higher exceedance in cluster 3, cluster 4 and cluster 5.

5.3.3.4 Elbow River

For Elbow River, the lowest to highest exceedance for the parameters were observed from cluster 2 towards cluster 5 (See **Fig. 5.4**). The mean exceedance for FC was above: (i) 1 for cluster 2, (ii) 2 for cluster 3 and cluster 4, and (iii) 4 for cluster 5. The mean exceedance for TUR was above: (i) 0.05 for cluster 2, (ii) 0.10 for cluster 3, (iii) 0.3 for cluster 4, and (iv) 0.5 for cluster 5. TP, TC and TN showed exceedance above 0.10 for cluster 5. The most prominent parameters in terms of exceedance were FC and TUR. Cluster 5 showed exceedance in six parameters (i.e., FC, TUR, TP, TC, TN and DOC). Cluster 4 had exceedance in three parameters (i.e., FC, TUR and TN). Cluster 2 and cluster 3 showed exceedance only in FC and TUR.

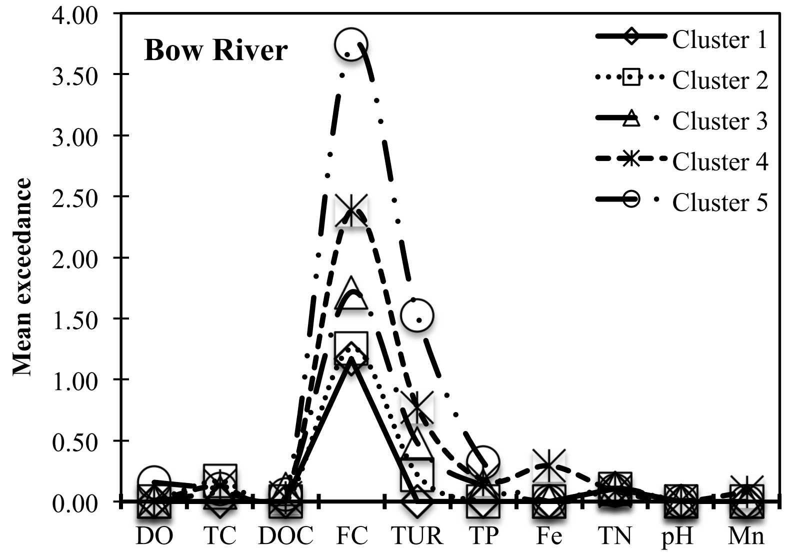

Figure 5.3: Mean exceedance of parameters in the clusters of Bow River.

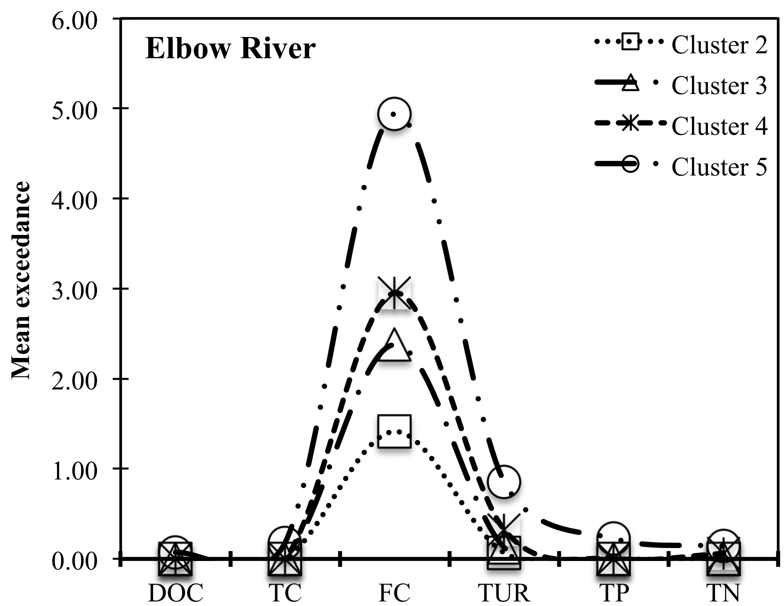

Figure 5.4: Mean exceedance of parameters in the clusters of Elbow River.

5.3.3.5 Milk River

The mean exceedance of parameters for the clusters of Milk River is given in **Fig. 5.5**. The exceedance values for the exceeded parameters are higher for cluster 5 as compared to cluster 4. The mean exceedance for: (i) FC and Fe was above 2, (ii) Mn and TUR was above 1.TP, DOC, TC, TN, and TDS had mean exceedance above 0.10 in cluster 5. The parameters in the order from the highest to lowest mean exceedance were Mn, FC, Fe, TUR, TDS and DO respectively in cluster 4. The mean exceedance for Mn, FC and Fe were above 1. The parameters of Cluster 5 for Milk River from the highest to lowest values were: FC, Fe, Mn, TUR, TP, DOC, TC, TN, TDS, pH and DO.

5.3.3.6 North Saskatchewan River

The mean exceedance of parameters for the clusters of North Saskatchewan River is given in **Fig. 5.6**. The lowest to highest exceedance for the parameters were observed from cluster 3 towards cluster 5.The mean exceedance for FC was above: (i) 1 for cluster 1, (ii) 2 for cluster 4, and (iii) 3 for cluster 5. The mean exceedance for TUR was above: (i) 2 for cluster 5, (ii) 0.5 for cluster 4, and (iii) 0.3 for cluster 3. The mean exceedance was above 0.1 for TP, TC, DOC and TN fin cluster 5. TP and TC had the exceedance above 0.1. TP in cluster 1 showed exceedance above 0.50 and TN had exceedance above 0.1. The mean exceedance was higher for FC, TUR, TP and TC as compared to other parameters for North Saskatchewan River.

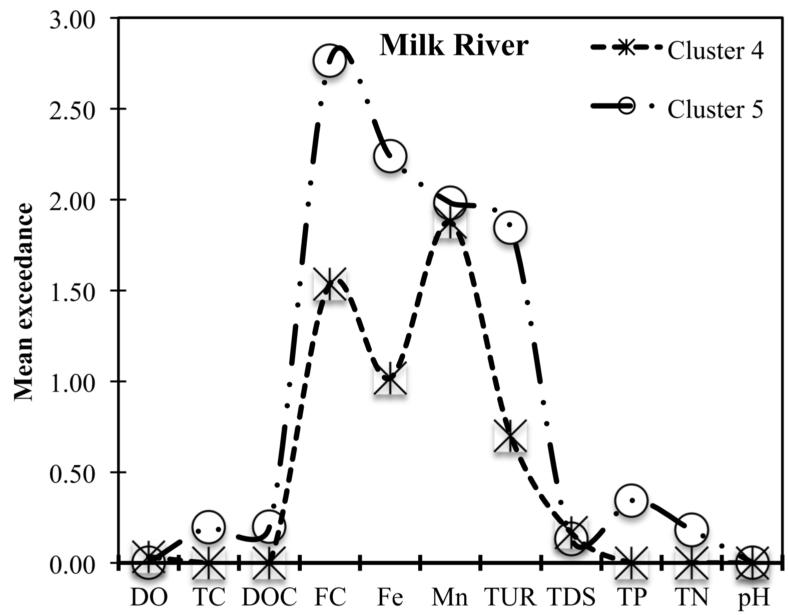

Figure 5.5: Mean exceedance of parameters in the clusters of Milk River.

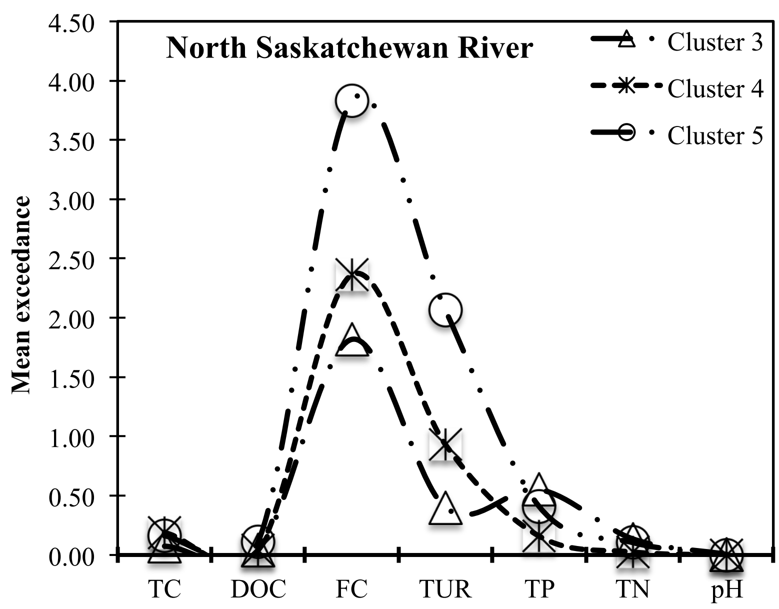

Figure 5.6: Mean exceedance of parameters in the clusters of North Saskatchewan River.

5.3.3.7 Oldman River

The mean exceedance of parameters for the clusters of North Saskatchewan River is given in **Fig. 5.7**. The lowest to highest exceedance for the parameters were observed from cluster 1 towards cluster 5.The mean exceedance for FC was above: (i) 1 for cluster 1, cluster 2 and cluster 3, (ii) 2 for cluster 4, and (iii) 4 for cluster 5. The mean exceedance for TUR was above: (i) 0.30 for cluster 2, (ii) 0.50 for cluster 3, (iii) 0.70 for cluster 4, and (iv) 2 for cluster 5. The mean exceedance was above (i) 0.40 for TP, and (ii) 0.10 for TN. The exceedance was above 0.10 for FE in cluster 4. The parameters with higher exceedance in: (i) cluster 5 was FC, TUR, TP, and TN, (ii) cluster 4 was FC, TUR and Fe, (iii) cluster 2 and cluster 3 were FC and TUR, and (iv) cluster 1 was FC.

5.3.3.8 Peace River

The mean exceedance of parameters for the clusters of Peace River is given in **Fig. 5.8**. The mean exceedance from the lowest to highest range was observed from cluster 3 towards cluster 5. The mean exceedance for TUR was above: (i) 0.50 for cluster 3, (ii) 1 for cluster 4, and (iii) 3 for cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 3, cluster 4 and (ii) 2 for cluster 5. For cluster 5, the exceedance for (i) TP was above 0.50, (ii) TC was above 0.20, (iii) DOC was above 0.10. The mean exceedance for TP was above 0.40 for cluster 4 and it was above 0.10 for TN. The parameters with the higher mean exceedance were TUR, FC, and TP in all three clusters.

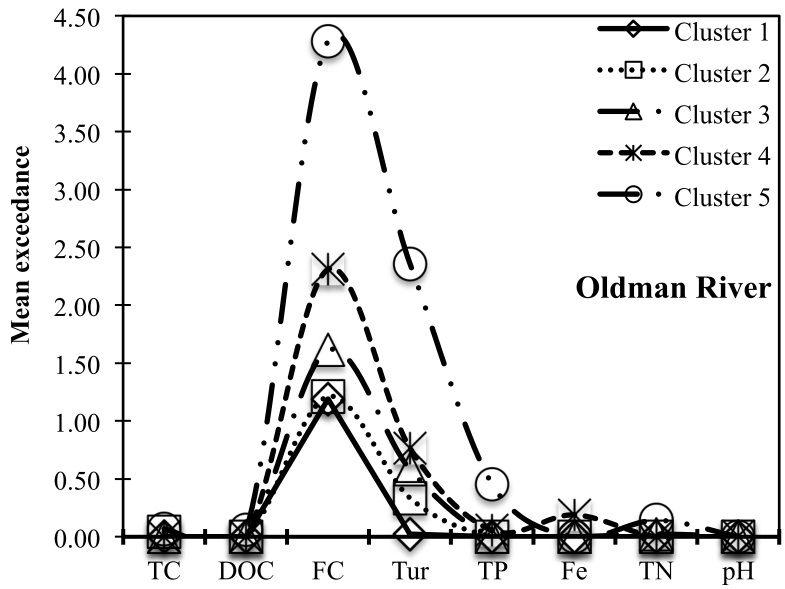

Figure 5.7: Mean exceedance of parameters in the clusters of Oldman River.

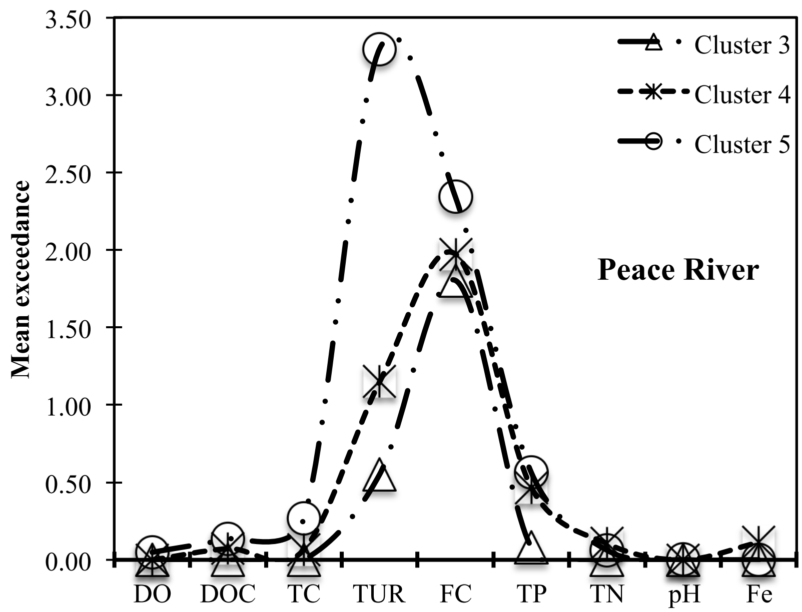

Figure 5.8: Mean exceedance of parameters in the clusters of Peace River.

5.3.3.9 Red Deer River

The mean exceedance of parameters for the clusters of Red Deer River is given in **Fig. 5.9**. The mean exceedance from the lowest to highest range was observed from cluster 1 towards cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 1, cluster 2 and cluster 3, and (ii) 4 for cluster 4. The mean exceedance for TUR was above: (i) 0.20 for cluster 2, cluster 3, (ii) 0.50 for cluster 4, (iii) 1 for cluster 5. For cluster 5: (i) TP was above 0.30, (ii) TC was above 0.20, (iii) DOC and TN was above 0.10. TP was above 0.10 for cluster 4. The parameters with the highest mean exceedance were: (i) FC, TUR, and TP in cluster 4 and cluster 5, (ii) FC, TUR, and TC in cluster 2 and cluster 3. FC was the only exceeded parameter in cluster 1.

5.3.3.10 Smoky River

The mean exceedance of parameters for the clusters of Smoky River shown in **Fig. 5.10**. The mean exceedance from the lowest to highest range was observed from cluster 3 towards cluster 5. The mean exceedance for TUR was above: (i) 0.50 for cluster 3, (ii) 1 for cluster 4, (iii) 2 for cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 3, (ii) 2 for cluster 4 and cluster 5. For cluster 5, the exceedance was above: (i) 0.40 for TP, (ii) 0.20 for TC, (iii) 0.10 for DOC. TC and DOC had exceedance above 0.10 in cluster 4. TP was above 0.40 in cluster 3 and TC was above 0.10 for cluster 3. The parameters with the highest mean exceedance were: (i) FC, TUR, and TP in cluster 3 and cluster 5, (ii) FC, TUR, and TC in cluster 2 and cluster 3. FC was the only parameter in cluster 1. The parameters with the highest mean exceedance were: (i) FC, TUR, and TP in cluster 3 and cluster 5, (ii) FC and TUR in cluster 2 and cluster 3.

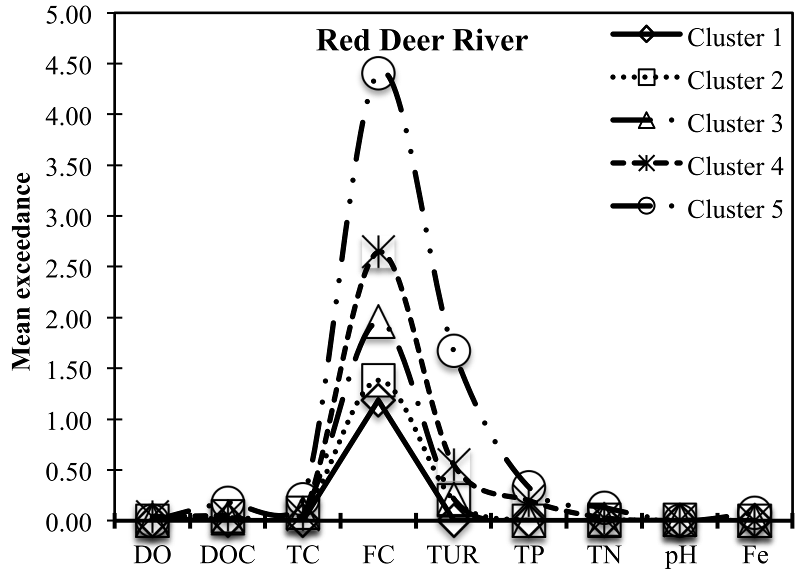

Figure 5.9: Mean exceedance of parameters in the clusters of Red Deer River.

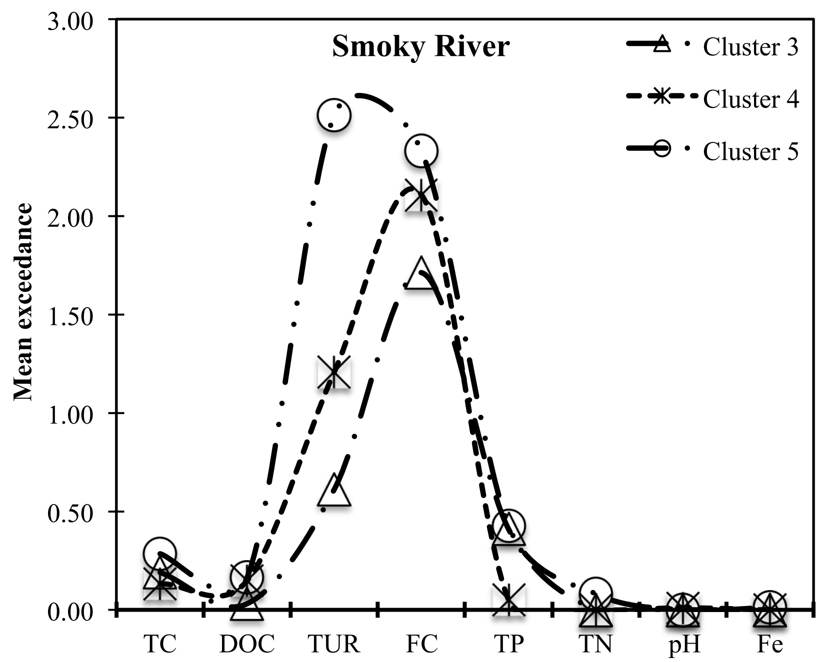

Figure 5.10: Mean exceedance of parameters in the clusters of Smoky River.

5.3.3.11 South Saskatchewan River

The mean exceedance of parameters for the clusters of South Saskatchewan River is shown in **Fig. 5.11**. The mean exceedance from the lowest to highest range was observed from cluster 2 towards cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 2, cluster 3, and cluster 4, (ii) 3 for cluster 5. The mean exceedance for TUR was above: (i) 0.30 for cluster 2, (ii) 0.70 for cluster 3, (iii) 1 for cluster 4 and cluster 5. The exceedance was above 0.30 for TP and it was more than 0.10 for DO and TN. For cluster 4, the mean exceedance for Fe was above 0.20 and it was above 0.10 for TN and TP. TN was also found above 0.10 for both cluster 2 and cluster 3. The parameters with the highest mean exceedance were: (i) FC, TUR, and TN in cluster 2 and cluster 3, (ii) FC, TUR and Fe in cluster 4, and (iii) FC, TUR and TP in cluster 5.

5.3.3.12 Wapiti River

The mean exceedance of parameters for the clusters of Smoky River is given in **Fig. 5.12**. The mean exceedance from the lowest to highest range was observed from cluster 1 towards cluster 5. The mean exceedance for FC was above: (i) 1 for cluster 3, (ii) 2 for cluster 4 and cluster 5. The mean exceedance for TUR was above: (i) 0.10 for cluster 1, (ii) 0.30 for cluster 3, (iii) 1 for cluster 4, and (iv) 2 for cluster 5. For cluster 5, the mean exceedance for: (i) TP was above 0.30, (ii) Mn was above 0.20, (iii) TC, TN, Fe and DOC was above 0.10. The mean exceedance for TP, TC and DOC was above 0.10 in cluster 4. The exceedance for TC and DOC was above 0.10 for cluster 3. The parameters with the highest mean exceedance were: (i) TUR in cluster 1, (ii) FC and TUR in cluster 2, (iii) FC, TUR and TC in cluster 4, and (iv) FC, TUR, and TC in cluster 5.

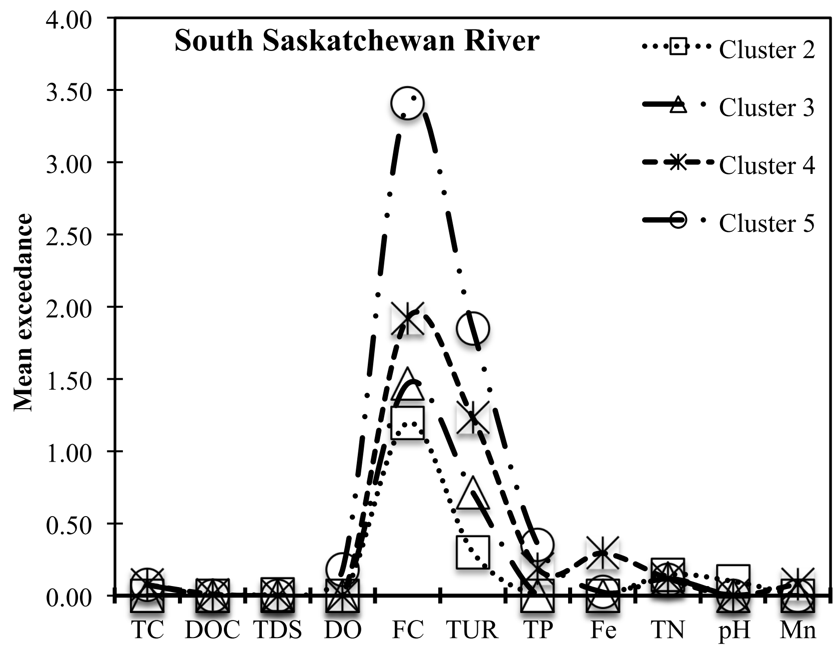

Figure 5.11: Mean exceedance of parameters in the clusters of South Saskatchewan River.

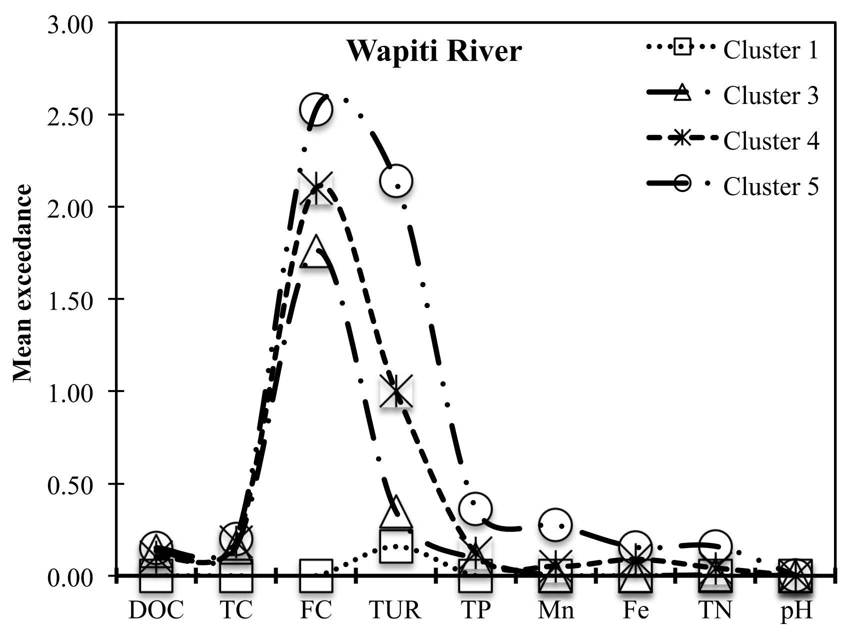

Figure 5.12: Mean exceedance of parameters in the clusters of Wapiti River.

5.4 Treatment technologies

We found that the parameters, which were exceeded for all the rivers, were TUR, FC, TC, Fe, Mn, TN, TP, TDS and DOC. On the basis of literature review, we have discussed the treatment technologies for these nine parameters in the subsequent sub sections:

5.4.1 Treatment technologies for turbidity

Turbidity is treated using the filtration technologies, which are: (i) chemically assisted filtration, (ii) slow sand filtration, (iii) diatomaceous earth filtration, and (iv) membrane filtration. Chemically assisted filtration includes chemical mixing, coagulation, flocculation, sedimentation and rapid gravity filtration. The coagulants are used for coagulation. Aluminum and ferric salts are the examples of coagulants. In coagulation, the particles are filtered out when water is passed through filters. With all these processes of filtration, turbidity level of 0.2-0.3 NTU can be achieved. The Pennsylvania Department of Environmental Protection evaluated 150 surface water treatment plants, which used filtration from 1988 to 1990 and found that turbidity level of 0.2 was successfully achieved by most of the plants (Consonery et al. 1991). Another study was conducted to test the treatment of low-turbidity surface water in Boston, it was found that the turbidity target of 0.1 NTU was obtained in more than 90% of the tests (Johnson et al. 1995). The turbidity of treated water was less than 1 NTU using slow sand filters (Fox et al. 1984). In another study, it was found that 50% of the 27 slow sand filter plants showed turbidity of 0.4 NTU or low and 15 % showed water with turbidity of 1 NTU or higher (Slezak and Sims 1984). Diatomaceous earth filtration was useful for treating waters with low turbidity. A study reported that the turbidity reduction was 56-78% using diatomaceous earth for the raw water having turbidity in the range from 0.95 to 2.5 NTU (Logsdon et al. 1981). For the same technology another study reported turbidity reduction to 75% with turbidity of 0.5 NTU (Pyper et al. 1985). Four membrane treatment processes can be applied depending upon the source water quality, treatment requirement and membrane pore size (Jacangelo 1991). Reverse osmosis can be used for salt removal, nanofiltration can be used for removal of cations (calcium and magnesium ions), ultrafiltration can be used for removal of dissolved organics and particulates, microfiltration can be used for removing particulates including pathogens (Jacangelo 1991; Taylor 1990).

5.4.2 Treatment technologies for fecal coliform

The disinfectants play important role for controlling bacteria, virus and other organisms. The most commonly used disinfectants are chlorine, chloramine, UV light, ozone, and chlorine dioxide. Chlorine the most widely used disinfectant kills bacteria and virus but ineffective for protozoans and organisms of biofilms. Whereas chloramine stays longer in the distribution system and hence more effective against coliforms (LeChevallier et al. 1990). UV light was found very affective disinfectant against different pathogens including protozoa (Wilczak et al. 1996). Ozone is more effective against all types of bacteria and viruses as compared to chlorine based disinfectant. A study reported that total coliforms were removed during the processes of pre-disinfection, clarification, and coagulation and the remaining coliforms were eliminated in filtration. The post-disinfection using chlorine or ozone had removed them completely (Payment et al. 1985). Different types of filtration systems are effective for removal of fecal coliforms. The examples of filtration technologies are conventional, direct, diatomaceous earth and slow sand. Direct filtration and diatomaceous earth filtration are good for high quality source waters. Slow sand filtration were found effective for removal of coliform bacteria and it was observed that total and faecal coliform removal was approximately 99% by using a biologically mature filter (Bellamy et al. 1985). The nanofiltration and reverse osmosis were useful in removal of bacteria and viruses (Taylor et al. 1990). Ultrafiltration (pore size 0.01 μm) and microfiltration (pore size 0.1 μm) are effective for partial removal of bacteria and viruses (Jacangelo et al. 1991).

5.4.3 Treatment technologies for true color

There can be different reasons for the color in the water. It can be due to: (i) coloured organic substances due to natural vegetation and soil runoff (Research Committee on Color Problems 1967), (ii) presence of iron or manganese which can be due to weathering, corrosion of distribution system, industrial wastes (American Water Works

Association 1971). In Canada, 90% of drinking water is obtained from surface water (Department of Fisheries and the Environment 1977). The color in the surface water is related to natural organic substances (Black and Christman 1963). Depending upon the source of color, water needs to be treated for that specific associated problem. For example dissolved air flotation is useful for treating water having high color due to particulate matter (U.S. NRC 1987). Direct filtration is applicable for water with maximum 40 color units (U.S. NRC 1987). High quality water with high color or where pre-treatment is done, slow sand filtration is beneficial (Adham et al. 1996). For example, the most suitable technology for the removal of organic contaminant carbon is Granular Activated Carbon (GAC). Different types of GAC are available for removing organics. The most commonly used carbon for the treatment of surface water in USA is coal-based carbon because of its hardness and adsorption capacity. For removal of color due to iron and manganese disinfectants like chlorine, chlorine dioxide, potassium permanganate and ozone can be used.

5.4.4 Treatment technologies for iron and manganese

There are different treatment methods that can be used for removal of iron and manganese from drinking water. These include: (i) oxidations using oxygen, chlorine, chlorine dioxide, potassium permanganate and ozone, (ii) ion exchange, (iii) lime softening, and (iv) sequestering chemicals (Crittenden et al. 2005). The process of oxidation with air is called aeration. In this process, DO is provided to water for converting Fe and Mn into $Fe(OH)_3$ and MnO_2. The oxidation of iron and manganese using chlorine or chlorine with potassium permanganate is the common method. After this oxidation, the water is processed for coagulation, clarification and filtration. The oxidation of Fe^{2+} and Mn^{2+} is faster with chlorine dioxide as it is a strong oxidant as compared to chlorine. The oxidation time is very fast (i.e., less than 20 seconds) if potassium permanganate ($KMnO_4$) is used. The cost of $KMnO_4$ is higher as compared to chlorine. If source water consists of both iron and manganese then iron is oxidized first using chlorine and then $KMnO_4$ is used to oxidize manganese. The oxidation process may not be effective for removal of Fe^{2+} in case of higher natural organic matter. In such situation the process, which are used for removal of NOM, can be effective for the removal of Fe. Ozone is effective method but costly as compared to other oxidation methods. This method is common in Europe for removal of Fe and Mn. Nanofiltration membranes are effective for removal of Fe and Mn (Kartinen and Martin 1995).

5.4.5 Treatment technologies for nitrate and nitrite

The nitrates and nitrites can be removed from drinking source waters using treatment technologies of (i) chemical denitrification, (ii) biological denitrification, (iii) reverse osmosis, (iv) electrodialysis, and (v) ion exchange (Crittenden et al. 2005; WHO 1992; Department of National Health and Welfare 1993). In chemical denitrification, nitrate is reduced to nitrogen gas using iron and aluminum. This process is very costly (Murphy 1991). In biological denitrification, nitrate is converted into nitrogen gas in anoxic conditions. Reverse osmosis can be used to reduce nitrate levels in the source water. Reverse osmosis is expensive to treat only nitrates and it can only be cost effective for nitrate removal if there it is used for other issues like TDS concentrations (Crittenden et al. 2005). In electrodialysis, electric current is passed through semi-permeable membranes to remove nitrates (Crittenden et al. 2005). Ion exchange is an efficient way for the removal of nitrate. It is more cost effective as compared to reverse osmosis.

5.4.6 Treatment technologies for phosphate

The adsorption technique can be used for removal of phosphate from drinking water. A study demonstrated that the removal efficiency for total phosphorus using the conventional treatment process was between 66% to 69% (Jiang et al. 2012). The efficiency of TP removal improved with the processes of coagulation-sedimentation and filtration. The removal of TP can be further improved by enhanced coagulation (Jiang et al. 2012). The processes of electrodialysis and reverse osmosis could also be used for removal of phosphate but these technologies are expensive and the removal efficiency was up to 10% (Yeoman et al. 1988). Another study mentioned the use of electrocoagulation process for the removal of phosphate from drinking water and the removal efficiency was 98% (Vasudevan et al. 2008).

5.4.7 Treatment technologies for total dissolved solids

TDS cannot be removed using the conventional water treatment processes. The addition of chemicals during conventional water treatment could increase the TDS (Department of National Health and Welfare 1993). Demineralization is required for TDS removal but it can be very expensive (Canadian Council of Resource and Environment Ministers 1987). The most economical technologies for removing TDS are reverse osmosis and electrodialysis (Clark 1977).

5.4.8 Treatment technologies for dissolved organic carbon

The natural organic matter (NOM) is measured as total organic carbon or dissolved organic carbon. It can be removed by enhanced coagulation, adsorption, ion exchange and reverse osmosis. Enhanced coagulation is performed at higher coagulation doses and low pH values. It is an economical treatment technology (Crozes et al. 1995). GAC adsorption and post filtration are useful in removing NOM. The limitation of ion exchange method is production of high TDS. RO is expensive and other limitation is concentrate disposal issue.

5.5 Conclusions

In this study, we developed a mean exceedance model to obtain the exceedance patterns of parameters in the clusters for 12 major rivers in Alberta using the data of 17 parameters for 23 sampling sites during the period 2004-2008. The parameters with higher mean exceedance were: (i) FC and TUR for Athabasca River, (ii) FC, TUR, Mn, DO and TP for Battle River, (iii) FC, TUR and TP for Bow River, (iv) FC and TUR for Elbow River, (v) FC, Fe, Mn and TUR for Milk River, (vi) FC, TUR, TP and TC for North Saskatchewan River, (vii) FC, TUR, TP, TN, and Fe for Oldman River, (viii) TUR, FC, and TP for Peace River, (ix) FC, TUR, TP and TC for Red Deer River, (x) FC, TUR, and TP for Smoky River, (xi) FC, TUR, TN and TP for South Saskatchewan River, and (xii) FC, TUR and TC for Wapiti River. The mean exceedance was highest for FC and TUR in all the rivers and these were also dominant parameters (Akbar et al. 2013). The exceedance in FC, TUR, TP, TN, TC, and DO was related to anthropogenic activities of land cover/uses. The exceedance in Fe and Mn was due to natural mineralization (Akbar et al. 2013; Anderson 1999; Bruneau et al. 2009; Sosiak and Dixon 2006). The higher trend of mean exceedance for TUR was related to snow melting (Akbar et al. 2013). The exceedance modeling was useful in: (i) identifying the parameters exceeded above the water quality guidelines in each cluster, (ii) obtaining the patterns of exceedance for the exceeded parameters, (iii) obtaining the exceedance level for each parameter of a cluster, (iv) targeting the parameters for specific treatment on the basis of exceedance level, (v) identifying the source of water pollution, and (vi) deciding the targeted treatment technology.

CHAPTER 6

DEVELOPMENT OF REMOTE SENSING BASED MODELS FOR SURFACE WATER QUALITY

Akbar T. A., Hassan Q. K., Achari G., Development of Remote Sensing Based Models for Surface Water Quality, *CLEAN-Soil, Air, Water* **2013**. DOI: 10.1002/clen.201300001 (In Press).

Abstract

The objectives of this paper were to develop, evaluate and apply the remote sensing based models for Canadian Water Quality Index (CWQI) and turbidity for the Bow River of Alberta. We used 31 scenes of Landsat-5 TM satellite data to establish the relationship between the planetary reflectance and the monthly ground measured data for the period of five years (i.e., 2006-2010). The four spectral bands (i.e., blue, green, red and near infrared) were used to obtain the most suitable models from 26 different band combinations. The co-efficients of determination (r^2) on the basis of red band were 0.91 for the CWQI model and 0.82 for the turbidity model. The best-fit models were validated with ground measured data and found that: 72% of the data showed 100% matching for the CWQI model and 83% of the data for the turbidity model. The Landsat-5 TM based CWQI and turbidity models were applied on all the scenes to obtain five CWQI classes (i.e., excellent, good, fair, marginal and poor), and six classes of turbidity (NTU) (i.e., 0 to 10, 10 to 20, 20 to 30, 30 to 40, 40 to 50, and >50). On the basis of percentages obtained for CWQI and turbidity classes, the ranks of years in terms of water quality from best to worst were: 2009, 2006, 2008, 2010, and 2007, respectively. The variation of river water quality in different years of interest was associated with the climatic changes. The most deteriorated water quality noted in two natural subregions included Mixed grass and Dry mixed grass which could be related to irrigation-based farming.

Keywords: Bow River; Canadian Water Quality Index; Landsat-5 TM; Remote sensing modeling; Turbidity

6.1 Introduction

Water extends approximately 71% of earth's surface and it is also imperative for the existence and sustainability of living organism on the earth surface (UNDESA 2005). The freshwater is just 2.5% of the earth's water. About 0.3% of freshwater is found in rivers, lakes and atmosphere (Gleick 1993). In general, the understanding of the water quality plays a critical role prior to utilize for various purposes including drinking (Environment Canada 2012). In this paper, we opted to understand the surface water quality for the Bow River, which is a major river in the Canadian province of Alberta having a total length of 587 km, and a main source of drinking water for many communities of the province (Telang 1990).

The surface water quality of the Bow River is measured every month at three fixed sampling sites (i.e., Carseland, Cluny and Ronalane) for different water quality variables using the traditional methods. In general, these methods provide accurate measurements, however, these may not be feasible means to sample the entire river due to the huge involvement of labour and cost. Currently, the measured data of water quality variables at the sampling sites of the Bow River are grouped into five classes (i.e., excellent, good, fair, marginal and poor) using the framework of Canadian Water Quality Index (CWQI: see details in in section 2.5) (CCME 2001). These classes are obtained on the basis of fixed-point locations, which does not represent the spatial dynamics of the entire river.

In another study, we classified the surface water quality of major rivers of Alberta on the basis of clusters. We observed higher (deteriorated water quality) clusters (i.e., 4 and 5) for the rivers during the growing season (April 1 – September 30) as compared to lower clusters (i.e. 1, 2, and 3) in winter months (Oct 1 – March 31). During the growing season, the snowmelt wash various materials from the land surface into the rivers due to anthropogenic activities related to different types of land use/cover. Turbidity was found to be a dominant parameter associated with the deterioration in water quality during the growing season (Akbar et al. 2013). On this basis, we considered turbidity separately besides CWQI in this study. For the Bow River, the turbidity is measured at fixed sampling location, which does not represent the mean turbidity for the whole water body (Moreno-Madrinan et al. 2010).

In order to address the spatial variability in water quality real time data, remote sensing-based methods were found to be alternative and efficient ones (Sládeček 2006; Olmanson et al. 2013; Stisen et al. 2008). The remote sensing methods are suitable to analyze: (i) spatial variability over a large geographic area, (ii) temporal trends over certain periods of interests, and (iii) the conditions of the water bodies in remote areas. In remote sensing, optical remote sensors are used for monitoring the water quality-related variables. The most commonly used sensors include the use of Landsat-7 ETM (Alparslan et al. 2007; Bustamante et al. 2009), Landsat-5 TM (Nas et al. 2010; Hellweger et al. 2004), MODIS (Wu et al. 2009), NOAA AVHRR (Bolgrien et al. 1995), and SPOT HVR (Dekker et al. 2002) among others. In most of the instances, the spectral bands used in these studies included blue, green, red and near infrared (Alparslan et al. 2007; Bustamante et al. 2009; Nas et al. 2010; Hellweger et al. 2004; Wu et al. 2009; Bolgrien et al. 1995; Dekker et al. 2002). The observed planetary reflectance from these bands was used to study water quality variables including suspended sediment, turbidity, Secchi disk depth, and chlorophyll-*a* (Bustamante et al. 2009; Nas et al. 2010; Oyama et al. 2009; Wang et al. 2006).

In another study we classified and analyzed the surface water quality for 12 major rivers of Alberta. We developed a surface water quality classification system using principal component analysis (PCA), total exceedance model and clustering technique. From PCA, we identified seven major principal components (PCs), which were the indicators of watershed geology, mineralization and anthropogenic activities related to land use/cover. The PCs were used to identify the dominant parameters. The normalized data of dominant parameters were used to develop a total exceedance model. The exceedance values were used to determine the patterns for the development of five clusters. The water quality deteriorates as the cluster number increased from cluster 1 to cluster 5. The clusters showed reasonably strong agreements (i.e., 80-90%) against the classes of CWQI. The dominant clusters during the growing and winter seasons were used for the spatial and temporal patterns of the surface water quality of rivers (Akbar et al. 2013).

In the present study, we have tested remote sensing-based methods for acquiring CWQI and turbidity classes for assessing both spatial and temporal dynamics of the Bow River. The specific objectives of this paper are to: (i) develop and evaluate remote sensing based models to acquire CWQI classes using the planetary reflectance of Landsat-5 TM and ground measured data, (ii) develop and evaluate remote sensing based models to retrieve turbidity using the planetary reflectance of Landsat-5 TM and *in-situ* data, (iii) apply the selected models to classify the source waters of the Bow River into CWQI and turbidity classes for spatial and temporal analysis, and (iv) study the impact of natural subregions on Bow River water quality.

6.2 Materials and methods

6.2.1 Study area

The Bow River originates from Bow Glacier located on the north of Lake Louise in Alberta. It flows in southeastern direction and merges with Oldman River to make South Saskatchewan River. The Bow River is surrounded by three natural regions including Grassland, Parkland, and Rocky Mountain. These regions are classified into six natural subregions (i.e., Dry mixed grass, Mixed grass, Foothills fescue, Foothills parkland, Montane, and Sub-Alpine) as shown in **Fig. 6.1**. For these subregions the range for: (i) the estimated length of the Bow River flowing through each natural subregion is from 39 km to 171 km, (ii) the mean annual temperature is from -0.1°C to 4.4 °C, and (iii) the mean annual precipitation is from 333 mm to 755 mm as given in **Table 6.1**. The main vegetation type for each of the subregions is also mentioned in **Table 6.1**.The drainage area for the Bow River is 25000 km^2 (Telang 1990). The major municipality along the river is the City of Calgary (i.e., having a population of 1,096,833 according to 2011 census), which receives drinking water from this river (Statistics Canada 2012). The mean annual flow of the river near Calgary is 91.1 m^3/s (Seneka 2004). The water flow is controlled by two dams (Bearspaw Dam and Ghost Dam) constructed on the Bow River for the supply of electricity to the City of Calgary (Jeffries et al. 2008). The surface water of Bow River allocated for various purposes include: (i) 71% for irrigation, (ii) 18% for municipal, (iii) 4% for water management,

(iv) 2% for management of wildlife, (v) 2% for dewatering, and (vi) 2% for commercial (Bennett and Murray 2010).

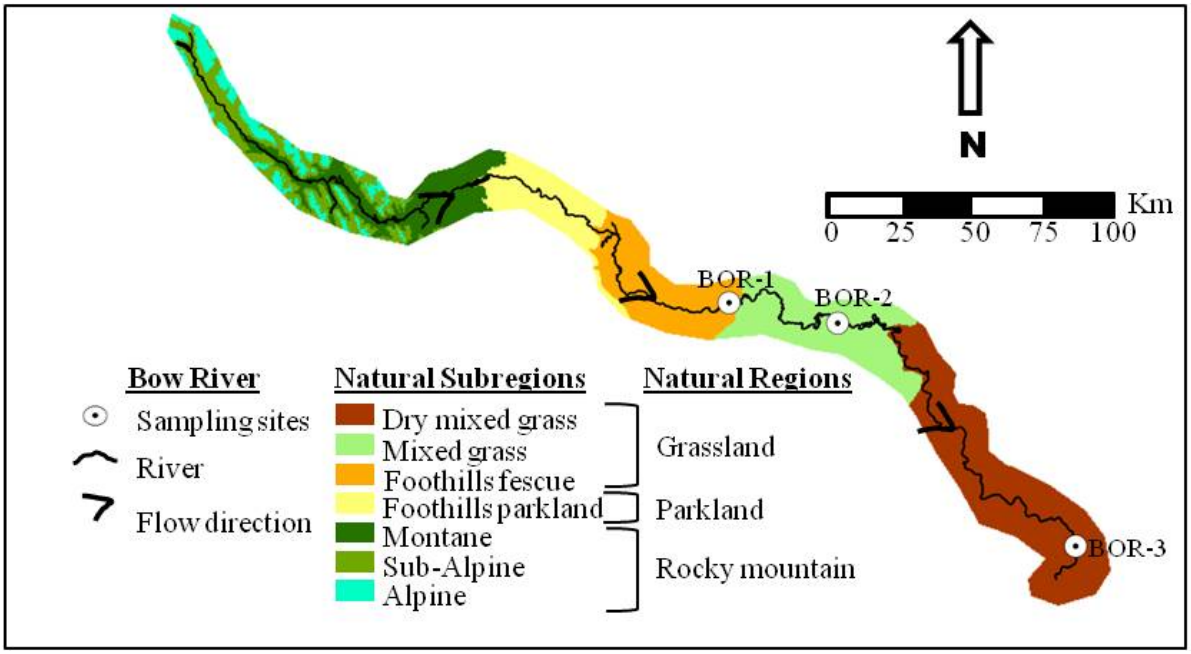

Figure 6.1: Natural subregions for the Bow River of Alberta.

Table 6.1: Characteristics of natural subregions for Bow River.

Natural subregions	Bow River length (km)	Mean annual temperature (°C)	Mean annual precipitation (mm)	Main vegetation
Dry mixed grass	171	4.2	333	Grasslands and shrublands
Mixed grass	107	4.4	394	Agriculture and native grassland
Foothills fescue	87	3.9	470	Mountain perennial and wheat grasses
Foothills parkland	63	3.0	517	Aspen forests and grasslands
Montane	120	2.3	589	Aspen, pine, fir and spruce forests and grasslands
Sub-Alpine	39	-0.1	755	Mixed conifer forests

6.2.2 *Satellite and water quality data*

We used 31 scenes of Landsat-5 TM multispectral image data for the different dates during the period 2006-2010 as listed in **Table 6.2**. The size of each scene was 185 km x 172 km. The spectral bands which were used in this study were (i) blue, (ii) green, (iii) red, and (iv) near infrared. The spatial resolution for each of these spectral bands was 30 m. The raw satellite data was downloaded from the United States Geological Survey Global Visualization Viewer (USGS GloVis) in GeoTIFF format with the Level 1T correction (USGS 2012). The Level 1T is the standard terrain correction in which systematic radiometric and geometric accuracy is provided using the ground control points and the topographic accuracy is obtained by using the Digital Elevation Model (DEM) (NASA 2012). The scenes of Landsat-5 TM were selected on the basis of the least cloud cover, least snow, and closeness to the sampling days. In total, we used ground measured data for 37 days at three sampling locations of the Bow River in 2006-2010 to develop and validate models using the planetary reflectance of 31 scenes of Landsat TM-5. The sampling locations and the dates for the ground water quality data and Landsat-5 TM scenes are given in **Table 6.2**. The water quality data were obtained from Alberta Environment and listed in **Table 6.3** with Alberta River Water Quality Index objectives (Alberta Environment and Sustainable Resource Development 2011; Alberta Environment 1999; Health Canada 2010; Ontario Drinking Water Standards 2006).

Table 6.2: Data used for development and evaluation of models for CWQI and Turbidity during 2006-2010.

River water sampling			Landsat-5 TM scene			
No.	Site	Date	No.	Acquisition date	Path	Row
1	Carseland	26-Apr-06				
2	Cluny	25-Apr-06	1	24-Apr-06	41	25
3	Carseland	23-May-06	2	17-May-06	42	24
4	Carseland	12-Jul-06				
5	Cluny	12-Jul-06	3	13-Jul-06	41	25
6	Ronalane	25-Jul-06	4	22-Jul-06	40	25
7	Carseland	31-Aug-06				
8	Cluny	31-Aug-06	5	30-Aug-06	41	25
9	Ronalane	19-Sep-06	6	24-Sep-06	40	25
10	Carseland	23-Nov-06				
11	Cluny	23-Nov-06	7	18-Nov-06	41	25
12	Carseland	19-Jun-07	8	21-Jun-07	42	24
13	Ronalane	25-Jun-07	9	23-Jun-07	40	25
14	Cluny	26-Jun-07	10	30-Jun-07	41	25
15	Carseland	23-Jul-07	11	23-Jul-07	42	24
16	Carseland	28-Apr-08	12	29-Apr-08	41	25
17	Carseland	14-May-08				
18	Cluny	14-May-08	13	15-May-08	41	25
19	Carseland	17-Jun-08	14	23-Jun-08	42	24
20	Ronalane	16-Jul-08	15	11-Jul-08	40	25
21	Carseland	15-Jul-08	16	18-Jul-08	41	25
22	Ronalane	21-Aug-08	17	28-Aug-08	40	25
23	Ronalane	14-Oct-08	18	15-Oct-08	40	25
24	Carseland	26-May-09	19	25-May-09	42	24
25	Ronalane	25-May-09	20	27-May-09	40	25
26	Carseland	15-Jun-09	21	10-Jun-09	42	24
27	Carseland	20-Jul-09	22	21-Jul-09	41	25
28	Carseland	17-Aug-09				
29	Cluny	17-Aug-09	23	22-Aug-09	41	25
30	Cluny	15-Sep-09	24	7-Sep-09	41	25
31	Carseland	14-Sep-09	25	14-Sep-09	42	24
32	Carseland	19-Oct-09	26	25-Oct-09	41	25
33	Cluny	21-Apr-10	27	19-Apr-10	41	25
34	Carseland	19-Apr-10	28	26-Apr-10	42	24
35	Carseland	10-May-10	29	12-May-10	42	24
36	Carseland	15-Jul-10	30	15-Jul-10	42	24
37	Ronalane	18-Aug-10	31	18-Aug-10	40	25

Table 6.3: Alberta River Water Quality Index objectives for 17 variables (Alberta Environment and Sustainable Resource Development 2011; Alberta Environment 1999; Health Canada 2010; Ontario Drinking Water Standards 2006).

Variable	Objective	Variable	Objective	Variable	objective
WT	15°C	TP	0.05 mg/ L	Na	200 mg/L
DO	6.5 mg/L	TN	1 mg/ L	F	1.5 mg/L
TUR	5 NTU	pH	6.5 and 8.5	FC	100 /100 ml
TC	15 Pt Co units	TH	500 mg/ L	Mn	0.05 mg/L
DOC	5 mg/L	CI	250 mg/ L	Fe	0.3 mg/L
TDS	500 mg/L	SO_4	500 mg/ L		

6.2.3 Image processing

The satellite scenes were processed to make them workable for the purpose of this research. The operations applied for processing are briefly explained in the following sub-sections:

6.2.3.1 Conversion of digital numbers into spectral radiance

In the first step, we converted the raw digital numbers of all the Landsat-5 TM images into spectral radiance using Eq. 6.1 (NASA 2012) as follows:

$$L_\lambda = \left[\left(\frac{L_{max\lambda} - L_{min\lambda}}{DN_{max} - DN_{min}}\right) \times \left(DN - DN_{min}\right)\right] + L_{min\lambda} \qquad (6.1)$$

where,

L_λ = spectral Radiance at the sensor's aperture in watts/m^2/ster/µm,

DN = quantized calibrated pixel value,

$L_{min\lambda}$ = spectral radiance that is scaled to QCALMIN in watts/m^2/ster/µm,

$L_{max\lambda}$ = spectral radiance that is scaled to QCALMAX in watts/m^2/ster/µm,

DN_{min} = minimum quantized calibrated pixel value (corresponding to $L_{min\,\lambda}$) in DN, and

DN_{max} = maximum quantized calibrated pixel value (corresponding to $L_{max\lambda}$) in DN.

The values of $L_{min\lambda}$, $L_{max\lambda}$, DN_{min} and DN_{max} in Eq. 6.1 were obtained from the metadata files.

6.2.3.2 Conversion from spectral radiance to planetary reflectance

In the second step, the spectral radiance were converted into planetary reflectance using Eq. 6.2 (NASA 2012) as follows:

$$\rho_\rho = \left[\frac{\pi \times L\lambda \times d^2}{ESUN\lambda \times \cos\theta s} \right] \tag{6.2}$$

where,

ρ_ρ = unitless planetary reflectance,

π = 3.141592654,

L_λ = spectral radiance at the sensor's aperture,

d = earth-sun distance in astronomical units,

$ESUN_\lambda$ = mean solar exo-atmospheric irradiance, and

Θ_s = solar zenith angle in degrees

The value of: (a) d was obtained from Science Data Users Handbook (NASA 2011), (b) $ESUN_\lambda$ for all bands of TM sensors were obtained from Chander and Markham (Chander and Markham 2003), (c) Θ_s was obtained from the formula (i.e., Θ_s = 90° - sun elevation angle), where the sun elevation angle was obtained from the metadata file of each satellite images.

6.2.3.3 Normalized Difference Vegetation Index (NDVI)

Finally, we calculated normalized difference vegetation index (NDVI: a measure of vegetation greenness) using Eq. 6.3 (Tucker 1979) as follows:

$$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \tag{6.3}$$

where,

ρ_{NIR} = reflectance of near infrared band, and

ρ_R = reflectance of red band.

Such NDVI calculations were performed over the sampling sites (BOR-1, BOR-2, BOR-3) in all the scenes of 37 data records in order to determine the possible contamination of sampling site pixels from other landuses (e.g. roads, agriculture, vegetation, and barren land, etc).The negative NDVI values (i.e. between 0 to -1) indicated the presence of water in the pixels whereas positive NDVI values showed the possible contamination due to other landuses (Weier and Herring 1999). In case of a positive NDVI value for any sampling site pixel, we considered the reflectance value of a nearest neighboring water pixel.

6.2.3.4 Canadian Water Quality Index (CWQI)

Using Eq. 6.4, we calculated CWQI for all three sampling sites of Bow River during the period 2006-2010 using the measured data of seventeen variables on the basis of Alberta River Water Quality Index objectives as given in **Table 6.3** (CCME 2001):

$$CWQI = 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \right) \tag{6.4}$$

where, F_1, F_2, F_3 are scope, frequency and amplitude, respectively. The equations to calculate these three factors are given in **Table 6.4**.The quantitative values (i.e. 0 to 100) obtained from Eq. 6.4 were divided into five classes: (i) 95 to 100 = 1 (excellent), (ii) 80 to 94 = 2 (good), (iii) 60 to 79 = 3 (fair), (iv) 45 to 59 = 4 (marginal), and (v) 0 to 44 = 5 (poor) (CCME 2001). The CWQI classes were produced for 37 data records as given in **Table 6.2**.

Table 6.4: Equations used for calculation of CWQI and identifying classes using the data.

$F_1\,(\text{Scope}) = \left(\dfrac{\text{Number of failed parameters}}{\text{Total number of parameters}}\right) \times 100$	$F_2\,(\text{Frequency}) = \left(\dfrac{\text{Number of failed tests}}{\text{Total number of tests}}\right) \times 100$
$\text{excursion}_i = \left(\dfrac{\text{Objective}_j}{\text{Failed test value}_i}\right) - 1$	$\text{excursion}_i = \left(\dfrac{\text{Failed test value}_i}{\text{Objective}_j}\right) - 1$
$\text{nse} = \left(\dfrac{\sum_{i=1}^{n} \text{excursion}_i}{\text{Number of tests}}\right)$	$F_3\,(\text{amplitude}) = \left(\dfrac{\text{nse}}{0.01\text{nse} + 0.01}\right)$

6.2.3.5 Models for CWQI and turbidity from planetary reflectance of Landsat-5 TM data

On the basis of literature review for the relationship of bands with the variables of water quality, we developed 26 individual empirical models in determining both CWQI and turbidity as a function of the spectral bands of B, G, R and NIR (Stisen et al. 2008; Alparslan et al. 2007; Bustamante et al. 2009; Nas et al. 2010; Hellweger et al. 2004; Wu et al. 2009; Bolgrien et al. 1995; Dekker et al. 2002; Oyama et al. 2009). The specific inputs of these models were: B, G, R, NIR, G/B, B/R, R/B, NIR/B, R/G, NIR/G, B+G, B+R, B+NIR, G+R, G+NIR, R+NIR, B+G+R, B+G+NIR, G+R+NIR, B+G+R+NIR, (B/NIR)+G, (B/NIR)+B, (B/R)+R, (B/R)+G, (B/R)+B, and (NIR/B)+NIR. We used regression analysis technique to obtain the quantitative relationship beween the satellite based planetary reflectances and water quality variables to develop the empirical models (Sládeček 2006; Olmanson et al. 2013; Stisen et al. 2008; Alparslan et al. 2007; Bustamante et al. 2009; Nas et al. 2010; Hellweger et al. 2004; Wu et al. 2009; Bolgrien et al. 1995; Dekker et al. 2002; Oyama et al. 2009). We used 23 data records (i.e., Landsat-5 TM as well as the ground data) for the development of models to obtain CWQI classes and turbidity from the planetary reflectance. The remaining 14 data records were used to validate the selected best models. In all these models, CWQI and turbidity were the dependent variables whereas the bands were the independent variables. CWQI and turbidity were plotted on the vertical axis and bands on the horizontal axis. Each model represents: (i) a unique band formula, (ii) an intercept, and (ii) a slope. The intercept of dependable variables is the distance from the origin to the point where the line crosses the vertical axis. Slope is the amount of change for dependent variables corresponding to one-unit increase in band formula. The intercept and slope were calculated using the least square regression method to obtain a line of best fit. The strength of correlations between the planetary reflectance and *in-situ* turbidity and CWQI for the development and validation of the models were obtained based on co-efficient of determination (r^2) (CliffsNotes 2013). On the basis of r^2 values, we identified the significant empirical models for CWQI and turbidity.

6.2.3.6 Spatial and temporal analysis for the Bow River

We subset all the scenes of interest to extract the Bow River. The NDVI was calculated for the Bow River to extract the pixels with water and remove the pixels contaminated with other land use types. We selected the most suitable empirical models for CWQI and turbidity on the basis of r^2 values. The selected CWQI model was applied on 31scenes of Landsat-5 TM to obtain the spatial distribution for five classes (i.e., as described in section 2.4) along the Bow River. Similarly the selected turbidity model was applied on these scenes to obtain the spatial distribution of turbidity. The turbidity values were divided into six classes which are: (i) 0 to 10 NTU, (ii) 10 to 20 NTU, (ii) 20 to 30 NTU, (iv) 30 to 40 NTU, (v) 40 to 50 NTU and (vi) >50 NTU (Cox et al. 1998). CWQI classes could vary from 1 to 5 with the increase in the concentrations of variables. For example lower turbidity classes might indicate lower cluster numbers whereas higher turbidity classes could represent higher cluster numbers. For both turbidity and CWQI, the percentage accumulated by each class was obtained by dividing the pixels of each class by the total pixels of all classes during each year of interest during 2006-2010. Finally we overlaid maps for CWQI classes for the selected period on the natural subregions (i.e., **Fig. 6.1**).

6.3 Results and discussion

6.3.1 Empirical models for determining CWQI classes

We developed 26 empirical models for determining CWQI classes as given in **Table 6.5**. These empirical models could be used to obtain the spatial distribution of CWQI classes using the planetary reflectance of bands for any periods. The slopes in CWQI were the constant numbers, which were multiplied with the band formulae (e.g., 28.072 in model 1 of **Table 6.5**). The intercepts are the positive or negative number in each model (e.g., + 0.5785 in model 1 of **Table 6.5**). Please see other slopes and intercepts for model no. 2 to model no. 26 in **Table 6.5**. The range of r^2 for all empirical models was from 0.01 to 0.91. The correlation coefficients less than 0.50 were considered weak due to which we regarded only the models with $r^2 > 0.50$ (model no.1 to model no.14 in **Table 6.5**) as significant (Roberts and Roberts 2013). Among these significant models, r^2 was higher (i.e., 0.73 to 0.91) for the models with red band (e.g., model no. 1 to model no. 8 in **Table 6.5**) whereas it was lower (i.e., 0.54 to 0.72) in the models without red band (e.g., model no. 10 to model no. 14 in **Table 6.5**). Some of the models (i.e., model 20 to model 26) had very low values of r^2 (i.e., in the range 0.01-0.17) as shown in **Table 6.5**. These models exhibited weak relationship between planetary reflectance and CWQI calculated on the basis of ground-measured data. The range of r^2 was from 0.30 to 0.36 for model 15 to model 19 as obvious from **Table 6.5**, which was better as compared to model 20 to model 26 but the models were still not significant for the application purpose. The best model was the use of the spectral band R (i.e., $r^2 = 0.91$, see model no. 1 in **Table 6.5**). The scatter plot and deviation plot of this model are shown in **Fig. 6.2 (a)** and **(b)**. The figure shows that 10 data records matched 100% of the modeled values, whereas 4 data records had a deviation of 1 from the modeled values (**Fig. 6.2 (b)**). The result of this validation indicates the usefulness of this model for obtaining CWQI from the reflectance of the red band.

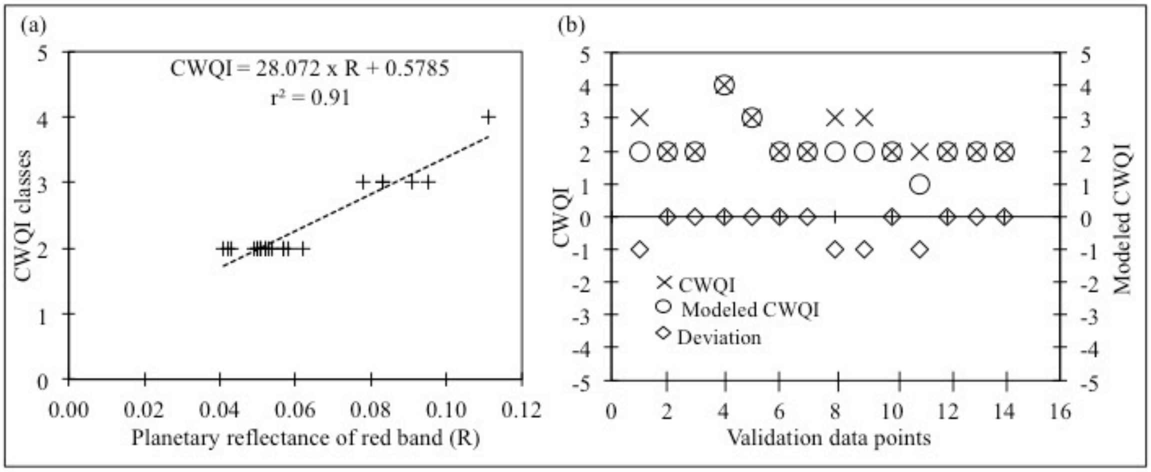

Figure 6.2: (a) Development, and (b) evaluation of most suitable model for obtaining CWQI classes using the planetary reflectance of red band for the Bow River.

Most of the previous studies showed the development of remote sensing based models for individual water quality variables (e.g., Sládeček 2006; Olmanson et al. 2013; Stisen et al. 2008; Alparslan et al. 2007; Bustamante et al. 2009; Nas et al. 2010; Hellweger et al. 2004; Wu et al. 2009; Bolgrien et al. 1995; Dekker et al. 2002; Oyama et al. 2009). A limited number of studies showed the application of remote sensing for the development of indices (Chen et al. 2005; Vignolo et al. 2006). Composite pollution index (CPI) was developed using band 1 (0.402-0.422 μm), band 2 (0.433-0.453 μm), band 3 (0.480-0.500 μm) and band 4 (0.500-0.520 μm) to obtain the five classes of water quality. The co-efficient of determination for CPI was 0.93 (Chen et al. 2005). In another study, remote sensing based water quality index was developed on the basis of blue and green bands. The r^2 was 0.82 for this water quality index (Vignolo et al. 2006). In our study, we found that the models with: (i) blue and green bands (e.g., model 13 in **Table 6.5**), (ii) green band (e.g., model 10 in **Table 6.5**), (iii) blue band (e.g., model 19 in **Table 6.5**) showed r^2 of 0.57, 0.71 and 0.30 respectively. It was also noticed that the models having blue and green band with: (i) red band (e.g. model 7 in **Table 6.5**), and (ii) red and near infrared bands (e.g. model 6 in **Table 6.5**) showed higher values for r^2 (i.e., 0.75, and 0.76 respectively).

Table 6.5: Models developed for mapping spatial distribution of CWQI classes for the Bow River using the first four spectral bands (i.e. blue, green, red and near-infrared) of Landsat-5 TM satellite data.

Model no.	Models	r^2	Model no.	Models	r^2
1	28.072 x R + 0.5785	0.91	14	15.044 x (B + NIR) + 0.2322	0.54
2	14.816 x (G + R) + 0.1715	0.84	15	-0.861 x (B/R) + 3.6939	0.36
3	16.031 x (R + NIR) + 0.6253	0.82	16	-0.8493 x [(B/R) + G] + 3.7451	0.33
4	10.789 x (G + R + NIR) + 0.2845	0.81	17	-0.8478 x [(B/R) + R] + 3.7247	0.32
5	5.3855 x (R/G) - 1.6506	0.77	18	-0.8108 x [(B/R) + B] + 3.6895	0.32
6	8.2823 x (B + G + R + NIR) - 0.0287	0.76	19	17.588 x B + 0.6427	0.30
7	10.241 x (B + G + R) - 0.1374	0.75	20	0.982 x (R/B) + 1.648	0.17
8	14.834 x (B + R) - 0.0091	0.73	21	-0.2439 x [(B/NIR) + G] 2.8986	0.12
9	16.705 x (G + NIR) + 0.2038	0.72	22	-0.2427 x [(B/NIR) + B] + 2.8986	0.12
10	29.187 x G - 0.103	0.71	23	1.0725 x [(NIR/B) + NIR] + 1.7535	0.11
11	10.852 x (B + G + NIR) - 0.0855	0.65	24	1.7162 x (NIR/G) + 1.4118	0.10
12	30.554 x NIR + 0.9825	0.58	25	0.9607 x (NIR/B) + 1.8524	0.08
13	13.808 x (B + G) - 0.1391	0.57	26	0.2513 x (G/B) + 2.0777	0.01

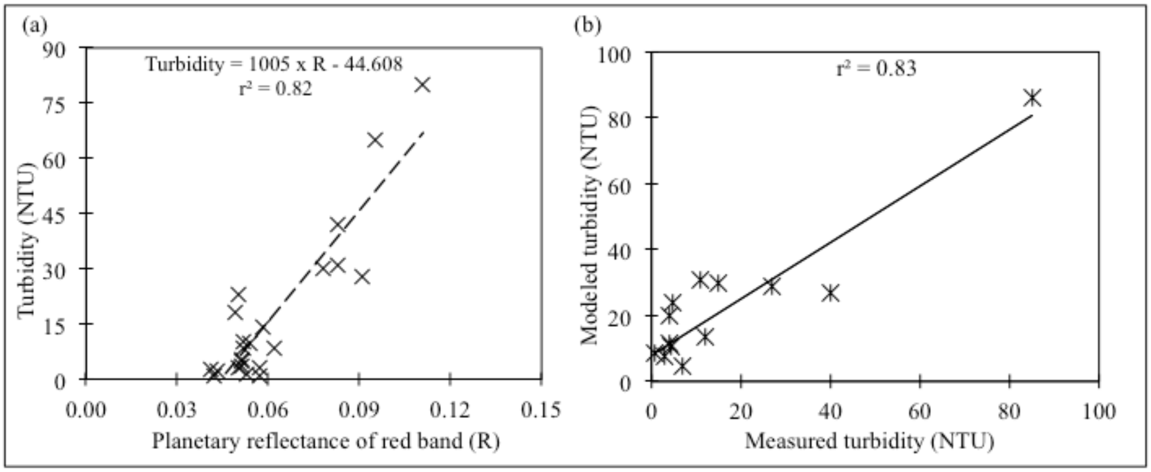

Figure 6.3: (a) Development, and (b) evaluation of most suitable model for obtaining turbidity using the planetary reflectance of red band for the Bow River.

6.3.2 Empirical models for obtaining turbidity classes

We created 26 empirical models for turbidity as given in **Table 6.6**, which could be used to obtain the spatial distribution of turbidity using the planetary reflectance of bands for a period of interest. The slopes and intercepts for the empirical models (i.e., model no. 1 to model no. 26) of turbidity are given in **Table 6.6**.The range of r^2 for the models was from 0.01 to 0.82. Similar to CWQI empirical models, we considered only the models with $r^2 > 0.50$ (model no. 1 to model no. 12 in **Table 6.6**) as significant and the all models with $r^2 < 0.50$ were weak (Roberts and Roberts 2013). Similar to CWQI models, r^2 was higher (i.e., 0.66 to 0.82 for the considerable models with red band (e.g., model no. 1 to model no. 8 in **Table 6.6**), whereas it was lower (i.e., 0.52 to 0.66) for the models without red band (e.g., model no. 9 to model no. 12 in **Table 6.6**). From **Table 6.6**, we found that the model 20 to model 26 showed very low values of r^2 (i.e., in the range 0.01-0.13). These models had weak relationship between planetary reflectance and *in-situ* turbidity. The values of r^2 were in the range 0.27-0.32 for model 15 to model 19, which were higher as compared to the values of model 20 to model 26 but still insignificant for the site application. The values of r^2 for model 13 and model 14 were 0.49 and 0.47 respectively. These values were closed to the significant value of r^2. On the basis of which these models could be applied for mapping turbidity of Bow River.

Among all the models, the best model was the use of the spectral band R (i.e., $r^2 = 0.82$, see model no. 1 in **Table 6.6**) and its development is also described in **Fig. 3(a)**. We evaluated the turbidity model (i.e. model no. 1 given in **Table 6.6**) using the validation data of the ground measured turbidity data as shown in **Fig. 3(b)**. The validation indicated a strong correlation of modeled turbidity with the measured turbidity by giving $r^2 = 0.83$ which is even higher as compared to r^2 obtained for turbidity model. These results suggest the usefulness of this model (i.e. model no.1 given in **Table 6.6**) for mapping turbidity from Landsat-5 TM satellite data for the Bow River. The red band also correlated well with *in-situ* turbidity in other studies. r^2 in these studies were 0.78, 0.76, and 0.57 respectively (Moreno-Madrinan et al. 2010; Bustamante et al. 2009; Nas et al. 2010).

Table 6.6: Models developed for mapping spatial distribution of turbidity for the Bow River using the first four spectral bands (i.e. blue, green, red and near-infrared) of Landsat-5 TM satellite data.

Model no.	Models	r^2	Model no.	Models	r^2
1	1005 x R - 44.608	0.82	14	530.92 x (B + NIR) - 55.95	0.47
2	533.51 x (G + R) - 59.624	0.77	15	-30.533 x (B/R) + 66.455	0.32
3	567.88 x (R + NIR) - 42.301	0.73	16	-30.041 x [(B/R) + R] + 67.509	0.29
4	385.17 x (G + R + NIR) - 54.932	0.73	17	-30.076 x [(B/R) + G] + 68.198	0.29
5	368.42 x (B + G + R) - 70.666	0.69	18	-28.74 x [(B/R)+B] + 66.277	0.29
6	295.94 x (B + G + R + NIR) - 66.185	0.69	19	630.78 x B - 42.417	0.27
7	187.38 x (R/G) - 120.43	0.66	20	32.406 x (R/B) - 4.4814	0.13
8	531.38 x (B + R) - 65.694	0.66	21	-7.1225 x [(B/NIR) + G] + 34.532	0.07
9	1059.2 x G - 70.18	0.66	22	-7.1171 x [(B/NIR) + B] + 34.604	0.07
10	595.17 x (G + NIR) - 57.657	0.65	23	33.057 x [(NIR/B)+NIR] + 0.1987	0.08
11	387.37 x (B + G + NIR) - 68.127	0.59	24	51.506 x (NIR/G) - 9.6089	0.07
12	516.81 x (B + G) - 73.428	0.52	25	28.779 x (NIR/B) + 3.641	0.05
13	1062.7 x NIR- 28.796	0.49	26	6.8317 x (G/B) + 11.017	0.01

6.3.3 Application of models for spatial and temporal analysis

We applied the best (i) CWQI model (i.e., model no. 1 in **Table 6.5**), and (ii) turbidity determination model (i.e., model no. 1 in **Table 6.6**) over all 31 scenes of Landsat-5 TM during the period 2006-2010 for generating the spatial distribution of CWQI and turbidity classes for the Bow River. The examples of classes for CWQI and turbidity are shown over a portion of the Bow River in **Fig. 6.4** and **Fig. 6.5** respectively. The percentages for five CWQI classes and six turbidity classes observed in each year during the period 2006-2010 are given in **Table 6.7** and **Table 6.8** respectively. The deteriorated quality of water could be estimated from the percentages accumulated in each year for the CWQI classes of 4, and 5. Those were: (i) 2.62% in 2006, (ii) 32.75% in 2007, (iii) 4.77% in 2008, (iv) 1.46% in 2009, (v) 6.94% in 2010, and (vi) 9.71% during 2006-2010 on an average. On this basis, we might rank the years in order from the best to the worst water quality, such as: 2009, 2006, 2008, 2010, and 2007. Turbidity also showed similar ranks for the respective years on the basis of percentages for the worst turbidity class (i.e., >50 NTU). The variation in the water quality for different years could be related to surface runoff from different amount of precipitations due to climatic factors like snow melt and rainfall (Akbar et al. 2013).

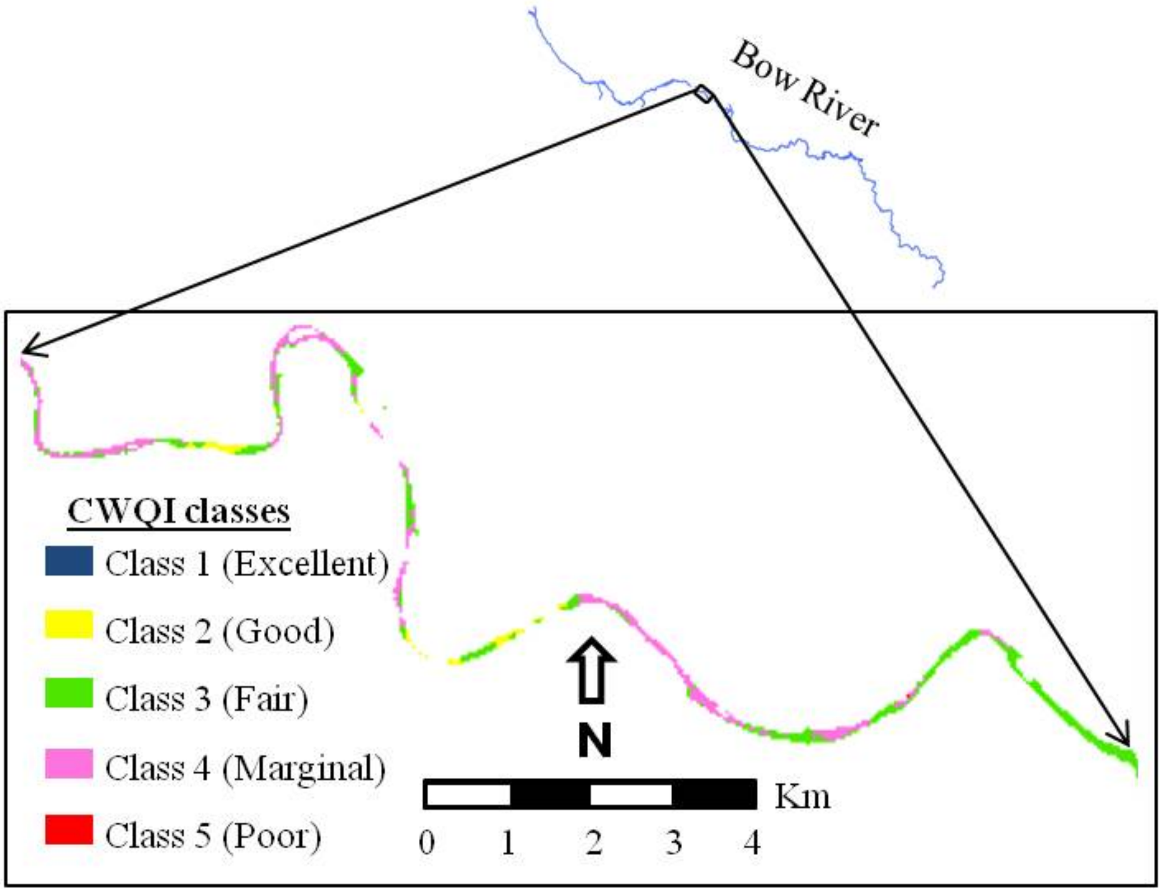

Figure 6.4: An example of CWQI classes for ~14 km long portion of the Bow River obtained by application of the most suitable empirical model (i.e., model no.1 in Table 6.4) on Landsat-5 TM satellite image dated 21st June 2007.

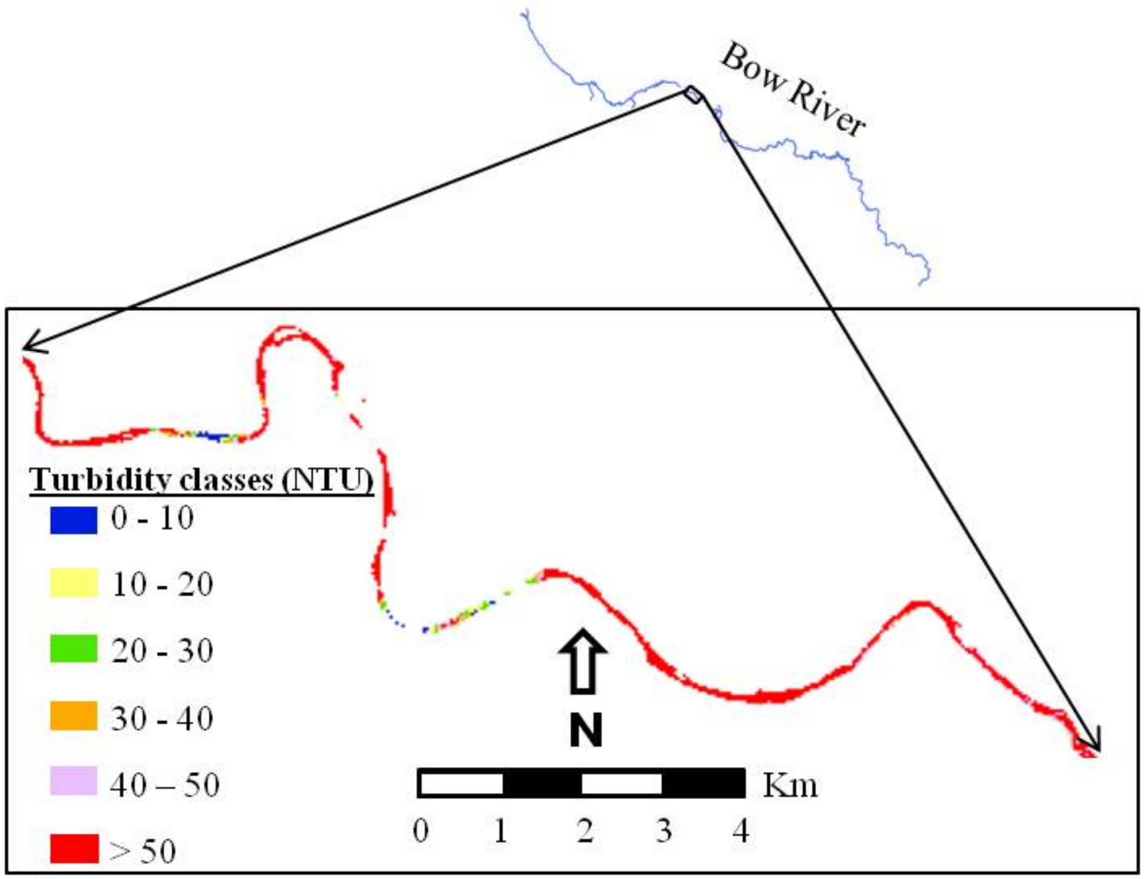

Figure 6.5: An example of turbidity classes for ~14 km long portion of the Bow River obtained by application of the most suitable empirical model (i.e., model no. 1 in Table 6.5) on Landsat-5 TM satellite image dated 21st June 2007.

Table 6.7: Percentages of CWQI classes for each year during 2006-2010.

CWQI classes	Percentage (%) of CWQI classes					
	2006	2007	2008	2009	2010	2006-2010
1	0.10	0.45	0.06	0.05	1.51	0.44
2	72.66	40.83	39.42	73.78	62.69	57.88
3	24.62	25.97	55.74	24.71	28.85	31.98
4	2.29	32.01	4.16	1.06	6.87	9.28
5	0.33	0.74	0.62	0.40	0.08	0.43

Table 6.8: Percentages of turbidity classes for each year during 2006-2010.

Turbidity classes (NTU)	Percentage (%) of turbidity classes				
	2006	2007	2008	2009	2010
0 - 10	24.40	11.18	10.91	19.37	12.86
10 - 20	36.22	19.64	18.72	41.90	28.17
20 - 30	24.08	16.60	16.24	18.90	27.05
30 - 40	6.88	9.29	21.30	10.66	9.15
40 - 50	3.92	4.27	22.57	6.06	8.99
> 50	4.49	39.02	10.26	3.10	13.78

The impact of natural subregions was reflected on the river water quality classification in **Table 6.9**. We found that the prominent CWQI classes were (i) class 3 or class 4 for Mixed grass, and (ii) class 3 for Dry mixed grass. The deteriorated water quality for the Bow River in both of these natural regions could be related to irrigation-based farming (Downing and Pettapiece 2006). During the summer months, we observed class 3 and class 4 for Bow River in Foothills parkland, Foothills fescue and Montane. This deterioration in the Bow river water quality in these three subregions could be due to till cropping (i.e., short-season crops) (Downing and Pettapiece 2006).

Table 6.9: Dominant CWQI classes for Bow River in natural subregions for selected scenes of Landsat-5 TM.

Landsat-5 TM scene dates	Dominant CWQI classes for				
	Dry mixed grass	Mixed grass	Foothills fescue	Foothills parkland	Montane
24th April 2006	2	3	2	-	-
21st June 2007	4	4	4	4	3
23rd July 2007	-	3	2	3	3
29th April 2008	3	4	3	-	-
18th July 2008	2	3	2	-	-
21st July 2009	2	2	2	-	-
25th Oct 2009	3	2	2	-	-
15th Jul 2010	-	3	2	2	2
19th April 2010	3	3	2	-	-

6.4 Concluding remarks

In this research, we developed empirical models for Canadian Water Quality Index (CWQI) and turbidity using the planetary reflectance data from the first four bands (i.e., blue, green, red and near infrared) of Landsat-5 TM for the Bow River of Alberta. The data utilized for the development and evaluation of these models included 31 scenes of Landsat-5 TM multispectral images, CWQI classes based on the monthly measured ground data for 17 water quality variables, and *in -situ* monthly measured turbidity data for a period of five years (i.e., 2006-2010). For CWQI, we created 26 models of which 14 were significant based on the co-efficient of determination (r^2) ranging from 0.54 to 0.91. Likewise for turbidity, we developed 26 models of which 12 were significant based on r^2 ranging from 0.52 to 0.82. For both CWQI and turbidity, the models with highest r^2 (i.e., 0.91 and 0.82 respectively) were evaluated and applied on all 31 scenes to obtain classes for CWQI and turbidity for the Bow River during 2006-2010. The empirical models for CWQI and turbidity were site and condition specific. These models can be used for other sites having characteristics similar to our site but we suggest evaluation of these models using their site-specific data. The red band was found to be the most important as it dominated in 8 CWQI models and 8 turbidity models with higher range of r^2 values with its solitary contribution in the best models.

The river water quality was deteriorated due to agricultural activities and climatic factors. The limitation of using 30m resolution satellite data was the contamination of river water pixels caused by influence of nearby land covers/uses. To overcome this in our research, we made use of NDVI to recognize such pixels and eliminate them from the images. The benefits of remote sensing based empirical modeling in water quality studies include: (i) simplicity in algorithm, (ii) easy implementation, (iii) easy interpretation, (iv) user-friendly output, and (v) efficient management, and (vi) targeted decisions.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

Based on the results of this research, the following major conclusions are drawn on the basis of four objectives:

Objective 1: *"Develop methodologies and models to cluster Alberta waters based on water quality".* New methods and models for clustering 18 lakes (Chapter 3) and 12 rivers (Chapter 4) were developed. For lakes, three PCs were identified which were indicators of hardness, salinity and biological activities. The dominant parameters under these PCs were total dissolved solids, total phosphorus, and chlorophyll-a. The normalized dominant parameters were used to obtain the generalized characteristics for five clusters using K-means clustering technique. The water quality deteriorated as the cluster number increased from 1 to 5. For classifying the water quality of 12 rivers, the clusters were developed and evaluated using CWQI. The normalization models were developed on the basis of Canadian water quality guidelines and the data was normalized to obtain PCA. PCA revealed seven PCs, which were the indicators of watershed geology, mineralization and anthropogenic activities related to land use/cover. The normalized dominant parameters (i.e., total dissolved solids, true color, pH, iron, fecal coliform, dissolved oxygen, and turbidity) of PCs were used to develop total exceedance model. 70% of the exceedance values were used to develop five clusters while 30% of the values were used to evaluate them. The matching of cluster 1, cluster 2, cluster 3, cluster 4, and cluster 5 was 85.71%, 83.54%, 90.22%, 80.74%, and 83.40% with their respective CWQI classes.

Objective 2: *"Analyze the spatial patterns and temporal trends of surface water quality".*

The clusters, obtained for lakes and rivers, were used to analyze the spatio-temporal patterns as discussed in Chapter 3 and Chapter 4. For lakes, it was found that 50% of the lakes showed stability in water quality while others changed over the time. The most deteriorated water quality was observed for five lakes (i.e., Cardinal, Moonshine, Winagami, Miquelon Lake, and Saskatoon). For rivers, the snow melting decreased the water quality due to anthropogenic activities from different land use/cover. The water quality was worse in the growing season. The most deteriorated water quality was observed for Battle River and Milk River.

Objective 3: *"Obtain exceedances of parameters in each cluster".* A mean exceedance model was developed to obtain the exceedance patterns of parameters in the clusters of rivers (Chapter 5). For all the rivers there was increasing trend for the mean exceedance of the parameters as the cluster number increased from low to high. The mean exceedance was higher for FC, TUR, TP, TN, TC, DO, Fe and Mn in various years. The exceedance in FC, TUR, TP, TN, TC, and DO could be related to anthropogenic activities of land cover/uses while the exceedance in Fe and Mn was due to natural mineralization.

Objective 4: *"Develop remote sensing based models for Canadian Water Quality Index (CWQI) and turbidity".* The empirical models were developed for Canadian Water Quality Index (CWQI) and turbidity using the planetary reflectance and ground measured data for the Bow River of Alberta (Chapter 6). For CWQI, 14 models were significant in which the co-efficient of determination (r^2) was in the range 0.54-0.91. Similarly 12 models were found significant with r^2 ranging from 0.52 to 0.82 for turbidity. The r^2 for the best-fit models were 0.91 for the CWQI model and 0.82 for the turbidity model. After validation of these best models with ground-measured data, 100% matching was found for 72% and 83% of data in CWQI and turbidity models respectively. Among bands, the red band was most prominent as it was present in 8 CWQI models and 8 turbidity models. The surface water quality from best to worst were: 2009, 2006, 2008, 2010, and 2007, respectively. The variation in water quality could be due to changes in weather conditions during the period of interest. In addition, activities related to irrigation could be the reason for the deteriorated water quality in Mixed grass and Dry mixed grass natural sub-regions.

7.2 Contribution to science

The specific contributions include following:

1. Several techniques (that included principal component analysis, normalization, and clustering) were implemented to obtain the clusters for analyzing the surface water quality. The outcomes demonstrated that small number of parameters would be sufficient in understanding the water quality, which is a new finding. The knowledge about such small number of parameters will be critical for the water quality monitoring agencies in order to: (i) reduce the cost of operation for monitoring water quality, and (ii) design new water quality monitoring stations.
2. The clusters generated in this research were used to analyze the impact of climate (i.e., snow melting in particular) and land use activities qualitatively on the surface water quality. According to our knowledge, such analyses of Alberta water bodies are conducted for the first time.
3. The outcomes of this research (i.e., clusters and parameter exceedance) can be considered as foundation for the development and designing of drinking water treatment plants. The results from this research could contribute in the sense that specific parameters with required treatment level can be identified for the targeted treatment trains of plants.
4. In Alberta, site-specific CWQI is calculated by considering the water quality data of the specific sampling site. To the best of our knowledge, remote sensing (RS) based CWQI models were developed for the first time. In comparison to existing site-specific approach, RS-based-CWQI model can be applied to delineate the spatial distribution of CWQI classes along the whole river. The outcome of these models can be utilized for the effective and efficient management of rivers.
5. It may be quite possible that the width of different rivers may be in order of several meters. As such, it will not be possible to use Landsat TM with 30 m spatial resolution. Keeping this approach as a base, high spatial resolution satellite data (i.e., 1m, 5m, and 10m) can be utilized.

7.3 Recommendations for future work

The recommendations for future research work are given below:

1. In this research large amount of ground measured water quality and satellite data was used. If data gaps are identified due to the reasons such as: (i) insufficient data, (ii) missing values, (iii) seasonal collection, (iv) cloud cover, and (v) snow cover, then new techniques and protocols should be explored and considered for filling them (Chowdhury and Hassan 2013).
2. Loss of information might have occurred while implementing the various algorithms and processes at different stages of this research work. The uncertainty analysis is recommended to explore uncertainty about: (i) model structure, (ii) model input, and (iii) model output (Beck 2010).
3. In this research remote sensing based CWQI model was developed and implemented for analyzing the water quality of one river. In a similar way, RS-based CWQI models can be developed, evaluated and applied for other rivers of Alberta. Prior to the development of such models, GIS-based layers for watershed characteristics, precipitation, soil, geology, sewage discharge, and industrial discharge can be considered to analyze the impact on the surface water quality.
4. Remote sensing based models were developed for mapping the spatial distribution of turbidity along the rivers. Similar types of models can be developed for other parameters like total suspended solids, total phosphorus and chlorophyll-a etc.

REFERENCES

Adham S. S., Jacangelo J. G., Laine J. M., Characteristics and Costs of MF and UF, *Plants. J. Am. Water Works Assoc.* **1996**, *88,* 22-31.

Akbal F., Gürel L., Bahadır T., Gürel İ., Bakan G., Büyükgüngör H., Multivariate Statistical Techniques for the Assessment of Surface Water Quality at the Mid-Black Sea Coast of Turkey, *Water Air Soil Pollut.* **2011**, *216,* 21-37.

Akbar T. A., Akbar R. A., Pesticide Health Risk Mapping and Sensitivity Analysis of Parameters in Groundwater Vulnerability Assessment, *Clean – Soil, Air, Water*. doi:10.1002/clen.201200232

Akbar T. A., Hassan Q. K., Achari G., A Methodology for Clustering Lakes in Alberta on the basis of Water Quality Parameters, *Clean – Soil, Air, Water* **2011,** *39*, 916-924.

Akbar T. A., Hassan Q. K., Achari G., Clusterization of Surface Water Quality and its Relation to Climate and Land Use/Cover, *JEP* **2013,** *4*, 333-343.

Akbar T. A., Lin H. S., GIS Based Arcprzm-3 Model for Bentazon Leaching towards Groundwater, *J Environ Sci.* **2010,** *22,* 1854-1859.

Alberta Environment and Sustainable Resource Development, *Alberta River Water Quality Index: Objectives, Water Data*, Government of Alberta, Canada, **2011**.

Alberta Environment, *Surface Water Quality Guidelines for use in Alberta*, Science and Standards Branch, Environmental Assurance Division, Edmonton, Alberta, 1999.

Alberta Environment, *Facts About Water in Alberta*, Water for Life, Alberta Environment Information Centre, Government of Edmonton, Alberta, Canada, **2010**.

Alberta Environment, *Lake Water Quality Monitoring Reports for Beauvais, Cardinal, Crimson, Dillberry, Elkwater, Gregg, Gregoire, Jarvis, Long, Mcleod, Miquelon, Moonshine, Reesor, Saskatoon, Spruce Coulee, Steele, Sturgeon and Winagami*, Alberta Environment Information Centre, Government of Edmonton, Alberta, Canada, **2006**.

Alparslan E., Aydöner C., Tufekci V., Tüfekci H., Water Quality Assessment at Ömerli Dam Using Remote Sensing Techniques, *Environ. Monit. Assess.* **2007,** *135*, 391–398.

Alvarez-Cobelas M., Catalan J., García-Jalón D., *Impacts on Inland Water Ecosystems in Moreno JM (Ed) A Preliminary General Assessment of the Impacts in Spain Due to the Effects of Climate Change*, Spanish Ministry of Environment, Madrid **2005**, p. 113–146.

American Water Works Association, *Water Quality and Treatment*, 3rd Edition, McGraw-Hill Book Company, Toronto **1971**.

Anderson A. –M., *River Water Quality of Battle River*, Water Sciences Branch, Water Management Division, Natural Resources Service, Alberta Environment, Canada, 1999.

Areerachakul S., Sanguansintukul S., Clustering Analysis of Water Quality for Canals in Bangkok, Thailand, *Comput. Sci. Appl.* **2010**, *6018*, 215-227.

Aronoff S., *Geographic Information System: A Management Perspective*, WDL Publications Ottawa, Canada **1993**.

Astel A., Biziuk M., Przyjazny A., Namieśnikb J., Chemometrics in Monitoring Spatial and Temporal Variations in Drinking Water Quality, *Water Res.* **2006**, *40*, 1706-1716.

Baborowski M., Buttner O., Einax J. W., Assessment of Water Quality in the Elbe River at Low Water Conditions Based on Factor Analysis, *Clean – Soil Air Water* **2011**, *39*, 437-443.

Barale V., Cipollini P., Davidov A., Melin F., Water Constituents in the North-Western Black Sea from Optical Remote Sensing and In Situ Data, *Est. Coast. Mar. Sci.* **2002** *54*, 309–320.

Barreto W. J., Ishikawa D. N., Scarminio I. S., Costa J. de S., Nora P. dos S., Soares M. de F., Nicolau R. M., et al., Fe, Mn, P and S Speciation in Sediments from the Capivara Hydroelectric Dam Lake (Brazil) as an Indicator of Anthropogenic Influences, *Clean – Soil Air Water* **2008**, *36*, 353-359.

Baruah P. J., Tamura M., Oki K., Nishimura H., *Neural Network Modeling of Surface Chlorophyll and Sediment Content in Inland Water from Landsat Thematic Mapper*

Imagery Using Multidate Spectrometer Data in Gilbert, G.D., Frouin, R. J. (Eds.), Ocean Optics: Remote Sensing and Underwater Imagery, Proceedings of SPIE the International Society for Optical Engineering 4488, Seattle, **2002** p. 205–212.

Beck M. B., Water Quality Modeling: A Review of the Analysis of Uncertainty, *Water Resour. Res.* **1987** *23*, 1393–1442.

Bellamy W. D., Silverman G. P., Hendricks G. W., *Filtration of Giardia Cysts and Other Substances, Vol. 2, Slow Sand Filtration*, EPA 600/2-85/026, U.S. Environmental Protection Agency, Cincinnati, OH, USA **1985**.

Ben-Hur A., Guyon I., *Detecting Stable Clusters Using Principal Component Analysis in Brownstein M. J., Khodursky A., Eds., Methods in Molecular Biology*, Humana Press, Totowa, New Jersey, USA **2003** p. 82-159.

Bengraïne K., Marhaba T. F., Comparison of Spectral Fluorescent Signatures-Based Models to Characterize DOM in Treated Water Samples, *J Hazard Mater.* **2003**, *B100*, 117-130.

Bennett M., Murray M., *Bow River Basin*, *State of the Watershed Summary 2010,* Bow River Basin Council, Calgary Water Centre, Calgary, Alberta, Canada **2010**, p. 5.

Binding C. E., Jerome J. H., Bukata R. P., Booty W. G., Suspended Particulate Matter in Lake Erie Derived from MODIS Aquatic Colour Imagery, *Int. J. Remote Sens*. **2010**, *31*, 5239-5255.

Black A. P., Christman R. F., Characteristics of Colored Surface Waters, *J. Am. Water Works Assoc.* **1963,** *55*, 753-770.

Bolgrien D. W., Granin N. G., Levin L., Surface-Temperature Dynamics of Lake Baikal Observed from AVHRR Images, *Photogramm. Eng. Rem. S.* **1995,** *61*, 211-216.

Bolstad P. V., Swank W. T., Cumulative Impacts of Land-Use on Water Quality in a Southern Appalachian Watershed, *J. Am. Water Resour. Assoc.* **1997,** 33, 519-534.

Bolton P.W., An Index to Improve Water Quality Classification, *Water Pollut. Control* **1978**, *77*, 271-284.

Bordalo A. A., Teixeira R., Wiebe W. J., A Water Quality Index Applied to an International Shared River Basin: The Case of the Douro River, *Environ. Manage.* **2006**, *38*, 910-920.

Boyce M. S., Vernier P. R., Nielsen S. E., Schmiegelow F. K. A., Evaluating Resource Selection Functions, *Ecol. Model.* **2002**, *157*, 281-300.

Brivio P. A., Giardino C., Zilioli E., Determination of Chlorophyll Concentration Changes in Lake Garda Using an Image-Based Radiative Transfer Code for Landsat TM Images, *Int. J. Remote Sens.* **2001**, *22*, 487-502.

Brown R. M., McLelland N. I., Deininger R. A., O'Connor M. F., A Water Quality Index - Crashing the Psychological Barrier, Indicators of Environmental Quality, Environmental Science Research **1972**, *1*, 173-182.

Bruneau J., Corkal D. R., Pietroniro E., Toth B., Kamp G. V. D., Human Activities and Water Use in the South Saskatchewan River Basin, *Prairie Forum* **2009,** *34,* 129-152.

Bustamante J., Pacios F., DÍaz-Delgado R., Aragonés D., Predictive Models of Turbidity and Water Depth in the Doňana Marshes Using Landsat TM and ETM+ Images, *J. Environ. Manage.* **2009**, *90*, 2219-2225.

Canadian Council of Ministers of the Environment (CCME), *Canadian water quality index 1.0 technical report and user's manual*, Canadian Environmental Quality Guidelines Water Quality Index Technical Subcommittee, Gatineau, QC, Canada **2001**.

Canadian Council of Resource and Environment Ministers, Total Dissolved Solids in Canadian Water Quality Guidelines Prepared by the Task Force on Water Quality Guidelines, Environment Canada, Ottawa **1987**.

Chambers P. A., Guy M., Roberts E. S., Charlton M. N., Kent R., Gagnon C., Grove G., Foster N., *Nutrients and their Impact on the Canadian Environment*, Environment Canada, Hull, QC, Canada **2001**.

Chander G., Markham B., Revised Landsat-5 TM, Radiometric Calibration Procedures and Postcalibration, Dynamic Ranges, *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2674-2677.

Chen C., Pan Z., Shi P., Zhan H., Application of Satellite Data for Integrated Assessment of Water Quality in the Pearl River Estuary, China, Geoscience and Remote Sensing Symposium, IGARSS Proceedings, *IEEE International* **2005,** *4*, 2550-2553.

Chuntao R., Changyou L., Keli J., Sheng Z., Weiping L., Youling C., Water Quality Assessment for Ulansuhai Lake using Fuzzy Clustering and Pattern Recognition, *Chin. J. Oceanol. Limnol.* **2008,** *26*, 339-344.

Chowdhury E. H., Hassan Q. K., Use of Remote Sensing-Derived Variables in Developing A Forest Fire Danger Forecasting System, *Nat. Hazards.* **2013,** *67*, 321-334.

Clark J. W., Viessman W., Hammer M. J., *Water Supply and Pollution Control*, 3rd Edition, Harper & Row Publishers, New York, NY **1977**.

CliffsNotes, *Simple Linear Regression*, Houghton Mifflin Harcourt Publishing Company, Boston, USA **2013.**

Consonery P. J., Greenfield D. N., Filter Plant Performance Evaluation: A Summary of Procedures and Results, Internal Report for the Bureau of Water Supply and Community Health, Pennsylvania Department of Environmental Protection, Harrisburg, PA **1991**.

Cox R. M., Forsythe R. D., Vaughan G. E., Olmsted L. L., Assessing Water Quality in Catawba River Reservoirs Using Landsat Thematic Mapper Satellite Data, *Lakes Reservoirs: Res. Manage.* **1998**, *14*, 405–416.

Crittenden J. C., Trussell R. R., Hand D. W., Howe K. J., Tchobanoglous G.,Water Treatment:Principles and Design Second edition, John Wiley & Sons, Inc. New Jersey, USA **2005**.

Crozes G., White P., Marshall M., Enhanced Coagulation: Its Effect on NOM Removal and Chemical Costs. *J.AWAA* **1995**, *87*, 78–89.

Cude C., Oregon Water Quality Index: A Tool for Evaluating Water Quality Management Effectiveness, *J. Am. Water Resour. As.* **2001**, *37*, 125-137.

Cunjie Z., Bourque C. P.-A., Landong S., Hassan Q. K., Spatiotemporal Modeling of Monthly Precipitation in the Upper Shiyang River Watershed in West Central Gansu, Northwest China, Adv. Atmos. Sci. **2010**, *27*, 185.

Cushing C. E., *Regional Assessment of Freshwater Ecosystems and Climate Change in North America*, Advances in Hydrological Processes, John Wiley and Sons Ltd., **1997.**

Dekker A. G., Vos R. J., Peters S. W. M., Analytical Algorithms for Lake Water TSM Estimation for Retrospective Analyzes of TM and SPOT Sensor Data, *Int. J. Remote Sens.* **2002**, *23*, 15–35.

Dekker A. G., Vos R. J., Peters S.W. M., Comparison of Remote Sensing Data, Model Results and In Situ Data For Total Suspended Matter (TSM) in the Southern Frisian Lakes, *Sci. Total Environ.* **2001**, *268*, 197–214.

Delpla I., Jung A. -V., Baures E., Clement M., Thomas O., Impacts of Climate Change on Surface Water Quality in Relation to Drinking Water Production, *Environ. Int.* **2009,** *35*, 1225-1233.

Dempster A. P., Laird N. M., Rubin D. B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Stat. Soc., B* **1977**, *34*, 1-38.

Department of Fisheries and Environment, Water Quality Branch, Surface Water Quality in Canada, An overview, Ottawa **1977**.

Department of National Health and Welfare, *Water Treatment Principles and Applications: A Manual for the Production of Drinking Water*, Canadian Water and Wastewater Association **1993**.

Do H. T., Lo S-L., Thi L. A. P., Calculating of River Water Quality Sampling Frequency By the Analytic Hierarchy Process (AHP), *Environ. Monit. Assess.* **2013,** *185*, 909-916.

Dong J., Zhang Y., Zhang S., Wang Y., Yang Z., Wu M., Identification of Temporal and Spatial Variations of Water Quality in Sanya Bay, China By Three-Way Principal Component Analysis, *Environ. Earth Sci.* **2010,** *60*, 1673–1682.

Dow C. L., Zampella R. A., Specific Conductance and pH as Indicators of Watershed Disturbance in Streams of the New Jersey Pinelands, USA, Environ. Manage. **2000** *26*, 437-445.

Downing D. J., Pettapiece W. W., *Natural Regions and Subregions of Alberta, Natural Regions Committee*, Government of Alberta, Alberta, Canada, **2006**.

Dubiella-Jackowska A., Astel A., Polkowska Z., Staszek W., Kudlak B., Namiesnik J., Atmospheric and Surface Water Pollution Interpretation in the Gdańsk Beltway Impact Range by the Use of Multivariate Analysis, *Clean – Soil Air Water* **2010**, *38*, 865-876.

Eneji I., Onuche A., Sha'Ato R., Spatial and Temporal Variation in Water Quality of River Benue, Nigeria, *JEP* **2012**, *3,* 915-921.

Environment Canada, Water Quality, *Fresh Water Quality Monitoring & Surveillance*, Gatineau, Quebec, Canada **2012.**

Fielding A. H., *Cluster and Classification Techniques for the Biosciences*, Cambridge University Press, Cambridge, UK **2007,** p. 58-65.

Fielding A. L., Bell J. F., A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models, *Environ. Conserv.* **1997,** *24*, 38-49.

Fox K. R., Miltner R. J., Logsdon G. S., Dicks D. L., Drolet L. F., Pilot Plant Studies of Slow-Rate Filtration, *J. Am. Water Works Assoc.* **1984,** *76*, 62-68.

Fraser R. N., Multispectral Remote Sensing of Turbidity among Nebraska Sand Hills Lakes, *Int. J. Remote Sens.* **1998,** *19*, 3011–3016.

Furtula V., Osachoff H., Derksen G., Juahir H., Colodey A., Chambers P., Inorganic Nitrogen, Sterols and Bacterial Source Tracking as Tools to Characterize Water Quality and Possible Contamination Sources in Surface Water, Water Res. **2012,** *46,* 1079-1092.

Gabellone N. A., Solari L., Claps M., Neschuk N., Chemical Classification of the Water in a Lowland River Basin (Salado River, Buenos Aires, Argentina) Affected by Hydraulic Modifications, *Environ. Geol.* **2008,** *53,*1353-1363.

García-Reiriz A., Magallanes J., Vracko M., Zupan J., Reich S., Cicerone D., Multivariate Chemometric Analysis of a Polluted River of a Megalopolis, *JEP* **2011**, *2*, 903-914.

Garg A., Achari G., Joshi R. C., Application of Fuzzy Rule-Based Logic to Estimate Flow of Methane for Energy Generation at a Sanitary Landfill, *J. Energy Eng.* **2007**, *133*, 212-223.

Giardino C., Pepe M., Brivio P. A., Ghezzi P., Zilioli E., Detecting Chlorophyll, Secchi Disk Depth and Surface Temperature in a Sub-Alpine Lake using Landsat Imagery, *Sci. Total Environ.* **2001,** *268*, 19–29.

Gleick P.H., *Water in Crisis: A Guide to the World's Freshwater Resources*, Oxford University Press, Inc., New York **1993,** p. 13.

Guan X., Li J., Booty W. G., Monitoring Lake Simcoe Water Clarity Using Landsat-5 TM Images, *Water Resour. Manage*. **2011**, *25*, 2015-2033.

Haag I., Westrich B., Processes Governing River Water Quality Identified By Principal Component Analysis, *Hydrol. Process.* **2002,** *16*, 3113–3130.

Health Canada, *Guidelines for Canadian Drinking Water Quality*, Federal-Provincial-Territorial Committee on Drinking Water of the Federal-Provincial-Territorial Committee on Health and the Environment, **2010**.

Hellweger F. L., Schlossera P., Lalla U., Weissel J. K., Use of Satellite Imagery for Water Quality Studies in New York Harbor, *Estuar. Coast. Shelf Sci.* **2004**, *61*, 437–448.

Horton R. K., An Index Number System for Rating Water Quality, *J. Water Pollut. Con. F.* 1965, *37*, 300-306.

Huang F., Wang X., Lou L., Zhou Z., Wu J., Spatial Variation and Source Apportionment of Water Pollution in Qiantang River (China) using Statistical Techniques, *Water Res.* **2010**, *44*, 1562-1572.

Hunchak-Kariouk K., Nicholson R. S., Watershed Contributions of Nutrients and Other Nonpoint Source Contaminants to the Barnegat Bay-Little Egg Harbor Estuary, *J. Coast. Res.* **2001** *32,* 28-81.

Inhaber H., An Approach To A Water Quality Index for Canada, *Water Res.* 1975, *9*, 821-833.

Jacangelo J. G., Laine J. M., Carns K. E., Cummins E.W., Mallevialle J., Low Pressure Membrane Filtration for Removing Giardia and Microbial Indicators, *J. Am. Water Works Assoc.* **1991,** *83,* 97-106.

Jayakumar R., Siraz L., Factor Analysis in Hydrogeochemistry of Coastal Aquifers – a Preliminary Study, *Environ. Geochem*. **1997,** *31*, 174-177.

Jeffries K. M., Nelson E. R., Jackson L. J., Habibi R. H., Basin-Wide Impacts of Compounds With Estrogen-Like Activity on Longnose Dace (Rhinichthys Cataractae) in two Prairie Rivers of Alberta, Canada, *Environ. Toxicol. Chem.* **2008**, *27*, 2042–2052.

Jiang D., Chen Y., Ni G., Phosphorus in Drinking Water and It's Removal in Conventional Treatment Process, *Adv. Mat. Res.* **2012** *461*, 453-456.

Johnson B. A., Gong B., Bellamy W., Fran T., Pilot Plant Testing of Dissolved Air Flotation for Treating Boston's Low Turbidity Water Supply, *Water Sci. Technol.* **1995***, 31,* 83-92.

Johnson M. L., Watt M. K., *Hydrology of the Unconfined Aquifer System, Mullica River Basin, New Jersey, 1991-1992*. U.S. Geological Survey, Water-Resources Investigations Report 94-4234 **1996.**

Juahir H., Zain S. M., Yusoff M. K., Hanidza T. I. T., Armi A. S. M., Toriman M. E., Mokhtar M., Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques, *Environ. Monit. Assess.* **2011,** *173*, 625-641.

Kambe J., Aoyama T., Yamauchi A., Nagashima U., Extraction of a Parameter as an Index to Assess Water Quality of the Tamagawa, Tokyo, Japan, by Using Neural Networks and Multivariate Analysis, *J. Comput. Chem. Jpn*. **2007**, *6*, 19-26.

Kanungo T., Mount D., Netanyahu N., Piatko D. C., Silverman R., An Efficient K-Means Clustering Algorithm Analysis and Implementation, *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *7*, 881-892.

Kartinen E. O., Martin C. J., An Overview of Arsenic Removal Processes, *Desalination* **1995** *103*, 79-88.

Kaufman L., Rousseeuw P. J., *Finding Groups in Data—An Introduction to Cluster Analysis*, John Wiley & Sons Inc., New York **1990**.

Khan A. A., Paterson R., Khan H., Modification and Application of the Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI) for the Communication of Drinking Water Quality Data in Newfoundland and Labrador, *Water Qual. Res. J. Can.* **2004**, *39*, 285-293.

Khorram S., Cheshire H., Geraci A. L., Rosa G. L., Water Quality Mapping of Augusta Bay, Italy from Landsat-TM Data. *Int. J. Remote Sensing* **1991**, *12*, 803-808.

Kloiber S. M., Brezonik P. L., Olmanson L. G., Bauer M. E., A Procedure for Regional Lake Water Clarity Assessment Using Landsat Multispectral Data, *Remote Sens. Environ.* **2002,** *82*, 38–47.

Koç C., A Study on the Pollution and Water Quality Modeling of the River Buyuk Menderes, Turkey, *Clean – Soil Air Water* **2010**, *38*, 1169-1176.

Koning C. W., Saffran K. A., Little J. L., Fent L., Water Quality Monitoring: The Basis for Watershed Management in the Oldman River Basin, Canada, *Water Sci. Technol.* **2006**, *53*, 153-161.

Koponen S., Pulliainena J., Kalliob K., Hallikainena M., Lake Water Quality Classification with Airborne Hyperspectral Spectrometer and Simulated MERIS Data, *Remote Sens. Environ.* **2002**, *79*, 51-59.

Lathrop R. G., Landsat Thematic Mapper Monitoring of Turbid Inland Water Quality, *Photogramm. Eng. Rem. S.* **1992**, *58*, 465-470.

LeChevallier M.W., McFeters G. A., *Microbiology of Activated Carbon in Drinking Water Microbiology: Progress and Recent Developments*, G. A. McFeters (Ed.), Springer-Verlag, New York, NY **1990.**

Lillesand T. M., Johnson W. L., Deuell R. L., Lindstrom O. M., Meisner D. E., Use of Landsat Data to Predict the Trophic State of Minnesota Lakes, *Photogramm. Eng. Rem. S.* **1983**, *49*, 219-229.

Liu C. W., Lin K. H., Kuo Y. M., Application of Factor Analysis in the Assessment of Groundwater Quality in a Blackfoot Disease Area in Taiwan, *Sci. Total Environ.* **2003**, *313*, 77-89.

Liu D., Chin C., Gong J., Fu D. Remote Sensing of Chlorophyll-a Concentrations of the Pearl River Estuary from MODIS Land Bands, *Int. J. Remote Sens*. **2010**, *31*, 4625-4633.

Liu Y., Guo H., Yang P., Exploring the Influence of Lake Water Chemistry on Chlorophyll a: A Multivariate Statistical Model Analysis, *Ecol. Model.* **2010**, *221*, 681-688.

Logsdon G. S., Symons J. M., Hoye R. L., Arozarena M. M., Alternative Filtration Methods for Removal of Giardia Cysts and Cyst Models, *J. Am. Water Works Assoc.* **1981***, 73*, 111-118.

Lu P., Mei K., Zhang Y., Liao L., Long B.,Dahlgren R. A., Zhang M., Spatial and Temporal Variations of Nitrogen Pollution in Wen-Rui Tang River Watershed, Zhejiang, China, *Environ. Monit. Assess.* **2011,** *180*, 501-520.

Lumb A., Halliwell D., Sharma T., Application of CCME Water Quality Index to Monitor Water Quality: A Case of the Mackenzie River Basin, Canada, *Environ. Monit. Assess.* **2006**, *113*, 411-429.

Mahapatra S., Mitra S., Managing Land and Water under Changing Climatic Conditions in India: A Critical Perspective, *JEP* **2012**, *3*, 1054-1062.

Mancino G., Nolè A., Urbano V., Amato M., Ferrara A., Assessing Water Quality by Remote Sensing in Small Lakes: The Case Study of Monticchio Lakes in Southern Italy, *IForest* **2009**, *2*, 154-161.

Mayo M., Gitelson A., Yacobi Y. Z., Ben-avraham Z., Chlorophyll Distribution in Lake Kinneret Determined from Landsat Thematic Mapper Data, *Int. J. Remote Sens.* **1995** *16*, 175-182.

Mazlum N., Özer A., Mazlum S., Interpretation of Water Quality Data by Principal Components Analysis, *Tr. J. of Engineering and Environmental Science* **1999,** *23*, 19-26.

Mcknight D. M., Brakke D. F., Mulholland P. J., Fresh Water Ecosystems and Climate Change, *Limnol. Oceanogr.* **1996,** *41*, 815-1149.

McLachlan G. J., Do K. A., Ambroise C., *Analyzing Microarray Gene Expression Data*, John Wiley & Sons, Inc., New York **2004**.

Ministry of the Environment, *Technical Support Document for Ontario Drinking Water Standards, Objectives and Guidelines*, Government of Ontario, Canada, **2006**.

Mitchell P., *Water quality of the North Saskatchewan River in Alberta*, Surface Water Assessment Branch, Technical Services and Monitoring Division, Alberta Environmental Protection, Canada, **1994**.

Montes-Botella C., Tenorio M. D., Water Characterization and Seasonal Heavy Metal Distribution in the Odiel River (Huelva, Spain) by Means of Principal Component Analysis, *Arch. Environ. Contam. Toxicol.* **2003,** *45*, 436-444.

Moreno-Madrinan M. J., Al-Hamdan M. Z., Rickman D. L., Muller-Karger F. E., Using the Surface Reflectance MODIS Terra Product to Estimate Turbidity in Tampa Bay, Florida**,** *Int. J. Remote Sens*. **2010**, *2*, 2713-2728.

Moss B., *Ecology of Fresh Waters: Man and Medium, Past to Future*, 3rd Edition, Blackwell Science, Oxford, UK **1998**.

Murdoch P. S., Baron J. S., Miller T. L., Potential Effects of Climate Change on Surface-Water Quality in North America, *J. Am. Water Resour. Assoc.* **2000,** 36, 347–366.

Murphy A. P., Chemical Removal of Nitrate from Water, *Nature* **1991,** *350,* 223-225.

Nas B., Ekercin S., Karabörk H., Berktay A., Mulla D. J., An Application of Landsat-5TM Image Data for Water Quality Mapping in Lake Beysehir, Turkey, *Water Air Soil Poll.* **2010**, *212*, 183-197.

National Aeronautics and Space Administration (NASA), *Landsat 7 Science Data Users Handbook,* USA, **2011.**

National Aeronautics and Space Administration (NASA), *Landsat Processing Details*, United States Geological Survey, USA, **2012**.

National Aeronautics and Space Administration (NASA), *MODIS/Terra Land Cover Type Yearly L3 Global 1km SIN Grid V004, Land Cover Product, MOD12Q1, LP DAAC User Services*, U.S. Geological Survey, Center for Earth Resources Observation and Science (EROS), Sioux Falls, South Dakota, USA, **2004**.

National Aeronautics and Space Administration (NASA), *MODIS/Terra Land Cover Type Yearly L3 Global 1km SIN Grid V004, Land Cover Product, MOD12Q1, LP DAAC User Services*, U.S. Geological Survey, Center for Earth Resources Observation and Science (EROS), Sioux Falls, South Dakota, USA, **2004**.

National Aeronautics and Space Administration (NASA), *The Thematic Mapper,* USA, **2013**.

Olmanson L. G., Brezonik P. L., Bauer M. E., Airborne Hyperspectral Remote Sensing to Assess Spatial Distribution of Water Quality Characteristics in Large Rivers: The Mississippi River and its Tributaries in Minnesota, *Remote Sens. Environ.* **2013**, *130*, 254-265.

Ontario Drinking Water Standards, *Technical Support Document for Ontario Drinking Water Standards, Objectives and Guidelines*, Ministry of the Environment, Drinking Water Ontario Government of Ontario, Canada **2006**.

Osborne L. L., Wiley M. J., Empirical Relationships Between Land Use / Cover and Stream Water Quality in an Agricultural Watershed, *J. Environ. Manage.* **1988,** *26,* 9-27.

Ostlund C., Flink P., Strombeck N., Pierson D., Lindella T., Mapping of the Water Quality of Lake Erken, Sweden, From Imaging Spectrometry and Landsat Thematic Mapper, *Sci. Total Environ.* **2001,** *268*, 139-154.

Ouyang Y., Evaluation of River Water Quality Monitoring Stations By Principal Component Analysis, *Water Res.* **2005,** *39*, 2621-2635.

Oyama Y., Matsushita B., Fukushimaa T., Matsushige K., Imai A., Application of Spectral Decomposition Algorithm for Mapping Water Quality in a Turbid Lake (Lake Kasumigaura, Japan) from Landsat TM Data, *ISPRS J. Photogramm.* **2009**, *64*, 73-85.

Panda U. C., Sundaray S. K., Rath P., Nayak B. B., Bhatta D., Application of Factor and Cluster Analysis for Characterization of River and Estuarine Water Systems – A Case Study: Mahanadi River (India), *J. Hydrol.* **2006,** *331*, 434-445.

Patrick R., *Rivers of the United States of America, Volume III: The Eastern and Southeastern States*, John Wiley and Sons, Inc., New York, New York, USA **1996**.

Payment P., Trudel M., Plante R., Elimination of Viruses and Indicator Bacteria at Each Step of Treatment During Preparation of Drinking Water at Seven Water Treatment Plants, *Appl. Environ. Microbiol*. **1985**, *49*, 1418-1428.

Prikler S., Chesy P., Parczewsk A., Einax J. W., Analytical and Chemometric Evaluation of the Situation of Waterbodies in the Polish Tatra Mountains, *Clean – Soil Air Water* **2011**, *39*, 55-59.

Pyper G. R., *Slow Sand Filter and Package Treatment Plant Evaluation: Operating Costs and Removal of Bacteria, Giardia and Trihalomethanes*, EPA/600/2-85/052, U.S. Environmental Protection Agency, Cincinnati, OH, USA **1985**.

Ragno G., Luca M. D., Ioele G., An Application of Cluster Analysis and Multivariate Classification Methods to Spring Water Monitoring Data, *Microchem. J.* **2007**, *87*, 119-127.

Razmkhah H., Abrishamchi A., Torkian A., Evaluation of Spatial and Temporal Variation in Water Quality by Pattern Recognition Techniques: A Case Study on Jajrood River (Tehran, Iran), *J. Water Resource Prot.* **2010**, *91,* 852-860.

Reghunath R., Murthy T. R. S., Raghavan B. R., The Utility of Multivariate Statistical Techniques in Hydrogeochemical Studies: An Example from Karnataka, India, *Water Res*. **2002**, *36*, 2437-2442.

Ren C., Li C., Jia K., Zhang S., Li W., Cao Y., Water Quality Assessment for Ulansuhai Lake Using Fuzzy Clustering and Pattern Recognition, *Chin J. Oceanol. Limnol.* **2008,** *26,* 339-344.

Research Committee on Color Problems, Report for 1966, *J. Am. Water Works Assoc.* **1967,** *59,* 1023-1035.

Rhodes A. L., Newton R. M., Pufall A., Influences of Land Use on Water Quality of a Diverse New England Watershed, *Environ. Sci. Technol.* **2001**, *35*, 3640–3645.

Roberts F., Roberts D., *Correlation Coefficient*, Mathbits, New York, USA **2013.**

Rocchini R., Swain L. G., *The British Columbia Water Quality Index*, Water Quality Branch, EP Department, B.C., Ministry of Environment, Land and Park, Victoria, B.C., Canada, **1995**.

Rock L., Mayer B., Identifying the Influence of Geology, Land Use, and Anthropogenic Activities on Riverine Sulfate on A Watershed Scale by Combining Hydrometric, Chemical and Isotopic Approaches, *Chem. Geol.* **2009**, *262*, 121-130.

Rosemond S., Duro D. C., Dubé M., Comparative Analysis of Regional Water Quality in Canada Using the Water Quality Index, *Environ. Monit. Assess.* **2009**, *156*, 223-240.

Salim N. A. A., Wood A. K., Yusof A. M., Hamzah M. S., Elias M. S., Rahman S. A., A Study of Arsenic and Chromium Contamination in Sediments of Freshwater Bodies, *Fresen. Environ. Bull.* **2009**, *18*, 1618-1623.

Sargaonkar A., Deshpande V., Development of an Overall Index of Pollution for Surface Water Based on a General Classification Scheme in Indian Context, *Environ. Monit. Assess.* **2003**, *89*, 43-67.

SAS/STAT 9.2 Users Guide, *The Cluster Procedure: Clustering Methods*, SAS Institute, Cary, NC, USA **2009**.

Sawyer C. N., McCarty P. L., *Chemistry for Sanitary Engineers*, 2nd Edition, McGraw-Hill Series in Sanitary Science and Water Resources Engineering, McGraw-Hill, Toronto, Canada **1967**.

Schiebe F. R., Harrington J. A., Ritchie J. C., Remote Sensing of Suspended Sediments: The Lake Chicot, Arkansas Project, *Int. J. Remote Sens.* **1992**, *13*, 1487–1509.

Sekhon N. S., Hassan Q. K., Sleep R. W., Evaluating Potential of MODIS-Based Indices in Determining "Snow Gone" Stage Over Forest-Dominant Regions, *Remote Sens.* **2010,** *2, 1348*-1363.

Seneka M., *Trends in Historical Annual Flows for Major Rivers in Alberta*, Environmental Monitoring and Evaluation Branch, Environmental Assurance, Alberta Environment, Government of Alberta, Alberta, Canada **2004**.

Senevirathna D. G. M., Achari G., King F., Fuzzy Rule-Based Models to Predict the Partition Coefficients of Nickel and Zinc in Aquifer Material, *Pract. Period. Hazard. Toxic Radioact. Waste Manage.* **2011**, *15*, 2-12.

Shaban M., Urban B., ElSaadi A., Faisal M., Detection and Mapping of Water Pollution Variation in the Nile Delta Using Multivariate Clustering and GIS Techniques, *J. Environ. Manage.* **2010**, *91*, 1785-1793.

Shrestha S., Kazama F., Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan, *Environ. Model. Software* **2007**, *22*, 464-475.

Singh K. P., Malik A., Mohan D., Sinha S., Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality of Gomti River (India) – a Case Study, *Water Res.* **2004**, *38,* 3980-3992.

Singh Y., Kumar M., Interpretation of Water Quality Parameters for Villages of Sanganer Tehsil, by using Multivariate Statistical Analysis, *J. Water Resource Prot.* **2010**, *2*, 860-863.

Sládeček V., Satellite Imagery in Distinguishing the Water Quality, *Acta hydrochim. hydrobiol.* **2006**, *17*, 609-617.

Slezak L. A., Sims R. C., The Application and Effectiveness of Slow Sand Filtration in the United States, *J. Am. Water Works Assoc.* **1984,** *76*, 38-43.

Smith R. A., Alexander R. B., Wolman M. G., Water Quality Trends in the Nation's Rivers, *Science* **1987** *235*, 1607-1615.

Sosiak A., Dixon J., Impacts on Water Quality in the Upper Elbow River, *Water Sci. Technol.* **2006,** *53,* 309-316.

Srivastava P. K., Mukherjee S., Gupta M., Singh S. K., Characterizing Monsoonal Variation on Water Quality Index of River Mahi in India using Geographical Information System, *Water Qual. Expo. Health* **2011,** *2*, 193-203.

St-Hilaire A., Brun G., Courtenay S. C., Ouarda T. B. M. J., Boghen A. D., Bobée B., Multivariate Analysis of Water Quality in the Richibucto Drainage Basin (New Brunswick, Canada), *J. Am. Water Resour. Assoc.* **2004**, *40*, 691-703.

Statistics Canada, *Population and dwelling counts, for Canada, provinces and territories, and census subdivisions (municipalities), 2011 censuses*, Alberta, Statistics Canada **2012**.

Stisen S., Jensen K. H., Sandholt I., Grimes D. I. F., A Remote Sensing Driven Distributed Hydrological Model of the Senegal River Basin, *J. Hydrol.* **2008**, *354*, 131–148.

Su S., Li D., Zhang Q., Xiao R., Huang F., Wua J., Temporal Trend and Source Apportionment of Water Pollution in Different Functional Zones of Qiantang River, China, *Water Res.* **2011**, *45*, 1781-1795.

Sundaray S. K., Panda U. C., Nayak B. B., Bhatta D., Behaviour and Distribution Pattern of Nutrients in River–Estuarine Waters of Mahanadi, Orissa, India, *Asian J. Water Environ. Pollut.* **2004,** *2*, 77-84.

Székely G. J., Rizzo M. L., Hierarchical Clustering Via Joint Between-Within Distances: Extending Ward's Minimum Variance Method, *J. Classif.* **2005,** *22*, 151-183.

Taylor J. S., Duranceau S. J., Barrett W. M., Goigel J. F., *Assessment of Potable Water Membrane Applications and Research Needs*, Prepared for the American Water Works Association Research Foundation, Denver, CO **1990**.

Telang S. A., Effects of Reservoir-Dam, Urban, Industrial, and Sewage Treatment Run-Off on the Presence of Oxygen and Organic Compounds in the Bow River, *Water Air Soil Poll.* **1990**, *50*, 77–90.

The Water Chronicles, *Guide to Canada's drinking water*, The Water Chronicles, Ottawa **2008**.

Theodoridis S., Koutroumbas K., *Pattern Recognition,* Third Edition, Academic Press, Inc. Orlando, FL, USA **2006**.

Tokalıoğlu Ş., Yılmaz V., Kartal S., An Assessment on Metal Sources by Multivariate Analysis and Speciation of Metals in Soil Samples Using the BCR Sequential Extraction Procedure, *Clean – Soil Air Water* **2010**, *38*, 713-718.

Toth B., Corkal D. R., Sauchyn D., Kamp G. V. D., Pietroniro E., The Natural Characteristics of the South Saskatchewan River Basin: Climate, Geography and Hydrology, *Prairie Forum* **2009**, *34,* 95-127.

Tucker C. J., Red and Photographic Infrared Linear Combinations for Monitoring Vegetation, *Remote Sens. Environ.*, **1979,** *8*, 127–150.

Tufford D. L., Samarghitan C. L., McKellar H. N., Porter D. E., Hussey J. R., Impact of Urbanization on Nutrient Concentrations in Small Southeastern Coastal Streams, *J. Am. Water Resour. As*. **2003**, *39*, 301-312.

U.S. NRC (National Research Council), *Drinking Water and Health*, Vol. 7, National Academy Press, Washington, DC, USA **1987**.

United Nations Department of Economic and Social Affairs (UNDESA), International Decade for Action ‘Water for Life’ 2005-2015, *Why a 'Water for Life' Decade?*, Zaragoza, Spain **2005**.

United States Geological Survey (USGS), *USGS Global Visualization Viewer, Earth Resources Observation and Science Center (EROS)*, USA **2012**.

Usunoff E. J., Guzman-Guzman A., Multivariate Analysis in Hydrochemistry: An Example of the Use of Factor and Correspondence Analyses, *Ground Water* **1989**, *27*, 27-34.

Vasudevan S., Sozhan G., Ravichandran S., Jayaraj J., Lakshmi J., Sheela S. M., Studies on the Removal of Phosphate from Drinking Water by Electrocoagulation Process, *Ind. Eng. Chem. Res.* **2008**, *47*, 2018-2023.

Verdin J. P., Monitoring Water Quality Conditions in a Large Western Reservoir with Landsat Imagery, *Photogramm. Eng. Rem. S.* **1985**, *51*, 343-353.

Vignolo A., Pochettino A., Cicerone D., Water Quality Assessment using Remote Sensing Techniques: Medrano Creek, Argentina, *J. Environ. Manage.* **2006**, *81*, 429-433.

Walski T. M., Parker F. L., Consumer’s Water Quality Index, *J. Environ. Eng.* **1974,** *100*, 593-611.

Wang F., Han L., Kung T., Van Arsdale R. B., Applications of Landsat-5 TM Imagery in Assessing and Mapping Water Quality in Reelfoot Lake, Tennessee, *Int. J. Remote Sens.* **2006,** *27*, 5269-5283.

Watson R.T., Zinyowera M.C., Mos R. H., *Climate Change 1995 Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses,* Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, Massachusetts, USA **1996**.

Weier J., Herring D., *Measuring Vegetation (NDVI & EVI)*, NASA earth observatory, USA **1999**.

Wenning R. J., Erickson G. A., Interpretation and Analysis of Complex Environmental Data Using Chemometric Methods, *Trends Anal. Chem.* **1994**, *13*, 446-457.

Westra S., Brown C., Lall U., Sharma A., Modeling Multivariable Hydrological Series: Principal Component Analysis or Independent Component Analysis?, *Water Res.* **2007**, *43*, 1-11.

Wilczak A., Jacangelo J. G., Marcinko J. P., Odell L. H., Kirmeyer G. J., Wolfe R. L., Occurrence of Nitrification in Chloraminated Distribution Systems, *J. Am. Water Works Assoc.* **1996**, *88*, 74.

Wilson C., Weng Q., Assessing Surface Water Quality and Its Relation with Urban Land Cover Changes in the Lake Calumet Area, Greater Chicago, *Environ. Manage.* **2010,** *45*, 1096–1111.

Winn N., Williamson C., Abbitt R., Rose K., Renwick W., Henry M., Saros J., Modeling Dissolved Organic Carbon in Subalpine and Alpine Lakes With GIS and Remote Sensing, *Landscape Ecol.* **2009**, *24* 807-816.

Wolfe B. B., Karst-Riddoch T. L., Hall R. I., Edwards T. W. D., English M. C., Palmini R., McGowan S., Leavitt P. R., Vardy S. R., Classification of Hydrological Regimes of Northern Floodplain Basins (Peace–Athabasca Delta, Canada) from Analysis of Stable Isotopes ($d^{18}O$, $d^{2}H$) and Water Chemistry, *Hydrol. Process.* **2007,** *21*, 151-168.

World Health Organization (WHO), *Analytical and Water Treatment Methods for Health-Based Guideline Values*, Report on a WHO consultation, Medmenham, U. K., WHO Regional Office for Europe, Copenhagen **1992**.

Wrona F. J., Carey J., Brownlee B., McCauley E., Contaminant Sources, Distribution and Fate in the Athabasca, Peace and Slave River Basins, Canada, *J. Aquat. Ecosyst. Stress Recovery* **2000,** *8,* 39-51.

Wu M., Zhang W., Wang X., Luo D., Application of MODIS Satellite Data in Monitoring Water Quality Parameters of Chaohu Lake in China, *Environ. Monit. Assess.* **2009**, *148*, 255–264.

Yeoman S., Stephenson T., Lester J. N., Perry R., The Removal of Phosphorus during Wastewater Treatment: A Review, *Environ. Pollut.* **1988**, *49*, 183-233.

Yidana S. M., Ophori D., Banoeng-Yakubo B., A Multivariate Statistical Analysis of Surface Water Chemistry Data – the Ankobra Basin, Ghana, *J. Environ. Manage.* **2008**, *86*, 80-87.

Zampella R. A., Characterization of Surface Water Quality Along a Watershed Disturbance Gradient, *Water Resour. Bull.* **1994** *30*, 605-611.

Zampella R. A., Procopio N. A., Lathrop R. G., Dow C. L., Relationship of Land-Use⁄Land-Cover Patterns and Surface-Water Quality in the Mullica River Basin. *J. Am. Water Resour. As.* **2007**, *43*, 437–445.

Zanderbergen P. A., Hall K. J., Analysis of the British Columbia Water Quality Index for Watershed Managers: A Case Study of Two Small Watersheds, *Water Qual. Res. J. Can.* **1998**, *33*, 519-549.

Zhang Q., Stanley S., Forecasting Raw-Water Quality Parameters for the North Saskatchewan River by Neural Network Modeling, *Water Res.* **1997**, *31*, 2340-2350.

Zhao W., Zhao Y., Wang Q., Zheng M., Zhang P., He Z., Spatial and Temporal Patterns of Plankton Assemblage Structure in a High Altitude Saline Lake, Namuka Co in Northern Tibet, China, *Clean – Soil Air Water* **2010**, *38*, 960-968.

Zhu Z., Broersma K., Mazumder A., Impacts of Land Use, Fertilizer and Manure Application on the Stream Nutrient Loadings in the Salmon River Watershed, South-Central British Columbia, Canada, *JEP* **2012**, *3*, 809-822.

APPENDIX-I

APPLICATION OF CLUSTERS FOR SEVEN MAJOR RIVERS OF ALBERTA

Summarized results and discussion

The classified water quality for the five major rivers was already described for the five major rivers in section 4.5.4 of chapter 4. In this appendix, we discussed the water quality for the remaining seven major rivers on the basis of monthly clusters during the period 2004-2008. The clusters for these rivers are given in Table 1 to Table 4. For Battle River, the dominant cluster for: (i) BR-1 was cluster 5, and (ii) BR-2 was cluster 4 and cluster 5 in the growing season. The dominant clusters were cluster 3 and cluster 4 for BR-1 and cluster 5 for BR-2 in winter. In case of Elbow River, the dominant cluster for ER-1 was cluster 5 in summer and it was cluster 4 in winter. The dominant cluster was cluster 5 for RDR-1 and RDR-2 of Red Deer River in growing season. The dominant cluster was cluster 3 for RDR-1 and it was cluster 4 for RDR-2 in winter. In case of Smoky River, the dominant cluster was cluster 5 in growing season and it was cluster 3 in winter for SR-1. For Oldman River, the dominant cluster was (i) cluster 3 for OR-1, (ii) cluster 4 and cluster 5 for OR-2, (iii) cluster 2 and cluster 3 for OR-3 in growing season. Cluster 3 was dominant for OR-1, OR-2, and OR-3 in winter. For South Saskatchewan River, the dominant cluster was cluster 4 and cluster 5 during growing season and it was cluster 3 and cluster 2 in winter. For Wapiti River, the dominant cluster was cluster 5 for WR-1 and WR-2 in growing season and it was cluster 3 for both sites in winter. During the growing season, the deteriorated water quality for Battle River, Red Deer River, Oldman River and South Saskachewan River could be related to: (i) agriculture activities as all of these rivers are dominated by cereal crops and grasses, and (ii) early snow melting (i.e., before 5-April). The anthropogenic and natural mineralization could impact the quality of Smoky River, Wapiti River, and Battle River.

Table 1: Clusters for two sampling sites of Battle River.

	Battle River									
	04		05		06		07		08	
Month	BR1	BR2	BR1	BR2	BR1	BR2	BR1	BR2	BR1	BR2
Jan	5	5	4	5	3	3	3	N	3	4
Feb	5	5	3	5	3	3	N	3	3	4
Mar	5	5	5	5	3	3	3	3	3	3
Apr	4	4	5	5	4	5	5	5	4	4
May	N	5	4	4	5	5	5	5	5	5
June	4	N	5	4	4	4	4	4	N	2
July	4	5	4	4	3	4	4	5	3	5
Aug	N	5	5	N	4	3	5	N	3	3
Sep	4	4	4	4	N	5	3	4	3	3
Oct	5	4	3	5	3	3	3	3	3	3
Nov	3	N	4	N	3	3	3	N	3	3
Dec	4	N	3	4	3	4	1	4	3	N

Note: N: No data.

Table 2: Clusters for (i) one sampling site of Elbow River, (ii) two sampling sites of Red Deer River, and (iii) one sampling site of Smoky River during the period 2004-2008.

	Elbow River					Red Deer River										Smoky River				
	04	05	06	07	08	04		05		06		07		08		04	05	06	07	08
Month	E1	E1	E1	E1	E1	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	S1	S1	S1	S1	S1
Jan	5	5	5	3	4	3	4	3	4	3	3	3	3	4	3	3	4	3	3	3
Feb	3	4	3	N	4	3	3	2	3	3	4	2	4	4	4	3	3	3	3	3
Mar	4	4	4	4	4	3	3	5	4	N	5	4	5	3	4	3	5	3	3	3
Apr	3	3	4	5	2	2	2	2	N	2	2	N	5	2	2	5	5	3	3	5
May	N	3	4	4	N	1	2	3	3	2	3	5	5	5	5	4	5	5	5	5
June	4	5	N	5	5	4	4	5	5	5	5	5	5	5	5	5	5	4	5	N
July	4	4	4	5	4	N	3	4	4	4	3	4	N	4	4	N	5	5	5	4
Aug	4	5	5	4	5	4	4	5	5	4	3	5	4	5	4	5	5	N	4	3
Sep	N	5	3	5	5	3	3	5	5	5	5	3	3	3	3	5	N	5	3	N
Oct	4	4	4	3	4	4	2	3	2	3	4	3	2	3	2	5	3	3	3	3
Nov	5	3	N	4	4	2	2	4	2	3	3	2	4	2	1	5	4	3	3	3
Dec	4	4	5	4	4	2	1	1	1	2	3	N	2	2	4	4	5	3	3	3

Note: N: No data; E1:ER-1; R1:RDR-1; R2:RDR-2; S1: SR-1.

Table 3: Clusters for three sampling sites of Oldman River during the period 2004-2008.

	2004			2005			2006			2007			2008		
Month	OR-1	OR-2	OR-3	OR-1	OR-2	OR-3	OR-1	OR-2	OR-3	OR-1	OR-2	OR-3	OR-1	OR-2	OR-3
Jan	N	1	N	5	4	4	3	3	2	3	3	3	3	2	3
Feb	3	3	2	4	3	3	3	3	N	3	3	3	3	3	4
Mar	3	3	2	4	3	2	3	3	N	5	4	1	3	3	1
Apr	3	2	2	3	3	2	5	5	2	3	3	2	3	3	2
May	2	3	2	3	4	2	3	3	2	5	5	2	5	5	3
June	4	4	3	5	5	5	5	5	3	5	5	2	5	5	4
July	4	4	2	4	4	N	4	4	3	4	5	3	4	5	2
Aug	5	5	3	4	4	3	3	4	3	3	4	3	3	4	3
Sep	4	3	3	5	5	5	3	4	3	4	5	3	3	3	3
Oct	2	3	2	4	4	3	4	N	2	3	4	2	3	2	2
Nov	3	3	3	3	4	3	3	3	2	3	3	2	2	N	3
Dec	3	3	3	3	3	3	4	3	3	3	3	4	N	N	4

Note: N: No data.

Table 4: Clusters for: (i) one sampling site of South Saskatchewan River, and (ii) two sampling sites of Wapiti River during the period 2004-2008.

	South Saskatchewan River					Wapiti River									
	04	05	06	07	08	04		05		06		07		08	
Month	SS1	SS1	SS1	SS1	SS1	W1	W2	W1	W2	W1	W2	W1	W2	W1	W2
Jan	3	N	2	1	2	3	5	3	N	3	3	3	N	3	4
Feb	3	3	2	3	2	3	5	3	4	3	3	N	3	3	4
Mar	4	3	3	4	3	4	4	4	N	3	3	3	3	3	3
Apr	3	3	N	3	3	4	4	5	5	4	5	5	5	4	4
May	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5
June	4	5	5	5	5	5	5	4	5	4	4	4	4	N	2
July	4	4	4	N	5	5	5	5	4	3	4	4	5	3	5
Aug	5	5	4	5	3	4	N	4	5	4	3	5	N	3	3
Sep	3	4	5	3	5	5	5	4	5	N	5	3	4	3	3
Oct	4	3	3	3	3	4	4	3	5	3	3	3	3	3	3
Nov	2	2	4	3	3	5	5	3	4	3	3	3	N	3	3
Dec	N	N	2	2	2	3	3	3	5	3	4	1	4	3	N

Note: N: No data; SS1: SSR-1; W1: WR-1; W2: WR-2.