### Environmental Modelling

(ENGO 583/ENEN 635)

---

**Lecture Note on:**

Surface Water Quality Modelling

Dept. of Geomatics Engineering; and Centre for Environmental Engineering Research and Education
Schulich School of Engineering
University of Calgary

---

**Review of Last Topics**
Topics of Discussion:
Surface Water Quality Modelling

- Case studies:
  - “Clusterization of surface water quality and its relation to climate and land use/cover”
  - “Development of remote sensing based models for surface water quality”

“Clusterization of Surface Water Quality and Its Relation to Climate and Land use/cover”

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Introduction (1)

- 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.

Introduction (2)

- In Alberta, 17 water quality-related parameters (see next slide for details) are periodically measured for 12 major rivers at 23 fixed sampling sites as shown in the right hand figure.

- The lengths of rivers are shown in the parenthesis.

- Directions of rivers’ flow is shown by arrows.

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
### Introduction

- There are guideline values for each of the 17 parameters in the context of determining the water quality are are summarized in the right hand Table (Health Canada 2010; Ministry of the Environment 2006; Alberta Environment and Sustainable Resource Development 2011).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Non-compliance if guideline value:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water temperature (WT)</td>
<td>&gt;15°C</td>
</tr>
<tr>
<td>Dissolved oxygen (DO)</td>
<td>&lt;6.5 mg/l</td>
</tr>
<tr>
<td>Turbidity (TUR)</td>
<td>&gt;1 NTU</td>
</tr>
<tr>
<td>True color (TC)</td>
<td>&gt;15 Pt Co units</td>
</tr>
<tr>
<td>Dissolved organic carbon (DOC)</td>
<td>&gt;5 mg/l</td>
</tr>
<tr>
<td>Total dissolved solids (TDS)</td>
<td>&gt;500 mg/l</td>
</tr>
<tr>
<td>Total phosphorus (TP)</td>
<td>&gt;0.05 mg/l</td>
</tr>
<tr>
<td>Total nitrogen (TN)</td>
<td>&gt;1 mg/l</td>
</tr>
<tr>
<td>pH (pH)</td>
<td>&lt;6.5 or &gt;8.5</td>
</tr>
<tr>
<td>Total hardness (TH)</td>
<td>&gt;500 mg/l</td>
</tr>
<tr>
<td>Chloride (Cl)</td>
<td>&gt;250 mg/l</td>
</tr>
<tr>
<td>Sulfate (SO₄)</td>
<td>&gt;500 mg/l</td>
</tr>
<tr>
<td>Sodium (Na)</td>
<td>&gt;200 mg/l</td>
</tr>
<tr>
<td>Fluoride (F)</td>
<td>&gt;1.5 mg/l</td>
</tr>
<tr>
<td>Fecal coliforms (FC)</td>
<td>&gt;0</td>
</tr>
<tr>
<td>Manganese (Mn)</td>
<td>&gt;0.05 mg/l</td>
</tr>
<tr>
<td>Iron (Fe)</td>
<td>&gt;0.3 mg/l</td>
</tr>
</tbody>
</table>

### Introduction

- The water quality-related parameters 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).
- The CWQI may be categorized into five classes, i.e., (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).
- 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.

---

**F_1(Scope)** = \( \frac{\text{Number of failed variables}}{\text{Total number of variables}} \) \times 100

\[ F_1 = \frac{\text{Number of failed variables}}{\text{Total number of variables}} \times 100 \]

**F_2(Frequency)** = \( \frac{\text{Number of failed tests}}{\text{Total number of tests}} \) \times 100

\[ F_2 = \frac{\text{Number of failed tests}}{\text{Total number of tests}} \times 100 \]

\[ F_3(Frequency) = \frac{\text{Number of failed tests}}{\text{Total number of tests}} \times 100 \]

\[ F_4(Scope) = \left( \frac{\text{Number of failed variables}}{\text{Total number of variables}} \right) \times 100 \]

\[ F_5(Exclusion) = \left( \frac{\text{Failed Test Value}}{\text{Objective}} \right) - 1 \quad \text{(if test value < objective)} \]

\[ F_6(Exclusion) = \left( \frac{\text{Objective}}{\text{Failed Test Value}} \right) - 1 \quad \text{(if test value > objective)} \]

Normalized sum of excursions, \( nse = \sum_{i=1}^{# \text{of tests}} \text{excursion}_i \)

\[ nse = \frac{\sum_{i=1}^{# \text{of tests}} \text{excursion}_i}{\# \text{of tests}} \]

\[ F_3(Exclusion) = \left( \frac{\text{Failed Test Value}}{\text{Objective}} \right) - 1 \quad \text{(if test value < objective)} \]

\[ F_4(Exclusion) = \left( \frac{\text{Objective}}{\text{Failed Test Value}} \right) - 1 \quad \text{(if test value > objective)} \]

**F_3(Amplitude)** = \( 0.01 nse + 0.01 \)

\[ F_3(Exclusion) = \left( \frac{\text{Failed Test Value}}{\text{Objective}} \right) - 1 \quad \text{(if test value < objective)} \]

\[ F_4(Exclusion) = \left( \frac{\text{Objective}}{\text{Failed Test Value}} \right) - 1 \quad \text{(if test value > objective)} \]

\[ F_3(Exclusion) = \left( \frac{\text{Failed Test Value}}{\text{Objective}} \right) - 1 \quad \text{(if test value < objective)} \]

\[ F_4(Exclusion) = \left( \frac{\text{Objective}}{\text{Failed Test Value}} \right) - 1 \quad \text{(if test value > objective)} \]

\[ CWQI = \left( \frac{F_1^2 + F_2^2 + F_3^2}{1.732} \right) \]
Objectives

- Develop clusters for major rivers in Alberta on the basis of monthly water quality data;
- Evaluate the clusters using Canadian Water Quality Index (CWQI) system;
- Apply clusters for spatio-temporal analysis; and
- Study the impact of climatic factor (i.e., snow-melting) and land use activities on the water quality of the rivers.

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)

Study Area and Data Requirements

- Monthly values of the 17 parameters for the period of 2004-2008 from 12 major rivers at 23 sampling sites
- Land use/cover map (MODIS-based annual composite at 1 km spatial resolution).
- Snow melting time period map (MODIS-derived at 500 m spatial resolution during 2008).

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Akbar et al. (2013) executed the following four steps for the development of clusters:

- Normalizing water quality data,
- Obtaining dominant parameters,
- Developing total exceedance model, and
- Identifying the cluster patterns.

Data was normalized for WT, TUR, TC, DOC, TP, TN, TDS, TH, Cl, SO₄, pH > 8.5, Na, F, Mn and Fe using the following expression:

\[
\text{Normalization} = \left( \frac{\text{measured}}{\text{guideline}} \right)^{0.25}
\]

Data was normalized for DO and pH < 6.5 using the following expression:

\[
\text{Normalization} = \left( \frac{\text{guideline}}{\text{measured}} \right)^{0.25}
\]

The power of a constant number (i.e., 0.25) was used in both expressions to reduce the spread between the parameters due to large variations in their measured values. Also, as the guideline was 0 for FC, so it was normalized using the exponent equal to 0.25.
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. In fact, PCA may be used to reduce the parameter/variable of a dataset. Thus, Akbar et al. (2013) used PCA to identify the major PCs and obtain the dominant parameters using the normalized data. 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).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PC-1</th>
<th>PC-2</th>
<th>PC-3</th>
<th>PC-4</th>
<th>PC-5</th>
<th>PC-6</th>
<th>PC-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO</td>
<td></td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUR</td>
<td></td>
<td></td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOC</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDS</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH&gt;8.5</td>
<td></td>
<td></td>
<td></td>
<td>-0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH&lt;6.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO₄</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td>Mn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.96</td>
</tr>
<tr>
<td>Var. (%)</td>
<td>31.5</td>
<td>20.8</td>
<td>12.6</td>
<td>9.1</td>
<td>6.1</td>
<td>5.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Cum. (%)</td>
<td>31.5</td>
<td>52.3</td>
<td>64.9</td>
<td>74.0</td>
<td>80.1</td>
<td>85.7</td>
<td>89.1</td>
</tr>
</tbody>
</table>

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Obtaining Dominant Parameters:
Use of Principal Component Analysis

PCA analysis of seven principal components (PCs) revealed the seven dominant parameters, and their indicators.

<table>
<thead>
<tr>
<th>No.</th>
<th>PCA</th>
<th>Parameter</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PC-1</td>
<td>TDS</td>
<td>watershed geology (<em>Anderson, 1999</em>)</td>
</tr>
<tr>
<td>2</td>
<td>PC-2</td>
<td>TC</td>
<td>natural and anthropogenic mineralization (<em>Anderson, 1999; Wolfe et al., 2007</em>)</td>
</tr>
<tr>
<td>3</td>
<td>PC-3</td>
<td>pH</td>
<td>anthropogenic activities related to different types of land use/cover (<em>Bruneau et al., 2009</em>)</td>
</tr>
<tr>
<td>4</td>
<td>PC-4</td>
<td>Fe</td>
<td>natural mineralization (<em>Anderson, 1999</em>)</td>
</tr>
<tr>
<td>5</td>
<td>PC-5</td>
<td>FC</td>
<td>anthropogenic activities, like PC-3 (<em>Bruneau et al., 2009</em>)</td>
</tr>
<tr>
<td>6</td>
<td>PC-6</td>
<td>DO</td>
<td>natural mineralization, like PC-4 (<em>Anderson, 1999</em>)</td>
</tr>
<tr>
<td>7</td>
<td>PC-7</td>
<td>TUR</td>
<td>anthropogenic activities, like PC-3 &amp; PC-5 (<em>Bruneau et al., 2009</em>)</td>
</tr>
</tbody>
</table>

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)

Developing Total Exceedance Model and Identifying the Cluster Patterns

- The normalized values of dominant parameters were used to calculate the total exceedance for each monitoring day during the period 2004-2008.

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

- Upon calculating the total exceedance values, Akbar et al. (2013) used 70% of the total exceedance data with the respective CWQI-values (as calculated using all the 17 parameters) to generate 5 clusters; and the remaining 30% of the data were used to validate the modelling schema.

- In generating the clusters, for example, for CQWI class 1, all the total exceedance values were considered to calculate the min, max, and mean-values; and named as “Cluster 1”. See the next slide for each of the 5 cluster-specific patterns.

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Developing Total Exceedance Model and Identifying the Cluster Patterns (2)

- Produced from minimum, maximum and mean of the exceedance values of dominant 7 parameters as derived using the PCA during the period 2004-2008, and shown in the figure.

- The exceedance values were calculated using the total exceedance equation shown in the last slide.

- Five clusters were generated using the five CWQI classes.

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)

Agreement Between Clusters and CWQI Classes

- Cluster development: 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.

- Cluster evaluation: 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.

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Comparison Between Clusters and CWQI Classes

Sampling site (BOR-1) of the Bow River during the period of 2004-2008.

Spatial and Temporal Trends (Bow River)

Four sampling sites (BOR-1, BOR-2, BOR-3, and BOR-4) of the Bow River using the clusters during the period of 2004-2008.
Land use/cover Classes and Snow Melting Periods (Bow River) (1)

Land cover:
- Water (1.87%)
- Shrubs (2.02%)
- Savannah (1.23%)
- Broad leaf crops (3.4%)
- Broad leaf forest (5.35%)
- Needle leaf forest (57.57%)
- Non vegetated (0.45%)
- Urban (0.17%)
- Grasses/cereal crops (30.11%)

Bow River

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)

Land use/cover Classes and Snow Melting Periods (Bow River) (2)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster Year</td>
<td>Cluster Year</td>
</tr>
<tr>
<td></td>
<td>3  2004, 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5  2005-2006, 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3  2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5  2007</td>
<td></td>
</tr>
<tr>
<td>BOR-4</td>
<td>3  2005</td>
<td>3  2004-2007</td>
</tr>
<tr>
<td></td>
<td>5  2006-2008</td>
<td></td>
</tr>
</tbody>
</table>

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)
Land use/cover Classes and Snow Melting Periods (Bow River) (3)

- In 2008, the change in clusters from winter to growing season for all sampling sites was related to snow melting period.
  - 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 contributed the deterioration of surface quality of Bow River in 2004-2007.
  - 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.
  - It is related to the agricultural activities of cereal and broad leaf crops as these three sites are located in adjacent agricultural areas.
  - In comparison, BOR-1 is located near needle leaf forests.

“Development of Remote Sensing Based Models for Surface Water Quality”

(Adopted from ©Akbar et al., 2013 licensed under CC BY 4.0)

(Adopted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)
Introduction

- In general, the understanding of the water quality plays a critical role prior to utilize for various purposes including drinking (Environment Canada 2012).

- Here, the aim was 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).

- Currently, the measured data of water quality variables at three 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) (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 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).

Objectives

- Akbar et al. (2014) evaluated remote sensing-based methods for acquiring CWQI and turbidity classes for assessing both spatial and temporal dynamics of the Bow River. The specific objectives were to:
  - Develop and evaluate remote sensing based models to acquire CWQI classes using the planetary reflectance of Landsat-5 TM and ground measured data;
  - Develop and evaluate remote sensing based models to retrieve turbidity using the planetary reflectance of Landsat-5 TM and in situ data; and
  - Apply the selected models to classify the source waters of the Bow River into CWQI and turbidity classes for spatial and temporal analysis.
Study Area

Data Used

- Used 31 scenes of Landsat-5 TM multispectral image acquired in different dates during the period 2006–2010.

- Used spectral bands: blue (B), green (G), red (R), and Near InfraRed (NIR).

- Ground measured data for 37 days at three sampling locations (BOR-1, BOR-2, BOR-3) of the Bow River in 2006–2010.
Image Processing

- Calculated normalized difference vegetation index (NDVI) as a measure of vegetation greenness:

\[
NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}
\]

where, \(\rho_{NIR}\) = reflectance of NIR band, and \(\rho_R\) = reflectance of R band,

- Calculations were performed over the sampling sites in all scenes of 37 data records to determine the possible contamination pixels from other landuses (e.g. roads, agriculture, vegetation, and barren land, etc.).

- The negative NDVI values (i.e. between 0 and -1) indicated the presence of water in the pixels.

- The positive NDVI values (i.e. between >0 and +1) showed the possible contamination due to other landuses, and considered the reflectance value of a nearest neighboring water pixel).

(Adopted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)

Models for CWQI and Turbidity from Planetary Reflectance Data

- Developed 26 individual empirical models in determining both CWQI and turbidity as a function of the spectral bands of B, G, R and NIR.


- Used regression analysis technique to obtain the quantitative relationship between the satellite based planetary reflectances and water quality variables to develop the empirical models.

- Used 23 data records (i.e. Landsat-5 TM as well as the ground data) for the development of models (calibration); and the remaining 14 data records were used to validate the selected best models.

- In all models, CWQI and turbidity were the dependent variables, whereas the bands were the independent variables.

(Adopted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)
Empirical Models for Determining CWQI Classes

○ Among the 26 models, the reflectance-values of R band showed the best results.

○ 10 data records matched 100% of the modeled values, whereas four data records had a deviation of 1 from the modeled values.

(Assorted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)

Empirical Models for Obtaining Turbidity Classes

○ Among the 26 models, the reflectance-values of R band showed the best results.

○ A strong correlation of modeled turbidity with the measured turbidity: $r^2 = 0.83$.

(Assorted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)
Maps of CWQI and Turbidity Classes

- The river water quality was deteriorated due to agricultural activities and climatic factors.

(Adopted from ©Akbar et al., 2014 licensed under CC BY-NC 4.0)

References (1)

References (2)


Sample Review Questions

- What are the factors influencing the water quality in rivers and lakes?
- What could be the potential sources of pollutants that impact the water quality?
- Is it possible to use lesser number of water quality-related parameters (i.e., instead of all 17 parameters) to produce equivalent CWQI-values? If yes, how?
- How would you identify the dominant water quality parameters from a set of given data, and identify the cluster patterns?
- Draw a schematic diagram to develop and evaluate:
  - water quality assessment framework with reduced number of parameters; and
  - a remote sensing based model to generate CWQI classes.