Multi-sensor Data Fusion and Mining for Sustainable Urban Water Infrastructure Systems Analysis

Ni-Bin Chang
Department of Civil, Environmental, and Construction Engineering, University of Central Florida

INTRODUCTION
SELECTED SITES
WATER QUALITY PARAMETERS & SURFACE REFLECTANCE
IDFM METHOD
RESULTS (STUDY SITE#1)
RESULTS (STUDY SITE#2)
RESULTS (STUDY SITE#3)

What is Remote Sensing?

Remote sensing is defined as the
• acquisition,
• processing,
• and interpreting
of images that are remotely obtained from sensors recording the interaction between electromagnetic energy and the Earth’s surface

Remote Sensing Platforms:
Ground-based
Airplane-based
Satellite-based
How can surface reflectance quantify water quality parameters?

- The amount of radiation that is emitted and reflected from an object is a function of wavelength.

Thus, materials can be identified based upon their unique spectral patterns.

SPECTRAL REFLECTANCE CURVES

INTRODUCTION

SELECTED SITES

WATER QUALITY PARAMETERS & SURFACE REFLECTANCE

IDFM METHOD

RESULTS (STUDY SITE#1)

RESULTS (STUDY SITE#2)

RESULTS (STUDY SITE#3)

METHOD COMPARISONS

Field Investigation

1. Analysis result is more accurate.
2. More variables can be analyzed.

Model Using RS Data

1. Get result in short time
2. Easily build up a long term monitoring system.

Disadvantage

1. Spend lots of money and manpower, and analysis in laboratory takes time.
2. Data only represents water quality at a point location.

1. Fewer variables can be analyzed.
2. Require field investigation data to adjust parameters in model.
3. Optical data is affected by clouds.

PHOTO COURTESY:
1. New York State Department of Environmental Conservation; 2. NASA; 3. Jackson Lieblein Team
Develop a cost-effective, forward-looking, and risk-informed nowcasting and forecasting scheme for water quality with the aid of remote sensing and sensor networks.

INTRODUCTION
SELECTED SITES
WATER QUALITY PARAMETERS & SURFACE REFLECTANCE
1DFM METHOD
RESULTS (STUDY SITE#1)
RESULTS (STUDY SITE#2)
RESULTS (STUDY SITE#3)

Remote Sensing Resolutions

- **Spatial Resolution**
  - The measure of the smallest object that can be detected by the sensor
  - Characterized by the pixel size
  - Images with high spatial resolution show greater detail

1. Vannah, 2013
2. Ahmed, 2010
**Temporal Resolution**

- The time required for a satellite to revisit the same location

**Spectral Resolution**

- The wavelength width of the recorded bands
- The finer the resolution, the narrower the wavelength range for a band
Assessment metrics and associated satellites for monitoring different water quality constituents

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Fusion levels</th>
<th>Enhanced properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Pixel</td>
<td>Spatial</td>
</tr>
<tr>
<td>IHS</td>
<td>Pixel</td>
<td>Spectral</td>
</tr>
<tr>
<td>HPF</td>
<td>Pixel</td>
<td>Spatial</td>
</tr>
<tr>
<td>Brovey Transform</td>
<td>Pixel</td>
<td>Spectral</td>
</tr>
<tr>
<td>Pyramid method</td>
<td>Pixel</td>
<td>Spectral</td>
</tr>
<tr>
<td>Wavelet transform</td>
<td>Pixel</td>
<td>Spatial</td>
</tr>
<tr>
<td>Curvelet transform</td>
<td>Feature</td>
<td>Spectral</td>
</tr>
<tr>
<td>HIS-Wavelet</td>
<td>Pixel</td>
<td>Spatial/Spectral</td>
</tr>
<tr>
<td>PCA-Wavelet</td>
<td>Pixel</td>
<td>Spatial</td>
</tr>
<tr>
<td>Semi-physical fusion</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>Unmixing based fusion</td>
<td>Pixel</td>
<td>Spatial/Spectral</td>
</tr>
<tr>
<td>STARFM</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>ESTARFM</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>STAARCH</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>SPSSTFM</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>STDFM</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>ESTDFM</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>Bilinear mixing model</td>
<td>Pixel</td>
<td>Spatial/Temporal</td>
</tr>
<tr>
<td>One-pair Image learning</td>
<td>Pixel</td>
<td>Spatial</td>
</tr>
</tbody>
</table>
**INTRODUCTION**

SELECTED SITES

WATER QUALITY PARAMETERS & SURFACE REFLECTANCE

IDFM METHOD

RESULTS (STUDY SITE#1)

RESULTS (STUDY SITE#2)

RESULTS (STUDY SITE#3)

**The chronological order of different types of the developed image fusion methods**

- **Spatial Domain**
  - 1980s: Brovey Transform
  - 1990s: PCA, IHS, HPF
  - 2000s: PCA, Wavelet, IHS, Wavelet, HPF Wavelet

- **Transform Domain**
  - Pyramid Transform
  - Wavelet Transform
  - Curvelet Transform

- **Unmixing-Based**
  - STDFM
  - ESTDFM

- **STAR-FM-Based Algorithm**
  - STDFM
  - ESTDFM
  - STAFM
  - SPSTFM

***Red arrows show the complementary effect of each image fusion method compare to previous one***

Chang et al. (2015)

---

**Landsat ETM+ Bandwidth and MODIS Bandwidth**

<table>
<thead>
<tr>
<th>Landsat ETM+ Band</th>
<th>ETM+ Bandwidth (nm)</th>
<th>MODIS Land Band</th>
<th>MODIS Bandwidth (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>450-520</td>
<td>3</td>
<td>459-479</td>
</tr>
<tr>
<td>2</td>
<td>530-610</td>
<td>4</td>
<td>545-565</td>
</tr>
<tr>
<td>3</td>
<td>630-690</td>
<td>1</td>
<td>620-670</td>
</tr>
<tr>
<td>4</td>
<td>780-900</td>
<td>2</td>
<td>841-876</td>
</tr>
<tr>
<td>5</td>
<td>1550-1750</td>
<td>6</td>
<td>1628-1652</td>
</tr>
<tr>
<td>7</td>
<td>2090-2350</td>
<td>7</td>
<td>2105-2155</td>
</tr>
</tbody>
</table>

MODIS Terra and Landsat 5 TM/7 ETM+ satellite imagery can be acquired free of charge.

MODIS has coarse spatial resolution.

Landsat has a lengthy 16 day revisit time.
One of challenges in water quality monitoring by remote sensing is associated with selecting a suitable inversion method to find the strong relationship between water quality parameters and remotely sensed data.

**Which inversion model?**

Empirical Methods

Analytical Methods

Semi-Analytical/Semi-Empirical Methods

MODIS Terra and Landsat 5 TM/7 ETM+ satellite imagery can be acquired free of charge.

MODIS has coarse spatial resolution.

Landsat has a lengthy 16 day revisit time.
INTRODUCTION

FACTORS AFFECTING WATER QUALITY
DATA SOURCE
SPATIAL AND TEMPORAL CHANGE OF TSS
IDFM ALGORITHM
RELIABILITY OF IDFM RESULTS
LINKAGE BETWEEN TSS LEVELS AND FOREST FIRES

Inversion Methods and Retrieval Algorithms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Empirical Methods</td>
<td>• Analytical Methods</td>
<td>• Semi-Empirical Method</td>
</tr>
<tr>
<td>• Statistic Regression</td>
<td>• MIM</td>
<td>• Semi-Empirical Algorithm</td>
</tr>
<tr>
<td>• ANN</td>
<td></td>
<td>• Semi-Analytical Garver-Siegel-Maritorena (GSM) Model</td>
</tr>
<tr>
<td></td>
<td>• Purely Analytical Method</td>
<td>• Semi-Analytical Method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Semi-Analytical Remote Sensing Method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Modifed (GSM) Semi-Analytical Inversion Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Semi-Analytical Method</td>
</tr>
</tbody>
</table>

The chronological trend of applying inversion methods for monitoring water quality by remote sensing

1990s
- Statistical Regression
- AN

2000s
- Statistical Regression
- GP
- Radial Basis Function Neural (RBFN) Network
- MIM
- Back-Propagation (BP) Artificial Neural Network
- Parallel Grammatical Evolution and Genetic Algorithm (GEG)

2010s
- Fuzzy Integral Based Fusion Method
- AN
- GP
- Genetic Algorithm and Partial Least Square (GA-PLS)
- Bio-optical model
- Analytical Optical Modeling
- Analytical Model of Underwater light-climate
- Semi-Empirical Method

The Integrated Data Fusion and Mining (IDFM) Technique

LANDSAT or MERIS Images → ArcGIS → MODIS Images

Matlab → STARFM

Ground Truthing Data → GP (Discipulus) → Nutrient Output Maps

STARFM = Spatial and Temporal Adaptive Reflectance Fusion Model
**GENETIC OPERATION: MUTATION**

- Mutation is an alteration that change the genetic structure.
ANIMATION: MUTATION

Inversion Methods and Retrieval Algorithms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll&lt;sub&gt;abs&lt;/sub&gt;</td>
<td>CDOM</td>
<td>Chlorophyll&lt;sub&gt;abs&lt;/sub&gt;</td>
</tr>
<tr>
<td>Pigment Concentration</td>
<td>Phytolanium</td>
<td>SDD</td>
</tr>
<tr>
<td>TSS</td>
<td>Chlorophyll&lt;sub&gt;abs&lt;/sub&gt;</td>
<td>CDOM</td>
</tr>
<tr>
<td>Chlorophyll&lt;sub&gt;abs&lt;/sub&gt;</td>
<td>TSS</td>
<td>Dissolved Organic Matter</td>
</tr>
<tr>
<td>TSS</td>
<td>DO</td>
<td>Cyanobacteria</td>
</tr>
<tr>
<td>TP</td>
<td>TP</td>
<td>SDD</td>
</tr>
<tr>
<td>TN</td>
<td>Sediment</td>
<td>Chlorophyll&lt;sub&gt;abs&lt;/sub&gt;</td>
</tr>
<tr>
<td>Turbidity</td>
<td>Turbidity</td>
<td>CDOM</td>
</tr>
<tr>
<td></td>
<td>TP</td>
<td>Phytoplankton</td>
</tr>
<tr>
<td></td>
<td>TSS</td>
<td>CDOM</td>
</tr>
<tr>
<td></td>
<td>CDOM</td>
<td>TSS</td>
</tr>
<tr>
<td></td>
<td>Tripton(TR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SDD</td>
<td></td>
</tr>
</tbody>
</table>

Application of different inversion methods for monitoring differing water quality constituents
Water quality mapping by remote sensing consists of 5 major tasks:

1. To acquire satellite images
2. To prepare the images for data fusion
3. Data fusion
4. Assimilating the ground truth data and fused image bands data into linear or non-linear model
5. Pixel based concentration map generation
INTRODUCTION
SELECTED SITES
WATER QUALITY PARAMETERS & SURFACE REFLECTANCE
IDFM METHOD
RESULTS (STUDY SITE#1)
RESULTS (STUDY SITE#2)
RESULTS (STUDY SITE#3)

Study Site #1 Lake Harsha

- Total Trihalomethanes (TTHMs) and Haloacetic Acids (HAA5s) above Drinking Water Standards may cause **public health impact**.

- The presence of TOC is a reliable **indicator of TTHMs and HAA5s**.

- The removal of TOC is **costly in drinking water treatment plants**.

- A near real–time monitoring system using remote sensing is in an acute need.
Study Site #2: Lake Erie

- The eutrophic conditions in Lake Erie favor the formation of algal blooms, some of which are toxic due to the presence of Microcystis—a cyanobacteria which produces the hepatotoxin microcystin. **Harmful algal blooms present a danger to both the environment and human health.**

- The aberrant toxicity of microcystin is accredited to their strong ability to bond to protein phosphates.

- Acute exposure to microcystin can lead to liver cancer, liver failure, and death. Minor yet more common symptoms include rashes, vomiting, diarrhea, and tissue inflammation.

- The WHO has listed 1 µg·L⁻¹ of microcystin as an acceptable level for infrequent consumption, yet it warns that there is a significant health concern for chronic exposure to low concentrations.

Study Site #3: Lake Mead

- 98 percent of inflow into the Lake Mead is originated from Colorado River which transports discharged sediments from the Colorado River Basin. Amount of sediment transported into the Lake Mead from the upper Colorado River has declined by 90% since the completion of the Glen Canyon Dam in 1963.

- Forest fire events in the recent decade in the upper part of the lower Virgin River watershed, having caused drastic land-use and land-cover changes.

- Las Vegas Wash transports urban stormwater runoff and treated wastewater effluent into Lake Mead and deteriorates the water quality in the lower Las Vegas Wash.

**Effects of TSS**

- Increases water temperature
- Decreases levels of dissolved oxygen
- High TSS in source water is always a threat leading to delay filtration process in the water treatment plant by clogging filters and disruption of the conventional treatment processes.
Remote Sensing of Organic Carbon

- Chromophoric dissolved organic matter (CDOM) is the light absorbing fraction of dissolved organic carbon (DOC).

<table>
<thead>
<tr>
<th>CDOM Spectral Peaks</th>
<th>Sampling Locations</th>
<th>Sampling Instrument</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>570</td>
<td>25</td>
<td>Spectron Engineering SE-590 spectrometer</td>
<td>Menken et al., 2005</td>
</tr>
<tr>
<td>550, 571, 670, 710</td>
<td>8</td>
<td>Scanning spectrophotometer</td>
<td>Arenz et al., 1995</td>
</tr>
<tr>
<td>560, 650-700</td>
<td>5+</td>
<td>Spectron Instrument CE395 spectroradiometer</td>
<td>Vertucci, 1989</td>
</tr>
</tbody>
</table>

- The majority of the sampling locations in these studies were lakes and reservoirs. The sites were chosen for their high levels of CDOM.
Remote Sensing of Microcystin and its Indicators

- Chlorophyll-a and phycocyanin concentrations are directly correlated with microcystin levels.
- Spectral reflectance curves for chlorophyll-a (top curve) and a mix of chlorophyll-a and phycocyanin (bottom curve) are presented.
- The first three bands of Landsat TM are shown as dotted lines (Vincent et al., 2004)

Wang et al. (2007) presented a laboratory study to assess the relationship between different concentrations of suspended sediments with reflectance spectra, and they found out that wavelengths in the range of b1 (430-500 nm) and b3 (670-735 nm) are defined as the best wavelength range to determine the lower suspended sediment concentration, while the wavelength range between b4 (780-835 nm) is considered the optimal range for high concentration of suspended sediments.

Ritchie et al. (1976) which identified an optimal wavelength range between (700-800 nm).
Theoretical Hyperspectral Advantages

INTRODUCTION
SELECTED SITES
WATER QUALITY PARAMETERS & SURFACE REFLECTANCE
IDFM METHOD
RESULTS (STUDY SITE#1)
RESULTS (STUDY SITE#2)
RESULTS (STUDY SITE#3)

Star-FM

• Input Data Streams:
  2 MODIS
  1 Landsat
• STARFM Output:
  1 Fused Image

• Goodness of Fit:
  • $R = 0.7696$
  • $R^2 = 0.5892$

• This process allows for near real-time monitoring and event detection
**Star-FM**

**Input Data Streams:**
- 3 MODIS
- 2 Landsat
- STARFM Output:
  - 1 Fused Image

**Goodness of Fit:**
- \( R = 0.9098 \)
- \( R^2 = 0.8278 \)

---

**Satellites Sensors**

**Multi-spectral and Hyper-spectral**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Hyperspectral Sensor Pair</th>
<th>Multispectral Sensor Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MERIS</td>
<td>MODIS TERRA (ocean bands)</td>
</tr>
<tr>
<td></td>
<td>MODOCL2</td>
<td>Landsat TM</td>
</tr>
<tr>
<td>Product</td>
<td>MER_FR_2P</td>
<td>LT5</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td>300 m</td>
<td>250/500 m</td>
</tr>
<tr>
<td>Temporal Resolution</td>
<td>1-3 days</td>
<td>16 days</td>
</tr>
</tbody>
</table>

Comparison between Multi-spectral and Hyper-spectral product (adapted from Vannah, 2013)
INTRODUCTION

SELECTED SITES

WATER QUALITY PARAMETERS & SURFACE REFLECTANCE

IDFM METHOD

RESULTS (STUDY SITE #1)

RESULTS (STUDY SITE #2)

RESULTS (STUDY SITE #3)

* Input Data Streams:
  * 3 MODIS
  * 2 Landsat

* STARFM Output:
  * 1 Fused Image

<table>
<thead>
<tr>
<th></th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.89</td>
<td>0.71</td>
<td>0.68</td>
</tr>
</tbody>
</table>

INTRODUCTION

SELECTED SITES

WATER QUALITY PARAMETERS & SURFACE REFLECTANCE

IDFM METHOD

RESULTS (STUDY SITE #1)

RESULTS (STUDY SITE #2)

RESULTS (STUDY SITE #3)

![Fused Image](image)

11 Oct 2006

27 Oct 2006

A

B

C

D

E

F

Moderate

MODIS

Landsat

Results

Study Site #1:
Lake harsha
**TOC: GP MODEL PREDICTIONS**

- The fused GP model outperforms the unfused MODIS image at predicting peak TOC values.

![Graph showing observed and predicted TOC concentrations](image)

**TOC: STATISTICAL ANALYSIS**

- Inversion model comparison:
  - The more computationally powerful data mining methods yield more accurate TOC predictions in the inland water body.
- Data fusion analysis:
  - The fused model outperforms the unfused MODIS GP model.

<table>
<thead>
<tr>
<th>Metric</th>
<th>2-Band</th>
<th>MODIS GP</th>
<th>Fusion-based GP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOC Obs mean (mg·L⁻¹)</td>
<td>7.276</td>
<td>7.276</td>
<td>7.276</td>
</tr>
<tr>
<td>TOC Pred Mean (mg·L⁻¹)</td>
<td>7.276</td>
<td>7.085</td>
<td>7.433</td>
</tr>
<tr>
<td>Percent Difference of the Means (%)</td>
<td>0.000</td>
<td>2.629</td>
<td>-2.166</td>
</tr>
<tr>
<td>Root Mean Square Error (mg·L⁻¹)</td>
<td>1.716</td>
<td>1.248</td>
<td>0.900</td>
</tr>
<tr>
<td>Ratio of St. Dev.</td>
<td>0.394</td>
<td>0.851</td>
<td>0.855</td>
</tr>
<tr>
<td>Mean Percent Error (%)</td>
<td>5.046</td>
<td>-0.448</td>
<td>3.921</td>
</tr>
<tr>
<td>Square of the Pearson Product Moment Correlation Coefficient</td>
<td>0.1974</td>
<td>0.5628</td>
<td>0.7680</td>
</tr>
<tr>
<td>Computational Time (min)</td>
<td>&lt;1</td>
<td>1800</td>
<td>4200</td>
</tr>
</tbody>
</table>
SPATIOTEMPORAL TOC VARIATIONS

- Seasonal TOC maps provide a visual representation into the dynamics of TOC throughout the year.
- Per literature, winter had one of the lowest TOC values, due to low temperatures and decreased microbial activity.
- Agren et al. (2008) observed spring TOC values peaking slightly, yet late summer and fall TOC values were still higher due to increased temperature and increased precipitation.
- This study shared similar results with the exception of summer TOC values being as low as winter values.

<table>
<thead>
<tr>
<th>Season</th>
<th>TOC (mg L⁻¹)</th>
<th>Season</th>
<th>TOC (mg L⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>8.0</td>
<td>Summer</td>
<td>7.2</td>
</tr>
<tr>
<td>Fall</td>
<td>7.4</td>
<td>Winter</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Results
Study Site #2: Lake Erie

47 days of interruption of water supply at Toledo in Aug. 2014.
Inversion model comparison:
- The more computationally powerful data mining methods yield more accurate TOC predictions in the inland water body.

Multispectral vs. Hyperspectral:
- Hyperspectral has longer model training times.
- Hyperspectral is more accurate overall.

<table>
<thead>
<tr>
<th></th>
<th>Fused Multispectral Input*</th>
<th>Fused Hyperspectral Input**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two-Band Ratio Model</td>
<td>Spectral Slope Model</td>
</tr>
<tr>
<td>Observed Microcystin Mean (µg·L⁻¹)</td>
<td>0.6718</td>
<td>0.6718</td>
</tr>
<tr>
<td>Predicted Microcystin Mean (µg·L⁻¹)</td>
<td>2.226</td>
<td>0.1792</td>
</tr>
<tr>
<td>Root Mean Square Error (µg·L⁻¹)</td>
<td>1.348</td>
<td>1.340</td>
</tr>
<tr>
<td>Ratio of St. Dev.</td>
<td>0.8270</td>
<td>0.1238</td>
</tr>
<tr>
<td>Mean Percent Error (%)</td>
<td>87.57</td>
<td>5.251</td>
</tr>
<tr>
<td>Square of the Pearson Product Moment Correlation Coefficient</td>
<td>0.02393</td>
<td>0.09625</td>
</tr>
<tr>
<td>Computational Time (Seconds)</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

- Multispectral Time Series (TOP):
  - Accurate peak value prediction.
  - Necessary for detecting hazardous blooms.
  - Unable to predict low microcystin values.

- Hyperspectral Time Series (Bottom):
  - Accurate peak value prediction.
  - More accurate predictions at low values.
  - Useful for observing bloom formation.
SPATIOTEMPORAL MICROCYSTIN VARIATIONS

(A) Hyperspectral GP model
(B) Multispectral GP model
(C) MODIS true color

- Notice the unique features identified by the finer spatial resolution in Figure B.

TRACKING MICROCYSTIS BLOOMS (NOAA HAB BULLETIN)
IDFM BLOOM IDENTIFICATION

For a functional early warning system, IDFM should be able to identify a toxic HAB throughout all stages of its life cycle:

- Clear water
- IDFM properly determines that no toxic HABs are present
- Bloom formation and expansion
- Tracking bloom's movement due to wind and currents

The perfect storm (July 16th, 2011). Factors leading to one of the largest HABs:

- Low rainfall in 2010 led to nutrient accumulation
- Fertilizer sales hit a decadal high in 2010
- Spring 2011 was the wettest in record

In conclusion, heavy precipitation flushed considerable loads of nutrients into Maumee bay, leading to one of the largest HABs in Lake Erie.

IDFM BLOOM TRACKING

MODIS image on 31-Jul-2011

MODIS image on 12-Sep-2011

Microcystin Map (μg L⁻¹)
Results
Study Site #3: Lake Mead
R-squared value between the spectral reflectance value of the fused image and the actual Landsat image

<table>
<thead>
<tr>
<th>Band</th>
<th>Band1</th>
<th>Band2</th>
<th>Band3</th>
<th>Band4</th>
<th>Band5</th>
<th>Band7</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.64</td>
<td>0.76</td>
<td>0.80</td>
<td>0.85</td>
<td>0.83</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Reliability of the produced fused images is sufficient for the application in machine learning.

Assess the performance of the applied data mining techniques

<table>
<thead>
<tr>
<th>Model</th>
<th>Training R</th>
<th>RMSE</th>
<th>Testing R</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>0.54</td>
<td>0.73</td>
<td>0.51</td>
<td>0.75</td>
</tr>
<tr>
<td>ANN</td>
<td>0.68</td>
<td>0.62</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>ELM</td>
<td>0.69</td>
<td>0.60</td>
<td>0.20</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Training: ELM and ANN show the best results.

Testing: Error of ELM is unacceptable.

Testing: Although GP model showed the best results, the error is just slightly less than the ANN model.

Selected Model: ANN model was selected for the forecasting procedure.

COMPARING THE PERFORMANCE OF SELECTED DATA MINING METHODS USING STATISTICAL INDICES

COMPARISON BETWEEN THE ESTIMATED AND OBSERVED VALUES OF TOC AT LVB 6.7 DURING AUG 14th - OCT 14th 2001

<table>
<thead>
<tr>
<th>Date</th>
<th>Observed TOC (mg.L⁻¹)</th>
<th>Estimated TOC (mg.L⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28-Aug-2001</td>
<td>4.69</td>
<td>3.9</td>
</tr>
<tr>
<td>4-Sep-2001</td>
<td>3.98</td>
<td>3.86</td>
</tr>
<tr>
<td>25-Sep-2001</td>
<td>3.83</td>
<td>3.44</td>
</tr>
<tr>
<td>7-Oct-2001</td>
<td>3.74</td>
<td>4.13</td>
</tr>
</tbody>
</table>
### Reference Method No. of hidden neurons

<table>
<thead>
<tr>
<th>Reference</th>
<th>Method</th>
<th>No. of hidden neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang (2003)</td>
<td>$N_h = 2\sqrt{(m + 2)N}$</td>
<td>120</td>
</tr>
<tr>
<td>Jinchuan and Xinzhe (2008)</td>
<td>$N_h = \frac{(n + N)}{L}$</td>
<td>20</td>
</tr>
<tr>
<td>Shibata and Ikeda (2009)</td>
<td>$N_h = \sqrt{n \times m}$</td>
<td>2</td>
</tr>
<tr>
<td>Arai (1993)</td>
<td>$N_h = 2^n/3$</td>
<td>20</td>
</tr>
<tr>
<td>Wang et al. (2013)</td>
<td>$\frac{nN}{n + m} \leq N_h \leq N$</td>
<td>180</td>
</tr>
</tbody>
</table>

$N_h$ is the optimal number of neurons in the hidden layer, $m$ is number of outputs, $N$ is number of samples, $L$ is number of hidden layers, and $n$ is number of inputs.

### Finding the optimal number of neurons

**R-SQUARED VALUES CORRESPONDING TO DIFFERENT NUMBER OF NEURONS**

Number of neurons equal to 180 leads to the highest R-squared value between the trained outputs and the desired ultimate output of ANN model.

### Assessing the performance of the nowcasting model

**Cloudy Days**
Introduction

Problem Statement, and research contribution

Research Objective

Proposed Methodology

Case Study

Results and Conclusion

TOC Measurement

<table>
<thead>
<tr>
<th>Station</th>
<th>LWLVB-B</th>
<th>LVB6.7</th>
<th>LVB2.7</th>
<th>CR346.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Value (mg.L⁻¹)</td>
<td>4.598</td>
<td>2.748</td>
<td>3.993</td>
<td>2.607</td>
</tr>
<tr>
<td>Estimated Value (mg.L⁻¹)</td>
<td>TOC&gt;4</td>
<td>2&lt; TOC &lt;4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The developed nowcasting model is able in to estimate the defined range that the values would fall into.

Scenario | 1 | 2 | 3 | 4 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral reflectance Value</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Meteorological Parameters</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water Level in the Reservoir</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assessing the performance of the nowcasting model

The developed nowcasting model is able in to estimate the defined range that the values would fall into.

Sensitivity Analysis

SUMMARY OF THE RESULTS OF ERROR INDICES

<table>
<thead>
<tr>
<th>Scenario</th>
<th>R²</th>
<th>RMSE</th>
<th>MAE</th>
<th>PBIAS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.78</td>
<td>0.24</td>
<td>0.13</td>
<td>1.38</td>
</tr>
<tr>
<td>2</td>
<td>0.81</td>
<td>0.26</td>
<td>0.12</td>
<td>-0.45</td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.23</td>
<td>0.10</td>
<td>-0.91</td>
</tr>
<tr>
<td>4</td>
<td>0.83</td>
<td>0.35</td>
<td>0.13</td>
<td>-1.06</td>
</tr>
</tbody>
</table>

COMPARISON OF PREDICTED VALUES IN DIFFERENT SCENARIOS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.89</td>
<td>2.76</td>
<td>4.68</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>3.49</td>
<td>3.02</td>
<td>4.18</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>3.67</td>
<td>2.91</td>
<td>4.79</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>2.98</td>
<td>2.00</td>
<td>4.26</td>
<td>0.77</td>
</tr>
</tbody>
</table>

To reduce the uncertainty, multiple runs (10 runs) for cross-validation were performed for each scenario, and the average of calculated error indices over the rounds were reported.

Scenario 1 based on only spectral reflectance value showed the lowest R-squared value and the highest errors.

Scenario 4 including meteorological and reservoir elevation was not result in the best outcome.

Scenario 3 showed the highest R-squared value and the lowest errors.

This finding rules out the importance of the water level in the reservoir.

RECORDED GROUND TRUTH DATA

(4.74 mg/L)
INTRODUCTION
SELECTED SITES
WATER QUALITY PARAMETERS & SURFACE REFLECTANCE
IDFM METHOD
RESULTS (STUDY SITE#1)
RESULTS (STUDY SITE#2)
RESULTS (STUDY SITE#3)

Effect of each inflow on TSS concentration

VR25.1

Colorado River

Overton Arm

LVB2.7

Las Vegas Wash

Estimated TSS Concentration (mg/L)

Observed TSS Concentration (mg/L-1)

R² = 0.995

Observed
Model-ANN
TOC Concentration at Each Arm
Based on Ground-Truth Data

Estimated TOC Concentration at Each Arm

White Color indicates no data value
Future Work & Perspectives

Smart Information Reconstruction (SMIR) via Time-Space-Spectrum Continuum for Cloud Removal in Satellite Images

Spectral Information Adaptation and Simulation Scheme (SIASS) for Merging Cross-mission Ocean Color Reflectance Observations from VIIRS and MODIS.

Big Data Issues

Thank You!

Ni-Bin Chang
nchang@ucf.edu