A Novel approach for monitoring cyanobacterial blooms using an ensemble based system from MODIS imagery downscaled to 250 meters spatial resolution

Anas El Alem[1] (anas.el_alem@ete.inrs.ca), **Karem Chokmani[1], Isabelle Laurion[1] and Salah-Eddine El Adlouni[2]**

[1]Centre Eau Terre Environnement, INRS, 490 De la Couronne Street, Québec, QC G1K 9A9

[2]Mathematics and Statistics Department, Moncton University, 18 Antonine-Maillet Avenue, Moncton, NB E1A 3E9

Context

Due to standard sampling programs limitations (spatial distribution and temporal frequency [1]), remote sensing data have become increasingly used for monitoring Harmful algae blooms (HAB), called also Cyanobacterial blooms (CB), in freshwaters by detecting the bio-optical activity of their principal pigment, the Chlorophyll-a (Chl-a). Recent studies demonstrated that waters loaded in Chl-a exhibit different apparent optical properties enabling them to be discriminated and used different classification approaches for water surveillance [2]. Such classifiers are however local and unstable [3]. On the other hand, Classifiers based on ensemble systems are more general and stable [4], but their biggest limitation is time consumption due to their conceptualization based on re-sampling techniques. To reduce this time consumption, we used the Gaussian quadrature formula which has the potential to convert this re-sampling problem to probabilistic numerical calculations that are simpler and are accurate and approved [5].

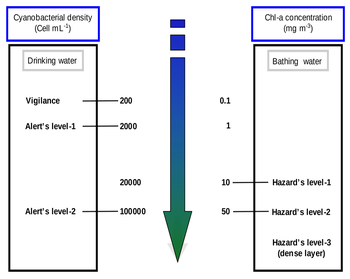
Objective

The objective of this study was then to develop a GAussian Quadrature Ensemble Classifier (GAQEC) and to compare its performance to Classification and regression tree (CART) in a context of HAB monitoring in Quebec freshwaters using MODIS images downscaled to 250 m spatial resolution. The performance of the two approaches was evaluated based on the same validation database using confusion matrices for a classification accuracy assessment and a standard-deviation computation for a robustness analysis.

Study area and calibration data

The calibration data set used in this work was collected by the environmental Ministry of Quebec called *MDDELCC* (*Ministère du Développement Durable, Environnement et Lutte contre les Changements Climatiques*), between the years 2004 and 2010 over 22 freshwaters. This data set is composed by three classes of cyanobacterial density: 1) waters with cell densities lower than 20,000 cells mL^{-1}, 2) waters with densities between 20,000 and 100,000 cells mL^{-1}, and 3) waters with densities higher than 100,000 cells mL^{-1}.

Geographic location of the water bodies used for model calibration and validation



Chl-a concentration and cyanobacterial densitie thresholds used in the decision tree published by the World health organization

Pre-processing images

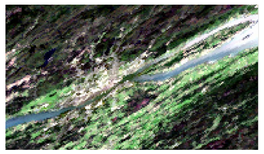
Concurrent MODIS imageries to dates of *in situ* measurements were firstly loaded from the NASA's website [6] and were pre-processed as follow:

1.The HKM and QKM Level 1B product were used for downscaling the spatial resolution of the bands 3 to 7 to 250 meters;

2.The MOD03 Level 1B product was used to re-project images from SIN to LCC projection;

3.The MOD08_D3 Level 3 product was used for atmospheric correction using the SMAC (Simplified model for atmospheric correction) algorithm.

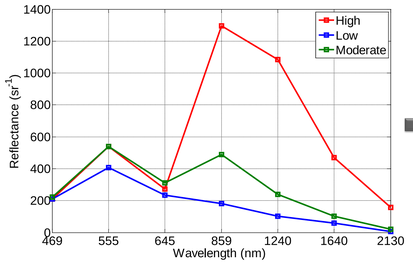
All pre-processing steps were done using a tool developed at the Canadian center for remote sensing [7].



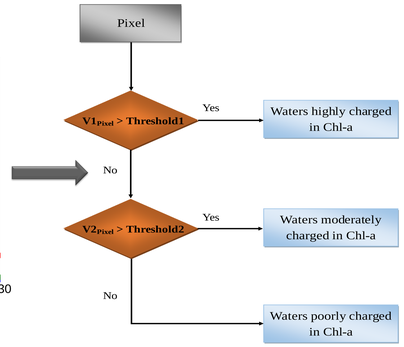
RGB MODIS original product (500 m spatial resolution)

RGB downscaled images (250 m spatial resolution)

Data analysis

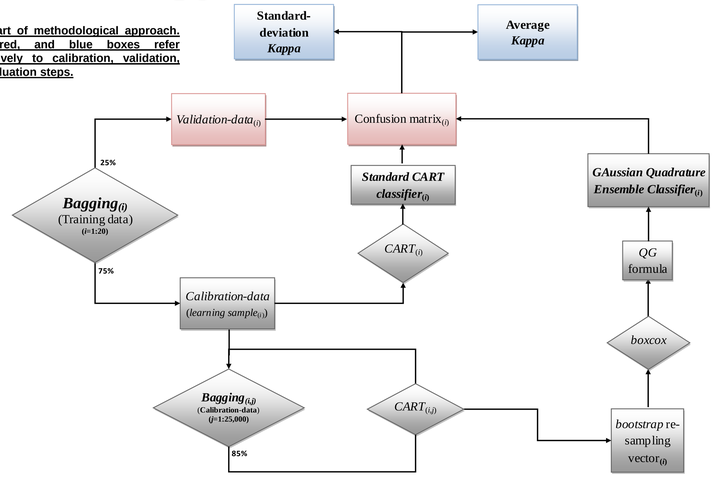


Averages of spectral signature picked up over training data for waters poorly, moderately, and highly charged in Chl-a.



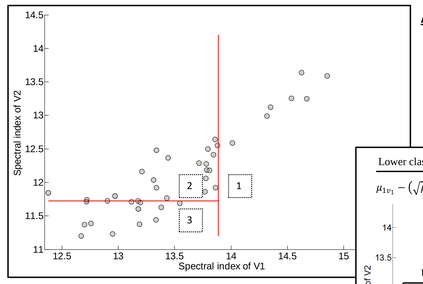
General Scheme of discrimination between three classes of water charged in Chl-a using the Classification and Regression Three algorithm.

Methodology



Flowchart of methodological approach. Gary, red, and blue boxes refer respectively to calibration, validation, and evaluation steps.

Methodology (continued)



CART classifier

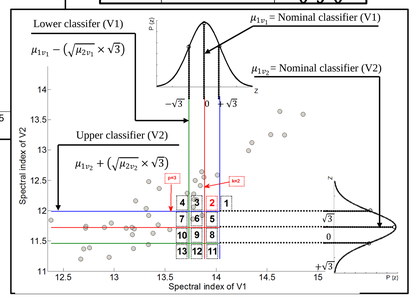
$$P(High) = \sum_{i=1}^{k} \omega_i$$

$$P(Moderate) = [1 - P(High)] \times \sum_{j=1}^{p} \omega_j$$

$$P(Low) = 1 - [P(High) + P(Moderate)]$$

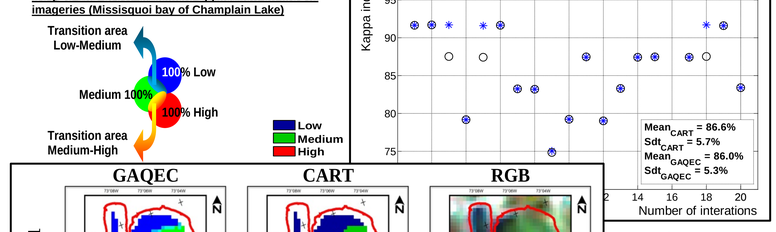
Abscissas and weights for the standard normal distribution

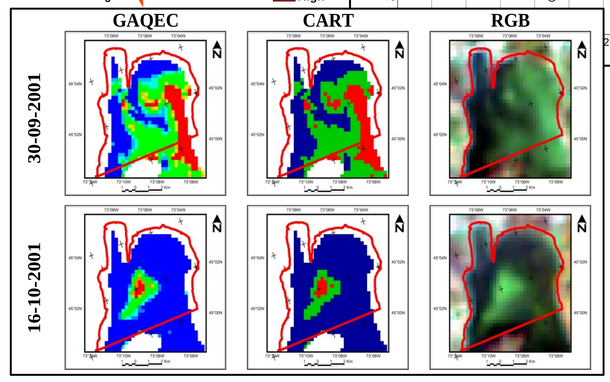
n	z_i	ω_i
0	0	1
2	$-1, +1$	$\frac{1}{2}, \frac{1}{2}$
3	$-\sqrt{3}, 0, +\sqrt{3}$	$\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$



Gaussian quadrature ensemble classifier

Results





CART and GUQEC accuracy assessment and robustness analysis.

Conclusions

This study was conducted to develop a classifier based on ensemble method using MODIS images downscaled to 250 m spatial resolution and to test its performance in terms of accuracy and robustness. Even the developed approach is relatively less accurate (Kappa index = 86% versus 86.6%) then a standard classifier, its robustness was higher. A further asset of the developed method is its ability to map transition areas between classes. by the present work, it was possible to highlight the potential of remote sensing data to monitor HAB in freshwaters with an acceptable error rate.

Acknowledgment

We thank Sylvie Blais and Nathalie Bourbonnais from the MDDELCC for providing the in situ database, The FQRNT for funding the research program, the NASA/GSFC for making MODIS data available free of charge, and Rasim Latifovic and Alexander P. Trishchenko from the Canadian Centre for Remote Sensing for providing the tool for downscaling the spatial resolution of MODIS bands and the algorithm for atmospheric correction.

[1]. Eija, R.; Rvi, R.O.; Seija, H.L.; Juha-Markku, L.N.; Mika, R. Effect of sampling frequency on detection of natural variability in phytoplankton: Unattended high-frequency measurements on board ferries in the baltic sea. J. Marine Sci. 1998, 55, 8.
[2]. Mackie, T.N. Public Health Surveillance of Toxic Cyanobacteria in Freshwater Systems Using Remote Detection Methods. Ph.D. Thesis, University of California, Berkeley, CA, USA, 2010.
[3]. Li, R.-H. and G.G. Belford, Instability of decision tree classification algorithms, in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. 2002, ACM: Edmonton, Alberta, Canada. p. 570-575.
[4]. Polikar, R., Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 2006. 6(3): p. 21-45.
[5]. Tørvi, H. and T. Hertzberg, Estimation of uncertainty in dynamic simulation results. Computers & Chemical Engineering, 1997. 21, Supplement(0): p. S181-S185.
[6]. http://ladsweb.nascom.nasa.gov/data/search.html
[7]. Trishchenko, A.P.; Luo, Y.; Khlopenkov, K.V. A method for downscaling MODIS land channels to 250 m spatial resolution using adaptive regression and normalization. Proc. SPIE 2006, 6366, 36607–36607.

INRS Université d'avant-garde

GRIL Groupe de recherche interuniversitaire en limnologie et en environnement aquatique

CENTRE D'ÉTUDES NORDIQUES CEN Centre for Northern Studies

Fonds de recherche sur la nature et les technologies Québec